

ARTÍCULO ORIGINAL

# Análisis de sentimientos orientado al comercio electrónico de CITMATEL: proyecto de voz del cliente

*E-Commerce-Oriented Sentiment Analysis of CITMATEL:  
Voice-Of-Thecust*

*Carlos Mar Rodríguez*

*cmar@citmatel.inf.cu* • <https://orcid.org/0000-0002-2367-45565>

EMPRESA DE TECNOLOGÍAS DE LA INFORMACIÓN  
Y SERVICIOS TELEMÁTICOS AVANZADOS, CITMATEL, CUBA

*Patricia Montañez Castelo*

*patriciacastelo99@gmail.com* • <https://orcid.org/0000-0002-6776-9434>

*Alfredo Simón-Cuevas*

*asimon@ceis.cujae.edu.cu* • <https://orcid.org/0000-0003-4506-5054>

UNIVERSIDAD TECNOLÓGICA DE LA HABANA "JOSÉ ANTONIO ECHEVERRÍA", CUJAE, CUBA

Recibido: 2023-01-26 • Aceptado: 2023-03-02

## RESUMEN

El comercio electrónico desempeña un papel fundamental en la transformación digital. El uso generalizado de las redes sociales ha propiciado que los clientes y consumidores opinen libremente sobre los productos y servicios, y expresen sus emociones y experiencias, lo que tiene gran repercusión en los negocios de venta en línea, de ahí el surgimiento de lo que se conoce como Voz del Cliente (VoC). El procesamiento eficaz y análisis inteligente del volumen de información textual no estructurada, que se genera para capturar el grado de satisfacción de los usuarios, es una tarea desafiante que demanda la aplicación de tecnologías de Procesamiento de Lenguaje Natural (PLN), específicamente, de Análisis de Sentimientos. Este trabajo constituye un primer acercamiento de las tecnologías de Análisis de Sentimientos en el comercio electrónico que desarrolla CITMATEL. Se muestra aquí un modelo de procesamiento y análisis computacional de información textual, que guía la implementación de un proyecto de VoC con impacto en el comercio electrónico. Como parte de ese modelo se desarrolla una solución concreta de Análisis de Sentimientos, cuyas

evaluaciones parciales sobre opiniones de productos en datos de prueba permitieron obtener resultados con valores de precisión y exhaustividad muy beneficiosos para su despliegue en dicha empresa.

**PALABRAS CLAVE:** comercio electrónico, procesamiento de lenguaje natural, análisis de sentimientos, Voz del Cliente.

## ABSTRACT

*E-commerce plays a key role in digital transformation programs. The extensive use of social networks has allowed customers and consumers to freely express their opinions about products and services, and to express their emotions and experiences, which has had a great impact on online sales businesses, hence the emergence of what is known as Voice of the Customer (VoC). The efficient processing and intelligent analysis of the entire volume of unstructured textual information generated for capturing the user-satisfaction degree is a challenging task, and demands the application of Natural Language Processing technologies, specifically, Sentiment Analysis. This work constitutes a first approach of the Sentiment Analysis technologies in e-commerce developed by CITMATEL. A model of computational processing and analysis of textual information that guides the implementation of a VoC project, with impact on e-commerce, is presented. As part of this model, a concrete Sentiment Analysis solution is developed, whose partial evaluations on a test dataset of product reviews where were obtained high precision and recall which makes it suitable for deployment in the enterprise.*

**KEYWORDS:** e-commerce, natural language processing, sentiment analysis, Voice of the Customer.

## INTRODUCCIÓN

Desde hace varios años el proceso de transformación digital a nivel mundial ha venido desarrollándose de forma acelerada y el comercio electrónico desempeña un papel fundamental, tanto para las empresas como para los clientes finales de los productos y servicios que estas ofrecen (Kian, 2021). El acceso a los portales de comercio electrónico y las compras en línea se han convertido en los nuevos mercados de la sociedad. La generalización del uso de Internet, acompañado del surgimiento y desarrollo de las redes sociales (Facebook, Twitter, etc.), y el incremento de otros espacios de comunicación impulsados por la Web 2.0, ha facilitado que

los clientes y consumidores opinen libremente sobre los productos y servicios. Estos expresan sus emociones y experiencias, lo que abre paso a una nueva forma de percepción de su satisfacción.

El comercio electrónico puede definirse como las actividades relacionadas con la compra y venta de productos o servicios a través de Internet. Las empresas se dedican cada vez más al comercio electrónico, debido a la creciente demanda de servicios en línea por parte de los clientes y su capacidad de crear una ventaja competitiva. Sin embargo, se ven obligadas a adaptar constantemente sus modelos de negocio, a las cambiantes necesidades de los clientes y a cómo estos asumen las tecnologías de información que están a su alcance. Las tecnologías de IA han ido transformando esta forma de comercio, gracias a las capacidades que brindan para interpretar correctamente datos, aprender de ellos y utilizar esos aprendizajes para lograr objetivos y tareas específicas mediante una adaptación flexible, creando oportunidades para que las empresas obtengan una ventaja competitiva (Kian, 2021). Algunas de las principales ventajas del uso de tecnologías de IA en el comercio electrónico son (Saraswat, Kumar, & Abhishek, 2022): búsqueda intuitiva centrada en el cliente, filtrado de reseñas e información falsa, automatización, personalización, asistentes virtuales para las compras, segmentación de clientes y logística inteligente.

Los consumidores suelen evaluar los productos y servicios en función de las opiniones que generan otros consumidores de dicho producto. Por tanto, hoy en día uno de los factores que más influyen en las ventas de productos y servicios son dichas reseñas, ya que ayudan a mitigar el miedo a ser engañado y aumentan la confianza entre consumidores y empresas (Saraswat, Kumar, & Abhishek, 2022). Los portales de comercio electrónico generan cada día muchos datos en forma de opiniones de clientes. El análisis de los datos de comercio electrónico ayuda a los vendedores minoristas en línea a comprender las expectativas de los clientes, ofrecer una mejor experiencia de compra y aumentar las ventas.

En este contexto, surge lo que se conoce como Voz del Cliente (VoC), término que describe el proceso exhaustivo de captar e interpretar con regularidad las expectativas, preferencias y experiencias del cliente con los productos y servicios (Mehtab, 2021). Un programa de VoC es una fuente inagotable de retroalimentación para detectar fortalezas, oportunidades de mejora y alimentar con iniciativas a cada área de la empresa, en síntesis, permite conectar la opinión de los clientes de la empresa con los decisores.

La inteligencia Artificial (IA) es la tecnología con las mayores tendencias dentro del comercio electrónico en los próximos años. Según Gartner, la mayoría de las organizaciones que utilizan la IA en sus sistemas de comercio electrónico lograrán un aumento de al menos 25 % en satisfacción de los clientes para 2023, junto a un aumento de los ingresos y una reducción de los costes (Kian, 2021). Enfrentar los desafíos que representa el procesamiento eficaz del gran volumen de información textual no estructurada que generan los usuarios, hace imprescindible la aplicación de tecnologías de IA y, entre ellas, las relacionadas con el Procesamiento del Lenguaje Natural (PLN) y la Minería de Textos (MT), con énfasis en el Análisis de Sentimientos o Minería de Opinión, dado que los objetivos y retos de los programas VoC y las

soluciones de procesamiento inteligente de textos son esencialmente los mismos. El Análisis de Sentimientos es el método más utilizado para analizar datos en forma de texto, identificar el contenido de sentimiento del texto e identificar información positiva, negativa y neutra de las opiniones de los clientes (Mehtab, 2021).

La Empresa de Tecnologías de la Información y Servicios Telemáticos Avanzados (CITMATEL) mantiene en funcionamiento varias plataformas de comercio electrónico que ofrecen diversos productos y servicios. Un grupo de especialistas de la empresa extrae y analiza, de forma manual, información textual vertida por los usuarios proveniente de varios canales.

Este trabajo tiene como objetivo plantear una primera aproximación de la introducción de las tecnologías de Análisis de Sentimientos o Minería de Opinión en el contexto del comercio electrónico de CITMATEL, como parte de la implementación de un proyecto de VoC y de transformación digital basado en IA. En este sentido, se propone una primera aproximación de la concepción general de un modelo de procesamiento y análisis computacional inteligente de información textual no estructurada, desde la perspectiva del Análisis de Sentimientos basado en aspectos y orientado al comercio electrónico, que sirve de guía para implementar el proyecto de VoC en la empresa. Además, se muestra la concepción de una solución de análisis de sentimientos, específicamente para la detección de polaridad basada en aspectos, cuyas evaluaciones parciales sobre textos de opinión generados en una plataforma de venta en línea, proporcionan resultados alentadores y de valor en un entorno de comercio electrónico.

## METODOLOGÍA

La metodología planteada para abordar el objetivo fundamental de este trabajo se puede resumir de la siguiente manera:

1. Estudio y asimilación de los fundamentos teóricos y, por un lado, del Procesamiento del Lenguaje Natural (PLN), la Minería de Texto y el Análisis de Sentimientos, y por el otro, de lo que se conoce como VoC, así como la caracterización de soluciones y tecnologías afines ya reportadas en la literatura.
2. Definición de una primera aproximación de la concepción general de un modelo de procesamiento y análisis computacional inteligente de información textual no estructurada, que ofrezca una guía para la implementación de un proyecto de VoC, desde la perspectiva del Análisis de Sentimientos orientado al comercio electrónico, como ámbito de la empresa sobre la que impactarán los resultados.
3. Concepción y evaluación preliminar de una solución de Análisis de Sentimientos, específicamente para la detección de polaridad basada en aspectos, para crear las bases y mostrar una vía para acercarse a la práctica uno de los servicios contemplados dentro del modelo de VoC propuesto, así como ejemplificar resultados (datos textuales) que pueden aportar a la toma de decisiones sobre productos que se ofertan en una venta *online*.

## RESULTADOS Y DISCUSIÓN

### COMERCIO ELECTRÓNICO EN LA ERA DE LA INTELIGENCIA ARTIFICIAL

El procesamiento de los contenidos de opinión generados por los usuarios sobre productos y servicios, tanto en las propias plataformas de comercio electrónico como en las redes sociales, hoy desempeña un papel fundamental para lograr el éxito en este tipo de empresa. Las organizaciones están desarrollando un nuevo modelo de marketing centrado en las relaciones entre la empresa y sus clientes, donde la IA encuentra un extraordinario campo de aplicación en la gestión de la información, en la atención al cliente y en el análisis de la experiencia de compra. En este sentido, el PLN y la Minería de Texto constituyen áreas de trabajo e investigación de las IA, desde donde se ofrecen las tecnologías necesarias para llevar a cabo el procesamiento y análisis eficaz de ese tipo de contenido, y más específicamente de lo que se conoce como Minería de Opinión o Análisis de Sentimientos. Precisamente, estos temas figuran entre los más investigados dentro de la aplicación de la IA en el campo del comercio electrónico (Kian, 2021).

### ANÁLISIS DE SENTIMIENTOS

El Análisis de Sentimientos o Minería de Opinión es el estudio computacional sobre las actitudes, los pensamientos y los sentimientos de las personas hacia determinadas entidades y sus características (Bawack, Wamba, Carillo, & Akter, 2022), que mediante el uso de la lingüística computacional y el análisis de textos, es capaz de procesar y extraer información de datos representados de forma textual (Liu & Zhang, 2012). Por lo general, en la literatura se hace una distinción entre opiniones y sentimientos (Varathan, Giachanou, & Crestani, 2016). Cuando se habla de opiniones, se refiere al punto de vista de una persona hacia un determinado asunto, mientras que el sentimiento o la polaridad se refiere a la emoción que experimenta la persona hacia ese asunto, la cual puede ser positiva o negativa (Varathan, Giachanou, & Crestani, 2016). La Minería de Opinión es dividida, por lo general, en tres categorías, según la extensión de texto que se va a analizar y los elementos que se quieren extraer de este. Las categorías son (Brauwerts & Frasincar, 2021):

- A nivel de documento (*Document-Level*): se toma todo el documento como opinión y se clasifica su polaridad general sin tener en cuenta otros detalles (Nazir, Rao, Wu, & Sun, 2022). En este nivel, la clasificación de polaridad se hace basándose en la cantidad de palabras con sentimiento positivo o negativo que contenga el texto, lo cual no resulta factible para la toma de decisiones, ya que un mismo usuario puede hacer referencia a varios elementos y expresar diferentes sentimientos en una misma opinión (Subhashini *et al.*, 2021).
- A nivel de oración (*Sentence-Level*): se asume que toda la oración contiene una única opinión y se clasifica en su conjunto de acuerdo a una polaridad (Nazir, Rao, Wu, & Sun, 2022). En este caso, un único aspecto y polaridad es considerado para cada oración del documento (Aboelela, Gad, & Ismail, 2021).

- A nivel de aspecto (*Aspect-Level*): se considera el nivel de más baja granularidad, ya que se clasifican por separado las entidades, los aspectos y las relaciones presentes en las opiniones (Nazir, Rao, Wu, & Sun, 2022). Para este nivel se tienen en cuenta todos los aspectos que constituyen el documento y se determina la polaridad de cada uno de ellos (Aboelela, Gad, & Ismail, 2021). El análisis a nivel de características o aspectos se podría decir que es el más cercano a dar un análisis preciso de las opiniones de los usuarios, siendo recomendado en muchos estudios como la mejor técnica de Minería de Opinión (García, Cuadros, & Rigau, 2018).

### **ANÁLISIS DE SENTIMIENTOS BASADO EN ASPECTOS**

El Análisis de Sentimientos basado en aspectos es un tipo de Minería de Opinión, que consiste en clasificar el sentimiento de la opinión que se está expresando sobre diferentes características o aspectos de productos o servicios que están siendo evaluados (Liu *et al.*, 2019). En este proceso, a partir del procesamiento de los textos de opinión, se es capaz de extraer información relevante, logrando ir más allá de las calificaciones en base a una puntuación que dan los usuarios en los sitios de reseñas (García, Cuadros, & Rigau, 2018). Este tipo de soluciones son de las que más podrían aportar comercio electrónico, sobre todo si se inserta en una plataforma de tienda virtual y se vincula a los productos que se ofrecen. Un ejemplo son las tiendas virtuales de CITMATEL.

La tarea de Análisis de Sentimientos basado en aspectos, según algunos autores, se compone de tres subtarefas fundamentales (Brauwers, & Frasincar, 2021): Extracción de Aspectos (*Aspect Extraction*, AE), Análisis de Sentimientos del Aspecto (*Aspect Sentiment Analysis*, ASA) y Evolución del Sentimiento (*Sentiment Evolution*, SE), otros trabajos (Geetha, & Karthika, 2021; Phan, & Ogunbona, 2020) suelen dividirla en dos subtarefas: Extracción de Aspectos (*Aspect Extraction*, AE) y Clasificación del Sentimiento del Aspecto (*Aspect Sentiment Classification*, ASC); la primera parte tiene como objetivo identificar y extraer el aspecto acerca del cual se está emitiendo una opinión, y la segunda parte se encarga de detectar la polaridad del sentimiento asociada a ese aspecto en la opinión (Geetha, & Karthika, 2021).

Existen cuatro métodos fundamentales para la extracción de aspectos (Akhoundzade, & Devin, 2020):

1. Basados en frecuencia de aparición de las palabras.
2. Basados en sentimientos y relaciones entre aspectos.
3. Utilizando métodos no supervisados.
4. Mediante modelado de tópicos y redes neuronales.

### **COMERCIO ELECTRÓNICO EN CITMATEL ENFOCADO EN EL ANÁLISIS DE SENTIMIENTOS**

La empresa CITMATEL ha desarrollado y mantiene en funcionamiento varias plataformas de comercio electrónico, entre las que se encuentran: [www.superfacil.cu](http://www.superfacil.cu); [www.bazar-virtual.ca](http://www.bazar-virtual.ca); [www.moviles.compra-dtodo.com](http://www.moviles.compra-dtodo.com); [www.ticom47.com](http://www.ticom47.com), donde se ofrecen productos y servicios

que constituyen una línea de negocio estratégica. El departamento comercial de la entidad, con el grupo de marketing y publicidad, analiza diferentes fuentes de información que brindan opiniones textuales de usuarios y clientes, para obtener retroalimentación sobre el funcionamiento en general de venta y en consecuencia tomar acciones. Entre esas fuentes de información están los canales en redes sociales de la empresa y otros específicos vinculados con las plataformas de venta en línea, además de los sistemas de quejas e incidencias internos de la empresa. La información obtenida por diversos canales es generalmente estadística y no se encuentra centralizada. Todo el proceso de consolidación se realiza de forma manual, demorando la trasmisión de la información a los decisores finales y desaprovechando un gran volumen de datos en forma de texto que son generados.

Los clientes expresan diariamente en las plataformas de redes sociales, sus criterios acerca de características, beneficios, deficiencias de los productos y servicios que se ofertan en las plataformas de CITMATEL. El seguimiento y la medición de la satisfacción de los clientes tiene el propósito de monitorear la percepción sobre la calidad de los productos y servicios adquiridos. Por tanto, analizar las opiniones de forma eficaz ayuda a las empresas a mejorar la satisfacción y la fidelidad de los clientes, al mismo tiempo podría aportar un valor añadido al servicio que ofrece a los proveedores con reportes sistemáticos sobre evaluación de la satisfacción de sus productos. Necesariamente, para enfrentar este procesamiento y análisis, hay que aplicar tecnologías de PLN, MT y Análisis de Sentimientos o Minería de Opinión, las cuales posibilitarían obtener datos textuales de valor para la empresa y dotarlos de significado, así como conocer y comprender mejor el estado de satisfacción de los clientes, y orientar las decisiones a mejorar sus experiencias y satisfacer más las necesidades de los clientes.

En este sentido, se muestra a continuación la primera aproximación de un modelo de VoC concebido para la actividad de comercio electrónico en CITMATEL, basado en la aplicación del Análisis de Sentimientos, e inspirado en el esquema general de las soluciones de MT. El modelo VoC-CITMATEL fue concebido mediante una arquitectura en cuatro capas (figura 1):

1. Escucha del cliente
2. Preparación y comprensión de los datos
3. Análisis y descubrimiento
4. Visualización

En el filtrado de la información, así como para su comprensión y análisis, resultaría conveniente disponer de un recurso de conocimiento, por ejemplo, donde los productos y servicios que se ofrecen sean semánticamente descritos, sus características y propiedades, lo que se podría lograr mediante unas ontologías. A través de esta base de conocimiento se definirá y representará el vocabulario de términos relevantes del dominio de interés en el ámbito del comercio electrónico de CITMATEL.

Luego de disponer de la información estructurada y almacenada, se llevarían a cabo los procesos de análisis y descubrimiento de conocimiento, en este caso enfocados en el Análisis

de Sentimientos sobre ese contenido textual. Los resultados se deben representar visualmente para hacerlos accesibles y comprensibles por ejecutivos, directivos y decisores en general.

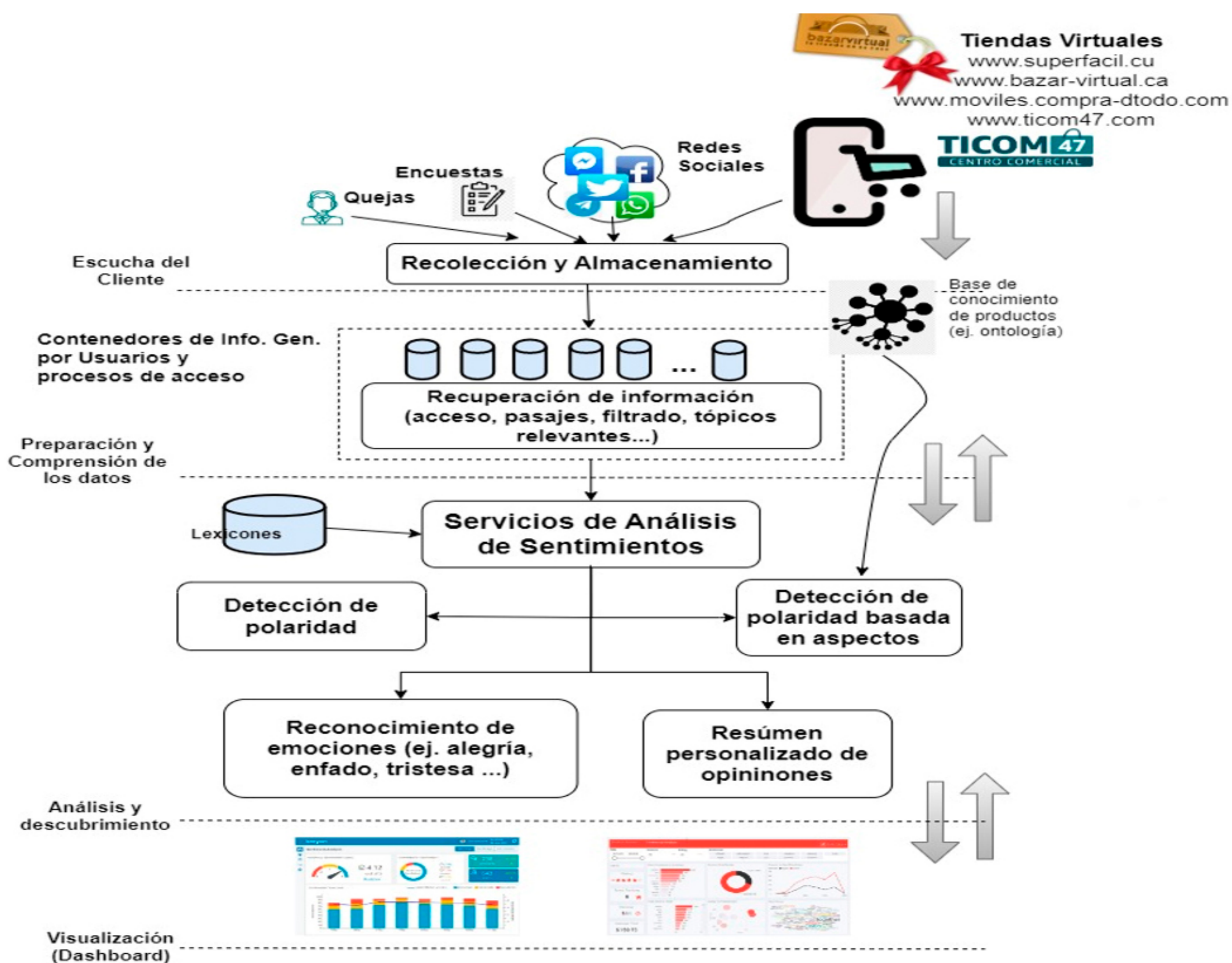


Fig. 1 Esquema general del modelo VoC-CITMATEL, orientado al comercio electrónico de CITMATEL.

1. Escucha del cliente: escuchar a los clientes es una de las formas clave para que en un modelo de negocio de comercio electrónico los decisores se mantengan informados sobre la satisfacción de los usuarios en relación con los productos y servicios que se ofrecen, ya que las opiniones y los comentarios son un medidor de esa satisfacción y proporcionan información procesable para mejorar la experiencia del cliente. En esta capa del modelo se representan las diferentes fuentes de información que los decisores de CITMATEL podrían aprovechar, algunas de las cuales pueden ser parte de sistemas ya implantados como sistemas de quejas, las propias tiendas virtuales (por ejemplo, los clientes podrán compartir sus opiniones sobre los productos y servicios que compran), y fuentes más abiertas como las redes sociales. La diversidad de fuentes contempladas responde a la necesidad de dotar a la organización de una gestión multicanal para la retroalimentación del cliente, y cada una de ellas, por su naturaleza (tipo y estructura de la información, objetivos de su recolección, plataforma origen, etc.), requieren servicios de recolección



- con características específicas, los cuales también forman parte del necesario proceso de homogenización de la estructuración del contenido desde la fuente primaria hacia su comprensión y análisis posterior.
2. Preparación y comprensión de los datos: en esta capa se comienza a enfrentar el desafío de dotar de significado a los datos textuales embebidos en el gran volumen de información recolectada. En ese empeño comienzan a jugar un rol primordial las técnicas de PLN, considerando como punto de partida la transformación de los contenidos textuales en datos estructurados, de tal forma que puedan ser procesados computacionalmente de una manera más eficaz. Como parte de este propósito se incluye un conjunto de tareas básicas de PLN, como: construcción de los índices asociados a cada una de las fuentes; servicios de recuperación de información, modelado de tópicos y frases relevantes; y la construcción de una base de conocimiento (ontología) del dominio que contribuirá a dotar de significado a los datos que se extraigan. En esta capa se deben incluir servicios que preparan los datos para tareas y objetivos más complejos de análisis y descubrimiento en textos.
  3. Análisis y descubrimiento: en esta capa se concentran las tareas de Minería de Texto de más alto nivel de complejidad y alcance, las cuales se enfocan fundamentalmente en el Análisis de Sentimientos o Minería de Opinión. El enfoque en este tipo de tareas está en correspondencia con el impacto de los resultados, en cuanto a conocer la satisfacción de los usuarios. Todas estas tareas se debían desarrollar como servicios y deben incluir la detección de polaridad (para conocer valoraciones generales sobre positividad o negatividad del contenido), la detección de polaridad basada en aspectos (para llevar el Análisis de Sentimientos a nivel de aspectos o características de productos), el reconocimiento de emociones (para descubrir otros tipos de información del contenido, por ejemplo, estados de alegría, enfado, sorpresa, entre otros), la generación de resúmenes personalizados de opiniones (para sintetizar el gran volumen de información generada y hacerla manejable por los decisores), entre otros servicios.
  4. Visualización: el uso de cuadros de mando visuales ayuda a realizar un seguimiento de los datos, la información y el conocimiento obtenido del análisis de la información textual, lo que permite ir más allá de las métricas y empezar a comprender la experiencia emocional de los clientes. Esto ayuda a observar cómo los cambios están afectando a ciertos productos y sus características, y su correlación con las métricas de satisfacción general.

### **CONCEPCIÓN DE SOLUCIÓN PARA DETERMINAR POLARIDAD BASADA EN ASPECTOS**

A continuación se describe la concepción de un primer acercamiento de solución para la detección de polaridad basada en aspectos en opiniones de usuarios, la cual sigue un enfoque no supervisado y está inspirada en lo reportado en López, & Arcos (2019). Existen otras soluciones con enfoque supervisado (Sarkar, 2019); pero requieren ser entrenadas con colecciones de opiniones etiquetadas, las cuales no siempre están disponibles, como en el caso de

CITMATEL. Además, las soluciones no supervisadas suelen ser más generalizables, ya que utilizan solo información contenida en los documentos de entrada. La solución fue concebida en tres fases: Preprocesamiento, Extracción de aspectos y Determinación de polaridad, según se muestra en la figura 2. Es importante aclarar que los aspectos que se extraen están asociados a una determinada entidad o producto en específico.

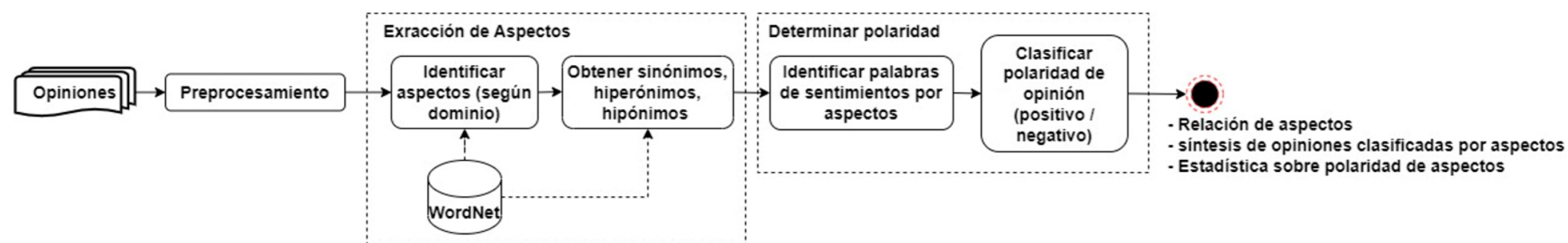


Fig. 2 Flujo de trabajo para determinar polaridad basada en aspectos de textos de opinión.

### Preprocesamiento

Este proceso inicia con la segmentación del texto en palabras a través de la extracción de los tokens, luego todas las palabras se llevan a minúsculas, se eliminan los caracteres especiales y las palabras sin significado relevante (o palabras de parada), y se realiza el etiquetado POS (*Part-of-Speech Tagging*). Para la determinación de polaridad todas las opiniones pasan un filtro y aquellas donde se encuentren los aspectos extraídos, son las que pasan a la fase de predicción de polaridad. Estas reseñas sin incongruencias son enviadas al proceso de extracción de aspectos.

### Extracción de aspectos

Una vez realizado el preprocesamiento de las opiniones, se extraen los aspectos candidatos en las opiniones. Este proceso se lleva a cabo un conjunto de patrones léxico-sintácticos, los cuales están definidos sobre la base de las etiquetas de sustantivo (NN) o frases sustantivas (NP). En la tabla 1 se relacionan los patrones para la extracción de aspectos. Las frases sustantivas que cumplan con alguno de los patrones se consideran como aspectos candidatos. Esta técnica es la más utilizada para la extracción de aspectos, fundamentalmente por su simplicidad y eficiencia.

Tabla 1. Patrones léxico-sintácticos para identificación de aspectos

Descripción	Patrón
sustantivo singular + sustantivo singular	<NN><NN>
sustantivo singular + sustantivo plural	<NN><NNS>
sustantivo plural + sustantivo singular	<NNS><NN>
adjetivo + sustantivo singular	<JJ><NN>
adjetivo + sustantivo plural	<JJ><NNS>

Luego, por cada aspecto se calcula su frecuencia de aparición en las opiniones y se seleccionan aquellos que aparezcan con mayor frecuencia. Los aspectos que resulten más frecuentes son filtrados según su relación semántica (similitud) con la entidad o el producto que está siendo objeto de análisis. El cómputo de esta relación semántica se lleva cabo

usando la medida de Wu & Palmer, que se encuentra disponible en el paquete de WordNet: Similarity (Shahzad, Pervaz, & Nawab, 2018). Todos los aspectos cuya relación semántica supere un umbral predefinido serán considerados aspectos del producto. El próximo paso consiste en encontrar, por cada uno de estos aspectos, su sinónimo, hiperónimo e hipónimo más cercano, usando la base de datos léxica WordNet. Los sinónimos en WordNet, agrupados en *synsets*, son conjuntos de palabras que comparten el mismo significado, mientras que los hipónimos e hiperónimos son las palabras que se encuentran semánticamente relacionadas dentro de la jerarquía, ya sea como superiores o inferiores.

El proceso descrito se aplicaría cuando solo se pueda utilizar el contenido de las opiniones. Otra alternativa, como la que se proyecta desarrollar en CITMATEL (figura 1), sería utilizando una ontología en la cual se describen semánticamente todos los productos y servicios que se ofrecen en la plataforma de venta *online* y sus características (aspectos). En este caso, la extracción de aspectos se guiaría por lo representado en la ontología y se complementaría con el uso de los patrones descrito para alcanzar mayor cobertura del texto.

### **Determinar polaridad**

La determinación de la polaridad inicia con el filtrado de todas las opiniones, seleccionando solo aquellas que contengan los aspectos elegidos en la fase anterior, o sinónimos, hipónimos e hiperónimos de los aspectos. Luego es necesario determinar cuáles son las palabras que se están utilizando para expresar la valoración o el sentimiento sobre el aspecto. Para ello, se estructuran las oraciones en n-gramas, cuyo tamaño se determina teniendo en cuenta dos factores fundamentales: la cantidad de palabras que conforman el aspecto y la longitud de la oración. A partir de ubicar el aspecto dentro de los n-gramas, se seleccionan las palabras asociadas en esa estructura y se procede a determinar su polaridad. En este caso se lleva a cabo la clasificación de la polaridad de la opinión utilizando VADER (*Valence Aware Dictionary for Sentiment Reasoning*) (Hutto, & Gilbert, 2014). VADER es una solución para determinar polaridad basada en lexicón, que devuelve cuatro valores fundamentales: positivos, negativos, neutros y compuestos, que significan la valencia y la magnitud de la emoción en el texto. VADER también identifica texto que no es emocional o neutro en su valencia (Hu, & Liu, 2004). El valor compuesto no se tuvo en cuenta, por tanto, se utilizaron los valores positivo, negativo y neutro. Cada palabra que se utiliza para hacer referencia a un aspecto es evaluada en VADER y se determina su orientación, considerando el valor más alto de los tres posibles. Si la orientación es neutra, se usan dos listas de palabras positivas y negativas para determinar su polaridad.

### **Evaluación y ejemplificación**

Esta solución propuesta fue evaluada en la clasificación de dos tipos de sentimientos: positivo y negativo, de opiniones escritas en idioma inglés sobre una cámara fotográfica y recolectadas de la tienda virtual Amazon (tabla 2) (Hu, & Liu, 2004). Los resultados se computaron utilizando las métricas *Precisión*, *Recall* y *F1-Score*, los cuales se muestran en la tabla 3 y reflejan

la calidad de los resultados de la solución. Además, en la figura 3 se ejemplifican resultados estadísticos que se obtienen con esta solución, para el análisis de la polaridad de una representación de los 15 aspectos (por ejemplo, zoom, flash, picture, quallity) extraídos, sobre los cuales hay opiniones en el conjunto de datos de prueba.

Tabla 2. Caracterización del corpus de evaluación

Conjunto de datos	Total de opiniones	Opiniones positivas	Opiniones negativas
Canon Powershot G3	286	225	61

Tabla 3. Resultados

Precisión	Recall	F1-Score
92,8	99,0	95,4

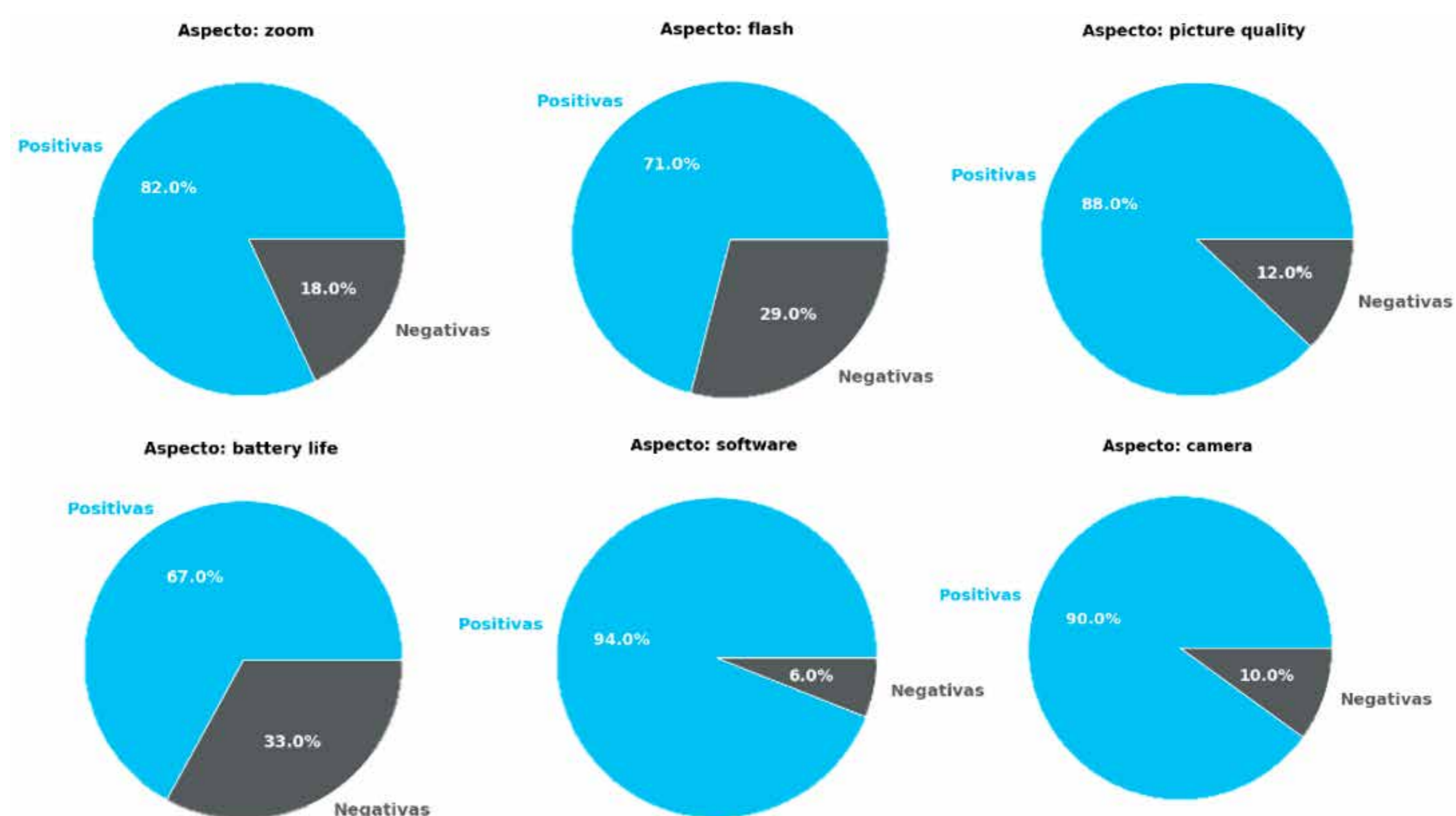


Fig. 3 Ejemplificación de resultados estadísticos de polaridad sobre aspectos en opiniones de la cámara Canon.

## CONCLUSIONES

La transformación del comercio electrónico en CITMATEL, mediante el uso de Inteligencia Artificial y, en particular, de PLN y Análisis de Sentimientos, resulta imprescindible para garantizar la excelencia en los servicios. El análisis de las opiniones de los clientes es vital para ofrecer soluciones más personalizadas y clave en la relación entre las decisiones de los directivos y los procesos de comercialización. En este trabajo se presentó un primer acercamiento para introducir estas tecnologías en CITMATEL, con la intención de implementar un proyecto de VoC, y para ello, se concibió y mostró el modelo de procesamiento y análisis que sirve de guía en este empeño, al tiempo que ofrece una panorámica de lo que se puede desarrollar en este sentido. La solución de detección de polaridad basada en aspectos presentada, permitió ejemplificar posibles datos de interés (por ejemplo, valoración sobre negatividad y positividad

de opiniones de una característica de la cámara), resultantes del análisis del contenido de opiniones. En esos resultados se aporta información valiosa sobre la satisfacción de usuarios acerca de elementos característicos de un producto, en este caso, una cámara, siendo estos resultados parciales muy alentadores.

## GLOSARIO DE TÉRMINOS

**Tecnologías de Inteligencia Artificial (IA):** combinación de algoritmos que se utilizan con el propósito de crear máquinas o sistemas informáticos que presenten las mismas capacidades que el ser humano.

**Procesamiento del lenguaje natural (PLN):** es un campo dentro de la IA y la lingüística aplicada que estudia las interacciones, mediante uso del lenguaje natural entre los seres humanos y las máquinas.

**Minería de texto (MT):** proceso de transformar un texto no estructurado en datos estructurados para proceder posteriormente al análisis.

## REFERENCIAS

- Aboelela, E. M., Gad, W., & Ismail, R. (2021). The impact of semantics on aspect level opinion mining. *PeerJ Computer Science*, 7, pp. 1-22.
- Akhoundzade, R., & Devin, K. H. (2020). Unsupervised aspect-based Sentiment Analysis in the Persian language: Extracting and clustering aspects. In *Proceedings of the 10th International Bawack, E. B., Wamba, S. F., Carillo, K. D. A, & Akter, S (2022). Artificial intelligence in E-Commerce: a bibliometric study and literature review. Electron Mark, 32(1): 297-338.*
- Brauwers, G. & Frasincar, F. (2021). A Survey on Aspect-Based Sentiment Classification. *ACM Computing Surveys*, 1(1).
- García-Pablos, A., Cuadros, M. & Rigau, G. (2018). W2VLDA: Almost Unsupervised System for Aspect Based Sentiment Analysis. *Expert Systems with Applications*, 91: 127-137.
- Geetha, M. P., & Karthika Renuka, D. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. *International Journal of Intelligent Networks*, 2, pp. 64-69.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168-177, Seattle, Washington, USA.
- Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1): 216-225, AAAI Press.
- Kian, R. (2021). Provide a model for an e-commerce system with the impact of artificial intelligence. *International Journal of Innovation in Management, Economics and Social Sciences*, 1(3): 88-94.

- Mehtab, A. M. (2021). Sentiment Analysis in E-Commerce. URL: <https://medium.com/federatedai/sentiment-analysis-in-e-commerce-e8a06a498a75>. [consultado: 1 de diciembre de 2022]
- Liu, B. & L. Zhang. (2012). A Survey of Opinion Mining and Sentiment Analysis. In Mining Text Data, C. C. Aggarwal and C. Zhai (Eds). Springer US: Boston, MA, pp. 415-463.
- Liu, N., Shen, B., Zhang, Z., Zhang, Z., & Mi, K. (2019). Attention-based Sentiment Reasoner for aspect-based sentiment analysis. *Human-centric Computing and Information Sciences*, 9(35): 1-17.
- López, D. & Arcos, L. (2019). Deep learning for aspect extraction in textual opinions. *Revista Cubana de Ciencias Informáticas*, 13(2): 105-145.
- Nazir, A., Y. Rao, L. Wu, & L. Sun, (2022). Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey. *IEEE Transactions on Affective Computing*, 13(2): 845-863.
- Phan, M. H., & Ogunbona, P. (2020) Modelling Context and Syntactical Features for Aspect-based Sentiment Analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3211-3220, Seattle, Washington, USA.
- Saraswat, A., Kumar Azad, H., & Abhishek, K. (2022). Towards improving e-commerce customer review analysis for sentiment detection. *Scientific Reports*, 12(1): 1-15.
- Sarkar, D. (2019). Text Analytics with Python. A Practitioner's Guide to Natural Language Processing, 2nd ed. APress.
- Shahzad, K., Pervaz, I., & Nawab, A. (2018). WordNet-based Semantic Similarity Measures for Process Model Matching. In *Proceedings of the 17th International Conference on Perspectives in Business Informatics Research*. Stockholm, Sweden, pp. 33-44.
- Subhashini, L. D. C. S., Li, Y., Zhang, J., Atukorale, A. S., & Wu, Y. (2021). Mining and classifying customer reviews: a survey. *Artificial Intelligence Review*, 54(8): 6343-6389.
- Varathan, K., A. Giachanou, & F. Crestani. (2017). Comparative Opinion Mining: A Review. *Journal of the Association for Information Science and Technology*, 68(4): 811-829. *Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 94-100. Mashhad, Iran.

