

# Un enfoque de Análisis Inteligente de Datos para Apoyar la Relación con los Clientes

Oswaldo Solarte Pabón<sup>1,2</sup>, José Heriberto Torres<sup>2</sup>, Víctor Andrés Bucheli<sup>2</sup>

**oswaldo.solartep@alumnos.upm.es; jose.heriberto.torres@correounivalle.edu.co;  
victor.bucheli@correounivalle.edu.co**

<sup>1</sup> Universidad Politécnica de Madrid, Campus de Montegancedo, 28660, Madrid España.

<sup>2</sup> Universidad del Valle, Calle 13 # 100-00, 760032, Cali, Valle del Cauca, Colombia

**DOI: 10.17013/risti.39.52-66**

**Resumen:** El análisis inteligente de datos es un componente importante en la cuarta revolución industrial. En particular, los datos generados por los clientes se pueden usar para mejorar los productos y servicios. Sin embargo, el análisis del cliente se ha enfocado más hacia el uso de datos estructurados. La mayoría de las empresas aún no aprovechan todo el potencial que ofrecen las fuentes de datos no estructuradas. Este artículo propone un enfoque para extraer indicadores de desempeño del servicio al cliente integrando técnicas de inteligencia de negocios y procesamiento del lenguaje natural (NLP), con el propósito de aprovechar fuentes de datos no estructuradas. Para validar la propuesta se realizó un caso de estudio en una empresa que ofrece servicios de salud. Se obtuvieron indicadores a partir del análisis de llamadas telefónicas y correos electrónicos escritos en español. Los resultados obtenidos muestran la utilidad del enfoque para apoyar las relaciones con los clientes.

**Palabras-clave:** Inteligencia de Negocios; Procesamiento del Lenguaje Natural (NLP); Gestión de relaciones con los clientes (CRM); Análisis inteligente de datos.

## *An Intelligent Data Analysis Approach to Support Customer Relationship*

**Abstract:** Intelligent data analysis is an important component of the fourth industrial revolution. In particular, data generated by customers can be used to improve products and services. However, customer analysis has focused more on the use of structured data. Most companies are still not taking advantage of the full potential of unstructured data sources. This paper proposes an approach to extract customer service performance indicators integrating business intelligence and natural language processing (NLP) techniques to take advantage of unstructured data. To validate the proposal, a study case was carried out in a company that offers health services. Indicators were obtained from the analysis of phone calls and emails written in Spanish. The results obtained show the utility of the approach to support customer relationships.

**Key words:** Business Intelligence; Natural Language Processing (NLP); Customer Relationship Management (CRM); Smart data analysis.

## 1. Introducción

El análisis de datos es un componente fundamental en la cuarta revolución industrial ya que la generación de grandes volúmenes de datos está creando oportunidades para mejorar la eficiencia de los sistemas productivos (Mittal, Khan, Romero, & Wuest, 2019), (Peres, Rocha, Leitao & Barata, 2018). Adicionalmente, en esta nueva revolución los clientes y socios comerciales se integran a los procesos productivos con el fin de aportar valor agregado a las industrias (Thoben, Wiesner & Wuest, 2017).

En ese sentido, uno de los grandes desafíos para las empresas es brindar una experiencia positiva a sus clientes (Hassan, Nawaz, Lashari & Zafar, 2015). Una de las estrategias más usadas para mantener las relaciones con los clientes es la utilización de sistemas CRM (*Customer Relationship Management*). Además, la integración de varios canales de comunicación y la sincronización de diferentes fuentes de datos crea oportunidades y retos para la mejorar los sistemas CRM (Picek, Peras, & Mekovec, 2018).

Con el objetivo de apoyar las relaciones con los clientes, se han presentado varias propuestas que integran capacidades de análisis de datos a los sistemas CRM, tal como se describe en (Rodrigues Chagas et al., 2019) y (Soltani & Navimipour, 2016). Sin embargo, muchas propuestas se han enfocado solo en el análisis de datos estructurados. Las empresas aun no aprovechan todo el potencial que ofrecen las fuentes de datos no estructuradas. De acuerdo con Müller, Debortoli, Junglas & Vom-Brocke, (2016), el 80% de los datos que se generan actualmente son datos no estructurados y frecuentemente se basan en texto. Adicionalmente, las propuestas que explotan fuentes textuales en su mayoría se han enfocado en el idioma inglés. Además, la extracción de información también representa sus propios desafíos en otros idiomas, además del inglés (Névéol, Dalianis, Velupillai, Savova, & Zweigenbaum, 2018).

En este artículo se propone un enfoque para extraer indicadores de desempeño del servicio al cliente integrando técnicas de inteligencia de negocios y procesamiento del lenguaje natural (NLP). Este enfoque saca ventaja de fuentes de datos estructuradas y de fuentes no estructuradas como documentos de texto escritos en español. El enfoque permite extraer información de productos, servicios, preferencias y las opiniones que el cliente tiene sobre estos. Los indicadores obtenidos se integran en un esquema dimensional orientado al soporte de consultas para el análisis de datos. Para validar la propuesta se aplicó un caso de estudio en una empresa que ofrece una red de servicios de salud para conectar a los consumidores y productores de estos servicios.

El resto de este documento está organizado de la siguiente manera: la Sección 2 describe algunos trabajos relacionados. En la sección 3 se presenta detalladamente el enfoque propuesto. En la Sección 4 se describe un caso de estudio que se realizó para validar la utilidad del enfoque. Finalmente, la Sección 5 incluye conclusiones y trabajos futuros.

## 2. Trabajos relacionados

En este artículo las propuestas que usan análisis de datos para apoyar las relaciones con los clientes se agrupan en tres categorías: las que integran sistemas CRM con bodegas de datos, CRM con Minería de datos y las que integran CRM con técnicas de NLP.

### 2.1. Integración de sistemas CRM con bodegas de datos

La integración de bodegas de datos con los sistemas CRM proporciona información integrada que facilita el análisis de datos del cliente. Una de las primeras propuestas para integrar los sistemas CRM con una bodega de datos se describe en (Cunningham, Song & Chen, 2004), donde se propone el uso del modelo dimensional para crear herramientas que soporten consultas del servicio al cliente. Continuando con el trabajo anterior, en (Cunningham & Song 2007) se proponen los requerimientos que debe cumplir una bodega de datos orientada al análisis de datos del cliente. De acuerdo con Khan, Rao & Shahzad (2019), la integración de los sistemas CRM con una bodega de datos reduce el costo de adquirir nuevos clientes y del proceso de ventas. Una bodega de datos orientada al cliente crea ventaja competitiva ya que aumenta la cantidad de clientes y la tasa de retención (Handzic, Ozlen & Durmic, 2014).

En Ishaya & Folarin (2012), se propone un enfoque para integrar varios sistemas operacionales con información de los clientes en una empresa de telecomunicaciones. En este enfoque, la información se analiza en tres pasos: la identificación del perfil de los clientes, segmentación y la identificación de campañas de marketing. Similarmente, en Maji & Sen (2016), se propone un enfoque basado en inteligencia de negocios para analizar el impacto que tienen las campañas de marketing en los clientes.

Por su parte, en Rodríguez, Flores & La Rosa (2016) se propone un enfoque de inteligencia de negocios para una empresa que distribuye y comercializa refrescos. El enfoque permite obtener indicadores comerciales para conocer las zonas y rutas donde se realizan más ventas y así poder asignar los vendedores de una forma más eficiente.

### 2.2. Integración de técnicas minería de datos y sistemas CRM

En los últimos años el uso de técnicas de minería de datos para analizar información de los clientes se ha expandido rápidamente (Rodrigues Chagas et al., 2019). En Farajian & Mohammadi (2010) se propone una estrategia para analizar el comportamiento del cliente usando algoritmos de *clustering* y reglas de asociación en el sector financiero. De manera similar, en Devi & Rajagopalan (2012) se analiza el comportamiento del cliente en el proceso de ventas.

En Kadiyala & Srivastava (2011) se propone un marco de trabajo para identificar las técnicas de minería de datos relacionadas con los sistemas CRM. Por su parte Ahn, Ahn, Oh, & Kim, (2011) proponen un modelo de clasificación para apoyar la venta cruzada en el dominio de las telecomunicaciones móviles. De acuerdo con Rahman & Kan (2017), las técnicas de minería de datos son usadas comúnmente por las empresas de telecomunicaciones, ya que este es un mercado donde los clientes se pueden cambiar

fácilmente de operador. Finalmente, en Bahari & Elayidom (2015), se propone un modelo predictivo aumentar la retención de clientes en el sector financiero.

### **2.3 Integración de técnicas de NLP y sistemas CRM**

Con el objetivo de extender el análisis de datos a fuentes no estructuradas, en los últimos años se han propuesto enfoques que integran técnicas de NLP con los sistemas CRM. En Gutfreund (2017) se propone un enfoque que usa minería de texto para analizar la gestión de relaciones con clientes de Elsevier<sup>1</sup>. Este enfoque busca predecir y reducir el abandono prematuro de los clientes. Por su parte, en Müller et al., (2016) se aborda el concepto de Social CRM (SCRM), que se refiere a la capacidad de integrar información de redes sociales con los sistemas CRM. En esta propuesta se desarrolló un prototipo para el servicio de tecnologías de información de la universidad de la Florida. El prototipo analiza descripciones textuales de casos reportados por los usuarios a través correo electrónico, llamadas telefónicas y redes sociales.

Una propuesta similar se describe en Zhou et al. (2018), donde se presenta un modelo para analizar opiniones que hacen los clientes a través de aplicaciones móviles de diferentes productos. El modelo sugiere las opiniones más relevantes que los desarrolladores de productos deberían tener para ajustar sus diseños y productos a las necesidades de los clientes. De igual manera, en Khan, Rao & Shahzad (2019) analizan datos de Twitter para extraer opiniones acerca una compañía aérea. Por último, en Kauffmann et al. (2019) se extrae opiniones de los clientes de comentarios escritos en Amazon.

La Figura 1 muestra un resumen de las propuestas descritas anteriormente. Para cada propuesta se muestra el proceso que se quería analizar (segmentación de clientes, retención de clientes, adquirir clientes, etc) y el dominio (telecomunicaciones, financiero, transporte, etc).

## **3. Enfoque para apoyar la relación con los clientes**

La Figura 2 muestra el enfoque propuesto para apoyar la relación con los clientes. Este enfoque combina técnicas de inteligencia de negocios y procesamiento del lenguaje natural para extraer indicadores de desempeño de diferentes fuentes de datos. Este enfoque está formado por cuatro componentes: Selección de fuentes de datos, Procesamiento de datos estructurados, Procesamiento de datos no estructurados y el componente de Integración de datos.

### **3.1. Seleccionar fuentes de datos**

Este componente permite la identificación de fuentes de datos tanto estructuradas como no estructuradas que se han generado a partir de la comunicación con los clientes. Los datos estructurados provienen principalmente de las bases de datos relacionales asociadas al CRM de la organización, hojas de cálculo u otros archivos estructurados. Para los datos no estructurados esta propuesta se enfoca en analizar información escrita en lenguaje natural como correos electrónicos, transcripción de registros de llamadas telefónicas, y otros documentos de texto escritos en español.

---

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

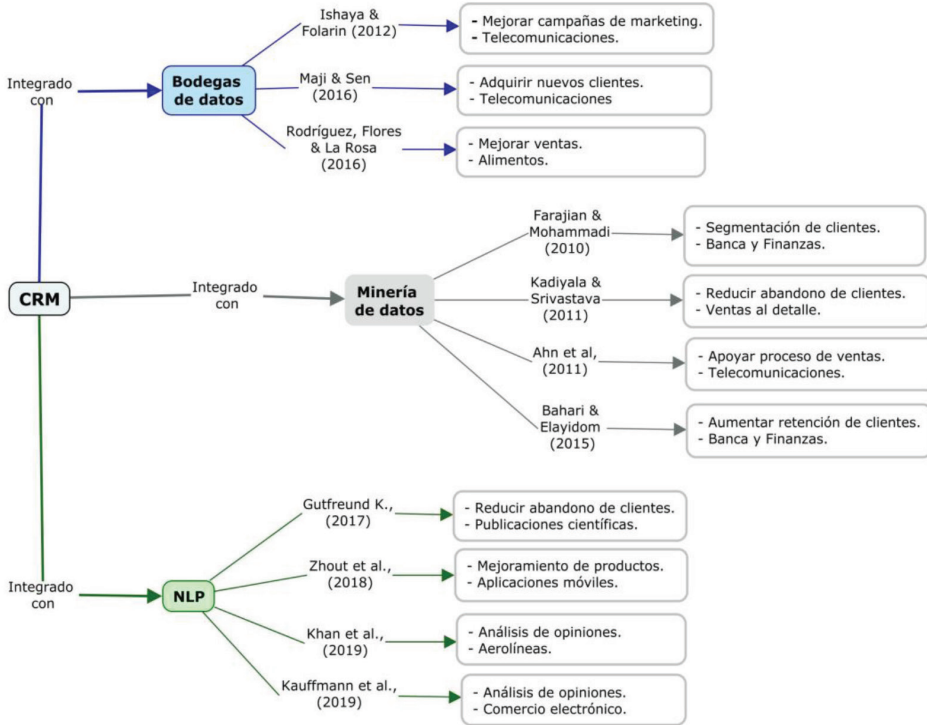


Figura 1 – Propuestas que utilizan análisis de datos para apoyar la relación con los clientes

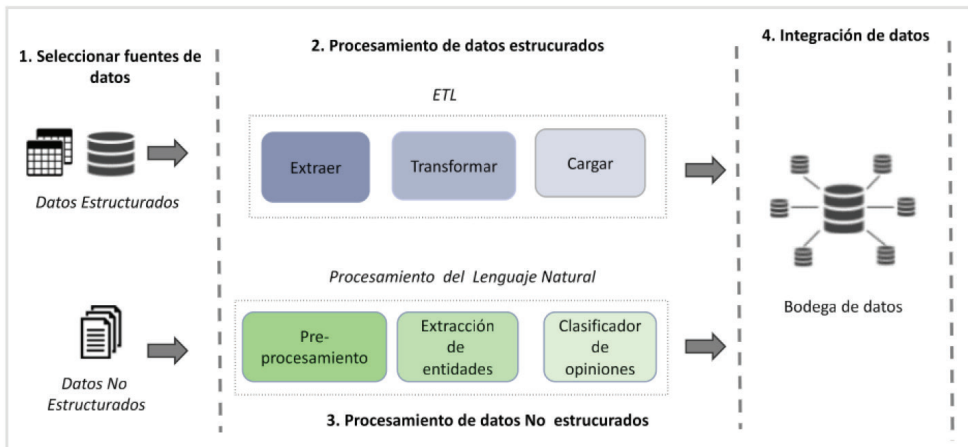


Figura 2 – Enfoque propuesto

### 3.2. Procesamiento de datos estructurados

En esta fase se aplican técnicas de inteligencia de negocios y bodegas de datos con el fin de obtener indicadores asociados a la relación con los clientes. Como entrada se toman diferentes registros estructurados y estos se transforman en indicadores orientados al análisis de información. La Figura 3 muestra los pasos que se ejecutan en el procesamiento de datos estructurados.

Los indicadores obtenidos se integran y almacenan en un modelo dimensional. Este modelo se describe detalladamente en la sección 3.4. Algunos indicadores que se obtienen a partir del análisis de datos estructurados son:

- Número de transacciones que se resuelven en el primer contacto o primera llamada.
- Número de transacciones que son escaladas a otro equipo que tomará la responsabilidad de resolver la transacción.
- Número de llamadas telefónicas en las que se obtuvo una venta.
- Tiempo promedio que lleva resolver una transacción.

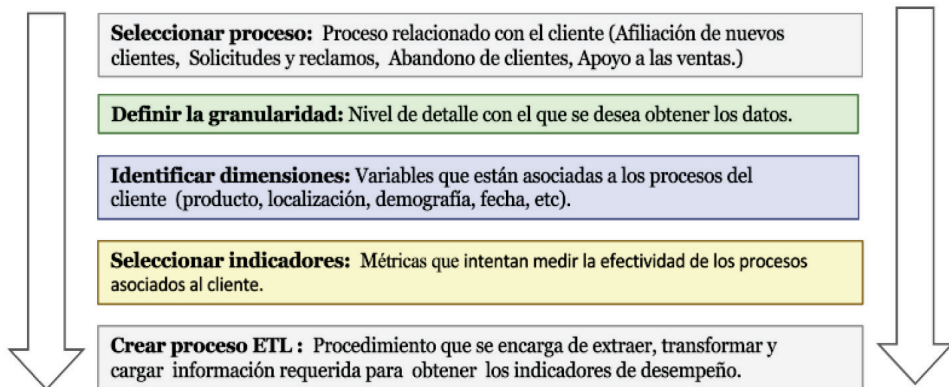


Figura 3 – Procesamiento de datos estructurados

### 3.3. Procesamiento de datos no estructurados

En este componente se aplican técnicas de NLP para extraer información de fuentes de datos basadas en texto, tal como se muestra en la Figura 4. Este componente puede procesar correos electrónicos, llamadas telefónicas previamente transcritas a texto, o cualquier otro archivo basado en texto.

De acuerdo con la Figura 4, este componente se divide en tres fases: Pre-procesamiento de datos, Extracción de entidades nombradas y un clasificador de opiniones. Los resultados obtenidos se almacenan en un modelo dimensional que sirve como base para el análisis de datos.

#### *Pre-procesamiento de datos*

En esta fase se aplican técnicas para limpiar y normalizar los documentos de texto:

- *Tokenización*: divide el texto en sentencias de texto y tokens. Un token es un elemento atómico de una sentencia como palabras, números o siglas.
- *Eliminar Stop Words*: se eliminan palabras que son muy frecuentes en el texto y no aportan valor semántico al texto (ejemplo: la, los, de, del, etc)
- *Normalización*: convierte el texto a minúsculas y luego elimina acentos.

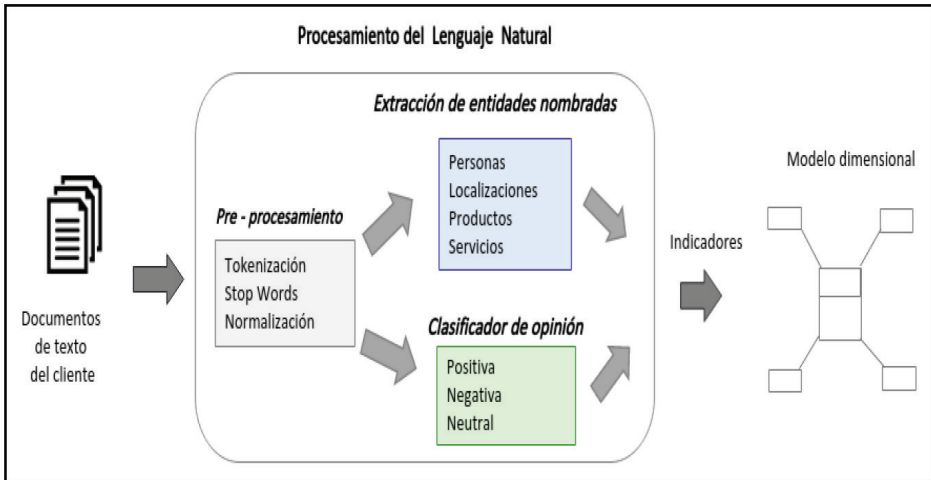


Figura 4 – Procesamiento de datos no estructurados

### Reconocimiento de entidades nombradas

En esta fase se aplica el reconocimiento de entidades nombradas (Named Entity Recognition o NER) con el objetivo de extraer personas, productos, servicios o localizaciones relacionadas con los clientes. Para hacer esta tarea se utilizó el estándar de clasificación de entidades propuesto en CoNLL-2003 para el idioma español (Sang & De Meulder, 2003). En el enfoque propuesto las entidades extraídas se clasifican usando las siguientes etiquetas:

- **PER**: etiqueta usada para hacer referencia a personas.
- **ORG**: identifica las organizaciones mencionadas por el cliente.
- **MISC**: productos y servicios ofrecidos por la organización
- **LOC**: localizaciones y ciudades mencionadas por el cliente.

Para extraer entidades se utiliza *Freeling*<sup>2</sup>, una librería de código abierto para el procesamiento de lenguaje natural y que tiene soporte para el idioma español. La Figura

<sup>2</sup> <http://nlp.lsi.upc.edu/freeling/index.php/node/1>

5 muestra un ejemplo donde se toma como entrada un correo electrónico enviado por el cliente y la salida es un conjunto de entidades extraídas del texto.

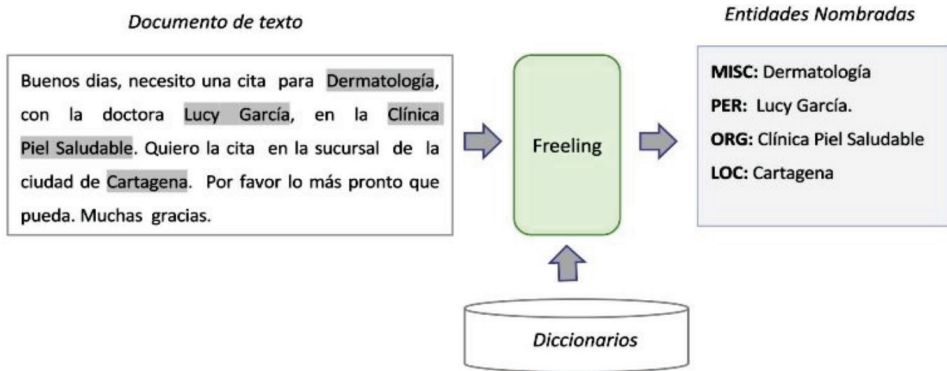


Figura 5 – Extracción de entidades nombradas

### Clasificador de opinión

En esta fase se extrae la opinión que los clientes tienen respecto a los productos o servicios que se identificaron en la fase anterior. Para extraer esta opinión se usa un clasificador automático que etiqueta un texto en tres categorías: positivo, negativo o neutral. El clasificador se basa en un modelo que se implementó usando los algoritmos Naive Bayes y Árboles de decisión.

El modelo de clasificación se entrenó con un conjunto de 3200 textos escritos en español. Estos textos fueron extraídos de correos electrónicos y llamadas telefónicas del centro de atención al cliente de la organización Previser, que ofrece servicios de salud a sus clientes. En esta organización se aplicó un caso de estudio del cual se hablará con detalle en la sección 3.4. El conjunto de datos fue anotado y validado manualmente por cinco personas expertas del área de servicio al cliente. Al finalizar el proceso de anotación se hizo un consenso entre todos los anotadores para determinar la etiqueta final. La Tabla 1 muestra tres ejemplos de textos con su respectiva etiqueta de opinión.

Texto	Etiqueta
No ha sido posible contactar con usted, esta <b>muy mala</b> la comunicación. Para pedir una cita siempre suena ocupado, está <b>pésimo</b> el servicio.	Negativo
Para pedir una cita médica con el cardiólogo. Tenemos disponible con el doctor Pedro García. <b>Perfecto</b> , el siempre me ha atendido <b>amablemente</b> .	Positivo
Señora, buenas tardes. Su cita se agendó para el viernes a las 2:00 PM. Esperamos su confirmación y llegar 10 minutos antes.	Neutral

Tabla 1 – Textos anotados manualmente con su respectiva etiqueta



### 3.4. Integración de datos

En este componente se integran los indicadores obtenidos en las fases anteriores. La integración se basa en el uso de elementos del modelo dimensional como *datamarts*, tablas de hechos y dimensiones. Con estos elementos se modela una bodega de datos que contiene información de fuentes estructuradas y no estructuradas.

El componente de integración de datos busca extender la capacidad de análisis de datos usando el potencial de los datos no estructurados. La integración se hace a través de dimensiones compartidas, es decir aquellas dimensiones que aparecen tanto en los datos estructurados como en los no estructurados, tal como se muestra en los siguientes ejemplos. La Figura 6 muestra un *datamart* que representa el proceso “Apoyo a las ventas”. Este *datamart* se genera a partir de datos estructurados y se pueden obtener indicadores como el porcentaje de ventas por productos, por localización, productos más y menos vendidos, etc.

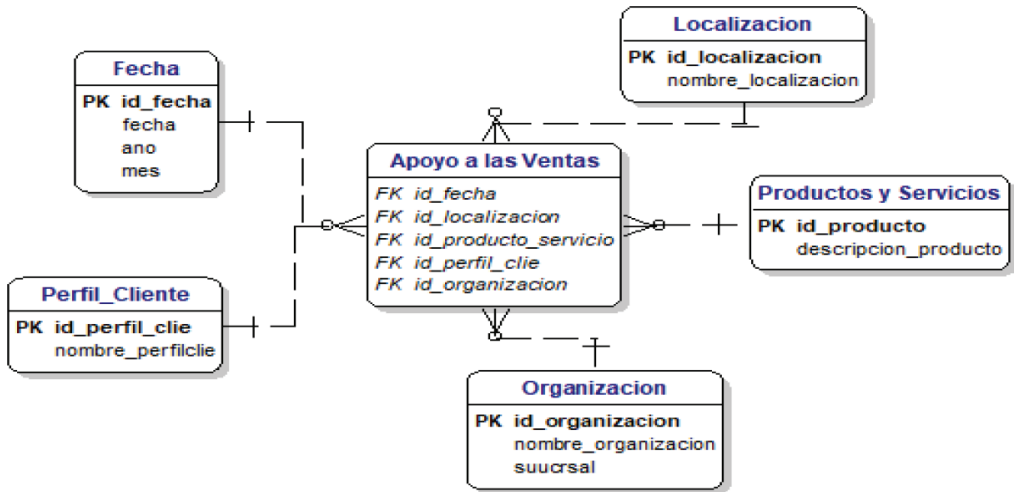


Figura 6 – *Datamart* para el proceso de apoyo a las ventas

Los indicadores obtenidos a partir de la Figura 6 describen cantidades y porcentajes de los productos y servicios vendidos, sin embargo, dichos indicadores son limitados ya que no permiten conocer los factores que podrían estar influyendo en el proceso de ventas.

Por su parte, la Figura 7 muestra un *datamart* que representa el proceso de “Quejas y reclamos del cliente”, el cual se genera a partir del análisis de datos no estructurados. Este *datamart* contiene indicadores de las opiniones que el cliente ha expresado acerca de productos, servicios o del proceso de ventas en general. Estos indicadores se obtienen mediante la extracción de entidades nombradas y el clasificador de opiniones. De este

*datamart* se pueden obtener indicadores como el porcentaje de opiniones positivas y negativas, productos con más quejas o reclamos, quejas más comunes mencionadas por los clientes, productos mejor calificados, etc.

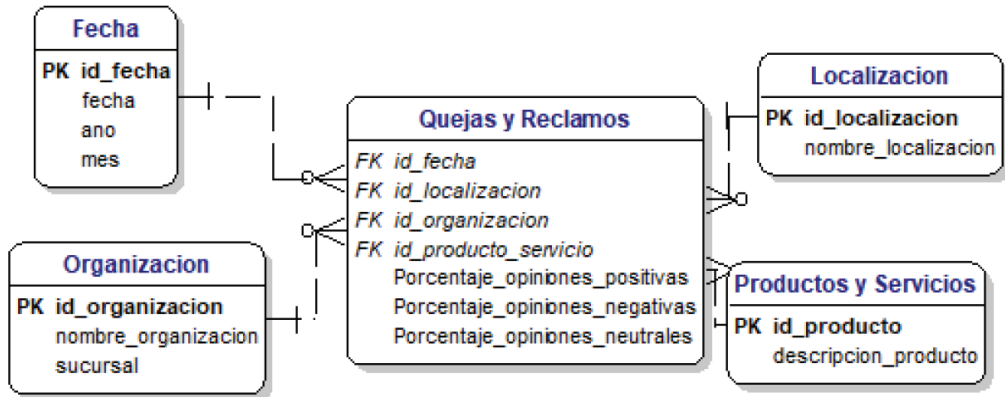


Figura 7 – *Datamart* para el proceso quejas y reclamos

Los *datamart* de las Figuras 6 y 7 se pueden integrar mediante el uso de dimensiones compartidas como productos, servicios o localización. Estas dimensiones aparecen tanto en el análisis de datos estructurados como no estructurados. A través de esta integración se extiende el proceso de apoyo a las ventas ya que al analizar las opiniones del cliente se conocen ciertos factores que podrían estar afectando el proceso de ventas. Usando la extracción de entidades nombradas y el análisis de opiniones se pueden establecer relaciones entre las opiniones del cliente y los resultados en el proceso de ventas. En este sentido, la integración de datos no estructurados permite obtener no solamente resultados de ventas, sino también los factores que están influyendo en esos resultados.

Así como el ejemplo descrito anteriormente, el enfoque propuesto permite modelar diferentes procesos asociados al cliente aprovechando el potencial de los datos no estructurados. Además, los indicadores obtenidos por medio del enfoque se pueden utilizar en diferentes tareas clave en la industria 4.0 como son el mejoramiento continuo de productos y servicios, la personalización de productos, la personalización de campañas de marketing y el constante mejoramiento de las relaciones con los clientes.

#### 4. Caso de Estudio

El caso de estudio se realizó en la organización Previser<sup>3</sup>, una compañía colombiana que ofrece una red de servicios que conecta a los consumidores finales con diferentes proveedores de productos de salud. Previser tiene un área de servicio al cliente que cuenta con un sistema CRM. Además, diariamente reciben cientos de llamadas telefónicas y correos electrónicos con las solicitudes de los clientes. Para validar el enfoque se hicieron

<sup>3</sup> <https://www.previser.com.co>

dos experimentos. El primero para evaluar el reconocimiento de entidades nombradas y el segundo, el clasificador de opiniones.

#### 4.1. Experimento 1: extracción de entidades nombradas

- **Objetivo:** validar el reconocimiento de entidades nombradas en textos asociados al cliente. Como *dataset* de prueba se tiene un conjunto de 1000 textos previamente etiquetados, que corresponden al 20 % de corpus de entrenamiento que contiene alrededor de 5000 documentos de texto.
- **Resultados:** la Tabla 2 muestra los resultados obtenidos separados por cada etiqueta. El reconocimiento de entidades se hizo con la herramienta *Freeling*. De acuerdo con la Tabla 2, el rendimiento para extraer las diferentes categorías de entidades es similar en todas las etiquetas. Una de las dificultades al desarrollar el experimento fue no tener un corpus disponible en español para el dominio del servicio al cliente. Por lo tanto, la primera decisión fue construir un corpus de prueba. El tamaño de este corpus influyó en los resultados obtenidos, a medida que el corpus crezca, la capacidad para encontrar diferentes entidades asociadas al servicio al cliente puede mejorar.

	PER	ORG	MISC	LOC
Precision	0,89	0,86	0,90	0,87
Recall	0,88	0,88	0,87	0,86
F-score	0,88	0,87	0,89	0,85

Tabla 2 – Resultados Experimento 1

#### 4.2. Experimento 2: clasificador de opiniones

- **Objetivo:** validar el clasificador de opiniones que categoriza la polaridad del texto en: positivo, negativo o neutral. Se usó el mismo conjunto de datos utilizado en el primer experimento.
- **Resultados:** la Tabla 3 muestra los resultados obtenidos separados por cada etiqueta. Las pruebas se realizaron con dos clasificadores, Naive Bayes y *Decision Tree*. De acuerdo con la Tabla 3, el algoritmo basado en árboles de decisión tiene un rendimiento superior en la tarea de extracción de opiniones. Los parámetros utilizados para el árbol de decisión fueron (*criterion='entropy', min\_samples\_split=20, min\_samples\_leaf=5, max\_depth=4*).

La mayor dificultad en la extracción automática de opiniones del cliente, se presentó en textos demasiado largos. En este caso, el clasificador por lo general tiende a etiquetarlos como neutros. Por otro lado, en el idioma español se pueden encontrar gran variedad de expresiones para opiniones similares expresadas por el cliente. Por lo tanto, el mejoramiento del corpus de entrenamiento es un paso fundamental para mejorar los resultados obtenidos.

	Árboles de decisión	Naive Bayes
<i>Precision</i>	0,89	0,86
<i>Recall</i>	0,87	0,85
<i>F-score</i>	0,88	0,85

Tabla 3 – Resultados Experimento 2

Los resultados que se muestran en las Tablas 2 y 3 muestran una métrica F-score mayor al 85% en ambos casos, esto indica que las técnicas de NLP utilizadas fueron útiles para extraer información relacionada con los clientes. Aunque estos resultados se pueden mejorar, el principal objetivo del caso de estudio era probar la utilidad del enfoque propuesto para extraer entidades y opiniones asociadas al cliente.

En el experimento 1 se extraen las entidades mencionadas por el cliente (*e.g* productos o servicios). En el experimento 2 se evalúa la opinión que el cliente tiene sobre dichas entidades. De esta manera se forma la pareja “entidad: opinión”. Esta pareja se usa para extender la información obtenida con los datos estructurados, los cuales solo permiten visualizar la cantidad de servicios solicitados, pero no la opinión de los clientes sobre dichos servicios. Analizar datos no estructurados permite obtener indicadores adicionales como la identificación de servicios con un alto número de opiniones negativas, los servicios mejor calificados o las quejas más comunes expresadas por el cliente. La Figura 8 muestra los servicios con más opiniones negativas que se obtuvieron en el caso de estudio. Esta información se puede aprovechar para entender mejor los resultados del proceso de ventas y así generar estrategias de mejoramiento del servicio.

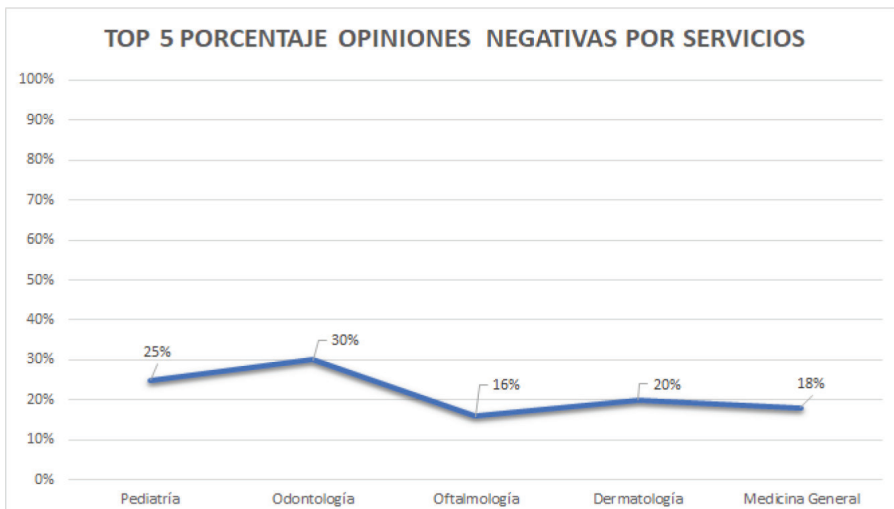


Figura 8 – Servicios con más opiniones negativas

La extracción de entidades y el clasificador de opiniones son técnicas del NLP utilizadas para extraer información relacionada con los gustos y opiniones expresados por los clientes en documentos de texto. Estas dos técnicas se complementan ya que los indicadores extraídos se integran en la bodega de datos, tal como se describió en la sección 3.3. Estos indicadores se pueden usar para el mejoramiento de procesos y productos. Así, el análisis de datos se transforma en ventaja competitiva, lo cual va en el camino de la cuarta revolución industrial.

## 5. Conclusiones y trabajo Futuro

En este artículo se propuso un enfoque para extraer indicadores de desempeño del servicio al cliente integrando técnicas de inteligencia de negocios y NLP. En particular, las técnicas de NLP aprovechan el potencial que tienen las fuentes de datos basadas en texto. Estas fuentes de datos contienen información que revela los gustos y opiniones del cliente. Extraer esta información e integrarla con otras fuentes de datos puede ser una ventaja competitiva que pueden aprovechar las empresas para mejorar sus procesos.

El análisis de diferentes fuentes de datos es primordial para apoyar el desarrollo de la cuarta revolución industrial. El enfoque propuesto mostró cómo se puede aprovechar fuentes de datos estructuradas y no estructuradas para obtener indicadores que pueden ser útiles para apoyar las relaciones con los clientes. Apoyar esas relaciones es importante ya que, en la nueva revolución industrial, los clientes pueden convertirse en fuentes que generan valor agregado para las empresas.

La gran variedad de datos que se generan actualmente implica que se deben combinar diferentes técnicas y estrategias para cumplir con los requerimientos de las empresas. En este artículo se propuso un enfoque que combina técnicas de inteligencia de negocios, bodegas de datos y NLP para tratar el problema de los datos no estructurados. Las técnicas de la inteligencia artificial como el NLP son importantes para el desarrollo de aplicaciones en la cuarta revolución industrial. Como trabajo futuro se planea probar otros algoritmos para el reconocimiento de entidades y opiniones con el fin de aumentar el rendimiento de nuestro enfoque. También seguir probando otras técnicas de NLP que se puedan aplicar al análisis de datos de los clientes.

## Referencias

- Ahn, H., Ahn, J. J., Oh, K. J., & Kim, D. H. (2011). Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques. *Expert Systems with Applications*, 38(5), 5005–5012. <https://doi.org/10.1016/j.eswa.2010.09.150>
- Bahari F., & Elayidom S., (2015). An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour. *Procedia Computer Science*, 46, 725–731. <https://doi.org/10.1016/j.procs.2015.02.136>
- Cunningham, C., Song, I. Y., & Chen, P. P. (2004). Data warehouse design to support customer relationship management analyses. *DOLAP: Proceedings of the ACM Int. Workshop on Data Warehousing and OLAP*, 17(June), 14–22. <https://doi.org/10.4018/978-1-60566-058-5.ch042>

- Cunningham, C., & Song, I.-Y. (2007). A Taxonomy of Customer Relationship Management Analyses for Data Warehousing. *The 26th International Conference on Conceptual Modeling*, 83, 97–102. Retrieved from <http://dl.acm.org/citation.cfm?id=1386972>
- Devi, P. I., & Rajagopalan, S.P. (2012). Analysis of Customer Behavior using Clustering and Association Rules. *International Journal of Computer Applications*, 43(23), 19–26.
- Farajian, M. A., & Mohammadi, S. (2010). Mining the Banking Customer Behavior Clustering and Association Rule Methods. *International Journal of Industrial Engineering & Product Research*, 21(4), 239–245. <http://ijiepr.iust.ac.ir/article-1-241-en.html>
- Gutfreund, K. (2017). Big Data Techniques for Predictive Business Intelligence. *Journal of Advanced Management Science*, 5(2), 158–163. <https://doi.org/10.18178/joams.5.2.158-163>
- Handzic, M., Ozlen, K., & Durmic, N. (2014). Improving Customer Relationship Management Through Business Intelligence. *Journal of Information & Knowledge Management*, 13(2).
- Hassan, R. S., Nawaz, A., Lashari, M. N., & Zafar, F. (2015). Effect of Customer Relationship Management on Customer Satisfaction. *Procedia Economics and Finance*, 23, 563–567. [https://doi.org/10.1016/s2212-5671\(15\)00513-4](https://doi.org/10.1016/s2212-5671(15)00513-4)
- Ishaya, T., & Folarin, M. (2012). A service oriented approach to Business Intelligence in Telecoms industry. *Telematics and Informatics*, 29(3), 273–285. <https://doi.org/10.1016/j.tele.2012.01.004>
- Kadiyala, S. S., & Srivastava, A. (2011). Data Mining for Customer Relationship Management. *International Business & Economics Research Journal*, 1(6), 61–70.
- Khan, D. M., Rao, T. A., & Shahzad, F. (2019). The Classification of Customers' Sentiment using Data Mining Approaches. *Global Social Sciences Review*, IV(IV), 146–156. [https://doi.org/10.31703/gssr.2019\(iv-iv\).19](https://doi.org/10.31703/gssr.2019(iv-iv).19)
- Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2019). Managing marketing decision-making with sentiment analysis: An evaluation of the main product features using text data mining. *Sustainability*, 11(15), 1–19. <https://doi.org/10.3390/su11154235>
- Maji, G., & Sen, S. (2016). Data warehouse based analysis on CDR to retain and acquire customers by targeted marketing. *5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*, Noida, India. <https://doi.org/10.1109/icrito.2016.7784955>
- Mittal, S., Khan, M. A., Romero, D., & Wuest, T. (2019). Smart manufacturing: Characteristics, technologies and enabling factors. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 233(5), 1342–1361. <https://doi.org/10.1177/09544405417736547>

- Müller, O., Debortoli, S., Junglas, I., & Vom-Brocke, J. (2016). Using text analytics to derive customer service management benefits from unstructured data. *MIS Quarterly Executive*, 15(4), 243–258.
- Névóel, A., Dalianis, H., Velupillai, S., Savova, G., & Zweigenbaum, P. (2018). Clinical Natural Language Processing in languages other than English: Opportunities and challenges. *Journal of Biomedical Semantics*, 9(1), 1–13. <https://doi.org/10.1186/s13326-018-0179-8>
- Peres, R. S., Rocha, D. A., Leitao, P., & Barata, J. (2018). IDARTS Towards intelligent data analysis and real-time supervision for industry 4.0. *Computers in Industry*, 101, 138–146. <https://doi.org/10.1016/j.compind.2018.07.004>
- Picek, R., Peras, D., & Mekovec, R. (2018). Opportunities and Challenges of Applying Omnichannel Approach to Contact. In *Proceedings of 2018 4th International Conference on Information Management (ICIM)*, 231–235.
- Rahman, A., & Khan, M. N. A. (2017). An Assessment of Data Mining Based CRM Techniques for Enhancing Profitability. *International Journal of Education and Management Engineering*, 7(2), 30–40. <https://doi.org/10.5815/ijeme.2017.02.04>
- Rodrigues Chagas. B. N., Nogueira V. J. A., Reinhold, O., Lobato, F., Jacob, A. F. L., & Alt, R. (2019). Current Applications of Machine Learning Techniques in CRM: A Literature Review and Practical Implications. In *Proceedings IEEE/WIC/ACM International Conference on Web Intelligence*, 452–458. <https://doi.org/10.1109/WI.2018.00-53>
- Rodríguez Flores, F., Flores Pulido, L., & La Rosa, E. D. (2016). Inteligencia de negocios y minería de datos aplicado a la industria refresquera. *Research in Computing Science*, 126(1), 63–71. <https://doi.org/10.13053/rcs-126-1-6>
- Sang, E.T.K., & De Meulder, F. (2003). Introduction to the conll-2002 shared task: Language independent named entity recognition. In *Proceedings of CoNLL-2002*, 155–158.
- Soltani, Z., & Navimipour, N. J. (2016). Customer relationship management mechanisms: A systematic review of the state of the art literature and recommendations for future research. *Computers in Human Behavior*, 61, 667–688. <https://doi.org/10.1016/j.chb.2016.03.008>
- Thoben, K. D., Wiesner, S. A., & Wuest, T. (2017). “Industrie 4.0” and smart manufacturing—a review of research issues and application examples. *International Journal of Automation Technology*, 11(1), 4–16. <https://doi.org/10.20965/ijat.2017.p0004>
- Zhou, S., Qiao, Z., Du, Q., Wang, G. A., Fan, W., & Yan, X. (2018). Measuring Customer Agility from Online Reviews Using Big Data Text Analytics. *Journal of Management Information Systems*, 35(2), 510–539.