

# Maestría en Tecnologías de Información para los Negocios

LGAC Inteligencia de Negocios Productividad (2018- 2020)

## Resumen

	Tipo	N	Participación Estudiantes (N)	Participación Estudiantes (%)
01	Revistas	36	13	36%
02	Libros	5	0	0%
03	Capítulos de libro	6	1	17%
04	Artículos de Congresos (arbitrados)	14	0	0%
05	Talleres, consorcios, simposios	0	0	0%
06	Registros de Propiedad intelectual (IMPI o INDAUTOR)	2	0	0%
	<b>TOTAL</b>	<b>63</b>	<b>14</b>	<b>22%</b>

**\*\* Participación estudiantes se refiere a los artículos en los que participan estudiantes adscritos a la Maestría en Tecnologías de Información para los Negocios**

Profesores que conforman la LGAC de Inteligencia de Negocios

- Dra. Cynthia Beatriz Pérez Castro (SNI Nivel C)
- Dr. Luis Felipe Rodríguez Torres (SNI 1)
- Dr. Luis Adrián Castro Quiroa (SNI 1)
- Dr. Guillermo Mario Arturo Salazar Lugo

# Hacia la construcción de una herramienta de visualización de datos en centros geriátricos

## *Towards the construction of a data visualization tool for geriatric centers*

Amanda Tapia Zazueta  
Instituto Tecnológico de  
Sonora (ITSON)  
Ciudad. Obregón, México  
amandatapiaz@hotmail.com

Luis A. Castro  
Instituto Tecnológico de  
Sonora (ITSON)  
Ciudad. Obregón, México  
luis.castro@acm.org

Cynthia B. Pérez  
Instituto Tecnológico de  
Sonora (ITSON)  
Ciudad. Obregón, México  
cynthia.perez@itson.edu.mx

Jessica Beltrán Márquez  
CONACYT-CITEDI  
Tijuana, México  
jbeltran@citedi.mx

**Resumen** — En la edad adulta pueden surgir enfermedades neorogenerativas que provocan que quienes las padecen dependan de personal capacitado en estancias geriátricas. La principal contribución del presente trabajo es proponer un modelo conceptual que busca delinear las implicaciones de diseño de una herramienta de visualización de datos para estancias al cuidado del adulto mayor. El modelo propuesto se crea a través de teoría fundamentada. Los resultados obtenidos en el análisis cualitativo enfatizan el papel primordial del personal para identificar los factores contextuales involucrados en el proceso de cuidado y monitoreo de adultos mayores en estancias geriátricas.

**Palabras Clave** – Herramienta de visualización de datos; cuidado del adulto mayor; visualización de información.

**Abstract** — Older adults can suffer from neogenerative diseases, which makes them dependent on trained personnel in geriatric centers. The main contribution of this paper is a conceptual model that seeks to delineate the implications for the design of a data visualization tool for older adults's care centers. The proposed model is created through grounded theory. The results obtained in the qualitative analysis emphasize the primary role of staff in identifying the contextual factors involved in the care and monitoring process of older adults in geriatric centers.

**Keywords** - Data visualization tool; care of the older adults; information visualization.

### I. INTRODUCCIÓN

Los adultos mayores que sufren de una enfermedad crónica o demencia podrían no ser capaces de realizar sus actividades diarias por sí mismos y necesitan asistencia de un cuidador [1]. Las organizaciones como las estancias de adultos mayores, brindan servicios de cuidado y atención a enfermedades neurodegenerativas como Alzheimer o Parkinson, y enfermedades propias de la edad como diabetes, hipertensión, problemas circulatorios, entre otros [2]. Debido a esto, algunos miembros de la familia deciden ingresar a los adultos mayores

para que puedan ser atendidos por personal capacitado, cuidadores, enfermeras y/o médicos [2].

Un procedimiento común utilizado en estancias al cuidado del adulto mayor, requiere que el personal realice anotaciones manuales de las actividades realizadas por los residentes, la información anotada es útil ya que permite tomar conciencia sobre el estado, el bienestar y contar con un expediente completo de los adultos mayores [3]. Tal procedimiento puede ocasionar pérdida de información debido a los cambios de turno, rotación de personal, estrés por la carga de trabajo, necesidades de atención y registro manuales [3]. Más aun, es común que el número de adultos mayores sobrepase al número de cuidadores, es decir que un cuidador está a cargo de varios residentes. Los cuidadores en estancias geriátricas registran constantemente grandes cantidades de datos sobre las actividades de los residentes. Si bien, estos datos son valiosos, por lo general, no se le da seguimiento a las anotaciones debido a las rotaciones o porque al personal se le olvida consultar de nuevo [4], ocasionando que sea difícil tomar decisiones durante la atención y cuidado al adulto mayor.

Para llevar a cabo una gestión adecuada de la información, las herramientas de visualización podrían ayudar en el proceso de toma de decisiones en el ámbito del cuidado de la salud. Dichas herramientas proporcionan una comprensión más clara de los datos de los pacientes [5] a través de la identificación de necesidades de usuarios específicos. Por ejemplo, al representar información relevantes para los médicos [6], los pacientes [5], las enfermeras [7], miembros de la familia [8] o incluso toda una comunidad [9].

El presente trabajo tiene como objetivo de investigación, mostrar un modelo conceptual que sirva como marco de referencia para el diseño de una herramienta de visualización de datos centrado en el cuidador. Dicha herramienta debe ser capaz de mostrar datos registrados de los residentes, para que los cuidadores puedan tener conciencia sobre el bienestar general de los adultos mayores, permitiendo interpretar fácilmente los datos para el apoyo al proceso de toma de decisiones y obtención de conocimiento.

## II. TRABAJO RELACIONADO

Actualmente, gran parte del trabajo en el cual se encuentra el personal en las empresas o instituciones del sector salud, ha implicado un uso extenso de datos generados desde múltiples fuentes y/o información utilizada para dar sentido al apoyo a la toma de decisiones [10]. En el campo de la salud, los datos pueden provenir de las lecturas de dispositivos médicos que detectan los signos vitales [7], a partir de anotaciones manuales sobre las actividades diarias de los pacientes [11], y pruebas de laboratorio [12]. Las prácticas clínicas se han centrado cada vez más en el estudio de análisis de los datos de los pacientes para el desarrollo de sistemas de apoyo a la toma de decisiones clínicas con información digitalizada. Por lo tanto, existe una necesidad creciente de almacenar y organizar todos los datos de pacientes para contribuir a la evaluación constante de casos o estudios clínicos [13]. Una preocupación común ha sido cómo presentar, diseñar e implementar información eficaz pertinente a los usuarios de forma que las tareas y las actividades se apoyen eficazmente [10]. Por ejemplo, un trabajo ha determinado factores críticos de diseño para las tecnologías de información en apoyo a la gestión de diabetes tipo 1 [14]. Los descubrimientos llevan al desarrollo de consideraciones de diseño que proporcionan una visión sobre el desarrollo de tecnologías de la información y comunicaciones (TIC) con fines de autogestión de los adolescentes y compromiso informal de los cuidadores [14]. Dentro de las aplicaciones y plataformas desarrolladas en el área de cuidado de la salud se ha utilizado la integración de conocimiento basado en e-salud [15]. Este concepto alude a la práctica de cuidados sanitarios apoyado en las TICs, cuya función es la integración, la extracción y análisis de los conocimientos de diferentes fuentes en conjunto con la interacción con los participantes y el personal experto. Lo anterior ha mejorado el acceso, la eficiencia, la calidad de los procesos tanto clínicos como empresariales, y a su vez ha generado un impacto en la documentación, las recomendaciones clínicas, la relación del paciente-médico, paciente-cuidador, cuidador-cuidador, la comunicación médica, planeación, evaluación de datos así como la toma de decisiones clínicas [15].

## III. METODOLOGÍA

Una de las motivaciones de este trabajo es investigar cuales son las implicaciones de diseño que requiere una herramienta de visualización de datos en estancias al cuidado del adulto mayor. En ese sentido, el presente trabajo utiliza métodos cualitativos de investigación para responder a lo siguiente.

### A. Preguntas de Investigación

La pregunta principal que busca responder este trabajo es: ¿Cuáles son las directrices de diseño para una herramienta de visualización de datos para la toma de decisiones en estancias geriátricas? De igual forma, se tienen algunas sub-preguntas:

- ¿Cuál es el requerimiento principal a considerar en relación con el personal y la gestión del proceso de atención?
- ¿Cuál es el requerimiento principal a considerar en relación con el personal y la información?
- ¿Cuál es el requerimiento principal a considerar en relación con el personal y el proceso de comunicación?

- ¿Cuál es el requerimiento principal a considerar en relación con el personal y la gestión de la calidad integral del servicio?

### B. Informantes

Los informantes (I) fueron cuidadores profesionales tanto (M=1, F=9), involucrados en el proceso de atención y cuidado del adulto mayor en un contexto multidisciplinario que incluyó a enfermeros, psicólogos, fisioterapeutas, nutriólogos, y cuidadores independientes con experiencia en hospitales y servicio de atención a domicilio. El reclutamiento se realizó a través de citas con personal de una estancia geriátrica en el noroeste de México, así como cuidadores que trabajan de manera independiente.

### C. Recolección de datos

La recolección de los datos se realizó a través de entrevistas semi-estructuradas. El protocolo de entrevista consistió de 23 preguntas abiertas, con una duración promedio de 45 min. cada una. Todas las entrevistas fueron realizadas con el consentimiento escrito de los participantes, grabadas en audio, y transcritas textualmente.

### D. Análisis de datos

El análisis de las transcripciones se llevó a cabo utilizando Teoría Fundamentada (TF) [16]. La TF es un método que permite analizar de manera sistemática datos cualitativos, que permite descubrir patrones que emergen de los datos, y fundamentan para una teoría emergente. TF consiste en tres etapas en el proceso de codificación: abierta, codificación axial, y selectiva. En la codificación abierta se crean códigos a través del análisis de los datos recolectados, oración por oración. La codificación axial, consiste en descomponer, examinar, comparar, conceptualizar y categorizar datos. Finalmente, la codificación selectiva que permite la identificación de categorías base, estableciendo relaciones y validándolas a través de los patrones que se relacionan para la definición de nuevas categorías, conexión entre conceptos y su significado para la integración de modelos teóricos [17]. Lo anterior permitirá definir las principales directrices de diseño para la construcción de una herramienta de visualización de datos durante el desarrollo de la investigación propuesta.

## IV. RESULTADOS

El análisis arrojó cuatro factores contextuales (Figura 1) identificados en las aportaciones de los informantes. Los factores contextuales transmiten las necesidades de los cuidadores y gestores mostrando una conexión entre el personal y los elementos que forman parte del proceso de atención y cuidado de adultos mayores en estancias geriátricas, así como la identificación de la principal implicación de diseño definido como requerimientos de información (Tabla I).

### A. Factores contextuales

#### 1) Gestión del proceso de atención

Este factor se refiere a la relación del personal en el proceso diario del servicio de atención, cuyo proceso consiste en atender las necesidades particulares de los adultos mayores con problemas neurogenerativos por medio de actividades de apoyo y cuidado personalizado, de acuerdo con lo comentado por el informante I-3. Otra de los informantes (I-1) comenta sobre la

periodicidad de las valoraciones que conlleva a una serie de anotaciones ya sea de manera habitual, o de manera eventual necesarias para llevar a cabo dicha gestión: “Hacemos 3 valoraciones cuatrimestrales durante el año, el mes pasado ya se hizo como siempre, pero este año tenemos pensado renovar y mejorar un poco, urge definir las conductas, qué tipo de conductas se están viendo en el adulto mayor, por ejemplo somnolencia. Llevar un record de un proceso estructurado, llevar algo en el cuatrimestre donde nos indique que comportamientos está teniendo la persona, porque si ya revisamos el medicamento, damos por hecho que puede ser esto, pero no llevamos el registro” (I-1). Con lo anterior se ha identificado que las valoraciones permiten generar evaluaciones continuamente a través de ciertos indicadores de salud con el registro de actividades para monitorear el estado de los adultos mayores y su vez mejorar la gestión del proceso de atención.

## 2) Gestión de información del personal

Este proceso incluye la relación del personal con el manejo de la información. Se ha analizado que para llevar a cabo la gestión de la información en el proceso de atención de forma adecuada, dependerá de la información que se comparte, de acuerdo con lo expuesto por el I-6. La información que maneja el personal es sumamente importante para cada uno de las áreas del proceso de atención del adulto mayor ya que sin el conocimiento de datos revisado en su historial se complicaría las funciones diarias entre los cuidadores y apoyos externos para el personal. Por ejemplo I-2 comenta: “Cuando yo veo que hay una conducta o algo en el adulto mayor, reviso si estuvo con dolor, somnoliento o indispuesto, voy al cuaderno donde están las anotaciones, y veo que hay una caída, se fracturó tal fecha, llegó a la estancia de esta forma, y hasta que el doctor de indicaciones, realizo mis actividades al adulto mayor señaladas en un cuaderno donde está ahí registrado”. Por lo anterior se ha identificado el recurso de bitácoras utilizados para compartir datos específicos registrados por parte del familiar como apoyo a la toma de decisiones para la realización de funciones de los cuidadores establecidas durante el día, sin embargo no ha sido suficiente para la gestión de información del personal.

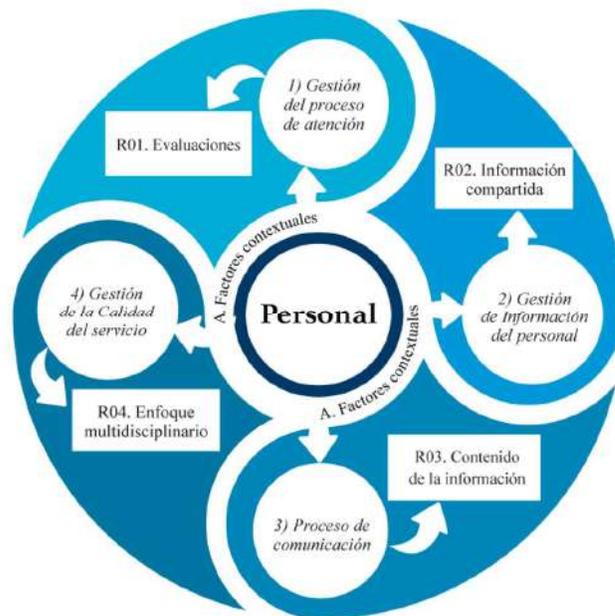
## 3) Proceso de comunicación

Este factor incluye la relación del personal con el proceso de comunicación entre cuidadores e interacción entre el personal y el familiar del adulto mayor. La información relevante para llevar a cabo una comunicación asertiva en las estancias geriátricas, dependerá del contenido de información presente en las bitácoras que es revisado constantemente por el personal encargado o asignado para el desarrollo de sus funciones. Los familiares que asumen la labor de tutor o encargado de cuidar a su adulto mayor en casa, muchas veces, desconocen los acontecimientos y cambios de la enfermedad necesarios a observar durante el día para el registro diario en bitácoras, por lo que a menudo los datos se muestran incompletos. Por otra parte, el personal presenta tareas simultáneas durante el proceso de registro, ocasionando descuidos y omisiones en el contenido de la información proporcionada por el familiar. Lo anterior ha generado retrasos o pérdida de tiempo durante el proceso de atención con el adulto mayor de acuerdo a lo comentado por el I-4: “Hay un

detalle, a veces la información debe de llegar y es donde yo siento un poco de falta de organización, porque tengo que ir a preguntar qué te dijeron, qué pasó, checar bitácoras, y veo que no ha dormido. Es ahí donde me frustro, o hasta cierta molestia porque la información del adulto es importante, ya que es una guía de trabajo para mí”.

## 4) Gestión de la calidad del servicio

Incluye la relación del personal con la calidad del servicio. Este factor consiste en el conjunto de propiedades y características del servicio de atención para satisfacer las necesidades de los adultos mayores en relación a las iniciativas y objetivos de las estancias geriátricas. Los I-5 y I7 han expuesto la importancia de contar con un equipo de trabajo con carácter multidisciplinario para brindar una calidad de servicio de manera integral, es decir, que responda a la necesidad de mejorar los procedimientos establecidos en las estancias geriátricas. Por ejemplo el I-5 menciona que el fisioterapeuta maneja información específica para sus procesos, pero también tiene que contar con información de las otras áreas como la de psicología o enfermería. Por otro lado, I-7 comenta: “Nosotros tenemos la preparación o enfermeras de cómo cuidar a un paciente, independientemente del tipo de paciente, de patología o enfermedad que tenga, capacitación para dar cuidados independientemente que sea en casa, en un hospital o asilos en conjunto con conocimientos interdisciplinario como la relación con médicos, la trabajadora social, una tanatología, químico, radiólogo, psicólogo igual que en un hospital, tenemos que relacionarnos con proveedores de insumos de alimentos, medicamentos, materiales, todo eso”. Esto ha originado consultar y revisar información complementaria entre el equipo de trabajo para



solucionar complicaciones durante el proceso de atención y cuidado del adulto mayor, con el apoyo de un enfoque multidisciplinario.

Figura 1. Modelo de la relación entre factores contextuales e implicaciones para el diseño.

## B. Consideraciones para el diseño de la herramienta

De acuerdo con la identificación de los factores contextuales enmarcados con anterioridad, se han descubierto los requerimientos de información (Tabla I) como principal directriz de diseño a considerar para definir los elementos que conformaran una herramienta de visualización de datos para el apoyo a la toma de decisiones en estancias geriátricas.

TABLA I. REQUERIMIENTO DE INFORMACIÓN

Req.	Descripción	Definición
R01	Evaluaciones	Las evaluaciones determinan en qué medida se han logrado los objetivos en el cuidado del adulto mayor. Los materiales de apoyo para realizar las evaluaciones en estancias geriátricas son: valoraciones, diagnósticos clínicos, revisiones periódicas e informes cuatrimestrales.
R02	Información compartida	Corresponde al análisis de datos específicos concentrados en bitácoras para todas las partes involucradas con información de apoyo durante el proceso de toma de decisiones. La información que se comparte son los registros de datos, informes, historial del paciente y anotaciones manuales.
R03	Contenido de la información	El contenido de la información en bitácoras es parte fundamental para el apoyo en la toma de decisiones de los cuidadores, en ocasiones se muestra incompleta y omitida. Por el contrario los expedientes de cada adulto mayor presenta exceso de información en archivos físicos que muestra datos generales y específicos revisado constantemente por el personal encargado.
R04	Enfoque multidisciplinario	Se refiere a los conocimientos de disciplinas relacionadas al área geriátrica para cubrir las necesidades de interacción dentro del equipo de trabajo para la solución de problemas de los adultos mayores que permita mejorar constantemente la gestión de la calidad del servicio.

## V. CONCLUSIONES

Este estudio busca avanzar en el conocimiento de elementos de diseño de una herramienta de visualización de datos para el apoyo a la toma de decisiones de los gestores y trabajadores en estancias geriátricas, que apoye a la integración de estrategias y gestión empresarial en proyectos de diseño centrados en el usuario.

El trabajo futuro incluye ampliar y transformar el modelo propuesto, con un nivel más profundo de análisis. La relación de los factores contextuales planteados en este trabajo, se retomará como base para el descubrimiento de nuevas implicaciones de diseño. Una vez determinado el modelo final será sujeto a iteraciones y verificación de los resultados del estudio con el personal en estancias geriátricas que manejan registro de anotaciones manuales. Lo anterior permitirá definir las directrices principales a seguir en el proceso de diseño para iniciar el desarrollo de prototipo de una herramienta de visualización de datos para el apoyo de decisiones, y su posterior evaluación con usuarios finales.

## AGRADECIMIENTOS

Agradecemos a las personas entrevistadas con experiencia en el área geriátrica de la región por el tiempo brindado. También,

agradecimientos especiales al Dr. Cayetano Cruz García, Dr. Francisco Chávez de la O y Dr. Francisco Fernández de la Vega por sus comentarios y recomendaciones.

## REFERENCIAS

- [1] S. Katz. "Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living". *Journal of the American Geriatrics Society*, 31(12), 721-727, 1983.
- [2] C. Fernández, J. M. García, J. Vicente, J. C. Naranjo, M. Robles, J. M. Benedi, and V. Traver. "Behaviour patterns detection for persuasive design in Nursing Homes to help dementia patients". In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 6413-6417). IEEE, August 2011.
- [3] V. Soto-Mendoza, J. Beltrán, E. Chávez, J. Hernández, and J. A. García-Macias. "Abnormal behavioral patterns detection from activity records of institutionalized older adults". pages 119-131, 2015.
- [4] V. Soto-Mendoza and J. A. García-Macias. "Lessons learned from a long-running assistive system for geriatric care". In International Workshop on Ambient Assisted Living, pages 83-90. Springer, 2014.
- [5] T. Le, B. Reeder, H. Thompson, G. Demiris, et al. "Health providers perceptions of novel approaches to visualizing integrated health information". *Methods Inf Med*, 52(3):250-258, 2013.
- [6] S. Amri, H. Ltifi, and M. B. Ayed. "Towards an intelligent evaluation method of medical data visualizations. In 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA), pages 673-678. IEEE, 2015.
- [7] M. Gorges, D. R. Westenskow, and B. A. Markewitz. "Evaluation of an integrated intensive care unit monitoring display by critical care fellow physicians. *Journal of clinical monitoring and computing*", 26(6):429-436, 2012.
- [8] C. Bock, G. Demiris, Y. Choi, T. Le, H. J. Thompson, A. Samuel, and D. Huang. "Engaging older adults in the visualization of sensor data facilitated by an open platform for connected devices". *Technology and Health Care*, (Preprint):1-10, 2016.
- [9] Be Well Placer community dashboard. <http://www.placerdashboard.org/index.php?module=indicators&controller=index note> = Accessed: 2016-09-06.
- [10] K. Sedig, and P. Parsons. "Interaction design for complex cognitive activities with visual representations: A pattern-based approach". 2013.
- [11] V. Soto-Mendoza, J. A. García-Macias, E. Chávez, A. I. Martínez-García, J. Favela, P. Serrano-Alvarado, and M. R. Z. Rojas. "Design of a predictive scheduling system to improve assisted living services for elders". *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(4):53, 2015.
- [12] P. C. Tang, J. S. Ash, D. W. Bates, J. M. Overhage, and D. Z. Sands. "Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption". *Journal of the American Medical Association*, 13(2):121-126, 2006.
- [13] S. Zillner, T. Hauer, D. Rogulin, A. Tsybmal, M. Huber and T. Solomonides. "Semantic visualization of patient information". In *Computer-Based Medical Systems, 2008. CBMS'08. 21st IEEE International Symposium on* (pp. 296-301). IEEE, June, 2008.
- [14] J. Bedrossian, L. Kerr, L. Robertson, A. Stewart, J. Suits, S. Patek, and R. Valdez. "Critical design factors for information technology supporting type 1 diabetes management". In *Systems and Information Engineering Design Symposium (SIEDS)*, (pp. 261-266). IEEE, April 2016.
- [15] S. Nasiri and M. Fathi. "Toward an integrated e-health based on acquired healthcare knowledge". In *Biomedical Engineering (MECBME), 2014 Middle East Conference on* (pp. 301-304). IEEE, February 2014.
- [16] J. M. Corbin and A. Strauss. "Grounded theory research: Procedures, canons, and evaluative criteria". *Qualitative sociology*, 1990, vol. 13, no 1, p. 3-21.
- [17] N.R. Pandit. "The creation of theory: A recent application of the grounded theory method". *The qualitative report*, 2(4), 1-15, 1996.

# Towards Human-Centric Interfaces for Decision Making Support in Geriatric Centers

Luis A. Castro<sup>1</sup>, Amanda Tapia<sup>1</sup>, Cynthia B. Perez<sup>1</sup>, Jessica Beltrán-Márquez<sup>2</sup>

<sup>1</sup> Sonora Institute of Technology (ITSON), Ciudad Obregon, Mexico

<sup>2</sup> CONACYT-CITEDI-IPN, Tijuana, Mexico

luis.castro@acm.org, amanda.tapiaz7@gmail.com,

cynthia.perez@itson.edu.mx, jbeltran@citedi.mx

**Abstract.** Older adults can suffer from neurodegenerative diseases, which can make them dependent on family members or trained personnel in geriatric centers. In this work, we aimed at identifying the main information elements needed for decision making in geriatric centers. We present the results of a qualitative study using grounded theory as the analytic lens for a contextual study carried out with workers of a geriatric center. The main contribution of this paper is a set of information elements that are needed for different types of decisions taken at a geriatric center. The results of this work can be used for designing better interfaces for the decision making support in geriatric centers.

**Keywords:** human-centric interface, older adults, caregivers, decision making support

## 1 Introduction

Older adults who suffer from chronic illness or neurodegenerative conditions may not be able to carry out several activities of daily life (ADL). Many of them are assisted by professional caregivers or family members who help them thrive. Although family members often volunteer as caregivers, there are institutions that can provide specialized care by physicians, psychologists, and nurses. These institutions are often residences, where older adults live, or community-dwelling adults who spend a few hours a day in the institution or geriatric center (GC).

In many developing regions, including Mexico, geriatric centers do not have much technology providing support to their internal processes. In such institutions in Mexico, instead of using technology, it is commonplace to take notes and jot down several observed behaviors of older adults or assessments. These data happens to be essential for assessing the progress of the older adult. However, although important, these notes can often become a liability for the institution due to shifts, personnel turnover, workload stress, and health care [1]. That is, in the case of a critical event, GCs do not have enough information to prove that the provided care was appropriate, let alone to determine that the decisions made were the right calls. Also, in many cases the number of older adults is greater than the number of caregivers, leading to constant overwork.

Then, GC's caregivers constantly maintain record of activities carried out by residents, since it is one of the main ways of keeping track of events. Even if these data are a valuable resource, generally, the follow-up of these notes is complex due to personal turnover or shifts. This makes it difficult for clinicians to make decisions regarding attention and care [1].

In this regard, human-centric interfaces can be of great help as a support in the decision making process pertaining to clinical or non-clinical decisions. Decision making can be regarded as a cognitive process in which an individual selects among a set of alternatives, yielding to reinforcing a belief or a course of action. In some occupations, such as in health care, decision making is critical, and are subject to strict regulations and accountability. In fact, a set of medical guidelines have been created as a way to guide decisions and criteria during medical practice in specific areas of healthcare (Visit <https://www.guideline.gov/>).

In this context, interfaces that can provide support in decision making processes can be of great help for healthcare professionals. Those said tools can turn instrumental in understanding patient data [2]. In particular, they can be fundamental for presenting relevant information to doctors, nurses, family members, or a community [2, 3, 4, 5, 6]. In this work, we aim at identifying information elements that can be useful for decision making in GCs. In particular, our ultimate goal is crafting design guidelines that can be useful for developers interested in providing human-centric interfaces. For this, we carried out a qualitative study with professionals who work regularly with older adults. Our results provide a framework for understanding what information elements are taken into account when making clinical and non-clinical decisions in a GC.

## **2 Related Work**

In the medical field, an extensive use of data generated from multiple sources is becoming commonplace, which is used to support, and make sense, of the decision making processes [7]. In particular, in the healthcare domain, these sources can be readings from medical devices [4], manual annotations by medical staff regarding the patients [8], lab results [9], or patients' behavioral data provided by self-report or family members.

Clinical practices have increasingly focused on analyzing digitalized patient data for developing decision support systems [10, 11]. There is a growing need for storing and organizing patient data to contribute with the constant assessment of clinical cases or studies [12]. A valid concern related to how to present, design, or implement information in an effective way, so that the user's activities can be effectively supported [7]. For example, in [13], the authors present the critical factors for designing information technologies and communication technologies (ICT) to support Type 1 diabetes management. The findings pointed toward developing ICTs' design insights for teenagers' auto management and commitment of informal caregivers [13]. Finally, some other works have been developed in which they integrate healthcare knowledge [14]. As such, decision support systems in health care are not new [10, 15, 16], how-

ever, previous works usually place particular emphasis on clinical decision making for physicians.

Although the literature for decision support systems in healthcare is extensive, to the best of our knowledge, this is one of the first works aimed at highlighting the information elements that are needed for decision making at a geriatric center, particularly in Mexico. These information elements can be the building blocks of appropriate human-centric interfaces that be used for supporting clinical or non-clinical decision making in geriatric centers. In particular, we present a study involving a geriatric center where technology uptake has been historically slow.

### **3 Methods**

In this section, we present the methods related to the qualitative study as well as a description of the GC and the participants:

#### **The geriatric center (GC) and informants.**

The GC where we mainly carried out this work is a non-for-profit, privately-funded geriatric center for patients diagnosed with dementia. It counts with 14 full-time employees, which includes 9 clinicians, 2 kitchen workers, 1 manager, 1 financial administrator, and 1 janitress. In addition, there are a number of volunteers who are usually family members or university students.

The GC provides attention to approximately 30-35 patients per month. Patients who apply for admission at the center undergo a pre-admission screening and baseline testing by the clinicians for eligibility. Although the center is privately funded, it charges family members an undisclosed fee for its services.

The GC is located in a mid-size city of approximately 400,000 inhabitants in Northwest Mexico. Physically, the premises were originally a social interest house, but it now includes the following spaces: a gym, a workout area, a patio for home-grown vegetables, rest areas, appropriate toilets, an infirmary, a kitchen, administrative offices, and a meeting room. It is fully gated, with a double check security point for safety, where visitors must register, and where family members meet and greet with the GC staff before leaving their family members for the day.

Our informants (see Table 1) work regularly with older adults as part of their professional practice. They were mainly recruited from the aforementioned GC, and professionals who work at hospitals, who also volunteer at older adult residences.

#### **Data collection and analysis.**

We carried out 10 semi-structured interviews. Interviews lasted in average 45 min each. All interviews were audio recorded and verbatim transcribed for their posterior analysis. The interview protocol consisted of 23 questions and involved the topics related to the activities carried out by adults, information management practices, health care, and decision making in the institution. For this work, only 7 interviews

were considered since 3 of them (I-8, I-9, I-10) did not provide useful information regarding decision making in the GC.

Our analysis aimed at uncovering current practices in the geriatric center in terms of decision making, which includes processes, actors, information needs, and timing. The findings of this work will be ultimately used for designing appropriate human-centric interfaces for decision making support in geriatric centers. For the analysis, we used Grounded Theory [17] for the data analysis. Grounded theory is a technique for systematic analysis of qualitative data, to uncover patterns and topics that are recurrent throughout the data. The coding process was carried out as follows: One author coded two interviews (open coding). Then, those codes were refined by two of the authors, and posteriorly used to code the rest of the interviews until saturation was reached. Then, axial and selective coding was iteratively carried out by two authors.

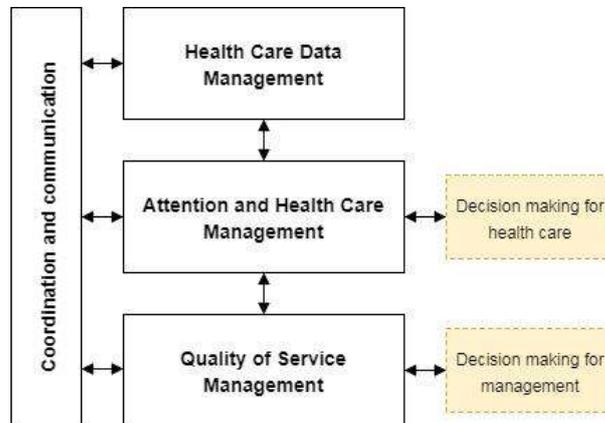
**Table 1.** Demographics of our informants

ID	Role	Gender	Age	Exp. (years)	Main work
I-1	Psychologist	F	30	4	Geriatric center
I-2	Physiotherapist	F	45	9	Geriatric center
I-3	Nurse	F	55	25	Hospital
I-4	Nutritionist	F	22	1	Geriatric center
I-5	Nurse	F	23	0.5	Geriatric center
I-6	Psychologist	M	24	3	Geriatric center
I-7	Nurse	F	26	2	Hospital
I-8	Caregiver	F	46	19	Residence
I-9	Caregiver	F	35	15	Residence
I-10	Manager	F	40	15	Hospital

#### 4 Results: Decision Making in the GC

After a result of the analysis, we identified four elements that are the building blocks of the geriatric center under study (see Figure 1):

1. **Quality of service management.** This element involves the development and refinement of business processes related to maintaining quality of service. This is pivotal to the center since it is a non-for-profit, privately-funded institution.
2. **Attention and health care management.** This element involves all the processes that are related to attention and health care of patients in the center. They are carried out by professionals of health care.
3. **Health care data management.** Policies and practices related management of data related to patients and their clinical assessments.
4. **Coordination and communication process.** Policies, practices, and processes related to maintain coordination among different shifts or roles of the GC's personnel, which have a direct influence on the patient's care.

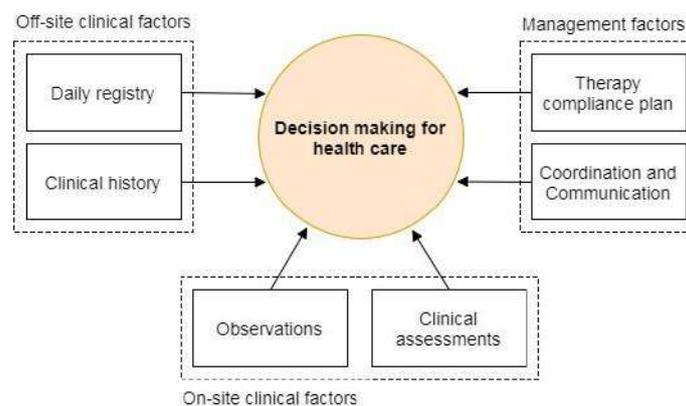


**Fig. 1.** Identified GC's elements and decision making

Within those elements, there were identified two main set of decision making processes that have an influence on the center's operations, and which is precisely the focus of this work: a) decision making in attention and health care, and b) decision making for management. We next describe these two processes, which are related to how data and information are created, collected, managed, and ultimately used for decision making.

#### 4.1 Decision making for health care

This process, which is part of the Attention and Health Care Management block (Figure 1), includes a set of factors that are taken into account when making decisions of clinical nature. There are primarily five factors that have an influence in clinical decision making in the GC, which are fundamentally records from clinical assessments and observations as well as therapy planning about patients.



**Fig. 2.** Factors influencing decision making for health care in the GC

Figure 2 shows the factors that influence decision making during attention and care in the GC under study. As seen, some of them relate to clinical data and others to observations or information of other kind that is needed for decision making. In fact, those factors can be grouped into three main classes: on-site clinical factors, off-site clinical factors, and management factors. We next explain each of them:

- **Off-site clinical factors:** These may refer to factors that pertain to health care, although they are not necessarily directly in charge of the GC, but they certainly need to be made aware of. These usually involve changes made by a physician or health care institutes such as changes in medication, diagnoses, and the like. In addition, it also involves incidents that happen when family members are in charge of patients such as falls, sleep deprivation, food intake, and so forth.
- **Daily registry:** This refers to a registry (i.e., logbook) that is updated every day as the patient arrives to the GC. It may be information provided by the family members such as sleep deprivation, or incidents/events that happened while at home. In addition, it also involves annotations by the physical therapists who record relevant information regarding the physical shape of the patient (e.g., lack of strength). In this regard, I-3 commented: *“Anything that happens such as a fall, [bone] fracture, sprain. In what refers to my physical area [her area of expertise], if I know that [the patient] fell, then I asked ‘what did the doctor say?’, about the prescription and indications. I make a copy and attach it.”* Here, I-3 is commenting on the importance of what the doctor outside the GC may say about the patient. Since the GC is not a registered primary care institution, the GC personnel must rely on prescriptions and recommendations of physicians. This is one of the reasons why the daily registry is an important element of information that all GC’s personnel must be aware of.
- **Clinical history:** This refers to the clinical history of the patients. It is usually collected when patients are first admitted to the GC, but they are essential for decision making of a clinical nature. It is usually provided by family members and maintained by clinicians affiliated to private or public healthcare institutions. In this regard, I-3 also commented: *“At the IMSS [Mexican Social Security Institute] or ISSSTESON [Institute for Social Security and Services for State Workers in the State of Sonora], they have their clinical record, their [clinical] history. They are pathological and non-pathological background. It’s like having all, in general, [like] name, type of dwelling. It’s a clinical history, but also their habitat, number of children, siblings, how their houses are built, the types of materials [with which their houses are built], if it has roof, floors, water service, electricity, sewage, all that. It provides a general picture of that history, and the pathological ones, [they] give an idea of the types of environmental effects. Also, [it provides an idea of] what could have triggered the disease. For instance, the dietary habits, physical activity. All that help us consider certain type of risks.”* This aspect is also important to have all those information elements at hand, particularly when a decision is about to be made.
- **On-site clinical factors:** These refer to factors that pertain to the core of the activities carried out in the GC. They are carried out by health care professionals such as

physicians, nurses, physical therapists or psychologists. They can be classified into two: Observations and Clinical assessment.

- **Observations:** This refers to data that is captured by GC staff who work directly with the patients such as nurses. Some of these data involve annotating behaviors, symptoms, last prescription, and information that may lead to better clinical decision making. Finally, it also involves annotations made by the nurses such as vital signs readings. For instance, I-7 commented: *“I introduce myself and ask the family member how he or she passed the night. You also make the patient tells you about it, talk it out, while you are assessing him or her. He or she thinks you are having a conversation but you are [in fact] assessing him or her. We measure her or his vital signs like the arterial tension, cardiac frequency, respiratory frequency, we use the Evans scale for pain assessment, but that’s just the beginning. Then, we see what’s been indicated [by the physician].”* In this case, I-7 works primarily for a hospital but she is assigned to the geriatric ward. The way she ‘assesses’ her patients is not uncommon among the GC’s personnel, since many use this informal approach to make them feel more comfortable. Still, although important for the quality of care, capturing many of these observations can be challenging due to the nature of her work. Although this information element is important for better clinical assessment, it may prove challenging if data annotation or collection tools are not that adequate.
- **Clinical assessment:** This refers to regular assessment carried by clinicians in the GC to measure the progress of the patients. Therapy compliance plans are designed, and then, progress is measured on a regular basis. In this center, they are carried out every four months, and use standard clinical instruments for measurement ranging from physiological, cognitive, to social. Regarding this, I-2 stated: *“For the four-month assessment, that’s when I reflect on everything... the patient had a fall, he or she was like this, he or she had a fall two months ago. It is documented in an evaluation. That’s the way it is carried out: such a date had a fall, the indication was this, or the doctor recommended this and that.”* Although not stated by I-2, it is important to mention that in this center they, all of the areas involved, make the assessment in a 2-week period, and then get together to discuss the assessment in a multi-disciplinary fashion. According to their own accounts, this is an important practice for them, since it helps provide with the necessary information from all parties involved, in case critical decisions have to be made. This assessment is important since it is used to classify patients in terms of the stage they are in.
- **Management factors:** These refer to factors that, although important for decision making, may not directly involve activities of a clinical nature. Some of them refer to supervising the compliance of the work plans of the caregivers, so they can provide a progress report to family members. Also, it involves the coordination and communication processes among caregivers.
- **Therapy compliance plan:** These refer to the compliance plan of physical, psychological, and clinical nature. In addition, it also involves a nutrition plan. The execution of these plans is then paramount for understanding the progress of patients. It is also important to make the necessary adjustment of these plans. In

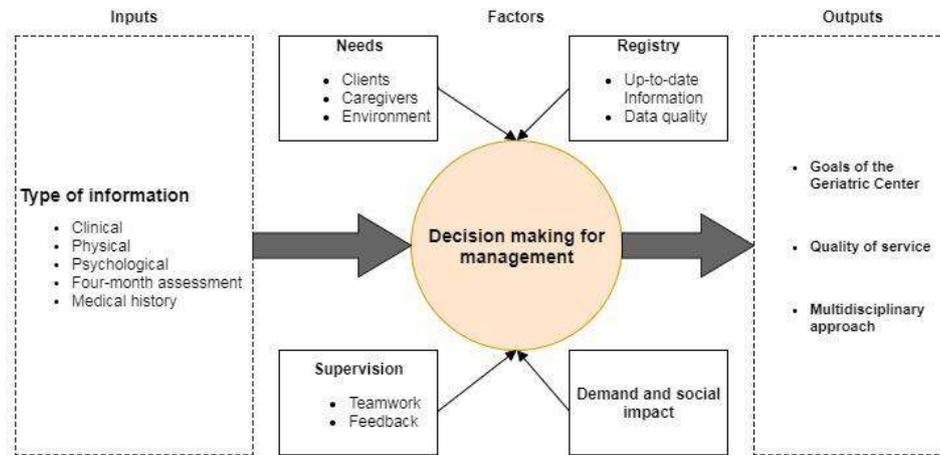
this regard, I-2 commented: *“I have a list [of patients] that are being assigned day to day. I [also] have a weekly intervention plan. The adult goes to the gym three times a week for 25 to 30 min, depending on their issue. [also] I daily have 12 to 13 personalized adults with a physical profile, and we work out on the priorities.”* This plan is a bit of information that is important for certain decisions that have to be made by the administration of the GC, particularly when it comes to evaluating performance of caregivers, as well as better understanding the workload. This aspect seems to be particularly important, as some caregivers expressed that some extra ‘hands’ are needed.

- **Coordination and communication:** This process is usually held during regular meetings between caregivers, including all the personnel related to health care and attention. In these sessions, comments are shared about the patients, so they can all be in the loop. It is during these sessions that critical decisions regarding health care are taken. It also involves communication with family members regarding the patient care as well as informal debriefing meeting where extraordinary events regarding patients are shared. Although regular meetings are crucial, informal debriefing among caregivers and family members seem to be decisive as information sharing can be limited or deficient. In this regard, I-4 expressed: *“Family members are really busy some times and do not inform us about everything. Some of them [do], but some others don’t. There are exceptions, like they change the medication, and do not inform is. That’s why one has to keep an eye on the patient, in the changes of behavior or something.”* In this sense, regular meetings among caregivers are expected, and they are professionals who know what information can be relevant to share. This may not be the case of all family members, who may disregard information that can be otherwise important.

## 4.2 Decision making for management

For management, decision making seems to be more complex since there are several factors involved. Also, there are some inputs and outputs that are expected, and they are fundamental for the institute. Decision making for management (see Figure 3) involves taking decisions that have to do with operations, planning, and financial resources for operation.

- **Inputs:** These refer to the elements needed for decision making for management. It usually involves different types of information needed for decision making.
  - **Type of information:** These usually refer to the types of information generated from the different areas of care: physical therapy, psychological therapy, and clinical care. In addition, some information is also obtained from the four-month assessment, and the medical history of the patient.



**Fig. 3.** Inputs, factors, and outputs related to decision making for management

- **Factors:** These refer to the factors that have some effect on decision making for management. Generally, they refer to needs and requirements by clients, caregivers, and the appropriate environment for geriatric care. Also, it involves a high-quality data in the registry. Third, it involves managerial practices of supervision in which they can make sure of teamwork and feedback among caregivers.
- **Needs:** They refer to the needs that are identified through frequent care of patients. Some of them refer to the needs expressed by their clients (i.e., family members), caregivers in terms of materials and tools needed for carrying out their work, and, finally, to the needs identified in the geriatric center for enabling spaces that can fulfill the needs of clients and patients. Regarding the needs of infrastructure, I-6 commented: *“Before, this place was not here. The place grew around here. The center was much smaller. I think there was demand, which is a good thing, and still is growing. It’s a bit complicated because of the space we used to have. Maybe because of the [increased] number of older adults. The personnel... we feel that we are short of two or three more personnel... to provide a better service to adults because at times I have to directly be with 13 adults. It’s a bit complicated and a lot of burden...”* As such, the GC is an organization that needs to fulfill the needs of its clients, patients, but also those of their personnel for providing adequate care. In this sense, some of the decisions to be made by the management have to take into account aspects that can ultimately relate to the kind of service that is being provided to patients.
- **Registry:** It is fundamental that the registry is up-to-date. Even more, making sure that the information is valid and complete is important for decision making. This is essential for critical decision making, as commented by I-1: *“If they change some medication, and at times it [the registry] is not updated or there is no registry because the family member [for instance] do not let us know if medication doses was increased.”* In this sense, the management has to make sure that the information they are collecting for the registry is up-to-date, correct, and

valid. Many of their managerial decisions have to be made considering this aspect, since is the basis for making the right decisions.

- **Supervision:** Supervision related to the managerial practices needed to help run the GC. This is important since decisions have to be made that contribute to achieving the goals of the GC, as well as fulfilling the needs of clients and patients. It also relates to supervising the behavior of adults on and off the GC, for which family members have an important role.
  - **Demand and social impact:** This factor helps determine the type of impact that the GC is having upon the community in which is located. In some way, as mentioned by some of our informants, this factor is reflected in the number of applications made by clients to be, as well as the fulfillment of the needs in the environment. This is important as it leads to the mission of the organization as such. Then, some decisions made by the administration go in this direction.
- **Outputs:** These refer to the outputs generated by the decision making process for management.
- **Goals:** It refers to the goals stated by the GC in terms of fulfilling the needs of adequate care for the community.
  - **Quality of service:** This refers to the quality of service that the GC aims at providing to their patients.
  - **Multidisciplinary approach:** One of the goals of the GC is offering a multidisciplinary care approach to patients, which involves multidisciplinary sessions for decision making.

## 5 Discussion

The purpose of this study was to understand the information elements needed for decision making in the GC. We also aimed at gaining insights into the design of adequate interfaces for decision making. Results from this study unraveled two broad categories for decision making –health care, and management–, which are deeply entrenched between each other. Among these, some of the emerging topics, directly related to decision making, include a) clinical information elements on and off the GC, b) therapy compliance plan, c) identification of patient needs, and those of caregivers for providing adequate care, and d) timely, quality data. Design considerations based on these topics, and directly linked to decision making are next presented:

### **Design considerations for interfaces for decision makers in geriatric centers.**

- Inclusion of primary caregiver's accounts. The interfaces must consider the inclusion of reports by family members. As continually reported by our informants, considering this information is paramount for making the right calls, or adjusting the therapy plan for the older adult due to unexpected events (i.e., falls).
- Collaborative nature of work. Understand the collaborative nature such as this one, in which managerial and clinical practices influence each other. Different points of view must come to a common understanding before taking final decisions.

- Identify a finite set of important decisions. This will enable the identification of adequate information elements that are needed for particular decisions, and the stakeholders that are involved in them. Interfaces should consider those decisions since some involve aggregated data from several months, and these types of decisions that merit careful consideration.
- Customization. The interfaces must account not only for individual differences among patients (i.e., disease stage), but also for the different roles and agenda of decision makers. In many cases, workers operate at different levels of care, in different areas of the geriatric center.
- Collective vs individual decisions. Critical decisions are made in full consensus among all the disciplines involved in the older adults care. Other decisions such as adjusting a physical therapy may be made individually.

#### **Design considerations for decision support systems in emergent regions.**

- Technology uptake low. Technology uptake such as information systems has been historically low in small GCs such as the one presented in this paper. This is particularly true of countries such as Mexico, which can derive in challenges such as delayed decision making. This may yield low-quality data derived from paper-based records, since data could be incomplete, unclear, invalid, or inaccurate.
- Process-driven data capture. Business processes must be taken into account for designing appropriate interfaces for capturing data. Supervising roles must ensure data quality such as those quality assurance departments found in manufacture.
- Automatic monitoring of behaviors. Sensors could help report and monitor – perhaps automatically– behaviors and occurrences on and off the GC.

#### **Limitations.**

There was the potential selection bias, by having only one male as an informant. A more representative sample could help provide more comprehensive design insights.

## **6 Conclusions**

Following a qualitative study approach, we unraveled some elements that are related to decision making for health care and management in a geriatric center. Although limited, we believe that this work helps shed some light into the issues that are related to decision making in a GC with similar characteristics. In this work, we provided valuable insights into the elements that need to be considered for appropriate interfaces that can help make decisions. Future work includes designing an interface that can be used by stakeholders when making collaborative decisions in a GC.

**Acknowledgements.** This work was partially funded by National Council for Science and Technology (CONACYT) in Mexico with a scholarship provided to the second author. We thank our informants for their support.

## References

1. Soto-Mendoza, V. and J.A. Garcia-Macias, Lessons learned from a long-running assistive system for geriatric care, in International Workshop on Ambient Assisted Living (IWAAL 2014), L. Pecchia, et al., Editors. 2014, Springer: Cham, Switzerland. p. 83-90.
2. Le, T., et al., Health providers' perceptions of novel approaches to visualizing integrated health information. *Methods of Information in Medicine*, 2013. **52**(3): p. 250-258.
3. Amri, S., H. Ltifi, and M.B. Ayed. Towards an intelligent evaluation method of medical data visualizations. in 15th International Conference on Intelligent Systems Design and Applications (ISDA 2015). 2015. Marrakesh, Morocco: IEEE.
4. Görges, M., D.R. Westenskow, and B.A. Markewitz, Evaluation of an integrated intensive care unit monitoring display by critical care fellow physicians. *Journal of clinical monitoring and computing*, 2012. **26**(6): p. 429-436.
5. Bock, C., et al., Engaging older adults in the visualization of sensor data facilitated by an open platform for connected devices. *Technology and Health Care*, 2016. **24**(4): p. 541-550.
6. Placer County Department of Health and Human Services. Be Well Placer: Community Dashboard. 2017; <http://www.placerdashboard.org/>.
7. Sedig, K. and P. Parsons, Interaction design for complex cognitive activities with visual representations: A pattern-based approach. *AIS Transactions on Human-Computer Interaction*, 2013. **5**(2): p. 84-133.
8. Soto-Mendoza, V., et al., Design of a Predictive Scheduling System to Improve Assisted Living Services for Elders. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2015. **6**(4): p. 53.
9. Tang, P.C., et al., Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *Journal of the American Medical Informatics Association*, 2006. **13**(2): p. 121-126.
10. Kawamoto, K., et al., Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*, 2005. **330**(7494): p. 765.
11. Musen, M.A., B. Middleton, and R.A. Greenes, Clinical decision-support systems, in *Biomedical informatics*. 2014, Springer. p. 643-674.
12. Zillner, S., et al. Semantic visualization of patient information. in 21st IEEE International Symposium on Computer-Based Medical Systems (CBMS 2008). 2008. Jyväskylä, Finland: IEEE.
13. Bedrossian, J., et al. Critical design factors for information technology supporting type 1 diabetes management. in *Systems and Information Engineering Design Symposium (SIEDS 2016)*. 2016. Charlottesville, Virginia, USA: IEEE.
14. Nasiri, S. and M. Fathi. Toward an integrated e-health based on acquired healthcare knowledge. in *Middle East Conference on Biomedical Engineering (MECBME 2014)*. 2014. Doha, Qatar: IEEE.
15. Hunt, D.L., et al., Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *Jama*, 1998. **280**(15): p. 1339-1346.
16. Miller, R.A., Medical diagnostic decision support systems—past, present, and future. *Journal of the American Medical Informatics Association*, 1994. **1**(1): p. 8-27.
17. Corbin, J. and A. Strauss, *Basics of qualitative research: Techniques and procedures for developing grounded theory*. 2008, Thousand Oaks, CA: Sage.

# Narrativa visual en inteligencia de negocios para apoyar al proceso de toma de decisiones

## *Visual stories in business intelligence to support decision making*

Andrés Gutiérrez  
Instituto Tecnológico de Sonora  
Sonora, México  
[gerardogtz.dg@gmail.com](mailto:gerardogtz.dg@gmail.com)

Cynthia B. Pérez  
Instituto Tecnológico de Sonora  
Sonora, México  
[cynthia.perez@itson.edu.mx](mailto:cynthia.perez@itson.edu.mx)

Luis A. Castro  
Instituto Tecnológico de Sonora  
Sonora, México  
[luis.castro@acm.org](mailto:luis.castro@acm.org)

Francisco Chávez  
Universidad de Extremadura  
Mérida, España  
[fchavez@unex.es](mailto:fchavez@unex.es)

Francisco Fernández de Vega  
Universidad de Extremadura  
Mérida, España  
[fcofdez@unex.es](mailto:fcofdez@unex.es)

**Resumen** — La visualización de información se está desarrollando en una nueva era en la que las consideraciones sociales, los datos complejos y la tecnología de vanguardia se fusionan para dar sentido a los datos y comunicar eficientemente información valiosa a la audiencia con la finalidad de dar soporte en la toma de decisiones. Aunque la investigación previa proporciona técnicas y modelos para la visualización de datos en general, su alcance se centra a menudo en la diversificación de los tipos de visualización en lugar de la presentación visual de datos complejos y el uso de las historias que se encuentran de forma natural detrás de éstos. La finalidad es que los usuarios logren percibir y entender la información visual para la toma de decisiones. Este artículo propone un modelo interactivo de visualización de información basado en narrativa visual (*storytelling*) para la inteligencia de negocios (BI). Para esto, se llevó a cabo un estudio cualitativo con seis usuarios novatos en visualización de información mediante el cual se identificaron cinco categorías y trece factores de información que fueron tomados en cuenta para el diseño del modelo propuesto.

**Palabras Clave** – *Visualización de información; narrativa visual; inteligencia de negocios; interacción humano-computadora; análisis de datos.*

**Abstract** — Information visualization is emerging in a new era in which social considerations, complex data and state-of-the-art technology are considered for making sense of data and then, effectively communicate valuable information to the audience for decision making. Although prior research provides frameworks for data visualization, their scope is often focused on diversifying visualization types instead of the complexity and the stories behind data. This paper proposes an interactive information visualization model for business intelligence using the storytelling technique. Thus, we conducted a qualitative study with six information visualization novices. As a result, we identify five categories and thirteen information factors that were included in the proposed model.

**Keywords** - *information visualization; storytelling; business intelligence; human-computer-interaction; data analysis.*

### I. INTRODUCCIÓN

En la última década, el uso de la información como factor clave en las organizaciones ha aumentado exponencialmente; y a su vez se han observado progresos notables en el campo de la visualización de datos y la analítica visual. Hoy en día, los dispositivos digitales han mejorado la capacidad de procesamiento, almacenamiento y conectividad, los cuales, a su vez, generan cantidades masivas de datos que deben ser recopilados, consultados, analizados y transmitidos a los usuarios para dar soporte en el proceso de la toma de decisiones. Para ello, la visualización facilita el descubrimiento de patrones, información de interés y proyección de tendencias que ayuden a comprender la información y con ello reducir la incertidumbre en la toma de decisiones. Sin embargo, con la creciente cantidad de datos, hay muchos elementos interrelacionados entre cada pieza de información que enriquece su contexto y significado.

De esta manera, es importante definir cómo presentar la información de una manera comprensible y memorable en los sistemas de visualización de BI para la toma de decisiones. Existe una amplia literatura en el área de visualización de datos, sin embargo, existen pocos estudios acerca de cómo perciben la información los usuarios que no están familiarizados con la visualización de datos. Grammel et. al [1] investigaron acerca de las barreras que se enfrentan los usuarios novatos en la construcción de visualizaciones. Kwon et. al [2] investigaron acerca de los retos que se enfrentan los usuarios novatos en el análisis de datos a través de una herramienta de visualización de datos. Peebles et. al [3] indagaron cómo las personas que no están familiarizadas con visualización de datos le dan sentido a los datos y descubrieron qué factores existen en su interpretación y, finalmente, Lee et. al [4] a través de de teoría fundamentada descubrieron cómo

los usuarios novatos daban sentido a las visualizaciones. En los estudios mencionados los autores coinciden en la importancia de dar estructura y secuencia lógica a las visualizaciones, esto se puede lograr con el poder de las historias, las cuales tienen el potencial de dar la estructura y secuencia necesaria otorgando al usuario un marco de referencia como apoyo para mejorar la comprensión de la visualización donde el objetivo se centra en la adquisición del conocimiento que dé soporte en la toma de decisiones. Desde tiempos antiguos las historias han sido utilizadas como estructura de comunicación para transmitir información y recientemente ha incrementado el interés en su estudio. Autores como Kosara y Macking [5] resaltan la importancia de la narrativa (*storytelling*) en la visualización de datos, la cual comentan puede ayudar a manejar los datos en una estructura fácil de recordar y que los analistas pueden utilizar con fines de presentación y comunicación. Por otra parte, Lee y Rechee [6] propusieron un modelo que identifica el proceso para convertir datos en historias visuales compartidas, el cual está limitado en el aspecto de cómo las organizaciones pudieran utilizar el poder de contar historias. En la actualidad, contar historias es una técnica potencial en la visualización de datos donde la información se presenta como pasos secuenciales e interactivos que mantienen la atención del espectador a través de visualizaciones de datos interactiva. De hecho, la técnica de narrativa visual ha sido considerada en el diseño de herramientas de visualización de datos (Tableau Public [7], GeoTime Stories [8], Qlik [9], SAP Lumira [10], Power BI de Microsoft [11]). Sin embargo, se limitan a incorporar diversas formas de comunicación narrativa y más niveles de interactividad. El uso de la técnica *storytelling* en las herramientas de visualización para soluciones de inteligencia ha recibido muy poca atención. Elias et. al [12] sustentan que los analistas de negocios identifican las relaciones entre múltiples piezas de datos que deben estar interconectados con secuencia lógica con el fin de comunicar sus resultados a los tomadores de decisiones. En su investigación emplean una técnica de narración de cuentos sólo para el proceso de diseño de un tablero de control (*dashboard*) que se centra en las capacidades de exploración y anotación según los resultados de cinco entrevistas con expertos en BI. En dicho estudio examinaron cómo las historias nos ayudan a comunicar conocimiento, compartir e interpretar experiencias y propusieron utilizar la técnica de *storytelling* en el área de BI limitada para el diseño de tableros de control bajo el contexto de exploración y anotación. Para proponer su diseño, realizaron una serie de entrevistas semi-estructuradas a expertos en el análisis de BI que les ayudaron a detectar la necesidad de incorporar la técnica de contar historias en solo una herramienta de visualización de datos utilizada en BI como los tableros de control.

En ese sentido, nuestro trabajo de investigación propone un modelo metodológico que permite utilizar la técnica de *storytelling* en las distintas herramientas de BI permitiendo con ello a las organizaciones contar con una vía de comunicación que exprese de manera efectiva el análisis de datos con el objetivo de comprender y entender mejor la información relevante para el proceso de toma de decisiones. Para esto, se planteó la siguiente pregunta de investigación, ¿Qué factores influyen en la comprensión de información?, y

como preguntas secundarias se establecieron las siguientes: ¿Cómo influye la información en la toma de decisiones?, ¿Cómo influye la presentación de información en la comprensión?, y ¿Cómo influyen los elementos visuales en la comprensión de información? La Sección II aborda la investigación cualitativa que fue utilizada para el diseño del modelo propuesto. La Sección III describe los resultados obtenidos y finalmente las conclusiones son resumidas en la Sección IV.

## II. METODOLOGÍA

Para llevar a cabo nuestro estudio, seguimos un enfoque de investigación cualitativa, bajo la metodología de teoría fundamentada [13], la cual permite a los investigadores producir resultados teóricos comprensibles [14] y dar sentido a una situación existente, esto con el fin de descubrir categorías y determinantes en el proceso de presentación de información que servirán de fundamento para el diseño del modelo. A continuación, se presentan los componentes para llevar a cabo el estudio:

### A. Recolección de datos:

Se recolectaron datos a través de entrevistas semi-estructuradas con seis usuarios novatos en visualización de información, tomadores de decisiones que presentan diariamente información. Para esto, las entrevistas abordaron cuatro categorías enfocadas en las preguntas de investigación como: toma de decisiones, comprensión de la información, presentación de la información y uso de elementos visuales. Se plantearon 38 preguntas abiertas y duraron entre 40 y 60 minutos. Estas entrevistas fueron grabadas y transcritas para su análisis.

### B. Proceso de análisis

El análisis de las transcripciones se realizó utilizando teoría fundamentada. En primera instancia, se analizaron línea por línea las transcripciones mediante codificación abierta que consiste en fragmentar los datos mediante códigos para comprender y desarrollar el concepto de datos; en segundo lugar, se utilizó la codificación axial para clasificar, generar secuencia y lógica con la estructura de datos y finalmente una codificación selectiva que dio como resultado las categorías de los factores que influyen en la comprensión de información.

### C. Descubrimientos

De acuerdo con los resultados de nuestro análisis de datos, para establecer los factores de la comprensión de información se obtuvieron las siguientes categorías:

1) **Factores de información:** Uno de los factores clave en comprensión gira entorno a la información que se presenta, su grado de utilidad, de dónde proviene, su disponibilidad inmediata, los elementos que la conforman y sus beneficios (ver Tabla 1). Como resultado de las entrevistas, uno de los informantes comenta: “No puedes tomar una decisión si no tienes la información actualizada o de manera, adecuada”, donde se observa la importancia de presentar información que esté actualizada y la forma en cómo se presenta, ver Figura 1.

2) **Factores del informante:** Los factores del informante juegan un papel importante en la comprensión, el escenario dónde presenta, la impresión que da a la audiencia y el conocimiento juegan un papel determinante en la comprensión de la información (ver Tabla 1). Para esto, uno de los informantes comenta: “Si el presentador no sabe transmitir aun cuando este bien la presentación habrá algunos que tengan la experiencia entonces si captan, pero si es a público en general si el presentador no sabe transmitir pues ya se perdió el mensaje”, mostrando con ello la importancia de conocer la información, cómo transmitir el mensaje y para quién va dirigido, es decir, conocer bien la audiencia a quién va dirigido el mensaje.

3) **Factores del mensaje:** El mensaje se refiere a lo que el informante desea transmitir a su audiencia, dentro de las propiedades de este factor destacaron su estructura y la claridad en que se transmite (ver Tabla 1). Por ejemplo, uno de los informantes comenta: “Recuerdo, básicamente las presentaciones, bien estructuradas y congruentes, son las que uno necesita”, marcando con ello la importancia que el mensaje debe ser estructurado, debe mantener una secuencia congruente para que éste sea memorable.

4) **Factores de la audiencia:** se refiere a los factores que intervienen en las personas que reciben el mensaje; las propiedades a considerar son su tamaño, preferencias personales, grado de conocimiento y forma de comunicación (ver Tabla 1). En ese sentido, uno de los informantes comenta “A mí en lo particular me llama más la atención lo gráfico, lo que esté asociado a una imagen, quizá haya personas que les llame la atención, los números o las letras”, denotando la importancia que tiene el aspecto visual en la presentación de información para captar la atención de la audiencia.

5) **Factores de interacción:** La comunicación entre el informante es esencial para la comprensión de la información. Los informantes coinciden en también que la retroalimentación es importante (ver Tabla 1). Por ejemplo, un informante menciona: “Hay métodos, para involucrar a las personas, regularmente preguntas directas, del expositor a la

audiencia, para ver qué les parece” denotando con ello que es importante buscar métodos que permitan involucrar a las personas dentro de la presentación de información para que ésta sea efectiva y memorable.

TABLA 1. FACTORES DE COMPRENSIÓN DE LA INFORMACIÓN

Categorías	Propiedades	Tipos, circunstancias, condiciones
Factores de Información	Utilidad	Esencia
	Fuente	Interna - Externa
	Elementos	Validez - Optima - Acceso
	Beneficios	Evidencia – Conocimiento - Reducir incertidumbre
Factores de Informante	Escenario	Medio
	Apariencia	Apariencia visual - Estructura
	Conocimiento	Contexto – Problema – Objetivo - Mercado
Factores de Mensaje	Estructura	Desarrollo - Orden - Conclusión
	Claridad	Elementos visuales – Sensibilización -Enganche
Factores de Audiencia	Magnitud	Masiva - Moderado
	Preferencias	Preferencias personales
	Conocimiento	Conocimiento previo - Experiencias previas - Habilidades
Factores de Interacción	Comunicación	Información - Audiencia Retroalimentación

Los resultados presentados en la Tabla 1 tienen como finalidad conocer y fundamentar los factores del proceso de comprensión que servirá como base para desarrollar el modelo interactivo de visualización de información para BI mediante la técnica *storytelling* para la presentación de la información.

### III. DISEÑO DE MODELO

Con el fin de visualizar la información de forma eficiente y lograr con ello una mejor comprensión de la información en los usuarios para reducir la incertidumbre en la toma de decisiones, proponemos un modelo metodológico a partir de los resultados obtenidos de la investigación cualitativa,



Figura 1. Modelo interactivo de visualización de información basado en narrativa visual (storytelling) para inteligencia de negocios.

además del estudio y análisis de la investigación que se ha llevado en periodismo [15, 16], analítica visual [17, 18] y BI [12, 19]. El modelo propuesto se compone de cuatro componentes básicos: gestión de datos, diseño de la historia, compartir la historia y recibir historias (ver Figura 1).

#### A. Gestión de datos

La etapa de gestión de los datos consiste en el grupo de actividades destinadas a explorar y analizar datos en herramienta de BI lo que constituye la fuente de contenido que esconden los datos. El resultado de gestionar datos una vez procesados por soluciones de BI es el de obtener información destilada que el diseñador de información utilizará para diseñar la presentación a través de la narrativa visual para comunicar efectivamente a su audiencia.

#### B. Diseño de la historia

En esta etapa, el diseñador de información diseña la historia partiendo del resultado del análisis de la información. Posteriormente, analiza el contexto que rodea a su audiencia: preferencias, necesidades y características especiales. Finalmente, se construye de forma creativa la historia. Para construir la historia utilizamos la información destilada en forma de narrativa, es decir, de manera creativa separamos la información en fragmentos clave que al darle una secuencia lógica darán sentido a la información en forma de historia. Algunos elementos que deben tener son: a) Escenario: Se presenta a la audiencia el contexto en el que se desarrolla la historia y se aporta conocimiento general sobre el tema. b) Contraste: Se presenta un punto de contraste en la visualización de forma comparativa, con el de captar la atención del espectador y crear intriga. c) Solución del problema: se proporciona una solución que permite al espectador comprender el mensaje, explorar la información y crear conclusiones con el fin de llevar a la audiencia a descubrir nuevo conocimiento. d) Atracción: Se deberá desarrollar bajo criterios estéticos para atraer la atención de la audiencia a través de una historia visualmente agradable.

#### C. Compartiendo la historia

Una vez diseñada la historia se procede a compartirla con la audiencia, para ello se analiza y escoge el medio apropiado para comunicar la información a través de una herramienta que permita visualizar la información preferentemente en la nube para que se pueda acceder a la información en tiempo real y desde cualquier dispositivo digital con acceso a internet. Sin embargo, si no se tiene acceso a internet se puede tener la aplicación *offline* en el dispositivo digital.

#### D. Recibiendo la historia

La audiencia recibe la historia, obtiene valor y ofrece retroalimentación para las mejores prácticas en el futuro. En esta etapa, se establece la comunicación entre el diseñador de información y la audiencia a través de la herramienta de visualización de información que le permite a la audiencia, explorar, interactuar y analizar con los elementos de la historia que ayudarán a reducir incertidumbre en el proceso de la toma de decisiones.

## IV. CONCLUSIONES Y TRABAJO FUTURO

Los resultados de esta investigación apoyan la idea sobre la importancia del uso *storytelling* en la comprensión de información. De esta manera, propusimos un modelo metodológico basado en la técnica de *storytelling* para proporcionar información que mejore la comprensión de información en el proceso de toma de decisiones en BI. Esto permitirá a los diseñadores de información un modelo interactivo para comunicar sus hallazgos a través de la técnica de *storytelling*. Como trabajo futuro se tiene la implementación del modelo propuesto para la difusión del problema del cambio climático en el estado de Sonora, México. Para ello, se están utilizando datos geoespaciales a través de una aplicación web de la plataforma arcgis [20] donde se pretende visualizar la relación del cambio climático con los golpes de calor a través de los años y las consecuencias que esto conlleva en la sociedad. De igual forma, se espera validar el modelo propuesto mediante la implementación y evaluación en diferentes casos de estudio que involucre organizaciones públicas y/o privadas.

### REFERENCIAS

- [1] Grammel I, Torv M & Storev M A "How information visualization novices construct visualizations" IFFF transactions on visualization and computer graphics, Vol. 16, pp. 943-952. 2010.
- [2] B Kwon B Fisher and I Yi "Visual analytic roadblocks for novice investigators" 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 3-11. 2011.
- [3] D Peebles D Ramdunv-Ellis G Ellis and I V H Ronner "The influence of graph schemas on the interpretation of unfamiliar diagrams" presented at the 27th British Comput Soc Human Comput Interaction Conf.: The Internet Things, Brunel Univ., London, United Kingdom, 2013.
- [4] S. Lee, S. Kim, Y. Hung, H. Lam, Y. Kang and J. Yi. "How do People Make Sense of Unfamiliar Visualizations? A Grounded Model of Novice's Information Visualization Sensemaking" IFFF Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 499-508, 2016.
- [5] R. Kosara and I. Mackinlay. "Storytelling: The Next Step for Visualization", Computer, vol. 46, no. 5, pp. 44-50, 2013.
- [6] B Lee N Riche P Isenberr and S Carnendale "More Than Telling a Story: Transforming Data into Visually Shared Stories" IEEE Computer Graphics and Applications, vol. 35, no. 5, pp. 84-90, 2015.
- [7] Tableau Public. <http://tableausoftware.com/public>, consultado en 2017.
- [8] Thomas Kanler and William Wright "Geo time information visualization.", Information Visualization, vol. 4, pp.136-146, July 2005.
- [9] <http://www.qlik.com/us/>, consultado en 2017.
- [10] <http://www.sap.com/product/analytics/lumira.html>
- [11] <https://powerbi.microsoft.com/es-es/>
- [12] M Elias M Anfaure and A Rezerianos "Storytelling in Visual Analytics Tools for Business Intelligence", Human-Computer Interaction - INTERACT 2013, pp. 280-297, 2013.
- [13] Strauss A L. & Corbin J "Basics of Qualitative Research: Grounded Theory Procedures and Techniques". 3rd Ed. Newbury Park, Ca: Sage, 2008.
- [14] Z. Pousman I. Stasko and M. Mateas. "Casual Information Visualization: Depictions of Data in Everyday Life" IEEE Transactions on Visualization and Computer Graphics, vol. 13, no. 6, pp. 1145-1152, 2007.
- [15] T. Munzner. "A Nested Model for Visualization Design and Validation" IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 921-928, 2009.
- [16] S. Bateman R. Mandryk C. Gutwin A. Genest D. McDine and C. Brooks "Useful junk?" Proceedings of the 28th international conference on Human factors in computing systems - CHI '10, 2010.
- [17] D. Sacha A. Stoffel F. Stoffel B. Kwon G. Ellis and D. Keim. "Knowledge Generation Model for Visual Analytics" IFFF Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 1604-1613, 2014.
- [18] Green, T. M., Ribarsky, W., & Fisher, B., "Building and applying a human cognition model for visual analytics.", vol. 8, pp. 1-30, 2009.
- [19] Neesh S. "Business Intelligence" Communications of the Association for Information Systems, vol. 13, pp. 177 - 195, 2004.
- [20] <http://www.arcgis.com/>, consultado en 2017.

# Data visualization process through storytelling technique in Business Intelligence

Andrés Gutiérrez  
Instituto Tecnológico de Sonora  
Antonio Caso S/N esq. Kino  
Cd. Obregón Sonora  
+52 6449978217  
gerardogtz.dg@gmail.com

Cynthia B. Pérez  
Instituto Tecnológico de Sonora  
Antonio Caso S/N esq. Kino  
Cd. Obregón Sonora  
+52 4109000 ext. 1517  
cynthia.perez@itson.edu.mx

## ABSTRACT

New technologies have allowed corporations to collect, store and manage huge quantities of data, being necessary in their daily management. Hence, business intelligence and data analytics have become an important and rapidly growing area of study that reflects the impact of data-related problems to be solved in business organizations. However, it is difficult to recognize patterns and draw conclusions from large amount of data and also, understand the stories of such information. We propose a methodology based on storytelling technique in order to improve the perception and interpretation of the data with the aim to make the information more memorable for people.

## CCS Concepts

• **Human Centered Computer** → **Data Visualization** • **Human Centered Computer** → **Interaction Design** • **Business Intelligence** → **Data analytics**

## Keywords

Data visualization, storytelling, business intelligence.

## 1. INTRODUCTION

New changes are emerging that will force organizations to work in different ways. Technology has increased, devices have improved the capacity to process, store and generate massive amounts of data, recognizing that value of companies have begun to push the use of information systems for the efficient management of information. Elias [1] argues “Business Intelligence (BI) deals with the collection of processes and software that supports organizations in understanding large business datasets, retrieving and analyzing information and making decisions”. Airinei & Homocianu [2] categorized BI technologies into reports, online analytical processing, analytics, data mining, business performance management, benchmarking, text mining, and predictive analytics. At the same time, a huge number of visualization systems have been used to help users view, explore, and analyze information [3]. However, such visualization systems have received little attention in organizations in the way to use them as a visual language in order to make information accessible to more people.

In recent years, there has been an increasing interest in the importance and value of narrative and anecdotal information in

the form of stories. According to M. Johnson [4] stories are one of the communication structures of humanity. The value of "storytelling" transcends language and culture, and as we move towards a future of improved connections between people and digital effects, stories continue to represent the most compelling platform for managing imagination, as they can provide interpretation by taking advantage of the results of analysis, trends and scenarios in spreadsheets. Since, analysts work with increasingly large data sets, data visualization has become an incredible important asset during sense-making analysis and to communicate results to other analysts, decision makers or to their audiences. Thus, the importance of storytelling in data visualization would be helpful in order to enhance visual analysis tools providing understandable information to a broader public.

The aim of this paper is to examine and propose a methodology based on storytelling technique in BI in order to provide insights and natural ways of perceiving complex information and makes it more memorable.

## 2. RELATED WORK

Kosara and Mackinlay [5] define a story as an ordered sequence of steps, with a clearly defined path through it, they argue that the stories in a structured context can be used not only to support discussion and decision-making, but also in the analysis process. By way of illustration, Heather Craig [6] discusses a web platform Face of Fracking, see Figure 1, about the impacts of fracking in California that uses storytelling technique through data visualization and interactive narrative form that facilitates understanding of the beholder.

Preliminary work on storytelling technique was undertaken by Lee and Reche [7], they propose a process of transforming data into visually shared stories. Although, they do not focus in business related-problems, we believe it is a fast growing application area in business that can provide significant competitive advantage to an organization. Thus, we contribute with a data-visualization methodology based on storytelling technique expecting to improve the perception and interpretation of data in business intelligence.

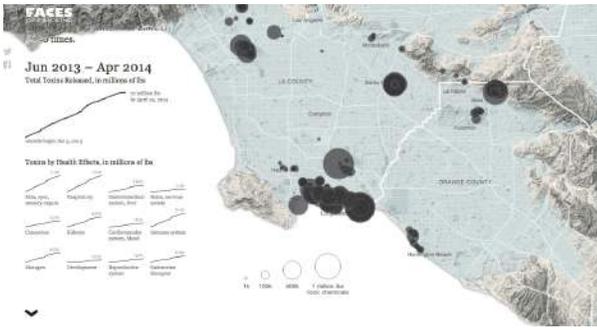


Figure 1. Example visual data story. Face of Fracking is available at <http://www.facesoffracking.org/data-visualization/>

### 3. PROPOSED METHODOLOGY

The process of turning data into a visual story has so far received little attention in business intelligence and analytics. Important stories live in our data; when we observe statistical information such as sales, profit by product, census population in tables or simple charts does not reflect the story behind data, the information is abstract, difficult to understand and communicate results. In order to visualize data effectively derived from an understanding of human perception, we propose to use storytelling technique in data visualization for business intelligence. Storytelling technique regulates the perception and interpretation of the data with the aim to make the information more memorable for people. The objective is to provide a methodology for the creation and presentation of stories from raw

### 3.1 Explore and Gain Insight

The context of the organization and the information obtained through the process of business intelligence is analyzed.

### 3.2 Create Story

Creating a story consists in identifying the context and creating a structure in order to get the viewer's attention, create intrigue, know the background and know the characters.

#### 3.2.1 Identify Context

At this stage, it is ensured that the audience knows the context in which the story unfolds. It stirs up interest from the viewer, explains general details and introduces the characters.

#### 3.2.2 Make Structure

The information is separated into small pieces and a sequence is created to which the point of contrast must be integrated, so that the audience becomes interested, and provide a solution that allows the viewer to explore information and create conclusions based on results.

### 3.3 Select Media

Once the type of audience is identified, implementing the story and setting its navigability and interactivity select a proper media for the presentation.

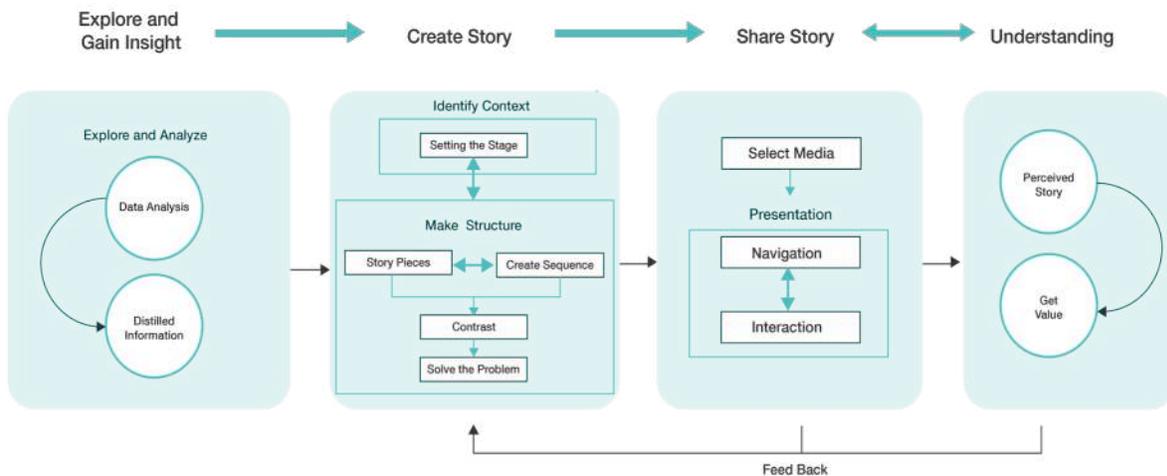


Figure 2. Data visualization process through storytelling technique in Business Intelligence.

data, offering a series of steps that allow people with technology skills use this proposal in dashboards, presentations, web pages, among others. The proposed methodology is shown in Figure 2 and comprises four steps as follows:

1. Explore and gain insight
2. Create story
3. Select Media
4. Understanding

### 3.4 Understanding

The user perceives the story, gets value and offers feedback for best practices in the future. The correlation between previous four steps will help us to provide a sequence of steps to tell information through stories and presenting information in an easy and remarkable way.

## 4. CASE STUDY

In this section we describe an example based on guidelines from the proposed methodology. This case study was conducted using information generated by the National Water Commission of México (CONAGUA)<sup>1</sup>. The mission of this organization is to preserve national waters and their inherent public goods for their sustainable management and ensuring water security with responsibility for the orders of government and society in general. According to the vision of the 2030 Water Agenda, CONAGUA,[8], its requires average annual investments greater than 4.16 billion dollars, turning its vision into a reality. Their principal objectives are focus on measures to increase efficiencies in the use of water in agriculture and public urban supply. Not acting implies growing opportunity costs, which for unmet industrial demand alone would reach orders of magnitude of 124.70 billion dollars per year up to 2030. The growing need to raise awareness about the sustainable use of water resources is noticeable. Therefore, it would be helpful if managers, educators, governments and society in general contribute to achieving the objectives and strategies proposed by the government. In this regard, our proposal as an interactive technological tool could contribute that people understand through visual information about the current hydrological problems and the possible consequences this would have in the near future.

In this case study, we analyzed a possible scenario to 2025 about the water consumption in different regions in Mexico according to CONAGUA and CONAPO<sup>2</sup> (National Population Council) projections.

was used to explored and analyze the information as shown in table below.

Administrative Region	Average natural availability per capita 2014	Average natural availability per capita 2025
Baja California Peninsula	1317	833
Northwest	3210	2491
North Pacific	6038	5517
Balsas	2703	2403
Sur Pacific	7782	7378
Rio Grande	1356	1007
North Central Basins	1726	1606
Lerna - Santiago - Pacific	1820	1583
North Gulf	4666	4200
Center Gulf	10574	9853
South border	24549	19790
Península de Yucatán	8255	5786
Mexico's Valley and Cutzamala System	188	162
Baja California Peninsula	1317	833
Northwest	3210	2491

Table 1. Average natural availability per capita in 2014 and 2025.

## 4.2 Create Story

We started with the question how will the future of water in Mexico? To hook the audience and introduce the context followed

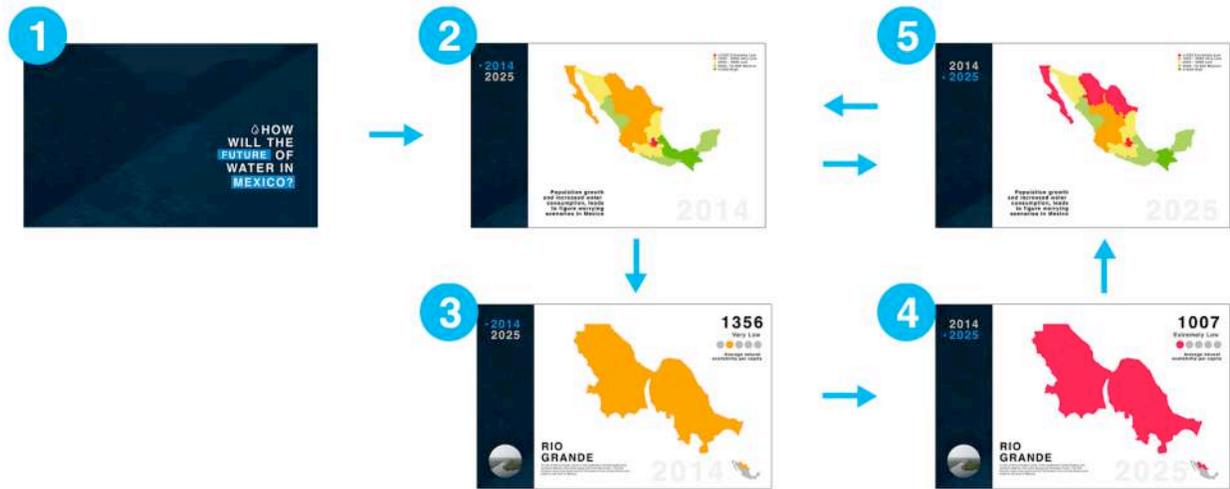


Figure 3. Data visualization process through storytelling technique using the proposed methodology for CONAGUA organization.

## 4.1 Explore and Gain Insight

Report on future scenarios of the water sector in Mexico average natural availability per capita in 2014 and 2025 of CONAGUA

by general information.

We use the regions as characters in this story, followed by the history divided into small parts and a logical sequence, see Figure 3, giving the meaning to the history; here the characters are gradually introduced up to the last scene where the contrast is shown and the water problem is revealed.

1 <http://www.conagua.gob.mx>

2 <http://www.gob.mx/conapo>

### 3.3 Select Media

In order to illustrate the use of storytelling technique in data visualization for this case study, we select adobe illustrator to create some transitions, about the water consumption per capita since 2014 through 2025, see Figure 3. The design of the interactive tool it is recommend to use a business intelligence software as Tableau, Qlik View, Pentaho, among others.

### 4.4 Understanding

Finally user perceives the story, gets value and offers feedback for best practices in the future, see Figure 3.

## 5. CONCLUSION

The results of this research support the idea about the importance of data visualization in business intelligence for decision-making. In this way, we propose a methodology based on storytelling technique in BI in order to provide information more remarkable for people to support the decision making process. Thus, we presented a case study to illustrate the use of our proposed methodology. The implementation of the interactive tool based on storytelling technique in a business problem is currently under development. Finally, we expect to assess the model in different scenarios for novice and expert users in order to validate our approach.

## 6. REFERENCES

- [1] Elias, M., Aufaure, M. A., & Bezerianos, A. (2013, September). Storytelling in visual analytics tools for business intelligence. In IFIP Conference on Human-Computer Interaction (pp. 280-297). Springer Berlin Heidelberg.
- [2] Airinei, D., & Homocianu, D. (2010). Data visualization in business intelligence. 2010 Proceeding of WSEAS MCBEC2010-RECENT ADVANCES in MATHEMATICS and COMPUTERS in BUSINESS, ECONOMICS, BIOLOGY & CHEMISTRY.
- [3] Gotz, D., Lu, J., Kissa, P., Cao, N., Qian, W. H., Liu, S. X., & Zhou, M. X. (2010, February). HARVEST: an intelligent visual analytic tool for the masses. In Proceedings of the first international workshop on Intelligent visual interfaces for text analysis (pp. 1-4). ACM.
- [4] Johnson, M. (1993). Moral imagination: Implications of cognitive science for ethics (Vol. 190). Chicago: University of Chicago Press.
- [5] Kosara, R., & Mackinlay, J. (2013). Storytelling: The next step for visualization. *Computer*, (5), 44-50.
- [6] Craig, H. (2015). Interactive data narrative: designing for public engagement (Doctoral dissertation, Massachusetts Institute of Technology).
- [7] Lee, B., Riche, N. H., Isenberg, P., & Carpendale, S. (2015). More than Telling a Story: A Closer Look at the Process of Transforming Data into Visually Shared Stories. *IEEE computer graphics and applications*, 35(5), 84-90.
- [8] Conagua. (2016). *2030 Water Agenda*. [online] Available at: [http://www.conagua.gob.mx/english07/publications/2030\\_water\\_agenda.pdf](http://www.conagua.gob.mx/english07/publications/2030_water_agenda.pdf) [Accessed 30 Jul. 2016].

# Decision Support System for a SME in the Restaurant sector: Development of a Prototype

Juan Edel Gutiérrez Neyoy  
Department of Computer Science  
and Design,  
Instituto Tecnológico de Sonora,  
México  
[Juanedelgtz@gmail.com](mailto:Juanedelgtz@gmail.com)

Luis-Felipe Rodríguez  
Department of Computer Science  
and Design,  
Instituto Tecnológico de Sonora,  
México  
[Luis.rodriguez@itson.edu.mx](mailto:Luis.rodriguez@itson.edu.mx)

Luis A. Castro  
Department of Computer Science  
and Design,  
Instituto Tecnológico de Sonora,  
México  
[Luis.castro@acm.org](mailto:Luis.castro@acm.org)

**Abstract** — Business intelligence serves a great purpose in any business, but is usually omitted or ignored by SMEs due to lack of knowledge or simply because the owners do not consider it necessary for a medium to long-term plan term. In the long run, businesses usually fail to make a profit or cannot continue, and end up without even knowing what happened. The information generated by a company could serve as a starting point for good and solid decision-making, but the culture that dominates the SMEs of any industry leaves important information without registration or connection, leaving them to work based only on evidence and errors and Mostly "luck". This paper proposes a prototype of decision support system for a SME in the restaurant sector. The results of this work show that the company (being a SME) can improve in terms of competitive advantage through this tool of business intelligence and cross the barriers where most SMEs are stagnant in Mexico.

**Keywords** - Business Intelligence, Decision support system, Data warehouse, SME, Dashboard, Key performance indicator.

## I. INTRODUCTION

The current business environment is the most complex and competitive in history. Surely, in the following years it will be even more complicated. Companies around the world are looking for new ways to gain a competitive advantage in order to face the aggressive competition that is currently being experienced.

Business Intelligence (BI) is a field whose main objective is to optimize the decision-making process in business by transforming data into information, and information into knowledge [1]. The implementation of business intelligence systems helps decision makers to achieve company's strategic goals. In fact, national governments around the world have known for some time the competitive advantage that can be achieved through an organized BI program [2]. However, many companies have not adopted a system for collecting data, for further analysis and intelligence. The development and use of this new management tool could significantly increase the competitiveness of companies [3].

There are several reasons for companies to consider developing a business intelligence system. First, each sector or department of the company needs reliable and timely information to perform its work [4]. Most companies are not aware of the amount of work and resources it takes to generate all that information and effectively distribute it to the people

who really need it [5]. Even more alarming, not many are aware of the cost of not doing so, as this leads to a lack of coordination, wrong decisions and misuse of resources available in the company. There is no more expensive information than that which is not available at the time needed. Second, the company can begin to face problems and projects that were previously impossible resolve due to lack of time or resources. In general, having relevant and timely information expands the company's horizon [6]. Third, the competition among companies has become more aggressive and sophisticated. Competitors are using national research and development programs to improve their technological strengths. Others are using alliance strategies to attack their competitors' customer portfolios. All are competing in terms costs and quality, which requires strategic thinking and quick action by the operational managers of the corporation. In addition, many of these competitors use competitive intelligence more effectively and more aggressively than others [2]. Fourth, business intelligence systems are not only capable of providing analytical, predictive, and reliable information quickly, these systems can also do it in an attractive and easy to understand visual way. This flexibility of presenting information allows decision makers to easily assimilate what is happening at the moment and to better detect crucial KPIs that may be useful to decision makers. For these reasons, business intelligence systems need to be developed and implemented in most types of business [7].

Everyday, small and medium enterprises are created in Mexico. According to ProMéxico, with data from INEGI, there are just over 4 million businesses in Mexico. Of those, 99.8% are SMEs, which generate 52% of gross domestic product (GDP) and 7 out of 10 jobs. In other words, small and medium enterprises are the support of development and industrial growth in Mexico [8].

This growth and the direct management of the business by their owners, together with the inexperience in information systems, have led companies to an incorrect implementation and use of data generated in daily operations, including average ticket amount, purchasing and acquisitions processes, sales by product, staff performance, among others. The lack of formal data entry and data analysis has led to several problems such as missing inventory, losses due to material waste, partial solutions and confusion from both staff and owners, despite

any success a company may have. All this adds up to problems and failure to comply with customer demands.

These are just some problems currently affecting SMEs: lack of strategic planning, problems between owners of a family-owned businesses, little or none business and financial background, no employee development, lack of a formal hiring and training process, excessive and unneeded expenditures, and low or null IT investment. These are some of the examples that affects SMEs and makes it harder for their owners and managers to visualize the need for a BI solution. Even the possibility of accessing specialized financial programs for SMEs can be greatly increased by having control and knowledge of operational and financial information readily accessible [9].

In the business environment, specifically in the restaurant sector, BI systems have been designed to provide support for decision making through Dashboards in order to provide value and advantage to the company. Evidence-based or historical decision-making generated by transactional systems relies on safe access to timely and accurate information [10].

This work was carried out in a SME of the restaurant sector, which has four branches in the same city. Through a situational analysis, we were able to identify several areas of opportunity that prevented the company from growing in the region, one of which was lack of control and vision of their sales. Managers could not know for sure the sales information of the branches; they could not know the income and expenses as well as the profits generated by the branches.

We believe that an implementation of a BI tool could directly attack this area of opportunity by providing managers with a clear and complete view of the sales area enabling them to view the historical data generated by their transactional systems in a visual way attractive to the user. We believe that business intelligence can help the organization to better manage its information for decision-making. We propose the development of a BI tool, which would be a Decision Support System. This system will be supported by a data warehouse, which will take as data source all the transactional databases of the company where sales histories are generated. Subsequently, with the data stored in the data warehouse, we will design a dimensional modeling of the information for later presenting it to the end user in a visual and attractive way to facilitate decision-making. The purpose of this work is to design a support system for decision making for the sales area of a SME in the restaurant sector, using a public database.

## II. METHOD

We followed a practical approach in the development of this work. In particular, we followed the methodology proposed by Ralph Kimball [11].

The Kimball Methodology is used for the construction and development of data warehouses (DW) for BI solutions. A data warehouse is a collection of data oriented to a specific domain (e.g., company or organization), integrated, non-volatile and variable over time. A data warehouse helps the decision making process in the company in which it is used [11]. The methodology is based on what Kimball calls the Dimensional Business Life Cycle, which can be seen in Figure 1.

In Figure 1, we can observe the tasks that Kimball proposes in his methodology, which consists of three working paths:

1. Technology (Top Route). It involves tasks related to use of specific software such as Microsoft SQL Analysis Services.
2. Data (Middle Route). In which the dimensional model is designed and implemented, and the Extraction, Transformation, and Load (ETL) subsystem to load the DW is developed.
3. Business Intelligence Applications (Bottom Route). This route includes tasks in which we design and develop business applications for end users.

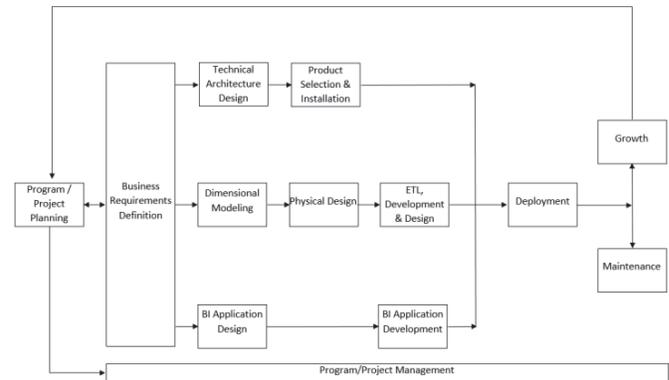


Figure 1. Ralph Kimball Methodology Diagram.

It should be mentioned that these routes are combined when the system is finally installed and the tasks are chosen depending on the magnitude of the project to be developed. Therefore, it may not be possible to carry out all the tasks of the methodology. The general project management activity is shown in the bottom part of the figure.

## III. DESIGN AND DEVELOPMENT

To carry out the development of the prototype we take as reference the tasks proposed by Ralph Kimball as part of his methodology (these tasks can be seen in Figure 1). We chose only the appropriate tasks to achieve our objective, which were:

- Business requirements
- Design of the technical architecture
- Product Selection
- ETL and dimensional modeling
- Development of the application

The following describes what was done in each of the tasks that we chose for the project.

### A. Business requirements

In the situational analysis carried out in the company the needs and areas of opportunity of the business were detected. During the situational analysis the business owners and some employees were interviewed in several weekly meetings. In

addition, visits were made to the different areas of the restaurant to observe what a normal day of work was like and how they performed their activities. We reviewed the growth history of the company, the basic business processes, and the structure of the organization chart.

The results of this analysis revealed different areas of opportunity that the company could take advantage of, but due to the company's current goal, which is to expand to other regions outside the city, we decided to attack the part of the control of revenues and expenses, as well as improving the technological infrastructure in which the company's transactional systems works. That is why we decided to carry out this project to implement a prototype to support the decision making, as it may impact several areas of opportunity that were detected in the company.

This analysis helped to identify the possible indicators for different areas of the company, which would help us to develop the Dashboard design for the sales area of the company. In Table I we can see the different indicators and their respective areas.

TABLE I. INDICATORS FOUND IN THE AREAS OF THE COMPANY.

Area	Indicators
Sales	<ul style="list-style-type: none"> <li>- Sales volume.</li> <li>- Sales by type.</li> <li>- Sales by product.</li> <li>- Total sales.</li> <li>- Sales over time.</li> <li>- Average ticket.</li> <li>- Ticket per hour.</li> </ul>
Kitchen	<ul style="list-style-type: none"> <li>- Preparation time.</li> <li>- Delivery time.</li> <li>- Average by type.</li> <li>- Average per kitchen.</li> </ul>
Internal Processes	<ul style="list-style-type: none"> <li>- Total processes.</li> <li>- Defined processes.</li> <li>- Documented processes.</li> </ul>
Learning and innovation	<ul style="list-style-type: none"> <li>- Training.</li> <li>- Certified employees.</li> <li>- Outstanding training.</li> <li>- Training required.</li> </ul>

*B. Design of technical architecture*

The design of the technical architecture refers to the set of hardware and software on which will run the different services that the company needs to achieve our objective.

Through an analysis of technological infrastructure, we were able to identify some deficiencies. The company did not have a dedicated server; they did not even use one. All the data of the transactional systems were registered in different personal computers, which in turn, function as a point of sale. Figure 2 shows the current technological infrastructure of the company.

*C. Product Selection*

For the development of the prototype presented in this paper, we used the Qlik View tool. It is a European product with more than 20 years in the market, which reports about 33,000 customers in more than 100 countries. Qlik View allows users to develop and deploy applications completely adapted to their needs in fast and efficient way. This tool promotes unrestricted analysis of application data, helping users discover hidden trends and make discoveries that drive decisions. Some key characteristics are the following:

- Fast deployment
- High cost of licensing
- Good interactive visualization
- Extraction of data through different sources such as Excel and database.

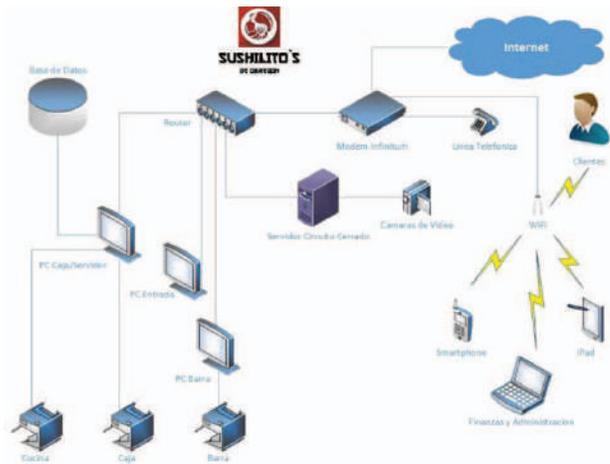


Figure 2. Current technological infrastructure.

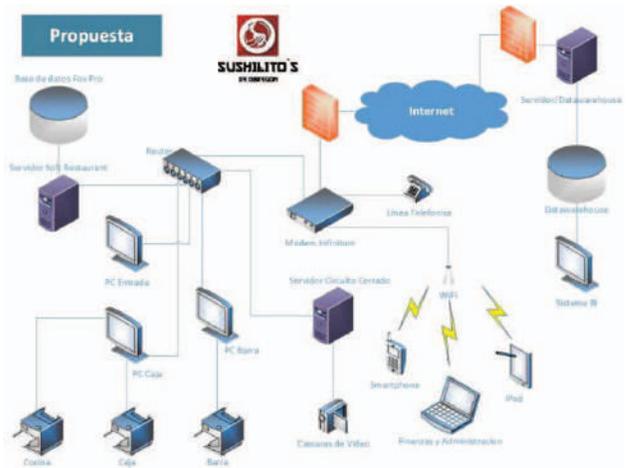


Figure 3. Proposed technological infrastructure.



In Table II we can observe the test that was applied to the participants of the study. Following Table II it can be seen that the percentage of participants with the highest items were "I would find system useful in my job", "I would find system to be flexible to interact with", "I would find it easy to get system to do what I want it to do," and the less for the item "Using system would improve my job performance." This means that our participants generally perceived that the proposed system complies with established requirements. That it would be easy for them to do what the users want, but it would be difficult for them to increase their performance through the use of the system. None of the participants disagreed at any level. These results can be seen in Figure 7.

TABLE II. TEST UTILITY PERCEIVED AND FACILITY OF USE

PERCEIVED USEFULNESS		1	2	3	4	5	6	7
1. Using system in my job would enable me to accomplish tasks more quickly.	UNLIKELY							LIKELY
2. Using system would improve my job performance.	UNLIKELY							LIKELY
3. Using system in my job would increase my productivity.	UNLIKELY							LIKELY
4. Using system would enhance my effectiveness on the job.	UNLIKELY							LIKELY
5. Using system would make it easier to do my job.	UNLIKELY							LIKELY
6. I would find system useful in my job.	UNLIKELY							LIKELY
PERCEIVED EASE OF USE		1	2	3	4	5	6	7
7. Learning to operate system would be easy for me.	UNLIKELY							LIKELY
8. I would find it easy to get system to do what I want it to do.	UNLIKELY							LIKELY
9. My interaction with system would be clear and understandable.	UNLIKELY							LIKELY
10. I would find system to be flexible to interact with.	UNLIKELY							LIKELY
11. It would be easy for me to become skillful at using system.	UNLIKELY							LIKELY
12. I would find system easy to use.	UNLIKELY							LIKELY

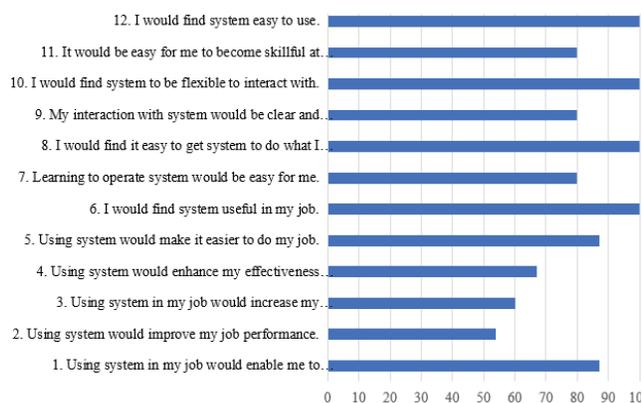


Figure 7. Results test Questionnaire for User Interface Satisfaction.

## V. RELATED WORK

Previous studies and proposals let us to observe and analyze cases that have been carried out to optimize or improve performance in key business processes. For example, Trkman demonstrates how the supply chain is optimized by including the use of Business Analytics in the redesign of the supply chain process, obtaining several benefits such as improvement

in performance, better use of resources and avoiding downtime [14].

We can also mention the case of Bayer, as they needed to consolidate the processes carried out in different countries where they operate and reduce errors by manual data entry in different formats. It had to be able to unify and manage multiple tools and data sources, resulting from the large number of acquisitions made since 2005 by Bayer. IBM offered Cognos BI solutions and implemented them alongside its business associate, KM. The platform has been deployed in 8 countries and there are plans to continue expanding. This optimization allowed the immediate availability of market information for quick decision-making, standardization of reports, cubes and dashboards, reduced infrastructure costs, licensing and consulting. Thanks to the IBM solution, negotiating with suppliers is centralized and support costs have been reduced [15].

In 2005, Sbarro Inc. installed in more than 400 stores a reporting system called Aloha Enterprise. This reporting system works in conjunction with the chain's Aloha by Radiant Systems point of sale software and hardware, and it worked by copying, on an hourly basis, in-restaurant data to a centrally hosted database augmented by analytical tools and reporting capabilities. Before the implementation, the data was received once a week, provoking late reactions and decisions. Now they have the information available to analyze every hour, and a reduction on food and labor costs was met, while also detecting dishonest employees, theft and employee coaching needs. Nonetheless, they recognize that "the most important change was to corporate culture" [16].

The Einstein Noah Restaurant Group moved managers from reports on spreadsheets to functional dashboards that displayed store performance data and the implementation achieve positive feedback. With nearly 7,000 employees and approximately 860 locations in the US, it's the nation's largest operator of bagel bakeries. As part of a 2013 BI initiative to migrate from Excel-based reporting, they started to build and populate a new enterprise data warehouse by employing agile development methods, adding also a mobile enterprise analytics project with Microstrategy. Qlik view and Microsoft SQL will still be used, but with Microstrategy as the focus point of the BI strategy. The plan is to move all data sources into a data warehouse, enabling better data governance, including master data management. Future plans are now focused on incorporating information, such as store audit data in the data warehouse to have it readily available through the MicroStrategy app, thus allowing business managers to plug in information from a store site into a dashboard on an iPad easily.

Red Robin International Inc. had different cross-data databases in Microsoft Access, which referred to the same product with different codes. This made it nearly impossible to compare sales and costs, which in turn caused a mismatch of thousands of dollars weekly when calculating costs. By carrying out a process that involved cleaning information, while also providing quality data, financial reports were balanced, and managers were able to generate reports and menus based on customer preferences, calculate the ideal cost

of each product, review profit margins and even savings by tracking lower costs of raw materials with different providers [17].

In 2012, El Pollo Loco inc, was able to review the latest data on limited-time offers, as well as the performance of new products. The company invested \$250,000 in 2010 with IBM Corp.'s Cognos Business Intelligence software to be able to "rack and stack numbers." This translated in easier, faster data comparison across the company's 401 restaurants, 58% of which are franchised. After the implementation, the company showed a 9.9% increase in same-store sales and 9% boost system-wide in 2012, reaching \$607 million in sales, regaining the company's profitability [18].

## VI. CONCLUSIONS

Decision support systems facilitate and make the decision-making process more efficient in the phases of design, selection and intelligence. This is carried out by transforming data into useful information (intelligence) for the application of the model (design) of information necessary to solve a problem by providing possible solution options (selection) in the environment that the organization lives in the administrative, financial, productive, and other areas. The proposed prototype generated a positive impression according to the results obtained using the validation instrument. It is a possible solution to attack an area of opportunity that was detected in the company. The next step is to take this prototype to the implementation using the actual data of the company to facilitate the decision making process in the sales area.

## REFERENCES

- [1] Withee, K., Microsoft Business Intelligence For Dummies. 2010: Wiley.
- [2] Herring, J.P., Building a business intelligence system. Journal of Business Strategy, 1988. 9(3): p. 4-9.
- [3] Ngai, E., et al., Decision support and intelligent systems in the textile and apparel supply chain: An academic review of research articles. Expert Systems with Applications, 2014. 41(1): p. 81-91.
- [4] Demirkan, H. and D. Delen, Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. Decision Support Systems, 2013. 55(1): p. 412-421.
- [5] Rooke, W., Business intelligence system. 2001, Google Patents.
- [6] Popovič, A., P.S. Coelho, and J. Jaklič, The impact of business intelligence system maturity on information quality. 2009.
- [7] Bocij, P., A. Greasley, and S. Hickie, Business information systems: Technology, development and management. 2008: Pearson education.
- [8] ProMéxico, P., eslabón fundamental para el crecimiento en México, 2014. 2015.
- [9] Domenge, R. and I. Belausteguigoitia, Nuevas PYMES: problemas y recomendaciones. Extraído de <http://direccionestrategica.itam.mx/nuevaspymes-problemas-y-recomendaciones/consulta>, 2010. 20: p. 08-14.
- [10] Foshay, N. and C. Kuziemsky, Towards an implementation framework for business intelligence in healthcare. International Journal of Information Management, 2014. 34(1): p. 20-27.
- [11] Kimball, R. and M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 2013: Wiley.
- [12] Moss, L.T. and S. Atre, Business intelligence roadmap : the complete project lifecycle for decision-support applications. 2006, Boston, Mass.; Munich [u.a.]: Addison-Wesley.
- [13] Inmon, W.H., Building the Data Warehouse. 2002: Wiley.
- [14] Trkman, P., et al., The impact of business analytics on supply chain performance. Decision Support Systems, 2010. 49(3): p. 318-327.
- [15] IBM. Bayer optimiza sus operaciones con tecnología de IBM Argentina. 2013 [cited 2016 July 7th]; Available from: [https://www-03.ibm.com/marketing/ar/casos\\_de\\_exito/bayer/](https://www-03.ibm.com/marketing/ar/casos_de_exito/bayer/).
- [16] Liddle, A.J. Above-store reporting benefits highlighted by FS/TEC panel. 2007 August 13, 2007 [cited 2017 February 15th]; Available from: <http://www.nrn.com/corporate/above-store-reporting-benefits-highlighted-fstec-panel>.
- [17] Rice, V., HOUSE SPECIALTY: FRESH DATA. Eweek, 2001. 18: p. 51 - 52.
- [18] Hamanaka, K., El Pollo Nuevo. Orange County Business Journal, 2014. 37.

# Financial Model for Pricing Decisions Support in SMEs of the Restaurant Sector in Mexico

Jorge Luis Fernandez<sup>1</sup>, Cynthia B. Pérez<sup>1</sup>, Luis A. Castro<sup>1</sup>

<sup>1</sup>Dept of Computing and Design, Sonora Institute of Technology (ITSON), Ciudad Obregon, Sonora, Mexico  
(e-mail: jorgefdez@live.com; cynthia.perez@itson.edu.mx; luis.castro@acm.org)

**Abstract** - *Companies have always analyzed product prices. A good pricing strategy can help most businesses with profit margins. SMEs struggle with their pricing strategy, specially in Mexico where managers spent a lot of time deciding a product price. By taking advantage of sales and operational data, a SME could create a pricing strategy that helps it to maintain or improve customer sales along with the reduction of costs, time and improving profits. Hence, we propose a Financial Model based on input from small business owners which is included in a pricing decision support tool that analyzes costs, sales and production data, providing a vision of possible outcomes and also SMEs could visualize a What-if scenario helping managers in the decision-making process.*

**Keywords:** Pricing decision, SME, Financial Model

## 1 Introduction

One of the most important decisions of any company in terms of its sale and marketing strategy is the establishment of the price of the product or products offered, since it is one of the essential elements for the determination of profit. Within this decision, a price must be foreseen to ensure that production and sales numbers are in the same range as customer demand, as well as to what those customers are willing to pay for the good or service (Treviño Ayala, Villalpando Cadena, Treviño Ayala, & Lozano Treviño, 2013).

While making pricing decisions, one must keep in mind the concept of price and demand elasticity, thus considering the response of customers to the offered product in the face of a change in price (McConnell, C & Brue, S. 1997).

In many researches, it has been found that an increase in profits is based on an average price (O'Regan, Ghobadian, & Gallear, 2006, De la Garza, Ayub, Chein, & Banda, 2009), as well as a competitive and fair price (Arslan & Kivrak, 2008) of any good or service. An ideal price level should therefore be considered to determine the positive response of customers at a price high enough to achieve a profit, having covered costs, considering that customers will only pay the price if they consider that the good or service is worth it (Zikmund, W. 1998).

Depending on the industry, the market and its size, each company can consider different factors to establish their prices. In supermarkets, for example, we consider product

groups, wholesale costs, sales forecasts, promotions, market, sales strategies, among others, considering that these changes occur on a weekly basis (Montgomery, A. L. 2005). On the other hand, in the Hotel Industry, profit management systems are used to determine prices based on a forecast of supply and demand, as well as a comparison of characteristics, amenities, geographical position, between hotels, to mention but a few, (Kisilevich, Keim, Byshko, Tsibelman, & Rokach, 2011).

Although each industry may vary their price management processes, basic factors like fixed and variable costs are often the basis for pricing, especially in businesses such as Restaurants. Also, companies can use the information they generate in their sales systems to support decision making.

### 1.1. SMEs in Mexico

In Mexico, small and medium-sized enterprises (SMEs) generally lack some essential elements for their growth, either from professional support or financial knowledge, and focus only on seeking a return on their investment (ROI). Business Intelligence (BI) can have a great impact on the growth of any company, but it is often omitted in the operation of SMEs, due to lack of knowledge or because the managers do not consider BI necessary for their growth in the medium or long term. It is common for SMEs not to achieve profits or cannot continue operating, and may end up even ignoring what went wrong, and this is how 52.5% of SMEs in Mexico tend to go out-of-business before their 2-year mark, mainly due to their inability to be profitable (Flores Kelly, 2016).

To comply with customer requirements and market challenges, SMEs must continuously improve their products, services and/or processes, and must also incorporate or develop technological assets that allow them to position themselves ahead of their competition. In fact, it has been reported that most SMEs that only make new investments once the competitors get good results are destined to fail (Rubio Bañón y Aragón Sánchez, 2006).

The information generated by a SME can serve as a starting point for good and solid decision-making, but the culture that prevails in small and medium-sized enterprises in any industry leaves out important information. As Legendre (2005) mentions, "Unlike what many might think, SMEs can use the benefits of systematizing their Business Intelligence."

Accessibility to information systems in recent years in SMEs has provided challenges and opportunities for both owners and managers. The most accessible data in PyMEs is transactional, which can be captured through barcode readers or, in the case of restaurants, when registering customer order in the system. Therefore, it is possible to track each of the sales movements of any item through the use of this data, providing a wealth of information on the reaction of customers related to prices and promotions. Hence, it is promising when this information is analyzed to improve any pricing strategy (Montgomery, A. L., 2005).

Price and product, as well as the service offered, provide a competitive advantage within the markets in which each SME can participate. Each company intends to provide a differentiated service in each product or service offered to create that competitive advantage, and at the same time an optimum price that attracts a greater number of customers. A pricing strategy for any differentiated service in a competitive market is highly important since, regardless of the quantity of products or services, the price needs to be the most appropriate, and at the same time it must be periodically revised (Cassaigne, N., & Singh, M. G., 2001).

## 1.2. Costs in the restaurant sector in Mexico

In restaurants, the prices are usually influenced by the competition and the target market. However, the basis of the unit cost of each product has its origin in the cost of the raw materials used, based on its yield and the its cost of preparation. Also, the fixed costs of operation must be considered when calculating the proportional cost of each product, and based on this and the desired profit we can generate a price in line to the ideal demand and the market, taking in consideration any applicable taxes. It is also important to consider any price adjustments in order to avoid a high difference between products that could cause a high flow of "cheap" products that affect the sales of "expensive" products, which usually provide that differentiation that is sought in the market in which we are competing (Cuevas, J. F., 2016).

Commonly, pricing decisions in SMEs in the restaurant industry in Mexico are carried out by hand, using only a calculator to sum up the average input costs per unit and operating costs based on the financial statements, as well as the desired profit without "punishing" the customer or the products profit margin. In addition, external factors such as the competition's quality and prices are taken into account, as well as changes in tax or fuel costs that directly impact the finances of the company. This situation is not uncommon among SMEs in this sector, so it is necessary to evaluate the way in which the analysis of information, generated and captured daily in a system, can be used to make pricing decisions at the managerial level (Legendre, 2005).

Carrying out a pricing decision can result in an exhaustive task that leads to consuming more time than necessary, because it can be interrupted at different times,

making it tedious, highly time consuming and prone to errors. Based on this, the following question arises:

How could the analysis of management-level information on the raw materials, product preparation and operation costs of a SME in the restaurant sector be enabled, to support pricing decisions?

The objective of this research is to design a Financial Model that allows the analysis of the cost costs related to the operation and production areas, whether a cost is fixed or variable. Thus, the aim is to carry out a price and profit valuation showing different options to the user, based on trends, that allow a more accurate and agile pricing decision within SMEs of the restaurant sector in Mexico.

## 2 Financial Model

One of the challenges of any company is the optimal management of financial resources, since all efforts and processes need to be expressed in a solid and structured way that reflects the performance attained. Likewise, it is extremely important to plan and implement said process (Pastor, 2009).

Based on the above, a need for a Financial Simulation Model to support the decisions of management in business processes was found, specifically for product pricing. This model is planned to be implemented in a business intelligence system so that the person in charge can visualize the information through dashboards. This business intelligence tool can be used online, providing a competitive advantage for the company as it will be able to visualize its information from any mobile device at any time, as long as it has internet access. Also, the proposed model can be easily replicated for different products, branches and even companies of the same industry.

### 2.1 Input and Output Analysis

A Financial Model would help to define and analyze the variability of the prices of the products. It would also be able to offer an optimum perspective on profit generation based of the price of each product and its sales, as well as the approximate cost by volume based on tendencies.

To carry out a financial model, it is necessary to identify the basic concepts that comprise it, such as the inputs or the elements on which it works, and the outputs or results that the model must produce. This stage is known as the generation of the Black Box of the model, according to Eppen (Eppen, Gould, Schmidt, Moore, & Weatherford, 2000). Initially the logical relations between each input and output element within the model are unknown. Despite its simplicity, this black box causes the following to be considered from the beginning of the process: what should be included, what should be excluded and the classification of the different factors that make it up.

Based on the above, the input data is defined, as well as the desired output and, from this, a Black Box diagram is generated, see Figure 1.

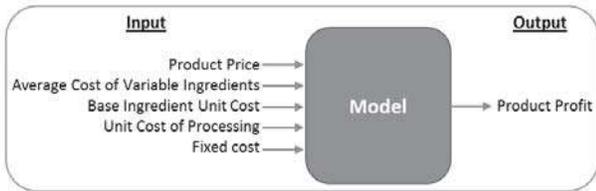


Figure 1. Black Box of the Financial Model.

## 2.2 Graphic Model Design

According to Eppen, et al. (2000) to define the logical relations, a very useful technique of model construction from data can be used, in which the desired relationships must be illustrated and it is not necessary to begin with the final mathematical equation. However, the graph would help us to deduce an equation for the final model. In this regard, using the same graphic model design technique it is possible to perform a primary data analysis, useful for the estimation of parameter values.

In order to have a clear definition of the relationships between the input and output data, it is advisable to generate the model as a graphic. Initially, the performance measure is chosen, in this case the Profit is decomposed into Revenues and Total Costs, see Table 1, and these in turn are decomposed into each of their elements until they reach the originally defined input elements. There are no strict rules for detailing the graphic model variables, since their sole purpose is to assist in the construction of the model more easily, not in its final definition (Eppen, et al., 2000). Once the model is constructed graphically, the relationships between each basic element, such as price and costs, and its evolution to profit can be observed, see Figure 2. To identify the price as the data *manipulated* by the user, it is marked in a different way than the rest of the variables within the graph, since this is the initial input data of the model. This data can be changed by the user in order to obtain different results that allow a better view of the possible costs and profit that could result from the selected price.

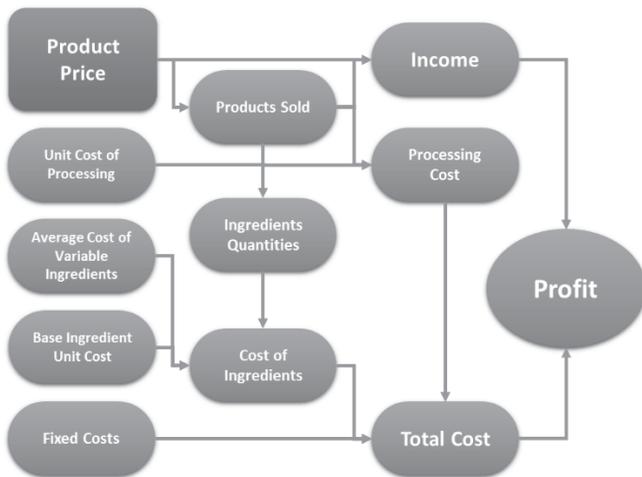


Figure 2. Graphic Model with relationships.

## 2.3 Equations

Once the model is constructed, the basic equations that indicate the relationship between the variables can be defined for the calculation of Profit, see Table 1. These equations are the basis for building the Financial Model since the Profit is calculated using this information. Hence, each equation solves a variable and is placed in reverse order in the table starting with Profit (output) and ending with the input variables, see Table 1. This way, Figure 2 illustrates the relationships between the variables used for the profit calculation.

Table 1. Equations Table

Variable	=	Equation
<b>Profit</b>	=	Income – Total Cost
<b>Income</b>	=	Product Price x Products Sold
<b>Total Cost</b>	=	Processing Cost + Cost of Ingredients + Fixed Costs
<b>Cost of Ingredients</b>	=	(Variable Ingredients Quantity x Average Cost of Variable Ingredients) + (Base Ingredient Unit Cost x Base Ingredient Quantity)
<b>Processing Cost</b>	=	Products Sold x Unit Processing Cost

## 3 Case Study

### 3.1 General info

This case study is carried out in a SME of the restaurant sector in Mexico. This company has 4 branches in the same city with about 100 full time employees. The growth of the company over 11 years of operation has been carried out informally, finances have historically been handled on paper, and sometimes digitally through spreadsheets. After 6 years of operation, a Point of Sale (POS) software with several management features was implemented, but it is only used to accelerate the process of capturing customer orders. None of the remaining business processes are currently being managed through the existing system, despite having the functionalities to do so.

### 3.2 Price/Sales Limit Equation

In order to generate the proposed Financial Model, an equation was defined together with the business owners, where it is assumed that for each \$10 pesos between prices, there would be a difference of 200 daily product sales, based on their experience. At higher prices, they have experienced less demand or fewer products sales. For this matter, it was defined by business owners that if the product costs \$100 pesos, it would not sell. Thus, the data presented in Figures 3, 4 and 7 are test data for model elaboration purposes, this data was also provided by the business owners and managers. This equation is defined as follows:

$$Products\ Sold = 2000 - (20 \times Product\ Price) \quad (1)$$

Once these parameters were defined, the base model was elaborated using a spreadsheet. Hence, the current price used in the base model was \$75 pesos, see Figure 3.

Financial Model (Daily Sales)	
<b>Decision Variable</b>	
Product Price	\$ 75.00
<b>Parameters</b>	
Average Cost of Variable Ingredients (\$ per Product)	\$ 20.00
Base Ingredient Unit Cost (\$ per Product)	\$ 15.00
Unit Cost of Processing (\$ per Product)	\$ 13.00
Fixed Costs	\$ 1,000.00
<b>Sold Products Equation</b>	
Intersection	2000
Slope (linear coefficient)	-20
<b>Cost Processing Equation</b>	
Linear Coefficient	
Quadratic Coefficient	
<b>Physical Results</b>	
Products Sold	500
<b>Financial Results</b>	
Income	\$37,500.00
Processing Cost	\$10,000.00
Cost of Ingredients	\$14,000.00
General Costs	\$ 1,000.00
Total cost	\$25,000.00
Profit (before taxes)	\$12,500.00

Figure 3. Base Financial Model.

### 3.3 Processing Cost Trend

When analyzing processing costs, it was found that the sold products equation results matched with the branch's current numbers for the amount of products sold. However, when comparing different quantities of sold products, the branch's numbers did not match the ones showed by the equation. To adjust this mismatch, a comparison was made with current data on the cost of processing for different levels of production, see Figure 4.

Cost Data (Thousands of Pesos)

Product Quantity (00)	Processing Cost (Real)	Processing Cost (Model)
2	\$ 6.00	\$ 7.00
3	\$ 8.00	\$ 9.00
4	\$ 10.00	\$ 11.00
5	\$ 12.00	\$ 13.00
6	\$ 15.00	\$ 15.00
7	\$ 18.00	\$ 17.00
8	\$ 21.00	\$ 18.00
9	\$ 24.00	\$ 20.00
10	\$ 27.00	\$ 22.00

Figure 4. Comparison Chart of Processing Costs.

In order to obtain a Financial Model that is close to reality, the comparative chart was graphed, see Figure 4, and a trendline was applied to adjust the information to the current data. From this graph, a cost-processing equation was obtained. When applying the trend line between cost of processing and the number of products, a polynomial cost processing equation was chosen for use, which automatically appears on the same chart, see Figure 5.

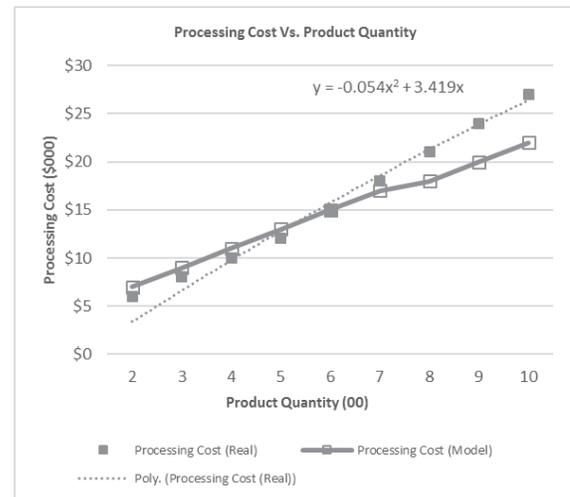


Figure 5. Trend of Processing Costs.

From Figure 5 the polynomial cost processing equation was obtained, see Equation 2. From it, the Linear Coefficient (x) and Quadratic Coefficient (x<sup>2</sup>) are extracted and are added to the proposed model in order to obtain a real trend of processing costs.

$$y = -0.054x^2 + 3.419x \tag{2}$$

### 3.4 Linear and Quadratic Coefficients

Both coefficients are applied to the model, replacing the Processing Cost previously considered. This gives the financial model accuracy on price calculation, profits and products sold. This projection becomes closely similar to the current numbers generated at the case study branch, since the calculations are based on the actual cost trend, calculated from the linear coefficient and the quadratic coefficient, see Figure 6.

Financial Model (Daily Sales)	
<b>Decision Variable</b>	
Product Price	\$ 75.00
<b>Parameters</b>	
Average Cost of Variable Ingredients (\$ per Product)	\$ 15.00
Base Ingredient Unit Cost (\$ per Product)	\$ 10.00
Fixed Costs	\$ 1,000.00
<b>Equation Coefficients</b>	
<b>Sold Products Equation</b>	
Intersection	2000
Slope (linear coefficient)	-20
<b>Cost Processing Equation</b>	
Linear Coefficient	3.419
Quadratic Coefficient	0.054
<b>Physical Results</b>	
Products Sold	500
<b>Financial Results</b>	
Income	\$ 37,500.00
Processing Cost	\$ 15,209.50
Cost of Ingredients	\$ 12,500.00
General Costs	\$ 1,000.00
Total cost	\$ 28,709.50
Profit (before taxes)	\$ 8,790.50

Figure 6. Financial Model adding Coefficients.

Financial Model (Daily Sales)								
Decision Variable								
Product Price	\$ 65.00	\$ 70.00	\$ 75.00	\$ 80.00	\$ 85.00	\$ 90.00	\$ 95.00	
Parameters								
Average Cost of Variable Ingredients (\$ per Product)	\$ 15.00	\$ 15.00	\$ 15.00	\$ 15.00	\$ 15.00	\$ 15.00	\$ 15.00	
Base Ingredient Unit Cost (\$ per Product)	\$ 10.00	\$ 10.00	\$ 10.00	\$ 10.00	\$ 10.00	\$ 10.00	\$ 10.00	
Fixed Costs	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	
Equation Coefficients								
Sold Products Equation								
Intersection	2000	2000	2000	2000	2000	2000	2000	
Slope (linear coefficient)	-20	-20	-20	-20	-20	-20	-20	
Cost Processing Equation								
Linear Coefficient	3.419	3.419	3.419	3.419	3.419	3.419	3.419	
Quadratic Coefficient	0.054	0.054	0.054	0.054	0.054	0.054	0.054	
Physical Results								
Products Sold	700	600	500	400	300	200	100	
Financial Results								
Income	\$ 45,500.00	\$ 42,000.00	\$ 37,500.00	\$ 32,000.00	\$ 25,500.00	\$ 18,000.00	\$ 9,500.00	
Processing Cost	\$ 28,853.30	\$ 21,491.40	\$ 15,209.50	\$ 10,007.60	\$ 5,885.70	\$ 2,843.80	\$ 881.90	
Cost of Ingredients	\$ 17,500.00	\$ 15,000.00	\$ 12,500.00	\$ 10,000.00	\$ 7,500.00	\$ 5,000.00	\$ 2,500.00	
General Costs	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	
Total cost	\$ 47,353.30	\$ 37,491.40	\$ 28,709.50	\$ 21,007.60	\$ 14,385.70	\$ 8,843.80	\$ 4,381.90	
Profit (before taxes)	\$ (1,853.30)	\$ 4,508.60	\$ 8,790.50	\$ 10,992.40	\$ 11,114.30	\$ 9,156.20	\$ 5,118.10	

Figure 7. "What-If" Analysis

With these adjustments applied to the model, it is possible to observe a profit more in line with current financial statements from the branch, based on the product price provided by the business owners.

### 4 Results

Once the model is elaborated, an analysis of different prices can be carried out by means of a projection with "What-If" scenarios, which provide a framework for comparing the possibilities of profits, based on a price range, see Figure 7.

Thanks to this analysis, business owners can compare the behavior of each of the possible prices, including the current price (\$ 70.00 pesos). Each option shows every parameter considered, fixed and variable, and the financial impact they

have on the results, including the assumption that a higher price would represent a smaller number of products sold. This analysis also shows the sales income behavior and the fluctuation in the costs for the preparation of each sold product. With this, the entrepreneur can quickly observe that the optimal profit range is between \$80 and \$85 pesos, by showing the projected profit for each of these prices.

Finally, a graph is generated showing the behavior of each element analyzed by the business owners. Here, the "What-If" analysis shows the break-even point at the intersection of the Profit line with the X-axis of prices, which is an approximate price of \$ 66.30 pesos, and the optimum price predicted at the highest point of the same graphed profit line, which again can be considered between \$80 and \$85 pesos per product, see Figure 8.

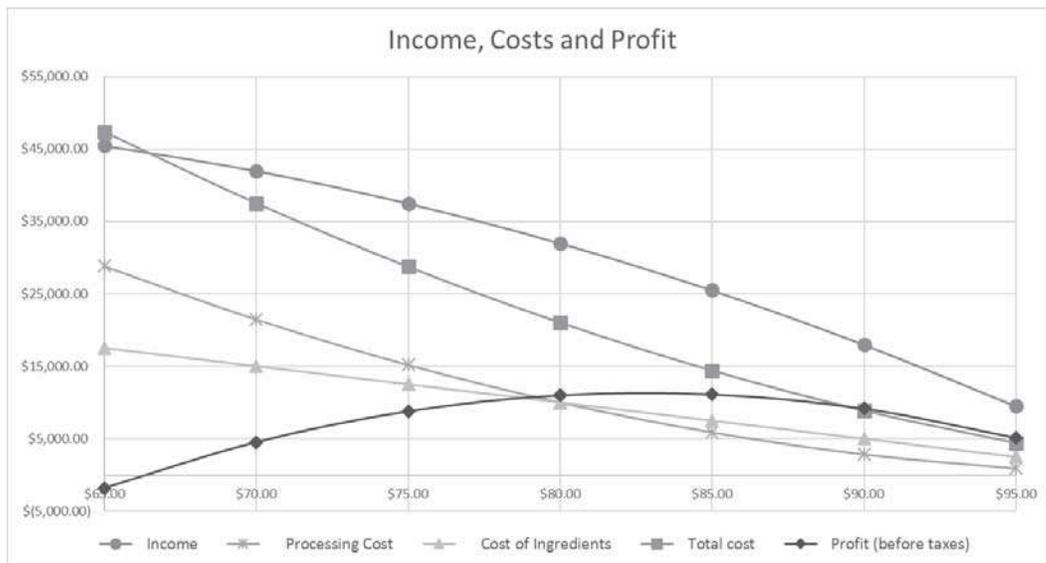


Figure 8. Price, Cost and Profits Analysis.

## 5 Conclusions

In this contribution, we have presented a Financial Model that offers a clear and concise view of the possible scenarios to which a SME may be exposed by making changes in the prices of its products. This model was applied in a case study related to a restaurant in Mexico showing a “What-if” scenario using a spreadsheet, helping business owners in the decision-making process and also saving time while reducing uncertainty. Thanks to the model provided, business owners were able to create pricing and promotion strategies faster and easier. Having different pricing and profit scenarios helped them save time when deciding new prices, and also to evaluate the current ones.

The Financial Model can be modified to suit different SMEs. Easy-input controls should be added to avoid undesired manipulation of the data within the spreadsheet. Thus, we are working on the development of a business intelligence system where the proposed Financial Model will be integrated, helping business owners in the decision-making process anywhere, anytime as long as they are connected to the internet. The idea is to provide through this system not only the ability to perform a constant analysis of profit behavior but also allowing business owners to make price adjustments without having to wait for monthly or quarterly results in order to make a decision.

## 6 References

- [1] Kelly, J. F. (2013, March 20). México necesita menos Pymes. Retrieved July 13, 2016, from <http://www.forbes.com.mx/mexico-necesita-menos-pymes/>
- [2] Rubio Bañón, A., & Aragón Sánchez, A. (2006). Competitividad y recursos estratégicos en las pymes. (Spanish). *Revista De Empresa*, (17), 32-47.
- [3] Legendre, R. (2005). ¿Es la Inteligencia de Negocio aplicable a las PYMEs? *PUZZLE*, Año 4, Edición No 17(Mayo-Junio), 4–9.
- [4] Montgomery, A. L. (2005). The implementation challenge of pricing decision support systems for retail managers. *Applied Stochastic Models in Business and Industry*, 21(4–5), 367–378. <http://doi.org/10.1002/asmb.572>
- [5] Cassaigne, N., & Singh, M. G. (2001). Intelligent decision support for the pricing of products and services in competitive consumer markets. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 31(1), 96–106. <http://doi.org/10.1109/5326.923272>
- [6] Cuevas, J. F. (2016). Control de Costos y Gastos en los Restaurantes. México D.F., México Limusa
- [7] Treviño Ayala, M. E., Villalpando Cadena, P., Treviño Ayala, R. A., & Lozano Treviño, D. F. (2013). La mercadotecnia en las PYMES y su influencia en el crecimiento de utilidades. *Innovaciones de Negocios*, 19(10), 125–144.
- [8] Zikmund, W. (1998). Investigación de mercados. México, D.F: Pearson Educación.
- [9] McConnell, C & Brue, S. (1997). *Economía*. México, D.F McGrawHill.
- [10] O'Regan, N., Ghobadian, A., & Gallea, D. (2006). In search of the drivers of high growth in manufacturing SMEs. *Technovation*, 26(1), 30-41.
- [11] Arslan, G., & Kivrak, S. (2008). Critical factors to company success in the construction industry. *Proceedings of World Academy of Science*, 35, 405-408.
- [12] Kisilevich, S., Keim, D., Byshko, R., Tsibelman, M., & Rokach, L. (2011). Developing a price management decision support system for hotel brokers using free and open source tools. In *ICEIS* (pp. 147-156).
- [13] Pastor, R. A. T. (2009). Modelo de gestión financiera para una organización. *Perspectivas*, (23), 55–72.
- [14] Eppen, G. D., Gould, F. J., Schmidt, C. D., Moore, J. H., & Weatherford, L. R. (2000). Investigación de operaciones en la ciencia administrativa: construcción de modelos para la toma de decisiones con hojas de cálculo electrónicas. México; Prentice Hall Hispanoamericana.

# Implementación de un sistema de almacén de datos para preservación de datos de origen climático

Mitchel Paola Hermosillo Perea, Luis A. Castro, Luis-Felipe Rodríguez

Instituto Tecnológico de Sonora (ITSON)  
Ciudad Obregón, Sonora, México  
mitchel.hermosillo@gmail.com, luis.castro@acm.org  
luis.rodriguez@itson.edu.mx

**Resumen.** En este artículo se explica el avance de proyecto de desarrollo de un almacén de datos como solución a problemas de almacenamiento de datos climáticos que generan sensores dispuestos en distintas localizaciones. Con la implementación del almacén se espera la mejora de los tiempos en extracción de la información y la creación de reportes, proveyendo a la organización de un punto de referencia singular para consultar su información. Actualmente los datos generados por parte de los sensores se encuentran dispersos en distintas partes de la organización, lo que ocasiona que aumente la posibilidad de pérdida u omisión de estos al momento de crear reportes.

**Palabras clave:** Datos, Almacén de datos, ETL, Almacén, tecnologías de información, TI, sensores.

## 1 Introducción

Existen distintas definiciones de almacén de datos (*data warehouse*, en inglés). La primera definición de Bill Inmon [1], considerado el padre de los almacenes de datos, define el almacén de datos como plataformas orientadas a temas, que cambian con el tiempo, donde se guarda información no volátil, y que integra información de diversas áreas de la organización. Por otro lado, Ralph Kimball otro autor dedicado al tema define el almacén de datos como una copia de datos específicamente transaccionales estructurados para consultas y análisis [2].

Dentro de las actividades que se realizan en las organizaciones, se encuentran el diagnóstico, pronóstico y toma de decisiones, las cuales son sustentadas con la información generada en sus bases de datos principales. Es importante recalcar que la clave para que se puedan tomar decisiones bien fundamentadas, es la calidad de los datos que están en las bases de datos operacionales. Sin embargo, para la toma de decisiones, en muchos casos, se utiliza un bajo porcentaje de la información disponible en bases de datos, ya que existen muchos más datos de los cuales sacar provecho para resolver otro tipo de problemáticas y no son aprovechados de manera adecuada.

Hay ejemplos citados en la literatura donde se toman en cuenta muchos datos a analizar. En [8], aplicado a ciudades inteligentes, se toma en cuenta que los datos que producen los sensores están conectados en una red extensiva de cosas. De manera similar, en [9] su perspectiva es la selección y agregación de múltiples fuentes de datos.

Sin embargo, en comparación con los ejemplos anteriores, en el caso de este trabajo, los datos se obtienen de manera manual, además de que se cuentan con millones de datos que no son propiamente almacenados. En este trabajo de tesis, muchos de los datos con los que se trabaja son generados de forma automática por sensores que miden ciertas variables de la naturaleza. Muchos de estos se encuentran instalados en áreas protegidas donde no existe una red inalámbrica que los comuniquen con un servidor externo. Uno de los elementos importantes a tomar en cuenta es que los sensores generan datos de forma constante.

De manera adicional, en este trabajo, uno de los problemas actuales es que los datos se recolectan directamente desde la reserva natural, ya que es necesario revisar el estado de los sensores y extraer la información por medio de una memoria USB para posteriormente replicarla en distintas computadoras para ser analizada y formateada. Esto debido a reglamentaciones federales que impiden uso del espectro radioeléctrico para transmisión de señales inalámbricas en dichas áreas. Como es de esperarse, esto puede representar un problema para la recolección y la integración de los datos. Los sensores miden distintas variables de comportamiento de la naturaleza que pueden ser: Temperatura, precipitación, humedad, aire, luz, etc. Esta información es capturada a 5 Hertz. De igual forma, otro de los problemas es que los sensores generalmente son de diferente fabricante, y proveen datos en distintos formatos. Debido a que los datos están dispersos en múltiples archivos, se tiene que realizar mucho trabajo de pre-procesamiento antes de tener listo un set de datos con el cual trabajar.

En el caso particular de este trabajo de tesis, como se comenta, los datos se encuentran dispersos en distintas ubicaciones provocando que no estén integrados, sean duplicados o a veces omitidos. Una de las ventajas de contar con un almacén de datos climáticos es que los datos que se tengan dispersos en distintas fuentes se encontrarán en un solo lugar, dando como resultado que estos no lleguen a omitirse o perderse y por supuesto disminuir el tiempo de espera en la búsqueda de información.

## **2 Trabajos relacionados**

En la literatura se encuentran distintos tipos de trabajos relacionados al desarrollo e implementación de almacenes de datos [3, 4, 5, 7,10,11]. En ese sentido, un almacén puede ser creado con cualquier tipo de necesidad siempre y cuando se cuente con datos dentro de la organización.

El primer ejemplo es el de un almacén de datos para gestión en construcción [3] provee información a los gerentes de construcción sobre datos existentes para tomar decisiones de forma eficiente sin interrumpir el trabajo diario de una base de datos organizacional que se está utilizando en tiempo real. Uno de los principales retos que se presentaron mientras se diseñaba el almacén de datos fue el determinar cuáles serían

los datos que se cargarían en él, por lo que se determinaron dos acercamientos. El primero basado en necesidades y el segundo basado en la disponibilidad de los datos, ya que por cuestiones de costos en tiempos es recomendable cargar información útil para el futuro que volver a recolectar los datos después.

Otros trabajos que se han hecho se relacionan los almacenes de datos en el área clínica para cuidado de la salud [4] el proyecto CATCH provee métodos sistemáticos para la comunidad como indicadores para evaluar el proceso de producir una lista ordenada y calificada de retos críticos en cuanto a la salud de la comunidad en que se aplique. Así como en el proyecto de construcción uno de los retos mencionados en el proyecto CATCH fue la disponibilidad de los datos particularmente en este proyecto fue identificar un nivel natural de agregación ya que los datos venían en grandes cantidades y de distintas fuentes y por supuesto lograr estándares altos de calidad de datos que apoyen a las políticas de formulación para asistencia médica.

Por otro lado, se encuentra el proyecto eSagu donde se implementa un almacén de datos para agricultura [5] utilizado para mejorar el desempeño y la utilización de tecnología de agricultura para evitar visitar en persona los campos de cultivo, obteniendo estatus, fotografías digitales y otro tipo de información. Como resultado de la investigación del proyecto [5] se presentaron los siguientes retos: habilitar a los expertos en agricultura a mejorar la calidad y eficiencia de la entrega de consejos, para ayudar a los científicos sobre el tema a preparar sugerencias de alta calidad. Otro reto muy importante es habilitar el acceso a distintos tipos de lenguaje y traducción automática ya que se poseía solo lenguaje local y finalmente presentar las variaciones de los distintos cultivos y sus plagas.

Otro trabajo relacionado es el proyecto “Discovery Net”, el cual provee un ejemplo donde diferentes usuarios pudieron desarrollar sus propios flujos de trabajo con colecciones de datos especificando cómo los datos de un sensor se procesarán antes de guardarlos en el almacén de datos [7]. Además, en [7] se presentan algunos retos cuando una red de sensores está siendo construida para medir la contaminación del aire, uno de ellos es utilizar técnicas estandarizadas que permitan la seguridad y la autenticación relacionados con el acceso y el control de los sensores.

Como parte de la implementación del data warehouse incluye entre sus pasos el desarrollo de un proceso de extracción, transformación y carga de datos (ETL). En [10] se presenta un ETL como un tratamiento previo a los datos, que ayuda a mejorar la consistencia de los mismos, los desafíos más remarcables de este proceso es su alto costo ya que este proceso tiene que adaptarse a ciertas características de los datos y filtraje de correcciones.

Según la literatura actual existe un nuevo enfoque en cuanto implementación de data warehouse enfocado a evaluar indicadores cualitativos; en [11] se hace hincapié en que es usual encontrar datos de naturaleza difusa, los objetivos de la implementación es implementar un data warehouse difuso donde permite realizar un análisis cualitativo de datos de estudiantes y además que apoye a la toma de decisiones. Además, en [11] mencionan que el trabajo ayudo a definir adecuadamente ciertos parámetros que obtienen buenos resultados a las consultas, también como en calidad de datos menciona que los parámetros no son universales más bien dependen del contexto.

Aun cuando se han diseñado e implementado almacenes de datos en diversos dominios de aplicación, hay retos asociados en cada una de esas áreas. En particular, debido a que el diseño debe estar basado en requerimientos que se especifican con el cliente. Esto, puede ser un problema grande si los requerimientos no son demasiado claros.

### **3 Objetivo de la investigación**

Diseñar e implementar un sistema de almacén de datos para contar con un punto de referencia en la organización para la extracción de la información que viene desde sensores instalados en áreas protegidas naturales de forma integrada, formateada, actualizada y lista para la creación de reportes o tableros de control que permitan tomar decisiones adecuadas en tiempo y forma.

#### **3.1 Objetivos específicos**

- Investigar de metodologías para el desarrollo de almacén de datos para selección de la mejor opción adaptada a las necesidades del usuario.
- Analizar datos de las distintas fuentes de datos para el diseño del almacén de datos.
- Diseñar un almacén de datos que permita el fácil y rápido acceso a la información.
- Evaluar el almacén de datos para verificar su correcto funcionamiento.

### **4 Metodología**

En esta sección se presenta la metodología que se sigue en el desarrollo de este proyecto de tesis. Utilizando los siguientes puntos se puede llevar a la conclusión del proyecto de tesis, explicando que se hará por cada punto.

1. Levantamiento de requerimientos. Se llevarán a cabo distintas entrevistas semi-estructuradas con los interesados en la implementación del almacén de datos, donde se establecerán las necesidades principales y las limitaciones del diseño.
2. Implementación de servidor con espacio suficiente de almacenaje. Para llevar a cabo la implementación del almacén de datos se necesita un espacio de almacenaje en línea que se definirá por parte del departamento de recursos naturales, dependiendo de la cantidad de datos a analizar.
3. Creación del proceso de carga de datos. Como parte de la implementación del almacén de datos, se codificarán procesos que permitan la automatización de las actualizaciones de la información.
4. Diseño de almacén de datos. Dependiendo de los acuerdos que se realicen en el levantamiento de requerimiento se definirá la metodología del diseño de almacén de datos, las disponibles son las siguientes:
  - (1) Diseño de almacén de datos para organizaciones [1]
  - (2) Diseño de modelo dimensional [2]
  - (3) Diseño ágil de almacén de datos [6]

5. Evaluación de diseño de almacén de datos: En esta etapa, se evaluará la efectividad del almacén de datos basados en las técnicas utilizadas en las metodologías. se realizarán pequeños tableros que muestren la información totalizada y detallada que compruebe la veracidad de los datos. De manera adicional, se analizará la utilidad percibida de los datos del almacén. Esto es, se validará el almacén con base a la respuesta de los usuarios llevando a cabo una evaluación de calidad de datos basada en la utilidad de los mismos.

## 5 Avances en la Investigación

Este es un proyecto de maestría, en el cual se inicia con el segundo año en el que se estará trabajando con una organización que obtiene múltiples datos de sensores distribuidos en el estado de Sonora, México.

Tras el desarrollo de la investigación se esperan tres principales beneficios: a) Reducción en tiempo de respuesta en la creación de reportes, ya que actualmente los reportes que se generan con los datos crudos son tardados de hacer, ya que los datos tienen que ser formateados y analizados a mano; b) Preservar datos limpios e integrados sin posibilidad de pérdidas, debido a que se pretende crear un proceso que permita de forma automática la limpieza de los datos y su formato para descartar errores que provengan de los sensores; y, c) tener un almacén de datos actualizado según requerimientos de usuario. Esto, debido a que el proceso que se creará para formato y limpieza se dispondrá para actualizaciones por intervalos de tiempo para mantener la información actualizada en todo momento para su consulta.

Hasta el momento, se nos proporcionan datos provenientes no solo de sensores si no de forma satelital que vienen de la plataforma LDAS (*Land Data Assimilation Systems*) que proporciona sets de datos con las mejores observaciones disponibles sobre precipitación, superficie meteorológica, flujos y cobertura de nieve. Se están analizando las dimensiones de cada uno de estos datos, así como la manera de integrar el proceso de extracción de las fuentes de datos.

De manera simultánea, se diseñó un instrumento que permite medir el grado de utilidad de los datos que estarán depositados en el almacén. Este instrumento consiste de una serie de reactivos que se utilizarán para medir la percepción del grado de utilidad de los datos que provee el almacén de datos, uno de los indicadores a evaluar sobre utilidad de datos fue la aplicabilidad de los mismos donde se cuestiona la forma en la que los datos están representados dentro de un data warehouse o en una base de datos operacional. Además, también se cuestiona sobre las soluciones que proveen los datos para la organización entre otros reactivos. El instrumento se validó primeramente con tres expertos que tienen entre 10 a 25 años de experiencia con datos y que han sido líderes de proyectos de software. La validación consistió en retroalimentación sobre cada uno de los reactivos, mismo que fueron incorporados al instrumento. Posteriormente, se aplicó a 30 participantes de distintas empresas, por ejemplo, SOA Software, ENI Networks, MAPCO, se reclutaron vía e-mail donde se consensuaron juntas en las empresas para explicar el contenido de los reactivos en caso de haber explicaciones adicionales para medir la fiabilidad de los ítems y del instrumento. Para la fiabilidad se

utilizó el alfa de Cronbach (0.838). Una vez verificada la confiabilidad, será aplicado a usuarios finales en etapas posteriores de este trabajo de tesis.

## 6 Conclusiones

En este artículo se presenta los inicios del proyecto de desarrollo de almacén de datos para resolver las necesidades de una organización que no cuenta con un punto de referencia para su información y que sus datos se encuentran dispersos en distintas fuentes de datos. Aun cuando estos están siendo utilizados, los costos en tiempo y el formato de los mismos son altos. En el futuro, con el instrumento diseñado, se medirá el grado en el que los datos que se encuentran en el almacén de datos son útiles para los usuarios para futuras mejoras en los procesos extracción y carga de los mismos. Con esto, se busca la disminución de tiempos de carga para aumentar el grado en el que estos están actualizados.

## 7 Referencias

1. W. Inmon, Building the Data Warehouse, John Wiley & Sons., 2002.
2. R. Kimball and M. Ross, The data warehouse toolkit: the complete guide to dimensional modelling, New York: John Wiley & Sons, 2002.
3. K. Chau, Y. Cao, M. Anson and J. Zhang, "Application of Data Warehouse and Decision Support System in Construction Management," *ELSEVIER*, vol. 12, no. 2, pp. 213-224, 2002.
4. D. Berndt, A. Hevner and J. Studnicki, "CATCH/IT: a data warehouse to support comprehensive assessment for tracking community health.," *Systems*, pp. 0114-8, 2003.
5. P. Krishna Reddy, G. Ramaraju and G. Reddy, "eSagu TM: A Data Warehouse Enabled Personalized Agricultural Advisory System," *ACM*, pp. 910-914, 2007.
6. L. Corr and J. Stagnitto, Agile Data Warehouse Design: Collaborative Dimensional Modeling, from Whiteboard to Star Schema, DecisionOne Consulting, 2011
7. Richards, M., M. Ghanem, M. Osmond, Y. Guo, and J. Hassard. "Grid-based analysis of air pollution data." *Ecological modelling* 194, no. 1 (2006): 274-286. Costa, C., & Santos, M. Y. (2017).
8. The SusCity Big Data Warehousing Approach for Smart Cities. In Proceedings of the 21st International Database Engineering & Applications Symposium on - IDEAS 2017 (pp. 264–273). New York, New York, USA: ACM Press. <https://doi.org/10.1145/3105831.3105841>
9. Jara, A. J., Bocchi, Y., & Genoud, D. (2013). Determining human dynamics through the Internet of Things. In Proceedings - 2013 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IATW 2013 (Vol. 3, pp. 109–113). IEEE. <https://doi.org/10.1109/WI-IAT.2013.161>
10. Duque Méndez, N. D., Hernández Leal, E. J., Pérez Zapata, Á. M., Arroyave Tabares, A. F., & Espinosa Gómez, D. A. (2016). Modelo para el proceso de extracción, transformación y carga en bodegas de datos. Una aplicación con datos ambientales. *Ciencia E Ingeniería Neogranadina*, 26(2), 95–109. <http://doi.org/10.18359/rcin.1799>
11. Zambrano Matamala, C., Sepúlveda, A. U., & Contreras, M. V. (2017). Análisis de rendimiento académico estudiantil usando Data Warehouse Difuso. *Ingeniare. Revista Chilena de Ingeniería*, 25(2), 242–254. <http://doi.org/http://dx.doi.org/10.4067/S0718-33052011000300007>

## Estudio Cualitativo sobre el Comportamiento del Consumidor en las Compras en Línea

**Kathya E. Mercado<sup>1</sup>, Cynthia B. Perez<sup>2\*</sup>, Luis A. Castro<sup>1</sup> y Adrian Macias<sup>1</sup>**

(1) Instituto Tecnológico de Sonora, Campus Náinari, Antonio Caso S/N y E. Kino, Colonia Villa ITSON. C.P. 85130. Ciudad Obregón, Sonora. México. (e-mail: [kathya.mercado@hotmail.com](mailto:kathya.mercado@hotmail.com); [luis.castro@acm.org](mailto:luis.castro@acm.org); [adrian.macias@itson.edu.mx](mailto:adrian.macias@itson.edu.mx))

(2) Instituto Tecnológico de Sonora, Unidad Guaymas, Carretera al Aeropuerto Km 3, C.P. 85400. Apartado Postal 870. Guaymas, Sonora. México. (e-mail: [cynthia.perez@itson.edu.mx](mailto:cynthia.perez@itson.edu.mx))

*Recibido May. 17, 2018; Aceptado Jul. 23, 2018; Versión final Sep. 14, 2018, Publicado Feb. 2019*

---

### Resumen

Se estudia las características principales del comportamiento del consumidor en plataformas de compra en línea en el estado sur de Sonora. Esto se hace mediante un estudio cualitativo basado en la metodología de la Teoría Fundamentada. La idea principal es identificar y categorizar variables sobre la compra en línea basándose en factores relevantes como la motivación de compra, preferencias, hábitos de consumo y patrones de compra. Para llevar a cabo el estudio, se realizaron entrevistas semi-estructuradas a consumidores del estado sur de Sonora, México, con experiencia en compras en línea. De esta manera, al concluir el análisis, se obtuvieron tres categorías principales: motivación de compra, experiencias del consumidor y comportamiento. También se identificaron una serie de variables que tienen mayor influencia en las preferencias, decisión de compra y el patrón de compra en los consumidores. Los resultados obtenidos proporcionan información de utilidad para la creación de estrategias que permitan tomar decisiones efectivas mediante su aplicación en el diseño de plataformas en línea ofreciendo un servicio personalizado a los consumidores.

*Palabras clave: análisis cualitativo; teoría fundamentada; compras en línea; comportamiento del consumidor; entrevista semi-estructurada*

## Qualitative Study of Consumer Behavior in Online Shopping

### Abstract

The main characteristics of consumer behavior in online shopping platforms at Sonora, Mexico have been studied. A qualitative study based on the methodology of Fundamental Theory is used to identify and categorize variables about online purchases based on relevant factors such as purchase motivation, preferences, consumption habits and purchasing patterns. Semi-structured interviews to consumers with experience in online purchases were conducted. In this way, three main categories were obtained: purchase motivation, consumer experiences and behavior. Also, some variables with a greater influence on preferences, purchase decision and consumer purchase pattern were identified. The obtained results provide useful information for strategies planning that allow making effective decisions using them in the design of online platforms with a personalized service to consumers.

*Keywords: qualitative analysis; grounded theory; online purchases; consumer behavior; semi-structured interview*

## INTRODUCCIÓN

El imparable desarrollo y expansión de las nuevas tecnologías de la información, en especial en el comercio electrónico ha causado el creciente interés en el comportamiento del consumidor en compras en línea (Zhou et al., 2007). Asimismo, la sociedad actual demanda más comodidad y menos tiempo en la compra de productos o servicios por Internet, por lo cual, están conscientes que es el medio más rápido para la búsqueda de información y muestra la importancia de que las empresas tengan cada vez más capacitación para mantener una efectiva presencia en Internet (Gallaughier, 1997; Grandon y Pearson, 2004; Barrera, 2017). El comercio electrónico consiste en la compra-venta de productos o servicios, en donde los consumidores obtienen información y adquieren productos a través de medios electrónicos (Olson y Olson, 2000). Por esta razón, Zarco y Ruiz (2006), plantean que la estructura comercial, el sistema de distribución y los canales de distribución han tenido considerables transformaciones en los últimos años, de modo que han sido ampliamente promocionados los beneficios potenciales del uso del comercio electrónico (Gefen y Straub, 2000). Sin embargo, para que se obtengan estos beneficios se necesita examinar la interacción del consumidor en el proceso de compra en una plataforma de compra en línea (Pavlou y Fygenson, 2006).

Hoy en día, el estudio del comportamiento del consumidor ha sido ampliamente estudiado desde diferentes perspectivas y áreas (Solomon et al., 2014). La mayoría de los estudios se han centrado principalmente en la acción de compra de productos o servicios y medir la satisfacción del consumidor, enfocándose en las experiencias posteriores a la compra de los consumidores con el fin de encontrar una relación con variables que enriquezcan más el conocimiento acerca del consumidor (Westbrook et al., 1991; Homburg y Annette, 2001). Por otra parte, existen varios estudios, que se centran en el análisis de datos tanto cuantitativos como cualitativos, para demostrar la relación entre la motivación de los consumidores y su comportamiento de compra (Joines et al., 2003; Goldsmith y Horowitz, 2006). Por ejemplo, Hausman (2000), probó que la decisión de compra de los consumidores es una consecuencia de los impulsos generados por la motivación a las recompensas que se obtendrán por realizar una compra. Asimismo, Karbasivar (2011), demuestra que existe una relación fundamental entre la publicidad de productos, las actividades promocionales (descuentos, productos o servicios gratuitos) y el comportamiento de compra del consumidor. Por lo cual, se afirma que el comportamiento del consumidor no es monolítico, puesto que, antes de realizar una compra los consumidores primero participan en la obtención de información de un producto o servicio y en otros factores que proporciona el sitio web. De modo que se plantean retos de investigación relacionados con el análisis y modelización del comportamiento del consumidor en las compras en línea, por ejemplo, 1) ¿Cuáles son los factores más relevantes para la comprensión del comportamiento del consumidor? Y 2) ¿Cómo se relacionan entre sí los factores encontrados?, usando los resultados como la base principal sobre la que se formulan estrategias organizacionales.

En este artículo, se presenta un estudio cualitativo bajo el enfoque de la Teoría Fundamentada para estudiar el comportamiento, actitudes y percepciones de los consumidores que compran productos y servicios en línea. Para esto, el objetivo del estudio es identificar las variables del comportamiento del consumidor de mayor relevancia en el proceso de decisión de compra en línea por medio del análisis de entrevistas de consumidores que realizan compras de productos y servicios en línea. Dentro de esas variables consideramos la interpretación y medición de la relación de los diferentes factores relevantes para las compras en línea, tales como la percepción de los productos, motivación de compra, preferencias, hábitos de consumo, influencias y comportamientos de compra. Asimismo, con la finalidad de identificar si el proceso de decisión de compras se ha visto alterado y cuáles han sido las variables del comportamiento del consumidor que han sido influenciadas. Los resultados obtenidos de esta investigación demuestran la importancia de 3 categorías para definir el comportamiento del consumidor en las compras en línea las cuales son: 1) la motivación de compra, 2) las experiencias, y 3) el comportamiento del consumidor. Además, se presenta cómo la motivación y la experiencia influyen en el comportamiento del consumidor.

## METODOLOGIA

La Teoría Fundamentada nos permite construir teorías, hipótesis y conceptos acerca de un fenómeno en particular, partiendo directamente de los datos mediante la utilización de un método comparativo constante (Glaser y Strauss, 1967). En este trabajo se realizó un estudio cualitativo para analizar e identificar las variables del comportamiento del consumidor de mayor relevancia en el proceso de decisión de compra.

### *Preguntas de investigación*

Este proyecto de investigación se basó en tres preguntas enfocadas a consumidores de productos y servicios en línea: P1. ¿Cuáles son las prácticas de compras por Internet que ejercen los consumidores?; P2. ¿Cuáles son los motivos por lo que los consumidores compran a través de Internet?; y P3. ¿Qué caracteriza la experiencia y el comportamiento de consumidores en sus compras por Internet?

### Informantes

El estudio se llevó a cabo en el sur del estado de Sonora, en México, donde participaron personas adultas con experiencia en compras en línea. Para el proceso de selección, los participantes debían cumplir los siguientes criterios de elegibilidad: residir en el sur del estado de Sonora, tener entre 18 y 45 años de edad, contar con experiencia en compras en línea, tener suficiente capacidad de destreza cognitiva, visual y manual para el uso de la computadora o dispositivo móvil, tener acceso diario a una computadora personal con conexión a Internet y contar con cualquier nivel ingreso, educación y ocupación. Para el proceso de selección de participantes, se realizó una breve entrevista telefónica a catorce personas para determinar la elegibilidad de los participantes. Al finalizar las entrevistas telefónicas, se seleccionaron once personas que cumplieron con los criterios de elegibilidad quienes se comprometieron a participar en el estudio, ver Tabla 1. En esta tabla se presentan los participantes seleccionados y sus características en cuanto a la edad, ocupación y sexo donde se seleccionaron 4 de las entrevistas realizadas, dos de estudiantes (P5 y P10) y dos profesionistas (P2 y P3).

Tabla 1: Descripción de participantes.

No. Participante	Ocupación	Edad	Sexo
P1	Estudiante de Posgrado	24	F
P2	Técnico en Electrónica	35	M
P3	Ejecutivo de Ventas	42	F
P4	Técnico de Sistemas	25	M
P5	Estudiante de Posgrado	23	F
P6	Estudiante de Posgrado	36	M
P7	Estudiante de Carrera	20	F
P8	Desarrollador Web	24	M
P9	Profesor de Español	33	M
P10	Estudiante de posgrado	27	M
P11	Contador Público	32	M

Con el objetivo de determinar un adecuado tamaño muestral para este trabajo de investigación, se empleó el método del muestreo selectivo basado en la Teoría Fundamentada donde la estructuración de la muestra de los participantes se realiza gradualmente a lo largo del proceso, por lo que el número y los rasgos de la población no se conocen *a priori*. Por ello, Glaser y Strauss (1967) definen que en este tipo de muestro la elección de individuos a estudiar se produce durante el proceso de interpretación de datos. Asimismo, Flick (2012) considera que el muestreo teórico puede partir de personas específicas donde la extensión de la población no se conoce, el tamaño de la muestra no está definido y el muestro acaba cuando se alcanza la saturación teórica, a diferencia del muestreo estadístico donde todo esto se conoce *a priori*. Glaser y Strauss (1967), definen como saturación teórica el criterio para determinar cuándo dejar de muestrear los distintos individuos, grupos o casos. Por lo tanto, se considera que se ha llegado a la saturación cuando no se hallan datos, o bien, cuando ya no emerge nada nuevo (Valles, 2009). De esta manera, la muestra no se obtiene a través de un muestreo aleatorio, sino que los individuos son seleccionados mediante las expectativas que generan en el aporte a ideas nuevas en relación con la teoría que se desea desarrollar. Adicionalmente, es importante considerar el tiempo disponible para llevar a cabo la investigación cualitativa debido a que el análisis de los datos cualitativos recolectados es muy riguroso y toma demasiado tiempo para llevarlo a cabo de forma manual. Para este estudio, se recolectaron once entrevistas de las cuales se analizaron cuatro, debido a que se encontró saturación teórica, tomando un tiempo de tres meses para la codificación y análisis de las cuatro entrevistas. Así mismo, con la codificación de los datos y el análisis de los mismos se consideró que la información recogida resultó suficiente en relación con el objetivo del estudio.

### Protocolo de entrevista

Los datos se obtuvieron de las entrevistas realizadas a los participantes. Cada una de las entrevistas semiestructuradas con preguntas abiertas fueron conducidas por el investigador con cada participante para documentar sus experiencias sobre las compras realizadas por Internet. Para esto, la entrevista se dividió en bloques de preguntas para facilitar el análisis, se obtuvieron diez bloques con diferentes temas a tratar: compras por Internet en general, acerca de su última compra realizada, compras en tiendas extranjeras, servicio de paquetería seleccionado, formas de pago utilizadas, tipo de dispositivos en los que realizan compras en línea, la confianza, la utilidad y control percibido a la hora de realizar compras en línea, la utilidad de la información proporcionada en los sitios de compras en línea, ver Tabla 2.

Tabla 2: Estructura de la guía de entrevista.

<i>Bloques de preguntas</i>	<i>Número de preguntas</i>
Acerca de compras en línea	9
Acerca de su última compra	9
Compras en tiendas extranjeras	8
Utilidad de los sitios de compra por internet	5
Control percibido	10
Envío de productos	6
Formas de pago	9
Tipo de dispositivos en las compras por internet	9
Confianza percibida	5
Utilidad de la información de un sitio de compra en línea	10

La entrevista consistió en llevar a cabo preguntas generales de inicio y posteriormente profundizar con preguntas cada vez más específicas como por ejemplo: “¿Con qué regularidad compras en línea?”, “¿Qué es lo que más valoras a la hora de comprar en línea?” y “¿Cuáles han sido tus experiencias en la compra en línea?”. Cada entrevista duró aproximadamente de 30 a 45 minutos, las cuales fueron transcritas para su posterior análisis. Una vez transcritas fueron codificadas de forma abierta acorde a la Teoría Fundamentada.

#### *Análisis de datos*

Se utilizó la técnica de Teoría Fundamentada, la cual consiste en un procedimiento sistemático para analizar y generar teorías a partir de datos cualitativos (Strauss y Corbin, 1998). Para Strauss et al. (2002), el propósito principal de la Teoría Fundamentada en los datos consiste en generar o encontrar modelos explicativos sobre determinados fenómenos, cuyos principios teóricos se encuentran apoyados en el análisis sistemático y posteriormente a la interpretación de los datos obtenidos de un conjunto de informantes mediante entrevistas. Por lo cual, se debe basar en el análisis comparativo constante entre las opiniones de los informantes en relación con la situación de cada pregunta con respecto a sus compras realizadas, con el fin de identificar patrones de recurrencia y sus relaciones. Como resultado del uso de esta metodología es la generación de teoría a partir de la inducción, de esta manera, ayuda a la explicación exacta del fenómeno estudiado aportando al investigador la habilidad de explicar un suceso. De esta manera, la aplicación de esta metodología permite crear teoría basada en la realidad, usando lo expresado por los informantes buscando mantener la esencia de las palabras que nos otorgaron.

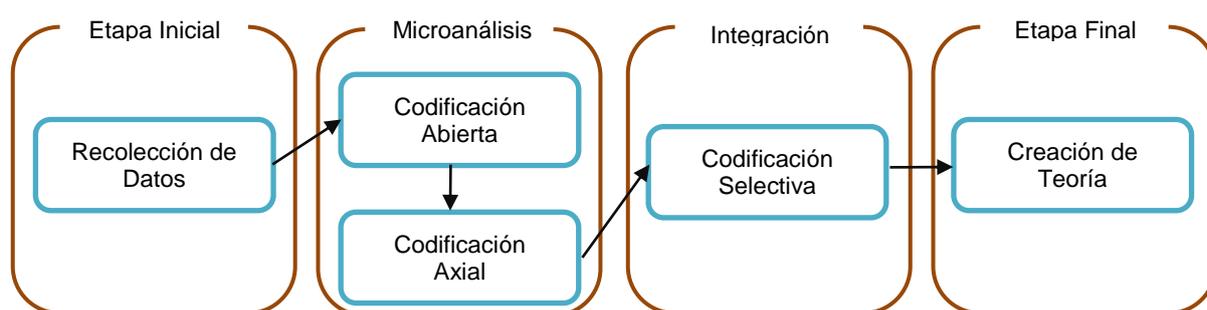


Fig. 1: Pasos de la Teoría Fundamentada

La metodología utilizada se basa en un modelo de cuatro etapas presentadas en la Figura 1. En este modelo se identifican las 4 etapas y 5 procesos a seguir para la creación de teoría. Primeramente en la etapa inicial se comienza con la identificación de área de interés a explorar y la selección de las fuentes de información, subsiguiente con la recolección de los datos. En la etapa de Microanálisis se comienza con el análisis de los datos y se van identificando que información relevante para el estudio, en esta etapa se realiza la codificación abierta y axial. La codificación abierta de los datos, se entiende como el análisis minucioso para identificar y conceptualizar los significados que el texto contiene. El resultado es una lista de códigos y al compararlos respecto de sus propiedades y dimensiones, se obtiene una clasificación, denominada categoría. Posteriormente, se lleva a cabo un análisis de codificación axial, en la cual, de acuerdo con Spiggle (1994), se identifican las relaciones entre las categorías obtenidas en la codificación abierta y sus subcategorías, esta relación está determinada por

las propiedades y dimensiones de las subcategorías y categorías que se quieren relacionar; por ejemplo, el código A causa el código B, el código A confirma el código B.

Una vez llevada a cabo la etapa de microanálisis, se procede a la etapa de integración donde se realiza la codificación selectiva, la cual, es una extensión de la codificación axial, pero con un mayor nivel de abstracción. Su propósito es obtener una categoría central que exprese el fenómeno de investigación para desarrollar una única línea narrativa con la cual todas las demás categorías relacionadas con la central estén cubiertas (Locke, 2001). Finalmente, en la última etapa se lleva a cabo el proceso de la creación de teoría, en la cual se empieza a delimitar las teorías emergentes y se formula una teoría con un grupo pequeño de conceptos de alta abstracción que permitirán construir conocimiento basado en las experiencias aportadas por los informantes. Esta metodología exige un laborioso trabajo de selección de contenidos relevantes, en función de lo expresado por los participantes, lo cual implica un alto nivel de abstracción donde el investigador no debe perder el sentido subjetivo que el informante transmitió.

#### Codificación abierta

Para llevar a cabo la codificación abierta de los datos obtenidos mediante las entrevistas, se analizaron detalladamente las transcripciones de éstas, línea por línea para identificar códigos y definir las categorías mediante la agrupación de conceptos. En la Figura 2, se muestra la codificación abierta de la transcripción de la entrevista del participante dos (P2). De esta manera, se detectan códigos abiertos sobre los datos recogidos, por ejemplo, en la respuesta del participante dos se identificó un patrón de navegación, al indicar la utilización de un buscador en los sitios de compra en línea, a su vez, manifiesta la necesidad de un catálogo en los sitios de compras en línea y se detecta un problema durante la navegación. Estos códigos seleccionados sirvieron posteriormente para la identificación de patrones e información importante para la siguiente fase.

154	E2	Yo utilizo el buscador y ya con eso me es suficiente pero no todos tienen
155		buscador eh! [Patrón de navegación] yo digo que deben tener un catálogo
156		que estén dividido por categorías que sea simple y sencillo y que también
157		tengan un buscador [necesidades en compras]. Porque es muy difícil
158		navegar cuando no tienen y solo los dividen por marca [Problemas: difícil
159		navegación]. Por ejemplo si yo entro a Privalia no le voy poner una "blusa
160		roja" no, o sea es lo que este actual las ofertas actuales y ver todo el
161		catalogo que tienen.

Fig. 2: Codificación abierta: rotulación a entrevista

#### Codificación Axial

En esta fase del proceso de análisis se identificaron las relaciones entre las categorías obtenidas en la codificación abierta, en la cual se determinaron la variedad de condiciones, acciones/interacciones y consecuencias asociadas al comportamiento del consumidor en compras en línea. Por lo cual se llevó a cabo la jerarquización en subcategorías que son relacionadas con las categorías obtenidas previamente, ver Tabla 3.

Esta Tabla presenta el análisis y el desarrollo de las categorías relevantes para el estudio del comportamiento del consumidor, la cual está conformada por cuatro columnas: categorías, subcategorías, propiedades y dimensiones.

Tabla 3: Descripción de categorías, subcategorías, propiedades y dimensiones

Categorías	Subcategorías	Propiedades	Dimensiones
Comportamiento	Patrón de compra	Frecuencia de compras	Regularmente
			Ocasionalmente
		Horario	Mañana
			Tarde
			Noche

Tabla 3 (continuación)

		Tipo de compras	Planteadas
			Por impulso
			No previstas
		Tipo de productos	Tecnológicos
			Entretenimiento
			Vestimenta
		Medio de interacción	Computadora
			Dispositivos móviles
		Gastos	Limitados
			Ilimitados
Ubicación	Casa		
	Oficina		
Preferencias	Tipos de sitios	Nacional	
		Extranjero	
	Medios de envío	Servicio regular o exprés	
		Envío a sucursal o domicilio	
Motivación de compra	Necesidades existentes por satisfacer	Métodos de pago	Transferencia electrónica bancaria (débito/crédito)
			Transferencia electrónica PayPal
			Depósito OXXO
			Depósito bancario
		Promociones	Precio accesible
			Ofertas
			Descuentos
		Servicios	Rápidos
			Eficientes
			Existencia
	Información	Clara	
		Real	
		Detallada	
Aceptación	Personal		
	Social		
Necesidades latentes de los sitios	Recomendaciones	Personalizado	
	Avisos	Personalizado	
	contexto	Personalizado	
Experiencia del consumidor	Problemas	Proceso de compra	Método de pago
			Autorización de compra
			Error de cobro
		Búsqueda de productos	Descripción escasa
			Visualización falsa
	Proceso de búsqueda de sitios	Fuentes	Publicidad
			Búsqueda propia
			Recomendaciones
	Personal	Conocimiento	Experiencia en computación
			Experiencia en los métodos de pago
			Experiencia en la navegación de sitios
			Experiencia en comparación
		Experiencias	Positivas
Negativas			
Confianza	Alta, media, baja		
Influencias	Amigos, conocidos		
	Blogs		

En la Tabla 3 se muestra una de las categorías que se definieron para la interpretación del estudio del comportamiento del consumidor, la cual nos define el comportamiento del consumidor y da a conocer las relaciones causales donde el comportamiento del consumidor genera un patrón de compra y otras subcategorías

como lo son las preferencias del consumidor. Una vez que la categoría se identificó, se describieron sus propiedades y las características de dimensión.

### Codificación selectiva

En esta fase del proceso se selecciona una categoría para ser el núcleo, y se relacionan todas las demás categorías con la central. La categoría central en este estudio es la del *comportamiento* presentada como una categoría en la Tabla 3. Así mismo, se describen otras dos categorías, la motivación y la experiencia del consumidor, las cuales están directamente relacionadas con la categoría central y explican las condiciones y consecuencias de la misma. En conjunto, las tres categorías ayudaron de manera significativa para la creación de teoría que explica el comportamiento del consumidor en compras en línea. La recolección y análisis de datos en el proceso de investigación ocurrió simultáneamente y la teoría que surgió al concluir el proceso de análisis de los datos. En el cual, los códigos obtenidos de la recolección y análisis de los datos fueron comparados e interpretados dentro del contexto de transcripciones generales presentados en la siguiente sección.

## RESULTADOS

Los datos se presentaron utilizando el marco propuesto por Strauss y Corbin (1998) para el desarrollo de categorías y subcategorías. La codificación selectiva se concentró en la formulación de una categoría central para obtener la idea conceptual bajo la cual se agruparán todos los elementos de las categorías alrededor de la central. Las categorías seleccionadas y desarrolladas fueron: comportamiento, motivación de compra y experiencias del consumidor. Finalmente, se generaron ideas a partir de las relaciones de las categorías, las cuales dieron origen a la teoría que describe los factores que influyen en mayor medida el comportamiento del consumidor en compras en línea, presentada en la Figura 3. En la Figura 3 se muestra como la motivación y la experiencia influyen en el comportamiento del consumidor. Asimismo, se identifican las variables que conforman a cada uno de estas categorías. A continuación se describe cada uno de las categorías del modelo teórico – explicativo.

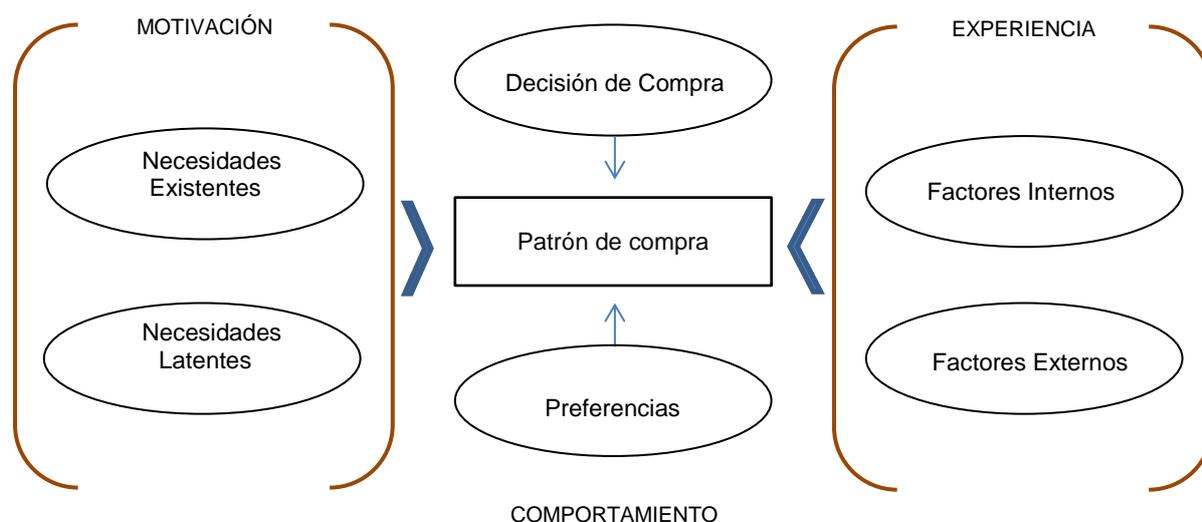


Fig. 3: Modelo teórico-explicativo del comportamiento del consumidor

### Categoría 1: Comportamiento

Esta categoría es la principal del estudio, en ella se define el comportamiento del consumidor. Por lo cual, el comportamiento es definido por las actividades del consumidor orientadas a la compra de productos y servicios, incluyendo los procesos de decisión y sus preferencias, influenciadas por las dos subcategorías (motivación y experiencia) que preceden y determinan el patrón de compras del consumidor.

#### Variable 1: Patrón de compras

Con base en los comentarios de los participantes se identificó que durante los últimos años los consumidores adquieren productos y servicios mediante una mayor búsqueda de información y procesos de selección que tiene mucho que ver con sus preferencias de consumo sobre qué, dónde y cuánto comprar. Por lo tanto, los participantes manifestaron un alto empoderamiento de sus decisiones de compra como se observa en el comentario del participante 3, en donde expone que son capaces de exigir mayores y mejores condiciones en sus compras en línea, realizando búsquedas previas en diferentes sitios de comercio electrónico para seleccionar

el más adecuado a sus preferencias. Comentario del participante 3 (P3): *“Pues depende que es lo que estoy comprando y en donde, por ejemplo, yo siempre por regla comparo los precios, nunca, nunca, nunca jaja, la verdad no compro la primera vez que entro al sitio, por ejemplo, la última compra que quería hacer la última semana... compare ese producto en las tiendas físicas, entonces ya sabía los precios de las tiendas que si tenía en mi ciudad y luego comparé en línea porque a veces ponen ofertas exclusivas en línea”*.

Asimismo, el comentario del P10 nos demuestra que no solo se limitan a obtener información de productos y servicios en general, sino que también analizan los comentarios de revisiones de otros consumidores que han adquirido un producto o servicio, esto indica que es una actividad que influye en gran medida en su decisión de compra. El comentario de este participante es el siguiente: *“Al tener el sitio donde es más barato reviso los comentarios de los clientes que han comprado el mismo producto para ver si me decido comprar”*. Por otra parte, en el comentario del participante 10 se identifica una fase de exploración obligatoria, en donde sienten una responsabilidad de indagar, explorar y encontrar información para sentirse seguros al realizar su compra, tal como sigue: *“Pues, yo creo depende de la personalidad de como sea la persona, por ejemplo yo soy muy paciente al buscar porque puedo durar 2 horas metida buscando artículos y porque quiero encontrar cosas que me gusten y estén baratas pero a mí se me hace muy fácil sobre todo si tiene los filtros para cambio rangos de precios, o por marcas y si esos filtros no los tuviera la verdad sería una flojera y así es más fácil encontrar algo que busco, mientras más detallado un filtro mejor y es lo que me ha gustado en los sitios que uso para comprar”*.

#### *Variable 2: Preferencias*

Con base en las experiencias de los participantes, se encontró que los consumidores tienen una lista priorizada de sus preferencias, como se muestra en los siguientes comentarios de los participantes 2, 3 y 5: *“Las compras por las páginas chinas usan correos de México, si he comprado en páginas así y he pagado extra el envío y ha sido DHL para que me llegue rápido, pero sólo en páginas en donde ofrecen eso porque no todos los vendedores te lo ofrecen.”* (P3) *“Uso más mercado libre por la rapidez de envío... lo que más valoró es la rapidez de entrega...”* (P2) *“También se me hace más seguro comprar en Amazon que en otro sitio, además que este sitio tiene el servicio de envío gratuito.”*(P5)

Por ejemplo, en cuanto a preferencias de los tipos de sitios, los medios de envío y los métodos de pago, están influenciadas directamente con las necesidades existentes por satisfacer y por la motivación de compra. Sin embargo, a la hora de la toma de decisión de compra se ven más inclinados a los medios de envío por encima de los precios y sitios de compra, como lo demuestra el comentario del participante 10: *“Por ejemplo, un día compré en una página donde si tu pedido era mayor a los 70 dólares tu envío era gratis pero como yo quería que me llegara pronto me valió pagar extra por el pedido.”* (P10)

#### *Categoría 2: Motivación de compra*

Esta categoría influye directamente sobre el comportamiento de compra, siendo identificada como el estímulo de comprar en línea.

#### *Variable 3: Necesidades existentes por satisfacer*

En este estudio, se describió esos estímulos como necesidades existentes por satisfacer incluyendo las promociones, servicios, información y aceptación que proporcionan estos sitios en línea.

Refiriéndose a esto, los participantes declararon: *“La variedad de los productos, que actualicen su catálogo de productos.”* (P5); *“Quiero encontrar cosas que me gusten y estén baratas.”* (P10); y *“Por los costos muy bajos y porque encuentras cosas que en otros sitios no encuentra.”*(P2)

#### *Variable 4: Necesidades latentes de los sitios*

También se detectaron necesidades relacionadas al medir, sentir o palpar el producto donde los participantes mostraron interés que deben ser atendidas de alguna manera para tener mejores experiencias de compra. A continuación comentarios del participante 10: *“...la experiencia de medirse, sentirlo y ver cómo te queda no lo tienes y poner fotos reales puede acercarnos un poco a esta experiencia de las tiendas físicas. También estaría padre tener a alguien ahí por la plataforma para poder preguntarle cosas sobre los productos.”* (P10)

#### *Categoría 3: Experiencias del consumidor*

En cuanto a la categoría de experiencias del consumidor, los participantes comentaron que es un factor importante para la definición del comportamiento de compra, debido a que los consumidores poseen conocimientos previos a la hora de navegar por sitios de compras en línea, esto indica que al momento de sentir el estímulo que anteriormente se define como motivación, la experiencia proporciona la información suficiente

que influirá en su comportamiento de compra. Esta experiencia se dividió en problemas (variable 5), proceso de búsqueda (variable 6) y personal (variable 7) donde cada una de estas variables se explican a continuación.

#### Variable 5. Problemas

De acuerdo a los comentarios de los participantes, las compras en línea implican un acto de fe en el que el consumidor pone su confianza en los sitios de compras en línea, para que sus pedidos sean enviados en tiempo y forma. Sin embargo, todos han pasado por desafortunadas situaciones que hicieron no obtener su producto o servicio y perder su dinero. Por lo cual, proporcionan sus experiencias y lecciones aprendidas de los eventos problemáticos a la hora de comprar, como lo expresa el siguiente comentario: “Lo debes hacer con moderación, checar muy bien el sitio que vas a utilizar, que sea un sitio web conocido, que tenga muchos usuarios y leer de perdida unos puntos de contrato de privacidad que te proporcionan algunos sitios.” (P2)

#### Variable 6. Proceso de búsqueda

Con base a las experiencias de los participantes, se encontró que la distribución y existencia de los elementos básicos de navegación en un sitio de compras en línea es de vital importancia para la decisión de compra, como se muestra en el comentario del participante 3: “Que este CLARO todo, que este todo completo, lo que te presentan y ¡claro!. Que sea fácil y que el buscador ¡híjole!, el buscador es algo que se busca, bueno por lo menos yo es lo primero que uso cuando quiero algo en específico, pero a veces entro a páginas nomas para ver qué hay de nuevo...”(P3)

#### Variable 7. Personal

Del mismo modo, las experiencias buenas y malas son un factor clave para el comportamiento de consumidor, porque son las responsables de influir en mayor medida en las preferencias y decisión de compra de los consumidores. El comentario del participante 3 expresa tener una vasta experiencia en compras en línea y que la mayoría han sido buenas: “Mis experiencias son bastante buenas, son que encuentras siempre lo que buscas, es económico, no batallé en el proceso de compra, al momento de buscar y hacer el pedido.” (P3). Por lo cual, a partir del establecimiento de las conexiones de las categorías relacionadas (motivación y experiencia) con la categoría central (comportamiento) se pueden enriquecer el conocimiento actual sobre el comportamiento del consumidor ante plataformas de compra en línea, mediante la implementación de estrategias de mejora con respecto a los resultados obtenidos, como se muestra en la Tabla 4.

Tabla 4: Estrategia de mejora en plataformas de compra en línea de acuerdo a las categorías definidas en el estudio

Categorías	Estrategia de mejora en plataformas de compra en línea
Motivación	<ol style="list-style-type: none"> <li>1) Actualización constante del catálogo de productos y servicios.</li> <li>2) Mantener siempre descuentos y promociones.</li> <li>3) Contar una descripción de productos detallada, al igual que fotos de productos reales que se pueden observar todas las características del producto.</li> <li>4) Proporcionar un chat online para resolver las inquietudes o miedos que le vayan surgiendo a los consumidores de la plataforma de compra en línea en tiempo real.</li> </ol>
Comportamiento	<ol style="list-style-type: none"> <li>5) Contar con una sección de comentarios para que los consumidores expresen sus opiniones acerca de los productos y servicios consumidos y ayuden en la decisión de compra de futuros consumidores.</li> <li>6) Proporcionar filtros de búsqueda de productos y servicios de diferentes tipos: marca, precio, departamento, entre otros.</li> <li>7) Ofrecer diferentes medios de pago y paquetería.</li> </ol>
Experiencia	<ol style="list-style-type: none"> <li>8) Proporcionar información acerca de garantías, políticas y contrato de privacidad.</li> <li>9) Contar con un buscador interno de productos.</li> <li>10) Ofrecer información acerca de devoluciones y reembolso.</li> <li>11) Contar con una sección de contacto para solución de problemas.</li> </ol>

La implementación de estas estrategias de mejora en las plataformas de compra en línea pueden garantizar un mayor conocimiento del comportamiento y patrón de compra del consumidor. Puesto que las plataformas que conformen todos los elementos que se proporcionan en la Tabla 4, podrán ir adaptándose de acuerdo al análisis del consumidor periódicamente consiguiendo una toma de decisiones efectiva asegurando una mejor aceptación y satisfacción del consumidor.

## DISCUSIÓN

En el comportamiento del consumidor existe una serie de influencias en el proceso de compra como se describe anteriormente: la motivación y las experiencias. Para esto, una de las influencias más importante está relacionada con la necesidad de existencia de productos, ya que, por ejemplo, cuando no existen los productos en tiendas locales, aumenta la tendencia de consumo en línea. Como ejemplo de ello, el participante 2 opinó *“Porque no hay localmente.”* (P2), el participante 10 comentó *“No hay existencia en las tiendas física.”* (P10) y el participante 5 *“a veces en las tiendas no hay lo que buscas y aparte que sale mucho más caro...”* (P5); Esta información demuestra que la tendencia de compra en línea podría aumentar al existir escasez de un producto y el alto costo en las tiendas locales. Puesto que en un estudio de la Asociación de Internet MX (2018), demuestra que las compras en línea han aumentado un 17% más que en el año 2017, manteniendo un 54% de usuarios que compran en línea. Sin embargo, este estudio afirma que hay una presencia de las páginas de comercio electrónico en la vida del usuario con un 39%, al contrario, nuestro estudio demuestra que el 75% de participantes están de acuerdo que el uso de plataformas de compra en línea cambiaron sus hábitos de compra ya que estas exhiben un mayor catálogo de productos y mejores ofertas.

En definitiva, los consumidores toman en primer lugar esta necesidad como su motivación principal para comprar en sitios de compra en línea, y esto es debido a la influencia de blogs y redes sociales actuales, los cuales han influido de manera notable en el proceso de decisión de compra. En ese sentido, algunos estudios se demuestra la importancia que han tomado las redes sociales por medio de la comunicación y la interacción de información de productos, lo cual afecta el comportamiento de compra favoreciendo la compra en línea (Guardia y Gómez, 2009; Vargas, 2015). Esto es debido a que la publicidad proporcionada en las redes sociales tiene una influencia directa en los aspectos intelectuales, sensitivos y emotivos del consumidor como se ha demostrado en estudios previos (Hsu, Chuan-Chuan y Chiang, 2013). Como ejemplo de ello, el participante 10 opinó: *“Las de ropa y todo eso, si fue por publicidad en Facebook, YouTube e Instagram, como yo me la llevo buscando y viendo en esas redes sociales noto esa publicidad...”* (P10). Por lo tanto, la influencia de las redes sociales y los blogs pueden influir en las distintas fases del proceso de decisión de compra. Más aún, en este estudio se encontró que los comentarios de los usuarios de los medios sociales tienen un mayor poder de influencia en la decisión de compra que los mensajes de comunicación comercial de cualquier otro medio, en el segmento de población estudiado nacidos entre 1975 y 1995, tomando en cuenta la generación X a la generación Z.

Agregando a lo anterior, los productos y servicios de entretenimiento y de moda son los más consumidos en sitios de compra en línea, puesto que el tiempo entre estímulo de compra (motivación) y la compra (decisión de compra) se ha disminuido en gran medida por causa de la influencia de las redes sociales mencionadas anteriormente, por lo cual los consumidores adquieren más productos en menos tiempo. Sin embargo, un estudio de la Asociación Mexicana de Venta Online (AMVO) demuestra las ventas generadas por el comercio electrónico, donde el sólo el 2% se genera en México. A pesar de ello, nuestro estudio señala que la población del estado de Sonora, cada vez adquieren mayor confianza para realizar compras en línea de acuerdo a su experiencia en compras previas, las cuales indicaron la existencia de mejores ofertas que en tiendas físicas, variedad de productos y marcas exclusivamente en línea; y de motivación, puesto a las influencias de redes sociales y blogs que promueven una reputación positiva a productos ofrecidos en línea y que las tiendas físicas empiezan a contar con su propio portal de compra en línea. Por otra parte, tanto los sitios web y aplicaciones de compras en línea, han mejorado su proceso de compra, para ofrecer una mayor facilidad y rapidez en la compra de los productos y servicios, lo cual estimulan con mayor intensidad la decisión de compra.

El tipo de perfiles de los participantes seleccionados para el análisis demostraron que las preferencias y patrón de compras son distintos, porque se manifiesta otro tipo de motivaciones y experiencias. El perfil del consumidor en línea generalmente conoce sus posibilidades en el nuevo medio de realizar compras, estas posibilidades de compra se expanden, desde el lugar de conexión, las valoraciones de productos y las recomendaciones personalizadas para la toma de decisiones de compra. Asimismo, las dos clasificaciones de los perfiles de los consumidores son 1) los que desean un ahorro de tiempo y 2) los que desean un ahorro de dinero. Sin embargo, nuestros resultados demuestran el origen del por qué un consumidor se clasifica en uno de los dos grupos y también, el cómo poder anticipar el perfil del consumidor para poder ofrecer un servicio personalizado de acuerdo a las características del usuario en la plataforma de compra en línea. Por otro lado, además de las características abordadas en este trabajo, otros estudios mencionan que una de las características que consideran más importantes para la compra en línea son la confianza y la fiabilidad que percibe el consumidor al momento de realizar la compra en el sitio en línea (Al-Debei, Akroush y Ashouri, 2015; Sánchez-Alzate y Montoya, 2017). Si bien, estas características no se abordan explícitamente como una categoría en este estudio, se consideraron implícitamente dentro del grupo de elementos fundamentales para la generación de estrategias que conlleve a la satisfacción del consumidor en las compras en línea. Por ejemplo, para los participantes que trabajan y adquieren productos y servicios para sus actividades laborales, el factor de mayor importancia es el precio del producto y servicio, por el contrario, si es de tipo personal el factor más importante es el medio de envío y en segundo lugar el precio. De esta manera, los datos obtenidos sobre el comportamiento del consumidor en compras en línea,

demuestran que los consumidores poseen cada vez más control y es necesario dar a conocer los factores encontrados y sus relaciones que influyen en el patrón y decisión de compra en línea. Asimismo, fue posible generar una teoría con información respecto a aquellos factores de mayor importancia en la decisión y patrón de compra por Internet en los consumidores del estado de Sonora en la compra de productos y servicios en los sectores de tecnología, vestimenta y entretenimiento.

## CONCLUSIONES

Los hallazgos de este estudio demuestran que el enfoque metodológico de la Teoría Fundamentada aplicado en el ámbito de compras por Internet garantizó la identificación de factores clave como la motivación, la experiencia y el comportamiento, los cuales se obtuvieron a partir de una reflexión crítica que permitió relacionar temas de interés para el estudio de comportamiento del consumidor a partir de la realidad observada. Con base en los resultados del estudio, se concluye que el análisis del comportamiento del consumidor puede ayudar a la creación de estrategias para la toma de decisiones efectivas mediante su aplicación en el diseño de plataformas de compra en línea. Estas plataformas podrían proveer al empresario una herramienta tecnológica donde podrá obtener un análisis de sobre el comportamiento de los consumidores para mejorar estrategias de mercadeo y venta.

## REFERENCIAS

- Al-Debei, M. M., M. N. Akroush y M. I. Ashouri, Consumer Attitudes towards Online Shopping: The Effects of Trust, Perceived Benefits, and Perceived Web Quality, doi: 10.1108/IntR-05-2014-0146, J. of Internet Research, 25(5), 707-733 (2015)
- Asociación de Internet.mx, 14 Estudio sobre los Hábitos de los Usuarios de Internet en México 2018, México (2018)
- Asociación Mexicana de Venta Online, Retos y Tendencias del Sector Logístico: "E Commerce y la Última Milla" México 2018, Daqua Strategic Intelligence, México (2018)
- Barrera, G. A., Relación de Capacitación con Adopción de Internet y E-Commerce: Diferencias entre Microemprendedores de Chile, doi: <http://dx.doi.org/10.4067/S0718-07642017000600008>, Revista de Información Tecnológica, 28(6), 61-70 (2017)
- Cheung, C.M., I. L. Liu y M. K. Lee, How Online Social Interactions Influence Customer Information Contribution Behavior in Online Social Shopping Communities: a Social Learning Theory Perspective, doi: 10.1002/asi.23340, J. of the Association for Information Science and Technology, 66(12), 2511-2521 (2015)
- Flick U., Introducción a la Investigación Cualitativa, 3ª Ed., 78-79, Ediciones Morata, Madrid, España (2012)
- Gallaughier, P.A. J.M., Factors Affecting the Adoption of an Internet-Based Sales Presence for Small Businesses, doi: 10.1080/019722497129287, J. of The Information Society, 13(1), 55-74 (1997)
- Gefen, D. y D. W. Straub, The Relative Importance of Perceived Ease of Use in IS Adoption: A Study of E-Commerce Adoption, ISSN: 1536-9323, J. of the Association for Information Systems, 1(8), 1-30 (2000)
- Glaser, B. y A. Strauss, The Discovery of Grounded Theory: Strategies for Qualitative Research, Aldine Publishing Company, Chicago, Estados Unidos (1967)
- Goldsmith, R.E. y D. Horowitz, Measuring Motivations for Online Opinion Seeking, doi: 10.1080/15252019.2006.10722114, J. of Interactive Advertising, 6(2), 2-14 (2006)
- Grandon, E. E. y J. M. Pearson, Electronic Commerce Adoption: an Empirical Study of Small and Medium US Businesses, doi: 10.1016/j.im.2003.12.010, J. of Information & Management, 42(1), 197-216 (2004)
- Guardia, M.L.G. y P.N. Gómez, Bloggers y su Influencia en la Imagen de una Marca, doi: <https://doi.org/10.7195/ri14.v7i1.343>, Revista ICONO14, Revista Científica de Comunicación y Tecnologías Emergentes, 7(1), 242-252 (2009)
- Hausman, A., A Multi-method Investigation of Consumer Motivations in Impulse Buying Behavior, doi: 10.1108/07363760010341045, J. of Consumer Marketing, 17(5), 403-426 (2000)
- Homburg, C. y A. Giering, Personal Characteristics as Moderators of the Relationship between Customer Satisfaction and Loyalty—an Empirical Analysis, doi: 10.1002/1520-6793(200101)18:1<43::AID-MAR3>3.0.CO;2-I, J. of Psychology & Marketing, 18(1), 43-66 (2001)
- Hsu, C. L., L. J. Chuan-Chuan y H. S. Chiang, The Effects of Blogger Recommendations on Customers' Online Shopping Intentions, doi: 10.1108/10662241311295782, J. of Internet Research, 23(1), 69-88 (2013)
- Joines, J. L., C. W. Scherer y D. A. Scheufele, Exploring Motivations for Consumer Web use and their Implications for e-commerce, ISSN: 0736-3761, J. of Consumer Marketing, 20(2), 90-108 (2003)
- Karbasivar, A. y H. Yarahmadi, Evaluating Effective factors on Consumer Impulse Buying Behavior, ISSN: 2222-1387, Asian Journal of Business Management Studies, 2(4), 174-181 (2011)
- Locke, K., Grounded theory in Management Research, SAGE Publications Inc., California, Estados Unidos (2001)
- Olson, J. S. y G. M. Olson, i2i trust in e-commerce, doi: 10.1145/355112.355121, Communications of the ACM, 43(12), 41-44 (2000)

- Pavlou, P. A. y M. Fygenson, Understanding and Predicting Electronic Commerce Adoption: An Extension of the Theory of Planned Behavior, doi: 10.2307/25148720, J. of Management Information Systems Quarterly, 30(1), 115-143 (2006)
- Sánchez-Alzate, J. A. y R. L. A. Montoya, La Confianza como Elemento Fundamental en las Compras a través de Canales de Comercio Electrónico: Caso de los Consumidores en Antioquia, doi: 10.15446/innovar.v27n64.62365, Revista Innovar, 27(64), 11-22 (2017)
- Solomon, M. R., D. W. Dahl y otros tres autores, Consumer behavior: Buying, Having, and Being, 6ª Ed., Pearson, Toronto, Canada (2014)
- Spiggle, S., Analysis and Interpretation of Qualitative Data in Consumer Research, doi: 10.1086/209413, J. of Consumer Research, 21(3), 491-503 (1994)
- Strauss, A. y J. Corbin, Basics of Qualitative research: Techniques and Procedures for Developing Grounded Theory, Thousand Oaks, California, Sage Publication (1998)
- Strauss, A.L., J. Corbin y E. Zimmerman, Bases de la Investigación Cualitativa: Técnicas y Procedimientos para Desarrollar la Teoría Fundamentada, 1ª Ed., Editorial Universidad de Antioquia, Medellín, Colombia (2002)
- Valles, M.S., Entrevistas Cualitativas, 2ª Ed., 68, Centro de Investigaciones Sociológicas, Madrid, España (2009)
- Vargas, S.A.M., Analizar la Influencia de los Mensajes en las Redes Sociales de Marcas de Prendas de Vestir, sobre la Decisión de Compra en Mujeres de 18 a 30 años de la Ciudad de Cali, Tesis de Licenciatura, Universidad Autónoma de Occidente, Facultad de Comunicación Social y Periodismo (2015)
- Westbrook, R. A. y R. L. Oliver, The Dimensionality of Consumption Emotion Patterns and Consumer Satisfaction, doi: 10.1086/209243, J. of Consumer Research, 18(1), 84-91 (1991)
- Zarco, A.I.J. y M.P.M. Ruiz, La Influencia de las TIC en la Distribución Comercial: Implicaciones Estratégicas para la Gestión Promocional Minorista, ISSN-e: 1681-5645, CTS+ I: Rev. Iberoamericana de Ciencia, Tecnología, Sociedad e Innovación, (7), 9 (2006)
- Zhou, L., L. Dai y D. Zhang, Online Shopping Acceptance Model-A Critical Survey of Consumer Factors in Online Shopping, ISSN: 1526-6133, J. of Electronic Commerce Research, 8(1), 41-62 (2007)

# Agile Dimensional Model for a Data Warehouse Implementation in a Software Developer Company

Kathya E. Mercado<sup>1</sup>, Cynthia B. Perez<sup>1</sup>, Laura P. Lopez-Arredondo<sup>1</sup>, Karina Caro<sup>2</sup>,  
Luis A. Castro<sup>1</sup> and Luis-Felipe Rodriguez<sup>1</sup>

<sup>1</sup> Sonora Institute of Technology (ITSON), Ciudad Obregon, Sonora,  
Mexico

<sup>2</sup> Drexel University,  
USA

{kathya.mercado, laurapaolalpz}@hotmail.com,  
{cynthia.perez, luis.rodriguez}@itson.edu.mx, luis.castro@acm.org, karinacaro@drexel.edu

**Abstract.** Nowadays, the increasing development of Business Intelligence (BI) solutions in organizations, has enabled executives achieve a better understanding of business information for timely and rapid decision-making in a tremendously dynamic market. Although there is an increasing interest in adopting an agile approach to the software development, the emergent need of using agile methodologies in BI solutions is undeniable. This paper discusses the importance of using agile methods in the design and development of data warehouses taking into account the business processes, requirements analysis, and organizational objectives. Thus, we present a case study derived from a real-world business project where the agile methodology Business Event Analysis and Modeling (BEAM) is used to design the data warehouse. The project is based on a billing system with about one million operations on a daily basis with more than 15,000 clients. Finally, the results of this paper include the design of the dimensional model using an agile approach, the construction of the data warehouse through the ETL processes and an interactive dashboard according to the key performance indicators defined by the business decision-makers.

**Keywords:** business intelligence, agile dimensional model, data warehouse.

## 1 Introduction

In the business world we aim to obtain greater profits and a greater competitive advantage, *hence*, appropriate timely decision-making plays a very important role in the fulfillment of the organizational objectives. Thus, the use of Business Intelligence (BI) systems can help meet these needs as it provides a set of methodologies, applications, and practices focused on the information management for accurate, timely decisions making in an organization.

A major component of any BI model is the design and implementation of a data warehouse in organizations, since it provides valuable and strategic information to support decision-making through real-time access to business transactions and advanced business analytics [1]. According to Inmon [2] and Imhoff [3], a data warehouse in organizations is an integrated data collection, non-volatile and variable over time. Hence, it has a complete history of the organization, beyond the transactional and operational information favoring the data analysis for decision-making.

Today, the two most widely used methodologies for the design and implementation of data warehouses are the model of Inmon [2] and Kimball [4]. They consider the data warehouse as the central repository of data for organizations that is used to present business reports. The difference between these two methodologies lies in how to make deliveries of progress (time) and how to manage changes during the process (see Table 1).

Recent studies tend to show that use of agile methodologies for the design and implementation of data warehouse in organizations is playing an important role to obtain value information to help decision-makers and to generate competitive advantage by improving the extraction and processing knowledge. These studies present a structured methodology, inspired from the agile development models as Scrum, XP and AP [5]. Thus, an increasing number of companies are choosing for an agile philosophy in software development due to the constant need to be flexible and adaptable to the technological changes and the new user demands [6-10].

For this reason, it is important to consider agile methods for the design and implementation of a data warehouse in a BI system (see Table 1). In this regard, it has been shown that agile development processes increase the potential for developing the success of a data warehouse by solving many of the typical problems presented in traditional methodologies [11, 8, 12].

In this paper, we use the BEAM methodology in the design, modeling, and implementation of a data warehouse for a management system. This system processes one million operations on a daily basis from more than 15,000 clients. The agile design and modeling of the data warehouse are presented through a case study of a Software Development Company located in Mexico, which offers software solutions focused on meeting the needs of development, implementation and support to any business sector.

## **2 Agile Dimensional Modeling**

Traditional data warehousing projects follow the waterfall structure to perform dimensional modeling [13]. However, its use is increasingly unlikely and alternatives of analyzing and designing similar to those used in software development projects such as agile methods are looked for [7, 12, 14]. In this respect, the agile dimensional modeling is being considered as a solution for BI systems since it allows developers to reduce the risks that the waterfall structure could produce [11, 14]. All this is possible by adopting a highly interactive, incremental and collaborative approach to the whole analysis, design and development activities of a data warehouse, such as the agile BEAM methodology [14] ( see Figure 1).

**Table 1.** Comparison of the Inmon, Kimball, and Corr methodologies.

	<b>Inmon [2] (Traditional)</b>	<b>Kimball [4] (Traditional)</b>	<b>Corr [14] (Agile)</b>
<b>Business Decisions</b>	Strategic	Tactical	Collaborative
<b>Scope</b>	Product owner	Project manager	Team
<b>Approach</b>	Top – down	Bottom – up	Bottom – up
<b>Objective</b>	Deliver a robust technical solution based on proven methods	Provide a solution that facilitates the end users to consult the data	Responding to change and user needs
<b>Data requirements</b>	Enterprise-wide	Business process	Individual business requirement (KPIs)
<b>Data modeling</b>	Normalized form (3NF)	Dimension model (Star or Snowflake)	Dimension model (Star or Snowflake)
<b>Orientation</b>	Enterprise-wide	Business process	People
<b>Communication</b>	Formal	Formal/Informal	Informal
<b>Time</b>	Longer start-up time	Shorter start-up time	Minimal start-up time
<b>Project schedule risk</b>	High	High	Low
<b>Ability to respond to change</b>	Low	Medium	High
<b>End users involvement</b>	Minimal	Oscillate depending on the project	High
<b>Cost to build</b>	High initial cost	Low initial cost	Minimal initial cost

## 2.1 BEAM Methodology

Corr [14] proposes the BEAM (Business Event Analysis and Modeling) methodology, an agile data modeling method for the design and development of data warehouses and data marts. This method combines analysis and modeling techniques to meet data requirements related to business events and data modeling for database design that is easy to understand by stakeholders and also, easy to translate into logical/physical models for IT developers. The BEAM methodology involves stakeholders who think beyond their current reporting requirements by describing data stories, that is, narratives that define the dimensional details of business activities necessary to be measured. In order to obtain these data stories, data modelers ask questions to stakeholders using a framework based on the 7Ws (who, what, where, when, how many, why and how) [14]. The way to find these answers of the 7Ws and make sure they inform data warehouse design is to ask end-users about the events that are happening in their business. Therefore, the enhanced Start Schema is used to generate and show schema of physical data bases, where are they involved Data Modelers, DBAs, DBMS, ETL Developers, BI Developers and Testers. This framework is one of the main activities of the BEAM methodology because it allows discovering and modeling data requirements and thus, to construct the table of dimensions and facts of the data warehouse depicted through the star model.

According to Corr [14], the BEAM methodology has several diagrams for the analysis and design of the data warehouse model, such as: BEAM Table, Hierarchy Chart, Timeline, Event Matrix and Enhanced Star Schema.

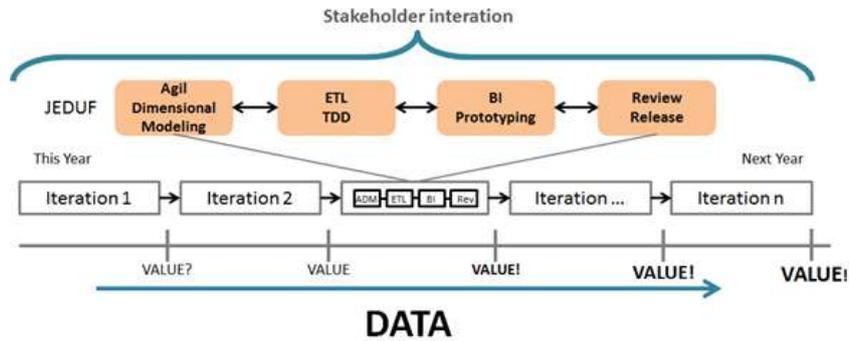


Fig. 1. Agile data warehouse development timeline. Source [14].

For instance, the BEAM Table is used to model business events and dimensions at the same time; people involved in this diagram are data modelers, business analysts, stakeholders and BI users. The Enhanced Star Schema is used to visualize the dimensional model for the implementation of the physical database schemas; the people involved in this diagram are Data Modelers, DBAs, DBMS, ETL Developers, BI Developers and Testers.

### 3 Agile Data Warehouse: A Case Study of a Billing System

This case study focuses on the design and implementation of a data warehouse using the BEAM methodology for a billing system of a company based in Mexico with operations in software development with around 15,000 active clients in 19 countries and more than a million operations on a daily basis. Despite all the information daily stored on the company's servers, this information is not used or analyzed so far by the working team, identifying an area of opportunity for the design and implementation of BI systems. Therefore, decision-makers could offer their customers significant knowledge through scorecards and thereby provide a competitive advantage. The management system access to a database based on the Entity-Relationship Model (ER) allowing to record, update, delete, and query information from the main business processes. This system has the following modules such as: billing, inventories, clients, payroll, branches, among others. Hence, in this case study, we focus on the billing module. The database used by the system contains around 70 tables using only the most relevant according to the key performance indicators (KPI).

#### 3.1 Analysis, Design and Implementation of the Data Warehouse

Unlike the development of software applications, where the requirements of the organizations are often relatively well defined by the result of the stability of business rules over time; create a data warehouse depends on the reality of the company and its current conditions.

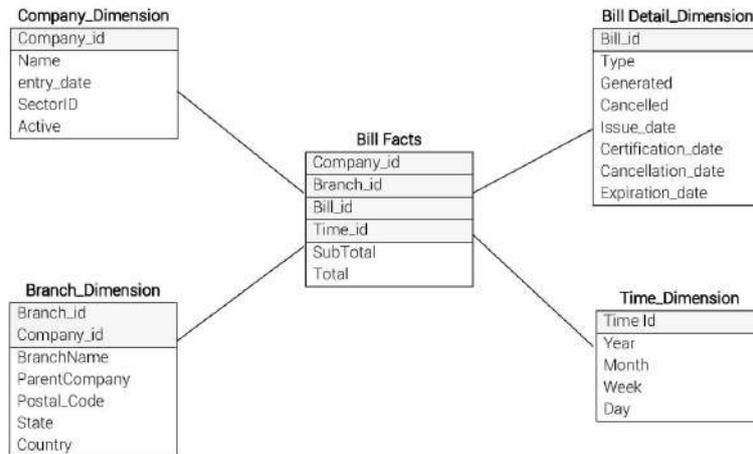


Fig. 2. Star schema for the proposed data warehouse.

Thus, the company requires identifying useful information in order to obtain significant information about its clients. In this way, the following key performance indicators (KPIs) were identified along with the business owners for the design and implementation of the data warehouse: a) Compare monthly growth of registered companies; b) Identify customer loyalty, through the use of the system more than 3 years; c) Identify quantity and list of branches by parent enterprise; d) Visualize the States where there are more than four companies using the management system; e) Measure different types of invoices volumes identifying the invoices variations not only by a time period but also, for company; f) Geographically view the top 10 companies that make the most invoices for a period of time; g) Compare the invoices by branch, time and state.

In order to model the data requirements, the 7Ws framework was used resulting in the identification of facts and dimensions in the star model, presented in Figure 2.

In this regard, the proposed dimensional model contains a fact table where the measurements or metrics of a specific event are recorded; for example, the invoice for a purchase and foreign keys referencing dimensional data tables (Company, Branch, Bill Detail and Time) which contain descriptive information. In order to implement the proposed model (see Figure 2), it is necessary to perform the ETL (Extraction, Transformation, Load) process which enables moving data from multiple sources, transform and load them into the data warehouse to analyze and thereby giving valuable information to organizations. In this way, Microsoft SQL Server Integration Services (SSIS) of Visual Studio 2015 was used to perform the ETL process. Finally, the information contained in the data warehouse was visualized through a scorecard developed using the Microsoft PowerBI tool. Thus, the KPIs were analyzed in order to identify the visual elements corresponding to each key indicator.

Once the visual elements are selected, the information about the clients is then shown in the scorecard by using queries in which decision-makers could use it interactively.

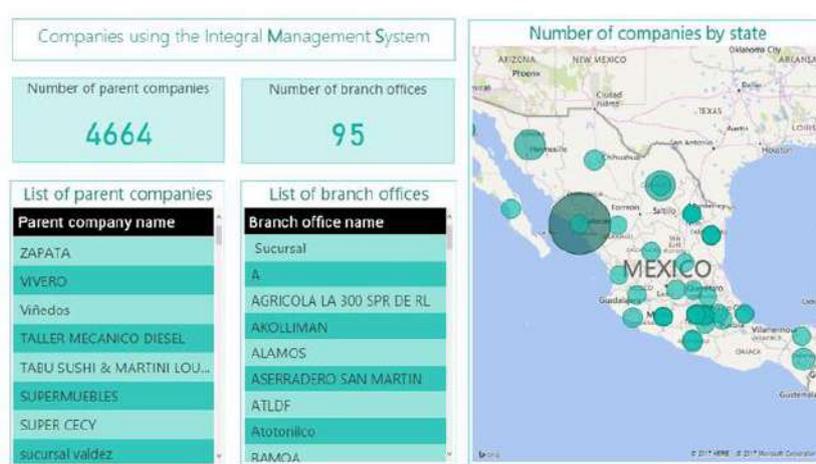


Fig. 3. Scorecard interface: Geographic coverage view of the system.

On the other hand, the software company's CEO was interested in seeing through graphs and trend lines, the situation and invoice generation behavior of those companies who use the management system.

In this way, it could be easily visualized when billing peaks are produced, i.e. sales generated by companies using the system. Figure 3 presents the view of the BI system that shows by state the number of companies who use the system, the bigger the circle the more companies that use the system are in the corresponding state. In this way, it is easy to identify the states where there is little or no presence of the system helping the decision-maker to pay attention in sale strategies. Likewise, Figure 4 presents the view related to invoice analysis where the number of invoices per type is visualized through an interactive list, a trend line to observe the behavior of the number of invoices is generated by each company who use the system and a pie chart showing the top 10 companies with the highest number of invoices issued by the system is shown.

In this way, it is possible to graphically observe the behavior of the number of invoices in a period of time by interacting with the system, allowing the decision-makers not only to know in what years more invoices per company are registered but also, detect those decreases that may indicate a risk in the strategic plan in order to develop an action plan.

### 3.2 Discussion of Results

The design and implementation of the proposed BI system using BEAM methodology allowed the analysis and design of the data warehouse through an agile method that focused on the users' needs and that easily respond to changes. Therefore, meetings with the working team, i.e., stakeholders, BI users, ETL developers, business analysts, among others, were held in order to compile information requirements during all stages of the project emphasizing uninterrupted communication and collaborative work.



Fig. 4. Number of invoices by time period.

On this basis, it ensures a greater understanding of the data warehouse information and the functionalities of the BI system. Moreover, the agile method maintains a logical data structure, scalable and adaptable to future functionalities such as the integration of other system's modules, predictive analysis, among others.

In this way, a robust and scalable BI system was designed and implemented where decision-makers can count on reliable, fast, flexible and easy-to-understand analyses through the scorecard, thereby facilitating the diagnosis of indicators and decision making. Accordingly, Figure 4 shows a scorecard of the invoices section, thus, the use of the scorecard provides reports of different participants in the decision-making process, representing an opportunity for homogenize and refine business processes. For this reason, it is expected to improve business opportunities through the use of key performance indicators by the extraction, processing and presentation of significant information according to the business strategic objectives. Eventually, the use of the BI system will positively impact the improvement of the company's value chain processes, its competitiveness and thus, the profitability of the business.

#### 4 Conclusions and Future Work

Today, entrepreneurs need to analyze and interact with real-time visual information in order to support decision-making. In this regard, the methodologies used in the design of BI systems should consider the current needs and challenges where business requirements are not static and change constantly. Hence, this paper proposed the use of an agile dimensional model for the design and implementation of a data warehouse based on the BEAM methodology applied to a case study for a Software Development Company. In order to complete the project successfully, the organizational requirements were defined, the star schema was modeled, the ETL process was

executed, the data warehouse was implemented and finally, the KPIs were graphically displayed into the scorecard for decision-making.

The results obtained from the use of an agile methodology are found as a model easy- to-understand for the stakeholders; for this reason, it is mandatory to involve them in the whole process.

By adopting this agile approach, flexibility is ensure, as well as, personal coordination with the stakeholders, consistency and simplicity in the whole process.

As a future work, we plan to incorporate a predictive analysis section into the BI system allowing decision-makers to discover patterns, opportunities and prevent risks by increasing the profitability of the business.

## References

1. Chaudhuri, S., Dayal, U.: An overview of data warehousing and OLAP technology. *ACM Sigmod record*, 26(1), pp. 65–74 (1997)
2. Inmon, W. H.: *Building the Data Warehousing*. John Wiley & Sons (2002)
3. Imhoff, C., Galemno, N., Geiger, J. G.: *Mastering data warehouse design: relational and dimensional techniques*. John Wiley & Sons (2003)
4. Kimball, R., Ross, M.: *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons (2011)
5. Wilson, N. B., Edgar, C. S.: Agile Methodology for Modeling and Design of Data Warehouses-AM4DW. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 9(9), pp. 2132–2137 (2016)
6. Golfarelli, M.: *From User Requirements to Conceptual Design in Data Warehouse Design*. IGI Global (2010)
7. Golfarelli, M., Rizzi, S., Turrinchia, E.: Modern software engineering methodologies meet data warehouse design: 4WD. In: *Data Warehousing and Knowledge Discovery*, pp. 66–79 (2011)
8. Collier, K.: *Agile analytics: A value-driven approach to business intelligence and data warehousing*. Addison-Wesley (2012)
9. Chow, T., Cao, D.: A survey study of critical success factors in agile software projects. *Journal of systems and software*, 81(6), pp. 961–971 (2008)
10. Dyba, T., Dingsoyr, T.: Empirical studies of agile software development: A systematic review. *Information and software technology*, 50(9), pp. 833–859 (2008)
11. Boutkhom, O., Hanine, M., Tikniouine, A., Agouti, T.: Integration approach of multicriteria analysis to OLAP systems: Multidimensional model. In: *Computer Systems and Applications (AICCSA), ACS International Conference on IEEE*, pp. 1–4 (2013)
12. Deshpande, K., Desai, B.: Model for Assessment of Agile Methodology for Implementing Data Warehouse Projects. *International Journal of Applied Information Systems*, 9(5), pp. 42–49 (2015)
13. Moody, D., Kortink, M.: From enterprise models to dimensional models: a methodology for data warehouse and data mart design. *DMDW*, pp. 5 (2000)
14. Corr, L., Stagnitto, J.: *Agile Data Warehouse Design: Colaborative Dimensional Modeling, from Whiteboard to Star Schema*. UK: Decision One Press (2011)

# Reengineering of the software development process in a technology services company

Software  
development  
process

Laura Lopez-Arredondo

*Instituto Tecnológico de Sonora, Ciudad Obregon, Mexico*

Cynthia Perez and Jesus Villavicencio-Navarro

*Instituto Tecnológico de Sonora, Guaymas, Mexico*

Kathya E. Mercado and Martin Encinas

*Instituto Tecnológico de Sonora, Ciudad Obregon, Mexico, and*

Patricia Inzunza-Mejia

*Universidad Autonoma de Sinaloa, Guamuchil Salvador Alvarado, Mexico*

Received 8 June 2018

Revised 14 June 2019

27 July 2019

Accepted 13 August 2019

## Abstract

**Purpose** – The purpose of this paper is to address the need of a reengineering of the software development process in a Mexican technology services company. In general, the main risk faced by small- and medium-sized software developers (SMEs) is the inability to meet delivery times or to adjust to project requirements, whether through lack of follow-up on the critical activities of the development process or through an inadequate distribution of workloads among members of the project team.

**Design/methodology/approach** – The methodology used to carry out the reengineering of the software development process is Hammer and Champy, which consists of six stages: introduction to the current situation of the company and the need for change; identification of the business processes; selection of the business process to be redesigned; understanding of the process selected; proposal for reengineering the business process selected; and the results of the comparison between the current situation of the process and the proposed reengineering of the process through the use of Business Process Model and Notation (BPMN 2.0).

**Findings** – Based on the results, the paper shows the importance of the analysis and implementation of the reengineering in a software development company. It describes step-by-step how to apply the methodology of Hammer and Champy in a business process through simulated scenarios, using BPMN 2.0. By carrying out the implementation of the proposed reengineering, the company would therefore save 45.12 percent of costs, 41.17 percent of time, and a better distribution of resources, at the same time guaranteeing the satisfaction of its clients.

**Originality/value** – The study addresses the current needs of small- and medium-sized software developers, providing a step-by-step guide to the implementation of a process reengineering methodology, performing an analysis and modeling of processes in BPMN 2.0 and providing results through a simulation of the critical process, with the aim to observe the flow of activities and the significant improvements that would be achieved by implementing the reengineering proposal. This simulation schema allows business owners to observe the potential of the changes and to verify the positive impact they would have on the company before beginning to make operational changes in the organization.

**Keywords** Business process, Software development, Process analysis, Software engineering, BPMN, Business process reengineering, Process reengineering

**Paper type** Case study

## Introduction

The information technology industry, currently one of the most dynamic and fast-growing industries in the global economy, is undergoing a major transformation at the Industry 4.0 (Lasi *et al.*, 2014; Wan *et al.*, 2016; Zhou *et al.*, 2015). The cause of the industry's remarkable global growth is that it is recognized as a generator of competitive advantages as a



---

knowledge industry identified with the rapid acceptance of change and the modernization of productive processes that integrates technological and administrative innovations in response to constant changes in the context. It offers competitive advantages by giving added value to its products and services. As a result, it has caused changes in the business environment, bringing about new challenges that demand greater managerial skills. The increase of competition in the technology sector itself has emphasized the need for these skills for companies that strive to keep up with their competitors (Zhou *et al.*, 2015). The software industry confronts a serious challenge due to competition, because of the constant emergence of small- and medium-sized software companies which are standing up to more established enterprises. These companies must make great efforts to improve their competitiveness and make them even more efficient. In order to strengthen these kinds of companies, efficient practices need to be adapted to their size and business type and processes need to be improved for increasing the quality and productivity of their services (Huang *et al.*, 2015; Fayad *et al.*, 2000).

The process for determining the quality of the product or services is one of the most important activities in the administration of a company; thus, it is advisable that all the company's processes are connected with the generation of products and service be aligned to a well-defined objective, and be constantly subject to review in response to changes, with the aim to improve internal processes and/or reduce defects in the products or services (Huang *et al.*, 2015). Hence, business processes reengineering (BPR) has helped companies to adapt to the highly competitive and rapidly changing market environments by incrementally or radically improving their processes. BPR involves a radical change in the company's manner of acting and it is defined as "fundamental reconsideration and radical redesign of organizational processes, in order to achieve dramatic improvements of current performance in cost, speed, and quality of service" (Hammer and Champy, 2009). However, BPR has also been considered as a process improvement approach in many projects over the years (Anand *et al.*, 2014; Attaran, 2003; Bertolini *et al.*, 2015; Bevilacqua *et al.*, 2009). According to Zellner (2011) since BPR were considered as an essential part of the business improvement, it could be found in the literature in many different ways such as business process redesign (Reijers and Mansar, 2005; Davenport and Short, 1990), business reengineering (Hammer and Champy, 1993), continuous improvement process (Deming, 2000), among others. BPR as a radical change is considered a high level of risk (Goksoy *et al.*, 2012) where about 70 percent of the projects failed (Omidi and Khoshtinat, 2016; Seethamraju and Marjanovic, 2009). In this case study, BPR is considered as a process improvement using the Hammer and Champy (1993) methodology.

Reengineering became popular in the early of 1990s when Hammer and Champy (1993) introduced this approach in their book titled *Reengineering the Corporation: A Manifesto for Business Revolution*. Hence, in order to respond to changes, increase profits, reduce costs, reduce process times, improve the quality of products and services or simply to improve productivity, the reengineering of company's processes has become an option since the 1990s (Attaran, 2003). However, according to Radhakrishnan and Balasubramanian (2010), the approach was not completely understood or appreciated at the beginning because of the poorly planned and executed improvement projects. In addition, it is well known that regardless of the application of reengineering as a radical change or process improvement, addressing the human factor is crucial to the success of any reengineering project. Thus, the explanation and persuasion of employees to understand and accept the changes within the company, especially the way they do their daily activities is necessary to obtain the expected results because they are the vehicle for change who will either embrace or resist change (Goksoy *et al.*, 2012; Anand *et al.*, 2014). In the last decade, many companies continue embracing reengineering with the aim to be more competitive and efficient integrating business websites, enterprise systems as well

---

as business intelligence systems where today's businesses demand the use of technology in the process redesign to obtain more information about their customers, suppliers, products, competitors, etc. According to Attaran (2003), the redesign of the processes will be between the company and its customers, suppliers and partners, this is known as process X-engineering proposed by James Champy (2002) where Cisco, Dell, Intel and Solecron have successfully implemented X-engineering.

Also, Tka and Ghannouchi (2014) considered that BPR is one of the existing methods in the literature that leads to evolutionary changes and adjustments that have become a necessity in today's business process improvement. According to Chan and Peel (1998), organizations carry out a reengineering project for external and internal reasons or factors. Internal factors occur within a company such as the need to improve technology, increase efficiency, reduce cost and define or redefine strategic goals. Otherwise, external factors are related to the pressures exerted by customers, competitors, market conditions and governmental regulations or political pressures. Thus, companies that aim to improve their competitiveness and increase profits decide to implement a reengineering but, since they often fear radical change, they choose for incremental improvement of their processes. For this, a business process model is a fundamental stage in the reengineering, which helps the organization to better understand its current processes and to plan the transition to the redesigned processes. Petrillo *et al.* (2018) propose a model as a valuable tool to facilitate a successful design of business process reengineering in project management which is intended to help companies to operate projects for transfer and optimization of production lines. It is expected that using Agile Reengineering Performance Model tools will facilitate to modify the evolution of the project, with the possibility of expanding or improving the application if necessary. In this way, case studies provide guidelines for BPR projects in different sectors such as financial institutions (Shin and Jemella, 2002), public administrations (Hesson *et al.*, 2007; Rinaldi *et al.*, 2015), agri-food sector (Bevilacqua *et al.*, 2009), seaport sector (Islam *et al.*, 2013), pharmaceutical industry (Bertolini *et al.*, 2015), among others. In the same way, contemporary companies responsible for software development are under enormous pressure to deliver their software products quickly, within the prescribed timeframe, and with the highest quality and lowest cost, which implies more dynamism, invisibility and exclusivity (Anand *et al.*, 2014).

In this regard, Mexican software companies are innovating in their internal processes to remain competitive and offer better products and services to their clients. In this paper, a case study is conducted to address a BPR project in a Mexican Technology Services Company, whose core activity is custom and commercial software development. Although the company is relatively new, its products are present in about 19 countries, reflecting the high commitment to hard work and teamwork. To improve its competitiveness, the company was interested in carrying out a reengineering project to redesign some of the company processes to better support the company's mission, strategic goals and customer needs. Mainly, the biggest concern was to reduce the delay time of software development, since it is the main business activity of the company. In addition, the reengineering project could help them align the mission and vision, standardize the processes, redefine the responsibilities of each role of the team members, among other aspects related to the organizational situation.

This study provides detailed information about a BPR process through the Hammer and Champy methodology for a Mexican technology-based company. Some of the actions that took place in the redesign are related to a better distribution of resources, while at the same time ensuring the quality of work and customer satisfaction. Even more, the benefits of the proposed approach are presented through simulation results where the company would save 45.12 percent in cost and the development time would be reduced by 41.17 percent.

---

### **Business process reengineering methodology**

In the last years, there have been technological innovations, new business models, rapidly changing market place and increasingly demanding customers, which has led companies to face difficult conditions in their daily operations, late deliveries and a continuously increasing competition. Hence, a business reengineering project is often a solution to improve companies' profit, customer satisfaction, costs reduction, respond better to competitive pressures, among others. Several methodologies have been proposed to support a business reengineering process, such as the Hammer and Champy, Davenport, Manganelli and Klein and Kodak (Radhakrishnan and Balasubramanian, 2010). Usually, each methodology shares three phases such as an analysis phase, a design phase and an implementation phase which shows similarities and differences. For example, in the analysis phase each approach differs in the scope of project preparation. Hammer and Champy (1993) focus on analyzing the company situation and the project preparation for a successful implementation. Davenport methodology emphasizes on information technology for innovating business processes. Manganelli and Klein's methodology suggests concentrating on those businesses that directly support the strategic goals focusing on a social and technical design which include organizational and personnel development plans. Finally, Kodak methodology is influenced by Hammer and Champy methodology, nevertheless, in the final step related to the change management, this is performed parallelly to the first four steps (project initiation, process understanding, new process design and business transition). Hammer and Champy (1993) describe the BPR as the radical change of business processes to obtain improvements related to cost, quality, service and quickness. However, Radhakrishnan and Balasubramanian (2010) point out that BPR could be a significant improvement to a business process or a radical change depending on the company needs. Michael Hammer proposed the Business Reengineering concept in the Harvard Business Review titled "Reengineering Work: Don't Automate, Obliterate." Hammer and Champy argue that the main problems of a fail project of process reengineering are the lack of efficient management, unclear objectives and resistance to change. Thus, they proposed a methodology for a successful reengineering project known as the Hammer and Champy methodology. Radhakrishnan and Balasubramanian (2010) synthesized this methodology in six steps: introduction to the reengineering of business processes; identifying business processes; selection of the business process to be redesigned; understanding of the selected business process; redesign of the selected business process; and implementation of the redesigned business process. In this regard, the selection of a methodology for the BPR project is according to the particular situation of the company taking into account strategic goals, economic situation and competitive requirements. On the other hand, other approaches, such as capability maturity model, capability maturity model integration, SPICE (ISO/IEC 15504), among others (Khan *et al.*, 2019), have been proposed for technology-based companies where the focus is the software process improvement. Those approaches are used when the company needs a maturity model to improve the activities carried out exclusively for software development, thus, this maturity model does not consider the mission and strategic objectives of the company during its transformation. Therefore, a reengineering methodology is more suitable for the company in this case study, since, the critical problems were at a strategic level and not in the software development itself. In this regard, Hammer and Champy methodology was used to conduct the reengineering project for the software development process because of the project preparation step for redesigning processes (Table I).

### **Case study**

This research was carried out on a software development company located in the southern part of the state of Sonora, in Mexico. The company was created in 2013 out of an

Step No.	Project steps	Objectives
1.	Introduction to the Reengineering of Business Processes	The CEO of the company begins the project. He describes briefly and programmatically the current situation of the business in order to initiate the necessary actions
2.	Identifying business processes	Examination of the general panorama of how processes interact within the company and in relation to the outside world. A deliverable of this process is a diagrammatic representation of all the processes
3.	Selection of the business process to be redesigned	Selection of one or more processes to be redesigned. All those are selected that provide significant value to the company's clients, as well as those which, by their nature, are easy to redesign
4.	Understanding of the selected business process	This is not limited to a detailed analysis of the functioning of the business process selected. It focuses rather on the performance of current processes in comparison to how they are expected to function in the future
5.	Redesign of the selected business process	According to Hammer and Champy, this is the most creative step, characterized by imagination, lateral thinking, and a touch of madness
6.	Implementation of the redesigned business process	This step covers the implementation phase of the business process reengineering project. Hammer and Champy do not say as much about implementation as they do about project planning. They believe in the success of the implementation once the five previous steps have been adequately carried out

**Source:** Radhakrishnan and Balasubramanian (2010)

**Table I.**  
Steps for the Hammer/  
Champy methodology

entrepreneurial idea that consisted of developing a system for administering the educational resources of a university. After a couple of years, several academic institutions in the region decided to try it out, and they were pleased with the results. Over time, what began as school software became a management system for human resources in transnational companies. Currently, the company is one of the best-known software companies in northwestern Mexico, with more than 15,000 clients in 19 countries. The company offers solutions and services through the use of successful and innovate agile methodologies and tools. Among the services offered by the company are warehouse and transport management systems, as well as systems for payroll, accounting and billing. The company also offers custom software development, contributing business value to its clients with speed and quality. The organizational structure of the company is without a strict hierarchy, with a focus on decentralization, few management levels, and horizontal communications, oriented toward multifunctional work and flexible limits between positions, and with an emphasis on teamwork and great interest in the client. Hence, the company offers technological solutions through the application of processes, models and standards. It is focused on satisfying client's needs in the areas of software development, software deployment and support in any business sector. Every two weeks, the company carries out a series of critical activities connected with the software development process, which must be completed within a timeframe agreed on with the client, by contractual obligation. It is currently difficult to maintain adequate control of these activities in order to ensure the delivery of the final product at the end of the two weeks. It is very common that, for the smaller projects, they keep the delivery date in mind, with all the necessary precautions taken. Nevertheless, it is more common for the larger projects to be delivered late, as requirements are constantly changing and there is no definite follow-up of the most important metrics in the process. This is because of the roles of various members of the work team are not well established. One of the main risks of continuing with the current workflow is, therefore, the possibility that the software not be published on time and the dissatisfaction of the client.

### *Research design*

The research design for the case study consisted of multiple visits to the company where a CEO and two software developers mainly participated in seven formal interviews. However, the authors had the opportunity to discuss with other developers about their activities during the software development process, either to clarify or explain and thereby informally collect data. For this purpose, there are about 15 developers who have different roles and responsibilities in the company, not only in the software development process. Moreover, multiple visits to the company were made in a period of eight months in order to collect data from various sources such as direct observation, directed questions, e-mail exchanges, document records and group discussions. First, a situational analysis of the company was carried out for the first four months where the mission and business vision were reformulated, the strategic objectives were redefined, and a value chain was designed, see Figures 1–3. In addition, the reengineering project was carried out for the following four months using the Hammer and Champy methodology.

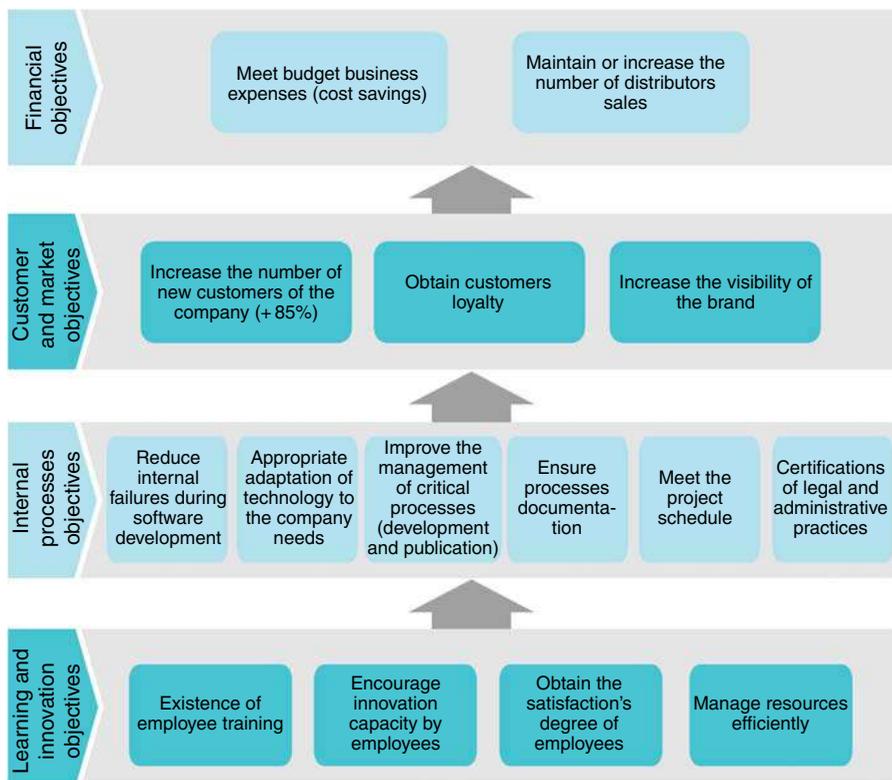
### *Reengineering of the software development process*

The software development process is redesigned following the Hammer and Champy methodology, described in Radhakrishnan and Balasubramanian (2010). In this manner, this methodology involves executing six steps: introduction to the reengineering, identify the main business processes, selection and understanding the selected business process, redesign and implementation. Therefore, an analysis of the organizational situation was made to know about the strategic planning of the company in order to redesign the critical process aligned with the company's objectives. A description of these steps is presented as follows.

*Step 1: introduction to the reengineering of business processes.* First, a situational analysis of the company was carried out in order to diagnose the situation and operation of the company, as well as, the identification of the opportunities and threats offered by the different factors in the environment and recognize the strengths and weaknesses of each of the elements that compose the organization. Furthermore, the reengineering project is



**Figure 1.**  
Vision, mission and  
values of the company

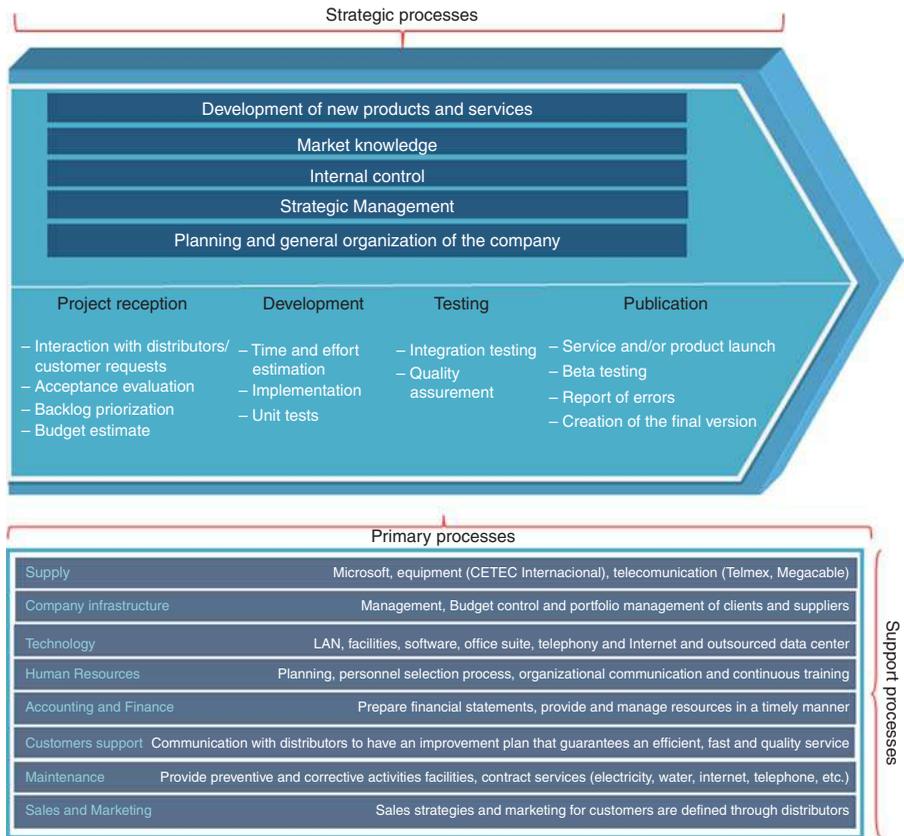


Software development process

**Figure 2.**  
Company's strategic objectives

delimited to a period of four months. For this purpose, the mission, vision and strategic objectives of the company were reformulated and the value chain was designed, see Figures 1–3. As can be seen in Figure 2, one of the objectives of the internal processes is to improve the management of critical processes, being the software development process, the main critical process for the company. The other strategic objective related to the internal processes is to reduce internal failures during the software development process; for example, adapt technology to the specific needs of the projects, ensure the availability of information about the process and meet the deadlines established when launching new projects or delivering advances on existing projects. The company therefore has a clear need of carrying out a reengineering of its software development process, since this is a critical process to which attention must be paid if the company wants to improve its competitiveness, increase its new client base, reinforce the loyalty and satisfaction of its existing clients and heighten the visibility of its brand.

*Step 2: identifying business processes.* After, the situational analysis, eight main processes were identified for accomplishing the company's daily activities, such as the software development process, marketing and sales, client support, recruitment and selection hiring process, human resources planning, finance and accounting, supply planning and maintenance services, see Table II. The company's business processes were identified and analyzed in a general way in order to select the critical one based on their needs. For this purpose, a list of priorities was established where the software development process was selected as the critical process to be redesigned.



**Figure 3.**  
Value chain of  
the company

*Step 3: selection of the business process for reengineering.* The selection of the critical business process for the reengineering project was the software development process due to the time-consuming of some activities represented thereby for the company money that could be invested in another project; additionally, some roles have a higher workload than others. The following criteria were used for the selection of the critical business process:

- state: identifying the processes that are ineffective or where problems have been detected in the activities such as time-consuming, work overload, lack of communication, among others;
- relevance: comparing the relative importance of each process, prioritizing those processes where there is a greater influence of the client and considering the objectives and strategic business plan of the company; and
- feasibility: considering the feasibility of the reengineering process by identifying the factors that determine the probability of success of the process changes in a period of time.

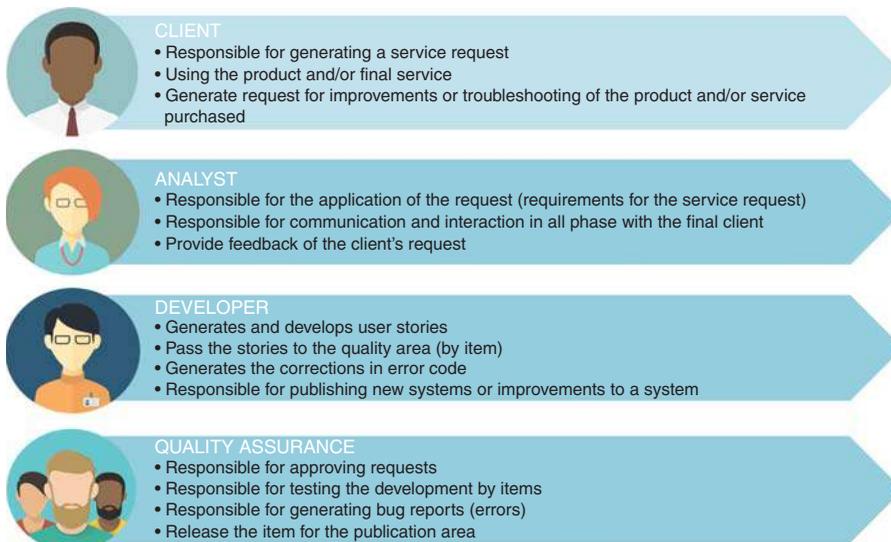
In other words, the critical processes for reengineering, in general, are those considered as inefficient or delayed, with a great impact on clients that contribute to the fulfillment the strategic objectives of the company. Thus, the selected process for this case study is the software development process, as it is the most relevant process for the company with a significant impact on the clients.

Software development process

No.	Area	Process	Description	Process owner
1	Software development	Software development process	This process begins when a request is registered by a client and continues with a series of activities involved in the development and final publication of software	Project management and software developer
2	Sales	Marketing and sales	Sales and marketing strategies are defined through distributors for increasing sales	Sales manager
3	Sales	Client support	The communication is established with the distributors in order to have an improvement plan for an efficient, fast and quality service	Sales manager
4	Human resources	Recruitment and selection hiring process	Recruitment of high-skilled professionals is crucial for the company success	Human resources manager
5	Human resources	Human resources planning	Growth planning, organizational communication and continuous training	Human resources manager
6	Finance	Finance and accounting	Prepare financial statements, provide and control resources with opportunity	CEO
7	Finance	Supply planning	Guarantee the necessary software, hardware and telecommunications equipment to perform daily activities	Project management
8	Finance	Maintenance service	Provide preventive and corrective activities of the facilities (electricity, water, internet, telephone)	CEO

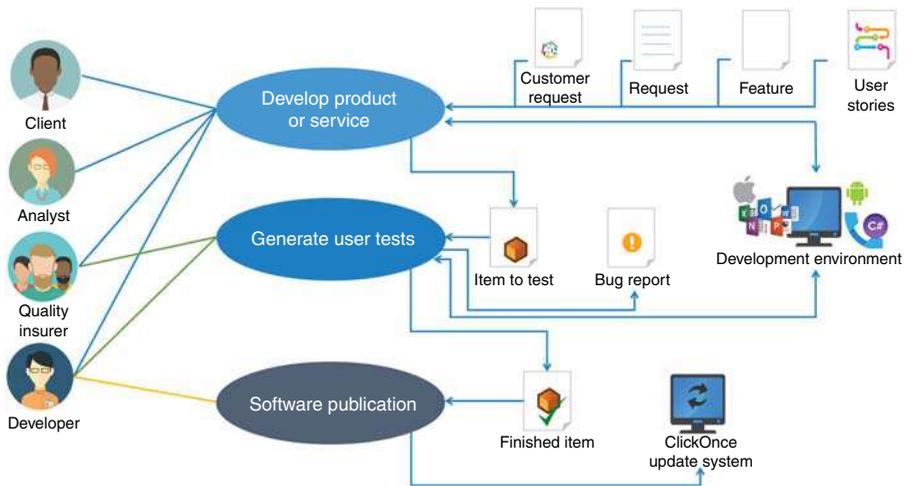
**Table II.**  
Company's main processes

*Step 4: understanding the selected business process.* The aim of this step of the Hammer and Champy methodology is to provide a high level view of the performance of the selected process in order to have the necessary knowledge and insight to create a totally new design that improves the current conditions of the process. Thus, four roles are involved in the selected software development process: the client, the analyst, the developer and the quality assurance, see Figure 4. In this respect, the client is the one in charge of requesting the service; the analyst is responsible for requesting the requirements and is responsible for

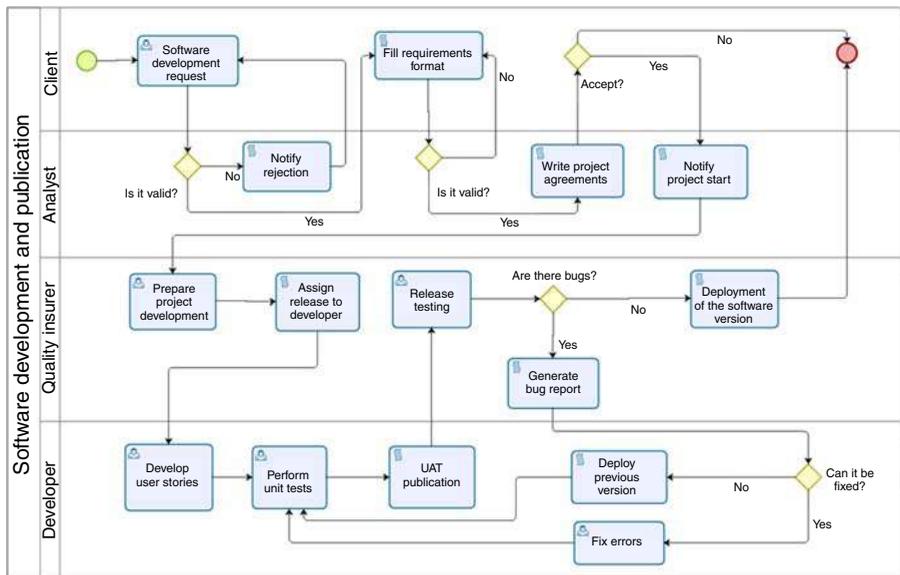


**Figure 4.**  
Roles involved in the process of developing software

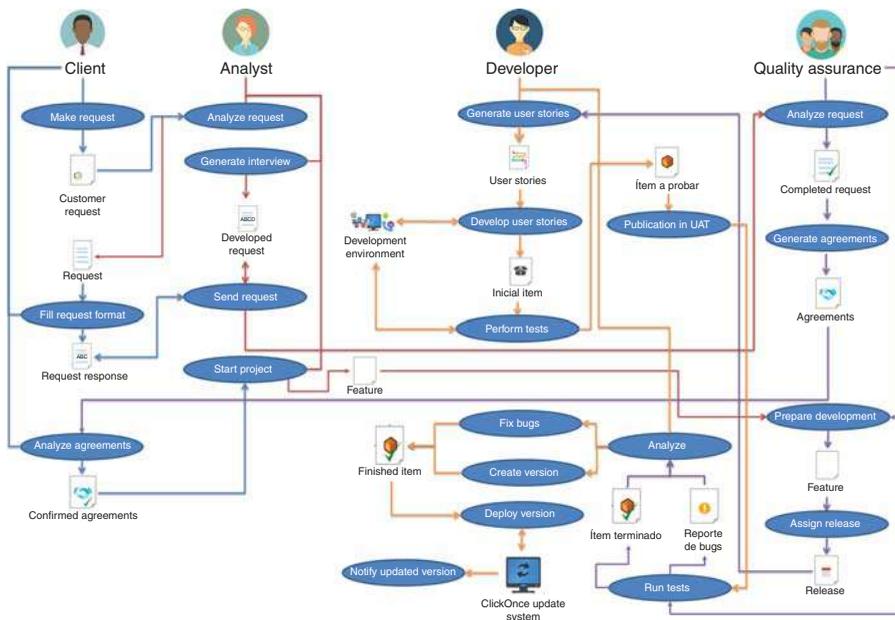
maintaining the communication and interaction with the client; the developer is responsible for generating and developing the user's stories as well as correcting code errors and modifications related to the project; finally, the quality assurance is responsible for testing the item, generating error reports and deciding when to send the item to the publication area. In this way, the interaction of each of these roles and their activities can be seen in Figures 4 and 5. In addition, Figures 6 and 7 show in detail the activities and artifacts involved in this process. On the other hand, the artifacts are referred to the input/output item indicated by an arrow; for example, in Figure 7 the client's request a document, user stories,



**Figure 5.** General vision of the current software development process



**Figure 6.** BPMN diagram of the current software development process



**Figure 7.** Rich picture of the current process of developing software

bugs reports, among others. Figure 6 shows the current workflow through a BPMN diagram (Allweyer, 2016) indicating the detail activities carried out by each role.

This process requires the fulfillment of four stages: project reception, construction, testing and publication, see Figure 3. Hence, when the projects are received, it is common to observe that the work team is stressed to meet the deadlines of the project delivery. The principal reason is because in this process there are activities carried out by people who have other assignments and chores, so it is difficult to comply within the established date. Sometimes, the employees forget the responsibilities of their role in the company, either due to the lack of documentation of the process and lack to communication. Also, through the interviews with the managers and the work team, it was found that the software development process has many areas of opportunity and improvement.

*Step 5: redesign of the selected business process.* In order to redesign, the selected business process is mandatory to understand the current process and identify the activities that must be improved. Hence, we propose the redesign of the software development process through the process modeling represented by a rich picture and BPMN 2.0. The Business Process Model and Notation (BPMN) is a language that has been developed on a solid mathematical foundation provided by the theory of process calculus, an essential requirement for automating execution and easily providing proofs of general consistency properties. In order to describe a workflow, BPMN provides a diagram of business process with a large set of elements and attributes (Cimino *et al.*, 2017). Thus, the notation allows the people involved in the business processes to understand how the flow of activities is carried out within a given process, and to observe the interactions among roles, activities and artifacts.

The activities identified as critical in the software development process actually are assigned to people who have other roles within the company, see the activities presented in Figure 7. On the other hand, Figure 9 presents the rich picture of the reengineered process where the role of project manager is included as well as the corresponding activities. In this respect, a list of the most important changes carried out in the redesigned process is

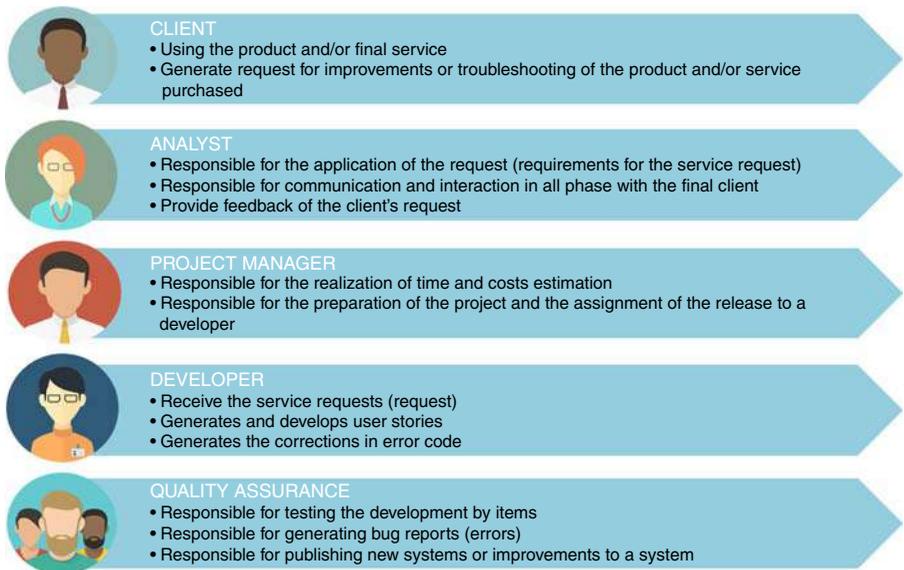
BPMJ

presented in Table III, where the activities are simplified and the roles are clearly defined along with their responsibilities for a better flow of the process. Moreover, during the visits to the company, a lack of process documentation was found, as well as, a lack of a clear definition of roles and responsibilities. It is therefore, we propose to include a new role named Project Manager, see Figure 8. The project manager will be responsible for the project preparation and team assignments, as well as, schedule the deliveries. Thus, some tasks were simplified to reduce time and costs, see Figures 7 and 9. The redesigned process in BPMN is presented in Figure 10 where the role of project manager is included improving the organization of the software project.

*Step 6: implementation of the redesigned business process.* This step consists the implementation of the redesigned business process according to the Hammer and Champy methodology. However, this case study presents simulation results for the redesigned process as well as the current process in order to quantitatively know about the benefits that the company could obtain by applying the proposed changes. According to Goksoy *et al.* (2012), the simulation of the reengineering process through a technological tool has proven

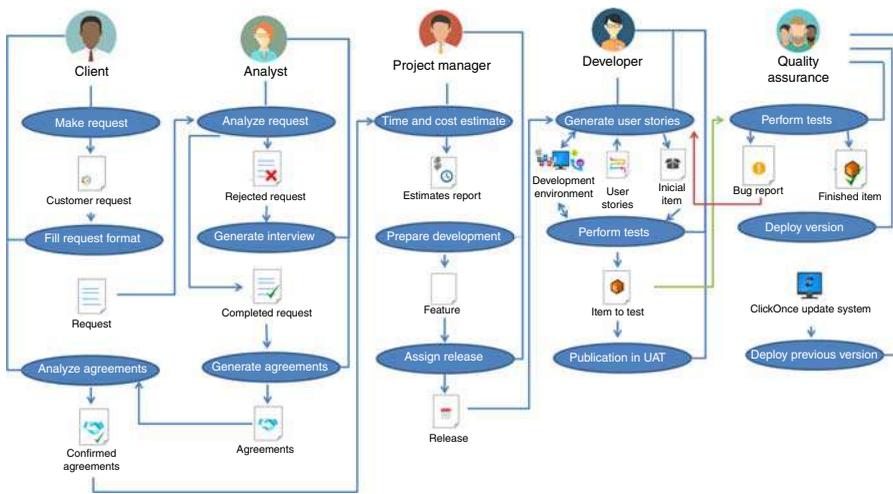
**Table III.**  
Main modified activities in the software development process

No.	Proposed activities
1	The client requests are now automatically approved when the requests are related to an existing system
2	The quality assurance activities are now limited to the generation of quality tests, detection of bugs, provide error reports and the deployment of updated versions
3	The role of project manager is now included in the process, as well as, their corresponding activities and responsibilities. These activities correspond to time and costs estimation, preparation of the project, team and member assignments, as well as, schedule the deliveries
4	Specific activities were established for each role, see Figure 8
5	Some tasks were simplified to reduce time and costs, see Figures 6 and 8
6	Some activities were eliminated and improved for a better process performance, see Figures 6 and 8

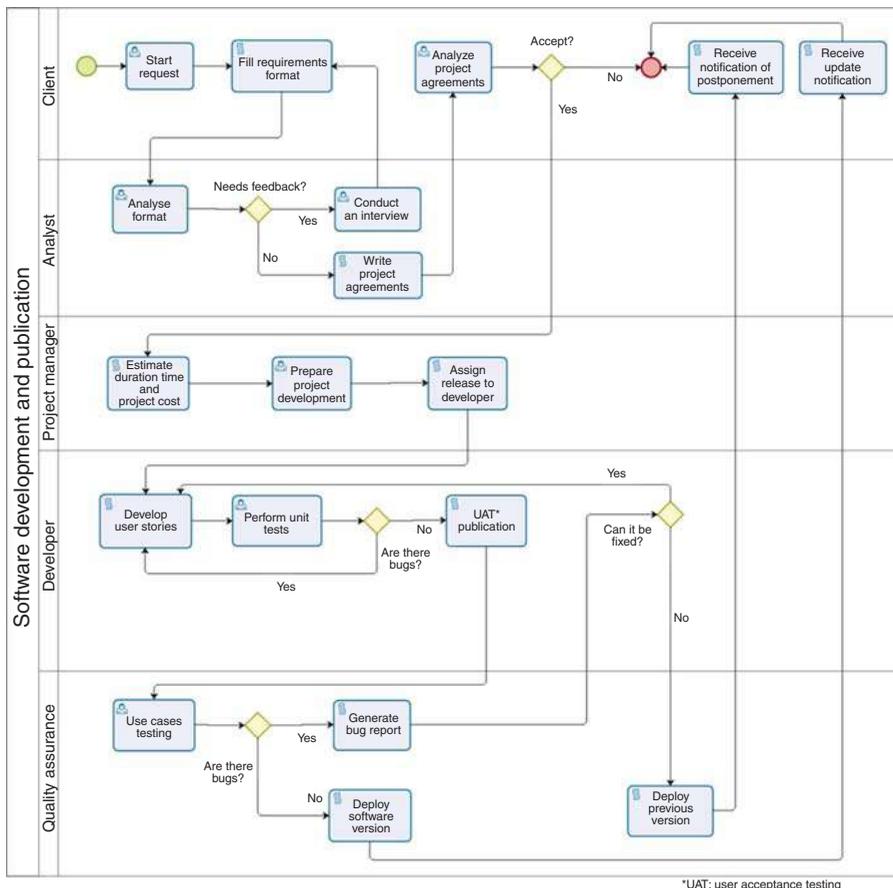


**Figure 8.**  
Proposed roles for the software development process

# Software development process



**Figure 9.** Rich picture of the reengineering software development process



**Figure 10.** Proposed BPMN modeling for the redesign of the software development process

to be an effective to determine before applying the changes to know the productive impact and also, clarify the problems presented in the process.

The simulation of the software development process, both, current process (as-is) and the redesigned process (to-be) was presented to the company in order to implement the proposed changes. The CEOs as well as the team were very satisfied with the proposal, therefore, they were interested in redesigning other company processes. In this way, they will decide when to implement the changes in the company in order to reduce the publishing time of the software development process, save costs in the company and increase customer satisfaction.

**BPMN simulation results**

The feasibility of the proposed redesign process is presented through two simulated scenarios using the software Bizagi ([www.bizagi.com/](http://www.bizagi.com/)) modeling the as-is and to-be processes in BPMN. A simulation is a useful tool, not only for identifying the activities that are causing delays and misuse of resources, but also, for obtaining comparisons between the processes and how the activities are carried out taking to account the required time to accomplish all the activities. Moreover, if bottlenecks are presented in the process, then, they could be localized in the simulation. In this regard, the data such as time and costs of each activity were collected for the simulation scenarios. The costs of the activities per hour in both scenarios were calculated based on the employee’s monthly payment and the hours worked in a month. Thus, the cost per hour of the activities is the same in both models due to the number of working hours and the employee’s monthly payment remains the same in the to-be model. Tables IV and V present the results in both scenarios where the reduction of time, cost and effort is noticeable. This reduction is mainly because some activities were simplified, improved and some others eliminated, setting clear and specific activities for each role. Moreover, the addition of the project manager as a role in the to-be model was indispensable since his activities and responsibilities were doing by the quality assurance and developer. In this sense, the workload of the roles in the to-be was better distributed.

*Scenario*

The software company has two principal business services, the first is the development of software projects (new project to be implemented) and the second is the improvement of an

**Table IV.**  
Simulation results of the current process

Resource	Use (%)	Total cost (\$)
Developer	44.03	7,951.42
Analyst	2.98	538.40
Quality assurance	50.91	9,140.19
Client	2.08	0.00

**Table V.**  
Simulation results of the redesign process

Resource	Use (%)	Total cost (\$)
Developer	39.83	3,950.63
Analyst	4.34	430.00
Quality assurance	39.56	3,923.75
Client	2.45	0.00
Project manager	13.82	1,370.63

existing system (modification of functionality and/or fixing bugs). For this case study, the most frequent situation faced by the company is the improvement of an existing system. Hence, we present comparisons of time, cost and effort measured in man-hours of the current and the proposed process using the same initial conditions. The parameters for the simulation scenario are the following.

*Scenario parameters*

Activities duration: 14 days

Currency: MXN – Mexican Peso

Initial time: 08.00

Duration of the working day: 8 h

Resources: analyst, quality assurance, project manager, developer, client.

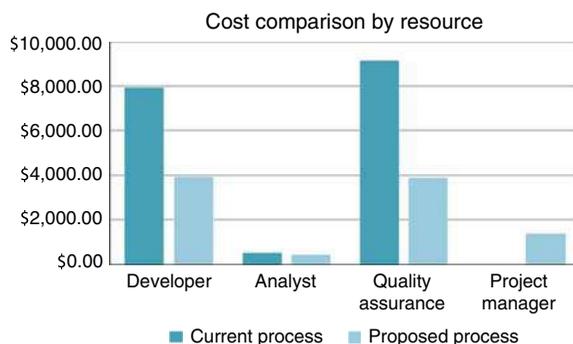
Cost: proportional to the hourly wage and the duration of each activity.

*Time, cost and effort comparisons*

The activities causing a delay in the delivery of the software will cause an additional cost for the company along with the client dissatisfaction. For example, the activities carried out by the quality assurance represent 50.91 percent of all the activities in the current software development process, see Table IV, while in the redesigned process represent 39.56 percent, see Table V, meaning a reduction of \$5,216.44MXN only for the quality assurance during the software development process. Consequently, a reduction of 45.12 percent of the costs in the overall process could be save the company implementing the redesigned process, see Table VI for an overall costs comparison. In this way, the savings are mostly due to the modification of the sequence of activities in order to reduce the cycle cost, the elimination of unnecessary activities and, the reduction of some tasks, see Figure 11.

Resource	Cost of current process (\$)	Cost of proposed process (\$)
Developer	7,951.42	3,950.63
Analyst	538.40	430.00
Quality assurance	9,140.19	3,923.75
Client	0.00	0.00
Project manager	0.00	1,370.63
Total cost	17,630	9,675

**Table VI.**  
Costs comparison



**Figure 11.**  
Cost comparison by resource between the current and proposed

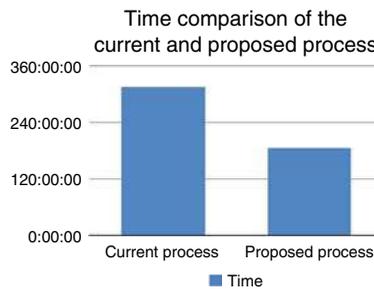
According to the simulation scenario, the current process requires 313:34:00 h to accomplish all the activities and deliver the software project, meanwhile, the proposed process would take 184:31:00 h, saving 129:03:00 h. This is a reduction of 41.17 percent of the total time required to finish the project, see Figure 12. The reduction in time is mainly due to the optimization of tasks and the new role added (project manager) in the redesigned process, see Figures 6, 7, 9 and 10, where the sequence of activities was modified to reduce the cycle time, unnecessary waiting times were reduced and activities corresponding to each role were reassigned.

According to the results, an inefficient process flow was found in the current process presenting an overload of activities in the quality area, with a 50.91 percent use of resources in order to complete all the activities of the process and deliver the development project, see Figure 13. Furthermore, an overload of activities and responsibilities of the quality assurance were detected in the current process. Finally, the lack of the project manager role was detected, who is responsible of the project planning, project cost and time estimation to mention but a few; thus, this role is included in the reengineering proposal since the project manager already work in the company, which means that there are no additional payments for this person.

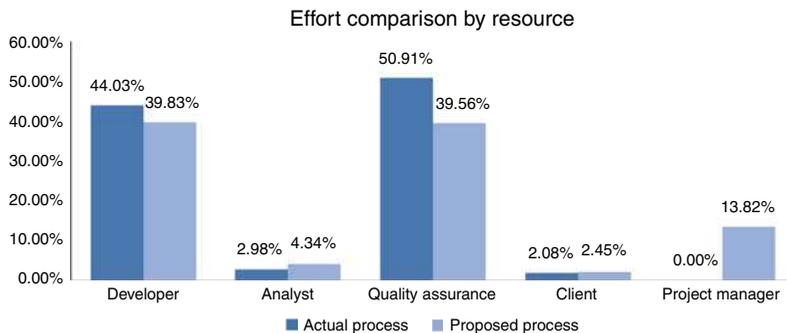
### Conclusions

This paper presents a case study of a reengineering process for a Mexican Software Development Company. According to the company mission and strategic goals, they are interested in creating high-quality solutions and services that exceed the expectations of its clients, thereby earning their loyalty and increasing the visibility of its brand. Currently, the company faces a challenge related to its software development process due to their constant concern about to meet deadlines agreed with the client and the continuing changes on the requirements. The methodology of Hammer and Champy was used to carry out the

**Figure 12.**  
Total time comparison  
of the current and  
proposed process



**Figure 13.**  
Effort comparison  
by resource for  
current and proposed  
process



reengineering project where the process modeling with the interactions of roles, activities and artifacts was represented through the BPMN diagram. Furthermore, two scenarios with the same initial conditions were formulated using Bizagi simulation software. One scenario simulates the current process and another simulates the reengineered process; as a result, the redesigned process presented a significant improvement in times and costs. Moreover, according to the results of the redesigned process, resources are redistributed in a balanced manner, in accord with the responsibilities of each role within the company allowing the optimization of task flows where:

- it is proposed that the activities of the quality assurance role correspond only to the generation of quality tests and detection of errors in the developed items, in addition to generating the error reports and the display of updated versions; and
- the set of activities that corresponds to the new role of project manager are established. It involves mainly those related to the project preparation and the assignment of the project release to a developer.

According to the results obtained from the analysis of the company's software development process, the reengineering proposal is a simple and a viable solution which would not require the company to invest in technology or human resources, but merely to make better use of the resources that already possesses. The reengineering of the software development process would have a positive impact reducing costs, times and effort in addition to guaranteeing quality and client satisfaction through development of projects that ensure an optimal flow of activities.

### Future directions

In order to accomplish a good performance of the software development process in the implementation of the reengineering proposal, it is suggested that the following actions are to be taken into account to obtain greater benefits:

- access to project information: allow the access to the project progress registry platform to all roles involved in a given ongoing project;
- follow-up and control of activities: the project manager should follow-up on employee activities in real time in order to maintain control of the employees' performance and so be able to detect delays before they become critical to the delivery of the project;
- use of data and data availability for new analyses: perform periodic reports showing the performance of the software development process in order to diagnose being able to make effective decisions; and
- design a mechanism for estimating costs and time: determine and document the estimated duration and cost of the all the activities required to carry out the process, which will make for more effective follow-up of the indicators as the project progresses.

### References

- Allweyer, T. (2016), "BPMN 2.0: introduction to the standard for business process modeling", in Demand, B.B. (Ed.), *Why a Notation?*, 2nd ed., ISBN: 978-3-8370-9331-5, Herstellung und Verlag, Books on Demand, Norderstedt, p. 9.
- Anand, G., Chandrashekar, A. and Narayanamurthy, G. (2014), "Business process reengineering through lean thinking: a case study", *Journal of Enterprise Transformation*, Vol. 4 No. 2, pp. 123-150, doi: 10.1080/19488289.2013.879681.

- 
- Attaran, M. (2003), "Information technology and business-process redesign", *Business Process Management Journal*, Vol. 9 No. 4, pp. 440-458.
- Bertolini, M., Bevilacqua, M., Ciarapica, F.E. and Postacchini, L. (2015), "Business process reengineering of drugs storage and distribution: a case study", *International Journal of Procurement Management*, Vol. 8 Nos 1-2, pp. 44-65, doi: 10.1504/IJPM.2015.066287.
- Bevilacqua, M., Ciarapica, F.E. and Giacchetta, G. (2009), "Business process reengineering of a supply chain and a traceability system: a case study", *Journal of Food Engineering*, Vol. 93 No. 1, pp. 13-22.
- Champy, J.A. (2002), *X-Engineering the Corporation: Reinventing Your Business in the Digital Age*, Warner Books, New York, NY.
- Chan, P.S. and Peel, D. (1998), "Causes and impact of reengineering", *Business Process Management Journal*, Vol. 4 No. 1, pp. 44-55.
- Cimino, M.G., Palumbo, F., Vaglini, G., Ferro, E., Celandroni, N. and La Rosa, D. (2017), "Evaluating the impact of smart technologies on harbor's logistics via BPMN modeling and simulation", *Information Technology and Management*, Vol. 18 No. 3, pp. 223-239.
- Davenport, T.H. and Short, J.E. (1990), *The New Industrial Engineering: Information Technology and Business Process Redesign*, Vol 31, No. 4, Sloan Management Review, Center for Information Systems Research, Massachusetts Institute of Technology, pp. 1-31.
- Deming, W.E. (2000), *The New Economics for Industry, Government, Education*, 2nd ed., ISBN: 9780262541169, MIT Press.
- Fayad, M.E., Laitinen, M. and Ward, R.P. (2000), "Software engineering in the small", *Communications of the ACM*, Vol. 43 No. 3, pp. 115-118, doi: 10.1145/330534.330555.
- Goksoy, A., Ozsoy, B. and Vayvay, O. (2012), "Business process reengineering: strategic tool for managing organizational change an application in a multinational company", *International Journal of Business and Management*, Vol. 7 No. 2, pp. 89-112.
- Hammer, M. and Champy, J. (2009), *Reengineering the Corporation: Manifesto for Business Revolution*, ISBN: 9780061808647, p. 272.
- Hammer, M.M. and Champy, J.A. (1993), *Reengineering the Corporation: A Manifesto for Business Revolution*, HarperBusiness Essentials, New York, NY.
- Hesson, M., Al-Ameed, H. and Samaka, M. (2007), "Business process reengineering in UAE public sector: a town planning case study", *Business Process Management Journal*, Vol. 13 No. 3, pp. 348-378.
- Huang, S.Y., Lee, C.H., Chiu, A.A. and Yen, D.C. (2015), "How business process reengineering affects information technology investment and employee performance under different performance measurement", *Information Systems Frontiers*, Vol. 17 No. 5, pp. 1133-1144.
- Islam, S., Olsen, T. and Daud Ahmed, M. (2013), "Reengineering the seaport container truck hauling process: reducing empty slot trips for transport capacity improvement", *Business Process Management Journal*, Vol. 19 No. 5, pp. 752-782.
- Khan, A.A., Keung, J., Niazi, M., Hussain, S. and Shameem, M. (2019), "GSEPIM: a roadmap for software process assessment and improvement in the domain of global software development", *Journal of Software: Evolution and Process*, Vol. 31 No. 1, pp. 1-12.
- Lasi, H., Fettke, P., Kemper, H.G., Feld, T. and Hoffmann, M. (2014), "Industry 4.0", *Business & Information Systems Engineering*, Vol. 6 No. 4, pp. 239-242.
- Omidi, A. and Khoshtinat, B. (2016), "Factors affecting the implementation of business process reengineering: taking into account the moderating role of organizational culture (case study: Iran Air)", *Procedia Economics and Finance*, Vol. 36, pp. 425-432.
- Petrillo, A., Di Bona, G., Forcina, A. and Silvestri, A. (2018), "Building excellence through the Agile Reengineering Performance Model (ARPM): a strategic business model for organizations", *Business Process Management Journal*, Vol. 24 No. 1, pp. 128-157, doi: 10.1108/BPMJ-03-2016-0071.

- 
- Radhakrishnan, R. and Balasubramanian, S. (2010), *Business Process Reengineering: Text and Cases*, 2nd ed., PHI Learning, Sonipat.
- Reijers, H.A. and Mansar, S.L. (2005), "Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics", *Omega*, Vol. 33 No. 4, pp. 283-306.
- Rinaldi, M., Montanari, R. and Bottani, E. (2015), "Improving the efficiency of public administrations through business process reengineering and simulation: a case study", *Business Process Management Journal*, Vol. 21 No. 2, pp. 419-462.
- Seethamraju, R. and Marjanovic, O. (2009), "Role of process knowledge in business process improvement methodology: a case study", *Business Process Management Journal*, Vol. 15 No. 6, pp. 920-936.
- Shin, N. and Jemella, D.F. (2002), "Business process reengineering and performance improvement – the case of Chase Manhattan Bank", *Business Process Management Journal*, Vol. 8 No. 4, pp. 351-363.
- Tka, M. and Ghannouchi, S.A. (2014), "Comparison of business process models as part of BPR projects", *Information Resources Management Journal*, Vol. 27 No. 1, pp. 53-66, doi: 10.4018/irmj.2014010104.
- Wan, J., Tang, S., Shu, Z., Li, D., Wang, S., Imran, M. and Vasilakos, A.V. (2016), "Software-defined industrial internet of things in the context of industry 4.0", *IEEE Sensors Journal*, Vol. 16 No. 20, pp. 7373-7380.
- Zellner, G. (2011), "A structured evaluation of business process improvement approaches", *Business Process Management Journal*, Vol. 17 No. 2, pp. 203-237.
- Zhou, K., Liu, T. and Zhou, L. (2015), "Industry 4.0: towards future industrial opportunities and challenges", *12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pp. 2147-2152.

### Further reading

- Chinosi, M. and Trombetta, A. (2012), "BPMN: an introduction to the standard", *Computer Standards & Interfaces*, Vol. 34 No. 1, pp. 124-134.
- Grant, D. (2016), "Business analysis techniques in business reengineering", *Business Process Management Journal*, Vol. 22 No. 1, pp. 75-88.
- Kotlarsky, J. (2007), "Re-engineering at LeCroy corporation: the move to component-based systems", *Journal of Information Technology*, Vol. 22 No. 4, pp. 265-278.

### About the authors

Laura Lopez-Arredondo is a Master's Degree student with experience in software development. She received the Bachelor Degree in Software Engineering from Sonora Institute of Technology in 2016. Currently, she is studying a Master's Degree in IT for Business at the Sonora Institute of Technology. Her research interests are data analysis, management of BI tools, machine learning and business processes.

Dr Cynthia Perez, PhD, received the Doctoral Degree with Honors at the University of Extremadura in 2014 and the MSc Degree in Computer Science at Scientific Research and Advanced Studies Center of Ensenada (CICESE). Her research interests are in the fields of artificial intelligence, machine learning and data science. Dr Cynthia Perez is the corresponding author and can be contacted at: cynthia.perez@itson.edu.mx

Jesus Villavicencio-Navarro, PhD, received the Electronic Engineering Degree at Autonomous University of Baja California, Mexico in 2001. He received the MSc Degree and PhD at Scientific Research and Advanced Studies Center of Ensenada, Mexico (CICESE). His research interest includes machine learning and signal processing.

Kathya E. Mercado received the Software Engineering Degree from Sonora Institute of Technology (ITSON), México. Her research interests include software development, data warehousing, data mining, knowledge management and business intelligence. She is currently a student of the Master's Degree in Information Technology for Business with specialization on Business Intelligence at the Sonora Institute of Technology, México.

Martin Encinas received the Bachelor Degree in Computer Science at Monterrey Institute of Technology and Superior Studies (ITESM), Obregón, Sonora, Mexico in 1990; the Master Degree in Management Information System (MSI) at ITESM, Monterrey, Nuevo León, Mexico in 1993 and the

---

BPMJ

Master Degree in Business Administration (MBA) at ITESM Virtual University in 1996. His areas of interest include Information Technology, Strategic Planning and Continuous Process Improvement (CPI).

Patricia Inzunza-Mejia is Professor and Researcher at the Autonomous University of Sinaloa. She received the Master Degree in Business and Senior Management and obtained the PhD with honors at the Autonomous University of Sinaloa in 2011. She worked as Post-doctoral Researcher at the University of Extremadura in Spain and the Department of Public Sector Economics and Public Finances of the Institute of Economic Research (IEE) at the Autonomous University of Mexico (UNAM). Currently, she is an Honorary member of the State System of Researchers and Technologists of Sinaloa. Her research interest includes Business Administration, Public Finance and Fiscal Policy.

---

---

For instructions on how to order reprints of this article, please visit our website:

**[www.emeraldgroupublishing.com/licensing/reprints.htm](http://www.emeraldgroupublishing.com/licensing/reprints.htm)**

Or contact us for further details: **[permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)**

## Estudio sobre la Percepción de los Factores Involucrados en la Estimación de Precios de Viviendas: El Caso de Cajeme

Laura P. Lopez-Arredondo<sup>1</sup>, Cynthia B. Perez<sup>2</sup>, Luis A. Castro<sup>1</sup> y Luis-Felipe Rodriguez<sup>1</sup>

(1) Instituto Tecnológico de Sonora, Campus Náinari, Antonio Caso S/N y E. Kino, Colonia Villa ITSON. C.P. 85130. Ciudad Obregón, Sonora. México. (e-mail: laurapaolapz@hotmail.com; luis.castro@acm.org; luis.rodriguez@itson.edu.mx)

(2) Instituto Tecnológico de Sonora, Unidad Guaymas, Carretera al Aeropuerto Km 3, C.P. 85400. Apartado Postal 870. Guaymas, Sonora. México. (e-mail: cynthia.perez@itson.edu.mx)

*Recibido May. 18, 2018; Aceptado Jul. 24, 2018; Versión final Sep. 13, 2018, Publicado Abr. 2019*

---

### Resumen

El objetivo de esta investigación es obtener la información necesaria para identificar la percepción sobre los factores de mayor y menor impacto durante el proceso de valuación de viviendas. Para ello, realizaron diez entrevistas semiestructuradas, las cuales se analizaron mediante Teoría Fundamentada. Los resultados de este trabajo sugieren que tanto los compradores como los vendedores de viviendas perciben a la ubicación, las proporciones y las condiciones de las viviendas como factores determinantes para estimar el valor de éstas. Los resultados de este trabajo pueden ser utilizados para mejorar el diseño de modelos predictivos para la estimación de precios de bienes raíces utilizados en plataformas en línea dedicadas a la compra-venta de bienes raíces. Así mismo, el estudio proporciona las bases necesarias para establecer las variables o parámetros que mejor se adapten al algoritmo predictivo para la estimación de precios de viviendas en la región sur del estado de Sonora, México.

*Palabras clave: investigación cualitativa; teoría fundamentada; factores determinantes; valuación de viviendas; mercado inmobiliario; compradores de inmuebles*

## Study about the Perception of Factors Involved in Housing Price Estimation: The Case of Cajeme

### Abstract

This work aims at obtaining the necessary information for identifying the perceived factors of higher or lower impact during the housing valuation process. Ten semi-structured interviews were conducted, and analyzed using Grounded Theory. The results of this work suggest that both buyers and sellers of households perceive the location, proportions and housing conditions as determining factors to estimate the value of these properties. The results of this work can be used to improve the design of predictive models for real estate which are used in online platforms. Furthermore, the study provides the bases to establish the variables or parameters used in the predictive algorithm for estimating housing prices in the southern region of the state of Sonora, Mexico.

*Keywords: qualitative research; grounded theory; determining factors; housing valuation; real estate market; real estate buyers*

## INTRODUCCIÓN

Para muchas personas, comprar o vender una casa es una de las decisiones más importantes en su vida e involucra un amplio umbral de incertidumbre acerca del valor a asignarle a la vivienda, puesto que, además de la asequibilidad del inmueble, otros factores como la ubicación o el panorama de la inversión a largo plazo afectan en mayor o menor medida al proceso de valuación. Debido a esto, la información relacionada con el valor de las viviendas es fundamental para los investigadores, organismos de gobierno, empresas de bienes raíces, vendedores de viviendas y para los propietarios de inmuebles. Esquer (2012) menciona que los métodos de valuación de inmuebles suelen ser bastante lentos y tediosos debido a que requieren de una investigación bien fundada y, en general, los procedimientos actuales deben basarse en información confiable para obtener resultados más precisos.

Asimismo, el impacto de los factores durante el proceso de valuación de viviendas varía dependiendo de la percepción que se tenga sobre estos. Carterette y Friedman (1982) definen la percepción como una parte primordial de la conciencia que consta de hechos abstraídos del mundo externo y, por tanto, se refiere a la realidad tal como es experimentada. Según Arias (2006), *“la percepción depende de la actividad de receptores que son afectados por procesos provenientes del mundo físico y puede entonces definirse como el resultado del procesamiento de información que consta de estimulaciones a receptores en condiciones que en cada caso se deben parcialmente a la propia actividad del sujeto”*. Con base en lo anterior, se puede decir que, ya sea que se desee vender o comprar una vivienda o inmueble, tener noción de cómo se determina su costo brinda control sobre el patrimonio en cuestión; además, proporciona seguridad en caso de que se necesite realizar una transacción con dicho inmueble.

Bell et al. (2015) mencionan que en el ámbito inmobiliario hay muchos métodos prescritos para obtener datos de mercado y otra información crítica; dichos métodos pueden ser de carácter cualitativo o cuantitativo. Los métodos más utilizados para este contexto son los cuantitativos debido a que, en muchos casos, los problemas inmobiliarios se prestan de manera predecible a datos cuantificables. Sin embargo, como ya se mencionó, los participantes del mercado inmobiliario tienen percepciones propias que pueden influir en los precios de venta y otros datos cuantificables. Debido a esto, el autor argumenta que el uso de métodos cualitativos puede proporcionar una mejor comprensión de los problemas y las motivaciones dentro del mercado. Durante la revisión de la literatura, se encontró que existe una considerable cantidad de investigadores interesados en el análisis cualitativo como método para obtener la percepción de las personas respecto a diversas cuestiones en áreas como la medicina, psicología, tecnología, gestión empresarial, estudio de mercados, entre otras (Figueira et al 2009; Bedoya et al., 2017; Festen et al. 2014; Riley et al. 2009; Bravo et al., 2018; Oliveira et al. 2016; Padilla et al., 2017; Pérez-Benedito et al., 2017; Kamagahara et al., 2016). Sin embargo, hay escasez de estudios en donde se lleve a cabo el análisis cualitativo para el contexto de bienes raíces.

En el ámbito inmobiliario, Kibet et al. (2016) llevó a cabo un estudio para evaluar el efecto de los factores socioculturales en las inversiones inmobiliarias. Ellos aplicaron una encuesta a una muestra de 300 propietarios de propiedades inmobiliarias comerciales en la ciudad de Kisumu en Kenia. Los datos fueron analizados tanto cualitativa como cuantitativamente y los resultados revelaron que los factores socioculturales influyen en gran medida en la inversión inmobiliaria. De igual forma, Kakulu (2014) describe una serie de estrategias para el análisis cualitativo y diversos métodos para el análisis de datos en el campo de bienes raíces en Nigeria. Asimismo, el autor argumenta que las futuras investigaciones de bienes raíces en dicho país deben tomar en cuenta las percepciones de las personas puesto que es un área que aún está poco investigada en dicha región. El autor recomienda ampliamente el uso de estrategias de investigación de fenomenología y aplicación de casos de estudio, métodos de recolección de datos de grupos focales y, el uso de una herramienta innovadora de análisis de contenido para el análisis de datos y la interpretación de los hallazgos, de tal manera que remarcan la importancia de este tipo de análisis. A pesar de ello, son muy pocos los estudios que se realizan bajo el contexto inmobiliario y sobre todo que utilicen métodos cualitativos de investigación.

Este trabajo tiene como objetivo identificar la percepción sobre los factores de mayor y menor impacto tomados en cuenta durante el proceso de valuación de una vivienda por parte de los compradores y vendedores de viviendas que residen en el municipio de Cajeme, ubicado en la región sur del estado de Sonora. El enfoque de teoría fundamentada, se basa en recopilar y analizar datos obtenidos mediante entrevistas semi-estructuradas. Una vez que se identificaron diversos patrones, categorías y similitudes en los datos analizados, los resultados demostraron que tanto los compradores como los vendedores de viviendas perciben a la ubicación, las proporciones y las condiciones de las viviendas como factores determinantes para estimar el valor de estas. Lo anterior proporciona beneficios para las personas de la región involucradas en la compra-venta de inmuebles, puesto que permitirá hacer aún más robusto el método de valuación a emplear al considerar el impacto que tiene cada uno de los factores. De igual manera, proporcionará las bases necesarias para establecer los parámetros de un algoritmo predictivo para la estimación del valor de las viviendas de la región.

## FENÓMENO BAJO ESTUDIO

Según datos del Instituto Nacional de Estadística y Geografía (INEGI, 2010), la población total del municipio de Cajeme fue de 409,310 habitantes en 2010, de los cuales más del 72% se concentraban en Ciudad Obregón, siendo ésta la cabecera municipal donde también se encontraba la mayor actividad económica. Así mismo, se calculó una tasa de crecimiento poblacional del 1.72% para el periodo 2011-2015 con base en históricos de 2005-2010. La agricultura es la principal base para el desarrollo económico del municipio, sin embargo, a través del tiempo se han introducido nuevas tecnologías y presentado nuevos productos al mercado (IMIP, 2017). La estructura urbana de la cabecera municipal se caracteriza por concentrar los edificios y monumentos más antiguos en el centro y por mantener gran parte de las viviendas de los habitantes en la periferia, propiciando el crecimiento urbano en la zona norte y sureste de la ciudad (Ayuntamiento de Cajeme, 2011). En cuanto al tema de seguridad pública, Cajeme fue el segundo municipio con mayor número de delitos registrados en 2012, lo cual representó un aumento gradual en la inseguridad con respecto en años anteriores (INEGI, 2012). Respecto al ámbito inmobiliario, en 2015 había un total de 128,490 viviendas particulares en el municipio, de las cuales más del 98% contaban con servicios básicos de energía eléctrica, agua de la red pública y drenaje. Además, el número de habitantes promedio por vivienda fue de 3.4 personas (INEGI, 2015). En el mismo año, se proyectó un notable crecimiento en el mercado inmobiliario del municipio, puesto que doce nuevos fraccionamientos se encontraban en proceso de desarrollo (CANADEVI, 2015).

En los últimos años, los problemas relacionados con el análisis del mercado inmobiliario han crecido muy rápido ya que este conocimiento es muy relevante para las predicciones inmobiliarias, las inversiones y las cuestiones relacionadas con los impuestos (Del Giudice et al. 2017). Debido a ello, la valuación de viviendas es un tema que ha sido constantemente explorado e investigado. Sin embargo, el impacto de los factores clave para determinar el valor de los inmuebles varía dependiendo de los aspectos socioculturales y de la percepción sobre el tema de valuación de viviendas que se tenga en cada región. Dicho impacto se puede ver afectado por el tipo de personas que viven en alguna comunidad, las condiciones de las viviendas de dicha zona, entre otros. Realizar una investigación de carácter cualitativo permite analizar la percepción que tienen los compradores y vendedores de cierta región sobre los factores clave para estimar el valor de las viviendas; a través de este estudio, se pretende identificar cuáles son los más relevantes en el municipio de Cajeme, Sonora. En la Tabla 1 se muestra una descripción de los factores estándar a tomar en cuenta al momento de realizar la estimación del valor de alguna vivienda. Dichos factores serán contemplados durante la realización de este trabajo de investigación.

Tabla 1: Factores para la valuación de viviendas.

<i>Factor</i>	<i>Concepto</i>	<i>Referencias</i>
Ubicación	Es uno de los factores con mayor influencia en el valor de las viviendas. El precio de venta o el costo del metro cuadrado de una vivienda varían dependiendo de la región en donde se encuentre. Esto, debido a la demanda que tiene la zona, las condiciones del vecindario, el grado de desarrollo a su alrededor, entre otros.	Basu et al. (1998); Bourassa et al. (2007).
Tipo de propiedad	El valor de la vivienda se puede ver afectado por el tipo de propiedad, es decir, si el inmueble se trata de una casa o un departamento. Aquí también se ven involucradas la cantidad de habitaciones y tipos de habitaciones con las que cuenta la propiedad.	Daniels et al. (1999); Zietz et al. (2008).
Condiciones de vivienda	En este punto se tocan aspectos como: el año de construcción, condiciones del techo, condiciones del piso, condiciones de las paredes, material exterior, tipo de tejado, reparaciones, vistas e iluminación.	Haurin et al. (1996); Zietz et al. (2008); Bourassa et al. (2007).
Proporciones de la vivienda	Involucra diversos aspectos de la vivienda, tales como: área, perímetro, número de pisos, cochera, estacionamiento y patio.	Daniels et al. (1999); Haurin et al. (1996); Zietz et al. (2008).
Instalaciones	La antigüedad de la vivienda se relaciona directamente con las condiciones de sus instalaciones debido al deterioro que pueden presentar. Los tipos de instalaciones van desde instalaciones eléctricas, de agua potable, red de saneamiento, de ventilación o servicios.	Zietz et al. (2008); Schram (2006).
Sitios cercanos	Los sitios cercanos al inmueble pueden tener influencias positivas o negativas en los precios de las viviendas. Entre los ejemplos de sitios cercanos se incluyen: escuelas, supermercados, hospitales o transporte público.	Basu et al. (1998); Zietz et al. (2008); Bourassa et al. (2007).

## METODOLOGÍA

El propósito de esta sección es describir el método de la teoría fundamentada que se utilizó para llevar a cabo el trabajo de investigación. A continuación, se presenta un listado de los pasos realizados durante la investigación con la finalidad de lograr los objetivos planteados. Posteriormente, se describen las preguntas de investigación, el proceso de recopilación de datos y el análisis de los mismos. Como se mencionó anteriormente, para llevar a cabo esta investigación se utilizó un enfoque de teoría fundamentada. Strauss y Corbin (1994) definen a la teoría fundamentada como una metodología generalizada para desarrollar teoría sobre un fenómeno de estudio en particular, basándose en datos recolectados y analizados sistemáticamente. La teoría evoluciona durante el proceso de investigación, y lo hace a través de la interacción continua entre el análisis y la recopilación de datos. Glaser y Strauss (2017) argumentan que la teoría fundamentada exige identificar categorías teóricas que son derivadas de los datos mediante la utilización de un método comparativo e iterativo. Es decir, se compara el contenido de diversas sesiones de entrevistas o de observación con los conceptos teóricos nacientes del esfuerzo por identificar los temas fundamentales (Wells et al. 1995; Barnes, 1996).

### Proceso de investigación

El proceso de investigación que se llevó a cabo, se basó en una serie de pasos definidos por Glaser y Strauss (2017), tal como se muestra en la Figura 1.



Fig. 1: Metodología de teoría fundamentada como generadora de teoría a partir de los datos.

### Preguntas de investigación

Agee (2009) menciona que muchos investigadores con enfoque en los estudios cualitativos ven una simple pregunta como un punto de partida para su investigación. Una vez que la pregunta se encuentra situada satisfactoriamente en su lugar, el estudio puede comenzar. Posterior a la detección del problema o fenómeno de estudio, se identificaron tres preguntas de investigación. En este caso, las preguntas de investigación buscan explorar todos aquellos factores que toman en cuenta los individuos al momento de estimar el valor de una vivienda, así como todos aquellos que ignoran. Se inicia planteando una pregunta general que se complementa con otras dos cuestiones específicas: 1) ¿Cuál es la percepción del individuo sobre los factores que determinan el valor de una vivienda en la región sur del estado de Sonora?; 2) ¿Cuáles son los factores de mayor y menor impacto durante la estimación de costo de una vivienda por parte de los compradores?; 3) ¿Cuáles son los factores de mayor y menor impacto durante la estimación del precio de venta por parte de los vendedores?

### Participantes

Con el objetivo de determinar un adecuado tamaño muestral para este trabajo de investigación, se empleó el método del muestreo selectivo propio de la investigación cualitativa. Martínez-Salgado (2012) menciona que, a diferencia del muestreo probabilístico, "la indagación cualitativa, en cambio, elige las unidades de estudio con un propósito: lograr un conocimiento intensivo, profundo y detallado de y sobre los casos en los que tiene lugar el fenómeno de interés, generalizable para otras situaciones en las que dicho fenómeno ocurre". Además

del propósito anterior, es importante tomar en cuenta otros factores al momento de delimitar el tamaño de la muestra; tal es el caso de las semanas disponibles para realizar la investigación, ya que el análisis de los datos cualitativos recolectados es bastante riguroso y toma mucho tiempo al ser llevado a cabo de forma manual (Basit, 2003). En este caso, se contaba solamente con 16 semanas para realizar el estudio. Por ello, se optó por establecer un número limitado de sujetos, los suficientes para representar al sector clave de compradores y vendedores de viviendas en la región estudiada.

Se decidió aplicar un total de 10 entrevistas a habitantes del municipio de Cajeme, ubicado en la región sur del estado de Sonora que compartían rasgos y estilos de vida predominantes en la región. De ellos, 5 correspondían a personas que cuentan con experiencia en cuanto a la compra de inmuebles o personas con planes de adquirir una vivienda. Los otros 5 participantes correspondían a personas con experiencia en la venta de inmuebles; dicha experiencia podía haber sido obtenida con apoyo por parte de profesionales de bienes raíces o por la venta directa de algún inmueble. Las entrevistas tuvieron una duración de 30 a 50 minutos, de las cuales se tuvieron que descartar 2 por falta de claridad y consistencia en la información obtenida (E5 y E6). Se identificó que las 8 entrevistas restantes compartían bastante información en común, lo que hacía posible acotar aún más el número de entrevistas a analizar. Finalmente se eligieron 4 para llevar a cabo el análisis cualitativo. La decisión se basó en la calidad de la información proporcionada, en la similitud de las ideas expuestas y en el rango de edad de los participantes, es decir, personas de 30 años o más. Para esto, los participantes de las entrevistas seleccionadas para el análisis fueron E1, E4, E7 y E9. Dos de las entrevistas son de personas con experiencia en la compra de viviendas y las otras dos entrevistas son de personas con experiencia en la venta de viviendas. Específicamente, el sujeto E1 realizó la compra de una vivienda unifamiliar aislada en un vecindario con nivel socioeconómico medio; el sujeto E4 llevó a cabo la venta de una vivienda unifamiliar aislada en un vecindario con nivel socioeconómico bajo; E7 adquirió una vivienda unifamiliar aislada en un vecindario con nivel socioeconómico alto y el sujeto E9 vendió una vivienda unifamiliar adosada en un vecindario con nivel socioeconómico medio. En la Tabla 2 se puede apreciar el perfil general de los participantes que fueron entrevistados.

Tabla 2: Perfil de los participantes.

<i>Participante</i>	<i>Edad</i>	<i>Plan de compra</i>	<i>Experiencia de compra</i>	<i>Presencia durante el proceso de venta</i>	<i>Experiencia de venta</i>	<i>Calidad en la información</i>
E1	44		X			X
E2	36	X				X
E3	25	X				X
E4	60				X	X
E5	38		X			
E6	37				X	
E7	34		X			X
E8	46			X		X
E9	30				X	X
E10	45			X		X

#### *Recolección de datos*

Para llevar a cabo la recolección de los datos se diseñó una guía de entrevista de tipo semiestructurada, tomando como base las preguntas de investigación mencionadas anteriormente. Cohen et al. (2006) menciona que las guías de entrevistas semiestructuradas proporcionan un conjunto claro de instrucciones para los entrevistadores y, puede generar datos cualitativos fiables y comparables. También brindan la oportunidad de identificar nuevas formas de ver y comprender el tema bajo estudio. Por otro lado, Clifford et al. (2016) menciona que las entrevistas semiestructuradas se desarrollan de una manera conversacional, ofreciendo a los participantes la oportunidad de explorar cuestiones que ellos crean relevantes. Este tipo de entrevista semiestructurada tiene como objetivo alentar al sujeto entrevistado para que narre algún evento importante de su vida y del contexto social (Muylaert et al. 2014). La guía o protocolo de entrevista llevada a cabo en este estudio, cubre los siguientes temas: i) Situación del individuo: Preguntas generales sobre antecedentes del participante, incluyendo su situación actual personal y de la vivienda, así como los conocimientos base que posee sobre la valuación de viviendas; ii) Experiencias con la compra o venta de viviendas: El propósito de estas preguntas es conocer cómo fue el proceso de compra o venta de la vivienda en cuestión; iii) Percepción de factores clave: Se realizan preguntas para conocer la percepción que tiene el participante sobre los diversos factores involucrados en la estimación del valor de las viviendas.

### Análisis de los datos

Una vez concluidas las entrevistas, se realizó su transcripción para proceder con el análisis de línea por línea. El objetivo de llevar a cabo el análisis de las transcripciones es profundizar aún más en las historias compartidas por los participantes, en busca de información relevante que responda a las preguntas de investigación planteadas. Por lo tanto, es necesario analizar e interpretar los datos para construir un sentido e ir generando la teoría a partir de estos. El proceso de análisis se puede ver reflejado en la Figura 2, en donde la codificación abierta se refiere a descomponer la información línea por línea, la codificación axial sugiere la creación de categorías, en la codificación selectiva se elige una categoría base y, por último, se procede con la generación de una teoría.

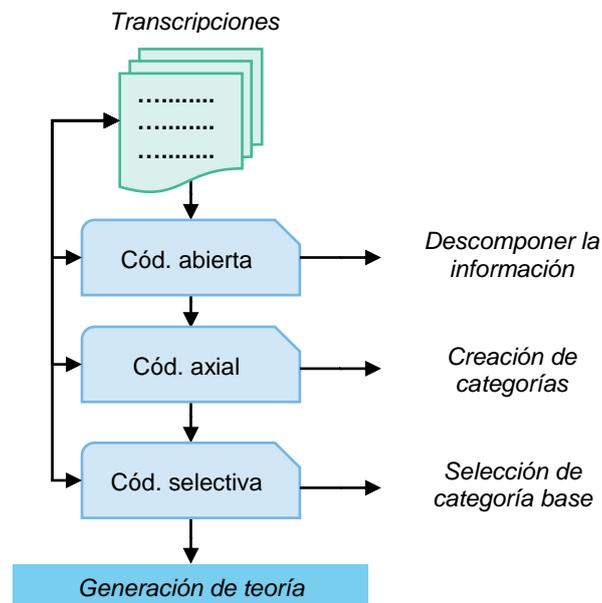


Fig. 2: Tipos de codificación para el análisis de datos.

#### Codificación abierta

Codificar los datos significa identificar todas y cada una de las ideas representadas en cada línea de texto de las transcripciones. Para esto, se debe dar un nombre clave a cada nueva idea, preferentemente que sean palabras claras y concisas (Grams, 2001). La codificación abierta incluye conceptos de etiquetado, definiendo y desarrollando categorías basadas en sus propiedades y dimensiones. Durante este proceso, los conceptos surgen de los datos brutos y luego se agrupan en categorías conceptuales. El objetivo es construir un marco preliminar descriptivo y multidimensional para su análisis posterior. El proceso por sí solo asegura la validez del trabajo debido a que se construye directamente a partir de los datos brutos (Khandkar, 2009). Durante esta fase, se fraccionaron los datos de las transcripciones para obtener diversos códigos que después se agruparían por categorías. En la Figura 3 se puede observar un fragmento de la transcripción de una entrevista y su respectiva codificación abierta, las palabras en rojo corresponden a las etiquetas resultantes según se analizaba el texto línea por línea.

#### Codificación axial

Strauss y Corbin (2002) definen la codificación axial como “el proceso de relacionar las categorías a sus subcategorías, denominado ‘axial’ porque la codificación ocurre alrededor del eje de una categoría, y enlaza las categorías en cuanto a sus propiedades y dimensiones”. Es por ello que el propósito de la codificación axial consiste en iniciar con la reagrupación de los datos que se fracturaron durante la codificación abierta. En la codificación axial, las categorías se relacionan con sus subcategorías para formar unas explicaciones más precisas y completas sobre los fenómenos (Strauss et al. 1998). En esta fase, se analizaron los códigos resultantes de la codificación abierta y se agruparon por categorías, subcategorías, propiedades y dimensiones, lo cual resultó en una mejor comprensión del escenario bajo estudio. En la Tabla 3 se muestran algunos datos resultantes del proceso de codificación axial; se encontraron tres categorías principales, doce subcategorías, sesenta propiedades y ciento veinte dimensiones.

Fig. 3: Fragmento de codificación abierta realizada a transcripción de E4.

6	L	¿Y validó de alguna manera el precio de venta?
7	E4	Sí lo validé, por el tipo de casa que tenía, más o menos le pregunté a alguien que sabía del
8		tema [Necesidad de asesoramiento] y ahí más o menos me orientó [Asesoramiento
9		valuación vivienda], me dijo “pues por lo que es, por la zona, en dónde está y la construcción
10		pues vale más o menos tanto, [Factores determinantes] puedes variar un poco más o poco
11		menos ese precio, pero eso es aproximadamente”. [Estimación de precio de venta]
12		Entonces yo tomé como referencia ese precio que me dio esa persona que sabe sobre
13		valuación de viviendas y modifiqué un poco el precio [Ajuste de precio de venta] ya viendo
14		y analizando bien la casa [Verificar condiciones de vivienda] y las ofertas de los compradores
15		que tenía en ese momento. [Analizar ofertas]

Las categorías principales se definen a continuación: i) Contexto actual del individuo: Concentra toda aquella información en torno a la situación actual del individuo bajo estudio. Es decir, involucra aspectos de su vida personal, características esenciales de su vivienda (ej. si es propia o rentada), así como su nivel de conocimientos sobre el tema de valuación de viviendas; ii) Experiencias del individuo: Se refiere a todos los aspectos relacionados con la experiencia, ya sea en la compra de viviendas o con la venta de estas. Así mismo, involucra propiedades como el tipo de experiencia, las ventajas, los problemas ocurridos, así como las habilidades y conocimientos puestos en marcha durante el proceso de compra o de venta; iii) Percepción de factores estimación: Aquí se concentra información importante sobre la percepción del individuo en cuanto a los diversos factores involucrados durante la estimación del valor de una vivienda, ya sea para su compra o venta, tales factores son: ubicación, sitios cercanos, vecinos, proporciones de la vivienda, condiciones de la vivienda, instalaciones, seguridad, terreno excedente de la vivienda. Así mismo, se determina el grado de importancia (alto o bajo) que percibe el individuo bajo estudio para cada uno de dichos factores, las ventajas y desventajas percibidas y demás información relevante que permite profundizar en la percepción que tienen de los mismos.

Tabla 3: Fragmento de codificación axial realizada durante análisis de datos.

Categorías	Subcategorías	Propiedades	Dimensiones
I. Contexto actual del individuo		Vida personal Características de la vivienda Habilidades y conocimientos sobre valuación	Edad, Estado civil, ... Rentada, Compartida, propia, remodelada, ... Alto, medio, bajo
II. Experiencias del individuo	Compra de viviendas	Tipo de experiencias Ventajas Problemas Habilidades y conocimientos Tipo plan de compra	Satisfactoria, insatisfactoria Asesoría, claridad en procesos de compra, facilidades de pago, ... Estimación de costos, ... Largo o corto plazo
	Venta de viviendas	Tipo de experiencia Ventajas Problemas Habilidades y conocimientos	Positiva o Negativa Asesoría, Uso de Tecnología, ... Periodos de espera, ... Intereses de compradores, ...
III. Percepción de factores estimación	Factores determinantes para compra	Nivel de impacto	Alto, Medio, Bajo
	Factores determinantes para venta	Nivel de impacto	Alto, Medio, Bajo

Tabla 3 (continuación)

	Ubicación	Grado de importancia Ventajas Desventajas Relación con plusvalía Relación con otros factores Relación con valor de vivienda	Alto, Bajo Buen mantenimiento, ... Precios elevados por ubicación, ... Aumento/Decremento Vecinos, calidad de vida, ... Alto, Medio, Bajo
	Sitios cercanos	Grado de importancia Grado de satisfacción Ventajas Percepción sobre sitios cercanos problemáticos Problemas Tipos de sitios cercanos Relación con valor de vivienda	Alto, Bajo Alto, Bajo Ahorro de gasolina, comodidad, ... Centros comerciales, deportivos y carreteras Tráfico, ruido, ... Comerciales, escolares, ... Alto, Medio, Bajo
	Vecinos	Grado de importancia Tipos de vecinos Conocimiento de comunidad Percepción de vecindario Confianza Tipo de relación con vecinos Relación con valor de vivienda Relación con otros factores	Alto, Bajo Agradable, tranquilos, ... Alto, nulo Tranquilo, ruidoso Alta, Nula Buena, mala Alto, Medio, Bajo Ubicación
	Proporciones de la vivienda	Grado de importancia Tipo de proporciones Problemas Percepción de utilidad Relación con valor de la vivienda	Alto, Bajo Tamaño, metros cuadrados, ... Necesidad de adaptación, ... Alto, Medio, Bajo Alto, Medio, Bajo
	Condiciones de la vivienda	Grado de importancia Nivel de deterioro Grado de satisfacción Descripción de construcción Condiciones de construcción Ventajas Desventajas Relación con valor de vivienda	Alto, Bajo Alto, Medio, Bajo Alto, Bajo Tipo de construcción, ... Buenas, en remodelación, ... Buena regulación de temperatura Reparación de techos, ... Alto, Medio, Bajo
	Instalaciones	Grado de importancia Tipos de instalaciones Ventajas Problemas Relación con valor de vivienda	Alto, Bajo Eléctricas, agua potable, ... Facilidad de reemplazo, ... Requiere conocimientos expertos Alto, Medio, Bajo
	Seguridad	Grado de importancia Nivel de seguridad Frecuencia de vandalismo Relación con valor de vivienda	Alto, Bajo Alto, Medio, Bajo Abundante, Nulo Alto, Medio, Bajo
	Terreno excedente	Grado de importancia Ventajas Desventajas	Alto, Bajo Proyecciones a futuro, ... Costos adicionales ...

## RESULTADOS

Una vez concluido el análisis mediante la codificación abierta y la codificación axial, se comenzó el proceso de integrar y refinar la teoría, también llamado codificación selectiva. Strauss y Corbin (1998) argumentan que, al igual que en todas las fases del análisis, la integración de la teoría requiere de la interacción constante entre el analista y los datos. Esta fase inicia con la selección de una categoría medular que representa el tema principal de la investigación. En este caso las preguntas de investigación no cambiaron conforme avanzaba el análisis, esto debido a que los datos continuaron relacionados con las mismas preguntas. Se eligió la tercera categoría como núcleo central para darle respuesta a las preguntas de investigación planteadas. Es decir, la categoría sobre la "percepción de los factores de estimación".

En la Figura 4 se puede observar la evolución del proceso de análisis de datos mediante las fases de codificación abierta, axial y selectiva. Se puede apreciar que los códigos resultantes se clasificaron en tres categorías principales (Contexto actual del individuo, Experiencias del individuo, Percepción de los factores) que interactuaban con algunas subcategorías (Factores determinantes en la compra/venta de viviendas, Experiencias con la compra/venta de viviendas), propiedades y dimensiones (Grado de importancia, Tipo de experiencia, Habilidades/Conocimientos, Contexto de vivienda actual, Contexto personal). Con la elección de

la categoría central, se reordenaron las relaciones entre categorías, subcategorías, propiedades y dimensiones, quedando la categoría sobre la percepción de los factores de estimación al centro y manteniendo una relación con las experiencias de los individuos, el contexto actual del individuo y el listado de los factores determinantes para la compra/venta de viviendas.

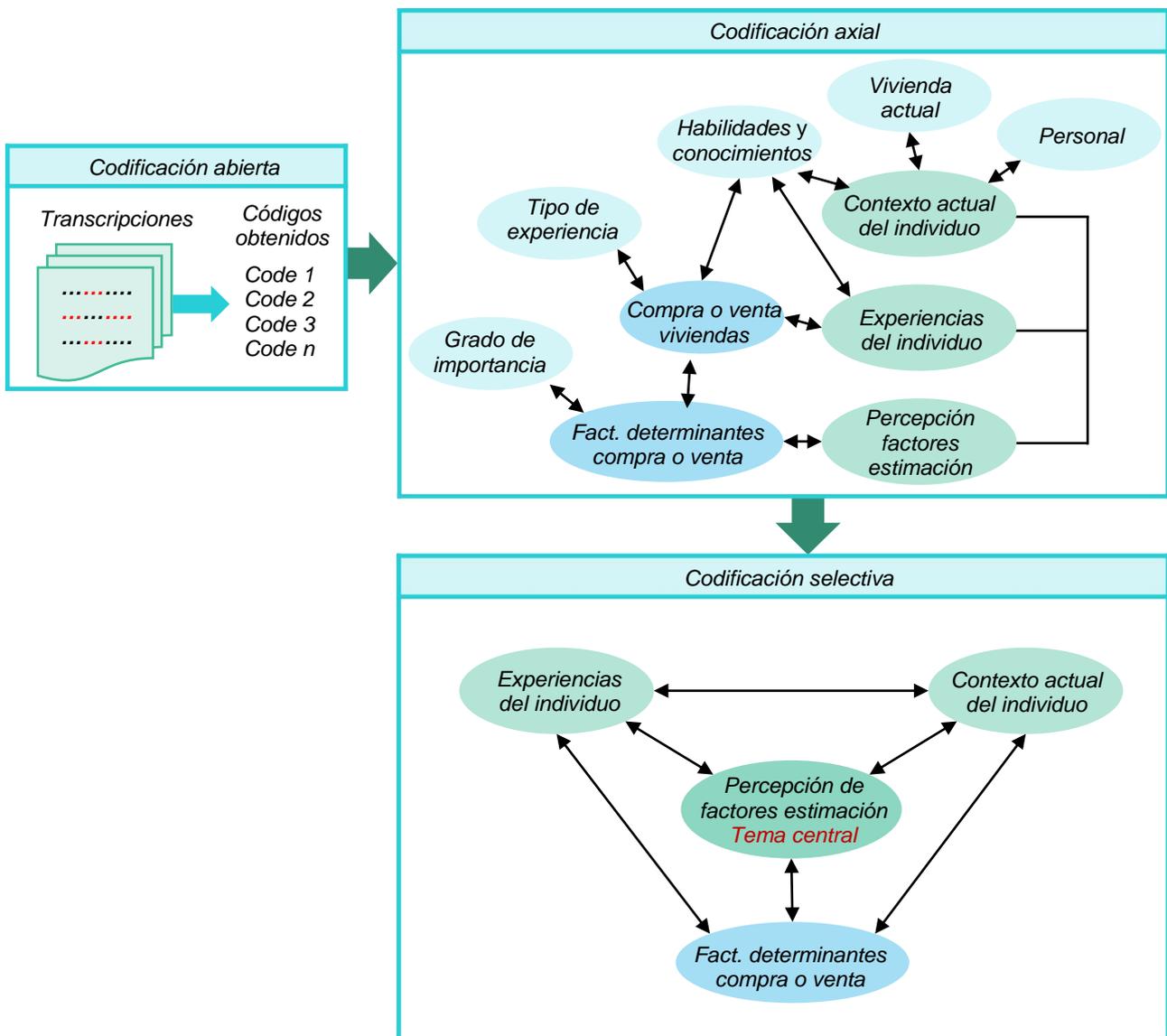


Fig. 4: Proceso de análisis de los datos: Codificación abierta, axial y selectiva.

*Percepción general sobre los factores que determinan el valor de una vivienda*

Tal como se menciona anteriormente, el análisis de las entrevistas arrojó a la categoría sobre la “percepción de los factores de estimación” como la central. Dicha categoría, en relación con las otras dos categorías identificadas, permite responder a las preguntas de investigación planteadas. En la Figura 5 se puede observar que los compradores y vendedores de viviendas de la región perciben tres factores como determinantes clave del valor de las viviendas, es decir: la ubicación, las proporciones y las condiciones de la vivienda. Todos los individuos entrevistados coincidieron en que dichos factores son los más importantes al momento de determinar, ya sea el costo o el precio de venta del inmueble. Por ejemplo, E7 (comprador) hizo énfasis en la relación directa que existe entre la ubicación de una vivienda y su costo: “si es en una zona de las afueras de la ciudad pues está más barata, si es céntrica ahí aumenta más el costo y si es en un residencial pues aún más cara todavía” (E7). Por otro lado, E9 (vendedor) menciona que, para los vendedores, la ubicación “es importante por la plusvalía que le puedes dar al inmueble” (E9). En cuanto a las proporciones de la vivienda, E4 (vendedor) comenta: “Considero muy importante las proporciones. Si tienen mucho que ver para sacar el precio de la vivienda, por los metros cuadrados. Tienes que ver el factor ese, es muy importante, el tamaño del terreno. Es primordial” (E4). Así mismo, E1 (comprador) afirma que las condiciones de la vivienda “Sí, tienen mucho que ver para la valuación de una vivienda, el tipo de material y construcción que

tenga es uno de los factores más importantes en los que todas las personas y evaluadores se fijan o toman en cuenta” (E1). Sin embargo, las opiniones de los participantes difieren en un par de factores, los compradores consideran determinantes para el costo de la vivienda a la seguridad y el terreno excedente. E1 (comprador) asegura que “por el riesgo que hay en el vecindario debería valer menos la casa” (E1). Es decir, si el nivel de seguridad en la zona es bajo, el costo de la vivienda debería disminuir por dicho factor. Por otro lado, E7 (comprador) comenta porqué piensa que el terreno excedente determina el costo: “La casa que elegí no se me hizo cara, tuve que dar un extra de dinero pero fue porque yo compré terreno excedente” (E7).



Fig. 5: Diagrama de Venn sobre la percepción de compradores y vendedores respecto a los factores que determinan el valor de una vivienda.

Por otro lado, los vendedores ni siquiera mencionaron esos factores, ellos hicieron énfasis en los sitios cercanos con los que cuentan las viviendas para determinar el precio de venta. En ese sentido, E4 (vendedor) justifica el precio de venta elevado de la siguiente manera: “Allá en dónde está la casa que vendí, el camión pasa por la casa, el seguro está cerca, la laguna, la escuela, el ITSON, las escuelas primarias o secundarias están ahí cerquitas (sic), tiene muchas cosas buenas que pueden influir en la vida de los compradores que vivan ahí, por el fácil acceso al transporte público y sitios cercanos que ya mencioné” (E4). Las opiniones entre compradores y vendedores difieren debido al tipo de experiencias y percepciones con los que cuentan. En ese sentido, la categoría sobre la percepción de los factores se ve relacionada con el contexto actual del individuo y sus experiencias. Los compradores consideran importante a la seguridad y al terreno excedente por sus intereses, es decir, buscan vivir en un sitio libre de delincuencia y en el que exista la posibilidad de expandir la construcción en un futuro, el siguiente fragmento de la entrevista a E1 (comprador) muestra su percepción sobre la importancia de la seguridad y el terreno excedente según sus intereses y experiencias: “Hubo un factor muy importante que no contemplé que fue la seguridad que había en la zona y al final terminé arrepintiéndome de dicha compra de la vivienda, por eso la regresé, por la poca seguridad que había ahí. ...El terreno excedente es uno de los factores más importantes al momento de determinar el valor o el costo de la vivienda... antes de comprarla se debe ver que sea lo suficientemente amplio como para expandirse o ampliarse de alguna manera la casa, es muy importante ver más allá, a largo plazo, por si crece la familia en algún momento tener terreno suficiente para poner uno o más cuartos” (E1). La percepción sobre las instalaciones (eléctricas, agua potable, saneamiento, servicios de telecomunicaciones) y los vecinos con los que cuenta la vivienda fueron bastante dispersas y dependen de otros subfactores más específicos dentro de cada tipo de participante, es por ello que no se incluyen en la Figura 5. Sin embargo, se explica la perspectiva de cada uno de esos factores como respuesta a las dos últimas preguntas de investigación.

#### Factores de mayor y menor impacto por parte de los compradores

Los compradores de la región sur del estado de Sonora que fueron entrevistados consideran como factores de mayor impacto a la ubicación, las proporciones, las condiciones, la seguridad y el terreno excedente. En la sección anterior se expusieron los motivos por los cuales dichos factores son catalogados como factores de mayor impacto o determinantes del valor de viviendas. Ahora bien, los factores considerados como factores de menor impacto son: Los sitios cercanos comerciales y de esparcimiento y las instalaciones de la vivienda (Ver Figura 6). E7 (comprador) comenta que no percibe relación entre los sitios cercanos y el costo de las viviendas: “Pues yo digo que con el costo no influyen, aunque ahora hace mucho auge eso de los sitios cercanos al fraccionamiento, ahora que hicieron la plaza comercial ahí cerca... No creo que influya en el costo pero en las personas sí, por ejemplo, dicen: ‘está cerca de la Plaza Sendero ahí voy y compro’...” (E7).

Las instalaciones de la vivienda son consideradas como un factor flexible que no influye realmente en el costo de la vivienda por parte de los compradores: "...No creo realmente que haya una relación entre las condiciones de las instalaciones y el costo de la misma. Es opcional al estimar el costo porque esas son cosas que se pueden reparar después... es algo que se puede reemplazar fácilmente. Claro que implica un costo pero es algo que se mira después de realizada la compra y se va arreglando poco a poco, conforme se tenga la posibilidad" (E1).

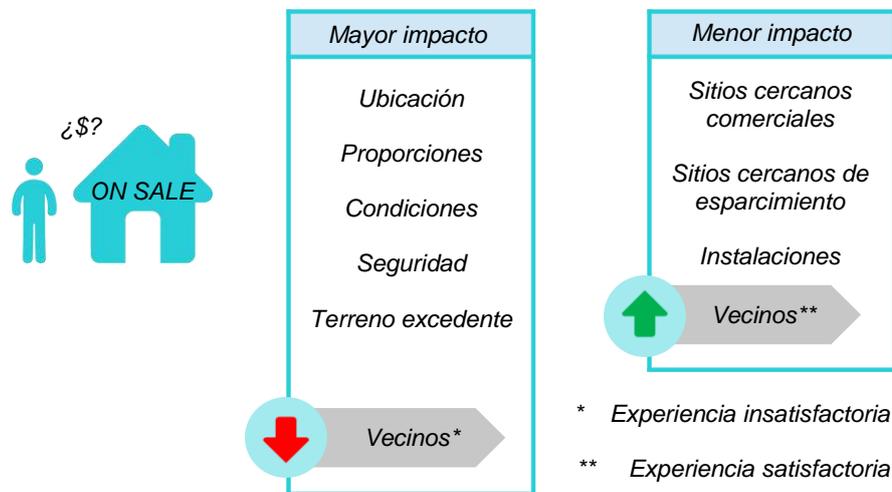


Fig. 6: Factores de mayor y menor impacto percibidos por compradores de viviendas.

La percepción y el nivel de impacto de los vecinos o el tipo de vecinos con los que cuenta una vivienda varía. Esto depende del tipo de experiencia previa con la que cuente el comprador en cuestión. Es decir, si el comprador ha tenido una experiencia insatisfactoria con los vecinos, entonces va a catalogar al "factor vecinos" como de mayor impacto para estimar el valor de alguna vivienda. E1 comenta que su experiencia con los vecinos ha sido insatisfactoria: "Mi experiencia no fue buena puesto que... me dejé llevar por la calle y no me fijé en los vecinos que tenía dicha vivienda o en el vecindario... tiempo después de que compré dicha casa...ya estaba pensando en mudarme ahí pero me di cuenta de que el barrio era realmente peligroso y los vecinos se veían peligrosos, me dio mucho miedo porque yo tengo dos hijas y preferí regresar esa casa y perdí la inversión que le había hecho con el piso y todo eso" (E1).

También menciona cómo eso ha influido en su percepción sobre la relación de los vecinos y el costo de las viviendas: "Hay un tipo de vecinos que son problemáticos... Por el riesgo que hay en el vecindario debería valer menos la casa, por eso te menciono que percibo esa relación entre esos dos puntos que son el tipo de vecinos y el costo de la casa" (E1). Por el contrario, si el comprador ha tenido una experiencia satisfactoria o neutral con los vecinos, entonces va a catalogar al "factor vecinos" como de menor impacto o como factor flexible para estimar el valor de alguna vivienda. E7 comenta que su experiencia con los vecinos ha sido satisfactoria: "Mi experiencia ha sido buena. Pues ahorita se ve que son vecinos tranquilos y que están al pendiente de las demás casas..." (E7). Debido a ello, no toma en cuenta o no cree que exista una relación entre los vecinos con los que cuenta una vivienda y el costo de la misma: "Pues no creo que los vecinos tengan algo que ver como tal con el costo de una vivienda... más bien segmentan por zonas de distintos niveles económicos en el fraccionamiento, haz de cuenta que por eso hay varias casas, hay unas más baratas y otras más caras, obviamente el tipo de persona no es la misma" (E7).

#### Factores de mayor y menor impacto por parte de los vendedores

Los vendedores de viviendas de la región sur del estado de Sonora que fueron entrevistados no mencionan la seguridad ni terreno excedente como posibles factores de mayor o de menor impacto, más bien se basan en los factores clave o estándar para determinar el valor de las viviendas. Consideran como factores de mayor impacto a la ubicación, las proporciones, las condiciones y los sitios cercanos con los que cuenta la vivienda. En la sección anterior se expusieron los motivos por los cuales dichos factores son catalogados como factores de mayor impacto o determinantes del valor de viviendas. Ahora bien, el factor considerado como factor de menor impacto fue el de vecinos de la vivienda (Ver Figura 7). E9 comenta que el tipo de vecinos no influye en el precio de venta del inmueble: "Yo creo que sería más importante sobre la zona como ya mencioné. Sobre los vecinos, no tanto. Porque si te vas a una zona de un nivel estable donde no haya mucho vandalismo ni delincuencia, creo yo que los vecinos importan muy poco... Ya es a criterio de cada quién, pero creo que no influye para estimar el precio de venta del inmueble." (E9).

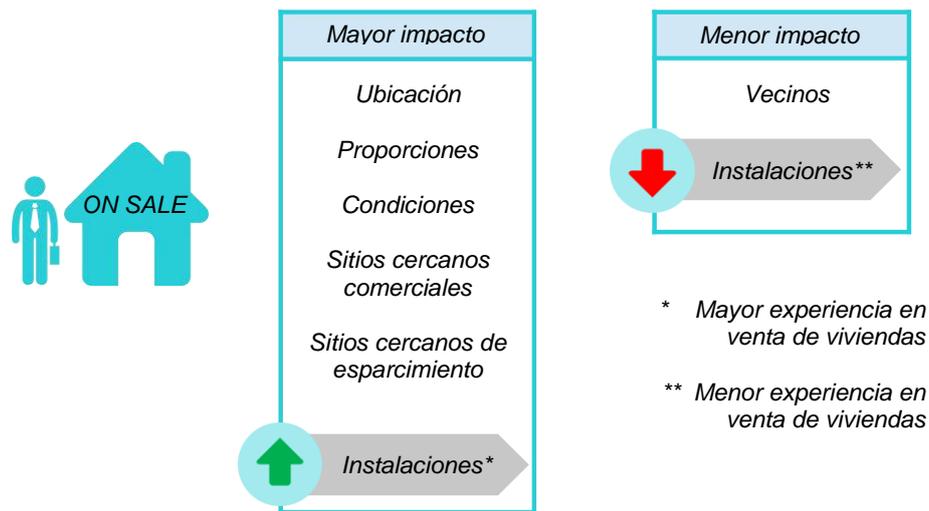


Fig. 7: Factores de mayor y menor impacto percibidos por vendedores de viviendas.

La percepción y el nivel de impacto en cuanto a las instalaciones (eléctricas, agua potable, saneamiento, servicios de telecomunicaciones) de la vivienda varía dependiendo del nivel de experiencia en la venta de viviendas del vendedor. Es decir, si el vendedor posee una mayor experiencia en cuanto a la venta de viviendas, entonces va a catalogar al “factor instalaciones” como de mayor impacto para estimar el valor de alguna vivienda. El participante E9 quién posee una mayor experiencia en cuanto a la compra de viviendas menciona lo siguiente: *“Es muy importante... las instalaciones son algo que se tiene que considerar. También, si tiene algunos servicios, por ejemplo, unos amigos acaban de comprar una casa y, ni Telcel ni Megacable llega para allá.... en los mismos valores catastrales te viene todo eso, si cuenta con todos los servicios e instalaciones o cuenta con algunos servicios, si tienes qué hacer fosa o no...”* (E9).

Por el contrario, si el vendedor posee una menor experiencia en cuanto a la venta de viviendas, entonces va a catalogar al “factor instalaciones” como de menor impacto para estimar el valor de alguna vivienda. E4 posee conocimientos básicos sobre el tema y comenta lo siguiente al respecto: *“...Sería secundario eso de las instalaciones de las tuberías o eléctricas, no lo considero tan importante porque hay otras cosas más importantes para estimar el precio de venta de la casa... porque eso es punto y aparte, se puede arreglar o mejorar poco a poco por parte de los compradores... es de los factores flexibles que mencionaba en la pregunta anterior...”* (E4).

## DISCUSIÓN

Durante el proceso de la revisión de literatura, se encontraron numerosas investigaciones cualitativas en donde los autores emplean diversos métodos de análisis de datos cualitativos, tales como la teoría fundamentada o el análisis de contenidos, para conocer la percepción de las personas respecto a temas de interés en diversas áreas como la medicina (Figueira et al., 2009; Bedoya et al., 2017), psicología (Festen et al., 2014), tecnología (Riley et al., 2009; Bravo et al., 2018), gestión empresarial (Oliveira et al., 2016; Padilla et al., 2017; Pérez-Benedito et al., 2017), estudio de mercados (Kamagahara et al., 2016), entre otros. Sin embargo, se identificó cierta escasez respecto a este tipo de estudios cualitativos enfocados al contexto de bienes raíces. Kibet et al. (2016) realizó un estudio híbrido con rasgos cualitativos y cuantitativos para demostrar que los factores socioculturales influyen en la inversión inmobiliaria. Por otro lado, Kakulu (2014) propone una serie de estrategias para llevar a cabo un análisis cualitativo en el contexto inmobiliario. Si bien ambos trabajos involucran el tema de las investigaciones cualitativas y de bienes raíces, ninguno de ellos lleva a cabo un estudio para conocer la percepción de las personas en cuanto a la compra y venta de viviendas. En contraste, haciendo uso del método de la teoría fundamentada, en el presente trabajo se analizaron las percepciones de compradores y vendedores de viviendas para conocer cuáles factores son considerados de mayor y menor impacto durante el proceso de valuación de inmuebles.

El análisis obtenido de este estudio indica que, tanto compradores como vendedores de viviendas en la región sur del estado de Sonora consideran a la ubicación, las proporciones y las condiciones de las viviendas como factores fundamentales al momento de llevar a cabo la estimación del precio de alguna vivienda. Más específicamente, los compradores perciben a la ubicación, las proporciones de la vivienda, las condiciones de la vivienda, la seguridad en la zona y al terreno excedente como factores de mayor impacto; los sitios cercanos a la vivienda y el tipo de instalaciones con las que cuenta la vivienda son considerados factores de

menor impacto. Por otro lado, los vendedores perciben a la ubicación, las proporciones, las condiciones y los sitios cercanos de la vivienda como factores de mayor impacto y consideran irrelevante o de menor impacto el tipo de vecinos. De igual modo, durante la fase de análisis de los datos se encontró que existen más dimensiones del fenómeno de estudio que pueden ser más profundamente analizadas en trabajos futuros.

Los resultados de este estudio beneficiarán a las personas de la región sur del estado de Sonora relacionadas con el contexto de bienes raíces; robusteciendo el método de valuación de viviendas al conocer la influencia percibida para cada uno de los factores involucrados en el proceso de estimación de precios. También proporcionará las bases necesarias para establecer las variables o parámetros a contemplar en un algoritmo predictivo para la estimación del precio de las viviendas en la región. Así mismo, este trabajo puede servir de guía para futuros estudios cualitativos relacionados con bienes raíces que deseen replicar el método utilizado y analizar datos obtenidos de muestras más grandes.

## CONCLUSIONES

De acuerdo con los resultados obtenidos, se concluye lo siguiente: 1) Las preguntas de investigación planteadas al inicio de este trabajo de investigación fueron resueltas satisfactoriamente mediante la aplicación del método de la teoría fundamentada; 2) además, se encontró que los factores determinantes del valor de una vivienda desde la perspectiva de compradores y vendedores de viviendas son principalmente: la ubicación, las proporciones y las condiciones de la vivienda; y 3) el impacto del resto de los factores analizados varía dependiendo de distintas situaciones, tal como el perfil del participante (vendedor o comprador), el nivel de experiencia (mayor o menor) en cuanto a la venta y valuación de viviendas por parte de los vendedores e incluso el tipo de experiencia con los vecinos (satisfactorio e insatisfactorio) por parte de los compradores de inmuebles.

## REFERENCIAS

- Agee, J., Developing Qualitative Research Questions, A Reflective Process, doi: 10.1080/09518390902736512, International Journal of Qualitative Studies in Education, 22(4), 431-447 (2009)
- Arias, C., Enfoques Teóricos sobre la Percepción que Tienen las Personas, ISSN: 2500-705X, Horizontes Pedagógicos, 8(1), 9-22 (2006)
- Ayuntamiento de Cajeme, Programa Parcial de Crecimiento Urbano Zona Sureste y Corredor Aeropuerto (2011)
- Barnes, D.M., An Analysis of the Grounded Theory Method and the Concept of Culture, doi: 10.1177/104973239600600309, Qualitative Health Research, 6(3), 429-441 (1996)
- Basit, T., Manual or electronic? The role of coding in qualitative data analysis, doi: 10.1080/0013188032000133548, Educational Research, 45(2), 143-154 (2003)
- Basu, S. y T. G. Thibodeau, Analysis of Spatial Autocorrelation in House Prices, doi: 10.1023/a:1007703229507, The Journal of Real Estate Finance and Economics, 17(1), 61-85 (1998)
- Bedoya, E. A., D. D. Sierra, C. A. Severiche y M. D. J. Meza, Diagnóstico de Bioseguridad en el Sector Sanitario del Departamento de Bolívar, Norte de Colombia, doi: 10.4067/S0718-07642017000500021, Información Tecnológica, 28(5), 225-232 (2017)
- Bell, R. y M. P. Bell, Real Estate Research Methods, Appraisal Journal, 83(4), 310-318 (2015)
- Bourassa, S.C., E. Cantoni y M. Hoesli, Spatial Dependence, Housing Submarkets, and House Price Prediction, doi: 10.1007/s11146-007-9036-8, The Journal of Real Estate Finance and Economics, 35(2), 143-160 (2007)
- Bravo, C. J., P. E. Ramírez y J. Arenas, Aceptación del Reconocimiento Facial Como Medida de Vigilancia y Seguridad: Un Estudio Empírico en Chile, doi: 10.4067/S0718-07642018000200115, Información Tecnológica, 29(2), 115-122 (2018)
- CANADEVI, Cámara Nacional de la Industria de Desarrollo y Promoción de Vivienda, Estadísticas y Publicaciones: Cajeme, (2015)
- Carterette, E.C. y M.P. Friedman, Manual de Percepción: Raíces Históricas y Filosóficas, ISBN: 9682411483-9789682411489, (1982)
- Clifford, N., M. Cope, T. Gillespie y S. French, Key Methods in Geography, ISBN: 978-1-4462-9858-9, Sage (2016)
- Cohen, D. y B. Crabtree, Qualitative Research Guidelines Project, Robert Wood Johnson Foundation (2006)
- Daniels, H. y B. Kamp, Application of MLP Networks to Bond Rating and House Pricing, doi: 10.1007/s005210050025, Neural Computing & Applications, 8(3), 226-234 (1999)
- Del Giudice, V., P. De Paola y F. Forte, Using Genetic Algorithms for Real Estate Appraisals, doi: 10.3390/buildings7020031, Buildings, 7(2), 31-40 (2017)
- Esquer, J.A., Estimación del Precio de Venta de la Vivienda en la Ciudad de Hermosillo Aplicando el Método de Precios Hedónicos, (U. d. Sonora, Editor), Tesis de grado, Posgrado en Ciencias de la Ingeniería, Ingeniería Civil, UNISON (2012)

- Festen, H., K. Schipper y otros cuatro autores, Parents' Perceptions on Offspring Risk and Prevention of Anxiety and Depression: A Qualitative Study, doi: 10.1186/2050-7283-2-17, BMC Psychology, 2(1), 1-17 (2014)
- Figueira, T.R., E. F. Ferreira, V. T. Schall y C. M. Modena, Women's Perceptions and Practices Regarding Prevention and Health Promotion in Primary Healthcare, doi: 10.1590/s0034-89102009005000081, Revista de Saude Publica, 43(6), 937-943 (2009)
- Glaser, B. y A. Strauss, Discovery of Grounded Theory: Strategies for Qualitative Research, ISBN: 9781351522168, Routledge (2017)
- Grams, G. D., Rudiments in the Use of 'Grounded Theory': A Working Guide, Unpublished Manuscript, University of British Columbia (2001)
- Haurin, D.R. y D. Brasington, School Quality and Real House Prices: Inter-and Intrametropolitan Effects, doi: 10.1006/jhec.1996.0018, Journal of Housing economics, 5(4), 351-368 (1996)
- IMIP, Instituto Municipal de Investigación y Planeación Urbana de Cajeme, Portal de transparencia, Antecedentes (2017)
- INEGI, Instituto Nacional de Estadística y Geografía, Censo de población y vivienda 2010, Banco de indicadores, Sonora, Cajeme (2010)
- INEGI, Instituto Nacional de Estadística y Geografía, Censo de población y vivienda 2015, Banco de indicadores, Sonora, Cajeme (2015)
- INEGI, Instituto Nacional de Estadística y Geografía, Censo Nacional de Gobiernos Municipales y Delegacionales, Seguridad Pública y Justicia, Sonora, Cajeme (2012)
- Kakulu, I. I., Qualitative Research Strategies and Data Analysis Methods in Real Estate Research - An innovative approach using the BB Model, Estate Management Department Workshop, At Federal Polytechnic, Nekede, Owerri (2014)
- Kamagahara, Y., T. Takeda y otros cinco autores, Qualitative Analysis of the Customer Satisfaction at the Dental Clinics, doi.org/10.1007/978-3-319-40247-5\_24, In International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management, Springer, Cham, 229-242 (2016)
- Khandkar, S.H., Open coding, University of Calgary, 23-31 (2009)
- Kibet, F., E. Nambuswa y G. S. Namusonge, Effect of Social-Cultural Factors on Real Estate Investment: A Survey of Kisumu City, European Journal of Business and Management, 8(29), 85-91 (2016)
- Martínez-Salgado, C., El muestreo en investigación cualitativa: principios básicos y algunas controversias, doi: 10.1590/S1413-81232012000300006, Ciência & Saúde Coletiva, 17, 613-619 (2012)
- Muylaert, C. J., V. Sarubbi y otros tres autores, Narrative Interviews: An Important Resource in Qualitative Research, doi: 10.1590/s0080-623420140000800027, Revista da Escola de Enfermagem da USP, 48(SPE2), 184-189 (2014)
- Oliveira, E., T. Conte, M. Cristo y E. Mendes, Software Project Managers' Perceptions of Productivity Factors: Findings from a Qualitative Study, doi:10.1145/2961111.2962626, In Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ACM (2016)
- Padilla, C.P., D. X. Arévalo, M.A. Bustamante y C.L. Vidal, Responsabilidad Social Empresarial y Desempeño Financiero en la Industria del Plástico en Ecuador, doi: 10.4067/S0718-07642017000400012, Información Tecnológica, 28(4), 93-102 (2017)
- Pérez-Benedito, M.A., L. Porcuna-Enguix y R. Porcuna-Enguix, Los Mapas Contables de Gestión de las empresas cotizadas chilenas: Análisis Cualitativo, doi: 10.4067/S0718-07642017000100016, Información Tecnológica, 28(1), 161-170 (2017)
- Riley, C., K. Buckner, G. Johnson y D. Benyon, Culture & Biometrics: Regional Differences in the Perception of Biometric Authentication Technologies, doi: 10.1007/s00146-009-0218-1, AI & society, 24(3), 295-306 (2009)
- Schram, J.F., Real Estate Appraisal, ISBN: 1887051252-9781887051255, Rockwell Publishing (2006)
- Strauss, A. y J. Corbin, Bases de la Investigación Cualitativa: Técnicas y Procedimientos para Desarrollar la Teoría Fundamentada, Medellín, Universidad de Antioquia, (2002)
- Strauss, A. y J. Corbin, Basics of Qualitative Research: Procedures and Techniques for Developing Grounded Theory, Sage (1998)
- Strauss, A. y J. Corbin, Grounded Theory Methodology, Handbook of Qualitative Research, 17, 273-285 (1994)
- Wells, K., The strategy of Grounded Theory: Possibilities and Problems, doi:10.1093/swr/19.1.33, Social Work Research 19(1), 33-37 (1995)
- Zietz, J., E.N. Zietz y G. S. Sirmans, Determinants of House Prices: A Quantile Regression Approach, doi: 10.1007/s11146-007-9053-7, The Journal of Real Estate Finance and Economics, 37(4), 317-333 (2008)

# Eventos de Vida y su Relación con el Padecimiento de Cáncer de Mama: Un estudio Exploratorio

Roberto Aguilar Arredondo, Luis A. Castro, Luis-Felipe Rodríguez

Instituto Tecnológico de Sonora (ITSON)

Ciudad Obregón, Sonora, México.

raguilar175318@alumno.itson.edu.mx, luis.castro@acm.org,  
luis.rodriguez@itson.edu.mx

**Resumen.** En este trabajo de tesis se estudian los eventos de vida de pacientes diagnosticadas con cáncer de mama. De manera retrospectiva, las pacientes cuentan aspectos importantes de su vida antes del diagnóstico. A través de la técnica de análisis cualitativo teoría fundamentada se busca explorar eventos de vida en pacientes sobrevivientes del cáncer de mama. Los resultados de este estudio exploratorio pueden servir como base para generar nuevos indicadores cuantitativos en el área de salud que permitan estudiar más a fondo el desarrollo de la enfermedad por medio de métodos computarizados.

**Palabras clave:** Cáncer de mama, teoría fundamentada, factores de riesgo.

## 1 Introducción

El cáncer de mama o cáncer de seno es un tumor maligno que se ha desarrollado a partir de células mamarias y es el tipo de cáncer más comúnmente diagnosticado en las mujeres. Los factores de riesgo tradicionales para el cáncer de mama incluyen el estado reproductivo, las mutaciones genéticas, la historia familiar y el estilo de vida [21]. En México, el gobierno federal, a través de la Secretaría de Salud, informa que a partir del año 2006, el cáncer de mama es la primer causa de muerte por cáncer en la mujer [18].

Más allá de los aspectos biológicos, en la literatura se ha planteado un vínculo entre el cáncer y los estados de ánimo particularmente estados afectivos que involucran pérdidas [10], depresión, preocupaciones y melancolía. De hecho, el cáncer de mama ha sido redefinido como una enfermedad biopsicosocial que involucra no solo al paciente, sino a su contexto y a quienes están en él [1]. A medida que la investigación se extiende más allá en el entorno psicosocial del paciente, las ideas con respecto a la causalidad pueden desarrollarse más completamente para incluir factores externos tales como el estrés que surge de las relaciones interpersonales [7].

En el presente documento se estudia de manera exploratoria acerca de los eventos de vida en mujeres mexicanas sobrevivientes del cáncer de mama. Los eventos de la vida y las

Aguilar Arredondo R, Castro LA, Rodríguez L-F (2018) Eventos de Vida y su Relación con el Padecimiento de Cáncer de Mama: Un estudio Exploratorio. Avances de Investigación en Ingeniería en el Estado de Sonora 4 (1):64-73

reacciones psicológicas y de comportamiento que los acompañan con frecuencia tienen un impacto en la vida cotidiana de las personas y se cree que los predisponen a esta enfermedad [6][16]. Los estudios de observación han establecido que los acontecimientos estresantes de la vida, a menudo definidos como una acumulación de acontecimientos de la vida ordinarios o duelo, aumentan los riesgos de trastornos mentales, infecciones agudas como el resfriado común llegando a causar la muerte [23]. Otros estudios sugieren una relación causal entre el maltrato infantil no sexual y una variedad de trastornos mentales, uso de drogas, intentos de suicidio, infecciones de transmisión sexual y conductas sexuales de riesgo. Todas las formas de maltrato infantil deben considerarse riesgos importantes para la salud con un impacto considerable en los principales contribuyentes a la carga de la enfermedad en todas partes del mundo [17].

En este trabajo se realizaron entrevistas semi-estructuradas con pacientes diagnosticados con cáncer de mama. Se utiliza la técnica de teoría fundamentada para el análisis de los datos recolectados. Los resultados pueden servir como base para generar nuevos indicadores cuantitativos en el área de salud que permitan estudiar más a fondo el desarrollo de la enfermedad por medio de métodos computarizados.

Esta investigación se organiza como sigue. En la sección 2 se presentan trabajos relacionados que se han realizado. En la sección 3 se presenta la Metodología de este trabajo de tesis. En la sección 4 se muestran y discuten los resultados. Finalmente, se presentan las Conclusiones y Trabajo Futuro.

## 2 Marco Teórico

En la actualidad, la orientación de las áreas médicas se centran en el diagnóstico temprano del cáncer de mama por medio de la realización de estudios radiológicos de mastografía en periodos recomendados según el rango de edad, antecedentes genéticos y otros factores [11]. De esta manera, se permite la temprana detección y tratamiento para este diagnóstico. Sin embargo, como se puede apreciar, mucho de esto se basa en la detección temprana, y no necesariamente en modelos de prevención adecuados.

La prevención inicia antes de la etapa de diagnóstico de cáncer de mama, por lo que se considera de suma importancia el estudio de diversos factores procedentes de diferente naturaleza como estado de salud, antecedentes genéticos, tabaquismo, proximidad a áreas geográficas donde se manejen productos químicos. Sin embargo, factores como los que se mencionan previamente como los eventos de vida normalmente no son considerados. En este estudio se centra en los eventos de vida del paciente como posibles factores con cierta influencia para este padecimiento.

Existen estudios donde se analizan factores de riesgo como son hábitos de vida [2], uso prolongado de tratamientos hormonales [5], relación o cercanía al uso de productos químicos o agroquímicos [3], [14], [25]. Hasta el mejor saber de los autores, pocos trabajos se han centrado en entender factores emocionales y su influencia en el padecimiento de cáncer de mama [6]. Más allá de eso, tampoco se han definido indicadores que permitan

cuantificar y estudiar a gran escala si es que en realidad existe una relación causal en cierta medida.

Otros trabajos han relacionado algunos aspectos emocionales derivados de la personalidad, y han estudiado algunos de estos aspectos en personas que padecen cáncer. La personalidad puede entenderse como un sistema interno automático de entornos únicos para cada individuo e incluye modelos de pensamientos, emociones y comportamientos junto con algunos patrones adquiridos [8].

A pesar de la relación bien establecida entre los rasgos de personalidad y la depresión y la ansiedad [1], se identificaron pocos estudios que examinaran esta relación en pacientes oncológicos [6]. Estudios sugieren que niveles más altos de neuroticismo aumentan el riesgo de depresión y ansiedad en pacientes con varios tipos de cáncer [15]. Por ejemplo, entre las mujeres con cáncer de mama que se sometieron a tratamiento quirúrgico, la personalidad neuroticista en estas pacientes aumentó el riesgo de depresión, de igual forma entre los pacientes con cáncer de pulmón. Muchos otros artículos han examinado la relación entre la personalidad y los síntomas físicos (por ejemplo, dolor, fatiga) en pacientes con cáncer [15] [4]. Como se puede apreciar, la personalidad, la manera de expresar las emociones, y la manera en que se afrontan las situaciones difíciles pueden tener efectos negativos en la salud, incluyendo el cáncer. Debido a esto, no es descabellado pensar que los eventos de vida negativos en la vida de las personas pueden tener efectos adversos en la salud de las mismas.

### 3 Métodos

#### 3.1 Pregunta de investigación y objetivo general

La pregunta principal que guía esta investigación de carácter exploratorio es ¿Qué patrones se pueden observar en cuanto a los eventos de vida negativos en los sobrevivientes de cáncer de mama?

Por otro lado, el objetivo general de este trabajo de tesis es identificar eventos de vida negativos con mayor incidencia para el grupo de mujeres sobrevivientes de cáncer.

#### 3.2 Participantes

Los participantes son mujeres pertenecientes a un grupo de apoyo para sobrevivientes de cáncer de mama. Se reclutaron nueve participantes. En la siguiente Tabla 1 se presentan los datos demográficos. Todas las participantes con diversas características educativas, ocupacionales y socio-económicas, a quienes se les había diagnosticado cáncer de mama, y con capacidad para hablar. Se firmaron cartas de confidencialidad y consentimiento informado para participar en el estudio.

**Tabla 1.** Datos demográficos de las participantes

ID	Edad	Peso Kg.	Ocup.	Hijos	Procedimiento quirúrgico	Seno afectado	Estado civil	Diagnóstico
E1	63	49	Ama casa	2	Mastectomía parc.	Izquierdo	Viuda	Dic 2016
E2	50	74	Ama casa	3	Mastectomía parc.	Izquierdo	Casada	Nov 2015
E3	41	85	Empleada	2	Ninguno	Derecho	Casada	Ene 2017
E4	47	64	Enfermera	1	Mastectomía parc.	Derecho	Divorciada	Feb 2015
E5	56	85	Ama casa	3	Mastectomía total	Derecho	Casada	Mar 2015
E6	44	78	Empleada	2	Mastectomía total	Izquierdo	Divorciada	Mar 2012
E7	54	59.5	Enfermera	2	Mastectomía total	Izquierdo	Casada	Feb 2018
E8	47	63	Empleada	3	Mastectomía total	Izquierdo	Divorciada	Jul 2017
E9	48	79	Empleada	3	Mastectomía total	Izquierdo	Unión libre	Sept 2017

### 3.3 Recolección de datos.

Se realizaron nueve entrevistas semi-estructuradas, las cuales fueron grabadas en audio. Las entrevistas tuvieron una duración promedio de 01:14 horas. El protocolo de entrevista consistió de 54 preguntas. Los tópicos incluidos son: aspectos de familia nuclear y extensa, relaciones laborales, relaciones sentimentales, situaciones fortuitas a lo largo de su vida. Previo a la entrevista, se realizó una breve conversación donde se comparte la razón y objetivos de la entrevista, posteriormente la sensibilización y motivos por los cuales debe haber apertura mental y emocional al compartir la información, las participantes llenaron un formulario de consentimiento previo al inicio de las preguntas.

### 3.4 Análisis de los datos.

Para el análisis se siguieron los lineamientos de la teoría fundamentada [9], [12]. La Teoría Fundamentada tiene por objetivo comprender la realidad a partir de la percepción o significado que cierto contexto u objeto tiene para la persona, generando conocimientos, aumentando la comprensión y proporcionando un guía significativo para la acción. Consiste en metodología de investigación cualitativa que extrae de las experiencias vividas por los actores sociales aspectos significativos, posibilitando conectar constructos teóricos, potencializando la expansión del conocimiento en enfermería y de otras áreas como psicología, sociología, medicina y otras [9], [20].

La primera etapa es la codificación de cada una de las transcripciones (la asignación de las unidades de texto a cada categoría de análisis, en cada una de las transcripciones). Esta etapa, conocida como codificación abierta, se ha realizado mediante el análisis individual de cada entrevista y la posterior asignación de códigos a las unidades de texto para terminar en la creación y agrupación de las categorías. En una segunda etapa, en la codificación axial, se redefinen y refinan las categorías depurando aquellas que según el criterio de investigación no sean de gran aporte al objetivo del estudio o bien que no se relacionen de

manera estrecha con el objetivo. En la tercera etapa, codificación selectiva, se selecciona una de las categorías principales para construir una teoría final.

## 4 Resultados

Con referencia a los resultados de la investigación cualitativa que arroja la codificación y categorización se encuentra lo siguiente. Se identificaron un total de 234 códigos en la primera etapa. Dichos códigos generaron 4 categorías principales que fueron agrupadas para su análisis, las cuales se describen a continuación:

### 4.1 La familia generadora de entornos influyentes en emociones

Existe una enorme coincidencia en la agrupación de los códigos encontrados sobre el cómo las relaciones familiares han creado los entornos que han influido en las emociones experimentadas en diferentes etapas de la vida, citaremos algunas de estas situaciones. Por ejemplo, E2 nos comenta: *“Mi relación con mi papá era muy buena, él era pescador y cuando llegaba a la casa yo lo abrazaba y me sentaba a platicar con él y yo era su preferida como fui muy enfermiza, yo me refugiaba mucho en mi papá. Cuando él se iba a pescar, duraba un mes o más y ahí era cuando pensaba que era mucho tiempo sin verlo. Mi mamá era muy estricta y nos pegaban a diario, y por cualquier cosa y nos decía ‘pobre de ti si lloras’, y después de eso me costaba mucho trabajo poder llorar, nos chicoteaba y eran fuertes sus golpes nos dolían mucho”*. En el caso particular de E2, se desarrolló en un entorno familiar donde predominaban las agresiones físicas y emocionales, así como la represión de sentimientos.

Por otro lado, E3 nos comenta de manera similar que: *“Mi papá siempre fue una persona muy seria. Él siempre trabajaba. Siempre estaba trabajando. Es chofer y fue un papá ausente y pues así crecí, y me acostumbré a esa situación. Mi papá es alcohólico y me daba miedo cuando llegaba tomado a la casa que discutiera con mi mamá. Eso me dolía mucho y siempre tuve ese temor. Así fue durante todo el tiempo que viví con ellos, que duró 19 años. También me dolió mucho el saber que mi papá tenía relaciones con otras mujeres, y pues me dolía mucho por mi mamá”*. Aquí podemos encontrar de igual forma un entorno familiar donde prevalecían emociones de incertidumbre, temor, agresiones verbales que mantuvieron por un largo periodo perturbando el desarrollo psicosocial de E3. Los entornos de agresiones físicas y verbales parecen ser algo frecuente en nuestras informantes. De igual forma E7 nos comparte *“Mi mamá nos pegaba mucho. Quería educar a base se golpes y mi papá no. Era todo lo contrario. Cuando estaba chica sentía que la odiaba”*. En este caso, E7 nos habla de un sentimiento de odio hacia su progenitora, pero también de comparación de su padre y su madre, y las diferentes estrategias que E7 percibía que usaban para tratar de imponer disciplina.

El comportamiento actúa recíprocamente con la salud del propio individuo de manera que el nivel de salud es, en parte, función del comportamiento del sujeto, y éste a su vez está condicionado por las características psicológicas, biológicas y de ecosistema del propio

individuo [24]. Por lo anterior, se puede concluir que en el núcleo familiar existía violencia física y emocional además de represión por expresar sentimientos y emociones mismos que fueron bloqueados y acumulados por largos periodos de tiempo pudiendo así afectar el desarrollo biopsicosocial, comprometiendo el futuro de la salud de los sujetos de estudio.

#### **4.2 Profanación de su cuerpo y pensamientos**

En su totalidad, todas las entrevistadas (9 de 9) comparten haber sido hostigadas y abusadas física y/o emocionalmente en algún momento de la vida. Por ejemplo, E2 nos comparte: *“En otra clínica donde trabajé, un doctor me acosaba. Según él me estaba dando mi tratamiento para embarazarme [pero] después cambie de médico, pero él me decía que mi esposo no podía y que él podía influir para que yo me embarazara y yo le dije que no. Primero pierdo el trabajo a que pasara algo. En una ocasión entré al consultorio a llevarle expedientes y ahí me agarro a la fuerza y... (llora fuertemente)”*. Casos como el que menciona E2 son comunes en nuestras informantes. En el caso de E2, fue víctima de agresión emocional y sexual por parte de un superior, influyendo de modo biopsicosocial en el desarrollo integral de modo permanente en su vida.

Por otro lado, E3 comenta de manera similar: *“Tengo un vecino. Es abogado. El llevó mi caso cuando demandé hace 12 años, [él] iba a mi casa y siempre me pedía dinero y como que iba también con otras intenciones, y me molestaba mucho, y eso duró como un año y hasta la fecha lo veo y me molesta”*. De igual forma E8 nos comparte: *“con mi primera pareja, pues él era un hombre muy prepotente. Tomaba mucho y había agresión verbal y física, aunque no de alto grado, pero si había también infidelidades de su parte y cuando llegaba a la casa en la madrugada me sacaba a empujones de la casa que porque no le tenía la comida lista, y pues yo me quedaba callada, y ahora veo que fue un error. Yo le tenía mucho resentimiento. Yo estaba consciente que fui una mujer abnegada y le tenía miedo, pero si deseaba que se muriera. El abusaba física, sexual y emocionalmente de mí”*. Aun cuando los casos de E2 y E8 difieren en que una era abusada por su esposo, y una por un superior, sus casos no son particularmente raros. Todas nuestras informantes reportaron haber sufrido abuso en algún momento de su vida. Tal como se ha reportado, el acoso emocional y el abuso sexual tiene repercusiones somáticas, psicológicas y psicofisiológicas [19]. De manera adicional, se identifican a la violación y a la violencia doméstica como causas significativas de trastorno por estrés postraumático y depresión en mujeres de todo el mundo afectando de manera significativa su salud.

#### **4.3 Muerte de familiares y seres allegados**

Todas nuestras informantes (9 de 9) reportaron haber sufrido una pérdida de familiares cercanos y/o seres allegados que representaban algo muy importante en su vida. Por ejemplo, E1 nos comenta: *“Tuve 2 hijas. Con la primera, excelente. Es lo máximo para mí, muy buena relación. Todo lo que guardo en mi corazón son puras cosas bonitas sobre ella. Es una hija excelente. Y la segunda nació prematura. A los 7 meses, y murió”*. E1 nos

comparte un enorme gozo, orgullo y satisfacción por su primera hija. Sin embargo, la gran pérdida de su segunda hija en el embarazo no llevado a término la sigue perturbando enormemente, aun en estos días después de más de 30 años de sucedido. Por otro lado, E2 dijo: *“Yo tuve 3 abortos. No se me lograban los bebés. Los 2 primeros de 4 meses y el otro fue provocado por una cuñada que me dio una patada en el abdomen en una discusión que tuvimos en la casa de mi suegra. Después de ese aborto, tuve una niña que nació y murió de insuficiencia cardíaca. Mi hermano, el de en medio, murió hace 23 años. Murió cuando tenía 22 años. Él era mi aliado. Yo lo quería mucho. Murió en un accidente automovilístico y más porque yo no sabía llorar y me tuvieron que hospitalizar. También la muerte de mi papa me pudo mucho. Murió hace 16 y también la muerte de mi hermano, el mayor, hace 8 años”*. En este caso, E2 expresa su gran dolor al perder sus 2 primeros bebés. Sin embargo, comparte que la pérdida del tercer bebe fue aún más dolorosa por las condiciones en que sucedió, resultado de una agresión física. De acuerdo a lo que comenta, el dolor parece incrementar en ella por la gran cantidad de pérdidas presentadas a lo largo de su vida. De manera adicional, E3 nos comparte: *“En noviembre 2017, falleció una sobrina de cáncer, hija de un primo hermano ella tenía 19 años y tenía cáncer en los pulmones. Su problema duró como 8 meses, me revivió mis miedos, volví a sentir todo por lo que pasé, cuando yo tuve mi problema”*.

Por otro lado, E4 nos comparte lo siguiente: *“Yo tenía como 9 o 12 años cuando una de mis hermanas falleció. Ella tenía como 18 años. Iba con un grupo de amigos al mar y a mí y a una prima nos mandaron con ella, y ahí ella se ahogó. No me di cuenta porque yo estaba jugando, y luego la empezaron a buscar y no la encontraban, y ahí fue cuando me di cuenta y pues nos asustamos, y no recuerdo yo hasta que nos dijeron que se había ahogado. En ese momento no era consciente. Ya en mi casa, mi mamá estaba llorando y llorando y recuerdo que la llevaron a mi casa [a mi hermana], y mi mamá la bañó y ahí la velaron. Yo me sentí culpable y más que dolor sentía miedo de eso que pasó”*. De acuerdo a [13], existen casos en que las reacciones iniciales posteriores al duelo se acentúan y prolongan por algunos años, que resultan ser autolimitadores para quien lo padece. Además, desencadenan alteraciones físicas o mentales de gran relevancia que pueden aumentar la probabilidad de morbilidad y mortalidad en determinados tipos de poblaciones. Lo anterior sustenta las afecciones mentales y emocionales sufridas por la muerte de familiares en nuestras entrevistadas volviéndolas vulnerables a múltiples padecimientos tanto físicos como emocionales.

#### **4.4 Abuso de sustancias tóxicas en integrantes de la familia**

Seis de nuestras informantes comparten que algunos de los integrantes de su familia consumieron sustancias tóxicas como alcohol, marihuana, o cocaína. E9 nos comparte: *“Cuando él [su papá] ganaba dinero, se iba a las cantinas y se lo acababa todo tomando. Aunque cuando yo crecí, ya no tomaba tanto y tengo recuerdos muy malos de esa etapa, y recuerdo que mis papás duraban meses sin hablarse a causa de borracheras e infidelidades de mi papá, dolía, aunque estaba muy chica”*. Además, E5 nos comparte de modo similar

respecto a su esposo: *“también tuvo problemas de alcoholismo y adicción a las drogas durante 15 años y teníamos muchos problemas por ese motivo”*. Además, E5 nos comenta: *“Una de mis hijas se fue a estudiar a EU, y si me dolió, sobre todo cuando supe que tenía problemas de drogadicción. En 3 ocasiones tuvo ese problema. Perduró por casi 6 años. Mostraba conductas violentas hacia todo mundo, pero ya salió adelante”*. Como se aprecia, la violencia puede ocurrir en cualquier etapa de la vida de la mujer y muchas mujeres experimentan múltiples episodios durante sus vidas, lo cual puede tener efectos inmediatos y acumulativos sobre su salud y desarrollo de sus capacidades cognitivas, afectivas, económicas y de relación [22].

## 5 Conclusiones y Discusión

En este trabajo se presentó un estudio exploratorio de mujeres sobrevivientes con cáncer de mama. Nuestras informantes reportan que habrían podido evitarse sufrimientos emocionales vistos como innecesarios durante algunas etapas de su vida. Tal reflexión puede ir encaminada hacia el manejo integral del cáncer, asociando al psiquiatra, para que este pueda tratar o prevenir los trastornos mentales que puedan surgir durante el desarrollo de la enfermedad y sus tratamientos.

De manera adicional nuestras informantes reportan sobre los eventos de vida negativos con mayor incidencia para el grupo de mujeres sobrevivientes de cáncer. En la gran mayoría de los casos convergen en situaciones de abuso físico y emocional, pérdida de integrantes de la familia con los que mantenían relaciones estrechas, y la gran influencia de las emociones generadas en los entornos de familia de origen y familia nuclear, sin perder de vista las afectaciones emocionales generadas por integrantes de familia que usaron sustancias tóxicas.

Los hallazgos reportados en este estudio muestran que todas las participantes presentan situaciones en cuando menos dos de las dimensiones encontradas en la investigación. Debido a eso, se requieren estudios más profundos para tratar de asociar de manera significativa estos eventos. Indicadores emocionales normalmente no son considerados dentro de los estudios clínicos, por lo que no se descarta que podrían aportar información de interés, y complementaria sobre este tema.

Una de las limitantes de este trabajo es el número reducido de sobrevivientes participaron en el estudio. Otra más es que al aun encontrarse en tratamientos paliativos, se programan entrevistas que frecuentemente son canceladas por las entrevistadas al ser llamadas para atención médica. Como trabajo futuro, se pretende explorar la relación de ciertos rasgos de personalidad con la intensidad con que afrontan diversos eventos de vida negativos.

### Agradecimientos

Un agradecimiento muy especial a las mujeres que integran el Grupo Reto Recuperación Total Los Mochis, A.C., ya que fueron las que brindaron información valiosa para esta investigación mostrando gran actitud y disposición por aportar.

## Referencias

1. Arbizu, J.P.: Factores psicológicos que intervienen en el desarrollo del cáncer y en la respuesta al tratamiento Psychological factors in the development of cancer and in the response to treatment. *An. Sis San Navarra Hosp. Navarra. Pamplona An. Sis San Navarra.* 24, 24, 173–178 (2000).
2. Blair, A. et al.: Causes of death among laundry and dry cleaning workers. *Am. J. Public Health.* 69, 5, 508–511 (1979).
3. De Brito Sá Stoppelli, I.M., Crestana, S.: Pesticide exposure and cancer among rural workers from Bariri, São Paulo State, Brazil. *Environ. Int.* 31, 5, 731–738 (2005).
4. Brown, F.: The Relationship between cancer and personality. *Ann. N. Y. Acad. Sci.* 125, 3, 865–873 (2006).
5. Calle, E.E. et al.: Breast cancer and hormone replacement therapy: Collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. *Lancet.* 350, 9084, 1047–1059 (1997).
6. Cano-Vindel, A.: Control emocional, estilo represivo de afrontamiento y cancer: ansiedad y cancer. *Archit. City Environ.* 2, 20, 131–148 (2012).
7. Carter, R.E. et al.: Emotional and personality types of breast cancer patients and spouses. *Am. J. Fam. Ther.* 20, 4, 300–309 (1992).
8. Casellas-Grau, A. et al.: Positive psychological functioning in breast cancer: An integrative review. *The Breast.* 27, 136–168 (2016).
9. Dantas, C. de C. et al.: Teoría fundamentada en los datos - aspectos conceptuales y operacionales: Metodología posible de ser aplicada en la investigación en enfermería. *Rev. Lat. Am. Enfermagem.* 17, 4, 573–579 (2009).
10. Eugenia, V., Facio, D.: Sexualidad , cuerpo y duelo : experiencia clínica con mujeres diagnosticadas con cáncer ginecológico o de mama. 7, 14, 155–160 (2010).
11. Hernández, D. et al.: Cáncer de mama en mujeres jóvenes. diciembre 2010 *Rev Venez Oncol.* 2222, 44, 216–221 (2010).
12. Jones, D. et al.: Grounded Theory Una aplicación de la teoría fundamentada a la salud. *Cinta de Moebio.* 19, 1–19 (2004).
13. López Rodríguez, A.: Importancia de la atención del paciente en duelo. *Rev. Espec. Médico-Quirúrgicas.* 14, 4, 153–4 (2009).
14. Mathur, V. et al.: Breast cancer incidence and exposure to pesticides among women originating from Jaipur. *Environ. Int.* 28, 5, 331–336 (2002).
15. Morgan, S. et al.: Association of personality profiles with depressive, anxiety, and cancer-related symptoms in patients undergoing chemotherapy. *Pers. Individ. Dif.* 117, 130–138 (2017).
16. Moscoso, M.S.: De la mente a la célula: impacto del estres en psiconeuroinmunoendocrinología. *Liberabit.* 15, 2, 143–152 (2009).

17. Norman, R.E. et al.: The Long-Term Health Consequences of Child Physical Abuse, Emotional Abuse, and Neglect: A Systematic Review and Meta-Analysis. *PLoS Med.* 9, 11, (2012).
18. Palacio-Mejía, L.S. et al.: Diferencias regionales en la mortalidad por cáncer de mama y cérvix en México entre 1979 y 2006. *Salud Publica Mex.* 51, SUPPL.2, (2009).
19. Ramos-Lira, L. et al.: Violencia sexual y problemas asociados en una muestra de usuarias de un centro de salud. *Salud Publica Mex.* 43, 3, 182–191 (2001).
20. Strauss, A., Corbin, J.: *Bases De La Investigación Cualitativa: Técnicas Y Procedimientos Para Desarrollar La Teoría Fundamentada.* 1, (1991).
21. Tyrer, J. et al.: A breast cancer prediction model incorporating familial and personal risk factors. *Stat. Med.* 23, 7, 1111–1130 (2004).
22. Valdez, R., Juárez, C.: *Impacto de la violencia doméstica en la salud mental de las mujeres: Análisis y perspectivas en México,* (1998).
23. Vázquez, C. et al.: Bienestar psicológico y salud: aportaciones desde la psicología positiva. *Anu. Psicol. Clínica y la Salud.* 5, 1, 15–28 (2009).
24. Vinaccia, S., Orozco, L.M.: Aspectos psicosociales asociados con la calidad de vida de personas con enfermedades crónicas. *Perspect. en Psicol.* I, 2, 125–137 (2005).
25. Wolkoff, P. et al.: Risk in cleaning: Chemical and physical exposure. *Sci. Total Environ.* 215, 1–2, 135–156 (1998).

**ISSN: 1870-4069**

---

Copyright © Instituto Politécnico Nacional 2018

Instituto Politécnico Nacional (IPN)  
Centro de Investigación en Computación (CIC)  
Av. Juan de Dios Bátiz s/n esq. M. Othón de Mendizábal  
Unidad Profesional “Adolfo López Mateos”, Zacatenco  
07738, México D.F., México

<http://www.rcs.cic.ipn.mx>

<http://www.ipn.mx>

<http://www.cic.ipn.mx>

The editors and the publisher of this journal have made their best effort in preparing this special issue, but make no warranty of any kind, expressed or implied, with regard to the information contained in this volume.

All rights reserved. No part of this publication may be reproduced, stored on a retrieval system or transmitted, in any form or by any means, including electronic, mechanical, photocopying, recording, or otherwise, without prior permission of the Instituto Politécnico Nacional, except for personal or classroom use provided that copies bear the full citation notice provided on the first page of each paper.

Indexed in LATINDEX, DBLP and Periodica

Printing: 500

Printed in Mexico

# Design and Implementation of a Data Warehouse to Support Decision-Making in a Health Environment

Simon G. Cornejo<sup>1</sup>, Karina Caro<sup>2</sup>, Luis-Felipe Rodriguez<sup>1</sup>, Roberto Aguilar A.<sup>1</sup>,  
Cynthia B. Perez<sup>1</sup>, Luis A. Castro<sup>1</sup>

<sup>1</sup> Sonora Institute of Technology (ITSON), Ciudad Obregon, Sonora,  
Mexico

<sup>2</sup> Drexel University, Philadelphia, Pennsylvania,  
USA

{scorejo175319 raguilal75318}@alumno.itson.edu.mx, karinacar@drexel.edu,  
{luis.rodriguez, cynthia.perez}@itson.edu.mx, @itson.edu.mx, luis.castro@acm.org

**Abstract.** One of the most common challenges in the management of electronic medical records (EMR) is to extract critical knowledge that could serve to enhance collaboration among medical doctors to support decision-making. In this paper, we present the design and implementation of a data warehouse designed to strengthen decision-making related to epidemiological patterns, trends, areas of influence and other pathologies. The data warehouse is based on a relational model that contains information about EMR of different states of Mexico. We used the Business Event Analysis and Modeling (BEAM) methodology for the design and implementation of the data warehouse, a novel methodology to design agile data warehouses. Using BEAM, we created a star dimensional model and defined the Extract, Transform and Load (ETL) processes to transfer the data to the new model. In order to show the potential of our data warehouse, an interactive dashboard with different indicators was built. We close discussing how the medical doctors could use our data warehouse to support the decision-making process.

**Keywords:** decision support systems, data warehouse, health environment, BEAM, ETL.

## 1 Introduction

In recent years, the use of electronic medical records (EMR) has increased considerably, changing the way traditional records are stored and managed [1]. In EMR, the data is managed and stored digitally, enabling health professionals to maintain information in one place. The use of these systems has enabled the accumulation of large amounts of data, opening up opportunities for analyzing and obtaining relevant information from them [2].

However, in most cases, besides data queries, these data are not analyzed, leaving aside all the knowledge that can be obtained from them. These large amounts of data and the knowledge that can be obtained from them open the possibility of supporting decision-making in the health environment, as well as promoting communication and collaboration among health professionals. Having a tool to support decision-making that could provide knowledge based on patients' historical data stored in EMR, might help to a great extent to improve health services.

An alternative to support this problem is the use of a data warehouse, a repository that preserves the historical context to accurately assess an organization's performance over time [3]. It is optimized for high-throughput queries, since user queries often require that hundreds or thousands of transactions are searched for and compressed into a set of responses. Data warehouses are the basis of the processing of Decision Support Systems (DSS). They facilitate the analysis since all data are concentrated in a single data source, which can integrate both structured and unstructured data, with different granularity. A data warehouse based on the data of an EMR could support health professionals to obtain relevant knowledge to enhance the decision-making process. However, the design and construction of a data warehouse is not an easy task. First, the preparation and cleaning of large amounts of data present some challenges such as the concentration of heterogeneous data, incomplete records, and integrity errors, among others. On the other hand, the design of the data warehouse has to be done in such way that the response time of the queries is fast and the results are correct.

An optimal design of the data warehouse is required to facilitate the extraction and analysis of the stored data. The use of agile methodologies (e.g., SCRUM [4]) have presented multiple advantages in software development such as customer satisfaction by the rapid and continuous delivery of useful software. Agile methodologies are characterized by emphasis on stakeholders and interactions rather than processes and tools [4]. On the side of the design of data warehouses, the Business Event Analysis and Modeling (BEAM) methodology [5] offers several advantages for designing data warehouses. For example, individuals and interactions over processes and tools, working software over comprehensive documentation, and customer collaboration over contract negotiation. BEAM upholds these values and the agile principle of data warehouse practitioners to work directly with stakeholders to produce data models rather than requirements documents, and working Business Intelligence (BI) prototypes of reports / dashboards rather than mockups.

This paper presents the design and implementation of a data warehouse using BEAM methodology, which information was obtained from EMR. This data warehouse provides information in the form of indicators that summarize what is happening in a health environment and support appropriate and timely decision making through the identification of diseases by geographic zones, diseases by stage of life and diseases by season of the year.

The illustration of the usefulness of the designed data warehouse, following, we present a usage scenario where health professionals use the information they visualize in an interactive dashboard to support decision making:

A health professional is concerned that many patients are presenting with a rare disease in their geographical area and he is not sure that the medications he is prescribing to his patients are the most appropriate. The health professional reviews the information provided by an interactive dashboard and he realizes that the disease which

he is dealing with is very common in another geographic area of Mexico. Thus, the health professional starts to get in touch with other health professionals in that geographical area to ask for opinions and share experiences in the treatment and intervention of that disease.

## **2 Related Work**

Research contributions have been made in the areas of data warehouse design, data staging for ETL processing, data quality assurance, and healthcare data warehouse applications, mainly in developed countries, such as the United States. Existing EMR [6] data are made available in a standardized and interoperable format, thus opening up a world of possibilities for semantic or concept-based reuse, consultation and communication of clinical data. The Community Health Applied Research Network (CHARN) [7], in the United States, represents more than 500,000 patients from diverse safety nets in 11 states, aims to create a national and centralized data warehouse with multiple partners from the Center for Community Health using different EMR systems.

The work [8] describes a virtual data warehouse (VDW) of the Health Maintenance Organization Research Network (HMORN), a public, research-centric data model implemented in 17 health care systems across the United States. At the Catholic Health Initiatives research institute, data consultation tools [11] are implemented to enable end users to access the VDW for simple consultation and research readiness activities, capture for collection of study-specific data and results reported by the patient. On the other hand, the decision support system [9] based on multi-criteria data analysis – Annalisa, is an online decision support tool for individuals and clinicians interested in making a shared decision.

In Mexico, there have been initiatives and programs of innovation and technological development of the public and private sectors that have begun to get involved in the subject of e-health. In the early 1990s, the State Basic Information System (SEIB) was centralized and encompassed all 32 states by The Ministry of Health (SSA, for its acronym in Spanish). In 1995, the National Epidemiological Surveillance System (SINAVE) was created. Its coordination is carried out by SSA and it is supported by the Single Information System for Epidemiological Surveillance (SUIVE).

In 2007, SSA initiated the development of the Mexican electronic clinical record standard. In 2011, a study was conducted in Mexican Institute of Social Security (IMSS, for its acronym in Spanish) [10] to develop Quality of Care Indicators (QCI) for Type 2 Diabetes Mellitus (T2DM). The goal was to determine the feasibility of constructing QCI using IMSS's EMR data and assessing the Quality of Care (QC) provided to IMSS patients with T2DM. As a result of this study [10], 18 QCIs were developed, of which 14 were possible to construct using available EMR data. ICQs comprised both the care process and health outcomes.

The related work shows that there is a potential of using EMR to enhance health care services. However, there are few studies that use the EMR data to obtain knowledge that could support the decision-making process in the healthcare environment. In this work, we propose to use EMR data to design and implement a data warehouse aimed at supporting the decision-making in the healthcare environment. In the next section,

we describe the BEAM methodology, following by the results of the design of the data warehouse and its implementation.

### **3 BEAM Methodology**

We used the Business Event Analysis and Modeling (BEAM) methodology to design the proposed data warehouse. To the best of our knowledge, BEAM is one of the most recent and the first agile methodology in the area of Data Warehouse and Business Intelligence (DW / BI) [5]. The BEAM methodology comprises a set of collaborative techniques for modeling BI data requirements and translating them into dimensional models on an agile time scale. Among the techniques used by the BEAM methodology are the 7W's Framework, BEAM\*tables, Event Matrix and Enhanced Star Schema.

#### **3.1 7W's Framework and BEAM\*tables**

The 7W's framework uses questions about who, what, where, when, how many, why, and how[5], data modelers design the model by asking BI stakeholders to tell data stories using these questions. BEAM uses tabular notation and data stories to define business events in a format that is easily recognizable and understandable to BI stakeholders. It uses spreadsheets that enable an easy translation into detailed star schemas.

BEAM\*tables help engage BI stakeholders to define reports that answer their specific business questions. They are used to define fact and dimension tables, and they use natural language enable BI stakeholders easily imagine, sort, and filter the low-level detail columns of a business event using the top-level dimensional attributes. BEAM\*tables can describe facts, events in terms of measures, and dimensions, descriptions of the facts, which can be used to filter, group and aggregate measurements.

#### **3.2 Event Matrix**

The Event matrix documents the relationships between all events and dimensions within a model. Event matrices record events in value chain sequences and promote the definition and reuse of conformed dimensions through dimensional models.

#### **3.3 Enhanced Star Schema**

A star schema consists of a central fact table surrounded by a series of dimension tables. The fact table contains facts: the numerical (quantitative) measures of a business event. Dimension tables contain mainly textual (qualitative) descriptions of the event and provide the context for the measurements. Enhanced star schemas are standard star schemas that use BEAM short codes to record dimensional properties and design techniques that are not directly supported by generic data modeling tools.

In the following section, we describe the obtained results of applying the above BEAM techniques to design a data warehouse, as well as, the data warehouse' implementation is described.

DiseasesxVectorxMunicipality [RE]

Patient	Disease	Municipality	State	Date	Amount
[who]	[what]	[where]	[where]	[when]	[how many]
Jaime Rodríguez	Peste bubónica	Cajeme	Sonora	29/01/2011	1
María Luisa Aceves	Tifus	Salina Cruz	Oaxaca	22/06/2018	3
Fernando López	Tifus	EL Fuerte	Sinaloa	20/01/2017	3
José García	Tifus	Ahome	Sinaloa	28/05/2017	3
Luis Soto	Peste neumónica	Huetamo	Michoacan	12/05/2017	1

Fig. 1. Event to show diseases by municipalities and states of Mexico.

TREATMENT [TF]

PRESCRIPTION_ID	PATIENT	DISEASE	DOCTOR	MEDICAL
	[who]	[what]	[who]	[what]
74525	Paciente 49697	Meningitis viral, sin otra especificacion	Medico 21263	AUTRIN 600
74525	Paciente 49697	Meningitis viral, sin otra especificacion	Medico 21263	BENAXIMA
70091	Paciente 105454	Fiebre exantematica enteroviral [exantema de boston]	Medico 21115	BLEMIL PLUS ARAC
73050	Paciente 218889	Vertigo epidemico	Medico 21354	DIDODOQUIN
74404	Paciente 105507	Otras infecciones virales del sistema nervioso central	Medico 13388	INVIRASE
317922	Paciente 105546	Otras infecciones virales del sistema nervioso central	Medico 21374	INVIRASE
70909	Paciente 105569	Otras fiebres virales transmitidas por mosquitos	Medico 20873	
361269	Paciente 218936	Fiebre del valle del rift	Medico 21181	PAMIGEN
74729	Paciente 218956	Fiebre del valle del rift	Medico 22178	PAMIGEN

Fig. 2. Fact table that contains the IDs that relate the dimensions.

EVENT (who does what)	Dimensions		Estimate	Doctor	Patient	Medical	Disease	Colony	Municipality	State	Treatment	Prescription	Doctor	Patient
	Importance													
Evaluate the patient	5	300		✓	✓			✓	✓	✓			*	✓
Diagnose the patient	7	200		✓	✓		✓	✓	✓	✓	✓		*	✓
Give medical treatment	6	100		✓		✓		✓	✓	✓		✓	*	✓

Fig. 3. Event matrix.

## 4 Results

The data repository used to design and implement the data warehouse is a relational database derived from EMR system, containing 316,295 records of medical consultations from different patients living in different cities and states of Mexico. The structure of this relational database was analyzed and only the required tables to create the data warehouse with the dimensional model obtained using BEAM were extracted.

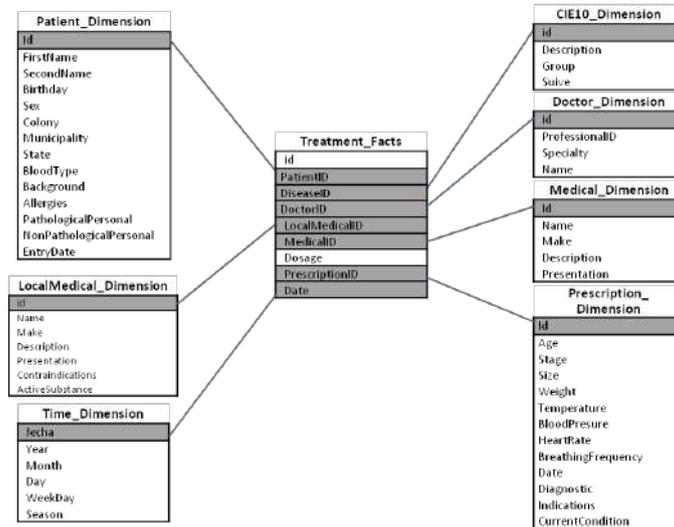


Fig. 4. Star Schema showing the dimensional model of the data warehouse.

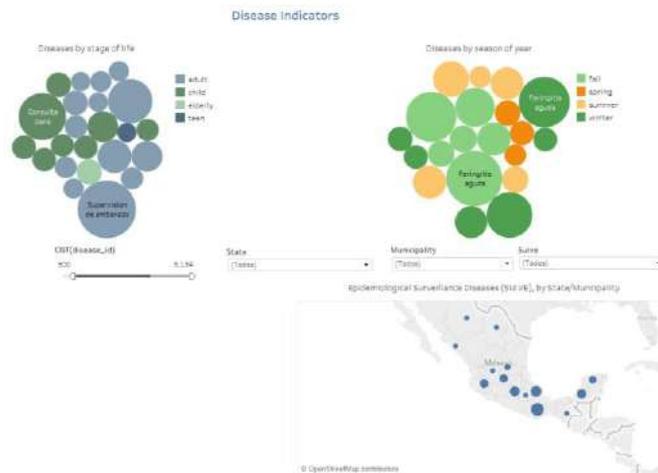


Fig. 5. Interactive dashboard connected to the dimensional model of the data warehouse.

The tables were identified according to the indicators that the stakeholders (doctors from Obregon city identified and prioritized the indicators) defined in the business events of the BEAM methodology, which would be useful for decision making. The indicators are disease by stage of life, diseases by season of the year, epidemiology surveillance diseases, all can be filtered by municipality and state. The relational database consists of 172 tables. The tables that contain the required data to create the

business events are only 7. These tables record data about patients, doctors, medications, diseases, treatments of the patients and the prescriptions that the doctors issue at the end of each medical appointment.

Once the relational database was analyzed, following the BEAM methodology, the 7W's framework technique was applied, several BEAM\*tables and the event matrix were created and finally the star dimensional model was created.

**7Ws Framework and BEAM\*Tables.** Events were defined with the help of stakeholders. In our case, the stakeholders are healthcare professionals. Examples of the defined events are Disease by stage of life, Disease by season of the year, Epidemiology surveillance diseases by municipality and state. The event shown in Fig. 1 is an example of a query to display records of patients' diseases, with granularity by municipalities and states of Mexico.

Fig. 2 shows an example of the fact table of the model that contains the IDs of the prescriptions, patients, diseases, doctors, medicines. All the descriptions corresponding to the IDs of this table are found in the derived dimension tables. This can be observed in the star dimensional model (Fig. 4), where the relationship between the IDs of the fact table and the dimension tables is defined.

**Event matrix.** Fig. 3 shows the event matrix with the events that happen during a medical consultation and relates them to the data of the dimensions that are involved in the development of the medical consultation.

**Star dimensional model.** As a result of using the BEAM techniques, we obtained the star dimensional model (Fig. 4). The dimensional model has seven dimension tables and a fact table, which contains the IDs that relate the facts to each of the dimensions. Patient dimension contains the records of the registered patients. Medical dimension contains the names, and specialty of health professionals. Medication and Medication Local dimensions are the Descriptions and brands of prescription drugs. CIE10 dimension includes the diseases and their classification SUIVE, in case of being considered epidemiological. Prescription dimension contains the symptoms of the patients at the time of the medical appointment. Time dimension is tied to the dates of the facts recorded and it enables to achieve the required granularity for the events. The relationships between the fact and the dimension tables enable agile and efficient queries, compared with the E-R model.

**Validation.** In order to validate the functionality of the dimensional model of the data warehouse, we performed the Extraction, Transformation and Load (ETL) processes of the data, as well as we created an interactive dashboard that shows the indicators resulting from the queries to the data warehouse.

*ETL processes.* The required tables were extracted from the relational database, completed and stored in temporary tables using SQL code. Next, there is an example of the code that we used to perform the extraction of the data:

```
select * into DW.dbo.paciente from mmanik_completa.dbo.paciente
select * into DW.dbo.municipio from mmanik_completa.dbo.municipio
select * into DW.dbo.estado from mmanik_completa.dbo.estado.
```

After the extraction process, the extracted data were transformed using the tables and fields of interest and loaded into the tables of the dimensional database. An example of the code used to perform the transformation and load processes is the following:

*select m.id,cedula,u.nombre,e.nombre as especialidad,cedula\_validada insert into DW.mm.medico from medico as m left join especialidad as e on especialidad\_id=e.id left join usuario as u on usuario\_id=u.id.*

*Interactive dashboard.* We designed and created an interactive TABLEAU <sup>1</sup> dashboard once the ETL processes were finished to validate the efficiency of the data warehouse design. This dashboard is connected with the data of the dimensional model and performs queries to extract the stored data in a rapid and simple way. The dashboard shows the diseases by stage of life presented by patients (Fig 5). The dashboard also shows the diseases by seasons of the year, with the same adjustments as the indicator of diseases by life stage. Additionally, the dashboard shows the diseases that are classified in the SUIVE in a geographical map.

Thus, the health professionals can observe the map with all the diseases or select one that is of his/her interest. All indicators can be filtered by state and municipality.

## 5 Conclusion

With our results it is possible to support the decision making related to epidemiological patterns, trends, areas of influence and other pathologies, based on the indicators that emerged in the business events of the proposed dimensional model.

This paper shows how the BEAM methodology can be applied to design a data warehouse based on EMR system. In the future, we plan to evaluate the use of the dashboard with health professionals to investigate the potential of the data warehouse in supporting decision making in a real-case scenario.

## References

1. Berner, E. S., Detmer, D. E., Simborg, D.: Will the wave finally break? A brief view of the adoption of electronic medical records in the United States. *J. Am. Med. Informatics Assoc.*, 12(1) pp. 3–7 (2005)
2. Abidi, S.: *Knowledge Management in Healthcare* 63, pp. 5–18 (2001)
3. Kimball, R., Ross, M.: *The data warehouse toolkit: the complete guide to dimensional modelling* (2011)
4. Skarin, M., Kniberg, H.: *Kanban y Scrum – obteniendo lo mejor de ambos* (2010)
5. Corr, L., Stagnitto, J.: *Agile Data Warehouse Design*, 1 th. Leeds, UK: DecisionOne Press (2012)
6. Marco-Ruiz, L., Moner, D., Maldonado, J. A., Kolstrup, N., Bellika, J. G.: Archetype-based data warehouse environment to enable the reuse of electronic health record data, *Int. J. Med. Inform.*, 84(9) pp. 702–714 (2015)
7. Laws, R., Gillespie, S., Puro, J., Van-Rompaey, S., Quach, T., Carroll, J., Chang Weir, R., Crawford, P., Grasso, C., Kaleba, E., McBurnie, M. A.: The Community Health Applied Research Network (CHARN) Data Warehouse: a Resource for Patient-Centered Outcomes Research and Quality Improvement in Underserved, Safety Net Populations, *eGEMS (Generating Evid. Methods to Improv. patient outcomes)*, 2(3), pp. 10–14 (2014)
8. Ross, T. R., Ng, D., Brown, J. S., Pardee, R., Hornbrook, M. C., Hart, G., Steiner, J. F.: The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support

---

<sup>1</sup> <https://www.tableau.com/>

- Collaboration, eGEMs (Generating Evid. Methods to Improv. patient outcomes), 2(1) (2014)
9. Dowie, J., Kjer-Kaltoft, M., Salkeld, G., Cunich, M.: Towards generic online multicriteria decision support in patient-centred health care, *Heal. Expect.*, 18(5), pp. 689–702 (2015)
  10. Pérez-Cuevas, R., Doubova, S. V., Suarez-Ortega, M., Law, M., Pande, A. H., Escobedo, J., Espinosa-Larrañaga, F., Ross-Degnan, D., Wagner, A. K.: Evaluating quality of care for patients with type 2 diabetes using electronic health record information in Mexico, *BMC Med. Inform. Decis. Mak.*, 12(1) pp. 50 (2012)
  11. Bailey, D., Weeks, J., Evans, E., Lowery, J., McFarland, L.: Technologies for Managing Virtual Data Warehouse Access and Identifying Appropriate Levels of Staffing at CHI Institute for Research and Innovation. *Journal of Patient-Centered Research and Reviews*, 4(3), pp. 197–198 (2017)

# Desafíos y oportunidades de aplicar tecnología de almacenamiento de datos como apoyo en la toma de decisiones gerenciales en una PyME de la industria fotovoltaica

Ana Laura García de León Villegas<sup>1</sup>, Cynthia B. Pérez Castro<sup>2</sup>

<sup>1</sup>Instituto Tecnológico de Sonora Campus, Unidad Nainari,  
Maestría en Tecnologías de la Información para los Negocios  
Antonio Caso 2266, Villa Itson, C.P. 85130, Cd. Obregón, Sonora, México.  
ana.garciadeleon@potros.itson.edu.mx

<sup>2</sup>Instituto Tecnológico de Sonora, Unidad Guaymas,  
Km 3, Carr. a Aeropuerto de Guaymas, 85400 Heroica Guaymas, Son.  
cynthia.perez@itson.edu.mx

**Resumen.** Las PyMES (Pequeñas y Medianas Empresas) mexicanas de la industria fotovoltaica, actualmente pasan por una etapa de evolución tecnológica. Por esta razón, es de su interés adoptar recursos tecnológicos para obtener una ventaja competitiva en el mercado. En el presente artículo se propone un análisis sobre los desafíos y oportunidades mediante el uso de tecnología de un sistema de inteligencia de negocios para el apoyo a la toma de decisiones estratégicas de una empresa de la industria fotovoltaica.

**Palabras clave:** Almacenamiento de datos, soporte para toma de decisiones, inteligencia de negocios, análisis de requerimientos.

## 1 Introducción

Los sistemas de inteligencia de negocios (BI, *Business Intelligence*), han sido una de las principales propuestas de solución tecnológica desde hace varias décadas ya que permiten analizar información crítica y con ello, entender desde otra perspectiva el negocio y el mercado [1-3]. Para esto, el diseño y construcción del almacén de datos (datawarehouse) es fundamental en este tipo de sistemas ya que es la fuente de información donde el sistema de inteligencia de negocios lleva a cabo las consultas estratégicas que facilitarán la toma de decisiones. De esta manera, existen diferentes metodologías para el diseño de un datawarehouse [4-6], no obstante la metodología de Ralph Kimball es la más utilizada por varias décadas en diferentes organizaciones por su centralización en la empresa y la naturalidad del proceso.

En ese sentido, en la Figura 1 se muestra la estructura de un almacén de datos, en donde se extrae la información de diversas fuentes (CRM, ERP, Web), para después ser transformados siguiendo el proceso de Extracción, Transformación y Carga conocido como

ETL (Extract, Transform and Load). La finalidad del ETL es contar con la información depurada y estructurada para llevar a cabo distintos análisis con los datos, para identificar patrones, relaciones y tendencias que son de utilidad para la toma de decisiones, alertando a los gerentes sobre posibles amenazas u oportunidades para el negocio, [7]. Los empresarios de las PyMEs en México por lo regular toman decisiones con base en la información que tienen en ese momento, que generalmente son reportes en físico, o bien, necesitan que sus empleados les hagan llegar la información requerida para tomar una decisión. Por lo tanto, resulta para el empresario inversión de tiempo y esfuerzo, así como muchas veces pérdida de dinero y de recursos.

En ese sentido, un sistema de BI resulta ser una excelente opción, mediante el cual podrían obtener diversos beneficios presentados en múltiples casos de éxito [1-3]. En el caso de las empresas dedicadas a la energía fotovoltaica en Ciudad Obregón, México, no cuentan con un sistema de inteligencia de negocio que les permita crear nuevas estrategias para obtener mayor ventaja competitiva. Para esto, se está trabajando con una empresa dedicada a proveer servicios de energía fotovoltaica quien será de las primeras en contar con esta tecnología en Ciudad Obregón.

Las empresas de este giro, brindan a los clientes soluciones solares para la reducción de dióxido de carbono en el ambiente y ahorro de energía mediante la auto suficiencia que permiten los sistemas fotovoltaicos siendo una tecnología emergente y en crecimiento. En ese sentido, los empresarios necesitan tomar decisiones que marcan el rumbo de la organización por lo que la información debe ser sólida y en tiempo real para que pueda cumplir sus objetivos empresariales de la mejor manera. El objetivo de este estudio es brindar una visión clara de los desafíos y oportunidades de aplicar la tecnología de almacenamiento de datos (datawarehouse) para la toma de decisiones de la alta gerencia en un PyME mexicana de la industria fotovoltaica.

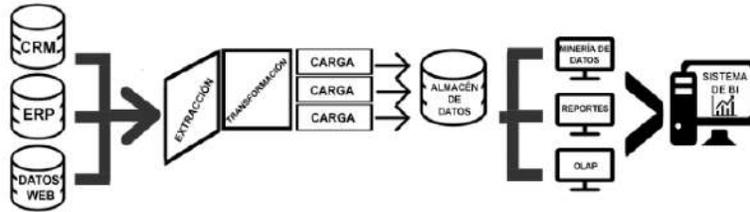


Figura 1. Esquema general de un sistema de inteligencia de negocios.

## 2 Marco teórico

La inteligencia de negocios es una fuente de información y conocimiento valioso para los responsables de la toma de decisiones al aprovechar una variedad de fuentes de datos, así como información estructurada y no estructurada. La información y los datos pueden residir dentro o fuera de la organización, también pueden obtenerse de múltiples fuentes y estructurarse de diferentes formas [8]. De esta manera, un almacén de datos

(datawarehouse) es un repositorio que almacena la información de forma estructurada que para su construcción, es necesario llevar a cabo el proceso de Extracción, Transformación y Carga denominado ETL. El proceso ETL es una parte de la integración de datos, cuya función completa el resultado de todo el desarrollo de la cohesión de aplicaciones y sistemas. Gracias a los procesos ETL es posible que cualquier organización mueva datos desde una o múltiples fuentes, formatee esos datos y los limpie, cuando sea necesario y los cargue en otro lugar como el almacén de datos, [9]. Una vez que el proceso ETL finaliza se carga la información al almacén de datos para ser usado por el sistema de inteligencia de negocios en el cual se mostrará información relacionada a los requerimientos solicitados por los empresarios.

### **3 Descripción del problema**

Hoy en día, las empresas hacen frente a la competencia globalizada donde la preparación científico-tecnológica y la capacidad de innovación del ser humano están haciendo la diferencia en esta época, presenciando rápidos desarrollos mediante el uso de la tecnología. Dentro de este contexto, tanto las grandes como las pequeñas empresas deben hacer frente al avance tecnológico para poder confrontar las amenazas generadas por la competitividad. En ese sentido, el soporte de toma de decisiones gerenciales es de vital importancia en una empresa en crecimiento, por lo que cada una de las decisiones que tome el empresario direcciona el rumbo al éxito o fracaso de la organización. Un sistema de BI no solo da el soporte necesario para que esta actividad sea sólida y segura, sino que además con la misma información se pueden llegar a desarrollar diferentes técnicas para un análisis más complejo sobre los datos de la organización. Para esto, se llevó a cabo un análisis de las empresas del giro fotovoltaico y de otros giros, en Ciudad Obregón, donde se identificaron desafíos y oportunidades al aplicar tecnología de almacenamientos de datos como apoyo en la toma de decisiones gerenciales, así como también, un análisis de requerimientos como inicio del diseño e implementación de un sistema de BI en una empresa del giro fotovoltaico.

### **4 Desafíos y oportunidades de aplicar tecnología de almacenamiento de datos en la toma de decisiones gerenciales.**

Una empresa se plantea objetivos medibles, claros y alcanzables pero a su vez desafiantes y coherentes con su misión y visión, con la finalidad de lograr un buen posicionamiento y perdurar en el mercado. Por tal motivo, para los empresarios es importante incrementar sus ventas, mantener e incrementar la fidelidad de clientes, tener mayor visibilidad en el mercado, mejorar su competitividad, entre otros.

En ese sentido, el apoyo de la tecnología es crucial para crecer y mejorar la toma de decisiones estratégicas que lleven al negocio a cumplir con su visión y misión. Para ello, los sistemas de BI, son una excelente opción que les permite contar con información en

tiempo real, conocer mejor a sus clientes, identificar patrones y tendencias, así como llevar a cabo reportes avanzados.

Es por ello, que este artículo introduce algunos desafíos y oportunidades referente a los sistemas de BI para la toma de decisiones estratégicas. Para ello, partimos de la identificación de algunas decisiones estratégicas que las PyMEs por lo regular se enfrentan como: administración de los recursos, inversión de nuevos productos o servicios, monitorear las ventas, monitorear la productividad de los empleados, fidelización de clientes, estrategias de mercadeo, entre otros.

En el caso de la empresa de estudio, el interés principal es en el área de ventas, donde interesa conocer clientes frecuentes, así como nichos de mercado desconocidos o en etapa de crecimiento como oportunidad a desarrollar estrategias, también conocer el número de ventas totales por sucursal, entre otros. Así mismo, la empresa se enfrenta a ciertos desafíos al momento de querer invertir en la implementación de un sistema que cuente con este tipo de tecnología, pero a su vez, el implementar estos tipos de sistemas en la organización representa oportunidades de crecimiento y posicionamiento en el mercado.

#### **4.1. Desafíos.**

##### **4.1.1 Grado de inversión de software.**

El costo de inversión en software dependerá de las funcionalidades que la empresa requiera. En la actualidad, se cuenta con una gran variedad de software para el desarrollo de aplicaciones de inteligencia de negocios tanto de acceso libre como de licencia. En el caso de las PyMEs regionales se conoce que actualmente en el estado de Sonora se cuenta con diez parques solares fotovoltaicos que se encuentran en constante incremento. Respecto al software en este tipo de empresas se sugiere comenzar con software libre o de licencia de bajo costo como PowerBI<sup>‡</sup>. Es importante mencionar que el número de usuarios es de vital importancia para la selección del software que se utilizará, ya que en algunos como PowerBI se paga por usuario, la metodología de Kimball es una excelente opción para una PyME regional, ya que su costo inicial es bajo y el tiempo de desarrollo es inferior [4], en comparación con otras metodologías [5].

##### **4.1.2 Infraestructura tecnológica.**

Por lo regular una PyME carece de información almacenada en bases de datos, o en el peor de los casos, no cuentan con información histórica de las operaciones diarias, registros de clientes, compras, etc. Por tal motivo, es un desafío para las PyMEs empezar a implementar estrategias para generar información histórica de tal forma que el sistema de BI pueda ser una realidad ya que muchas de las veces no cuentan con la infraestructura

---

tecnológica necesaria, y con ello, el desafío es mayor al incrementarse el costo de la inversión. En este caso, la metodología de Kimball [4] también se ajusta de manera conveniente a estas empresas, ya que el alcance puede ser a departamentos individuales con la tecnología existente de la empresa.

#### **4.1.3 Resistencia al cambio.**

El avance tecnológico muchas de las veces no va de la mano con la evolución de las organizaciones, tal es el caso de las PyMEs. La experiencia en este sentido, es que al principio la mayoría de las PyMEs desconocen la existencia de este tipo tecnología pero al conocer los beneficios, su interés es inmediato. Sin embargo, hay diversos factores de resistencia, entre los que destacan los cambios internos que eso conlleva, cambio de estrategias y de procesos, así como la resistencia por parte de los empleados quienes prefieren hacer las cosas como tradicionalmente lo hacen. Por lo que una buena metodología como la de HEFESTO [10] sería de gran ayuda, ya que involucra al usuario final en cada etapa para que tome decisiones respecto al comportamiento y funciones del datawarehouse.

### **4.2. Oportunidades.**

#### **4.2.1 Proporciona información clave para la toma de decisiones empresariales.**

Al llevar a cabo un análisis de requerimientos, se establece la importancia de las decisiones estratégicas que la empresa necesita llevar a cabo. De esa manera, se diseña el almacén de datos para que pueda responder a las consultas requeridas por los empresarios, la etapa de análisis de requerimientos la desglosan varios autores de metodologías, sin embargo el que más se centraliza en ello es HEFESTO [10] ya que los identifica con el fin de entender los objetivos de la organización y su estructura se adapta con facilidad y rapidez ante los cambios en el negocio.

#### **4.2.2 Permite una mayor flexibilidad y rapidez en el acceso a la información.**

El sistema permite llevar a cabo consultas avanzadas desde cualquier dispositivo electrónico (tabletas, teléfono inteligente, computadora, laptop) con conectividad a internet, mediante el cual se pueden generar reportes *ad-hoc*, gráficos personalizados que ayudarán a la alta gerencia contar con información en tiempo real para tomar decisiones tanto estratégicas como operativas. En ese sentido, la mayoría de las metodologías para el diseño del datawarehouse proveen de flexibilidad y rapidez al acceder a la información. Sin embargo, la metodología ágil propuesta por Corr [6], es la que cuenta con un mejor potencial de respuesta al cambio, brindando mayor satisfacción al cliente, a través de la entrega temprana y retroalimentación continua.

#### **4.2.3 Permite hacer planes de forma más efectiva.**

Al contar con información confiable, en tiempo real y analizada por el sistema, es posible que el empresario pueda llegar a conclusiones de comportamientos recurrentes ya sea en clientes, proveedores, ventas, etc., permitiéndole generar nuevas estrategias de forma efectiva y rápida, en vez de tomar decisiones por medio de la intuición y experiencia como ocurre actualmente en la mayoría de las PyMEs. De esta manera, con los entregables periódicos que se realizan por medio de la metodología ágil [6] se minimizan riesgos durante la realización del proyecto de manera colaborativa lo que incrementa la productividad y calidad mediante un seguimiento diario.

## 5 Conclusiones

Las empresas hoy en día se enfrentan a una competitividad cada vez mayor por lo que el buen uso de la tecnología como elemento impulsor de cambios, es crucial para sobresalir en el mercado y contar con procesos más eficientes. Por ello, se llevó a cabo un análisis sobre la situación actual de las empresas del giro fotovoltaico para conocer su percepción sobre la utilización de tecnología de almacenamiento de datos y conocer los principales desafíos a los que se enfrentan y las oportunidades que esta tecnología les ofrece. Se determinó que las empresas de este giro en Ciudad Obregón, no cuentan con esta tecnología y la mayoría toma decisiones gerenciales utilizando información desactualizada, incompleta, por intuición y experiencia. En este momento, se está trabajando con una empresa del giro fotovoltaico para el diseño e implementación de un sistema de inteligencia de negocios. Para ello, se obtuvieron los requerimientos del negocio, se acordaron indicadores relacionados al área de ventas para comenzar a trabajar en el diseño del almacén de datos. Finalmente, pese a los desafíos que las empresas enfrentan al adoptar este tipo de tecnología, hay empresas dispuestas a tomar retos para innovar, crecer y mejorar su competitividad.

## Referencias

1. Gutiérrez, A., Motz, R., Revello, B., & Silva, L.: Construcción de un sistema de apoyo a la toma de decisiones para el área gerencial del Hospital de Clínicas. In: Simposio de Informática en Salud-Jornadas Argentinas de Informática e Investigación Operativa, Buenos Aires (2001)
2. Lim, E. P., Chen, H., & Chen, G.: Business intelligence and analytics: Research directions. *ACM Transactions on Management Information Systems (TMIS)* vol. 3(4), pp. 17 (2013)
3. Vanegas Lago, E., Guerra Cantero, L. M.: Sistema de inteligencia de negocios para el apoyo al proceso de toma de decisiones. *Revista Ingeniería UC*, 20(3) (2013)
4. Kimball, R., Ross, M.: *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons, Indianapolis (2011)
5. Inmon, W. H.: *Building the data warehouse*. John Wiley & Sons, Indianapolis (2005)
6. Corr, L. Stagnitto, J.: *Agile Data Warehouse Design: Colaborative Dimensional Modeling, from Whiteboard to Star Schema*. DecisionOne Press, United Kingdom (2011)

Desafíos y oportunidades de aplicar tecnología de almacenamiento de datos como apoyo en la toma de decisiones gerenciales en una PyME de la industria fotovoltaica

7. March, S. T., Hevner, A. R.: Integrated decision support systems: A data warehousing perspective. *Decision Support Systems*. 43(3), 1031-1043 (2007)
8. Coronel, C., Morris, S., Rob, P.: *Bases de Datos, diseño, implementación y administración*. Cengage Learning Editores (2011)
9. Vassiliadis, P., Simitsis, A., Skiadopoulos, S.: Conceptual modeling for ETL processes. *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP*, pp.14-21 (2002)
10. Brizuela, E. I. L., Blanco, Y. C.: Metodologías para desarrollar Almacén de Datos. *Revista Arquitectura e Ingeniería*. 7(3), 3-12 (2013)

# Exploración de sentimientos asociados a sistemas de facturación electrónica

Laura Elena Cervantes, Luis A. Castro  
Departamento de Computación y Diseño  
Instituto Tecnológico de Sonora  
Ciudad Obregón, México  
laura.cervantes@potros.itson.edu.mx, luis.castro@acm.org

**Resumen**— Los sistemas para facturación electrónica en México se han tornado obligatorios para efectos fiscales. Existen numerosas empresas que se dedican a crearlos y comercializarlos, para quienes es importante conocer la manera en que se van a promocionar tales productos. El pago de impuestos y los procesos de facturación pueden ser temas sensibles para muchos, por lo que se requiere de diferentes métodos para entender mejor a usuarios de estos sistemas. Sin embargo, estos tópicos han sido rara vez estudiados. Este proyecto de tesis está guiado por la pregunta: ¿Cuáles son los sentimientos asociados a los sistemas de facturación electrónica? y propone el uso del método ZMET (Zaltman Metaphor Elicitation Technique), el cual, a través de un proceso de entrevistas genera los constructos que influyen en nuestra forma de pensar y actuar respecto a un producto, servicio, marca o tema en particular. En este caso, se busca explorar los pensamientos de una muestra de usuarios de sistemas de facturación electrónica en México con el fin de conocer lo que realmente sienten al utilizar estos sistemas para el diseño de campañas de mercadotecnia digital más personalizadas, con productos que se adecuen a resolver de manera satisfactoria las verdaderas necesidades de los clientes.

*Palabras clave*- ZMET, Percepción, Consumidor.

## I. INTRODUCCIÓN

En México, desde 2014, de acuerdo con una nueva disposición del Servicio de Administración Tributaria (SAT) de la Secretaría de Hacienda y Crédito Público (SHCP), la facturación se debe realizar de manera digital, lo cual ha favorecido la creación de un gran número de empresas dedicadas a la generación de facturación y timbrados fiscales [1]. La recaudación fiscal es un tema muy importante no solo en México, por lo que se han realizado diversos estudios alrededor del mundo para conocer el comportamiento de los contribuyentes [2, 3]. La literatura sugiere que el cumplimiento tributario está determinado por cinco factores generales: disuasión; normas (tanto personales como sociales); equidad y confianza (en la administración tributaria); complejidad del sistema tributario; y el papel del gobierno y el entorno económico más amplio [2]. Además, muestran que las identidades sociales de los contribuyentes existen y varían según los contextos situacionales y psicológicos en los que se encuentran los contribuyentes [4]. Como en muchos otros giros, en el ámbito de la facturación electrónica existen factores que afectan la utilización de los sistemas

de facturación electrónica. La utilización y comercialización de estos sistemas normalmente es afectada por la percepción de los ciudadanos respecto a los procesos de pago de impuestos y facturación. Sin embargo, hasta el momento no han identificado trabajos en México que sirvan de base para la generación de campañas de mercadeo efectivas para dichos sistemas.

En este trabajo se presentan los resultados de un estudio que se realizó con usuarios de sistemas de facturación electrónica para identificación de sentimientos asociados con estos procesos. Este estudio muestra resultados acerca de un análisis de constructos generados a partir de una muestra de usuarios de sistemas de facturación electrónica. Con base a los constructos generados en el análisis de cada una de las entrevistas, se puede decir que los usuarios de los sistemas de facturación en México encuentran la actividad como molesta y los sistemas que utilizan no son completamente del agrado de los usuarios. Dichos constructos se pueden realizar para realizar campañas de mercadeo o segmentación de mercado mucho más adecuada.

## II. MARCO TEÓRICO Y TRABAJOS RELACIONADOS

La segmentación de mercados tiene un rol clave en las estrategias de mercadotecnia al ser una herramienta que puede agrupar tanto a personas como organizaciones, ayuda a definir con precisión las necesidades y deseos de los clientes [5]. Philip Kotler define la segmentación como “una fase de reconocimiento en la que las empresas entienden que no se puede atender a todos los clientes con el mismo nivel de satisfacción” [6]. En Latinoamérica, se han realizado diversos estudios para conocer que estrategias de mercadeo y aplicación de segmentación de mercado utilizan las PyMES, [7, 8] los cuales concluyen que, en su mayoría, utilizan criterios demográficos simples para la selección del mercado meta. No obstante, en estudios aplicados a otros sectores como el turismo [9], se ha demostrado la importancia de una buena estrategia de segmentación donde se combinen diferentes técnicas para la obtención de resultados favorables para las empresas.

En [10], se habla de cómo la mercadotecnia y la investigación de mercadotecnia han sido basadas en conceptos de comportamiento del consumidor que se encuentran profundamente incrustados en nociones lineales de actividades de mercadotecnia y cómo con el incremento de marcos de referencia se han creado nuevas herramientas como ZMET que ayudan a los mercadólogos a entender la interacción entre actitudes, emociones y decisiones con respecto a marca/producto.

Christensen & Olson [11] demuestran el poder de la técnica de obtención de metáforas de Zaltman (ZMET, *Zaltman Metaphor Elicitation Technique*) para obtener esa percepción del consumidor al obtener primero y luego mapear las estructuras de conocimiento de los consumidores. Además, se proporciona ilustraciones de cómo se puede usar ZMET para crear un mapa cognitivo colectivo para un grupo de consumidores, y cómo los datos de ZMET se pueden mapear de diferentes maneras para brindar una mejor comprensión de las estructuras de conocimiento de los productos de los consumidores.

En [12] se plantea que, a pesar del crecimiento en la adopción de tecnología por parte de los gobiernos, la evaluación de la calidad en los servicios públicos prestados electrónicamente ha sido relativamente deficiente. Investigaciones anteriores sobre sistemas de información han identificado la calidad de la información y el servicio, el uso del sistema, la diversión y la calidad del diseño del sistema como factores críticos para la entrega exitosa del servicio en el comercio electrónico.

### III. OBJETIVOS

A través de esta investigación se pretende conocer las percepciones de los usuarios de sistemas de facturación y timbrado a través de herramientas de mercadotecnia emocional como el método ZMET para así poder definir segmentos de clientes para la futura elaboración de planes estratégicos de mercadotecnia.

### IV. MÉTODOS

En esta sección se describen los métodos utilizados en este trabajo.

#### A. Participantes

Se reclutaron 15 participantes (7 mujeres, 8 hombres) económicamente activos, pertenecientes a cualquier giro comercial que realicen el proceso de facturación fiscal electrónica por sí mismos sin importar si la plataforma utilizada era gratuita o de paga.

#### B. Recopilación y Análisis de Datos

Se utilizó el método ZMET, el cual permite sacar a la superficie los constructos que influyen en nuestra forma de pensar y actuar respecto a un producto, servicio, marca o tema en particular de un segmento de mercado específico [13, 14]. Con esta técnica se puede obtener información no factible de identificar utilizando las herramientas más tradicionales como grupos focos o encuestas ya que en las entrevistas ZMET se exponen aspectos no considerados como primordiales por los investigadores o incluso sentimientos que los propios entrevistados no conocían. Esta herramienta consta de ocho pasos e implica la recolección de imágenes que ayudan a la realización de metáforas y constructos para determinar sus pensamientos más profundos. El orden de los pasos consiste en: (1) relato de imagen, (2) imágenes perdidas, (3) descarte de imágenes, (4) provocación de constructos, (5) elaboración de constructos, (6) imágenes sensoriales, (7) montaje y (8) viñeta. Los investigadores pueden usar los constructos y sus relaciones para generar

diagramas que representen los modelos mentales y sus conceptos o constructos interrelacionados de un individuo o la muestra bajo estudio. [15].

#### C. Procedimiento de la Investigación

La recolección de los datos comenzó pidiéndole a cada participante que buscara imágenes que representaran sus pensamientos y sentimientos sobre la facturación digital. Después de dar las instrucciones iniciales, se programó una entrevista personal una semana más tarde.

Las imágenes presentadas por los participantes fueron metáforas que ayudaron al investigador a comprender los procesos cognitivos y sirvieron como enfoque para la entrevista. Al darles la libertad de elegir las imágenes desde un buscador, ellos mismos quedan a cargo de los estímulos utilizados con los que se guía la entrevista. Este tipo de entrevista le permite al investigador aprender sobre los pensamientos iniciales o lógicos de los participantes (los cuales se pueden obtener mediante encuestas o grupos focales), además, se aprende sobre sus pensamientos cognitivos más profundos.

#### D. Proceso de Análisis de los Datos

##### 1) Paso 1

Los participantes deben describir como cada una de las imágenes que eligieron se encuentra relacionada con el tema estudiado (en este caso, la facturación electrónica). La mayoría de los participantes no tenían una historia definida sobre las imágenes que recopilaron para la entrevista.

##### 2) Paso 2

Se le preguntó a los participantes si había alguna imagen que había querido utilizar para la entrevista, pero no había podido encontrarla. De ser así, el investigador se encargaba de buscar la imagen deseada, o en su defecto, dibujarla con ayuda del participante para que cumpliera con las expectativas de este. Este punto es sumamente importante ya que todas las imágenes deben ser expuestas.

##### 3) Paso 3

Consistió en realizar un descarte de imágenes. Esto en caso de que el participante hubiera traído consigo la imagen de una factura, la cual fue la única restricción que se indicó el día de las instrucciones previas.

##### 4) Paso 4

Se llevó a cabo la provocación de constructos; el investigador agrupa las imágenes de diferentes formas y le pide al participante que indique porque esas imágenes se parecen o que tienen en común, en todo caso, correspondiendo al tema de estudio que es la facturación electrónica. Si el participante no puede hacer ningún tipo de distinción entre las imágenes, la entrevista avanza al siguiente paso.

##### 5) Paso 5

Se llevó a cabo la realización de metáforas. Se le solicita al participante que cuente una historia con las imágenes que recolectó y que se incluya a el mismo en dicha historia.

##### 6) Paso 6

Consistió en solicitar al participante que provoque otras metáforas basadas en otros sentidos distintos a la vista, como

olores, sonidos y sentimientos. En este paso el participante debe volver a contar la historia, agregando los nuevos sentimientos y pensamientos.

7) Paso 7

Consistió en dejar que el participante realice una viñeta de la historia que acaba de contar con las imágenes que recolectó para la entrevista. De ser necesario, el investigador puede ayudar al participante a ordenar las imágenes para la realización de la viñeta.

8) Paso 8

El investigador revisa todos los constructos elaborados a lo largo de la entrevista y realiza una nueva viñeta con ayuda de herramientas digitales para plasmar de manera gráfica el producto final de la entrevista.

E. Análisis de los datos

Al terminar la entrevista, el investigador revisó toda la información recolectada para encontrar constructos clave o tópicos específicos. Los constructos son creados de manera abstracta y representan ideas comunes, temas y conceptos expresados por cada uno de los participantes del estudio. Estos constructos representan aspectos importantes del mapa mental de los participantes como emociones, sentimientos, valores, ideas, entre otros [15].

Siguiendo la recomendación de [15] para que al menos una cuarta parte de los participantes mencionaran un concepto para considerarlo significativo, se excluyó cualquier constructo que no fuera mencionado por al menos tres personas. Una vez establecido esto, se formularon dos niveles para la generación de constructos y relaciones acordadas por la mayoría de los participantes. El nivel secundario de importancia se realizó para desarrollar los mapas de constructos al igual que como un medio para utilizar los constructos mencionados por pocos participantes. Los mapas se realizaron con líneas continuas y punteadas. Las líneas continuas representaban los constructos generados por cinco o más participantes y las líneas punteadas se refieren a los constructos evidenciados por tres o cuatro de los participantes.

El mapa de consenso simboliza, en promedio el 80% de todos los constructos mencionados por los participantes y se representa mediante un diagrama de flujo que muestra los vínculos entre esos constructos. Las conexiones directas e indirectas (líneas continuas y punteadas) definen los procesos de pensamientos sobre como una idea conduce a otra.

Por ejemplo, si siete participantes mencionaron el constructo Sistema (refiriéndose a los sistemas de facturación electrónica), se marcó un número cinco (5) junto a este constructo en la lista. Después de marcar las frecuencias de los constructos, las relaciones que tuvieron una frecuencia mayor a dos se agregaron al mapa general de constructos. Cuando se mencionaba un solo constructo en dos relaciones diferentes, se creaba una cadena.

V. RESULTADOS

La Tabla I expone los constructos que son únicos para cada muestra dividido entre hombres y mujeres, así como los constructos compartidos por cada una de estas. Es necesario

recordar que cada uno de estos constructos se tomaron con base a los diálogos reales de los participantes durante la fase de entrevistas de esta investigación.

TABLE I. CONSTRUCTOS MENCIONADOS POR LAS MUESTRAS

Constructos comunes entre las dos muestras	Constructos específicos entre mujeres	Constructos específicos entre hombres
Sistema	Errores	Ventas
Cancelación	Pereza	Tiempo
Pagos	Enojo	Dinero
Cientes		Gente
Estrés		Correos
SAT		

La parte más importante para el desarrollo del mapa de constructos es la relación de estos ya que a partir de este punto el entrevistador se informa sobre como los diferentes conceptos se asocian y afectan entre sí. Si bien, se puede decir que estas relaciones pueden representarse como vínculos causales claros, es sumamente importante recordar que estas solo indican las percepciones de los participantes. Estos nexos se muestran en la Tabla II.

TABLE II. RELACIONES DE CONSTRUCTOS MENCIONADOS POR LAS MUESTRAS

Relaciones de constructos primarios comunes entre las dos muestras	Relaciones de constructos primarios entre mujeres	Relaciones de constructos primarios entre hombres
<b>Estrés → Sistema</b> <b>Cancelación</b> <b>→ SAT</b>	<b>Sistema → Pereza</b> <b>Cancelación</b> → <b>Enojo</b>	<b>Sistema → Ventas</b> <b>Pagos → Dinero</b> <b>Tiempo → Estrés</b>
Relaciones de constructos secundarios comunes entre las dos muestras	Relaciones de constructos secundarios entre mujeres	Relaciones de constructos secundarios entre hombres
<b>Estrés → SAT</b>	<b>Enojo → SAT</b> <b>SAT → Errores</b>	<b>Estrés → Gente</b>

A partir de los constructos comunes y las relaciones entre los constructos, se elaboraron dos mapas consensuales para cada una de las muestras las cuales muestran de forma gráfica dichas relaciones. Para diferenciar las relaciones, se utilizó una línea continua para los constructos primarios y líneas punteadas para los constructos secundarios.

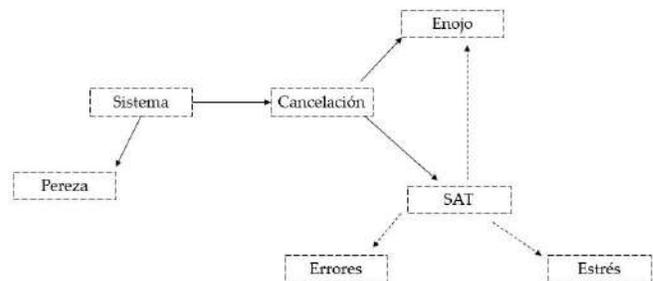


Figure 1. Mapa de consensos entre las mujeres

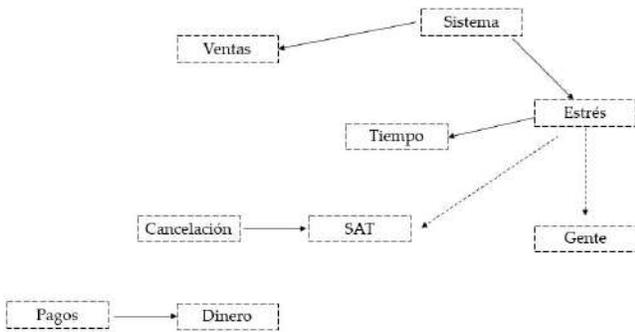


Figure 2. Mapa de consensos entre los hombres

Siguiendo las Figuras 1 y Figura 2, se puede decir que los usuarios de los sistemas de facturación en México encuentran la actividad como molesta y los sistemas que utilizan no son completamente del agrado de los usuarios. La muestra conformada por los hombres tuvo un mayor número de constructos específicos, así como las relaciones entre primarias entre estos. Sin embargo, el número de constructos comunes fue mayor lo cual indica que los pensamientos entre hombres y mujeres no consta de muchas diferencias entre sí. La mayoría de los constructos entre las muestras tiene que ver con sentimientos negativos, donde el *SAT* es el que más predominó a lo largo del estudio, seguido de *estrés*. El constructo *pereza* solo fue mencionado por las mujeres y representó la falta de ganas de llevar a cabo la generación de facturas puesto que en diversas ocasiones la muestra mencionó que preferían posponer el mayor tiempo posible la realización de esta. Por otra parte, el constructo *dinero* solo fue mencionado por los hombres y lo asocian al hecho de que, al generar una factura van a obtener como resultado un ingreso monetario en el corto plazo. En los constructos específicos también es importante señalar que las mujeres expresaron las emociones que sentían al momento de realizar la facturación electrónica (enojo, pereza) y los hombres se enfocaron más en la actividad a realizar (ventas, dinero, gente, correo, correos). Este es otro punto importante por considerar, ya que, si bien en los constructos comunes se creó el constructo estrés, este fue la única emoción presentada de manera explícita por los hombres.

Además, en las relaciones de los constructos entre las mujeres se puede observar como una acción a desarrollar en el sistema conlleva un sentimiento (Ej. Cancelación → enojo), mientras que en la mayoría de las relaciones de constructos entre hombres una acción desencadena en otra acción (Ej. Sistema → Ventas)

## VI. CONCLUSIONES

Este estudio constituyó una investigación acerca de un análisis de constructos generados a partir de una muestra de usuarios de sistemas de facturación electrónica. Los resultados de los constructos pueden ser de utilidad para la elaboración de campañas de mercadotecnia dirigidas a este sector, utilizando un enfoque más empático con el usuario y que incremente la efectividad de estas campañas. Sin

embargo, es recomendable que las empresas pertenecientes a este giro utilicen herramientas cuantitativas como la minería de datos para analizar a sus clientes actuales, segmentarlos y utilizar el método ZMET planteado en esta investigación para aumentar el grado de detalle de cada segmento y que esto se traduzca en planes estratégicos de mercadotecnia efectivos.

## REFERENCIAS

- [1] SHCP. (2014). "El Código Fiscal de la Federación y los servicios Electrónicos". Recuperado de: [http://www.tfjfa.gob.mx/media/media/index\\_files/miscelanea/Lic%20Fernando%20Martinez%20Coss.pdf](http://www.tfjfa.gob.mx/media/media/index_files/miscelanea/Lic%20Fernando%20Martinez%20Coss.pdf)
- [2] Walsh, Keith. (2012). The Economic and Social Review, Vol. 43, No. 3, Autumn, 2012, pp. 451–475
- [3] Saha, P., Nath, A. K., & Salehi-Sangari, E. (2012). Evaluation of government e-tax websites: an information quality and system quality approach. Transforming Government: People, Process and Policy, 6(3), 300–321.
- [4] Braithwaite, Valerie. (2003). Taxing Democracy.
- [5] Zárraga, L., Sandoval, V., Corona, E. (2013). Estudio De Caso: Análisis De La Aplicación De La Segmentación De Mercado Como Estrategia Para Las Pequeñas Empresas (Case Study: Analysis of Implementation of Market Segmentation as a Strategy for Small Business). Revista Internacional Administración & Finanzas, Vol. 6 (5) pp. 109-119 (2013).
- [6] Kotler, P. and Armstrong, G. (2012). Principles of marketing. 14th ed. Boston: Pearson Prentice Hall.
- [7] Alcaide, J., Bernués, S., Díaz-Aroca, E., Espinosa, R. Muñiz, R. & Smith, C. (2013) Marketing y PYMES, Las principales claves de marketing en la pequeña y mediana empresa. ISBN: 978-84-6957487-4.
- [8] Morales Ramírez, Mariher Yary, Estrategias de mercadeo de las PYMES del sector confección de la región zuliana. Multiciencias [en línea] 2006, 6 (marzo)
- [9] Wade & Eagles, (2010). The Use of Importance-Performance Analysis and Market Segmentation for Tourism Management in Parks and Protected Areas: An Application to Tanzania's National Parks. Journal of Ecotourism. Vol 2.
- [10] Christensen, G. L., & Olson, J. C. (2002). Mapping consumers' mental models with ZMET. Psychology and Marketing, 19(6), 477–501. doi:10.1002/mar.10021
- [11] Zárraga, L., Molina, V., Sandoval, E. (2013). "Caso de estudio: análisis de la aplicación de la segmentación de mercado como estrategia para las pequeñas empresas". Revista Internacional Administración & Finanzas, Vol. 6 (5) pp. 109-119.
- [12] Zaltman, G., & Higie, R.A. (1993). Seeing the voice of the customer: The zaltman 64 metaphor elicitation technique. Marketing Science Institute. Report Number: 93-114.
- [13] Zaltman, G., & Coulter, R.A. (1995). Seeing the voice of the customer: metaphor-based advertising research. Journal of Advertising Research, 35(4), 35-51.
- [14] Zaltman, G. (1996). Metaphorically speaking. Marketing Research, 8, 13-22.
- [15] Zaltman, G. (1997). Rethinking market research: Putting people back in. Journal of Marketing Research, 34(4), 424-437.

# AVANCES DE INVESTIGACIÓN EN INGENIERÍA EN EL ESTADO DE SONORA

**AVIIES 2019**

**Año 5, número 1**

**ISSN: 2448-7473**

**Responsable de la Edición del volumen**  
**Dr. Oscar Mario Rodríguez Elías**

**Colaboradores en la edición:**  
**Dr. Mario Barceló Valenzuela**  
**Dr. Alonso Pérez Soltero**  
**Dr. Ramón René Palacio Cinco**  
**Dr. Joaquín Cortez González**  
**M. C. Sonia Regina Meneses Mendoza**



Difusión vía red de cómputo

Noviembre del 2019



**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA



**TECNOLÓGICO  
NACIONAL DE MÉXICO**

**AVANCES DE INVESTIGACIÓN EN INGENIERÍA EN EL ESTADO DE SONORA, Año 5 Núm 1, octubre de 2019, es una revista anual, publicada y editada por el Tecnológico Nacional de México dependiente de la Secretaría de Educación Pública, a través del Instituto Tecnológico de Hermosillo, por la División de Estudios de Posgrado e Investigación, con domicilio en Av. Universidad Núm. 1200, Colonia Xoco, Delegación Benito Juárez, C.P. 03330, Ciudad de México, Tel. 5536017500, Correo electrónico: d\_vinculacion@tecnm.mx. Editor Responsable: Dr. Oscar Mario Rodríguez Elias. Reserva de derechos al uso exclusivo No. 04-2015-101310132700-203, con ISSN: 2448-7473, ambos otorgados por el Instituto Nacional del Derecho de Autor.**

**Responsables de la última actualización de este volumen: Dr. Oscar Mario Rodríguez Elias, en colaboración con, Dr. Mario Barceló Valenzuela, Dr. Alonso Pérez Soltero, Dr. Ramón René Palacio Cinco y Dr. Joaquín Cortez González, M.C. Sonia Regina Meneses Mendoza, en las instalaciones del Instituto Tecnológico de Hermosillo, Ave. Tecnológico y Periférico Poniente SN C.P. 83170, Colonia Sahuaro, Hermosillo, Sonora, México. Fecha de término de impresión, 31 de octubre de 2019.**

**Su objetivo principal es difundir los avances en investigación a nivel posgrado y licenciatura en diversas áreas de la ingeniería, realizados durante el lapso de un año, en las instituciones participantes de educación superior del estado de Sonora.**

**Los artículos son sometidos a un proceso de arbitraje, por lo que su contenido es responsabilidad exclusiva de sus autores, y no representa necesariamente el punto de vista de la institución.**

**Queda estrictamente prohibida la reproducción total o parcial de los contenidos e imágenes de la publicación sin previa autorización del Instituto Tecnológico de Hermosillo.**

**Enlace de acceso: [www.aviies.ith.mx](http://www.aviies.ith.mx)**

# Índice de Contenido

## A.- Resultados de Investigación

### Ciencias de la Computación .....1

<b>Análisis, Diagnóstico y Propuesta de Mejora para el Proceso de Compras de un Hospital Privado.</b> Cristel Marisela Fragoso Pacheco, Ramon René Palacio Cinco, Mario Alberto Nuñez Luna, Jorge Guadalupe Mendoza León.....	1
<b>Diseño de un Sistema en Línea para la Medición Automática de Habilidades Blandas.</b> Manuel Guerrero Garcia, Oscar M. Rodríguez Elías, María T. Serna Encinas, Abelardo Mancinas Gonzalez. ....	9
<b>Arquitectónico para un Sistema de Reconocimiento de Patrones en Bases de Datos Sobre Adicciones.</b> Edgar G. Estrada Rios, María T. Serna Encinas, Cesar E. Rose Gómez. Diseño.....	17
<b>Sistema Multi-Agente de Reconocimiento de Frases en LSM Utilizando CBR.</b> César René Martínez Aguirre, Ana Luisa Millán Castro, Juan Pablo Soto Barrera, César Enrique Rose Gómez, Abelardo Mancinas González. ....	23
<b>Arquitectura Propuesta para el Análisis de la Instrumentación Didáctica de Asignaturas de Educación Superior Tecnológica usando un Modelo de Conocimiento y Procesamiento de Lenguaje Natural.</b> Daniel Alfredo Hernandez Carrasco, Cesar Enrique Rose Gomez, Samuel Gonzalez López, Abelardo Mancinas Gonzalez, Ana Luisa Millan Castro. ....	30
<b>Diseño arquitectónico de un sistema para la detección de un carcinoma en biopsias de mama.</b> José Reynaldo Sánchez Quintero, María Trinidad Serna Encinas, César Enrique Rose Gómez, Minor Raúl Cordero Bautista, Fernando Javier Carrasco Guigón. ....	37
<b>Traductor de texto en español a texto en glosa LSM.</b> Juan Carlos Hernández-Cruz, César Enrique Rose- Gómez, Samuel González López, Ana Luisa Millán Castro, María Trinidad Serna Encinas.....	45

### Eléctrica y electrónica .....51

<b>Desarrollo de un API Rest Full para monitoreo del consumo de la energía eléctrica mediante una aplicación móvil en una instalación residencial con sistema fotovoltaico interconectado a la red.</b> Víctor Alfonso Tánori Ruíz, Fredy Alberto Hernández Aguirre, José Antonio Hoyo Montaña, José Manuel Chávez, Jesús Manuel Tarín Fontes. ....	51
<b>Sistema de Monitoreo de Posición de Recepción de Haz Láser de un Gimbal.</b> Eduardo Romero, Juan Ayala, Arturo Arvizu , Joel Santos. ....	55

### Ingeniería industrial.....61

<b>Factores de riesgo ergonómico ocasionados por manejo manual de materiales en operaciones logísticas en la industria automotriz.</b> Debbie Yemileth Vásquez Gómez, Javier Enrique De la Vega Bustillos.....	61
<b>Análisis de los tiempos de paro de producción en prensa de estampado.</b> Gilberto Orrantia Daniel, Jaime Sánchez Leal, Gloria María Velázquez Quijada, Jorge de la Riva Rodríguez, Manuel Rodríguez Medina. ....	67
<b>Análisis por medio de volumen finito y elementos finitos de un seguidor solar.</b> Luis Álvarez, Víctor Manuel Herrera.....	75
<b>Cadena de suministro: Control interno por la satisfacción del cliente.</b> Daniel Antonio Torres Coronado, Gil Arturo Quijano Vega. ....	83

### Ingeniería mecánica y mecatrónica .....90

**Diseño y construcción de un prototipo de órtesis mecatrónica auxiliar en el proceso de rehabilitación de mano para pacientes que sufrieron una ECV.** David Sotelo Valencia, Carlos Alberto Pereyda Pierre, Flor Ramírez Torres, Eliel Eduardo Montijo Valenzuela, Aureliano Cerón Franco. .... 90

## **B.- Avances de investigación**

### **Computación.....97**

<b>Propuesta para la evaluación plataformas Matlab y ROS en la implementación de algoritmos de control en vehículos aéreos no tripulados.</b> Julio C. Montoya-Morales, Guillermo Valencia-Palomo, Rafael A. Galaz-Bustamante, Rosalia C. Gutiérrez-Urquidez, María Eusebia Guerrero-Sánchez, Omar Hernández-González. ....	97
<b>Seguimiento de comportamientos de riesgo en jóvenes universitarios a través de publicaciones en Instagram.</b> Ivan Encinas, Luis A. Castro. ....	101
<b>Detección de factores asociados a la deserción de personal operativo mediante técnicas de minería de datos.</b> Abigail Márquez Hermosillo, Luis Felipe Rodríguez Torres, Guillermo Mario Arturo Salazar Lugo. ....	104
<b>Exploración de sentimientos asociados a sistemas de facturación electrónica.</b> Laura Elena Cervantes, Luis A. Castro.....	108
<b>Propuesta de rediseño de proceso en centros geriátricos mediante la implementación de tecnologías para el monitoreo del paciente.</b> Alma Leticia Chavez Quintero, Gilberto Borrego, Laura Elena Cervantes García, Luis Felipe Rodríguez Torres .....	112
<b>Diseño de un tablero de control para la toma de decisiones de productores de limón.</b> Berenice Mancilla Rojas, Luis A. Castro. ....	116
<b>Implementación de un sistema de almacenamiento de datos agroclimáticos para la predicción de cosecha de trigo.</b> María Monserrat Torres, Luis A. Castro. ....	120
<b>Victor Manuel Moreno García, Ana Luisa Millán Castro, Marcela Patricia Vázquez Arquitectura propuesta para sistema de recomendación de ejercicios a niños con dislexia mediante algoritmos de machine learning.</b> Valenzuela, María Trinidad Serna Encinas, César Enrique Rose Gómez. ....	124
<b>Desarrollo de una aplicación móvil, para medir los niveles de inteligencia de niños y niñas del sur de Sonora.</b> José de Jesús Soto Padilla, Ramón Rene Palacio Cinco, Carlos Jesús Hinojosa Rodríguez, Gilberto Manuel Córdova Cárdenas. ....	128
<b>Arquitectura propuesta para un Módulo Recomendador de un Sistema de Aprendizaje Adaptativo para la formación de docentes de Educación Superior.</b> Mirey Rocío García Mora, Abelardo Mancías González, César Enrique Rose Gómez, Oscar Mario Rodríguez Elías.....	132
<b>Técnicas de minería de datos para la gestión de abastecimiento de las pequeñas y medianas empresas (PyMEs).</b> Christopher Louis Vega Ruiz, Sonia Regina Meneses Mendoza, Oscar Mario Rodríguez Elías, César Enrique Rose Gómez. ....	136
<b>Desarrollo de un sistema de visión artificial para la evaluación de calidad de conectores en bolsas de aire automotrices.</b> Eduardo Rodarte Leyva, Victor Hugo Benítez Baltazar. ....	140
<b>Propuesta de Refinamiento de un Algoritmo de Minería de Datos para Detección de Pacientes.</b> Omar Fernando García Mora, Federico Miguel Cirett Galán, Raquel Torres Peralta. ....	143
<b>Arquitecturas de Redes Neuronales Convolucionales para la Clasificación de Imágenes en Celdas de Flotación.</b> Nelsón Romero García, Sonia Regina Meneses Mendoza, Oscar Mario Rodríguez Elías, César Enrique Rose Gómez. ....	146
<b>Propuesta de una plataforma de soporte a la capacitación en línea para la realización de estudios clínicos.</b> Maria Alejandra García Bayona, Oscar Mario Rodríguez Elías, Sonia Regina Meneses Mendoza, Hazael Gómez Encinas, Raúl Eduardo Rodríguez Ibañez.....	150
<b>Metodología para medir el logro del perfil de conocimiento de egresados de carreras de cómputo a través de un sistema valorador de perfiles de conocimiento.</b> Abraham Duarte Ruiz, Oscar Mario Rodríguez Elías, César Enrique Rose Gómez, Sonia Regina Meneses Mendoza.....	154

# DetECCIÓN DE FACTORES ASOCIADOS A LA DESERCIÓN DE PERSONAL OPERATIVO MEDIANTE TÉCNICAS DE MINERÍA DE DATOS

Abigail Márquez Hermosillo, Luis-Felipe Rodríguez, Guillermo Mario Arturo Salazar Lugo  
Maestría en Tecnologías de la Información para los Negocios, ITSON, Cd. Obregón, Sonora, México  
mha.9600@gmail.com, luis.rodriguez@itson.edu.mx, guillermo.salazar@itson.edu.mx

**Resumen** – Actualmente, las empresas se enfrentan a una problemática generalizada para retener el recurso humano habilidoso y capacitado para lograr sus objetivos empresariales. Esto plantea que se busquen mejores soluciones al problema de la deserción laboral, pues esta problemática impacta de manera directa a las empresas en cuanto a costos de selección, contratación y capacitación. Es por ello que en el área recursos humanos se ha comenzado a incorporar soluciones tecnológicas encaminadas a identificar los motivos por los cuales los empleados desertan de su trabajo. En la presente investigación, se tomará el caso de la empresa Malumex, una empresa outsourcing, que brinda servicios a empresas manufactureras en el parque industrial de Cd. Obregón, cuyos índices de rotación de personal operativo alcanzaron el 71.16% en el primer trimestre del 2019. En este artículo, se presenta el análisis de la situación de la empresa y la propuesta de la aplicación de la metodología CRISP-DM con el objetivo de identificar los perfiles de empleado susceptibles a la deserción temprana en puestos operativos.

**Palabras Clave:** Minería de datos, Deserción laboral, Rotación de personal.

## I. INTRODUCCIÓN

En un ambiente donde las empresas buscan la competitividad, contar con el recurso humano adecuado para alcanzar sus objetivos estratégicos y lograr la permanencia de este recurso dentro de sus equipos representa un gran reto. Es por ello que una alta tasa de rotación de personal es indeseable, lo cual lleva a carencias de personal habilidoso o entrenado, pérdida de productividad o los altos costos que se tienen que pagar en el reclutamiento y el entrenamiento [1].

Según Chiavenato [2], la rotación es “*el intercambio de personas entre la organización y el ambiente definido por el volumen de personas que ingresan y que salen de la organización*”. La rotación puede estar asociada a diversos factores como la edad, sexo y experiencia laboral de los trabajadores, así como otros factores referentes al clima y ambiente laboral de la organización. Identificar estos factores permite establecer soluciones más acertadas en relación a la contratación de recurso humano, ya que este tipo de datos sobre el comportamiento pasado de los trabajadores pueden ser analizados para predecir comportamientos futuros [3].

La incorporación de tecnologías de la información a las empresas ha tenido un auge en los últimos años, siendo el área de recursos humanos beneficiada por estas tendencias. Parte de las soluciones que se implementan en esta área, son guiadas hacia combatir y prevenir la deserción laboral mediante técnicas de minería de datos y modelos de ayuda en

el descubrimiento del comportamiento de los empleados y sus razones detrás de mismo [4]. Este tipo de técnicas ayudan a obtener conocimiento a partir de los datos de los empleados y así tomar mejores decisiones dentro del área de recursos humanos [4], [5].

Actualmente en Cd. Obregón se registra un incremento en los índices de rotación de personal en las empresas [6], lo que genera preocupación a los empresarios y particularmente a los responsables de recursos humanos. Esto los lleva a incorporar soluciones tecnológicas en el área, como lo es la inteligencia artificial y la minería de datos, encaminadas a descubrir patrones y establecer reglas dentro de los grandes volúmenes de datos [7].

Existen sectores que son más afectados por la rotación laboral, como son las industrias maquiladoras y las prestadoras de servicio de personal, organizaciones que tienen su principal activo en los recursos humanos y dependen de la estabilidad de su personal para alcanzar tanto objetivos como metas propuestas [1]. Tal es el caso de Malumex, una empresa regional de outsourcing especializada en contrataciones de nivel operativo. Malumex ha presentado un incremento en sus índices de rotación de personal desde el año 2016, medida utilizada para medir la deserción en proporción a la contratación de empleados. Al analizar esta medida, Malumex tuvo un índice de rotación de 71.16% durante el primer trimestre del 2019, lo cual le genera un incremento en los costos de reclutamiento, contratación y separación de la relación laboral del personal. Además se han tenido efectos negativos en el cumplimiento de contrato con sus clientes debido a estos altos índices de rotación.

## II. PLANTEAMIENTO DEL PROBLEMA

La empresa Malumex S.A. de C.V., se dedica a brindar servicios de subcontratación de recursos humanos para empresas maquiladoras. Malumex presenta un incremento de sus índices de deserción del personal que se incorpora en sus esquemas de subcontratación, especialmente en aquella que se da de manera temprana, es decir, antes de cumplir con el contrato inicial de tres meses. Dicho panorama le ha generado deficiencias en la prestación de sus servicios al incurrir en incumplimientos de contratos debido a la falta de personal. Lo anterior, lleva a que la empresa tenga sanciones por parte de sus clientes y sus indicadores internos de rotación de personal se vean afectados. Un ejemplo de los efectos de esta problemática sucedió en el año 2018, cuando al no mantener la plantilla mínima de empleados, se tuvo que terminar el

contrato con un cliente, lo que condujo a liquidaciones totales para los empleados y pago de sanciones con el cliente.

El aumento de los índices de rotación de personal operativo es un foco rojo para la gerencia de Malumex, por lo que se busca siempre reducirlo y minimizar el impacto. Sin embargo, esta situación se complica durante las temporadas donde aumenta la producción de sus clientes, pues al incrementarse los requerimientos de personal en lo que llaman de forma masiva, los procesos de reclutamiento y selección tienden a realizarse con más prisa y menores mecanismos de control. Esto desencadena en un incremento de los indicadores de deserción laboral.

La pregunta que se espera resolver en esta investigación es la siguiente: *¿Cuál es el perfil de empleado de puesto operativo susceptible a la deserción temprana en la empresa Malumex?*

### III. OBJETIVO

El objetivo de este proyecto es identificar un perfil de empleado susceptible a la deserción temprana en puestos operativos mediante técnicas de minería de datos, a través de la identificación de los factores principales que influyen en la deserción laboral.

### IV. METODOLOGÍA

Se propone aplicar la metodología *Cross Industry Standard Process for Data Mining* (Proceso estándar de la industria para la minería de datos en español) o CRISP-DM. La metodología, consiste en un ciclo de seis fases para el descubrimiento del conocimiento, que inicia con la comprensión del negocio, los objetivos a lograr, los datos a utilizar y termina con el despliegue del modelo seleccionado para alcanzar los objetivos planteados para el proyecto de minería de datos.

CRISP-DM se eligió a partir del objetivo de plantear una solución utilizando técnicas de minería de datos y por ser un referente en el desarrollo de proyectos de este tipo, pues proporciona a cualquier persona, desde principiantes hasta expertos en el tema, un plan completo para llevar a cabo proyectos de minería de datos [8].

Esta metodología, como se observa en la Figura 1, es un proceso cíclico e iterativo, donde las fases se pueden ir alternando y repitiendo a fin de obtener una mejor comprensión de los datos. En particular, se plantea desarrollar el proyecto aplicando las siguientes fases:

1. **Comprensión del Negocio.** En esta primera fase se comprenderá la situación que enfrenta Malumex, sus objetivos como empresa, la situación organizacional, la problemática a través de entrevistas y visitas a la empresa, la generación de un proyecto de minería de datos y los objetivos del mismo.

**Comprensión de Datos.** Se analizará la base de datos de los empleados de Malumex alojada en el sistema SuperNOMINA. Se incluirán, además, campos con información adicional, que han de ser capturados directamente de los expedientes de los empleados y ex

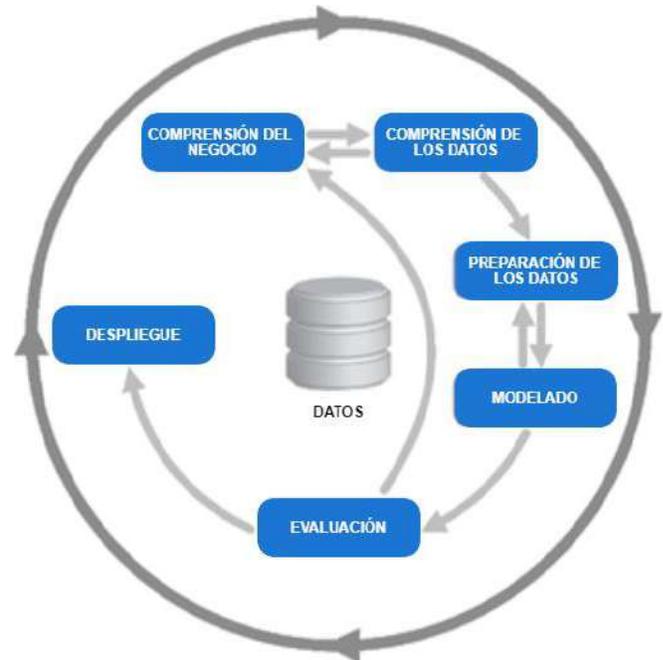


Figura 1. Ciclo de la Minería de Datos [10]

empleados de la empresa. Se presentará un análisis exploratorio descriptivo mediante SPSS y MS Excel, así como un informe de calidad de los datos.

2. **Preparación de Datos.** Se ejecutará un proceso de pre-procesamiento de los datos mediante el cual se pretende preparar la información y transformarla de manera que se asegure su confiabilidad. Este proceso involucra técnicas como limpieza, integración, transformación y reducción de datos [9]. Este proceso se llevará a cabo mediante el uso de herramientas como SPSS y Rapidminer, con lo cual se ejecutará la limpieza de datos, seleccionando cuáles son útiles, la exclusión de variables que no aportan significancia a la investigación.

3. **Modelado.** Se preparará el modelado de las técnicas a utilizar para obtener los resultados deseados, utilizando el programa SPSS, además de un software especializado para algoritmos de minería de datos, seleccionando los más adecuados al caso de estudio.

4. **Evaluación.** Se verificará la validez del modelo seleccionado mediante pruebas.

5. **Despliegue.** Se preparará un prototipo de un tablero de control basado en una plataforma de registro de entrevistas que incluya el modelo de minería de datos.

### V. TRABAJOS RELACIONADOS

Se han encontrado estudios que fundamentan la intención de aplicar técnicas de minería de datos con la finalidad de establecer los parámetros en los que la deserción laboral es más probable dentro de un perfil de empleado.

La investigación de Gutiérrez, Melendres, Ramírez & García [1] establece que *“la deserción laboral es asociada a*

*su vez a factores propios del ambiente de trabajo como son el clima laboral, el turno, el puesto, la remuneración, el plan de carrera que permite la empresa; además de tener relación con factores que conforman el perfil del empleado como su edad, sexo, escolaridad, experiencia laboral, habilidades”, esto después de aplicarse cuestionarios donde se evaluaron tanto el perfil sociodemográfico, académico y de trabajo de los empleados de una empresa maquiladora en Ensenada, Baja California.*

Partiendo del hecho que factores relacionados al perfil de empleado pueden influir en la deserción laboral, Sikaroudi & EsmaieeliSikaroudi [5] aplicaron técnicas de minería de datos para la predicción de la rotación voluntaria de personal en una planta manufacturera en la Ciudad de Arak, Irán. En este estudio se evaluaron características de los empleados como edad, habilidades técnicas y experiencia laboral. A través de la aplicación de minería de datos, dicho estudio generó un sistema de soporte de decisiones para el área de recursos humanos. Con ello se buscaba inferir el desempeño de candidatos a puestos en la empresa, con la finalidad de reducir pérdidas al evitar la contratación de candidatos inestables y poco capacitados.

Por otra parte, Attri [4] muestra en su investigación la aplicación de un modelo que ayuda a predecir la probabilidad de que el empleado deserte de su puesto, basándose en la información previa del área de recursos humanos, con la intención de aplicar estrategias que prevengan estos comportamientos. La propuesta del autor se basa en que el desempeño de los empleados es predecible y este conocimiento ayudaría a evitar la deserción laboral de empleados valiosos para una empresa. De esta forma, utilizó criterios dados por los expertos del área como la edad, sexo, departamento al que pertenece, satisfacción laboral, educación, entre muchos otros y probó que el modelo podía predecir en un 80% el comportamiento de los empleados.

En contraste a las investigaciones que buscan predecir el comportamiento de empleados sobre la deserción, el estudio realizado por Kimiri & Moturi [11] plantea la aplicación de técnicas de clasificación de minería de datos para predecir el desempeño de los empleados existentes en una empresa gubernamental en Nairobi, basándose en la información sociodemográfica y los análisis de desempeño previos. Con dicha información, se creó un modelo clasificador del desempeño laboral, el cual se utilizó para mejorar el proceso de evaluación del desempeño laboral. Como resultado, se obtuvo que la experiencia era el factor de mayor impacto positivo en el desempeño de los empleados, seguido de factores como la edad, calificaciones, sexo, estado civil y puntuación de la evaluación de desempeño.

## VI. RESULTADOS PARCIALES Y ESPERADOS

Con la aplicación de la metodología CRISP-DM en este caso de estudio se espera identificar los perfiles de empleado con tendencia a la deserción temprana, esto considerando que para la empresa Malumex, la deserción temprana es aquella

en la que el empleado no cumple con el contrato inicial de tres meses.

Como segundo punto y basándose en los resultados obtenidos en el punto previo, se plantea generar el prototipo de un modelo clasificador que ayude a la toma de decisiones en las entrevistas iniciales en los procesos de reclutamiento y selección de candidatos.

Como producto derivado se incluye la integración de un tablero de control en el área de recursos humanos que ayude al monitoreo de las actividades de reclutamiento y selección, así como la visualización de los indicadores de rotación y deserción de personal.

Algunos de los resultados parciales incluyen avances en el análisis del negocio y la comprensión de los datos. En particular, durante el análisis del negocio, se aplicaron entrevistas a la gerencia y la jefatura de recursos humanos en los periodos de Febrero a Abril de 2019, donde se estableció la problemática que la empresa presenta en cuanto un alto índice de rotación laboral, el cual fue de 71.16% en el primer trimestre del año 2019.

También se identificaron diversas estrategias que la empresa ha adoptado en los últimos tres años para prevenir la deserción entre sus empleados, las cuales van desde el estímulo monetario (bonos por contratación, asistencia y puntualidad) hasta los reconocimientos morales para los empleados y la modificación del perfil de empleado de los puestos con mayor índice de rotación. Sin embargo, Malumex nunca ha corroborado con un estudio la efectividad de los mismos, mientras que sus índices de rotación continúan elevados.

Por lo tanto, se estableció un objetivo para el proyecto de minería de datos, el cual es identificar un perfil de empleado susceptible a la deserción laboral.

En consecuencia a lo anterior, se obtuvo acceso a la información de los empleados de nivel operativo de la empresa Malumex, S.A. de C.V., ubicados en los centros de trabajo del Parque Industrial de Ciudad Obregón, comprendida de los años 2016 al mes de marzo de 2019. La información existente incluye los datos sociodemográficos de los empleados, así como anotaciones resultantes de las entrevistas de trabajo y de salida de los ex empleados, incluyendo las fechas de inicio y término de la relación laboral.

Se cuenta con una base de datos obtenida del programa SuperNOMINA la cual brindó información de los empleados de Malumex para el periodo de interés, siendo un total de 3736 registros. La información relevante para el estudio que se presenta en la Tabla 1.

Sin embargo, otras variables no se encuentran capturadas en el sistema pero se encuentran presentes en los expedientes físicos de los empleados y los repositorios de información de entrevistas iniciales y de salida de los ex empleados. Esta información se presenta en la Tabla 2 y debe ser capturada manualmente.

TABLA I. VARIABLES RESCATADAS DEL SISTEMA SUPERNOMINA

VARIABLE	DESCRIPCIÓN
Centro de Trabajo	Lugar de trabajo dentro del parque industrial
Equipo	Equipo de trabajo al que estaba asignado el empleado
Sueldo	Categoría de nivel salarial del empleado
Puesto	Puesto asignado al empleado
Fecha de Ingreso	Fecha en que inició la relación laboral
Fecha de Baja	Fecha en que terminó su relación laboral
Estado	Si está activo en nómina o de baja
Turno	Turno en que labora el empleado
Fecha de Nacimiento	Fecha de nacimiento del empleado
Sexo	Sexo del empleado
Tipo de Contrato	Tipo de contrato laboral con el empleado
Tipo de Jornada	Horas diarias que labora el empleado según su contrato

TABLA II. VARIABLES A CAPTURAR MANUALMENTE

VARIABLE	DESCRIPCIÓN
Motivo de Baja	Motivo por el cuál ocurrió la separación laboral
Procedencia	Lugar donde habita el empleado
Hijos	Información sobre si tiene hijos
Estado Civil	Estado civil del empleado
Nivel Educativo	Nivel educativo del empleado
Experiencia Laboral en términos de área	Area del último empleo previo a laborar para Malumex
Experiencia Laboral en términos de antigüedad	Antigüedad del último empleo previo a laborar para Malumex
Reingreso	Información sobre si ha laborado anteriormente con Malumex

## VII. CONCLUSIONES

La identificación de perfiles de empleado susceptibles a la deserción mediante la aplicación de una metodología de minería de datos en la empresa Malumex permitirá a los responsables de los procesos de selección de empleados tomar decisiones más certeras. Malumex podrá tomar decisiones más informadas sobre quiénes son los mejores candidatos posibles para los puestos requeridos. Esto significará además la reducción del índice de rotación y por consecuencia un mejor desempeño por parte de la empresa en cuanto al cumplimiento de sus contratos con los clientes.

En cuanto a la información que se obtuvo, existen irregularidades debido a la captura manual por parte de los colaboradores de Malumex, que incluyen duplicidad de datos, errores de captura y valores nulos en la información. Esta situación se corregirá en la fase de preparación de datos, mediante el uso de técnicas de limpieza de datos principalmente, para tratar los valores perdidos.

Otro aspecto que se debe considerar, son las diversas fuentes de información que se requerirán para tener una base de datos completa con la cual aplicar el modelado para la minería de datos, lo cual implica la captura manual de una gran cantidad de datos. Este proceso se realizará utilizando la información de los expedientes de empleados y ex

empleados, quedando a cargo del autor la captura manual de la información.

Para la continuidad del proyecto, se estima realizar un análisis exploratorio de la información presentada, previo a la limpieza de los datos y el modelado.

## REFERENCIAS

- [1] M. F. A. Gutiérrez, V. D. M. Carlos, M. C. R. Barón, y B. R. G. Rivera, "Los factores de la rotación de personal en las maquiladoras de exportación de ensenada, B.C.", vol. 6, núm. 2, p. 16, 2018.
- [2] I. Chiavenato, "Administración de recursos humanos. El capital humano de las organizaciones", p. 519.
- [3] H. Yousaf, "Analysing which factors are of influence in predicting the employee turnover", p. 25.
- [4] T. Attri, "Why an Employee Leaves: Predicting using Data Mining Techniques", p. 23.
- [5] A. M. E. Sikaroudi y A. EsmaeeliSikaroudi, "A data mining approach to employee turnover prediction (case study: Arak automotive parts manufacturing)", p. 16.
- [6] "Boletín Rotación y ausentismo.pdf".
- [7] H. Jantan, A. R. Hamdan, y Z. A. Othman, "Towards applying Data Mining Techniques for Talent Mangement", p. 6.
- [8] W. W. Eckerson, N. Hanlon, y R. Barquin, "DIRECTOR OF EDUCATION AND RESEARCH", vol. 5, núm. 4, p. 15, 2000.
- [9] S.A. Alasadi y W. S. Bhaya, "Review of Data Preprocessing Techniques in Data Mining", *Journal of Engineering and Applied Sciences*, vol. 12, núm 16, pp. 4102-4107, 2017.
- [10] "IBM SPSS Modeler CRISP-DM Guide", p. 50.
- [11] J. M. Kimiri y C. A. Moturi, "Application of Data Mining Classification in Employee Performance Prediction", *International Journal of Computer Applications*, vol. 146, núm. 7, pp. 28-35, jul. 2016.

# Diseño de un tablero de control para la toma de decisiones de productores de limón

Berenice Mancilla Rojas, Luis A. Castro

Depto. de Computación y Diseño, Instituto Tecnológico de Sonora (ITSON)

Ciudad Obregón Sonora, México

mancybe@gmail.com, luis.castro@acm.org

**Resumen**—El valor del limón se ha incrementado debido a la escasez del mismo, derivada a su vez de diversos factores. Dichas condiciones en conjunción con la poca tecnificación del cultivo han de alguna manera afectado la producción de dicho fruto. Este trabajo tiene por objetivo utilizar las tecnologías de información para diseñar una herramienta que apoye en la toma de decisiones de los productores de limón en México. Los resultados de este trabajo pueden beneficiar a los productores de tal manera que pueden hacer un mejor uso de los recursos utilizados en los cultivos, además de prevenir plagas que representan pérdidas para ellos.

**Palabras clave:** --Agricultura, Limón, Tecnología, Decisiones, Tablero, Indicador.

## I. INTRODUCCIÓN

En los últimos años el valor de limón ha incrementado debido a la escasez de producción que ha sido provocada en parte por plagas que atacan a los árboles productores. Tan solo en Colima, de 2010 a 2012 se estimó una caída en la demanda de mano de obra para labores de campo y cosecha por 2 435 empleos de tiempo completo [1].

El consumo de este fruto es indispensable y de suma importancia no solo para la alimentación, sino que, por su fuerte poder antioxidante y antiviral, el limón es un producto importante para la salud humana. De igual manera, forma parte de los Citroflavonoides los cuales desempeñan un papel esencial en la protección frente a los fenómenos de daño oxidativo, y tienen efectos terapéuticos en un elevado número de patologías [2]. Debido a esto, su producción y distribución no deben verse afectadas. Sin embargo, en muchas regiones en México la producción de limón ha ido en declive. Además, en muchos casos, los productores de limón no cuentan con herramientas tecnológicas necesarias que brinden información para contrarrestar plagas, utilizar sus recursos adecuadamente, con lo que se afecta la producción y distribución de este producto no solo en México sino en otros países a donde se exporta.

Debido a este rezago tecnológico que existe, muchas de las decisiones que toman los productores las hacen de acuerdo a su experiencia. En general, en México los productores no cuentan con sistemas de información que apoye a su toma de decisiones o los orienten en cuanto a diversos temas de interés para ellos como lo son: prevención de enfermedades en sus cultivos, precios en el mercado, utilización eficiente de pesticidas o fertilizantes, manejo del riego, entre otros.

En este trabajo se espera poder contribuir a esta actividad económica mediante la utilización de tecnologías de información y el manejo de datos diseñando un tablero de control que permitan a los productores tener una herramienta que apoye a la disminución de situaciones amenazantes para sus cosechas y así lograr una producción más segura y eficiente de limón.

Este artículo está organizado como sigue. En la sección II se muestra el marco teórico del trabajo. En la sección III y IV se muestran los objetivos y la metodología a utilizar. Finalmente, se presentan los resultados esperados y las conclusiones en la sección V y VI, respectivamente.

## II. MARCO TEÓRICO

Es importante entender y definir todos los conceptos acerca de los sistemas de información y el manejo de los datos, ya que son el soporte de las herramientas que apoyan en la toma de decisiones en los negocios. Es preciso destacar que existen agro negocios que también pueden implementar estas herramientas para contar con conocimiento que haga más eficiente la producción agroalimentaria además de asegurar grandes cultivos y prevenir enfermedades que puedan representar grandes pérdidas económicas para los productores y escasez de alimentos en el mercado.

### A. Agricultura de Precisión

La Agricultura de Precisión se define como un sistema empleado para analizar y controlar la variación espacio-temporal del terreno y el cultivo. La variación espacial comprende las diferencias en fertilidad de distintas secciones del terreno y las que se dan en el crecimiento de las plantas cultivadas. La variación temporal engloba las diferencias observadas en la producción de un mismo terreno entre una temporada y otra (Tabla I) [3].

TABLA I. VARIACIÓN ESPACIO-TEMPORAL EN CULTIVOS [3]

Tipo de variación	Sujeto de la variación	Factores involucrados
<i>Temporal</i>	Cosecha	<ul style="list-style-type: none"><li>– Variación productiva entre periodos de siembra distintos.</li><li>– Condiciones climatológicas (por ej. radiación solar o humedad ambiental) entre distintas temporadas.</li></ul>
<i>Espacial</i>	Fertilidad del suelo	<ul style="list-style-type: none"><li>– Condiciones fisicoquímicas (entre otras la acidez-alcalinidad o pH, y el contenido de nitrógeno o de metales).</li></ul>

		<ul style="list-style-type: none"> <li>- Contenido de humedad, materia orgánica y contaminantes.</li> <li>- Conductividad eléctrica e hidráulica.</li> <li>- Textura, fuerza mecánica y profundidad.</li> <li>- Salinidad.</li> <li>- El relieve o topografía del terreno.</li> <li>- Microbiota y fauna del suelo.</li> </ul>
	Desarrollo vegetal	<ul style="list-style-type: none"> <li>- Maleza (plantas oportunistas).</li> <li>- Plagas (insectos, virus y microorganismos).</li> <li>- Características genéticas del cultivo (como la resistencia a la sequía y velocidad de desarrollo).</li> </ul>

Se requieren tres etapas para la implementación de la agricultura de precisión: a) La recolección de datos, la cual se lleva a cabo con equipos especializados como satélites o sensores remotos; b) El análisis de los datos, en la que un experto analiza los datos y emite sugerencias para manejar adecuadamente la variación espacio-temporal detectada; y c) La implementación, en la que el productor cultiva el terreno según las recomendaciones.

### B. Inteligencia de Negocios

Por otra parte, la Inteligencia de Negocios (BI, del inglés Business Intelligence) es una herramienta bajo la cual diferentes tipos de organizaciones, pueden soportar la toma de decisiones basadas en información precisa y oportuna; Estos apoya a la generación del conocimiento necesario que permita escoger la alternativa que sea más conveniente para el éxito de la empresa [4].

Es entonces que a partir del surgimiento de esta división en las Tecnologías de Información (TI) la agricultura se ve contribuida por la ingeniería de sistemas y electrónica ofreciendo a usuarios la posibilidad de incrementar la producción y calidad de productos, incorporando sensores inteligentes capaces de procesar y transmitir información inalámbricamente a una base de datos, para su posterior análisis estadístico, usando técnicas de inteligencia de negocios para la toma de decisiones importantes en alta gerencia y de esta manera aumentar la producción a costos de implementación elevados recuperables a mediano o largo plazo [5].

La aplicación más común de las TIC en la agricultura es la creación de servicios de información a través de páginas Web especializadas. Noticias, consejos prácticos, reportes de clima, resultados de investigaciones, precios de productos, cotizaciones, son solo algunas de las secciones predilectas en estos sitios de Internet, que son impulsados por los gobiernos y por organizaciones que estudian el tema. La información a la que puede acceder un agricultor le permite hacer una planificación estratégica de sus refuerzos, reducir sus costos, mejorar sus negociaciones colectivas, capacitarse en línea, dar a conocer sus productos, expresar sus posturas en la formulación de leyes y reglamentaciones que les afecten, hacer conocimiento público sus necesidades locales y sus proyectos, y hasta contribuir en la investigación agrícola. Hoy en día, es posible asegurar que la incorporación de las

técnicas informáticas o tecnologías de la información a la agricultura, permite desarrollar herramientas que hacen posible aumentar la eficiencia y productividad de los sistemas de explotación, así como mejorar las condiciones organizativas en el sector agrícola. Al lograr la automatización de las actividades en algunos casos se puede facilitar el manejo de los datos, con un significativo ahorro de tiempo, mayor veracidad, confiabilidad y rapidez en la obtención de la información para llevar a cabo los objetivos propuestos [6].

### C. La tecnología en la producción de limón en México

En México existen cultivos que requieren de implementación de dichas herramientas tecnológicas y sistemas de información que auxilien a los productores en la toma de decisiones para prevenir enfermedades, utilizar eficientemente los recursos y para mejorar la sostenibilidad y productividad para su posterior venta dentro y fuera del país.

Uno de esos es el Limón Mexicano (*Citrus aurantifolia* Christ Swingle), del cual se cultivan 83 mil hectáreas. Dichas Hectáreas producen 745 mil toneladas de fruto, con un rendimiento promedio de 9.0 ton/ha. La mayor parte (90 %) de la producción ocurre en los meses de junio a septiembre, época que coincide con los precios más bajos en los mercados nacional e internacional. En el caso del limón, la producción es escasa en el resto del año, sobre todo en los meses de diciembre a abril, cuando se alcanzan los precios más altos, y el cultivo es más rentable. Además de ser bajo el rendimiento, la calidad de los frutos no es adecuada a causa del manejo deficiente del cultivo [7].

Por otro lado, la producción de alimentos y otros productos agrícolas normalmente se ve mermada por diversos siniestros, entre ellos las plagas, las cuales han sido un problema preocupante, tanto por sus efectos directos como por aquéllos indirectos que llegan a provocar. El Huanglongbing (HLB), enfermedad de origen bacteriano considerada en el mundo como la más destructiva para todas las especies de cítricos. Para México representa un verdadero reto fitosanitario, tal vez el más importante que se haya enfrentado, no solo por el daño que el HLB ocasiona a las plantas, sino porque se cuenta con más de medio millón de hectáreas de cítricos en 23 estados. Si este tipo de plagas llegara a los cultivos de limón mexicano y los productores no cuentan con información acerca de sus síntomas en las plantas, las pérdidas potenciales estimadas para la actividad citrícola primaria están consideradas de la siguiente manera: el país enfrentaría perjuicios en la producción de limón agrio en 183,168 toneladas, a las cuales Colima, al estar expuesta a un impacto alto del HLB, contribuiría con el 48% (87,765 toneladas); Michoacán frente a impacto moderado, con el 32% (59,071 toneladas); y otros 20 estados expuestos a impacto bajo, con el 20% restante (36,332 toneladas) [8]. Debido a esto, es importante contar con tecnología que permita a los productores estar preparados ante una eventual amenaza como las plagas.

### III. OBJETIVO

El objetivo de este trabajo es diseñar un tablero de control para el apoyo a la toma de decisiones de los productores de

limón, y así mediante su implementación apoyar para la prevención de cualquier situación desfavorable que amenace con afectar sus cultivos.

#### IV. METODOLOGÍA DE TRABAJO

El proyecto se está desarrollando en 4 etapas:

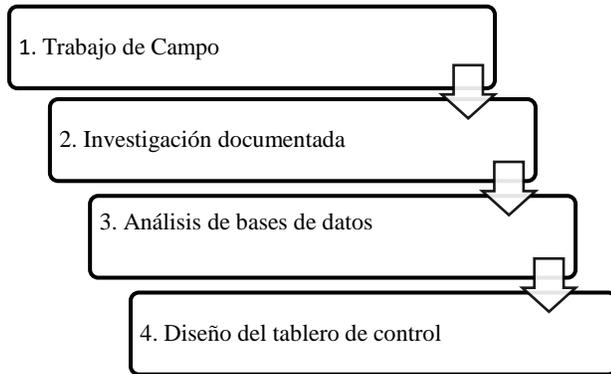


Figura 1. Diagrama de la metodología implementada.

##### A. Trabajo de campo

En la primera etapa se contempla un trabajo de campo con los productores del limón mexicano para definir reglas de comportamiento en las plantas. Los participantes incluyen a productores de limón en el Estado de Colima, los cuales cuentan con huertas de limón y que es su actividad económica principal.

Se tienen considerados dos tipos de instrumentos de recolección de datos. El primero fue entrevistas semi-estructuradas para conocer como es su relación y las necesidades de información que requieren para tomar decisiones acerca de la cosecha. El segundo instrumento son cuestionarios, que éstos fueron aplicadas a personas dueñas de empaques de limón y empresas de logística que se encargan de comprar dicho fruto en su mayoría en Colima, estos últimos con la finalidad de conocer los requerimientos del mercado y que tanto se adaptan los productores a las exigencias de sus principales compradores.

La primera herramienta fue la entrevista personal dirigida que como que es una conversación generalmente entre 2 personas, (uno el entrevistador y otro el entrevistado) Las preguntas pueden ir registradas en una boleta que se llama cuestionario o bien se puede auxiliar de una grabadora para registrar los datos obtenidos. La ventaja que tiene es que la gente generalmente responde cuando es confrontada en persona. El entrevistador puede notar reacciones específicas y eliminar malos entendidos sobre alguna pregunta hecha. Sus limitaciones a parte del costo, es el adiestramiento que tenga el entrevistador. Este es el método más completo para adquirir información por estar en contacto directo con la fuente, es por eso que se eligió para aplicarlo a los productores, ya que es ahí donde va a estar dirigido el funcionamiento del tablero de control [9].

Como segunda herramienta se utilizaron las encuestas, pero estos fueron dirigidos para otro grupo que influye en la cadena productiva del limón, que son los compradores, estos

son importantes porque ellos definen características de compra del fruto, a los cuales los productores de limón tienen que adaptarse para poder vender su limón.

##### B. Investigación documentada

Esta fase incluye investigación documentada con agrónomos acerca del comportamiento de las plantas de limón en su entorno con las distintas variables de interés.

La literatura publicada acerca del comportamiento de las plantas de limón, su crecimiento, características de desarrollo y cuáles son las enfermedades que más afectan a estas plantas, y esto con la finalidad de establecer reglas que ayuden de manera visual a generar los indicadores dentro del tablero.

##### C. Análisis de bases de datos

Esta etapa consiste en realizar búsqueda de bases de datos que incluyan variables de interés para poder ver el comportamiento de las plantas y el funcionamiento en la información brindada.

Las bases de datos que se tomarán en cuenta para introducirlas y ver su comportamiento en el tablero de control, son de la estación climatológica de Conagua delegación colima, instalada en el municipio de Tecomán, mismo que es sede de la mayoría de cultivos de limón en el Estado de Colima.

##### D. Diseño del tablero de control

En esta fase, se va a diseñar el tablero de control e introducir la información para analizar si en realidad la información que arroja apoya a las decisiones de los productores.

El diseño del tablero se realizará con base a los requerimientos arrojados en la investigación de campo con los productores y compradores, y con información del comportamiento de las plantas, todo eso con indicadores que se afecten al introducir información climatológica pueda enviar alertas de prevención a los productores y así puedan tomar decisiones sustentadas para cualquier asunto de interés de sus cultivos.

#### V. RESULTADOS ESPERADOS

En este trabajo se espera como resultado el diseño de tableros de control que permitan a los productores obtener orientación mediante información ahí plasmada para tomar decisiones respaldadas con el comportamiento de sus variables y así poder llevar un mejor manejo de sus cultivos y estos no se vean afectados por algún fenómeno inapropiado.

Como resultados esperados se busca que estos tableros tengan un gran impacto para la producción y para la economía de los productores mexicanos, debido a que facilitará información que respaldará las decisiones que tomen para cualquier tema que tenga que ver con sus cultivos y los recursos utilizados en la plantación, además de que apoyará para el rendimiento las tierras y la calidad en los frutos.

## VI. CONCLUSIONES

Este trabajo tiene como objetivo la implementación de tecnologías de información al ámbito de la agricultura en México, como bien se menciona en su marco teórico, la inteligencia de negocios es una herramienta emergente que está tomando fuerza en las pymes y empresas tradiciones de comercialización y transformación en la industria, pero cabe mencionar que no se ha dado un especial enfoque en el uso de dichas herramientas para la mejora de la industria agroalimentaria y que es hoy en día un área con muchas oportunidades de mejora ya que la población está creciendo exponencialmente y la cantidad de alimento que se está requiriendo cada vez va en aumento y las industria han buscado otros métodos por sustituir muchos alimentos provenientes de la tierra por cosas sintéticas que al final afectan la salud humana. Por lo cual es importante el enfoque de apoyar la agricultura con la tecnología para que su producción no se vea afectada y puedan satisfacer las necesidades de la sociedad.

## REFERENCIAS

- [1] M. M. Robles-gonzález, M. Orozco-santos, M. Á. Manzanilla-ramírez, y J. J. Velázquez-monreal, “Efecto del HLB sobre el rendimiento de limón mexicano en Colima , México \* Effect of HLB on Mexican key lime yield in Colima , Mexico Resumen Introducción”, vol. 8, pp. 1101–1111, 2017.
- [2] S. Martinez-florez y J. Culebras, “Flavonoids: Properties and antioxidizing action”, núm. November, 2002.
- [3] P. Zhang y Z. Zhou, “Postharvest ethephon degreening improves fruit color, flavor quality and increases antioxidant capacity in ‘Eureka’ lemon (Citrus limon (L.) Burm. f.)”, *Sci. Hortic. (Amsterdam)*, 2019.
- [4] U. F. De Paula, “INTELIGENCIA DE NEGOCIOS : ESTADO DEL ARTE BUSINESS INTELLIGENCE : STATE OF THE ART”, núm. 44, pp. 321–326, 2010.
- [5] F. E. J. R. Angel Andres Espinza Vega, “Unidad académica de ingeniería civil carrera de ingeniería de sistemas”, 2016.
- [6] M. M. y M. M. Adriana Perez, “Revisión bibliográfica”, 2006.
- [7] A. B. A. y Noe larcon C. Rafael Ariza Flores, Ruben Cruzaly Sarabian, Enrique Vazquez Garcia, “Efecto de las labores culturales en la producción y calidad del lim”. 2004.
- [8] D. Salcedo *et al.*, “DE LA ENFERMEDAD DE LOS CITRICOS HUANGLONGBING ( HLB ) EN LA CADENA Mayo 2010”, 2010.
- [9] P. Inga, M. Torres, I. Ing, y F. G. Salazar, “Metodos de recoleccion de datos para una investigación”, núm. 03, pp. 1–21.

# Implementación de un sistema de almacenamiento de datos agroclimáticos para la predicción de cosecha de trigo

María Monserrat Torres, Luis A. Castro

Depto. de Computación y Diseño, Instituto Tecnológico de Sonora (ITSON)

Ciudad Obregón, México

tcmonserrat90@gmail.com, luis.castro@acm.org

**Resumen**— La variabilidad del clima es de gran importancia para la producción agrícola en muchas regiones del mundo, la cual constituye uno de los factores principales de la falta de seguridad alimentaria. Se ha reportado una relación entre el rendimiento de trigo con el clima, debido a que el desarrollo de este cereal es influenciado por factores meteorológicos como la temperatura del aire y la duración de la luz del día. El objetivo de este trabajo es diseñar un almacén de datos para la predicción del rendimiento del trigo. Para la realización del trabajo se tomarán datos proporcionados por la Red de Estaciones Meteorológicas Automáticas de Sonora (REMAS) y el diseño de la estructura del almacén será con base en los requerimientos del usuario. Además, se evaluará el tiempo de carga de registros, realizando una comparación del tiempo de consulta de los datos anteriores con los actuales. Posteriormente, se realizará el análisis de la información para obtener un pronóstico de la producción en tiempo real. Se proyecta que los resultados de este trabajo de tesis permitan apoyar en la toma de decisiones en el sector agrícola.

**Palabras clave**- almacén de datos, datos agroclimáticos, pronóstico.

## I. INTRODUCCIÓN

La agricultura es uno de los recursos más importante con el que cuenta el hombre, y es también una de las actividades humanas que más dependen del clima. Según estadísticas de la FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura), el 40% de la población económicamente activa se dedica a la agricultura para su subsistencia [1]. El impacto de la variabilidad climática ha ocasionado pérdidas en la producción de trigo a nivel mundial [2], el cual vincula el rendimiento con el clima y se indica que la producción mundial de trigo disminuyó 5.5% en las últimas tres décadas. Científicos, organizaciones no gubernamentales (ONG) e instituciones de gobierno se han dado a la tarea de analizar diferentes escenarios del cambio climático, y las consecuencias negativas que se puedan tener para la agricultura a nivel mundial.

México representa el 4.85% de la producción global de trigo [3]. La Secretaría de Agricultura y Desarrollo Rural (SADER) en México, reportó una disminución de la producción nacional de trigo de 344 mil toneladas durante el año 2017 en relación al ciclo anterior, con base en estadísticas del Servicio de Información Agroalimentaria y Pesquera [4]. El estado de Sonora es líder en la producción

de trigo grano y representa el 51.7% de la producción nacional.

Los constantes cambios que ha presentado el clima han ocasionado pérdidas hasta un 29% en la producción de trigo en el sur de Sonora [5]. Se prevé que el aumento en la oscilación de los factores climáticos ejercerá un efecto negativo en el desarrollo fenológico del trigo, y predispondrá a la planta para el ataque por algunos insectos y enfermedades [5]. De igual forma, se pronostica una disminución en el rendimiento del trigo de 15% considerando los escenarios del cambio climático [6].

La predicción de la producción antes de sembrar tiene influencia en la gestión de las actividades de cosecha y almacenamiento. Este tema ha sido tratado por investigadores en diferentes países con el propósito de programar volúmenes de exportación y diagnosticar los efectos del clima en los cultivos [5]. Además, con el análisis de los factores limitantes y la situación de alerta de las condiciones climatológicas pueden emitirse estrategias de manejo agronómico para reducir el efecto adverso que podrían causar las interacciones de estos factores climatológicos en la producción.

En el 2008, se realizó un estudio sobre el comportamiento histórico del clima en el sur de Sonora [7], en el cual se seleccionaron estratégicamente estaciones climáticas representativas del valle del Yaqui y valle del Mayo, las cuales pertenecían a la red de monitoreo meteorológico de la Comisión Nacional del Agua (CONAGUA). La selección de las estaciones para el análisis se basó principalmente en la disponibilidad de la base de datos, ubicación geográfica, extensión histórica y consistencia de los datos registrados. La base de datos climatológicos se formó a partir de registros diarios, por lo que fue necesario recopilar la extensa información en hojas de cálculo, creando un archivo maestro o de trabajo, donde se agrupó la información de cada una de las estaciones en un arreglo especial. En dichas hojas, los datos fueron recolectados y procesados para la elaboración de cuadros y gráficos para el análisis y descripción del comportamiento histórico mensual y anual del clima.

Actualmente se encuentran las estaciones climáticas de la Red de Estaciones Meteorológicas Automáticas de Sonora (REMAS); los datos se reportan con una frecuencia de 10 minutos, horario y diario, los cuales son transmitidos a través de sensores que evalúan la temperatura del aire (mínima, máxima), humedad relativa, precipitación, radiación solar,

velocidad y dirección de viento, presión barométrica y evapotranspiración.

Mediante estos datos se puede realizar un análisis detallado de cada una de las variables; por ejemplo, el comportamiento de la temperatura, lo cual ayuda a determinar la cantidad de horas frío que se han presentado para el ciclo del cultivo del trigo. Sin embargo, el procesamiento de los datos para realizar este análisis es normalmente laborioso, ya que se carece de la tecnología adecuada para la preservación y consulta de dichos datos.

El desarrollo de un sistema de almacén de datos para la preservación y consulta de datos agroclimatológicos, es una gran ventaja para el apoyo a la toma de decisiones en el sector agrícola. Esto puede derivar en una mejora en la programación de volúmenes de exportación y diagnósticos en los efectos del clima en los cultivos. En el almacén de datos, los datos se encontrarán en un mismo lugar, con un mismo formato para mejorar el procesamiento por parte de los usuarios, considerando a productores, investigadores y empresarios. Además, el tiempo de respuesta será menor para el desarrollo de reportes o realizar consultas de años anteriores.

## II. MARCO TEÓRICO Y TRABAJO RELACIONADO

Actualmente se han desarrollado distintas definiciones de qué es y cuál es el uso de un almacén de datos dentro de la organización, pero la esencia de la definición permanece. Un almacén de datos (en inglés *Data Warehouse*) “es una colección de datos orientados por temas, integrados, variables en el tiempo y no volátiles que se emplea como apoyo a la toma de decisiones estratégicas” [8].

Por otro lado, el modelo “The Data Warehouse Toolkit” [9] ha sido adoptado por la industria de los almacenes de datos y el desarrollo de los mismos es mediante el modelado dimensional o estrella. En general, un almacén de datos es un sistema que integra información de distintas fuentes de datos para la consulta y el análisis de los datos, que a diferencia de las bases de datos relacional o conjuntos de datos no es posible realizar. Además, la razón para producir información de un almacén de datos, es ayudar a la toma de decisiones con la capacidad de realizar análisis en tiempo real.

Los almacenes de datos han sido implementados en distintos campos de trabajo, como por ejemplo en el campo de la educación [10], medicina [11], gobierno [12], agricultura [13]. En el caso de este proyecto de tesis, se tratará con trabajos relacionados con los almacenes de datos o bases de datos usados en el área climática.

De hecho, existen diversos sistemas que se han implementado para el uso de datos climáticos, con la finalidad de procesar y visualizar datos diarios provenientes de diferentes estaciones meteorológicas. Uno de estos sistemas es el sistema CLICOM (CLimate COMputing Project) del Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), el cual es un software de manejo de datos climatológicos desarrollado por las Naciones Unidas [14]. Esta herramienta ofrece un ambiente Web donde se realiza una selección por estado, estaciones y filtros sobre un mapa donde se procesan datos diariamente. Contrario a esto, los usuarios deben conocer de

antemano el nombre de la estación climática que desean analizar, para que la herramienta proporcione la información. La herramienta CLICOM permite hacer gráficas y descargarlas, así como los datos de las gráficas en texto. De manera adicional, otro software llamado MOCLIC (Monitoring Climate Change) es capaz de identificar y evaluar las tendencias de los cambios climáticos a causa del cambio climático global. Esta herramienta funciona a través de datos obtenidos por las estaciones meteorológicas, y puede identificar tendencias de cambio en las temperaturas extremas ocasionadas por el cambio climático de una comunidad en específico [15]. Finalmente, otra plataforma es la Red de Estaciones Meteorológicas Automatizadas de Sonora (REMAS), la cual tiene como propósito el desarrollo, gestión y mantenimiento de las estaciones que la integran, a fin de garantizar la sanidad vegetal y facilitar la exportación de los productos agrícolas cultivados en Sonora. Esta herramienta ofrece un sitio Web con el objetivo de ser útil para los productores del estado de Sonora [16].

## III. OBJETIVO

El objetivo del presente proyecto de tesis es el diseñar un almacén de datos agroclimáticos, con el fin de contar con datos de calidad en un tiempo corto para su análisis, y consecuentemente para favorecer el desarrollo de sistemas de pronósticos de rendimiento de cosecha de trigo.

### A. Objetivos específicos

- Investigar metodologías para el almacén de datos, con la finalidad de seleccionar la que se adapte mejor a las necesidades actuales.
- Diseñar un almacén de datos de las distintas fuentes, ordenando la información climática.
- Implementar el almacén de datos diseñado para analizar y comparar el tiempo de carga de los datos y para el pronóstico de cosecha.

## IV. METODOLOGÍA

En esta sección se presenta la metodología para el desarrollo de este trabajo.

### A. Procedimiento de la investigación

El desarrollo del trabajo de tesis se llevará a cabo en seis etapas como se muestra en la Figura 1 y las cuales se describen a continuación:

#### *Etapa 1.- Revisión de la literatura.*

Consiste en la investigación de las diferentes metodologías para el desarrollo de almacenes de datos, y realizar una comparativa para seleccionar la que más se adapte al usuario final.

#### *Etapa 2.- Requerimientos del usuario.*

Consiste en la realización de pláticas, entrevistas semi-estructuradas y conocimiento del área agrícola para el levantamiento de requerimientos por parte de los interesados, con la finalidad de generar un programa de trabajo y ordenar los resultados según la metodología seleccionada para el desarrollo.

#### *Etapa 3.- Diseño de la estructura del almacén de datos.*

En esta etapa se incluye el diseño de la estructura del almacén de datos mediante la metodología de Kimball [9], con la cual permitirá al usuario final entender el diseño de la estructura del almacén.

*Etapa 4.- Proceso de extracción, transformación y carga.*

Una vez diseñada la estructura del almacén, se crea un proceso de automatización de datos, pasando por una evaluación de calidad de los datos y posteriormente ser cargados a la estructura.

*Etapa 5.- Evaluación del desempeño de la estructura.*

Se realiza la evaluación del desempeño de la estructura mediante la carga de los datos y haciendo una comparativa de la carga de los datos anterior con la actual.

*Etapa 6.- Realización del pronóstico del rendimiento de trigo.*

Se realiza la estimación del rendimiento de trigo por hectárea en una primera aproximación mediante el modelo de simulación climática, considerando que el modelo es alimentado con la información de las estaciones meteorológicas [17].

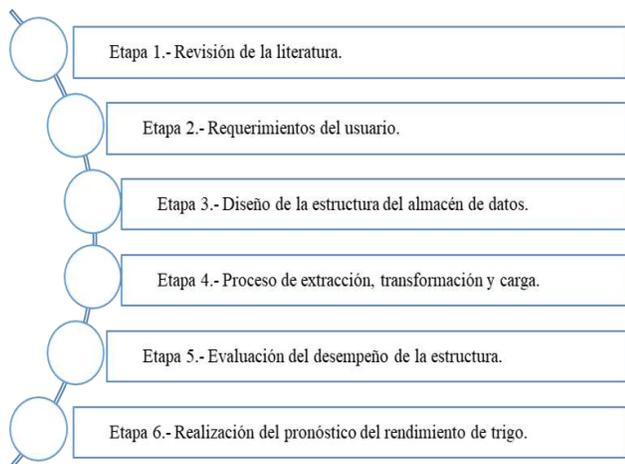


Figura 1. Procedimiento del trabajo de tesis

*B. Instrumentos y materiales*

*1) Set de datos 1*

Uno de los elementos a utilizar es un conjunto de datos de libre acceso por medio de la red de monitoreo meteorológico de la Comisión Nacional del Agua (CONAGUA) [7], el cual tiene 360 archivos de hojas de cálculo, con aproximadamente 17,856 registros cada una. Estos registros son generados por medio de sensores que miden distintas variables del clima (Tabla I).

TABLA I. LISTADO DE VARIABLES CONTENIDA EN EL SET DE DATOS 1

Variable	Descripción
Temperatura del aire	Grado de calor específico del aire.
Humedad relativa	Relación entre la cantidad de vapor de agua que tiene una masa de aire y la máxima que podría tener.
Precipitación	Caída de agua desde la atmósfera hacia la superficie terrestre

*2) Set de datos 2*

De igual manera, las estaciones climáticas de la Red de Estaciones Meteorológicas Automáticas de Sonora (REMAS) también se consideran, las cuales tienen como propósito generar, almacenar, procesar y diseminar datos meteorológicos en el estado de Sonora. Actualmente, se cuentan con 224 archivos de hojas de cálculo, con aproximadamente 30,355 registros cada una. Por medio de sensores automatizados son generados estos registros que miden 8 variables climáticas, pero para fines del trabajo de investigación, solamente se tomarán en cuenta 5 variables, las cuales son las siguientes:

TABLA II. LISTADO DE VARIABLES CONTENIDA EN EL SET DE DATOS 2

Variable	Descripción
Temperatura del aire	Grado de calor específico del aire.
Humedad Relativa	Relación entre la cantidad de vapor de agua que tiene una masa de aire y la máxima que podría tener.
Precipitación	Caída de agua desde la atmósfera hacia la superficie terrestre
Radiación Solar	Conjunto de radiaciones electromagnéticas emitidas por el Sol
Evapotranspiración	Cantidad de agua del suelo que vuelve a la atmósfera como consecuencia de la evaporación y de la transpiración de las plantas.

V. RESULTADOS ESPERADOS

Una de las contribuciones de este trabajo es el diseño de un almacén de datos que permitirá el apoyo a la toma de decisiones en el sector agrícola, considerando presentar la información en tiempo real y oportuna para los diferentes usuarios. Además, obtener un pronóstico dinámico del rendimiento de trigo para los diferentes ciclos.

VI. CONCLUSIONES

El análisis de los datos climáticos tiene una gran importancia en la actualidad, debido a que hay numerosos factores que pueden influir en sectores diversos a través del descubrimiento de patrones, tendencias o predicciones en relación con el clima.

La decisión de implementar un sistema de almacén de datos, ayudará a contar con datos de calidad antes de convertirse en información, y como resultado tener una visión más clara y automatizada para el análisis de los datos, generando conocimientos a partir de datos originales.

En este caso, el análisis de datos climáticos es una gran ventaja para la toma de decisiones en el sector agrícola, debido a que los datos se encuentran en un mismo lugar, con un mismo formato para mejorar el procesamiento por parte de los usuarios. El impacto de la información a priori, permite a la agricultura regional medir la perspectiva de exportación en cantidad y calidad y los precios del producto trigo.

REFERENCIAS

- [1] FAO (Organización de las Naciones Unidas para la Agricultura y la Alimentación), *El estado mundial de la agricultura y la alimentación 2010-11*. 2012.
- [2] S. Asseng et al., "Rising temperatures reduce global wheat production," *Nat. Clim. Chang.*, vol. 5, no. 2, pp. 143–147, Jan. 2015.
- [3] FAOSTAT, "PRODUCCION CULTIVOS," 2017. [Online]. Available: <http://www.fao.org/faostat/es/#compare>. [Accessed: 12-Jul-2019].
- [4] SIAP (Servicio de Información Agroalimentaria y Pesquera), "Anuario Estadístico de la Producción Agrícola," 2017. [Online]. Available: <https://nube.siap.gob.mx/cierreagricola/>. [Accessed: 20-Jun-2019].
- [5] P. Félix Valencia, J. E. Ortíz Enriquez, G. Fuentes Dávila, J. G. Quintana Quiroz, and J. Grajeda Grajeda, *Horas frío en relación al rendimiento de trigo*, no. 63. 2009.
- [6] L. Argenteal, J. Garatuzo-Payan, M. Armendariz Ontiveros, E. Yopez, J. T. Arredondo Moreno, and J. González Aguilera, "Estrés térmico en cultivo del trigo. Implicaciones fisiológicas, bioquímicas y agronómicas," *Cultiv. Trop.*, vol. 38, pp. 57–67, 2017.
- [7] P. Félix Valencia, J. C. Leyva Corona, J. E. Ortíz Enriquez, J. G. Quintana Quiroz, J. Grajeda Grajeda, and G. Jiménez Ortega, *Comportamiento histórico y tendencia del clima en la zona Agrícola y pecuaria del sur de Sonora. Expectativas ante un cambio climático*. Cd. Obregon, 2008.
- [8] A. Puerta, *Business Intelligence Y La Tecnología de la Información*. IT Campus Academy, 2015.
- [9] R. Kimball, "The data warehouse toolkit: the definitive guide to dimensional modeling / by Ralph Kimball, Margy Ross," no. c, p. 601, 2013.
- [10] S. Gorbea-Portal and M. de Jesús Madera-Jaramillo, "Diseño de un data warehouse para medir el desarrollo disciplinar en instituciones académicas," *Investig. Bibl.*, vol. 31, no. 72, pp. 161–181, May 2017.
- [11] M. A. Raebel et al., "Identifying preanalytic and postanalytic laboratory quality gaps using a data warehouse and structured multidisciplinary process," *Arch. Pathol. Lab. Med.*, vol. 143, no. 4, pp. 518–524, Apr. 2019.
- [12] R. K. Arora and M. K. Gupta, "e-Governance using Data Warehousing and Data Mining," *Int. J. Comput. Appl.*, vol. 169, no. 8, pp. 975–8887, 2017.
- [13] V. M. Ngo, N.-A. Le-Khac, and M.-T. Kechadi, "An Efficient Data Warehouse for Crop Yield Prediction," in *Proceedings of the 14th International Conference on Precision Agriculture*, 2018, pp. 1–12.
- [14] CICESE (Centro de Investigación Científica y de Educación Superior de Ensenada), "CLICOM," 2013. [Online]. Available: <http://clicom-mex.cicese.mx/>. [Accessed: 21-Aug-2019].
- [15] F. Bautista, D. A. Bautista-Hernández, O. Álvarez, M. Anaya-Romero, and D. De La Rosa, "Software para identificar las tendencias de cambio climático a nivel local: un estudio de caso en Yucatán, México," *Rev. Chapingo, Ser. Ciencias For. y del Ambient.*, vol. 19, no. 1, pp. 81–90, Jan. 2013.
- [16] CESAVE, "Manual De Usuarios Del Sitio Web Remas," pp. 1–18, 2015.
- [17] P. Félix Valencia, A. D. Baez González, V. M. Rodríguez Moreno, J. L. Ramos González, and M. A. González González, *Predicción de cosechas. Metodología y resultados de trigo en Sonora. Ciclo OI 2006-2007*. INIFAP, 2007.



## **COLDPower:** **El agente inteligente** **mexicano de** **mercadeo de energía**



Estimación de precios de vivienda mediante IA

El uso malicioso de la IA en Twitter

Interfaz gráfica del algoritmo basado en el forrajeo de bacterias

Conceptos básicos de RGN y PRGN

Módulo de rutas para la entrega de comida a través de un algoritmo PSO

Control de interferencia en una red de comunicaciones móviles





©Komputer Sapiens, Año XII Volumen II, mayo-agosto 2020, es una publicación cuatrimestral de la Sociedad Mexicana de Inteligencia Artificial, A.C., con domicilio en Ezequiel Montes 56 s/n, Fracc. los Pilares, Metepec, Edo. de México, C.P. 52159, México, <http://www.komputersapiens.org>, correo electrónico: [editorial@komputersapiens.org](mailto:editorial@komputersapiens.org), tel. +52 (833)357.48.20 ext. 3024, fax +52 (833) 215.85.44. Impresa por Sistemas y Diseños de México S.A.

de C.V., calle Aragón No. 190, colonia Álamos, delegación Benito Juárez, México D.F., C.P. 03400, México, se terminó de imprimir el 31 de agosto de 2020, este número consta de 1000 ejemplares.

Reserva de derechos al uso exclusivo número 04-2009-111110040200-102 otorgado por el Instituto Nacional de Derechos de Autor. ISSN 2007-0691.

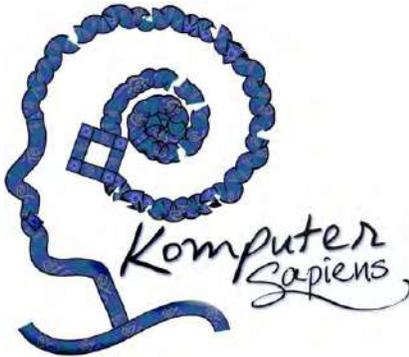
Los artículos y columnas firmados son responsabilidad exclusiva de los autores y no reflejan necesariamente los puntos de vista de la Sociedad Mexicana de Inteligencia Artificial. La mención de empresas o productos específicos en las páginas de Komputer Sapiens no implica su respaldo por la Sociedad Mexicana de Inteligencia Artificial.

Queda estrictamente prohibida la reproducción total o parcial por cualquier medio, de la información aquí contenida sin autorización por escrito de los editores.

Komputer Sapiens es una revista de divulgación en idioma español de temas relacionados con la inteligencia artificial. Creada en  $\text{\LaTeX}$ , con la clase **papertex** disponible en el repositorio *CTAN*: Comprehensive TeX Archive Network, <http://www.ctan.org/>

Indizada en el IRMDCT de CONACYT y en Latindex.

Presidente	<b>Directorio SMIA</b> Félix A. Castro Espinoza	<b>Directores Fundadores</b> Carlos Alberto Reyes García Ángel Kuri Morales	Jorge Rodas Osollo Grigori Sidorov
Vicepresidente	Ildar Batyrshin		Víctor Germán Sánchez Arias
Secretario	Oscar Herrera Alcántara		Aurora Torres
Tesorero	María de Lourdes G. Martínez Villaseñor	<b>Comité Editorial</b> Félix A. Castro Espinoza Jesús Favela Vara Sofía Natalia Galicia Haro Miguel González Mendoza Oscar Herrera Alcántara Raúl Monroy Borja Eduardo F. Morales Manzanares Leonardo Garrido Luna Carlos Alberto Reyes García Angélica Muñoz Meléndez Antonio Sánchez Aguilar Luis Enrique Sucar Succar Ángel Kuri Morales José A. Martínez Flores Juan Manuel Ahuactzin Larios Manuel Montes y Gómez Ofelia Cervantes Villagómez Alexander Gelbukh Grigori Sidorov Laura Cruz Reyes Elisa Schaeffer Ramon Brena Pinero Juan Humberto Sossa Azuela	Gregorio Toscano Pulido Tania Turrubiates López Esua Villatoro Tello Juan Carlos Osorio Gómez Julio Ponce
Vocales:	Sofía Natalia Galicia Haro Sabino Miranda Jiménez Roberto A. Vázquez Espinoza de los Monteros Hiram Erendin Ponce Espinosa Francisco Viveros-Jiménez Noé Alejandro Castro Sánchez Obdulia Pichardo Lagunas Omar Montaña Rivas Antonio Marín Hernández Néstor Velasco Bermeo Iris Iddaly Méndez Gurrola Gustavo Arroyo Figueroa Leobardo Morales Tiburcio Iván Olvera Rodríguez Juan Martínez Miranda		
Director general	<b>Komputer Sapiens</b> Félix A. Castro Espinoza	<b>Árbitros</b> Federico Alonso Pecina Giner Alor Hernández Ángel G. Andrade Héctor Avilés Ramón F. Brena Antonio Camarena Ibarrola Luis Castro Jaime Cerda Jacobo Ofelia Cervantes César Cárdenas Hugo Jair Escalante Luis Felipe Fernández Karina Figueroa Claudia Gómez Yasmín Hernández José Alberto Hernández Óscar Herrera José Francisco Martínez Trinidad Efrén Mezura Montes Daniela Moctezuma Raúl Monroy C. Alberto Ochoa Zezatti Francisco Javier Ornelas Nora Reyes	
Editora en jefe	Karina Mariela Figueroa Mora		
Editor científico	Héctor Gabriel Acosta Mesa		
Editores asociados	Marco Antonio Aguirre Lam Laura Cruz Reyes		
Coordinadora de producción e-Tlakuiló	Viridiana Mena Gómez Marco Antonio Aguirre Lam		
Estado del Arte	Reyna Carolina Medina Ramírez Jorge Rafael Gutiérrez Pulido Claudia G. Gómez Santillán		
Sakbe	Laura Cruz Reyes Laura Gómez Cruz		
IA & Educación	María Yasmín Hernández Pérez María Lucía Barrón Estrada J. Julieta Noguez Monroy		
Deskubriendo Konocimiento	Alejandro Guerra Hernández Leonardo Garrido Luna		
Asistencia técnica	Alan G. Aguirre Lam		
Corrección de estilo	Gilberto Rivera Zárate Marcela Quiroz Castellanos Ruth Esmeralda Barreda Guajardo Denisse Alvarado Guillermo de Jesús Hoyos Rivera Susana Aurora Velasco Montiel Nicandro Cruz Ramírez Enrique Gómez Cruz Claudia Leticia Díaz González Leonardo Romero Muñoz Sonia Lilia Mestizo Gutiérrez Nelson Rangel Valdez Laura Gómez Cruz Irene Morales Pagaza, Mopi Diseño		
Edición de imagen			
Portada			



## Contenido

### ARTÍCULO ACEPTADO

#### COLDPower: el agente inteligente mexicano de mercadeo de energía

Ansel Y. Rodríguez González, Eduardo Morales y Luis Enrique Sucar

pág. 6 ⇒ Introducción de conceptos básicos y muestras de resultados alcanzados en el Torneo PowerTAC.

### ARTÍCULO ACEPTADO

#### Estimación automática de precios de vivienda mediante inteligencia artificial

Roberto Quintero, Cynthia B. Perez, Laura P. López-Arredondo, Jesus R. Villavicencio, Cesar Puente y Juan Villegas

pág. 12 ⇒ Proceso para la estimación automática de precios de viviendas.

### ARTÍCULO ACEPTADO

#### Interfaz gráfica del algoritmo basado en el forrajeo de bacterias para optimización numérica con restricciones

J. Adrian García-López, Betania Hernandez-Ocaña, Oscar Chávez-Bosquez y José Hernández-Torruco

pág. 19 ⇒ Diseño de una interfaz gráfica para el algoritmo TS-MBFOA.

### ARTÍCULO ACEPTADO

#### El uso malicioso de la Inteligencia Artificial en Twitter

Israel Morales Trigueros y Carlos Adolfo Piña García

pág. 25 ⇒ El contenido que nos presenta Twitter representa un beneficio propagandista para aquellos que logran controlar el flujo de información.

### ARTÍCULO ACEPTADO

#### Algoritmos inspirados en la naturaleza para control de interferencia en una red de comunicaciones móviles celular

Jesús Alberto Gómez Avilés, Anabel Martínez-Vargas y Ángel G. Andrade

pág. 30 ⇒ Algoritmos inspirados en la naturaleza para una red móvil celular.

### ARTÍCULO ACEPTADO

#### Conceptos básicos de los generadores de números aleatorios (RGN) y pseudoaleatorios (PRGN)

Elena Fabiola Ruiz- Ledesma, Elizabeth Acosta-Gonzaga y Tonáhtiu Arturo Ramírez-Romero

pág. 35 ⇒ La generación de números aleatorios tiene un campo amplio de uso y la selección de un programa PRNGs depende de diversos factores.

### ARTÍCULO ACEPTADO

#### Módulo de rutas para la entrega de comida a través de un algoritmo Optimizado por enjambre de partículas (PSO) y con tecnología de Google maps

Sergio Soto Rodríguez, Edmundo Bonilla Huerta, Alberto Portilla Flores, José Crispín Hernández Hernández y Oscar Atriano Ponce

pág. 40 ⇒ Módulo para la optimización de rutas de reparto.

### Columnas

Sapiens Piensa.  
Editorial [pág. 2](#)

e-Tlakuilo [pág. 3](#)

Estado del IArte [pág. 4](#)

Sakbe [pág. 5](#)

IA & Educación [pág. 47](#)

Deskubriendo  
Konocimiento [pág. 49](#)

## ARTÍCULO ACEPTADO

# Estimación automática de precios de vivienda mediante inteligencia artificial

Roberto Quintero, Cynthia B. Pérez, Laura P. López-Arredondo, Jesús R. Villavicencio, Cesar Puente y Juan Villegas

## La Importancia de la valuación de viviendas en México

En la actualidad, la industria inmobiliaria presenta un constante crecimiento en el mercado, ya que resulta esencial para diversos sectores económicos alrededor del mundo contar con tasaciones precisas de bienes inmuebles [1,2]; para esto, el proceso de valuación de viviendas consiste en establecer el valor que le corresponde tomando en cuenta distintos factores o atributos de la propiedad. Sin embargo, llevar a cabo la estimación del precio de las viviendas de manera acertada y precisa puede resultar bastante complejo debido a la variabilidad de los factores. Es por ello que la naturaleza volátil del mercado inmobiliario ha generado la necesidad de mejorar los métodos para valuar los precios de los inmuebles, involucrando características de viviendas, modelos estadísticos y predictivos confiables donde se busca reducir al máximo el error de las estimaciones, es decir, de la valuación del bien inmueble.

Para denotar su importancia global basta con analizar algunas cifras obtenidas de diversos estudios inmobiliarios realizados en distintos países [3]. En Estados Unidos, se ha registrado un incremento promedio anual del 2.2% del PIB (Producto Interno Bruto) desde el tercer trimestre de 2009. En Filipinas, el PIB presentó un crecimiento del 6.7% en 2016, lo cual representó un incremento bastante acelerado con respecto a los años anteriores, ya que muchos inversionistas alrededor del mundo han comprado diversas propiedades en todas las islas [4]. En México, la participación de los servicios inmobiliarios en la economía ha ido en aumento y en 2019 ya representaba el 11% del PIB total [5]. Así mismo, para finales de 2017 hubo un notable incremento en la construcción de viviendas multifamiliares dentro del núcleo urbano de las ciudades, correspondiente al 23% de todas las unidades en construcción y posicionando las mejores oportunidades de desarrollo en los suburbios [6]. De igual manera, existen muchos otros países en donde la industria y los proyectos inmobiliarios juegan un rol muy importante en la economía. Un estudio reciente [7] analiza el crecimiento de esta industria en Beijín, China, acorde a un modelo multivariable para poder hacer predicciones de precios en bienes raíces ante situaciones cambiantes y complejas del mercado. Debido a la importancia que sigue generando el mercado inmobiliario

en la economía de los países ha surgido la necesidad de implementar métodos de valuación de bienes inmuebles, que permitan estimar los precios de viviendas de forma rápida y precisa a través de plataformas tecnológicas. En la revista Expansión [8], se argumenta que el rol de los valuadores se ha complementado por el interés de las empresas de tecnologías en los procesos inmobiliarios. Por lo tanto, el estudio de modelos predictivos, así como el uso de tecnologías y plataformas que implementan algoritmos de aprendizaje automático para analizar cientos de datos sobre cada propiedad ha ganado popularidad, generando nuevas oportunidades para el crecimiento económico, la innovación y la construcción del contexto inmobiliario. Muestra de ello es que, en los últimos años en México, más del 50% de las búsquedas inmobiliarias fueron realizadas a través de plataformas de Internet pasando de 21 millones a 42 millones de búsquedas [4]. Lo anterior se debe a que dichas tecnologías sirven como apoyo a los propietarios de inmuebles, agentes de bienes raíces, bancos, instituciones gubernamentales, entre otros, para conocer el valor estimado de las viviendas en tiempo real a través de la automatización basada en herramientas de cómputo, marcando una tendencia clara hacia la implementación de la Inteligencia Artificial (IA).

## Estimación automática de precios de viviendas

La dificultad más relevante en el pronóstico de los precios inmobiliarios siempre ha sido la subjetividad de los juicios a procesar para obtener valores inmobiliarios acertados y confiables [9]. En ese sentido, Esquer [10] menciona que los métodos manuales de valuación suelen ser bastante lentos y tediosos debido a que requieren de una investigación fundamentada que suele tomar tiempo, además de la participación de personas capacitadas en valuación de viviendas. Schram (2006) y Diaz & Hansz (2002) [11,12] explican ocho pasos necesarios para llevar a cabo la valuación de viviendas de forma tradicional (ver Figura 1). En general, los procedimientos actuales deben basarse en información confiable para obtener resultados más precisos.

En Estados Unidos, los emprendedores han intentado incrementar la eficiencia de los corredores hipotecarios disminuyendo costos de transacción y el acceso a la información de los mercados. Esto conduce a la necesidad

de incorporar herramientas tecnológicas durante el proceso de valuación de tal forma, que permitan interpretar los fenómenos típicos del mercado inmobiliario [13,14]. De esta manera, los datos son transformados en conocimiento mediante el uso de inteligencia artificial con el fin de automatizar los procesos y visualizar las estimaciones en tiempo real. Es por ello que la implementación de algoritmos predictivos y de aprendizaje automático en el contexto de bienes raíces ha tenido un notable crecimiento en la industria inmobiliaria extranjera. No obstante, la búsqueda por robustecer el ajuste de precios de viviendas disminuyendo el margen de error en dichas predicciones es continuo, esto con la finalidad de ofrecer estimaciones más acertadas y cercanas a la realidad sobre el valor de los bienes inmuebles. En consecuencia, existen diversos estudios en la literatura que analizan la precisión de las estimaciones haciendo uso de los algoritmos de aprendizaje automático. Entre los más comunes destacan las redes neuronales artificiales (RNA) [15-18], los modelos de regresión [19], las máquinas de soporte vectorial [20], y los algoritmos evolutivos [9], entre otros. Para que este tipo de técnicas obtengan una estimación certera es crucial la selección adecuada de los factores implicados en la valoración de viviendas. Actualmente, el estudio de tecnologías y plataformas que implementan algoritmos predictivos y de aprendizaje automático para estimar el valor de propiedades es de gran interés para este mercado; aunque, son pocas las empresas inmobiliarias en México que hacen uso de este tipo de tecnología, así como escasos los estudios que realizan un análisis previo sobre la importancia de los factores involucrados en la estimación de precios de viviendas [21,22].

Es por ello que este trabajo se enfoca en analizar los principales factores para la estimación de precios de viviendas mediante el uso de herramientas de inteligencia artificial, utilizando solo aquellos factores predominantes sin afectar la precisión de la estimación en el contexto del mercado inmobiliario en México. Lo anterior, ayudaría a establecer la estimación automática por medio de una plataforma web donde el usuario (comprador, vendedor, agente) introduciría solamente la información de los factores más relevantes y con ello, se desplegaría automáticamente el precio que le correspondería a esa vivienda en particular.

El análisis del mercado inmobiliario no solo es relevante para las estimaciones de bienes raíces, sino también para las inversiones y cuestiones fiscales [9]. Es por ello, que existe un gran interés por el uso de algoritmos de inteligencia artificial, ya que ayudan a interpretar los fenómenos típicos del mercado inmobiliario reduciendo el tiempo del proceso de estimación de bienes inmuebles. Además de poder visualizar los resultados de las estima-

ciones desde una plataforma web a través de cualquier dispositivo con conexión a Internet. Para esto, algunas plataformas estadounidenses como Intelimétrica<sup>1</sup>, Redfin<sup>2</sup> y Zillow<sup>3</sup>, están revolucionando el mercado inmobiliario aplicando una filosofía de análisis intensivo de datos mediante el uso de algoritmos de inteligencia artificial para democratizar el acceso a la información facilitando la toma de decisiones.



Figura 1. Proceso de valuación tradicional en bienes raíces [11,12].

### Selección de factores para la estimación de precios de viviendas

El proceso relacionado con la estimación del precio de venta de las viviendas agrupa diversas características o factores relevantes que influyen directamente en la estimación del valor del bien inmueble. Esta información corresponde, en su mayoría, a los factores o atributos principales de la vivienda como aquellos relacionados con: la ubicación, el tipo de propiedad, las condiciones de la vivienda, las proporciones de la vivienda, sus instalaciones

<sup>1</sup>www.intelimetrica.com

<sup>2</sup>www.redfin.com

<sup>3</sup>www.zillow.com

y los sitios cercanos a la vivienda (ver Tabla 1). De manera más específica, Komagome-Towne [23], menciona que las principales características utilizadas para la predicción de los precios corresponden a los pies cuadrados de la vivienda, el tamaño del lote, el número de habitaciones, el número de baños, el año de construcción de la casa y la fecha de venta. No obstante, muchos otros factores pueden llegar a afectar en mayor o menor medida las estimaciones del precio de los inmuebles.

Tabla 1. Descripción de los factores.

Factor	Descripción de los factores
ID	Identificador del registro del avalúo de la vivienda.
Código Postal	Código postal de la vivienda
Calle	Calle en la que se sitúa la vivienda
Lote	Descripción del lote si es que lo requiere
Manzana	Descripción o nombre de la manzana
Colonia	Nombre de la colonia
Valor concluido	Representa el valor que definieron los valuadores de la empresa inmobiliaria
Numero	Número de casa
Clase	Nivel del interés social
Tipo Inmueble	Si se trata de departamento, casa, etc.
Años	Años de la vivienda
Vida Útil Remanente (VUR)	Periodo de tiempo probable, expresado en años, que se estima funcionará la vivienda
Recamaras	Número de recámaras
Baños	Número de baños
Niveles	Número de pisos
Estacionamientos	Si cuenta con estacionamiento y cuántos tiene
Superficie Terreno	Superficie en metros cuadrados
Superficie Construida	Superficie en metros cuadrados construido
Superficie Accesoría	Superficie de construcción en metros cuadrados (área no habitable: cocheras, porches, pórticos, terrazas)
Superficie Vendible	Total de metros cuadrados de la construcción (área habitable)
Valor Físico del Terreno	Valor en pesos (MXN) del terreno
Valor Físico Construido	Valor en pesos (MXN) de lo construido
Valor del Mercado	Valor en pesos (MXN) de acuerdo al criterio del mercado
Valor Unitario Homologado de Ventas	Valor unitario aplicable con afectación según los comparables de mercado (la afectación se deriva de la comparación del sujeto con los comparables según la edad, estado de conservación, ubicación, superficie de construcción etc.)
Valor Unitario Homologado del Terreno	Valor unitario aplicable con afectación según los comparables de mercado (la afectación se deriva de la comparación del sujeto con los comparables según la ubicación y deseabilidad del terreno etc.)
Valor Reposición Nuevo	Es el valor por metros cuadrado de construcción en estado de conservación nuevo. Se aplica según el tipo de construcción
Fecha Avalúo	Fecha en que se realizó el avalúo
Latitud	Latitud de la ubicación
Longitud	Longitud de la ubicación

En los últimos años, las técnicas predictivas han sido de gran utilidad para ayudar a facilitar y precisar la valoración de viviendas, así como conocer cuántos y cuáles factores considerar para la estimación del inmueble. Para esto, la IA provee de herramientas que ayudan a automatizar el proceso de selección de factores reduciendo

su número a aquellos que tengan un mayor impacto en la estimación de precios sin afectar su precisión. En ese sentido, existen diversos métodos para la selección de factores mediante la reducción de la dimensionalidad [24]. En este trabajo se utilizaron cuatro métodos: la técnica de filtrado (FM, Filter Method), envoltura por eliminación hacia atrás (WM-BE, Wrapper Method-Backward Elimination), envoltura por eliminación recursiva (WM-RFE, Wrapper Method-Recursive Feature Elimination) y método embebido (EM, Embedded Method). Para esto, la empresa inmobiliaria cuenta con 29 factores en su base de datos, los cuales describen las características de 1842 viviendas ubicadas en Ciudad Obregón, Sonora (ver Tabla 1). De estos 29 factores se redujo a 11 mediante la aplicación de los métodos mencionados anteriormente (ver Tabla 2). Si bien es deseable que la variable tiempo u otra relacionada con la depreciación del inmueble fuera considerada de forma explícita, ésta es contemplada en el cálculo del factor “valor del mercado”, presentado en la Tabla 1, la cual es calculada por el personal experto en el área de valuación de la empresa bajo estudio.

Tabla 2. Comparación de la selección de factores.

Factores	FM	WM-BE	WM-RFE	EM	Frecuencia de factores seleccionados
ID					
Código Postal		1	1		2
Calle					
Lote					
Manzana					
Colonia					
Tipo Inmueble					
Clase					
Años		1	1		2
Vida Útil Remanente (VUR)		1			1
Recamaras	1	1	1		3
Baños	1		1		2
Niveles			1		1
Estacionamientos			1		1
Superficie Terreno	1	1			2
Superficie Construida		1			1
Superficie Accesoría		1	1		2
Superficie Vendible	1	1	1		3
Valor Físico del Terreno	1	1			2
Valor Físico Construido		1			1
Valor del Mercado	1	1		1	3
Valor Unitario Homologado		1	1		2
Valor Unitario Homologado del Terreno	1	1	1		3
Valor Reposición Nuevo		1			1
Fecha Avalúo					
Latitud					
Longitud					

Como el número de factores es muy grande para establecer automáticamente el precio que le corresponde a cada una de las viviendas, es necesario reducir su dimensionalidad utilizando algoritmos de IA como el método de filtrado, de envoltura y embebido. Al hablar de reducción de dimensionalidad se entiende desde dos

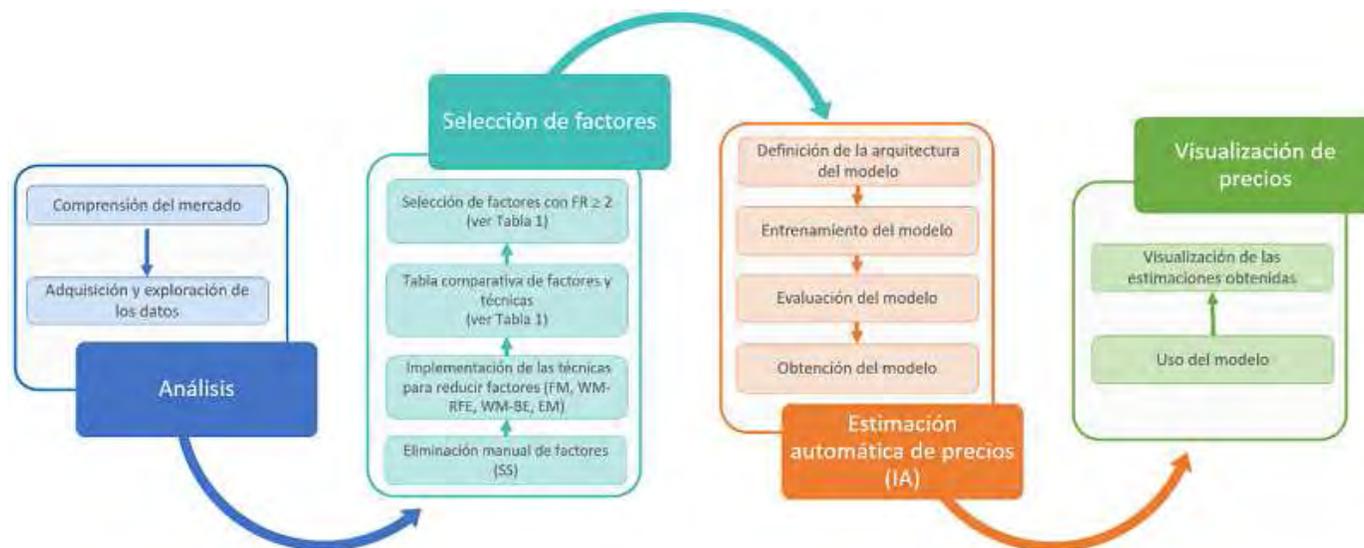


Figura 2. Proceso de estimación automática de precios de viviendas.

perspectivas, la primera es aplicar algoritmos del tipo PCA donde se hace una reducción de variables analizando las más representativas y a partir de ellas rehacer o declarar un nuevo conjunto de variables reducido; la segunda perspectiva es reducir la cantidad de variables por discriminación de representatividad, sin alterar los valores de las variables seleccionadas. En nuestra propuesta estamos considerando el segundo enfoque [25]. Para esto, se siguió la metodología que se presenta en la Figura 2, donde primero fue necesario llevar a cabo un análisis y comprensión del mercado inmobiliario, así como la exploración de los datos. Posteriormente, se realiza la selección de factores donde es necesario que primero se eliminen manualmente aquellos factores que contengan información nula o inconsistente. Una vez preprocesada la información se aplican cuatro métodos de reducción de dimensionalidad (FM, WM-BE, WM-RFE, EM), para seleccionar aquellos factores relevantes que serán utilizados en la estimación de precio. Esta selección se lleva a cabo mediante la comparación del número de factores seleccionados por cada uno de los métodos, donde la frecuencia sea igual o mayor a dos factores (ver Tabla 2). Por lo tanto, los factores con mayor influencia fueron: código postal, años, recámaras, baños, superficie del terreno, superficie accesoria, superficie vendible, valor físico del terreno, valor de mercado, valor unitario homologado y valor unitario homologado del terreno. Así mismo, el método de reducción de dimensionalidad seleccionado fue el WM ya que, en conjunto, el WM-BE y WM-RFE, es donde se presentaron la mayor cantidad de factores.

### Estimación automática de precios mediante redes neuronales artificiales (RNA)

La estimación automática de precios se llevó a cabo mediante el uso de RNAs (ver Figura 2). Para esto, los datos registrados en la base de datos se dividieron para la etapa de entrenamiento y para la etapa de prueba. En total son 1842 datos de los cuales, el 80 % fue utilizado para entrenamiento y el 20 % para prueba, con una arquitectura en la red como se muestra en la Tabla 3. Un ejemplo de los resultados obtenidos se presenta en la Tabla 4, ahí se observa la estimación del precio de 10 viviendas seleccionadas aleatoriamente de las 368 correspondientes al 20 % de pruebas. Estos resultados muestran la estimación obtenida mediante RNAs y la estimación obtenida por medio de los expertos en bienes raíces; en promedio la valoración realizada por el algoritmo de red neuronal tiene una correlación (R2) o efectividad del 95 % con un error del 5 % para todos los datos de prueba, lo que significa que el algoritmo puede dar una estimación con un error en promedio del 5 % con respecto al valor real de la propiedad (ver Tabla 5).

Tabla 3. Arquitectura de la RNA.

Característica	Descripción
No. de Neuronas por Capas	10-7-1
Función	ReLU
Método de optimización	L-BFGS-B
Épocas	600
Tasa de aprendizaje ( $\alpha$ )	0.0001

Tabla 4. Resultados de la estimación de precios de viviendas mediante redes neuronales.

Estimación del precio de viviendas			
	RNA	Valor Concluido	Diferencia
1	1983339.8	1905200	78139.8
2	655326.9	655190	136.9
3	972386.8	972790	-403.2
4	266524.9	265680	844.9
5	690236.8	690410	-173.2
6	1804916.9	1804900	16.9
7	2595231.3	2512300	82931.3
8	880107.9	880600	-492.1
9	523908.7	523700	208.7
10	337977.9	337900	77.9

Tabla 5. Resultados del error en la estimación de precios de viviendas mediante redes neuronales.

Modelo	MSE	RMSE	MAE	R2
RNA para la estimación de precios	60688.124	246.350	114.492	0.95
MSE= Error cuadrático Medio (Mean Squared Error) RMSE= Raíz del Error cuadrático Medio (Root Mean Squared Error) MAE= Error Absoluto Medio (Mean Absolute Error) R2= Correlación				

## Conclusiones

Este artículo presenta el proceso para la estimación automática de precios de viviendas mediante RNA. Para ello, se colaboró con una empresa de bienes raíces ubicada en Ciudad Obregón, Sonora, con más de 20 años de experiencia. En ese sentido, primero fue necesario seleccionar los factores más relevantes mediante cuatro métodos de reducción de dimensionalidad para llevar a cabo la estimación de forma automática mediante redes neuronales. Se entrenó la red neuronal con un total de 1474 datos históricos de la empresa para posteriormente probar el modelo con 368 datos.

Los resultados confirman que es posible llevar a cabo la estimación automática de precios de viviendas ya que el porcentaje de error máximo permitido por la empresa es del 10 % y se obtuvo mediante la técnica de IA un error del 5 % muy por debajo del establecido por la empresa. Adicionalmente, el tiempo de estimación fue reducido de 72 hrs a 2 segundos, tomando en cuenta el tiempo que tardó la plataforma al usar el modelo obtenido por la red neuronal y presentar la estimación del inmueble. Así mismo, se está trabajando en la implementación de una plataforma geovisual interactiva la cual utiliza el algoritmo de RNA descrito para estimar automáticamente los precios de viviendas. Se espera que esta plataforma pueda ser utilizada por otras inmobiliarias para estimar automáticamente los precios de viviendas reduciendo con ello, costos y tiempos requeridos en este proceso.\*

**Agradecimientos.** Publicación financiada con recursos PFCE 2019.

## REFERENCIAS

- Ghysels, E., Plazzi, A., Valkanov, R., y Torous, W. (2013). Forecasting Real Estate Prices. *Handbook of Economic Forecasting*, 509-580.
- The Economist (2019). For now, residential-property prices are likely to keep rising. Recuperado el 22 de octubre de 2019, de <http://www.economist.com/graphic-detail/2019/06/29/for-now-residential-property-prices-are-likely-to-keep-rising>.
- Guo, H., Jiang, H., Zhang, W., Cui, F., Zhang, R., He, Z., Lun, Y., Chen, Q., y Chen, Q. (2019). Real Estate Level Forecasting - Review. *Business, Management and Economics Research*. 5(54), 57-61.
- Lamudi. (2017). Informe del Mercado Inmobiliario 2017: Internet, penetración y tendencias. México. Recuperado el 8 de octubre de 2018, de <https://www.lamudi.com.mx/informe-mercado-inmobiliario-2017>.
- Balbuena, F., Vázquez S., y Serrano C. BBVA Research (2019). Situación Inmobiliaria México. Primer semestre 2019. Recuperado el 8 de octubre de 2019, de <https://www.bbvaesearch.com/publicaciones/situacion-inmobiliaria-mexico-primer-semestre-2019/>.
- CBRE Research. (2018). 2018 Real Estate Market Outlook - United States. Recuperado de <https://www.cbre-ea.com/docs/default-source/default-document-library/2018-u-s-real-estate-market-outlook.pdf>.
- Li, Y., Xiang, Z., y Xiong, T. (2020). The Behavioral Mechanism and Forecasting of Beijing Housing Prices from a Multiscale Perspective. *Discrete Dynamics in Nature and Society*, 2020, 1-13.
- Expansión (2017). Los start-ups que saben, exactamente, cuánto cuesta tu casa. Recuperado el 4 de marzo de 2017, de <https://www.pressreader.com/mexico/expansion-m%C3%A9xico/20170115/281621010028736>.
- Del Giudice, V., De Paola, P., y Forte, F. (2017). Using Genetic Algorithms for Real Estate Appraisals. *Buildings*, 7(2), 1-12.
- Esquer, J.A. (2012). Estimación del precio de venta de la vivienda en la ciudad de Hermosillo aplicando el método de precios hedónicos. Tesis de maestría en Ciencias de la Ingeniería. Universidad de Sonora.
- Schram, J.F. (2006). Real Estate Appraisal. Rockwell Publishing.
- Diaz, J., y Hansz, J. A. (2002). Behavioral Research into the Real Estate Valuation Process: Progress Toward a Descriptive Model. En *Real Estate Valuation Theory*, 3-29.
- Manganelli, B., De Mare, G., y Nesticò, A. (2015). Using genetic algorithms in the housing market analysis. En *International Conference on Computational Science and Its Applications*, 36-45.
- Muralidharan, S., Phiri, K., Sinha, S.K., y Kim, B. (2018). Analysis and prediction of real estate prices: A case of the Boston housing market. *Issues in Information Systems*, 19(2), 109-118.
- Bency, A.J., Rallapalli, S., Ganti, R.K., Srivatsa, M., y Manjunath, B.S. (2017). Beyond Spatial Auto-Regressive Models: Predicting Housing Prices with Satellite Imagery. En *IEEE winter conference on Applications of Computer Vision (WACV)*. Santa Rosa, California, USA, 320-329.
- Han, S., Ko, Y., Kim, J., y Hong, T. (2018). Housing Market Trend Forecasts through Statistical Comparisons based on Big Data Analytic Methods. *Journal of Management in Engineering*, 34(2), 1-12.

17. Wang, L., Chan, F. F., Wang, Y., y Chang, Q. (2016). Predicting public housing prices using delayed neural networks. En *IEEE Region 10 Conference (TENCON)*, Marina Bay Sands, Singapore, 3589-3592.
18. Lim, W. T., Wang, L., Wang, Y., y Chang, Q. (2016). Housing price prediction using neural networks. En *IEEE 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. Changsha, China, 518-522.
19. Ozgur, C., Hughes, Z., Rogers, G., y Parveen, S. (2016). Multiple Linear Regression Applications in Real Estate Pricing. *International Journal of Mathematics and Statistics Invention (IJMSI)*, 4(8), 39-50.
20. Mu, J., Wu, F., y Zhang, A. (2014). Housing value forecasting based on machine learning methods. En *Abstract and Applied Analysis*, 2014. Hindawi.
21. Ma, Y., Zhang, Z., Ihler, A., y Pan, B. (2018). Estimating Warehouse Rental Price using Machine Learning Techniques. *International Journal of Computers, Communications & Control*, 13(2).
22. Lopez-Arredondo, L. P., Perez, C. B., Castro, L. A., y Rodriguez, L. F. (2019). Estudio sobre la Percepción de los Factores Involucrados en la Estimación de Precios de Viviendas: El Caso de Cajeme. *Información tecnológica*, 30(2), 11-24.
23. Komagome-Towne, A. (2017). Models and Visualizations for Housing Price Prediction. Tesis Doctoral, California State Polytechnic University, Pomona, USA.
24. Guyon, I., y Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.
25. Fukunaga, K. (1990). Introduction to statistical pattern recognition. *Academic Press Professional*. San Diego, CA, USA.

## SOBRE LOS AUTORES



**Roberto Quintero** es estudiante de la Maestría en Tecnologías de la Información para los Negocios del Instituto Tecnológico de Sonora. Sus intereses de investigación incluyen aprendizaje automático y ciencia de datos.



**Cynthia B. Pérez** obtuvo el grado de Doctor en Ciencias de la Computación en la Universidad de Extremadura en Mérida, España en 2014. Actualmente es profesor investigador en el Instituto Tecnológico de Sonora, Unidad Guaymas. La Dra. Pérez es miembro del Sistema Nacional de Investigadores y participa como revisor de artículos científicos en el área de Inteligencia Artificial. Sus intereses de investigación incluyen aprendizaje automático, cómputo evolutivo, visión por computadora e inteligencia de negocios.



**Laura López-Arredondo** es maestra en Tecnologías de la Información para los Negocios por el Instituto Tecnológico de Sonora. Actualmente trabaja como ingeniero de pruebas y calidad en Novutek SC, Ciudad Obregón, Sonora.



**Jesús R. Villavicencio** obtuvo el grado de Doctor en Ciencias en Electrónica y Telecomunicaciones por el Centro de Investigación Científica y de Educación Superior de Ensenada en el 2015. Actualmente es profesor en el Instituto Tecnológico de Sonora, Unidad Guaymas. Sus intereses de investigación incluyen aprendizaje automático, internet de las cosas y procesamiento de señales.



**Cesar Puente** obtuvo el grado de Doctor en Ciencias de la Tierra por el Centro de Investigación Científica y de Educación Superior de Ensenada en el 2011. Actualmente es profesor investigador en la Universidad Autónoma de San Luis Potosí. El Dr. Puente participa como revisor de artículos científicos en el área de inteligencia artificial y percepción remota. Sus intereses de investigación incluyen reconocimiento de patrones, visión computacional, cómputo evolutivo y percepción remota.