



---

# EVALUACIÓN DE ALTERNATIVAS DE UN PRODUCTO EN PROCESO DE COMPRA UTILIZANDO WEB SCRAPING Y ANÁLISIS DE SENTIMIENTOS

---

*“Desarrollo de una herramienta que evalúa las alternativas de un producto según las opiniones de los clientes y, según las necesidades de cada cliente, recomienda la alternativa óptima”*

Ignacio Fabián Mondino

Universidad Católica Argentina

Facultad de Ingeniería y Ciencias Agrarias

Tutor: Emilio Picasso

Buenos Aires, Argentina

8 de Septiembre de 2022

## Resumen

La cantidad de opciones a la hora de comprar cualquier producto es inmensa y cambiante lo que implica que el consumidor invierta muchos recursos, principalmente tiempo, para poder definir cuál de las opciones elegir.

Si bien existieron algunas herramientas que intentaron facilitar este proceso, no se encontró ninguna que esté actualmente disponible al público. En el presente estudio se propone una herramienta capaz de automatizar el proceso de evaluación de alternativas u opciones en la compra de un producto y sugerir la alternativa óptima según las necesidades de cada cliente.

En esta investigación, la cual sigue la metodología CRISP-DM para un desarrollo organizado, se obtienen los datos de las alternativas y sus opiniones aplicando web scraping al sitio web de Mercado Libre. A partir de las reseñas del producto recolectadas, se descubren las necesidades del cliente. Luego, mediante un modelo de análisis de sentimientos basado en aspectos se discrimina el grado de satisfacción de cada necesidad del cliente en cada opinión. Gracias a la matriz de la calidad (QFD), se puede atribuir dicho grado de satisfacción a los atributos del producto. Finalmente, a partir de la importancia que le asigna el consumidor a cada necesidad del producto, se calcula una valoración final para cada alternativa y se sugiere aquella que es óptima para éste. Para que cualquier persona pueda utilizar la herramienta, se desarrolla una página web.

Se toma el caso de estudio de los celulares para poder ver los resultados. En evidencia de estos, se demuestra que la herramienta facilita considerablemente el proceso de compra principalmente por una gran reducción en el consumo de tiempo.

*Palabras clave:* alternativas, opiniones, necesidades del cliente, metodología CRISP-DM, análisis de sentimientos, matriz de la calidad.

## Índice

<b>1. INTRODUCCIÓN</b>	<b>5</b>
1.1. Problema a investigar	5
1.2. Estado del arte	7
1.3. Objetivos	12
1.4. Estructura del informe	13
<b>2. MARCO TEÓRICO</b>	<b>14</b>
2.1. Framework	14
2.1.1. Framework en minería de datos y su importancia	14
2.1.2. Metodología CRISP-DM	14
2.2. Web scraping	19
2.3. Necesidades del cliente, atributos del producto y su relación	20
2.4. Matriz de la calidad	21
2.5. Descubrimiento de necesidades del cliente a partir de opiniones del producto	23
2.5.1. Problemas a abordar	24
2.5.2. Enfoques	25
2.6. Identificación de necesidades del cliente en opiniones del producto	27
2.6.1. Menciones explícitas e implícitas	27
2.6.2. Riesgo	28
2.7. Análisis de sentimientos	28
2.7.1. Definición	29
2.7.2. Tipos de análisis de sentimientos	29
2.7.3. Herramienta para implementación	32
<b>3. MATERIALES Y MÉTODO</b>	<b>33</b>
3.1. Diseño de la investigación	33
3.1.1. Tipo de investigación	33
3.1.2. Enfoque de la investigación	34
3.1.3. Población y muestra	34
3.1.4. Variables	35
3.2. Técnicas de recolección de datos	38
3.2.1. Consideraciones	38
3.2.2. Pseudocódigo	40
3.3. Técnicas de procesamiento de datos	41
3.3.1. Preparación de los datos	43
3.3.1.1. Preprocesamiento de opiniones	44
3.3.1.2. Preprocesamiento de alternativas	47
3.3.1.3. Creación de matriz de la calidad	54
3.3.2. Modelado	55
3.3.2.1. Atribución de sentimiento a necesidades del cliente	56
3.3.2.2. Atribución de sentimiento a atributos	60
3.3.3. Despliegue	68
3.3.3.1. Asignación de pesos a necesidades del cliente	68
3.3.3.2. Cálculo de importancia técnica de atributos	69

	4
3.3.3.3. Cálculo de valoración final de alternativas	69
<b>4. RESULTADOS</b>	71
4.1. Preparación de los datos	71
4.1.1. Preprocesamiento de opiniones	71
4.1.1.1. Limpieza de opiniones	72
4.1.1.2. Descubrimiento de necesidades del cliente	72
4.1.2. Preprocesamiento de alternativas	75
4.1.2.1. Selección de atributos	76
4.1.2.2. Conversión de atributos con texto a atributos numéricos	77
4.1.2.3. Limpieza de alternativas	79
4.1.2.4. Eliminación de atributos constantes	80
4.1.2.5. Categorización de atributos numéricos continuos	80
4.1.3. Creación de matriz de relaciones	83
4.2. Modelado	84
4.2.1. Atribución de sentimiento a necesidades del cliente	84
4.2.1.1. Identificación de las necesidades del cliente en las opiniones	84
4.2.1.2. Asignación de sentimiento a necesidades del cliente	86
4.2.2. Atribución de sentimiento a atributos	88
4.2.2.1. Atributos con sentimiento por valor	88
4.2.2.2. Atributos con sentimiento por alternativa	91
4.3. Despliegue	93
<b>5. DISCUSIÓN</b>	102
5.1. Análisis de resultados	102
5.2. Limitaciones	105
5.3. Futuras investigaciones	107
<b>6. CONCLUSIONES</b>	108
<b>7. REFERENCIAS BIBLIOGRÁFICAS</b>	111
<b>8. ANEXO</b>	114

# 1. INTRODUCCIÓN

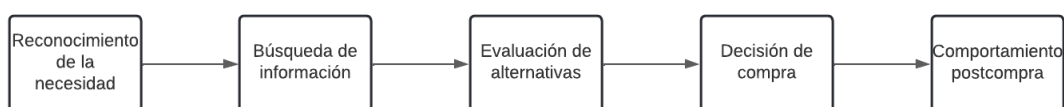
## 1.1. PROBLEMA A INVESTIGAR

Si bien el proceso de decisión de compra (*Buyer decision process*) es de especial interés para los especialistas en marketing, sirve a efectos de conocer el camino que todo consumidor recorre a la hora de comprar cualquier producto.

Existen varios modelos que intentan explicar en una serie de pasos cómo es que el consumidor parte desde una necesidad hasta que efectúa la compra. Sin embargo, el modelo más utilizado es el propuesto por el estadounidense John Dewey en 1910. Este consiste de 5 etapas: Reconocimiento de la necesidad, Búsqueda de información, Evaluación de alternativas, Decisión de compra y Comportamiento postcompra (Figura 1).

**Figura 1**

*Etapas de modelo de proceso de compra del consumidor*



Madhavan y Chandrasekar (2015) resaltan que el proceso de decisión de compra comienza mucho antes de la compra real y continúa incluso después de ésta. El punto de partida es cuando el consumidor reconoce una necesidad la cual Kotler (2017) define como una diferencia entre su estado real y algún estado deseado. Luego, es probable que el consumidor realice una búsqueda de información relacionada con la necesidad consultando fuentes ya sea personales (familia, amigos, vecinos, conocidos), comerciales (publicidad, vendedores, Internet, empaques, exhibidores), públicas

(medios de comunicación, organizaciones de calificación de consumidores) y/o experienciales (manipular, examinar, usar el producto). El siguiente paso y más importante a efectos del estudio, es la evaluación de alternativas en la cual el consumidor usa toda la información recopilada para evaluar las distintas alternativas posibles. Cuando el cliente escoge una de las alternativas, decide lugar y medios de compra y efectivamente la realiza, se lo identifica como la cuarta etapa, la decisión de compra. Por último, ya después de la adquisición, el consumidor toma medidas o realiza acciones según si está satisfecho o insatisfecho con ésta.

Como se mencionó, el modelo propuesto en la Figura 1 no es el único que existe, sin embargo, en todos ellos (Prasad y Jha, 2014) siempre se lleva a cabo la tarea de **evaluar alternativas**. Eso no quita que los consumidores puedan saltarse esta etapa. De hecho, Kotler (2017) afirma que esto podría suceder en las compras más rutinarias y da el ejemplo de una mujer que reconoce la necesidad de comprar pasta de diente y directamente realiza la compra, omitiendo la búsqueda de información y evaluación de alternativas, puesto que tiene una marca habitual o favorita. Sin embargo en muchos casos esto no es posible y a menudo estas decisiones son difíciles. Hung-Cheng y Shih-Wen (2004) indican que muchos consumidores son conscientes de sus necesidades básicas, pero no tienen idea de los tipos de productos, funciones, usos y terminologías que muestran información importante. Y a pesar de que los consumidores tengan conocimiento sobre el producto, normalmente, tal como Payne et al (1991) expone, se enfrentan a una gran cantidad de alternativas, que cambian constantemente debido a las nuevas tecnologías y las presiones competitivas. Sumado a todo esto, existe alguna evidencia que sugiere que los individuos pueden procesar óptimamente un máximo de sólo seis alternativas (Malhotra, 1982) cuando en el mercado de un solo producto suele haber cientos de éstas.

El desconocimiento del cliente sobre las alternativas del producto implica una inversión de tiempo muy grande por parte de éste exclusivamente para conocer el producto, sus atributos y/o las posibles alternativas, siendo que aún falta la selección de una de ellas. Adicionalmente, dada la incapacidad de las personas de analizar todas las alternativas de un producto, el consumidor selecciona una dentro de un conjunto reducido de éstas, lo que aumenta el riesgo de dejar fuera la alternativa “óptima” para el mismo.

Por tales motivos, es un gran aporte el desarrollo de una herramienta que automáticamente analice todas las alternativas del producto y, según las necesidades del consumidor, recomiende aquella que es óptima en cuestión de minutos.

## 1.2. ESTADO DEL ARTE

Los consumidores constantemente toman decisiones con respecto a la elección, compra y uso de productos y servicios, por ende, sería ideal escoger la alternativa óptima con alto grado de confiabilidad y en un corto periodo de tiempo. La siguiente revisión de la literatura confirma que el proceso de evaluación de alternativas no tiene una única forma de llevarse a cabo por los consumidores y que puede resultar muy difícil por diversos factores. Además, se discuten algunas soluciones posibles para facilitar dicha tarea.

Varios autores hablan acerca de los elementos o factores presentes en la decisión del consumidor. Punj y Stewart (1983), basándose en los procesos propuestos por otros autores, describen las características de dicho proceso en cuatro grandes categorías. Acorde al estudio, estas son:

1. Características de la solución

- a. Número de alternativas disponibles (multiplicidad de soluciones): cantidad de opciones de productos o servicios que el consumidor puede considerar en cualquier momento del proceso de toma de decisiones.
- b. Incertidumbre de la decisión (verificabilidad de la solución): grado en que el consumidor puede determinar que la elección de soluciones alternativas es correcta u óptima antes de la selección de la misma.
- c. Frecuencia de la decisión: El número de veces que el consumidor prevé elegir entre alternativas durante el proceso de toma de decisiones.
- d. Importancia de la decisión: grado en que una decisión incorrecta es costosa (en tiempo y dinero) para el responsable de ésta.
- e. Familiaridad de la población (experiencia con la tarea): la frecuencia con la que la población en general se ha enfrentado a la tarea en el pasado.

## 2. Características de la información

- a. Número de atributos por alternativa: número de características que describen una alternativa y que están a disposición del consumidor en algún momento del proceso de toma de decisiones.
- b. Compatibilidad de atributos: grado en que la información de los atributos está disponible, y de forma similar, para todas las alternativas.
- c. Formato de la información (presentación secuencial de la información): patrón espacial y temporal de la presentación de la información.
- d. Fuente de información: el medio por el que se obtiene la información.
- e. Forma de la información: el grado en que la información se presenta en forma verbal o numérica, en términos abstractos o concretos, en términos absolutos o relativos, y en forma resumida (nombres de marcas) o detallada.
- f. Nivel de ruido (distracción): cantidad de información disponible para el consumidor durante el proceso de decisión que no está relacionada con la elección entre las alternativas.



### 3. Características temporales

- a. Presión de tiempo: grado en que una decisión debe tomarse en un período de tiempo inferior al necesario para una toma de decisiones óptima.

### 4. Características de cooperación

- a. Características de cooperación: grado en que una decisión requiere la participación de más de un individuo.

De los trabajos que estudian las formas del consumidor de evaluar las alternativas se deduce que no hay una metodología única para llevarla a cabo. Kotler (2017) menciona que, desafortunadamente, los consumidores no utilizan un proceso de evaluación simple y único en todas las situaciones de compra. Depende del consumidor individual y de la situación de compra específica. El propio Kotler nombra algunos modelos. Primero, el modelo de valor de expectativa de elección del consumidor, en el cual la mayoría de los compradores consideran varios atributos, pero asignan una importancia diferente a cada uno. Otro es el modelo conjuntivo de elección del consumidor en el que evalúan únicamente alternativas que satisfagan un conjunto de niveles mínimos de atributos. Este último, también es nombrado por Madhavan and Chandrasekar (2015).

Los autores que tal vez aporten más información respecto de las formas de los consumidores de evaluar alternativas son Payne et al (1991). Estos nombran un gran número de estrategias, sin embargo, entre las más relevantes se encuentra la regla de la suma ponderada (WADD) y la Heurística de Satisfacción (SAT).

La primera, considera los valores de cada alternativa en todos los atributos relevantes y considera todas las importancias relativas de los atributos para el tomador de decisiones. Específicamente, desarrolla un valor ponderado para cada atributo multiplicando el peso por el valor del atributo y sumando todos los atributos para llegar a una evaluación general de una alternativa. Se supone que se elige la alternativa con

la evaluación global más alta. No obstante, el autor afirma que si bien las personas a veces toman decisiones de manera consistente con un procedimiento de este tipo, en la mayoría de los casos utilizan procesos de decisión más simples.

En la segunda, las alternativas se consideran una a la vez, en el orden en que ocurren en el conjunto. Se considera el valor de cada atributo de una alternativa para ver si cumple con un nivel de corte predeterminado. Si algún valor de atributo está por debajo del límite, se rechaza esa alternativa. Se elige la primera alternativa que tenga valores que cumplan con los requisitos de corte para todos los atributos. Si ninguna alternativa pasa todos los puntos de corte, el nivel de corte se puede relajar y repetir el proceso, o se puede seleccionar una alternativa al azar. Una implicación de la heurística satisfactoria es que la elección estará en función del orden en que los consumidores evalúan los productos. Es decir, si tanto la marca A como la marca B superan los niveles de corte, la elección de A o B dependerá de si se evalúa primero A o B.

Muchos son los autores que hablan de la dificultad de la toma de decisión del consumidor. No hay un acuerdo general de cuáles son los factores que aumentan dicha dificultad pero hay algunos en común.

Acorde a Payne et al (1991), la dificultad aumenta cuando:

1. Aumenta el número de alternativas.
2. Se acrecienta el número de atributos por alternativa.
3. Mayor es la incertidumbre de la decisión, por ejemplo, poca seguridad en los valores de muchos atributos.
4. Disminuye la compatibilidad de atributos, es decir, que el número de atributos compartidos se vuelve más pequeño.
5. Fuente de información.
6. Crece la importancia de la decisión. Payne asegura que para la mayoría de los consumidores, existe una enorme diferencia entre elegir una marca de mayonesa y comprar un automóvil. En el primer caso, la decisión suele ser

rutinaria, tiene relativamente pocas consecuencias y se toma casi automáticamente, con poco esfuerzo. En el último caso, las consecuencias son mucho mayores, y el consumidor a menudo dedicará un gran esfuerzo, buscando grandes cantidades de información, solicitando asesoramiento y agonizando por las difíciles compensaciones.

Otros autores como Madhavan y Chandrasekar (2015) asocian una mayor dificultad cuando:

1. Más infrecuente es la decisión, es decir, se compran con poca frecuencia (por ejemplo, una computadora, un televisor, un automóvil o una casa).
2. Crece la importancia de la decisión, principalmente, por el factor precio.
3. Menor es la familiaridad de la población, es decir, cuando se compran productos desconocidos.

A pesar de lo difícil que puede llegar a ser la toma de decisión del consumidor, no existieron muchas soluciones. En la literatura, se puede ver que se construyeron algunas herramientas pero actualmente ninguna de éstas está disponible.

Butler y Peppard (1998) mencionan el caso de "ComparaNet" que tenían un programa que, para una amplia gama de bienes y servicios, buscaba en la Web el producto que el consumidor especificaba. Luego, ayudaba al consumidor a elegir y ordenar los criterios relevantes para la evaluación de ese producto. Finalmente, producía instantáneamente datos tabulados que permitían al comprador comparar las opciones.

Hung-Cheng y Shih-Wen (2004) propusieron un método para ayudar a los consumidores a seleccionar el producto "adecuado" para satisfacer sus necesidades. El procedimiento llevado a cabo fue:

1. Adquisición de las necesidades de los clientes a partir de encuestas informales a los clientes.
2. Consultores de investigación identificaron 10 características con sus respectivos valores a partir de un análisis de las distintas alternativas del producto.
3. Luego de investigar las necesidades del cliente, diseñaron cuestionarios para poder obtener, mediante el criterio de los mismos consultores de investigación, las relaciones entre las necesidades de los clientes y los atributos del producto.
4. Finalmente usan un modelo llamado modelo borroso (*fuzzy model*) que, según la importancia que el cliente le da a cada necesidad, obtiene el valor ideal para cada atributo del producto.

Luego, al consumidor se le presentaba la mejor combinación de valores de los atributos del producto y este podía interactuar con la herramienta mediante una interfaz de consulta asistida por Internet.

La evaluación de alternativas es una tarea cada vez más difícil por la creciente y cambiante oferta. La literatura revisada en el estudio define dicho problema y muestra que actualmente no hay una solución.

### 1.3. OBJETIVOS

En este trabajo se buscará desarrollar una herramienta capaz de evaluar las alternativas de un producto y, según las necesidades de cada cliente, recomendar la alternativa óptima a cada uno.

Para ello se tendrán que responder las siguientes preguntas de investigación:

- ¿Cómo descubrir las necesidades del cliente de un producto?

- ¿Cómo identificar las alternativas de un producto y sus atributos?
- ¿Cómo relacionar las necesidades del cliente con los atributos del producto?
- ¿Cómo determinar qué valores de un atributo son mejores que otros?
- ¿Cómo comparar las alternativas entre sí?
- ¿Cómo permitir que el consumidor utilice la herramienta?

#### 1.4. ESTRUCTURA DEL INFORME

El resto del documento está organizado de la siguiente manera. El capítulo 2 define conceptos básicos y las bases teóricas de las herramientas y técnicas utilizadas en el trabajo. En el capítulo 3, se exponen los pasos para la extracción, organización y análisis de los datos. En el capítulo 4, se presentan los resultados de la herramienta a partir de un caso de estudio. En el capítulo 5 se exponen e interpretan los resultados. En el capítulo 6, se desarrollan las conclusiones del trabajo.

## 2. MARCO TEÓRICO

### 2.1. FRAMEWORK

#### 2.1.1. FRAMEWORK EN MINERÍA DE DATOS Y SU IMPORTANCIA

Un *framework* (o metodología) es un modelo estandarizado para poder llevar a cabo un proyecto. Existen varios de estos, inclusive, dentro del campo de Minería de datos.

Las principales utilidades de un *framework* se pueden resumir en que otorga una manera ordenada de resolver un problema minimizando la posibilidad de saltar una tarea importante y, además, sirve para la documentación. En fin, como Wirth y Hipp (2000) afirman, los proyectos son menos costosos, más confiables, más repetibles, más manejables y más rápidos. Además, los autores resaltan que la minería de datos necesita un enfoque estándar que ayude a traducir los problemas comerciales, sugerir transformaciones de datos y técnicas de minería de datos adecuadas, y proporcionar medios para evaluar la efectividad de los resultados y documentar la experiencia.

Azevedo y Santos (2008) aseguran que los *frameworks* más utilizados en trabajos científicos de minería de datos son: SEMMA y CRISP-DM. Según los autores, a pesar de que a primera vista CRISP-DM podría ser más completo, ambas son metodologías muy similares. No obstante, Huber et al (2019) ratifican que el estándar de facto para realizar análisis de datos en aplicaciones industriales es la metodología CRISP-DM.

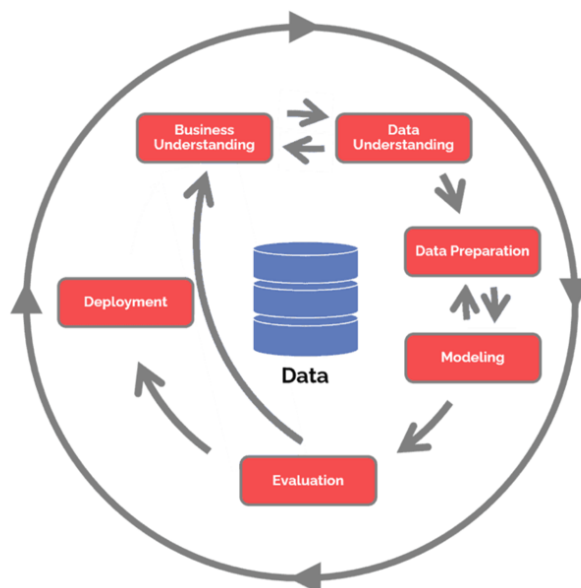
#### 2.1.2. METODOLOGÍA CRISP-DM

CRISP-DM (*Cross Industry Standard Process for Data Mining*) se puede definir como un modelo que proporciona un marco para llevar a cabo proyectos de minería de datos. Éste es independiente tanto del sector industrial como de la tecnología utilizada y se

compone de seis etapas o fases, las cuales se pueden ver en la Figura 2, cada una compuesta por tareas. A continuación se desarrolla cada etapa y sus tareas.

**Figura 2**

*Etapas de metodología CRISP-DM*



### **I. COMPRENSIÓN DEL NEGOCIO (*BUSINESS UNDERSTANDING*)**

La fase de comprensión del negocio se centra en comprender los objetivos y requisitos del proyecto. Aparte de la tercera tarea, las otras tres son actividades de gestión de proyectos fundamentales que son universales para la mayoría de los proyectos:

- **Determinar los objetivos empresariales:** En primer lugar, hay que comprender a fondo, desde una perspectiva empresarial, lo que el cliente realmente quiere conseguir y, a continuación, definir los criterios de éxito empresarial.
- **Evaluar la situación:** Determinar la disponibilidad de recursos, los requisitos del proyecto, evaluar los riesgos y las contingencias y realizar un análisis de costo-beneficio.
- **Determinar los objetivos de la extracción de datos:** Además de definir los objetivos de negocio, también se debe definir cómo es el éxito desde la perspectiva técnica de la minería de datos.

- **Elaborar el plan del proyecto:** Seleccionar las tecnologías y herramientas y definir planes detallados para cada fase del proyecto.

Mientras que muchos se apresuran a través de esta fase, el establecimiento de una fuerte comprensión del negocio es como la construcción de los cimientos de una casa, es decir, es absolutamente esencial.

## II. COMPRESIÓN DE LOS DATOS (*DATA UNDERSTANDING*)

En la actual etapa se busca identificar, recopilar y analizar los conjuntos de datos que pueden ayudar a lograr los objetivos del proyecto. Esta fase también tiene cuatro tareas:

- **Recoger los datos iniciales:** Adquirir los datos necesarios y cargarlos en su herramienta de análisis.
- **Describir los datos:** Examinar los datos y documentar sus propiedades superficiales, como su formato, el número de registros o las identidades de los campos.
- **Explorar los datos:** Visualizar los datos e identificar las relaciones entre estos
- **Verificar la calidad de los datos:** Identificar y documentar cualquier problema de la calidad de los datos.

## III. PREPARACIÓN DE LOS DATOS (*DATA PREPARATION*)

Una regla común es que el 80% del proyecto es la preparación de los datos. Esta fase prepara el conjunto de datos final para la modelización. Consta de cinco tareas:

- **Seleccionar los datos:** Determinar qué conjuntos de datos se utilizarán y documentar los motivos de inclusión/exclusión.
- **Limpiar los datos:** Suele ser la tarea más larga. Una práctica común durante esta tarea es corregir, imputar o eliminar valores erróneos.
- **Construir datos:** Derivar nuevos atributos que sean útiles.



- **Integrar datos:** Crear nuevos conjuntos de datos combinando datos de múltiples fuentes.
- **Formatear los datos:** Cambiar el formato de los datos según sea necesario. Por ejemplo, convertir los valores de texto que almacenan números en valores numéricos para poder realizar operaciones matemáticas.

#### IV. MODELADO (*MODELLING*)

Lo que se considera el trabajo más emocionante de la ciencia de los datos suele ser también la fase más corta del proyecto. Aquí es probable que se construyan y evalúen varios modelos basados en diferentes técnicas de modelado. Esta fase tiene cuatro tareas:

- **Seleccionar las técnicas de modelización:** Determinar qué algoritmos probar (por ejemplo, regresión, red neuronal).
- **Generar el diseño de la prueba:** A la espera de su enfoque de modelización, es posible que tenga que dividir los datos en conjuntos de entrenamiento, prueba y validación.
- **Construir el modelo.**
- **Evaluar el modelo:** Por lo general, varios modelos compiten entre sí, y el científico de datos tiene que interpretar los resultados del modelo basado en el conocimiento del dominio, los criterios de éxito predefinidos, y el diseño de la prueba.

Aunque la guía CRISP-DM sugiere "iterar la construcción y la evaluación del modelo hasta que se crea firmemente que se ha encontrado el mejor modelo o modelos", en la práctica se debe continuar iterando hasta que se encuentre un modelo "suficientemente bueno", proceder a través del ciclo de vida CRISP-DM, y luego seguir mejorando el modelo en futuras iteraciones.

## V. EVALUACIÓN (*EVALUATION*)

Mientras que la tarea de evaluación del modelo de la fase de modelado se centra en la evaluación técnica de éste, la fase de evaluación examina más ampliamente qué modelo se ajusta mejor a la empresa y qué hacer a continuación. Sus tres tareas son:

- **Evaluar los resultados:** ¿Los modelos cumplen los criterios de éxito de la empresa? ¿Cuál o cuáles se deben aprobar para la empresa?.
- **Revisar el proceso:** Revisar el trabajo realizado. ¿Se ha pasado algo por alto? ¿Se han ejecutado correctamente todos los pasos? Resumir los resultados y corregir lo que sea necesario.
- **Determinar los siguientes pasos:** Sobre la base de las tres tareas anteriores, determinar si se debe proceder a la implantación, iterar más o iniciar nuevos proyectos.

## VI. DESPLIEGUE (*DEPLOYMENT*)

Dependiendo de los requisitos, la fase de despliegue puede ser tan simple como generar un informe o tan compleja como implementar un proceso de minería de datos repetible en toda la empresa. Un modelo no es especialmente útil si el cliente no puede acceder a sus resultados. Esta fase final tiene cuatro tareas:

- **Planificar el despliegue:** Desarrollar y documentar un plan para desplegar el modelo.
- **Planificar la supervisión y el mantenimiento:** Desarrollar un plan de seguimiento y mantenimiento exhaustivo para evitar problemas durante la fase operativa (o posterior al proyecto) de un modelo.
- **Elaborar el informe final:** El equipo del proyecto documenta un resumen del proyecto que puede incluir una presentación final de los resultados de la extracción de datos.
- **Revisar el proyecto:** Realiza una retrospectiva del proyecto sobre lo que ha ido bien, lo que podría haber sido mejor y cómo mejorar en el futuro.

## CONSIDERACIONES

Es conveniente tener presente las siguientes consideraciones:

- la metodología se debe implementar desde el inicio del proyecto puesto que ayuda a definir correctamente el problema a abordar.
- las etapas se llevan a cabo secuencialmente desde la primera hasta la sexta. No se debe pasar a una etapa sin haber terminado la anterior (Wirth y Hipp, 2000).
- según cada proyecto en particular, una tarea puede no ser necesaria y por ende suprimida, se puede cambiar el nombre de algunas de ellas, e inclusive, agregar otras nuevas tal como Wirth y Hipp (2000) hicieron (para separar una tarea en otras más concretas).
- en la fase cinco del modelo se debe decidir si avanzar a la fase seis y desplegar el proyecto, o bien, si volver a iterar, es decir, regresar a la primera etapa para volver a recorrer todas las fases. Wirth y Hipp (2000) aconsejan encarecidamente planificar las iteraciones de forma explícita. Específicamente, sugieren planificar tres iteraciones en las que la segunda y tercera tengan la mitad del tiempo y la cuarta parte del tiempo de la primera iteración respectivamente. En efecto, el estudio de Moro et al (2011) concluye que iterar es de gran valor puesto que gracias a ello lograron mejorar considerablemente sus resultados.

## 2.2. WEB SCRAPING

*Web scraping* (Respaldo web) consiste en la extracción de datos significativos de una o varias páginas web de manera automática. Es preciso tener presente que lo que se conoce como web (o WWW, de las siglas en inglés de *World Wide Web*) no es más que una colección de datos a la que se accede a través de Internet.

Para implementar esta técnica lo más habitual es escribir un programa automatizado que consulte a un servidor web, solicite datos (normalmente en forma de HTML y otros archivos que componen las páginas) y, a continuación, analice esos datos para extraer la información necesaria. En otras palabras, se implementa escribiendo un programa que genera un bot de Internet el cual navega por la web tal como lo hace cualquier persona visitando distintas páginas como Wikipedia, un diario, un blog, etcétera. Sin embargo, conforme navega, solicita datos y los extrae.

Evidentemente, el acceso a datos casi ilimitados tiene muchas aplicaciones muy prácticas: las previsiones de mercado, la traducción automática de idiomas e incluso los diagnósticos médicos se han beneficiado enormemente de la capacidad de recuperar y analizar datos de sitios de noticias, textos traducidos y foros de salud, respectivamente. Incluso en el mundo del arte, el "web scraping" ha abierto nuevas fronteras a la creación. El proyecto de 2006 "*We Feel Fine*", de Jonathan Harris y Sep Kamvar, buscó en diversos sitios de blogs en inglés frases que empezaran por "*I feel*" ("me siento") o "*I am feeling*" ("me estoy sintiendo"). Esto dio lugar a una popular visualización de datos, que describía cómo se sentía el mundo día a día y minuto a minuto (Mitchell, 2018).

### 2.3. NECESIDADES DEL CLIENTE, ATRIBUTOS DEL PRODUCTO Y SU RELACIÓN

Hay una distinción importante entre las **necesidades del cliente** y los **atributos del producto**.

- Una **necesidad del cliente** es una declaración abstracta dependiente del contexto que describe los beneficios, en las propias palabras del cliente, que éste busca obtener de un producto o servicio. En otras palabras, son las

expresiones que los clientes utilizan para describir los productos y sus características deseables.

- Los **atributos** del producto son los medios para satisfacer las necesidades del cliente (Yacuzzi y Martín 2003, Timoshenko y Hauser 2019).

Por ejemplo, para el producto bebida energizante, el consumidor podría expresar la necesidad de “un sabor rico” la cual puede satisfacerse con varios atributos del producto (soluciones), incluidas la cantidad de cafeína o la cantidad de endulzante.

Considere, por ejemplo, el producto celular. En ese caso, el cliente podría desear que el dispositivo "tenga un funcionamiento rápido y no se trabe". Ésta se podría satisfacer con los atributos modelo del procesador, cantidad de núcleos, memoria RAM, etcétera.

## 2.4. MATRIZ DE LA CALIDAD

El despliegue de la función de calidad (o QFD, por las siglas en inglés de *Quality Function Deployment*) es una metodología de desarrollo de productos cuyo objetivo es "desplegar" la voz del cliente a lo largo de dicho proceso. Ésta consiste de una serie de una o más matrices, de las cuales, la primera y más común se conoce como Casa o Matriz de la Calidad (o HOQ, por las siglas en inglés *House Of Quality*) (Hauser et al, 2010). Hauser y Clausing (1988) definen que la base de la matriz de la calidad es la creencia de que los productos deben diseñarse para reflejar los deseos y gustos de los clientes.

La matriz de la calidad, núcleo del QFD, es un mapa conceptual que relaciona las necesidades de los clientes con los atributos del producto necesarios para satisfacerlos (Figura 3).

**Figura 3***Estructura de la matriz de la calidad*

<b>CARACTERÍSTICAS DE CALIDAD</b>	Característica A	Característica B	Característica C	...	Característica M
<b>REQUERIMIENTOS DEL CLIENTE</b>					
Requerimiento A	●	●			
Requerimiento B	●		△		○
Requerimiento C		●	△		
⋮					○
Requerimiento N	○		●		

● Relación fuerte
○ Relación media
△ Relación débil

*Nota.* Adaptado de Matriz de la calidad, por Aiteco Consultores, 2022, (<https://www.aiteco.com/diagrama-de-matriz/>).

El panel rectangular conformado por las intersecciones entre las filas de las necesidades del cliente (el "qué") y las columnas de los atributos (el "cómo") indica la correlación o relación entre estos. Esta correlación expresa cuán bien cada voz del cliente es considerada por un atributo determinado, o, en otras palabras, cuánto afecta a cada necesidad del cliente un atributo específico. Es usual utilizar los símbolos y ponderaciones asociadas de la Figura 4, aunque también se emplean otros símbolos y colores (Yacuzzi y Martín, 2003).

Durante la confección de la matriz de la calidad debe verificarse que para cada atributo exista siempre por lo menos un cruce con una necesidad del cliente, pues de lo contrario, no habría razón para incluir dicho atributo en la matriz. Análogamente, cada necesidad del cliente debe estar correlacionada con uno o varios atributos, porque si no, no se contemplaría, desde el punto de vista de la ingeniería, la voz del cliente (Yacuzzi y Martín, 2003).

**Figura 4**

*Símbolos y valores de relaciones o correlaciones*

<b>Grado de correlación entre RC y CT</b>	<b>Símbolo utilizado</b>	<b>Valor numérico asignado</b>
Muy correlacionados	⊙	9
Correlacionados	○	3
Poco correlacionados	△	1
Sin correlación	Blanco	0

*Nota.* Adaptado de Símbolos utilizados en la matriz de la calidad (p. 12), por Yacuzzi y Martín, 2013, Universidad del Centro de Estudios Macroeconómicos de Argentina (UCEMA)

No toda necesidad del cliente es igual de importante para cada consumidor. Conociendo estas prioridades, se puede calcular la importancia técnica de cada atributo. Ésta se calcula multiplicando el peso dado por los clientes a cada necesidad por la relación entre dicha necesidad y el atributo. Acorde a Yacuzzi y Martín (2003), esta información orienta de inmediato a los técnicos hacia los atributos que deben ser considerados con prioridad (aunque no con exclusividad), ya sea porque son críticos para un número pequeño de necesidades del cliente, o porque son relevantes para un gran número de estas, o cualquier otra combinación que produzca un resultado de ponderación total relativamente elevado.

## 2.5. DESCUBRIMIENTO DE NECESIDADES DEL CLIENTE A PARTIR DE OPINIONES DEL PRODUCTO

El proceso de descubrir las necesidades del cliente para un producto es un problema que data hace mucho tiempo. Tradicionalmente, se identifican a partir de las interacciones humanas con los clientes, como las entrevistas de experiencia y los

grupos de discusión. Normalmente, el proceso comienza con 20-30 entrevistas cualitativas basadas en la experiencia. Luego, varios analistas revisan las transcripciones, destacan las necesidades de los clientes y eliminan la redundancia ("depuración") para producir un conjunto básico de aproximadamente 100 declaraciones abstractas de éstas necesidades dependientes del contexto (Timoshenko y Hauser, 2019).

Sin embargo, los métodos tradicionales son caros y requieren mucho tiempo, por lo que, surge la necesidad de buscar otras formas. Timoshenko y Hauser (2019) relevan algunas formas que han sido propuestas por otros autores como la utilización de una interfaz web para pedir a los clientes que introduzcan directamente sus necesidades e historias, para luego estructurarlas y eliminar la redundancia mediante el juicio humano o, el juicio de los directivos o la revisión de comparaciones de productos en la web. Finalmente, los autores concluyen que los contenidos generados por los usuarios (CGU), como las reseñas en línea, las redes sociales y los blogs, es, como mínimo, una fuente de necesidades de los clientes comparable a la de las entrevistas experimentales, y probablemente una fuente mejor. En muchas categorías, el CGU es extenso; por ejemplo, sólo en Amazon hay más de 300.000 opiniones sobre productos de salud y cuidado personal.

### 2.5.1. PROBLEMAS A ABORDAR

El descubrimiento de las necesidades de los clientes a partir del CGU, como las opiniones de un producto, plantea múltiples problemas. Acorde a Timoshenko y Hauser (2019) estos son:

- la propia escala del CGU hace que sea difícil de procesar para los lectores humanos.



- gran parte del CGU es repetitivo o no es relevante. Frases como "Recomiendo encarecidamente este producto" no expresan las necesidades del cliente. Esto hace que un análisis manual tradicional sea ineficiente.
- la mayor parte del CGU se concentra en unas pocas necesidades de los clientes.
- los datos del CGU no están estructurados y se basan principalmente en el texto. Para identificar las necesidades abstractas de los clientes que dependen del contexto, los investigadores deben comprender los ricos significados que hay detrás de las palabras.
- a diferencia de los métodos tradicionales basados en una muestra representativa de clientes, los clientes se auto-seleccionan para publicar CGU. La autoselección puede hacer que los analistas pasen por alto importantes categorías de necesidades de los clientes.

### 2.5.2. ENFOQUES

Hay varios enfoques para identificar aspectos (en el caso del proyecto, necesidades del cliente) en las opiniones o reseñas del producto. Acorde a Feldman (2013), el enfoque clásico, utilizado por muchas empresas comerciales, consiste en extraer todas las frases sustantivas (PN) y, a continuación, conservar sólo aquellas cuya frecuencia esté por encima de algún umbral determinado experimentalmente. Otro enfoque planteado por el autor es utilizar un analizador de dependencia de frases que utiliza expresiones de sentimiento conocidas para encontrar aspectos adicionales (incluso infrecuentes).

### **Consideraciones**

Primero, la frecuencia con la que aparece una necesidad del cliente en el contenido generado por usuario (CGU) podría ser un indicador de su importancia dado que los clientes deciden si publican (Timoshenko y Hauser, 2019).

Segundo, es recomendable que las necesidades del cliente sean frases y no palabras sueltas para así captar más información. De hecho, Timoshenko y Hauser (2019) proponen que las frases describan con detalle los beneficios que los clientes desean obtener de los productos. Por ejemplo, la necesidad del cliente "batería" en los celulares, no nos permite diferenciar si la necesidad se trata de la duración de ésta, de su carga rápida o de la posibilidad de removerla, etcétera. Lo correcto sería, por ejemplo, "duración de la batería".

Tercero, una de las alternativas para poder identificar frases en un conjunto de texto, tal como una opinión, es extrayendo N-gramas. Un N-Grama es una secuencia de N-palabras en una cadena de texto. En este caso, N es un número entero que representa el número de palabras de la secuencia. Por ejemplo, si  $N=1$  entonces se denomina uni-grama. Si  $N=2$  entonces se trata de un bi-grama. Si  $N=3$  entonces es un trigramas. Y así. Los N-gramas captan el contexto en el que las palabras se utilizan juntas. Por ejemplo, podría ser una buena idea considerar bigramas como "Buenos Aires" en lugar de dividirlo en palabras individuales como "Buenos" y "Aires".

Considérese, por ejemplo, la opinión "Muy conforme con el producto". En la Tabla 1, se puede ver como son los N-gramas para distintos valores de N.

**Tabla 1**

*Ejemplo de N-gramas en una frase para distintos valores de N*

Valor de N	N-gramas generados
1	["Muy", "conforme", "con", "el", "producto"]
2	["Muy conforme", "conforme con", "con el", "el producto"]
3	["Muy conforme con", "conforme con el", "con el producto"]
4	["Muy conforme con el", "conforme con el producto"]

Cuarto, en los casos en que se implementa un modelo de *machine learning* para identificar posibles necesidades del cliente es importante que el último paso lo lleve a cabo un humano. Este consiste en hacer una revisión y eventualmente seleccionar las necesidades que correspondan. Según Timoshenko y Hauser (2019), los híbridos máquina-humano han demostrado su eficacia en un amplio conjunto de aplicaciones.

Quinto, las opiniones suelen ser calificadas por otros usuarios en función de su utilidad. El estudio de Timoshenko y Hauser (2019) muestra que las frases con “me gusta” o “no me gusta” no tienen más probabilidad de contener una necesidad del cliente.

## 2.6. IDENTIFICACIÓN DE NECESIDADES DEL CLIENTE EN OPINIONES DEL PRODUCTO

La identificación de necesidades del cliente en las opiniones consiste en identificar cuáles de estas necesidades son mencionadas en cada opinión.

### 2.6.1. MENCIONES EXPLÍCITAS E IMPLÍCITAS

El proceso de identificación de necesidades en opiniones no es trivial. Si bien una opinión puede mencionar una necesidad del cliente de forma explícita, también lo puede hacer de forma implícita. Ejemplos son “el teléfono es muy grande” para referirse a la necesidad de tamaño de pantalla, o “la cámara sobresale mucho y no es cómodo” para hablar de la necesidad de diseño.

Así, la identificación de **menciones implícitas** es considerablemente más difícil que la de las **menciones explícitas**. Por ejemplo, es relativamente sencillo identificar que una opinión habla del precio cuando aparece la palabra "precio" pero se empieza a complicar

si el cliente habla del precio sin mencionarlo propiamente, “el celular es muy caro”, “lo bueno es que es barato”, etcétera.

### 2.6.2. RIESGO

En la identificación de necesidades del cliente en las opiniones existe un **riesgo** el cual no es posible pasar por alto: identificar que una opinión habla de una necesidad del cliente cuando en realidad no lo hace.

Por ejemplo, la palabra "capacidad" se utiliza en la mayoría de los casos para referirse a la necesidad de memoria en frases como “Muy buena calidad de capacidad de almacenamiento”. Sin embargo, si bien sucede en menor medida, también se utiliza “capacidad” para referirse a la necesidad de la batería en frases del tipo “Salvo por el tamaño de pantalla y capacidad de batería, el celular es malo”.

Cabe destacar que este riesgo es más frecuente en adjetivos que en sustantivos puesto que un adjetivo se usa para modificar uno o varios sustantivos. Por ejemplo, el adjetivo "rápido" se usa en la gran mayoría de los casos para referirse implícitamente a la velocidad del dispositivo. No obstante, también se usa para referirse a otros sustantivos como se ve en las frases "teléfono carga rápido", "producto llega rápido", etcétera.

Uno podría creer que solo sucede con las menciones implícitas pero esto no es así. En el mismo ejemplo visto antes, “la cámara sobresale mucho y no es cómodo” se menciona la necesidad cámara y, sin embargo, no habla de ésta sino de la necesidad de diseño.

## 2.7. ANÁLISIS DE SENTIMIENTOS

### 2.7.1. DEFINICIÓN

El análisis de sentimientos es un proceso que permite que los programas de computadora comprendan si la opinión expresada en un texto es positiva, negativa o neutral.

Los textos sobre los que se aplica suelen ser reseñas de productos. Estas revisiones son importantes para los propietarios de negocios, ya que pueden tomar decisiones comerciales de acuerdo con los resultados del análisis de las opiniones de los usuarios sobre sus productos. El análisis de sentimientos no solo se aplica en revisiones de productos, sino que también se puede aplicar en mercados de valores, artículos de noticias o debates políticos. En los debates políticos, por ejemplo, se podría averiguar las opiniones de las personas sobre ciertos candidatos electorales o partidos políticos. Los resultados de las elecciones también se pueden predecir a partir de los puestos políticos (Medhat et al, 2014).

Las fuentes de estos textos suelen ser las redes sociales como Twitter, blogs, páginas de compra-venta como Mercado Libre, Amazon, EBay, etcétera. Estas son muy útiles puesto que las personas comparten y discuten sus opiniones sobre un tema determinado libremente.

### 2.7.2. TIPOS DE ANÁLISIS DE SENTIMIENTOS

Acorde a Feldman (2013) y Medhat et al (2014) los tres tipos de análisis de sentimientos más comunes son:

- Análisis de sentimientos a nivel de documento.
- Análisis de sentimientos a nivel de oración.
- Análisis de sentimientos basado en aspectos.

Antes de ver cada uno de los tipos de análisis de sentimientos, es conveniente definir los conceptos de **documento** y **entidad**.

- Se entiende por **documento** a toda la pieza de texto. Si los textos son reseñas de productos, entonces un documento es una reseña. En cambio, si los textos son artículos de diarios, un documento es uno de los artículos del diario.
- Una **entidad** puede representar personas, hechos o temas. Así, un documento puede hablar de una o varias entidades. En general, éstas pueden ser de dos categorías:
  - o sustantivos propios que se asignan a entidades únicas. Por ejemplo, personas y lugares específicos, etcétera.
  - o sustantivos comunes

Una buena práctica consiste en que si algo es un sustantivo, califica como una "entidad". Por mencionar algunos ejemplos de entidades: Lionel Messi, Buenos Aires, Boca Juniors, celular, televisión.

Es preciso notar que una entidad tiene varios aspectos o atributos. Por ejemplo, la entidad "celular" tiene atributos como batería, pantalla, procesador, sistema operativo, entre otros.

### **ANÁLISIS DE SENTIMIENTOS A NIVEL DE DOCUMENTO**

Esta es la forma más simple de análisis de sentimientos y se supone que el documento contiene una opinión sobre un objeto principal. Esto puede ser correcto en una opinión como "El mejor gama alta por lejos, el más agradable al tacto por ser compacto." Pero fallaría en otros casos como "Lindo celular, tamaño justo. Podría tener una mejor cámara, no me gusta el lector de huellas en el botón de encendido. Precio aceptable por lo que es el producto.". En ésta última frase se habla bien de algunos aspectos como el precio y el diseño, neutral de otros aspectos como el tamaño de la pantalla, y negativamente de la cámara y el lector de huellas.

### **ANÁLISIS DE SENTIMIENTOS A NIVEL DE ORACIÓN**

Un documento a menudo expresa múltiples opiniones incluso sobre las mismas entidades. Cuando se quiere tener una visión más detallada de las diferentes opiniones expresadas en el documento sobre las entidades, es preferible utilizar el nivel de la oración, el cual tiene como objetivo clasificar el sentimiento expresado en cada oración. Se supone aquí que se conoce la identidad de la entidad discutida en la oración y, además, que hay una sola opinión en cada oración. Esta suposición se puede relajar dividiendo la oración en frases donde cada frase contiene solo una opinión.

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

Los dos enfoques anteriores funcionan bien cuando el documento completo o cada oración individual se refiere a una sola entidad. Sin embargo, en muchos casos las personas hablan de entidades que tienen muchos aspectos (atributos) y tienen una opinión diferente sobre cada uno de los aspectos. Esto sucede a menudo en reseñas sobre productos o en foros de discusión dedicados a categorías específicas de productos (como automóviles, cámaras, teléfonos inteligentes e incluso medicamentos). Como ejemplo, aquí hay una revisión de Kindle Fire tomada del sitio web de Amazon: “Como fanático de Kindle desde hace mucho tiempo, estaba ansioso por tener en mis manos un Fire. Hay algunos aspectos geniales; el dispositivo es rápido y, en su mayor parte, muy fácil de usar. La pantalla es fantástica con buen brillo y excelente color, y un ángulo de visión muy amplio. Pero también hay algunas desventajas; el pequeño tamaño del bisel hace que sea difícil sostenerlo sin pasar de página inadvertidamente, la falta de botones dificulta los controles, la memoria de almacenamiento accesible está limitada a solo 5 GB” (Feldman, 2013)

Feldman (2013) afirma que clasificar esta revisión como positiva o negativa hacia el Kindle perdería por completo la valiosa información encapsulada en ella. Quién escribió

la opinión brinda comentarios sobre muchos aspectos del Kindle (como la velocidad, la facilidad de uso, la calidad de la pantalla, el tamaño del bisel, los botones y el tamaño de la memoria de almacenamiento). Algunos de estos aspectos se revisan positivamente, mientras que otros obtienen un sentimiento negativo. Así, el autor concluye que el análisis de sentimientos basado en aspectos (también llamado análisis de sentimientos basado en características) es el problema de investigación que se centra en el reconocimiento de todas las expresiones de sentimiento dentro de un documento determinado y los aspectos a los que se refieren.

Considérese, por ejemplo, la reseña "La calidad de voz de este teléfono no es buena, pero la duración de la batería es larga". El autor de la reseña tiene una opinión respecto de cada aspecto del celular, negativo para la calidad de voz y positivo para la batería.

### 2.7.3. HERRAMIENTA PARA IMPLEMENTACIÓN

Pérez et al (2021) reflejan una limitación muy importante para analizar textos en idioma español. Los autores mencionan que un problema que inhibe a los usuarios de estas tecnologías de minería de opinión es que no hay un gran espectro de librerías para estas tareas; principalmente, hay que recurrir a las APIs de pago proporcionadas por empresas, o confiar en modelos que están muy desactualizados o incluso no están disponibles para un idioma distinto del inglés.

Una de las posibles herramientas para implementar el análisis de sentimientos en Python es la librería pysentimiento. Esta tiene soporte actual de español e inglés y es un software libre y de código abierto para cualquier persona interesada en utilizarla con fines de investigación. En adición, utiliza modelos basados en transformadores, últimamente muy utilizados en el procesamiento de lenguaje natural, tanto para tareas de clasificación como de generación, desplazando a los modelos basados en redes recurrentes (Pérez et al, 2021).



### 3. MATERIALES Y MÉTODO

Habiendo identificado las razones fundamentales por las que es importante llevar a cabo el proyecto, se procede a informar cuáles fueron los procedimientos seguidos en la investigación. Con tal motivo, se definirá el enfoque y el tipo de investigación al cual se amoldó el proyecto, la población estudiada, cuál fue la muestra y cómo se determinó, qué variables se estudiaron, qué técnicas e instrumentos se aplicaron y qué experiencias se obtuvo con ello, en resumen, qué tratamiento se le dio al material recolectado (Rojas, 2021).

#### 3.1. DISEÑO DE LA INVESTIGACIÓN

Según Cerda (2000), el concepto de diseño de investigación sirve para designar el esbozo, esquema, prototipo, modelo o estructura que indica el conjunto de decisiones, pasos, fases y actividades para realizar en el curso de una investigación. En otras palabras, es el plan para responder a cada una de las preguntas de la investigación.

##### 3.1.1. TIPO DE INVESTIGACIÓN

Si bien, es difícil encasillar un estudio dentro de alguno/s de los tipos de investigación (Rojas, 2021), podría definirse que el presente trabajo será descriptivo de observación. La razón es que se intenta describir la realidad del objeto de estudio formulando preguntas como: ¿Qué es una necesidad del cliente? ¿Qué es un atributo de un producto? ¿Cuáles son las necesidades del cliente y atributos de un producto?

Dado que se busca observar las características y los comportamientos de las variables y los fenómenos tal como se dan dentro de su contexto, para después analizarlos y sacar conclusiones, el diseño de la investigación será no experimental. A su vez, será un estudio transversal y retrospectivo puesto que se obtienen los datos de las variables

en un solo momento del tiempo y con posterioridad al acontecimiento de la exposición y la respuesta de las variables.

### 3.1.2. ENFOQUE DE LA INVESTIGACIÓN

El presente trabajo estará diseñado bajo el enfoque del tipo cuantitativo-cualitativo aunque con principal foco en el primero. Del enfoque cualitativo, se utilizará la técnica del despliegue de la función de calidad (QFD) en la que se realizará una interpretación de la relación entre las necesidades del cliente de un producto y los atributos de éste. El resto del estudio utilizará el enfoque cuantitativo dado que los datos se someterán a un tratamiento basado en la estadística y sistematización. Se implementarán técnicas de estadística descriptiva como también un análisis de sentimientos basado en aspectos.

### 3.1.3. POBLACIÓN Y MUESTRA

La población son las alternativas u opciones de un producto disponibles en Argentina en un periodo de tiempo determinado. Como subconjunto de dicha población se seleccionarán aquellas que están en el sitio web de Mercado Libre Argentina. De esta manera, se trata de una muestra del tipo no probabilístico.

En el caso de estudio del producto celular, la población son todas las alternativas disponibles de celulares en Argentina en Julio de 2022 mientras que la muestra solo aquellas que están en el sitio de Mercado Libre Argentina para la misma fecha. Como resultado, la muestra está compuesta por 401 alternativas que toman distintos valores para 60 atributos del producto con un total de 6685 opiniones. Cabe destacar que, de las 401 alternativas, 193 tienen al menos una opinión.

Es necesario tener presente que normalmente una alternativa está “repetida” varias veces en las publicaciones de Mercado Libre. Esto se explica principalmente a partir de

que una misma alternativa la pueden vender varios vendedores distintos. Por ejemplo, no hay un solo vendedor del iPhone 7 en todo Mercado Libre. Con el fin de reducir las repeticiones, se filtraron las publicaciones por condición “Nuevo” descartando todas las publicaciones de condiciones “usado” o “reacondicionado”.

#### 3.1.4. VARIABLES

Por cada producto, se utilizarán dos conjuntos de datos:

- “Alternativas”: tiene como unidad de análisis la alternativa de un producto y sus columnas se muestran en la Tabla 2. Como se puede ver, todos los productos tendrán solo dos columnas en común (id\_alternativa y precio) puesto que el resto de ellas serán los atributos propios de cada producto. En la Tabla 3, a modo de ejemplo, se puede ver el conjunto "alternativas" para los celulares.
- “Opiniones”: tiene como unidad de análisis una opinión del producto y sus columnas se pueden ver en la Tabla 4.

**Tabla 2**

*Columnas del conjunto de datos “alternativas”*

<b>Columna</b>	<b>Descripción</b>
id_alternativa	Código numérico para localizar unívocamente a una alternativa
precio	Precio de venta de alternativa
atributo <sub>1</sub> del producto	Característica técnica <sub>1</sub> de alternativa
atributo <sub>2</sub> del producto	Característica técnica <sub>2</sub> de alternativa
...	
atributo <sub>n</sub> del producto	Característica técnica <sub>n</sub> de alternativa

**Tabla 3***Columnas del conjunto de datos “alternativas” para el producto celular*

<b>Nº</b>	<b>Atributo</b>
1	id_alternativa
2	precio
3	Marca
4	Línea
5	Modelo
6	Año de lanzamiento
7	Nombre del sistema operativo
8	Versión original del sistema operativo
9	Capa original de personalización del sistema operativo
10	Resolución de la cámara trasera principal
11	Resolución de video de la cámara trasera
12	Resolución de la cámara frontal principal
13	Características principales de las cámaras
14	Cantidad de cámaras traseras
15	Resolución de las cámaras traseras
16	Cantidad de cámaras frontales
17	Resolución de video de la cámara frontal
18	Con lector de huella digital
19	Con reconocimiento facial
20	Peso
21	Modelo del procesador
22	Modelos de CPU
23	Cantidad de núcleos del procesador
24	Velocidad del procesador
25	Modelo de GPU
26	Velocidad del GPU
27	Es Dual SIM
28	Tamaños de tarjeta SIM compatibles
29	Memoria interna
30	Memoria RAM
31	Con ranura para tarjeta de memoria
32	Tipos de tarjeta de memoria
33	Tamaño de la pantalla
34	Tipo de resolución de la pantalla
35	Resolución de la pantalla
36	Tecnología de la pantalla
37	Tipo de pantalla
38	Píxeles por pulgada de la pantalla
39	Frecuencia de actualización de la pantalla
40	Brillo máximo de la pantalla
41	Con pantalla táctil
42	Capacidad de la batería
43	Tipo de batería
44	Con carga rápida
45	Con batería removible

Nº	Atributo
46	Con teclado QWERTY físico
47	Red
48	Tipo de conector de carga
49	Con conector jack 3.5 mm
50	Con Wi-Fi
51	Con GPS
52	Con Bluetooth
53	Con carga inalámbrica
54	Con NFC
55	Capacidad máxima de la tarjeta de memoria
56	Zoom digital
57	Con reconocimiento de iris
58	Es resistente a salpicaduras
59	Es resistente al polvo
60	Con flash en la cámara frontal
61	Es resistente al agua

**Tabla 4**

*Columnas del conjunto de datos “opiniones”*

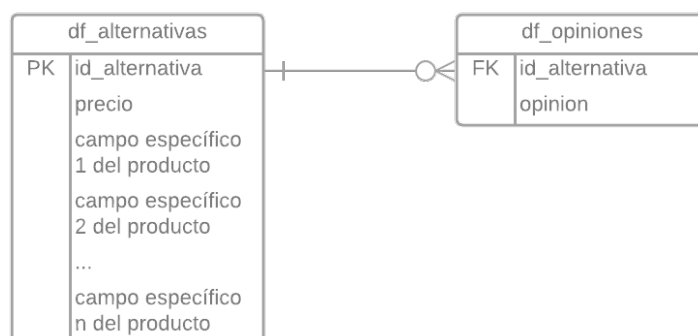
Columna	Descripción
id_alternativa	Código numérico para localizar unívocamente a una alternativa
opinion	Opinión sobre una alternativa

Como se ve en la Figura 5, la relación entre los conjuntos de datos “alternativas” y “opiniones” está dada por el campo “id\_alternativa” el cual en el primero es una clave primaria mientras que en el segundo es un clave foránea.

Una clave primaria es cualquier campo de un conjunto de datos que sirve para poder identificar una fila como única. Por otro lado, una clave foránea es la presencia en un conjunto de datos de la clave primaria de otro conjunto. Siempre que existe una clave foránea, entonces existe una relación entre conjuntos de datos. Es decir, si no estaría el campo “id\_alternativa” en el conjunto “opiniones” no se tendría una clave foránea, por lo que, no se podría establecer una relación con el conjunto de datos “alternativas”.

**Figura 5**

Relación entre conjuntos de datos “alternativas” y “opiniones”



## 3.2. TÉCNICAS DE RECOLECCIÓN DE DATOS

Las técnicas convencionales son la observación, la entrevista y la encuesta. No obstante, el presente estudio no utilizará ninguna de ellas sino que implementará la técnica conocida como *web scraping*<sup>1</sup> con el objetivo de extraer los conjuntos de datos de las Tablas 2 y 4.

### 3.2.1. CONSIDERACIONES

Es importante tener en cuenta algunas consideraciones al usar *web scraping* como técnica para extraer datos.

Primero, una alternativa del producto puede tener varias publicaciones de Mercado Libre asociadas, por lo que, si se extraen todas las publicaciones se tendrán alternativas repetidas. Esto trae innumerables problemas en el procesamiento de datos, por lo que, se evita dicha repetición a partir de extraer solo las publicaciones:

- con condición “Nuevo” evitando publicaciones con condición “usado” y “reacondicionado”

<sup>1</sup> Ver el punto 2.2 del marco teórico

- cuyo atributo “Modelo” sea nuevo. En algún producto, se podría considerar algún atributo adicional como en el caso de los celulares donde se tiene en cuenta no solo el modelo sino la memoria interna y la memoria RAM. Así, dos modelos iguales con distintas memorias son consideradas dos alternativas distintas.

Segundo, las publicaciones de Mercado Libre no están todas en una sola página sino que se distribuyen en varias de éstas. Dado que hay aproximadamente 50 publicaciones por página, los productos que tienen más de 50 publicaciones tendrán más de una página. O sea, un producto podría tener 20 páginas de 50 publicaciones cada una.

Tercero, cada publicación de Mercado Libre tiene un código de identificación. Sin embargo, dos publicaciones que venden la misma alternativa, podrían tener el mismo identificador (una de las formas de identificar alternativas repetidas).

Cuarto, hay publicaciones que no tienen opiniones, o sea, no tienen el botón “Ver todas las opiniones”.

Quinto, solo se pueden ver 5 opiniones por publicación a menos que se le indique al programa que cargue todas las opiniones haciendo *scroll down*, o sea, desplazándose hacia abajo.

Sexto, algunas publicaciones tienen las mismas opiniones, “opiniones repetidas”. Esto ocurre en algunas publicaciones que corresponden a un mismo modelo pero son:

- de distintos vendedores
- del mismo vendedor pero con distinto precio

Si no se evitase su extracción, se analizaría muchas veces los mismos datos introduciendo ruido en el análisis.

Séptimo, las páginas web cambian con el tiempo y con ello, la ubicación de los datos dentro de la página. Así, un programa que funciona en un determinado periodo de tiempo puede no hacerlo en un periodo posterior.

Octavo, el programa debe ser lo más “humano” posible para evitar ser detectado como un bot puesto que sino se interrumpe la extracción.

Noveno, se extraen datos de alternativas incluso cuando no tienen opiniones, o bien, son repetidas.

### 3.2.2. PSEUDOCÓDIGO

El pseudocódigo es una forma de expresar los distintos pasos que va a realizar un programa, de la forma más parecida a un lenguaje de programación. Su principal función es la de representar por pasos la solución a un problema o algoritmo, de la forma más detallada posible, utilizando un lenguaje cercano al de programación. El pseudocódigo no puede ejecutarse en un ordenador ya que entonces dejaría de ser pseudocódigo, como su propio nombre indica, se trata de un código falso (pseudo = falso), es un código escrito para que lo entienda el ser humano y no la máquina.



- CREACIÓN DEL NAVEGADOR WEB AUTOMÁTICO
- INGRESAR A PAGINA PRINCIPAL DE MERCADO LIBRE DEL PRODUCTO Y SELECCIÓN DE PUBLICACIONES CON CONDICIÓN="NUEVO"
- **POR PÁGINA DE LA PAGINACIÓN**
  - EXTRAER URLS DE LAS PUBLICACIONES DE LA PÁGINA Y LA URL DE LA SIGUIENTE PÁGINA
  - **POR CADA PUBLICACIÓN DE LA PÁGINA**
    - CLICK EN LA PUBLICACIÓN
    - OBTENER SU IDENTIFICADOR("id\_alternativa")
    - EXTRAER SUS VALORES PARA CADA ATRIBUTO
    - SI LA PUBLICACION CORRESPONDE A UNA ALTERNATIVA NUEVA
      - GUARDAR DATOS DE PUBLICACIÓN EN CONJUNTO DE DATOS "ALTERNATIVAS"
      - BUSCAR EL BOTÓN "VER TODAS LAS OPINIONES" DENTRO DE LA PUBLICACIÓN
      - SI EXISTE TAL BOTÓN (en ese caso, la publicación tiene opiniones)
        - CLICK EN EL BOTÓN
        - SI LAS OPINIONES SON NUEVAS
          - HAGO SCROLL DOWN PARA CARGAR TODAS LAS OPINIONES
          - EXTRAER OPINIONES Y GUARDAR EN EL CONJUNTO DE DATOS "OPINIONES"
        - SI LAS OPINIONES NO SON NUEVAS (ES DECIR, SE REPITEN)
          - NO EXTRAIGO OPINIONES
        - CLICK EN BOTÓN "VOLVER" PARA SALIR DE SECCIÓN "VER TODAS LAS OPINIONES"
      - SI NO EXISTE EL BOTÓN "VER TODAS LAS OPINIONES"
        - ENTONCES NO EXTRAIGO OPINIONES
    - CLICK EN BOTÓN "VOLVER" PARA SALIR DE LA PÁGINA DE LA PUBLICACIÓN
  - CLICK EN SIGUIENTE PÁGINA DE PAGINACIÓN
- FINALIZADA LA EXTRACCIÓN, CIERRO EL NAVEGADOR WEB AUTOMÁTICO

### 3.3. TÉCNICAS DE PROCESAMIENTO DE DATOS

Tal como se hizo desde el inicio del proyecto, se continuará usando la metodología CRISP-DM<sup>2</sup> para poder llevar a cabo un procesamiento de datos en forma sistemática y organizada. A esta altura las dos primeras fases de dicha metodología, Comprensión del negocio y Entendimiento de los datos, ya fueron llevadas a cabo y conforman la base para poder realizar el procesamiento de datos. Este último, es abordado en la tercera, cuarta y sexta fase, Preparación de los datos, Modelado y Despliegue respectivamente. La quinta fase, Evaluación, está relacionada con otras tareas como la revisión de los modelos y el proyecto.

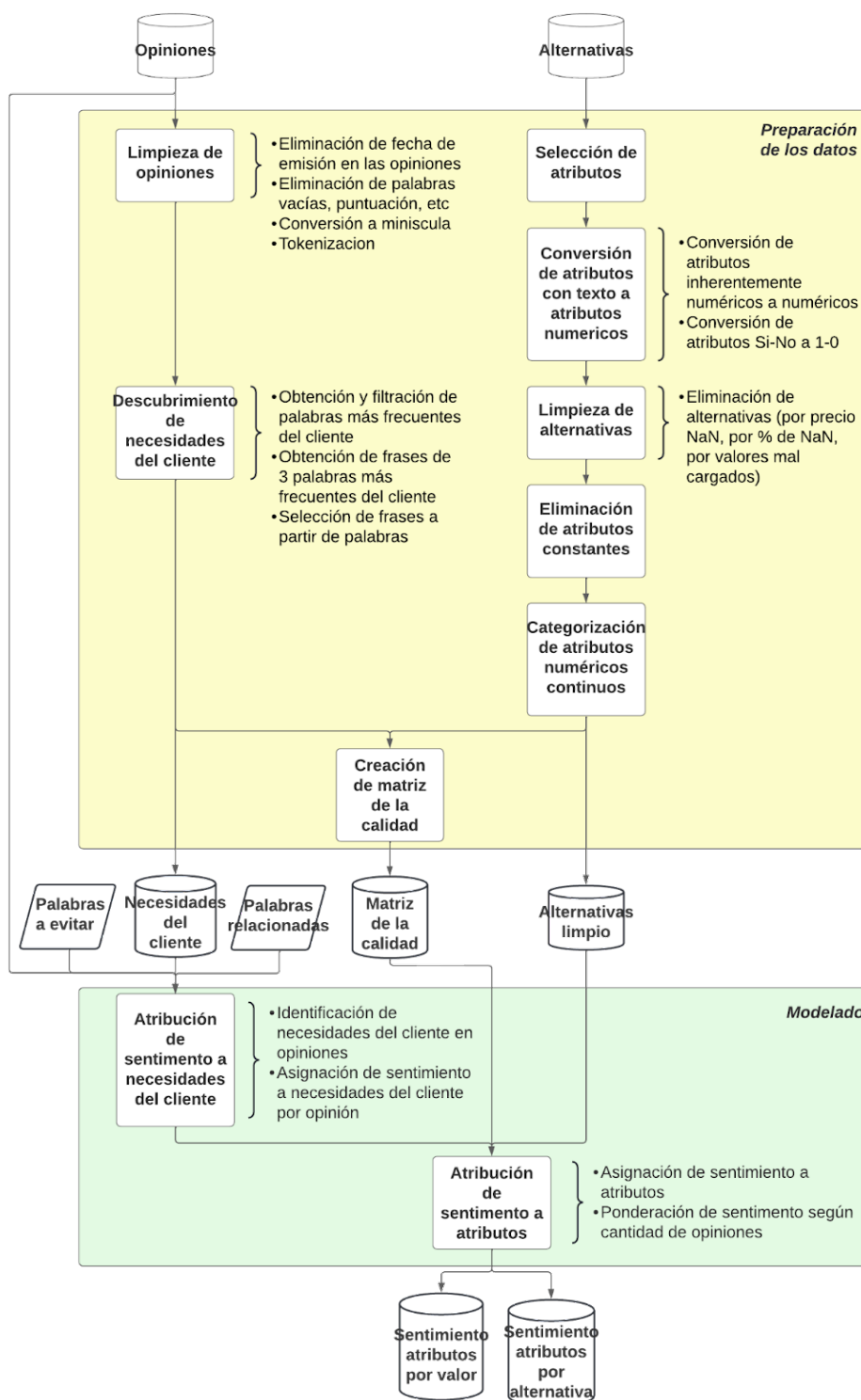
---

<sup>2</sup> Ver punto 2.1 del marco teórico

El procesamiento de datos realizado en las fases Preparación de los datos y Modelado (Figura 6) es llevado a cabo fuera de línea (*offline*) mientras que el de la fase Despliegue (Figura 7) es en línea (*online*) puesto que requiere de la interacción del consumidor.

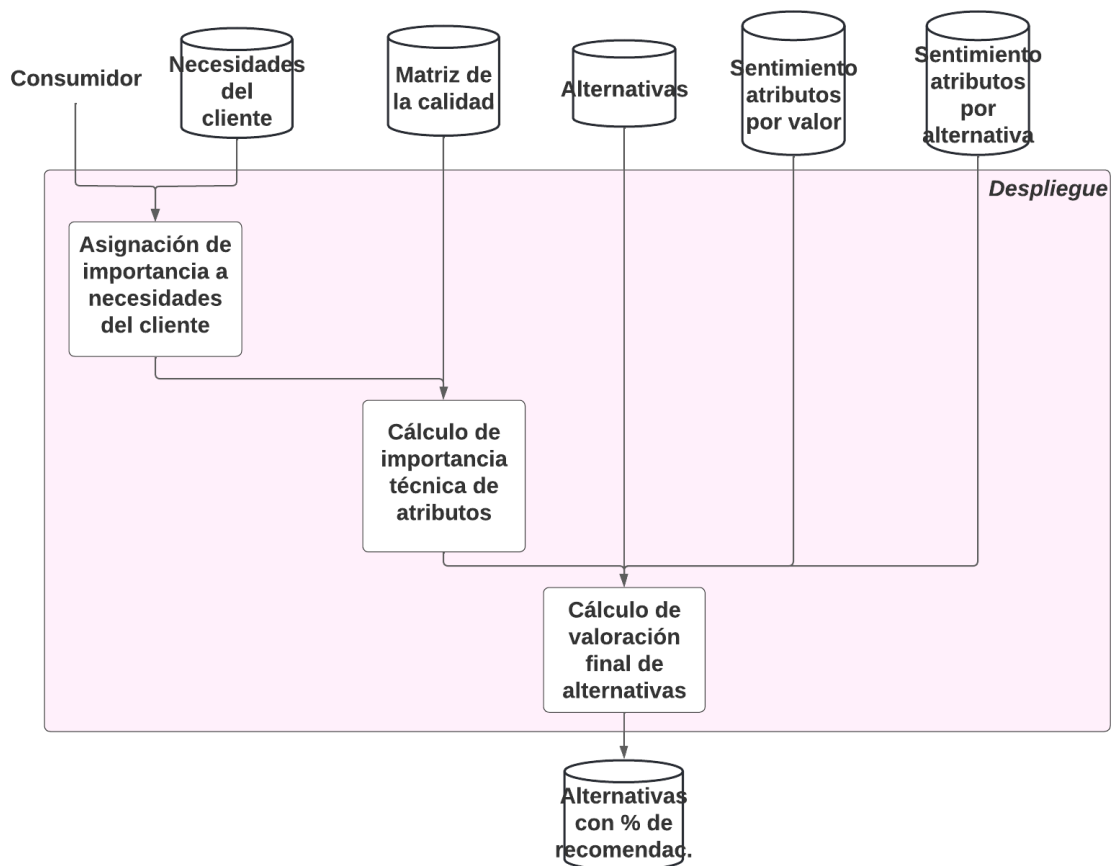
**Figura 6**

*Procedimiento en fases “Preparación de los datos” y “Modelado” realizado fuera de línea*



**Figura 7**

*Procedimiento en fase “Despliegue” realizado en línea*



A continuación se plantea el procesamiento ejecutado en cada una de las tres fases descritas.

### 3.3.1. PREPARACIÓN DE LOS DATOS

Tal como se muestra en el recuadro superior de la Figura 6, en la fase de Preparación de los datos se preparan los datos para que puedan ser analizados en fases posteriores.

Para facilitar la comprensión de lo realizado, es conveniente diferenciar 3 partes:

- 1) Preprocesamiento de opiniones
- 2) Preprocesamiento de alternativas
- 3) Creación de matriz de la calidad

### 3.3.1.1. PREPROCESAMIENTO DE OPINIONES

El objetivo de esta parte es descubrir las necesidades del cliente para un producto. Dado que éstas se descubrirán a partir de las opiniones del producto, es necesario previamente preprocesar dichas opiniones. Así, las actividades que comprende son “Limpieza de opiniones” y “Descubrimiento de necesidades del cliente”.

#### **Limpieza de opiniones**

Extraer información de cadenas de texto tal como las opiniones del producto, representa un desafío. Para poder lograrlo, es necesario previamente preprocesar el texto. Con tal objetivo se lleva a cabo:

- Eliminación de fecha de emisión de opinión. Al final de todas las reseñas del producto aparece la fecha de emisión como “Hace x meses”, “Hace x días”, “Ayer”, “Hoy”, etcétera. Por ejemplo, “Muy conforme con el producto. Hace 2 meses”.
- Conversión de texto a minúsculas.
- Eliminación de acentos.
- Eliminación de puntuación.
- Tokenización. Un texto está compuesto por un conjunto de palabras. La tokenización significa separar el texto en sus palabras, resultando una lista donde cada elemento es una palabra.
- Eliminación de palabras “vacías”, es decir, aquellas que no aportan a entender el significado del texto y dificultan el análisis. Algunos ejemplos son: artículos como “el”, “las”, “los”; pronombres como “yo”, “vos”, “mi”, “su”, “eso”, “ese”; algunas preposiciones como “a”, “en”, “de”, “sobre”; algunos verbos como “ser”, “haber”; conjunciones “y”, “o”, “pero”, “porque”.

### **Descubrimiento de necesidades del cliente**

Dado que las reseñas del producto son preprocesadas en la actividad anterior, es más sencillo ahora, analizar dichas opiniones de manera de descubrir las necesidades del cliente para tal producto. Esta actividad se compone de tres pasos secuenciales:

- 1) Obtener las frases más frecuentes utilizadas por consumidores en las opiniones
- 2) Descartar aquellas frases que son irrelevantes
- 3) De las restantes, seleccionar manualmente aquellas que son necesidades del cliente

#### ***1.- Obtener las frases más frecuentes utilizadas por consumidores en las opiniones***

De cada opinión, se extraen todos sus N-gramas (frases de largo N) y se guarda la frecuencia de cada uno. Una vez concluido el mencionado proceso, se selecciona un número determinado de éstos determinando así, las frases más frecuentes.

Como se deduce, hay dos parámetros que definir: valor de N de N-gramas y cantidad de frases frecuentes a seleccionar.

- **Valor de N de N-gramas.** Tras experimentar con valores de N de 1 hasta 4, se verificó que se maximiza el significado de las frases con N=3. Así, las frases son tri-gramas.
- **Cantidad de frases a extraer.** Considerando varios productos, prácticamente todas las necesidades del cliente se encontraron dentro de las primeras 1000 frases más frecuentes. Dado que no se pretende dejar fuera del análisis una necesidad del cliente, el parámetro se definió en 2000.

#### ***2.- Descartar frases más frecuentes que son irrelevantes***

Hay dos problemáticas que hacen que existan frases irrelevantes:

- Frases que se refieren a una misma característica del producto. Por ejemplo, considerando el producto celular, las frases "saca buenas fotos" y "tiene buenas cámaras" se usan para referirse a la característica de la cámara.
- Frases que no se refieren a ninguna necesidad del cliente. Por ejemplo, en el producto televisor podría ser "buen smart tv", o bien, en celulares frases como "teléfono gama media".

Para abordar dichas problemáticas,

- I. Se obtiene una determinada cantidad de las palabras más frecuentes utilizadas por los clientes. El parámetro de cantidad de palabras se definió en 200.
- II. De dichas palabras, se seleccionan aquellas que son:
  - A. sustantivos. En su mayoría, las necesidades del cliente son sustantivos pues son características del producto. Por ejemplo, para el producto celular, podrían ser batería, tamaño, pantalla, etcétera.
  - B. relevantes. Evitando palabras que no son necesidades del cliente como nombres de marcas, formas de referenciar al producto, etcétera.
  - C. no se refiere a la misma característica que otra de las palabras.
- III. Por palabra restante, se selecciona la frase más frecuente que la nombra.

### ***3.- Manualmente seleccionar las necesidades del cliente a partir de las frases frecuentes relevantes***

Tal y como se expuso, aún la mejor forma de terminar de seleccionar las necesidades del cliente cuando se extraen de contenido generado por usuario, sigue siendo manual, complementando así, el hombre y la máquina.

En el caso del proyecto, se revisan las diversas frases para identificar las necesidades del cliente y, en caso que sea necesario, se las reformula de manera que se comprenda aún mejor su significado.

### 3.3.1.2. PREPROCESAMIENTO DE ALTERNATIVAS

La segunda parte tiene como fin preprocesar el conjunto de datos “alternativas” para optimizar su análisis posterior. Primero, de todos los atributos del producto extraídos, se seleccionan aquellos que son relevantes. Segundo, se lleva a cabo una eliminación de todas las alternativas con algún inconveniente. Por último, se implementa una conversión de los valores de algunos atributos.

#### **Selección de atributos**

No todo atributo extraído en la recolección de datos, se usará en el análisis. El criterio de selección consiste en seleccionar aquellos atributos que son relevantes para explicar al menos una necesidad del cliente. En otras palabras, se descartarán todos aquellos atributos que no tienen relación o no ayudan a explicar las necesidades del cliente.

No obstante, sería un error pensar que los atributos descartados no son importantes. Si bien, es verdad que no son lo suficientemente relevantes como para relacionarse con alguna necesidad del cliente, podrían ser un factor clave en la evaluación de alternativas de algún consumidor.

Considere, por ejemplo, el producto celular. El atributo "Es dual sim" (posibilidad de tener dos chips en un mismo celular) no ayuda a explicar las necesidades del cliente como "duración de la batería", "relación precio calidad", "buena calidad de pantalla", etcétera. Pero, podría suceder que un consumidor necesite el celular tanto para trabajar como para uso personal, por lo que, uno de sus criterios es que el teléfono deba ser dual sim. En ese caso, es fundamental que el atributo este presente en la evaluación. Así, los atributos descartados quedan fuera del análisis pero serán guardados para ser mostrados al consumidor en la evaluación de alternativas.

### **Conversión de tipos de datos de atributos**

La idea de esta actividad consiste en transformar las columnas que contienen texto en columnas numéricas. Desde ya, esto no es posible en cualquier caso sino que solo en:

- Conversión de columnas inherentemente numéricas a numéricas
- Conversión de columnas Si-No a 1-0

Es conveniente que los valores de un atributo sean numéricos principalmente por la posibilidad de poder hacer operaciones matemáticas con ellos. Más adelante, se comprenderá la extremada importancia de esta conversión.

### ***Conversión de columnas inherentemente numéricas a numéricas***

Existen atributos cuyos valores son tratados como texto cuando en realidad son números. El motivo es simple, el número no está solo sino que está acompañado de su unidad, la cual es un texto. Por ejemplo, el atributo "Capacidad de la batería" toma valores como 6000 mAh, 4000 mAh, etcétera. Al ser la unidad "mAh" un texto, el programa entiende que los valores del atributo son textos.

El problema surge de la propia recolección de datos que al momento de extraer los valores del atributo, lo hace con las unidades puesto que así es como aparece en la página web. Evitarlo al momento de la extracción sería muy difícil dada la gran variedad de campos a extraer, por lo que, sería fácil cometer un error en la eliminación.

Por todo lo expuesto, se identificarán aquellos atributos cuyos valores sean del tipo número+unidad y se quitará ésta última. Dado que un atributo podría tener más de una unidad, es necesario previamente definir una de ellas para toda la columna y convertir todos los valores numéricos según la unidad elegida.



***Conversión de columnas Si-No a 1-0***

Un gran número de atributos toman valores “Si” o “No”. Esto sucede típicamente en atributos del tipo “Con ...” o “Es ...” como por ejemplo “Con lector de huella digital”, “Es Dual SIM”, “Con ranura para tarjeta de memoria”, “Con pantalla táctil”, “Con teclado QWERTY físico”, etcétera.

Como ya se mencionó, es conveniente convertir los atributos a numéricos, por lo que, los valores “Si” se reemplazarán por 1 mientras que los valores “No” por 0.

***Limpieza de alternativas***

Las causas que llevarán a la eliminación de una alternativa son:

- Alternativa sin precio
- Alternativa sin datos en mayoría de atributos
- Alternativa con valores mal cargados

***Alternativas sin precio***

Si bien es infrecuente, desafortunadamente, no toda publicación de Mercado Libre tiene un precio, por lo que, habrá alternativas sin éste. Dado que el precio suele ser un factor decisivo en la evaluación de alternativas, no sería conveniente mostrar dichas alternativas al consumidor, por ende, son eliminadas.

***Alternativas sin datos en mayoría de atributos***

En algunas publicaciones del producto se ofrece al consumidor muy poca información acerca de las características del producto. Por un lado, lo más probable es que el consumidor no considere dicha alternativa dentro de las posibles a elegir y, por otro, podría agregar interferencia en el análisis.

En consecuencia, se eliminan todas aquellas alternativas que no tienen datos para el 50% o más de los atributos.

***Alternativas con valores mal cargados***

Cada vendedor de Mercado Libre es el encargado de cargar los datos de su publicación. Por supuesto, que esto lo hacen siguiendo una cierta estructura brindada por la propia página web, por ejemplo, la obligación de indicar la marca del producto o el modelo, etcétera. Sin embargo, esto no quita que existan valores mal cargados por parte del vendedor.

Por ejemplo, una publicación de un celular podría indicar que el tamaño de la pantalla es de 667 pulgadas. Evidentemente, esto es un error en el que el vendedor no puso la coma entre los números seis, de manera de que el tamaño sea 6,67 pulgadas.

Estos valores mal cargados alteran varios procesos que siguen a continuación.

- Si el atributo es numérico y continuo (tal como el atributo tamaño de pantalla), alteraría la media y el desvío de éste lo que modificaría la categorización del atributo. En adición, las clases tendrán una amplitud distinta a la óptima.
- Confundiría al consumidor en la evaluación de alternativas
- No permitiría eliminar correctamente alternativas repetidas lo que introduciría ruido afectando todo el análisis.

Es preciso indicar que la detección de valores mal cargados en atributos no numéricos no fue abordada en el presente estudio dado que requeriría de un análisis manual demandante y complejo.

En cambio, en los atributos numéricos, si bien un valor atípico no necesariamente es un valor mal cargado, se pueden utilizar los métodos de detección de valores atípicos. Siguiendo el ejemplo anterior, un tamaño de pantalla de 667 pulgadas es un valor mal cargado mientras que un tamaño de 8 pulgadas podría ser identificado como un valor

atípico pero es totalmente plausible de que exista tal celular. Todo alternativa con un valor atípico identificado como:

- bien cargado, es tomada en cuenta en el análisis pero habiendo eliminado su/s valor/es atípico/s.
- mal cargado, es eliminada del análisis para evitar efectos no deseados.

En el presente trabajo, se utiliza el método del rango intercuartílico dado que el IQR es una medida robusta, es decir, no se ve afectado por valores atípicos o *outliers*. Sin embargo, éste método muestra dificultades para variables cuantitativas discretas ya que Q1 y Q3 podrían coincidir de manera que varios valores se identifican como atípicos cuando en realidad no lo son. Por esto, para dichas variables se utiliza el método de la media y desvío.

Así, el cálculo de los límites se realiza de la siguiente forma:

I.- Variables cuantitativas discretas:

- Para identificar valores atípicos que son:
  - bien cargados, se utilizan los límites dados por las Ecuaciones 1 y 2
 
$$\text{Límite inferior} = \bar{x} - 3 * s \quad (1)$$

$$\text{Límite superior} = \bar{x} + 3 * s \quad (2)$$
  - mal cargados por el vendedor, se utilizan las Ecuaciones 3 y 4
 
$$\text{Límite inferior} = \bar{x} - 6 * s \quad (3)$$

$$\text{Límite superior} = \bar{x} + 6 * s \quad (4)$$

donde  $\bar{x}$  se calcula mediante la Ecuación 5 y  $s$  mediante la Ecuación 6.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (5)$$

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (6)$$

## II.- Variables cuantitativas continuas:

- Para identificar valores atípicos que son:

- bien cargados, se utilizan los límites dados por las Ecuaciones 7 y 8

$$\text{Límite inferior} = Q1 - 1,5 * IQR \quad (7)$$

$$\text{Límite superior} = Q3 + 1,5 * IQR \quad (8)$$

- mal cargados por el vendedor, se utilizan las Ecuaciones 9 y 10

$$\text{Límite inferior} = Q1 - 5 * IQR \quad (9)$$

$$\text{Límite superior} = Q3 + 5 * IQR \quad (10)$$

Como se puede observar de las Ecuaciones 3, 4, 9 y 10, para definir que un valor *outlier* es un valor mal cargado se exige la máxima confiabilidad posible.

**Eliminación de atributos constantes**

Una variable es constante cuando toma siempre el mismo valor. Esto podría sucederle a algún atributo del producto y, en ese caso, éste no generaría diferencia entre las alternativas. En consecuencia, se trata de atributos que no son importantes en la evaluación y por ende es conveniente no considerar en el análisis.

**Categorización de atributos numéricos continuos**

Hay variables cuantitativas que toman tantos valores distintos que se las trata como continuas. Estas variables representan un problema para la herramienta dado que ésta asigna un sentimiento a cada valor de un atributo. Cuantos más valores tenga la variable, menos confiabilidad tendrá cada sentimiento puesto que hay menos opiniones para cada valor.

Se identifica si una variable cuantitativa es continua al tener más valores únicos que la cantidad dada por la Ecuación 11.

$$N^{\circ} \text{ de valores únicos} = \sqrt{N^{\circ} \text{ de observaciones}} \quad (11)$$

Si la variable es efectivamente continua, se procede a su categorización mediante alguno de los siguientes métodos:

### **1.- Creación de clases con amplitud constante**

Se define el número de clases a generar según la Ecuación 11 redondeando hacia arriba tal que sea un número entero. A continuación, se define la amplitud que debería tener cada una de las clases según la Ecuación 12.

$$\text{amplitud} = \frac{\text{rango}}{\text{Cantidad de clases a generar}} = \frac{\text{valor máximo} - \text{valor mínimo}}{\text{Cantidad de clases a generar}} \quad (12)$$

Observar que cada clase podría englobar una cantidad muy distinta de valores de la variable cuantitativa original.

### **2.- Creación de clases con amplitud variable**

Dado un percentil, la cantidad de clases a generar se calcula con la Ecuación 13 cuyo resultado es redondeado de manera tal que sea un número entero.

$$\text{Cantidad de clases a generar} = \frac{1}{\text{percentil}} \quad (13)$$

Observar que cada clase tendrá prácticamente la misma cantidad de valores de la variable cuantitativa original.

Con el objetivo de tener la mínima cantidad pero suficiente de valores para un atributo, se utiliza la creación de clases con amplitud constante a menos que la distribución de los valores sea deficiente ya sea por una gran concentración de valores en unas pocas clases, o bien, una presencia en la mayoría de éstas. En esos casos, se utiliza la

creación de clases con amplitud variable. Se define la cantidad de clases a generar como el 65% del resultado de la Ecuación 11 y en base a ello el valor del percentil.

### 3.3.1.3. CREACIÓN DE MATRIZ DE LA CALIDAD

En la tercera y última parte del procesamiento ejecutado en la fase Preparación de los datos, se establece la relación entre las necesidades del cliente (descubiertas en el punto 3.3.1.1. Preprocesamiento de opiniones) y los atributos del producto (ya habiendo descartado aquellos que son irrelevantes y aquellos que son constantes) mediante la matriz de la calidad<sup>3</sup>.

#### **Consideraciones**

Primero, es posible que una necesidad del cliente no tenga ningún atributo con el cual relacionarse y, en ese caso, se estaría descartando del análisis una característica deseable para el consumidor. Para evitar esto, se crea artificialmente un atributo, “atributo ficticio”.

Segundo, no todos los atributos del producto tienen necesidades del cliente con las cuales relacionarse. En ese caso, el atributo no ayuda a explicar ninguna de ellas y es conveniente no tenerlos en cuenta en el análisis. Con el fin de evitar preprocesar atributos que luego se descartaran, este problema ya fue abordado en la actividad de Seleccionar atributos del procesamiento de alternativas.

Tercero, los posibles valores de las relaciones entre las necesidades del cliente y los atributos se describen en la Tabla 5.

---

<sup>3</sup> Ver punto 2.4 del marco teórico

**Tabla 5**

*Posibles valores para relacionar atributos y necesidades del cliente*

<b>Relación</b>	<b>Valor numérico</b>
Nula	0
Débil	1
Media	3
Alta	9

Cuarto, el cliente podrá asignar la importancia que tiene para él cada necesidad del cliente mediante la página web utilizando una escala Likert. Dicha escala permitirá asignar algunas de las siguientes categorías: No es importante – Poco importante – Algo importante – Importante – Muy importante.

Quinto, para que cada necesidad del cliente dependa exclusivamente de la importancia que asigna el cliente a cada una, la suma de las relaciones de cada necesidad del cliente debe ser similar.

Esto es una pauta general para todo atributo que no haya sido creado artificialmente. Estos últimos, siempre tendrán el peso mínimo posible puesto que el sentimiento del atributo es mucho menos confiable que en los casos en que el sentimiento se asigna a cada valor del atributo.

### 3.3.2. MODELADO

El objetivo del análisis de los datos, los cuales ya fueron preprocesados en la fase anterior, es la asignación de sentimientos a los atributos del producto. En esta fase, el procesamiento aún es fuera de línea (*offline*), es decir, no se requiere de la interacción del consumidor (Figura 6).

Si bien el objetivo es asignar un sentimiento a los atributos, los datos que se disponen son reseñas del producto hechas por consumidores y estos no hablan un lenguaje técnico sino el suyo propio. Por ejemplo, los consumidores no dirán frases como “El procesador Snapdragon 888 es muy rápido pero la memoria RAM de 2 gb es un poco deficiente” sino “es rápido pero cuando abrís muchas apps se traba un poco”. O sea, no es posible conocer la opinión de los consumidores respecto de los atributos de forma directa.

Sin embargo, existe una forma indirecta de hacerlo. Tener presente que las necesidades de los clientes, por un lado, están en el lenguaje propio de los consumidores y, por otro, se relacionan con los atributos. Es decir, se puede conocer la opinión de los consumidores respecto a cada necesidad del cliente y luego atribuir dicha opinión a los atributos. Así, se podría saber el sentimiento de los consumidores respecto de los atributos en forma indirecta. En la frase de ejemplo “es rápido pero cuando abrís muchas apps se traba un poco”, se podría discriminar que el consumidor tiene un sentimiento neutral respecto a la velocidad y, mediante la matriz de la calidad, asignar dicho sentimiento a los atributos como “Memoria RAM”, “Velocidad del procesador”, etcétera.

Entonces el proceso es:

- I. Por opinión, identificar qué necesidades del cliente menciona y determinar si habla mal, neutral o bien de éstas.
- II. Atribuir el sentimiento de las necesidades del cliente a los atributos con que se relaciona

#### 3.3.2.1. ATRIBUCIÓN DE SENTIMIENTO A NECESIDADES DEL CLIENTE

La atribución de sentimiento a las necesidades del cliente del producto consiste de:

- a) Identificación de necesidades del cliente, y luego



b) Asignación de sentimiento a necesidades del cliente

### **Identificación de necesidades del cliente**

A pesar de haber descubierto las necesidades del cliente en las mismas opiniones sobre las que se trata de hacer la identificación de éstas, esto no es un proceso trivial. El desafío es que si bien una opinión puede mencionar una necesidad del cliente de forma explícita, también lo puede hacer de forma implícita.

Podría resultar sencillo identificar que el consumidor habla de la necesidad precio cuando dice “la relación precio-calidad es buenísima” pero no tan así cuando dice frases como “es demasiado caro”.

Para lograr la identificación de las menciones implícitas se vinculan “palabras relacionadas” a cada necesidad de manera de que se identifica que una opinión habla de una necesidad si menciona alguna de sus palabras relacionadas. En el ejemplo, consistiría en vincular "caro" a la necesidad "precio".

Un problema serio es la detección de que una opinión habla de una necesidad del cliente cuando en realidad no lo hace. Por ejemplo, suponga que una necesidad del cliente del producto celular es la calidad de la pantalla. En ese caso, sería un error identificar que “hay que comprarle un protector de pantalla urgente” habla de dicha necesidad cuando en ningún momento lo hace.

Para ello, se vinculan “palabras a evitar” a cada necesidad del cliente y sus palabras relacionadas de manera de que se identifica que una opinión menciona una necesidad del cliente solo si no menciona las palabras a evitar para dicha necesidad. En el ejemplo, "protector" estaría dentro de las palabras a evitar de la necesidad "pantalla", por lo que, si bien se menciona la necesidad no se identifica que así lo hace.

Otro punto a tener en cuenta es que un gran número de opiniones no menciona ninguna necesidad del cliente. Por ejemplo, “La verdad es muy buena y al amigo que se lo compre le resultó fantástico.”

### **Asignación de sentimiento a necesidades del cliente**

Una vez identificadas las necesidades del cliente que menciona una opinión, se procede a identificar si el consumidor lo hace negativa, neutral o positivamente, es decir, el sentimiento de éste.

Pseudocódigo:

- **POR OPINIÓN**
  - **POR FRASE DE LA OPINIÓN**
    - SI MENCIONA MÁS DE UNA NECESIDAD DEL CLIENTE Y EL SENTIMIENTO DE LA FRASE ES NEUTRAL
      - DIVIDO FRASE SEGÚN COMAS
      - **POR FRASE ENTRE COMAS**
        - SI MENCIONA AL MENOS UNA NECESIDAD DEL CLIENTE
          - NECESIDAD TOMA SENTIMIENTO DE FRASE ENTRE COMAS
    - SI MENCIONA UNA NECESIDAD DEL CLIENTE, O BIEN, MÁS DE UNA PERO EL SENTIMIENTO DE LA FRASE ES CATEGÓRICO
      - NECESIDAD TOMA SENTIMIENTO DE FRASE
    - SI NO MENCIONA NECESIDADES DEL CLIENTE
      - NECESIDADES NO TOMAN SENTIMIENTO
  - OBTENER SENTIMIENTO DE NECESIDADES DEL CLIENTE EN OPINIÓN (pues puede aparecer en más de una frase)

El primer paso consiste en dividir las opiniones en frases según los puntos (“.”) que ésta contenga.

Luego, se implementa el análisis de sentimientos basados en aspectos gracias a que aquellas frases que contienen más de una necesidad del cliente son divididas en sub frases a partir de algún elemento separador tal como las comas.

Sin embargo, se especifica que se hace tal división de las frases solo si el sentimiento de ésta es neutral dado que éste podría ser un indicador de que hay más de un sentimiento en la misma frase y, por lo tanto, es conveniente dividir.

Por el contrario, carece de sentido dividir las frases que mencionan más de una necesidad del cliente cuando el sentimiento es categórico, o sea, muy negativo o muy positivo. Por ejemplo, la frase “El celular está buenísimo, es lindo estéticamente, es rápido, tiene buena definición y otros detalles lindos destacables.” menciona varias necesidades como diseño, velocidad y pantalla y es claramente positiva, por lo que, no tiene sentido dividir dicha frase en sub frases.

Luego, todas aquellas frases que mencionan una sola necesidad del cliente, o bien, más de una pero con sentimiento categórico entonces se procede a asignar el sentimiento de la frase.

Por último, las frases que no mencionan necesidades del cliente son descartadas dado que no aportan información.

Tras revisar una opinión, se computa el sentimiento que tiene cada necesidad del cliente en ésta. Hay 2 casos:

- necesidades sin sentimiento pues no son mencionadas en la opinión
- necesidades con sentimiento pues son mencionadas en la opinión
  - o si es mencionada una vez, tendrá el sentimiento de la frase que la menciona
  - o si es mencionada más de una vez, tendrá el promedio de sentimientos de las frases que la mencionan.

Observaciones:

Primero, a pesar de identificar una necesidad del cliente en una frase, es posible que no se le asigne el sentimiento de ésta. Esto sucede en aquellos casos que la frase no contiene adjetivos y su sentimiento es neutral. Una frase que sirve como ejemplo es “Tiene refresco de pantalla de 120hz aunque los juegos y las app te las corre a máximo 60fps”.

Como se ve, son frases que mencionan una necesidad pero solo para hacer descripciones del producto.

Segundo, el modelo que lleva a cabo el análisis de sentimientos de las frases es proveniente de la librería `psentimiento` en Python. Como resultado del modelo se obtienen tres probabilidades cuya suma da 1: probabilidad de que la frase sea negativa, probabilidad de que sea neutral y probabilidad de que sea positiva. En consecuencia, el sentimiento de la frase se computa con la Ecuación 14 la cual se reduce a la Ecuación 15.

$$\text{Sentimiento} = 1 * P(\text{positiva}) + 0 * P(\text{neutral}) + (-1) * P(\text{negativa}) \quad (14)$$

Lo que se reduce a:

$$\text{Sentimiento} = P(\text{positiva}) - P(\text{negativa}) \quad (15)$$

De esta manera, el sentimiento de las frases, y por ende de las necesidades del cliente, es un valor comprendido entre -1 y 1. Valores cercanos a -1 serán frases con sentimiento negativo, valores cercanos a 0 serán neutrales y aquellos cerca del 1, positivos.

#### 3.3.2.2. ATRIBUCIÓN DE SENTIMIENTO A ATRIBUTOS

Ya se conoce el sentimiento de cada necesidad del cliente en cada opinión. Ahora resta atribuir dichos sentimientos a los atributos. Para ello será necesaria la matriz de la calidad, quien representa el vínculo entre necesidades y atributos.

La herramienta atribuye un sentimiento a los atributos de dos maneras: por valor del atributo o por alternativa. El primer caso es para los atributos que fueron extraídos de

Mercado libre mientras que el segundo, para aquellos que fueron creados artificialmente con el fin de evitar la eliminación de una necesidad del cliente. A continuación se ilustra cómo lo hace en cada caso.

### Atributo extraído de Mercado Libre

Por cada valor del atributo:

- De todas las opiniones recolectadas, se seleccionan aquellas que pertenecen a alternativas que toman tal valor en el atributo.
- Solo para aquellas necesidades del cliente que tienen relación con el atributo, se promedia su sentimiento en dichas opiniones.

Por cada necesidad del cliente se calcula su sentimiento promedio según la Ecuación 16.

$$sent_{prom_{necesidad}} = \frac{\sum_{i=1}^n sentimiento_{necesidad_i}}{cantidad\ de\ opiniones\ con\ sentimiento} \quad (16)$$

donde  $i = 1, 2, \dots, n$  es el número de opiniones

Observaciones:

- No toda opinión tiene un sentimiento para cada necesidad puesto que podría no mencionar alguna/s necesidad/es.
- Se promedia según la cantidad de opiniones con sentimiento para la necesidad y no según la cantidad de opiniones totales.
- Se asigna el sentimiento correspondiente al valor del atributo según la siguiente

Ecuación 17

$$sent_{valor} = \frac{\sum_{j=1}^n sent_{prom_{necesidad_j}} * relacion_{atrib\ y\ necesidad_j}}{\sum_{j=1}^n relacion_{atrib\ y\ necesidad_j}} \quad (17)$$

donde  $j = 1, 2, \dots, n$  es el número de necesidades del cliente relacionadas al atributo

Observaciones:

- Si una necesidad del cliente no tiene relación con el atributo, dicha necesidad no aporta al sentimiento del valor de tal atributo. Por ello directamente no se obtiene su sentimiento promedio en el paso anterior.

- Si una necesidad del cliente tiene relación con el atributo, aporta su sentimiento al valor de acorde al peso de la relación. Por ejemplo, si el atributo se relaciona fuertemente con una necesidad y débilmente con otra, el sentimiento del valor se basará en mayor medida en el de la primera necesidad.
- Si el atributo tiene relación con una sola necesidad, el sentimiento del valor de tal atributo es igual al de la necesidad.
- Cada valor del atributo tendrá un  $sent\ prom_{necesidad}$  distinto pues cambian las opiniones. El resto de los parámetros es igual.

### Atributo creado artificialmente (“atributo ficticio”)

Por alternativa del producto:

- De todas las opiniones recolectadas, se seleccionan aquellas que pertenecen a la alternativa.
- Solo para la necesidad del cliente para la cual fue creado el atributo, se promedia su sentimiento en dichas opiniones.

Para la necesidad del cliente, se calcula su sentimiento promedio según la Ecuación 16.

- Se asigna el sentimiento correspondiente al atributo según la Ecuación 18 la cual es prácticamente igual a la Ecuación 17 salvo que en lugar de obtener el sentimiento de un valor del atributo, en este caso, se trata del atributo propiamente.

$$sent_{atrib} = \frac{\sum_{j=1}^n sent\ prom_{necesidad_j} * relacion_{atrib\ y\ necesidad_j}}{\sum_{j=1}^n relacion_{atrib\ y\ necesidad_j}} \quad (18)$$

donde  $j = 1, 2, \dots, n$  es el número de necesidades del cliente relacionadas al atributo

Sin embargo, dado que todo atributo ficticio tiene relación con una sola necesidad del cliente cuyo peso de la relación es 1, la Ecuación 19 se reduce a:

$$sent_{atrib} = sent\ prom_{necesidad} \quad (19)$$

### Ejemplo 1

Considérese, por ejemplo, el atributo “Cantidad de cámaras traseras” en el producto celular. Suponga los siguientes datos:

- Atributo fue extraído de Mercado Libre, por lo que, se atribuirá un sentimiento por valor de éste.
- Atributo tiene relación solo con la necesidad del cliente “cámara” y que esta es media, es decir, de 3.
- Los valores que puede tomar el atributo son: 1, 2, 3, 4 o 5.
- La cantidad de opiniones recolectadas es 1000.

La herramienta asociaría un sentimiento a cada uno de estos de la siguiente forma:

- Para el valor *Cantidad de cámaras traseras* = 1:
  - De las 1000 opiniones recolectadas, se seleccionan aquellas que pertenecen a alternativas que toman el valor 1 en el atributo “Cantidad de cámaras traseras”, por ejemplo, 200.
  - Solo para la necesidad del cliente “cámara”, se promedia su sentimiento en las 200 opiniones aunque solo considerando aquellas que mencionan a la necesidad “cámara”.

Suponiendo que se habla de la cámara en 150 de las 200 opiniones y que se hace de manera negativa, el sentimiento promedio de la necesidad podría ser algo como -0,5.

$$sent\ prom_{necesidad} = -0,5$$

- Se asigna el sentimiento correspondiente al valor del atributo según la Ecuación 17. Dado que el atributo tiene relación con una sola necesidad, dicha ecuación se reduce a:

$$sent_{cant\ cam\ traseras=1} = \frac{sent\ prom_{cámara} * \cancel{relacion_{cant\ cam\ traseras\ y\ cámara}}}{\cancel{relacion_{cant\ cam\ traseras\ y\ cámara}}} \\ = sent\ prom_{cámara} = -0,5$$

- Para el valor *Cantidad de cámaras traseras* = 2:

- De las 1000 opiniones recolectadas, se seleccionan aquellas que pertenecen a alternativas que toman el valor 2 en el atributo “Cantidad de cámaras traseras”, por ejemplo, 50.
- Solo para la necesidad del cliente “cámara”, se promedia su sentimiento en las 50 opiniones aunque solo considerando aquellas que mencionan a la necesidad “cámara”.

Suponiendo que se habla de la cámara en 20 de las 50 opiniones y que se hace de manera positiva, el sentimiento promedio de la necesidad podría ser algo como 0,6.

$$sent\ prom_{necesidad} = 0,6$$

- Se asigna el sentimiento correspondiente al valor del atributo según la Ecuación 17. Dado que el atributo tiene relación con una sola necesidad, dicha ecuación se reduce a:

$$sent_{cant\ cam\ traseras=2} = \frac{sent\ prom_{cámara} * relacion_{cant\ cam\ traseras\ y\ cámara}}{relacion_{cant\ cam\ traseras\ y\ cámara}}$$

$$= sent\ prom_{cámara} = 0,6$$

- Y así sucesivamente para los valores 3, 4 y 5.

En el ejemplo, una conclusión plausible podría ser que los consumidores hablen mejor de la cámara cuantas más cámaras tiene. Esto se refleja en un sentimiento mayor para los valores más grandes.

## Ejemplo 2

Suponga que no hay ningún atributo de los celulares que explique la necesidad del cliente “calidad de sonido”, motivo por el cual, se decide crear el atributo “sonido”.

Considere los siguientes datos:

- Atributo tiene relación solo con la necesidad del cliente “calidad sonido” y que esta es débil, es decir, de 1.
- Se recolectaron 1000 opiniones



Así el procedimiento sería:

- Para alternativa 1 del producto:
  - De las 1000 opiniones recolectadas, se seleccionan aquellas que pertenecen a la alternativa, por ejemplo, 20.
  - Solo para la necesidad de “calidad sonido”, se promedia su sentimiento en dichas 20 opiniones, aunque solo considerando aquellas en que se menciona tal necesidad. Suponga que se habla positivamente de manera que el sentimiento sea 0,8.

$$sent\ prom_{calidad\ sonido} = 0,8$$

- Se asigna el sentimiento correspondiente al atributo según la Ecuación 19.

$$sent_{sonido} = sent\ prom_{calidad\ sonido} = 0,8$$

- Para alternativa 2 del producto:
  - De las 1000 opiniones recolectadas, se seleccionan aquellas que pertenecen a la alternativa, por ejemplo, 100.
  - Solo para la necesidad de “calidad sonido”, se promedia su sentimiento en dichas 100 opiniones, aunque solo considerando aquellas en que se menciona tal necesidad. Suponga que se habla negativamente de manera que el sentimiento sea -0,5.

$$sent\ prom_{calidad\ sonido} = -0,5$$

- Se asigna el sentimiento correspondiente al atributo según la Ecuación 19.

$$sent_{sonido} = sent\ prom_{calidad\ sonido} = -0,5$$

- Y así sucesivamente para el resto de alternativas.

**Observaciones generales**

1. Es importante notar que es mucho más robusto y confiable el sentimiento de los atributos cuando se asocia por valor en lugar de por alternativa. La razón de esto es por el simple hecho de que en el primer caso se reúnen las opiniones de todas las alternativas cuyo atributo toma tal valor, mientras que, en el segundo caso solo se reúnen las opiniones de una alternativa.
2. En los atributos "ficticios" no es posible la asignación por valor del atributo pues no tiene valores.
3. Un valor de un atributo podría no recibir un sentimiento si las alternativas que toman tal valor en el atributo no tienen opiniones asociadas, o bien, si tienen pero no mencionan a la/s necesidad/es del cliente con la/s que el atributo tiene relación.  
  
Análogamente, una alternativa podría no recibir un sentimiento para un atributo ficticio si ésta no tiene opiniones asociadas, o bien, si tiene pero no menciona a la/s necesidad/es del cliente con la/s que el atributo tiene relación.
4. Para la asignación de sentimientos de las necesidades del cliente a los atributos se usa la matriz de la calidad.

Los atributos podrían tener relación con más de una necesidad del cliente y, por ende, tener más de un sentimiento que asignar. En estos casos se hace un promedio ponderado según el peso de la relación del atributo con cada necesidad del cliente

Por ejemplo, el atributo "tamaño de pantalla" podría tener relación con las necesidades del cliente "tamaño de pantalla" y "calidad de imagen de pantalla". Una relación alta con la primera mientras que una relación débil con la segunda. Cualquier valor del atributo tamaño de pantalla, por ejemplo, 6" tomará un sentimiento basado principalmente en el sentimiento de la necesidad "tamaño de pantalla" y un poco menos en "calidad de imagen de pantalla" pues la relación es más alta con la primera.

5. En la atribución tanto por valor como por alternativa, cada sentimiento del atributo se basa en un número distinto de opiniones.

No es lo mismo el sentimiento de un valor o alternativa basado en 1000 opiniones que el sentimiento de otro valor o alternativa basado en 1 opinión. El de una sola opinión tiene bajo grado de confianza dado que depende de lo que dijo una sola persona.

Por esta razón, es necesario ponderar el sentimiento de cada valor del atributo o alternativa para atenuar el sentimiento cuando el valor o alternativa tiene muy pocas opiniones.

Cada sentimiento se ponderará siendo multiplicado por el factor resultante de la Ecuación 20.

$$\text{Factor final} = 0,1 * f_1 + 0,9 * f_2 \quad (20)$$

Donde:

$$f_1 = \frac{N^{\circ} \text{ opiniones}_{\text{val o alt}}}{N^{\circ} \text{ máximo de opiniones del atributo}} \quad (21)$$

$$f_2 = \frac{N^{\circ} \text{ opiniones}_{\text{val o alt}}}{N^{\circ} \text{ óptimo de opiniones del atributo}} \quad (22)$$

6. Los sentimientos que se basan en pocas opiniones serán muy afectados por la ponderación de manera tal que siempre valdrán 0 independientemente del

sentimiento original. Por tal motivo, todos los sentimientos de los valores de los atributos que estén basados en muy pocas opiniones son eliminados. Análogamente, los sentimientos de los atributos cuya alternativa tenga muy pocas opiniones, son descartados. El parámetro de corte se definió en el 20% del número óptimo de opiniones por valor o alternativa para el atributo.

### 3.3.3. DESPLIEGUE

Una vez ya hecho todo el procesamiento fuera de línea (*offline*) descrito en la Figura 6, se procede al procesamiento de la Figura 7 en el que el consumidor interactúa con la herramienta. Éste ingresa qué importancia tiene para él cada necesidad del cliente para poder calcular la importancia técnica de los atributos y, por último, la valoración final de cada alternativa.

#### 3.3.3.1. ASIGNACIÓN DE IMPORTANCIAS A NECESIDADES DEL CLIENTE

Al consumidor se le mostrará cada necesidad del cliente del producto utilizando una escala Likert de manera que éste pueda asignar la importancia que le da a cada una. Cada categoría de la escala Likert tendrá asociado un valor numérico (Tabla 6) para poder realizar operaciones matemáticas en el procesamiento que le sigue.

**Tabla 6**

*Categorías de escala Likert y su valor numérico*

<b>Categoría</b>	<b>Valor numérico</b>
No es importante	1
Poco importante	2
Algo importante	3
Importante	4
Muy importante	5

A modo orientativo, el consumidor podría utilizar unas importancias predeterminadas definidas para cada uso del producto.

### 3.3.3.2. CÁLCULO DE IMPORTANCIA TÉCNICA DE ATRIBUTOS

Una vez definida las importancias de las necesidades del cliente, ya se puede calcular la importancia técnica de cada atributo según la Ecuación 23.

$$Importancia\ técnica_{atributo} = \sum_{i=1}^n Importancia_{necesidad_i} * Relación_{nec_i\ y\ atrib} \quad (23)$$

donde  $i = 1, 2, \dots, n$  es cada una de las necesidades del cliente del producto.

Los atributos con mayor importancia técnica son aquellos que tendrán mayor influencia en la valoración final de cada alternativa.

### 3.3.3.3. CÁLCULO DE VALORACIÓN FINAL DE ALTERNATIVAS

Dado los sentimientos respecto de los valores de los atributos y la importancia de cada atributo, se calcula la valoración final de cada alternativa según la Ecuación 24.

$$Valoración\ final_{alt} = \sum_{i=1}^n Importancia\ técnica_{atributo_i} * Sentimiento_{atributo_i} \quad (24)$$

donde  $i = 1, 2, \dots, n$  es cada uno de los atributos del producto.

Tener en cuenta:

- 1.- El sentimiento del atributo puede ser por valor del atributo, o bien, por alternativa.
- 2.- Algunas alternativas podrían no tomar un valor para un atributo. Dado que una alternativa podría salir beneficiada ante esta situación pues no “suma” ni “resta” en la

valoración final, se estaría perjudicando a las alternativas cuyos valores tienen un sentimiento negativo. Por ejemplo, una alternativa que no tiene precio se vería favorecida respecto de todas las alternativas que tienen un precio con sentimiento negativo. Por este motivo, en esos casos, se opta por asignarles el peor sentimiento del atributo.

3- Algunas alternativas podrían tomar un valor de un atributo que no tiene sentimiento, o, si el atributo es ficticio, que la alternativa no tenga sentimiento para tal atributo. En cualquier caso, el sentimiento del atributo tiene el mismo efecto que si fuese cero, es decir, no interviene en el análisis.

4.- La valoración final de cada alternativa se traduce mediante la Ecuación 25 en un porcentaje de recomendación comparándola con la valoración final más alta de las alternativas.

$$\text{porcentaje recomendación}_{alt} = \frac{\text{valoración final}_{alt}}{\text{valoración final máxima}} \quad (25)$$

## 4. RESULTADOS

En esta sección se describe el caso de estudio del producto celular para demostrar el rendimiento de la aplicabilidad de la metodología propuesta. Se muestran los resultados en cada una de las fases que incluye un procesamiento de los datos “Preparación de los datos”, “Modelado” y “Despliegue” (misma estructura que en el inciso 3.3. Técnicas de procesamiento de datos).

Antes, es preciso indicar que se llevaron a cabo un total de 3 iteraciones de la metodología CRISP-DM en las cuales se fueron redefiniendo todas las fases incluidas la primera, la comprensión del negocio. Los resultados de cada una siempre fueron ampliamente mejores que los de la anterior.

### 4.1. PREPARACIÓN DE LOS DATOS

Aquí, es posible ver la preparación de los datos para el producto celular. De nuevo, para facilitar la comprensión de lo realizado, se diferencian 3 partes:

- 1) Preprocesamiento de opiniones
- 2) Preprocesamiento de alternativas
- 3) Creación de matriz de la calidad

#### 4.1.1. PREPROCESAMIENTO DE OPINIONES

A continuación, se muestra cómo se realizó el descubrimiento de las necesidades del cliente de los celulares tras haber limpiado las opiniones.

#### 4.1.1.1. LIMPIEZA DE OPINIONES

Para visualizar los resultados de la limpieza de opiniones se muestra una reseña del producto celular antes y después de dicho proceso.

- **Antes:** "Excelente producto. Buena calidad por un buen precio. La batería dura todo el día, ¡me encanta!. El cargador es genial, recupera la batería en una hora estando casi muerta. La cámara está buenísima también. Hace 12 días"
- **Después:** ["excelente", "producto", "buena", "calidad", "buen", "precio", "bateria", "dura", "me", "encanta", "cargador", "genial", "recupera", "bateria", "hora", "muerta", "camara", "buenisima"]

#### 4.1.1.2. DESCUBRIMIENTO DE NECESIDADES DEL CLIENTE

##### **1.- Obtener frases más frecuentes utilizadas por consumidores en las opiniones**

Se obtuvieron las 2000 frases más frecuentes utilizadas por los consumidores en las opiniones. Las primeras 10 de ellas son:

- |                              |                           |
|------------------------------|---------------------------|
| 1. relacion precio calidad   | 6. bateria dura 2         |
| 2. buena relacion precio     | 7. bateria dura muchisimo |
| 3. saca buenas fotos         | 8. ideal personas mayores |
| 4. bateria dura mucho        | 9. buena duracion bateria |
| 5. excelente relacion precio | 10. bateria dura monton   |

Observar que solo en las 10 primeras frases ya se pueden visualizar los dos problemas mencionados:

- Frases que se refieren a una misma característica como "bateria dura muchisimo", "buena duracion bateria", "bateria dura monton", "bateria dura mucho", "bateria dura 2".
- Frases que no mencionan necesidades del cliente como "ideal personas mayores"



## 2.- Descartar frases más frecuentes irrelevantes

Para poder filtrar la lista de 2000 frases más frecuentes se definieron las 200 palabras más frecuentes, de las cuales las primeras 10 son:

- |                 |                |
|-----------------|----------------|
| 1. "excelente", | 6. "bueno",    |
| 2. "bien",      | 7. "bateria",  |
| 3. "tiene",     | 8. "buena",    |
| 4. "celular",   | 9. "precio",   |
| 5. "buen",      | 10. "telefono" |

Se procedió eliminando las palabras:

- que no son sustantivos tal como verbos, adjetivos, adverbios, entre otros. Por ejemplo, "excelente", "bien", "tiene", "buen", "bueno", "buena"
- irrelevantes como "celular", "telefono".
- que se refieren a la misma característica. Por ejemplo, se removieron las palabras "fotos" y "definicion" dado que ya estaba "camara". Se eliminó "capacidad" y "almacenamiento" pues estaba la palabra "memoria".

Así, de las 200 palabras más frecuentes, quedaron las siguientes 25 palabras:

- |                    |                    |               |
|--------------------|--------------------|---------------|
| 1. "bateria",      | 10. "tamaño",      | 19. "huella", |
| 2. "precio",       | 11. "auriculares", | 20. "diseño", |
| 3. "camara",       | 12. "funciones",   | 21. "boton",  |
| 4. "pantalla",     | 13. "app",         | 22. "opcion", |
| 5. "cargador",     | 14. "apps",        | 23. "caja",   |
| 6. "memoria",      | 15. "señal",       | 24. "cable",  |
| 7. "carga",        | 16. "velocidad",   | 25. "brillo"  |
| 8. "aplicaciones", | 17. "sistema",     |               |
| 9. "sonido",       | 18. "duracion",    |               |

Finalmente, por palabra frecuente, se seleccionó la primera frase de las 2000 que la mencionaba y los resultados se muestran en la Tabla 7.

**Tabla 7**

*Frase seleccionada por palabra frecuente y su posición según frecuencia*

Palabra frecuente	Frase frecuente seleccionada	Posición en frecuencia
precio	relacion <b>precio</b> calidad	1
bateria	<b>bateria</b> dura mucho	4
camara	tiene buena <b>camara</b>	14
memoria	tiene buena <b>memoria</b>	25
sonido	tiene buen <b>sonido</b>	29
señal	tiene buena <b>señal</b>	45
carga	tiene <b>carga</b> rapida	55
duracion	tiene buena <b>duracion</b>	76
pantalla	<b>pantalla</b> tiene buena	106
aplicaciones	varias <b>aplicaciones</b> abiertas	167
cable	trae <b>cable</b> usb	178
caja	vino <b>caja</b> cerrada	191
boton	<b>boton</b> asistente google	200
cargador	viene <b>cargador</b> original	251
tamaño	tiene buen <b>tamaño</b>	257
funciones	cumple <b>funciones</b> basicas	279
auriculares	malo viene <b>auriculares</b>	319
sistema	<b>sistema</b> operativo nada	640
velocidad	celular buena <b>velocidad</b>	722
huella	desbloqueo facial <b>huella</b>	770
opcion	buena <b>opcion</b> economica	1075
brillo	grande buen <b>brillo</b>	1400
diseño	celular lindo <b>diseño</b>	1762
apps	redes sociales <b>apps</b>	1910

### 3.- Manualmente seleccionar necesidades del cliente a partir de las frases frecuentes relevantes

Tras un análisis exhaustivo, de las frases seleccionadas tras el descarte del paso 2, se seleccionaron las que efectivamente son necesidades del cliente (Tabla 8).

**Tabla 8**

*Necesidad del cliente por palabra frecuente y su posición según frecuencia*

<b>Palabra frecuente</b>	<b>Frase frecuente seleccionada</b>	<b>Posición en frecuencia</b>
precio	relacion precio calidad	1
bateria	bateria dura mucho	4
camara	tiene buena camara	14
memoria	tiene buena memoria	25
sonido	tiene buen sonido	29
pantalla	pantalla tiene buena	106
tamaño	tiene buen tamaño	257
sistema	sistema operativo nada	640
velocidad	celular buena velocidad	722
diseño	celular lindo diseño	1762

En este caso, se reformularon algunas necesidades del cliente para que se entienda mejor el significado de cada una (Tabla 9).

**Tabla 9**

*Reformulación de necesidades del cliente*

<b>Necesidad del cliente</b>	<b>Necesidad del cliente reformulada</b>
relación precio calidad	relación precio calidad
bateria dura mucho	duración de bateria
tiene buena camara	calidad de fotos y videos
tiene buena memoria	capacidad de almacenamiento
tiene buen sonido	calidad de sonido
pantalla tiene buena	calidad de imagen de pantalla
tiene buen tamaño	tamaño de pantalla
sistema operativo nada	sistema operativo
celular buena velocidad	velocidad de procesamiento
celular lindo diseño	calidad de diseño

#### 4.1.2. PREPROCESAMIENTO DE ALTERNATIVAS

En esta parte, se preprocesan las distintas alternativas dentro del producto celular para facilitar su análisis posterior.

## 4.1.2.1. SELECCIÓN DE ATRIBUTOS

De los 60 atributos de la Tabla 3, sólo 21 de ellos (el campo “id\_alternativa” no es un atributo) sirven para explicar las necesidades del cliente ya mencionadas. Estos se muestran en la Tabla 10 que sigue.

**Tabla 10**

*Conjunto de datos “alternativas” para el producto celular tras selección de atributos*

Nº	Atributo
1	id_alternativa
2	precio
3	Marca
4	Resolución de la cámara trasera principal
5	Resolución de video de la cámara trasera
6	Resolución de la cámara frontal principal
7	Cantidad de cámaras traseras
8	Resolución de video de la cámara frontal
9	Peso
10	Modelo del procesador
11	Cantidad de núcleos del procesador
12	Velocidad del procesador
13	Modelo de GPU
14	Memoria interna
15	Memoria RAM
16	Tamaño de la pantalla
17	Tipo de resolución de la pantalla
18	Tecnología de la pantalla
19	Píxeles por pulgada de la pantalla
20	Capacidad de la batería
21	Tipo de batería
22	Con carga rápida

Algunos de los atributos descartados efectivamente no servían para explicar alguna de las necesidades del cliente. Por nombrar algunos ejemplos “Línea”, “Modelo”, “Año de lanzamiento”, “Es dual SIM”, “Con lector de huella digital”, “Con wifi”, “Con GPS”, “Con bluetooth”, “Con NFC”, “Con batería removible”, “Red”.

No obstante, otro gran porcentaje de los atributos descartados en realidad si servían para explicar alguna necesidad del cliente pero

- Tenían un valor muy predominante, por lo cual, no generaban diferencia entre las alternativas lo cual es fundamental en cualquier evaluación de alternativas. Como ejemplo, “Nombre del sistema operativo”, el cual, si bien servía para explicar la necesidad del cliente relacionada al sistema operativo, en el 86% de las alternativas que toman un valor, “Android”. Otro caso es el de “Cantidad de cámaras frontales” que ayudaba a explicar la necesidad de la cámara de fotos pero fue descartado por tomar el valor de 1 en el 99% de las alternativas que tomaban un valor.
  
- Muy pocas alternativas tomaban un valor para el atributo. Esto perjudicaba mucho a las alternativas que no tenían valores y, sumado a ello, el sentimiento del atributo era muy poco confiable al tener poca información. Como ejemplo, “Capa original de personalización del sistema operativo” servía para explicar la necesidad del cliente respecto del sistema operativo pero solo en el 30% de las alternativas tomaba un valor. Lo mismo sucedió, en el caso de la necesidad de la calidad de la pantalla con los atributos “Brillo máximo de la pantalla” o “Frecuencia de actualización de la pantalla” donde solo el 32% y el 35% de las alternativas tomaba un valor respectivamente.

#### 4.1.2.2. CONVERSIÓN DE ATRIBUTOS CON TEXTO A ATRIBUTOS NUMÉRICOS

##### **Conversión de columnas inherentemente numéricas a numéricas**

De los 21 atributos de la Tabla 10, 9 de ellos son inherentemente numéricos. Estos son:

- Resolución de la cámara trasera principal
- Resolución de la cámara frontal principal
- Peso

- Velocidad del procesador
- Memoria interna
- Memoria RAM
- Tamaño de la pantalla
- Píxeles por pulgada de la pantalla
- Capacidad de la batería

Considere, por ejemplo, el atributo “Memoria RAM”. En la Tabla 11 se pueden visualizar los valores que toman las primeras 10 alternativas en el atributo de memoria RAM.

**Tabla 11**

*Valores de atributo “Memoria RAM” para las primeras 10 alternativas*

<b>Nºalternativa</b>	<b>Memoria RAM</b>
1	4 GB
2	4 GB
3	4 GB
4	2 GB
5	32 MB
6	4 GB
7	6 GB
8	4 GB
9	2 GB
10	4 GB

El atributo “Memoria RAM” no es considerado numérico por el programa al tener las unidades como “GB”, “MB”, “KB”, etcétera.

Observar que éste es el caso de un atributo con más de unidad y, como se mencionó, es necesario definir una unidad para toda la columna y, según ésta, convertir los valores numéricos. Seleccionando GB, dado que 1 GB son 1024 MB, el valor de 32 MB de la Tabla 11 correspondiente a la alternativa N°7, se convierte en 0,03 GB tal como se puede visualizar en la Tabla 12.

**Tabla 12**

Valores de atributo "Memoria RAM" para las 10 primeras alternativas luego de convertir la variable a numérica

Nºalternativa	Memoria RAM [GB]
1	4
2	4
3	4
4	2
5	0,3
6	4
7	6
8	4
9	2
10	4

#### Conversión de columnas Si-No a 1-0

En el caso de los celulares, solo 1 de los 21 atributos tiene valores "Sí" o "No". Este es el atributo "Con carga rápida". Simplemente, se reemplazó el valor "Sí" por el número 1 y el valor "No" por el número 0.

#### 4.1.2.3. LIMPIEZA DE ALTERNATIVAS

De las 401 alternativas iniciales, se eliminó el 30,7% de estas: 7,7% por no tener precio, un 20,4% por no tener datos en la mayoría de los atributos y un 2,5% por tener valores mal cargados. Así, tras la eliminación quedaron 278 alternativas (Tabla 13).

**Tabla 13**

Desglose de porcentaje de eliminación de alternativas según motivo

Nºalts totales	Nº alts sin precio	Nº alts sin datos en mayoría de atributos	Nºalts con valores mal cargados	Nºalts eliminadas	Nºalts restantes
401	31	82	10	123	278

#### 4.1.2.4. ELIMINACIÓN DE ATRIBUTOS CONSTANTES

En el caso del producto celular, a pesar de existir atributos con un valor muy predominante, no se encontraron atributos que tomen un solo valor.

#### 4.1.2.5. CATEGORIZACIÓN DE ATRIBUTOS NUMÉRICOS CONTINUOS

El precio junto con los 9 atributos a los cuales se les eliminó las unidades, son atributos numéricos continuos, por lo que, fueron categorizados o discretizados. Estos son:

- Precio
- Resolución de la cámara trasera principal
- Resolución de la cámara trasera principal
- Peso
- Velocidad del procesador
- Memoria interna
- Memoria RAM
- Tamaño de pantalla
- Píxeles por pulgada de la pantalla
- Capacidad de la batería

A continuación, se aborda un ejemplo en el que funciona bien la categorización por amplitud variable y otro en el que conviene la categorización por amplitud constante.

#### **Ejemplo 1: precio**

El atributo precio es categorizado pues tiene 224 valores únicos cuando lo recomendado, según la Ecuación 11, es como máximo 16. Según la Ecuación 12, la amplitud de las clases a generar es de 14.325. Así, se crearon 16 clases con amplitud de 14.325 como se muestra en la Tabla 14.

Como se ve en la Tabla 14, no es óptima la creación de las clases puesto que hay muchas clases que contienen al menos un valor, por ende, habrá valores que tendrán un sentimiento poco robusto. Así, es conveniente crear las clases con amplitud variable



mediante percentiles como se ve en la Tabla 15. Se crearon 10 clases pues es el 65% de 16 redondeando.

**Tabla 14**

*Clases con amplitud constante creadas para atributo "precio"*

Clase N°	Valor mínimo	Valor medio	Valor máximo	Cantidad de valores únicos
1	799	7.962	15.124	44
2	15.124	22.287	29.449	26
3	29.449	36.612	43.774	24
4	43.774	50.937	58.099	20
5	58.099	65.262	72.424	24
6	72.424	79.587	86.749	20
7	86.749	93.912	101.074	17
8	101.074	108.237	115.400	14
9	115.400	122.562	129.725	9
10	129.725	136.887	144.050	9
11	144.050	151.212	158.375	3
12	158.375	165.537	172.700	3
13	172.700	179.862	187.025	5
14	187.025	194.187	201.350	1
15	201.350	208.512	215.675	2
16	215.675	222.837	230.000	3

**Tabla 15**

*Clases con amplitud variable creadas para atributo "precio"*

Clase N°	Valor mínimo	Valor medio	Valor máximo	Cantidad de valores únicos
1	799	4.460	8.120	22
2	8.199	11.600	15.000	22
3	15.599	20.849	26.099	23
4	26.900	33.459	40.018	22
5	40.475	48.875	57.275	23
6	57.699	63.800	69.900	22
7	69.999	77.999	85.999	22
8	86.300	95.902	105.505	23
9	105.990	117.995	130.000	22
10	131.295	180.648	230.000	23

**Ejemplo 2: Memoria RAM**

Los 19 valores únicos del atributo “Memoria RAM” son: [0,002, 0,004, 0,008, 0,02, 0,03, 0,06, 0,1, 0,15, 0,25, 0,5, 0,51, 0,75, 1, 2, 3, 4, 6, 8, 12]

El atributo Memoria RAM es categorizado pues tiene 19 valores únicos cuando lo recomendado, según la Ecuación 11, es como máximo 16. El cálculo de la amplitud dado por la Ecuación 12 arroja un resultado de 0,75. En consecuencia, se procedió a crear 16 clases con amplitud de 0,75 y sus resultados se indican en la Tabla 16.

**Tabla 16**

*Clases con amplitud constante creadas para atributo “Memoria RAM”*

Clase Nº	Valor mínimo de clase	Valor medio de clase	Valor máximo de clase	Cantidad de valores únicos en clase
1	0,0	0,4	0,8	12
2	0,8	1,1	1,5	1
3	1,5	1,9	2,2	1
4	2,2	2,6	3,0	1
5	3,0	3,4	3,8	0
6	3,8	4,1	4,5	1
7	4,5	4,9	5,2	0
8	5,2	5,6	6,0	1
9	6,0	6,4	6,8	0
10	6,8	7,1	7,5	0
11	7,5	7,9	8,2	1
12	8,2	8,6	9,0	0
13	9,0	9,4	9,8	0
14	9,8	10,1	10,5	0
15	10,5	10,9	11,2	0
16	11,2	11,6	12,0	1

A diferencia del atributo precio, en este caso si funciono correctamente la creación de clases con misma amplitud. Todos los valores únicos entre 0,002 y 0,75 fueron muy convenientemente condensados en una sola clase, por lo que, habrá un solo sentimiento para todos estos valores y no 12 distintos.



**Consideraciones:**

Primero, solo se utilizaron los 21 atributos de la Tabla 10 pues son los que ayudan a explicar una necesidad del cliente.

Segundo, las necesidades del cliente "sonido", "sistema" y "diseño" no tienen atributos con los cual relacionarse. Con el fin de evitar dejarlas fuera del análisis, se crea artificialmente un atributo para cada una.

Tercero, para mantener la equidad en la importancia de cada necesidad del cliente, la suma de las relaciones de cada necesidad del cliente es similar, rondando entre 8 y 11. Esto es, por ejemplo, para la necesidad del cliente precio, la suma de los pesos de los atributos precio y marca los cuales son 9 y 1 respectivamente, o sea, 10.

Esta regla difiere en el caso de los atributos "ficticios" los cuales tienen el peso mínimo posible, en este caso, de 1.

## 4.2. MODELADO

En la actual fase, se procedió a analizar los datos. En primer lugar, se realizó la identificación y asignación de sentimientos a las necesidades del cliente. Finalmente, se atribuyen dichos sentimientos a los atributos del producto.

### 4.2.1. ATRIBUCIÓN DE SENTIMIENTO A NECESIDADES DEL CLIENTE

Como se mencionó, para poder determinar el sentimiento de las necesidades del cliente en las opiniones es necesario previamente identificar en cuales se mencionan.

## 4.2.1.1. IDENTIFICACIÓN DE LAS NECESIDADES DEL CLIENTE EN LAS OPINIONES

Como resultado de la identificación de las necesidades del cliente en las opiniones, se puede ver la cantidad de opiniones en que se identificó cada una respecto del total de opiniones.

**Tabla 18**

*Porcentaje de opiniones en que se identificó cada necesidad del cliente*

	precio	batería	memoria	cámara	sonido	pantalla	tamaño	sistema	diseño	velocidad
<b>Porcentaje de opiniones totales</b>	24%	16%	10%	23%	9%	11%	11%	12%	9%	25%

A continuación, se dispone de algunos ejemplos sobre cómo funciona la identificación tanto para las menciones implícitas como las explícitas.

**Ejemplo 1**

*“Excelente producto. Buena calidad por un buen **precio**. La **batería** dura todo el día, ¡me encanta!. El cargador es genial, recupera la **batería** en una hora estando casi muerta. La **cámara** está buenísima también.”*

La herramienta detecta sin dificultad la mención explícita de “precio”, “batería” y “cámara”.

**Ejemplo 2**

*“El celu me pareció muy bueno, muy práctico, con una **batería** que dura un montón de tiempo, es muy **fácil de usar**, saca unas **fotos** y graba vídeos espectaculares, con excelente definición y colores.”*

Por un lado, se detecta que la opinión habla de la “batería” porque la menciona explícitamente. Sin embargo, también identifica:

- “fácil de usar” como mención implícita del “sistema operativo”



La asignación es sencilla dado que el consumidor tiene un sentimiento muy bueno para todas las necesidades del cliente mencionadas.

### Ejemplo 2

*“Tuve un Xpería xa1 y buscaba uno parecido pero más **grande**, y éste fue el salto ideal, mucho mejor, de hecho. Tiene muchas **funciones** útiles, por ejemplo los atajos con sólo doble toque del botón de encendido, o con las cámaras, sobre todo por la integración con google lens. La carga es super rápida y la **batería** dura mucho aún con un uso intensivo. Lo único que hecho en falta es que no traiga auriculares y que no sea compatible con miracast, que es el que ocupó, aunque si es compatible con chromecast. Cada día más contento con esta compra.”*

### Tabla 20

*Sentimiento asociado a cada necesidad del cliente según opinión del ejemplo 2*

precio	batería	memoria	cámara	sonido	pantalla	tamaño	sistema	diseño	velocidad
	0,71					0,99	0,44		

La herramienta detecta que en todos los casos el sentimiento es positivo pero aún más cuando se habla del tamaño el cual resalta efusivamente.

### Ejemplo 3

*“Llevo como una semana con el celular y ya tuve varios problemas: **pantalla** que hace cosas raras (por momentos de congeló o se vieron cosas raras con los colores), es bastante **lento**. Es lindo, buena **memoria**, buenas **cámaras**. Pero **lento** y con algún problema, al menos el mío.”*

### Tabla 21

*Sentimiento asociado a cada necesidad del cliente según opinión del ejemplo 3*

precio	batería	memoria	cámara	sonido	pantalla	tamaño	sistema	diseño	velocidad
		0,98	0,98		-0,99				-0,58

Este es un ejemplo típico de que las reseñas pueden contener más de un sentimiento. El consumidor critica tanto la velocidad diciendo que el celular es “lento” como la pantalla pero resalta los aspectos de memoria y cámara.

#### Ejemplo 4

*“Lento. Si le sacas las animaciones funciona un poco mejor. Demasiado **caro** para lo que es. La **cámara** es buena, y la **memoria** interna alcanza. Lo voy a cambiar ni bien pueda. No lo recomiendo para nada.”*

#### Tabla 22

*Sentimiento asociado a cada necesidad del cliente según opinión del ejemplo 4*

precio	batería	memoria	cámara	sonido	pantalla	tamaño	sistema	diseño	velocidad
-0,99		0,44	0,44						-0,16

En la línea del Ejemplo 3, el consumidor tiene un sentimiento negativo hacia el precio al decir “demasiado caro”, negativo a neutral para la velocidad reflejado en la palabra “lento” y positivo tanto para la memoria como para la cámara las cuales cataloga como “buenas” y “memoria alcanza” respectivamente.

#### 4.2.2. ATRIBUCIÓN DE SENTIMIENTO A ATRIBUTOS

A continuación se mostrarán los resultados de la atribución de sentimientos de las necesidades del cliente a los atributos mediante la matriz de la calidad. Dado que la atribución puede ser por valor o por alternativa, se discriminan los resultados de cada uno.

##### 4.2.2.1. ATRIBUTOS CON SENTIMIENTO POR VALOR

En esta sección se incluirá el atributo que mejor explica cada necesidad del cliente: se utilizará el atributo “precio” para la necesidad del cliente “Relación precio calidad” (Tabla



23), “Memoria interna” para “Capacidad de almacenamiento” (Tabla 24), “Memoria RAM” para “Velocidad de procesamiento” (Tabla 25), “Capacidad de batería” para “Duración de batería” (Tabla 26), “Resolución de cámara trasera principal” para “Calidad de fotos y vídeos” (Tabla 27), “Tecnología de la pantalla” para “Calidad de imagen de pantalla” (Tabla 28), “Tamaño de pantalla” para “Tamaño de pantalla” (Tabla 29). En el Anexo se adjunta el resto de los atributos implicados en el análisis.

**Tabla 23**

*Sentimiento asociado a cada valor del atributo “precio”*

Valor [\$]	Nºopis	Sentimiento
48.738	189	0,70
33.059	323	0,60
20.799	188	0,51
62.983	161	0,46
4.460	99	0,39
95.495	114	0,34
77.240	98	0,32
11.600	63	0,20
174.950	70	0,18
117.000	45	0,08

**Tabla 24**

*Sentimiento asociado a cada valor del atributo “Memoria interna”*

Valor [GB]	Nºopis	Sentimiento
248	37	0,60
128	310	0,51
68	83	0,48
38	76	0,19
23	31	0,05
8	49	0,04

**Tabla 25**

*Sentimiento asociado a cada valor del atributo "Memoria RAM"*

Valor [GB]	Nºopis	Sentimiento
6	332	0,58
8	144	0,58
4	447	0,48
3	128	0,31
2	165	0,23
12	49	0,12
0	45	0,01
1	129	-0,12

**Tabla 26**

*Sentimiento asociado a cada valor del atributo "Capacidad de batería"*

Valor [mAh]	Nºopis	Sentimiento
4.970	405	0,46
4.497	94	0,40
4.100	125	0,32
855	50	0,14
2.798	47	0,12
3.400	66	0,10
2.000	55	0,07
1.458	21	0,05
7.465	3	
5.675	14	

**Tabla 27**

*Sentimiento asociado a cada valor del atributo "Resolución de la cámara trasera principal"*

Valor [Mpx]	Nºopis	Sentimiento
64	210	0,65
51	453	0,51
10	351	0,43
105	149	0,43
3	49	0,07
17	25	
30	19	
24	5	

**Tabla 28**Sentimiento asociado a cada valor del atributo “*Tecnología de la pantalla*”

Valor	Nºopis	Sentimiento
AMOLED	64	0,54
OLED	61	0,47
Super AMOLED	78	0,46
Dynamic AMOLED 2X	36	0,29
IPS	220	0,26
TFT	35	0,22
PLS	60	0,16
Dynamic AMOLED	23	0,10
LCD	2	
PLS LCD	0	
LTPS	9	

**Tabla 29**Sentimiento asociado a cada valor del atributo “*Tamaño de pantalla*”

Valor [pulgada]	Nºopis	Sentimiento
5,4	42	0,56
6,6	220	0,49
7,2	50	0,48
6,5	310	0,47
6,4	111	0,35
6,2	70	0,30
4,3	89	0,28
6,8	162	0,26
5,9	64	0,26
2,9	3	

#### 4.2.2.2. ATRIBUTOS CON SENTIMIENTO POR ALTERNATIVA

A continuación, se muestran las alternativas que mejor y peor se desempeñan en cada atributo “ficticio”: sonido (Tabla 30), diseño (Tabla 31) y sistema (Tabla 32).

**Tabla 30***Alternativas con mejores y peores cinco sentimientos asociados a atributo "Sonido"*

Precio	Marca	Línea	Modelo	Nºopis	Sentimiento
8.315	Panacom		MP-1112	6	0,98
240.870	Samsung	Galaxy S	S21 Ultra 5G	4	0,93
119.000	Xiaomi	Pocophone	Poco X3 GT	4	0,93
38.990	Samsung	Galaxy J	J2 Core	4	0,92
40.599	Motorola	Moto G	G31	7	0,92
...	...	...	...	...	...
7.550	Nokia	Lumia	106 (2018) Dual SIM	3	-0,59
40.018	Alcatel		5H	8	-0,62
16.999	Noblex		N405	8	-0,65
37.999	TCL		20B	4	-0,67
34.499	Motorola	Moto E	E40	5	-0,79

**Tabla 31***Alternativas con mejores y peores cinco sentimientos asociados a atributo "Diseño"*

Precio	Marca	Línea	Modelo	Nºopis	Sentimiento
131.295	Xiaomi	Mi	11T	4	0,99
60.140	LG	K Series	K62	6	0,98
99.999	Motorola	Moto G	G200 5G	5	0,95
67.786	Xiaomi	Pocophone	Poco M3 Pro 5G	4	0,92
96.902	Xiaomi	Mi	11 Lite 5G NE	11	0,87
...	...	...	...	...	...
89.298	Samsung	Galaxy	XCover Pro	6	-0,31
5.999	Energizer	Energy	E10+	4	-0,33
76.100	Samsung	Galaxy A	A31	3	-0,46
8.890	iPro	adult	F183	2	-0,51
9.499	iPro		A20	6	-0,61

**Tabla 32**

*Alternativas con mejores y peores cinco sentimientos asociados a atributo "Sistema"*

Precio	Marca	Línea	Modelo	Nºopis	Sentimiento
10.499	Panacom		MP-1106	6	0,85
32.999	Samsung	Galaxy A	A03	4	0,79
172.403	Apple	iPhone	iPhone 11	4	0,74
39.999	Samsung	Galaxy A	A13	6	0,74
25.999	TCL		TCL-L7+	8	0,73
...	...	...	...	...	...
9.166	Philco		P241	4	-0,20
40.018	Alcatel		5H	1	-0,20
26.900	P&S Mobile	Neo	NeoPro	1	-0,20
3.100	Blackview	BV Series	BV8000	2	-0,36
75.000	Blackview	BV Series	BV6300	2	-0,39

### 4.3. DESPLIEGUE

Se realiza un ejemplo para simular el procedimiento de operación cuando un consumidor interactúa con la herramienta. Como resultado, el consumidor obtiene las alternativas que más se adecuan a sus necesidades particulares.

#### EJEMPLO

Suponga que los padres de un niño/a le quieren comprar un celular a éste para jugar juegos. Desafortunadamente, los padres no tienen idea de qué alternativa elegir. Así, deciden ingresar a la página web, facilitada por la herramienta, para poder tomar una decisión rápidamente y con seguridad.

### (1) Asignación de importancias a necesidades del cliente

Como se muestra en las Figuras 8 y 9, al consumidor se le presentan las diez necesidades del cliente de celulares en la página web. Allí, deberá seleccionar la importancia de cada una en función de sus necesidades. En este caso, se utilizó las importancias predeterminadas para el uso “para jugar”.

#### Figura 8

Importancias asignadas por el consumidor a cada necesidad del cliente en ejemplo

AYUDA: Orientación de importancias según uso del producto

Para jugar

**1) RELACION PRECIO CALIDAD**

Ingrese que importancia tiene para vos "RELACION PRECIO CALIDAD":

Algo importante

No es importante      Muy importante

**2) DURACION DE LA BATERIA**

Ingrese que importancia tiene para vos "DURACION DE LA BATERIA":

Importante

No es importante      Muy importante

**3) CAPACIDAD DE ALMACENAMIENTO**

Ingrese que importancia tiene para vos "CAPACIDAD DE ALMACENAMIENTO":

Algo importante

No es importante      Muy importante

**4) CALIDAD DE FOTOS Y VIDEOS**

Ingrese que importancia tiene para vos "CALIDAD DE FOTOS Y VIDEOS":

Poco importante

No es importante      Muy importante

**5) CALIDAD SONIDO**

Ingrese que importancia tiene para vos "CALIDAD SONIDO":

Algo importante

No es importante      Muy importante

**Figura 9**

Continuación de Figura 8

**6) CALIDAD DE IMAGEN DE LA PANTALLA**  
Ingrese que importancia tiene para vos "CALIDAD DE IMAGEN DE LA PANTALLA":  
Importante  
No es importante      Muy importante

**7) TAMAÑO DE PANTALLA**  
Ingrese que importancia tiene para vos "TAMAÑO DE PANTALLA":  
Algo importante  
No es importante      Muy importante

**8) SISTEMA OPERATIVO**  
Ingrese que importancia tiene para vos "SISTEMA OPERATIVO":  
Poco importante  
No es importante      Muy importante

**9) CALIDAD DE DISEÑO**  
Ingrese que importancia tiene para vos "CALIDAD DE DISEÑO":  
No es importante  
No es importante      Muy importante

**10) VELOCIDAD DE PROCESAMIENTO**  
Ingrese que importancia tiene para vos "VELOCIDAD DE PROCESAMIENTO":  
Muy importante  
No es importante      Muy importante

**(2) Cálculo de importancia técnica de atributos**

Según los pesos seleccionados en las Figuras 8 y 9 (No es importante – Poco importante – Algo importante – Importante – Muy importante), se procede a calcular la importancia técnica de cada atributo del producto mostrados en la Tabla 33.

Considere, como ejemplo, el cálculo de la importancia técnica del atributo “Capacidad de la batería” mediante la Ecuación 23:

$$Importancia\ técnica_{cap\ bat} = \sum_{i=1}^n Importancia_{necesidad_i} * Relación_{necesidad_i\ y\ cap\ batería}$$

donde  $i = 1, 2, \dots, n$  es cada una de las necesidades del cliente del producto celular.

**Tabla 33**

*Importancia técnica de cada atributo de los celulares en ejemplo*

Atributo	Importancia técnica del atributo
Capacidad de la batería	36
Tamaño de la pantalla	31
Precio	27
Memoria interna	27
Cantidad de núcleos del procesador	15
Memoria RAM	15
Tipo de resolución de la pantalla	12
Tecnología de la pantalla	12
Píxeles por pulgada de la pantalla	12
Resolución de la cámara trasera principal	6
Resolución de la cámara frontal principal	6
Modelo del procesador	5
Velocidad del procesador	5
Modelo de GPU	5
Tipo de batería	4
Con carga rápida	4
Marca	3
Peso	3
Sonido	3
Resolución de video de la cámara trasera	2
Cantidad de cámaras traseras	2
Resolución de video de la cámara frontal	2
Sistema	2
Diseño	1

Dado que el atributo tiene relación solo con la necesidad del cliente "batería", todos los términos de la suma, salvo cuando la necesidad es "batería", son cero puesto que



independientemente del peso de la necesidad, la relación siempre es cero. Así, la ecuación anterior se reduce a:

$$\text{Importancia atributo}_{cap\ bat} = \text{Importancia necesidad}_{batería} * \text{Relación}_{batería\ y\ cap\ batería}$$

Dado que:

- la necesidad del cliente "batería" tiene un peso de "Importante", este se traduce en el valor numérico de 4 y,
- la relación entre la necesidad "batería" y el atributo "capacidad de batería" es alta, este se traduce en el valor numérico 9

Numéricamente,

$$\text{Importancia atributo}_{cap\ bat} = 4 * 9 = 36$$

### **(3) Cálculo de valoración final de alternativas**

Se toma el cálculo de la valoración final de dos alternativas al azar como ejemplo.

**Alternativa 1.** Se desglosa el cálculo de la valoración final de la alternativa en Tabla 34.

¿Qué significa cada columna de la Tabla 34?

- "Atributo": describe todos los atributos tenidos en cuenta en el análisis para todas las alternativas (Tabla 10).
- "Valor del atributo": muestra los valores que toma cada atributo en una alternativa en particular. En el caso del ejemplo, se tomó una alternativa que tiene un precio de 48.738, que pertenece a la marca Motorola y demás.
- "Sentimiento del atributo": define el sentimiento del atributo para la alternativa. Como se mostró en los resultados de la atribución de sentimiento de los atributos, el precio de 48.738 tiene un sentimiento de 0,7, la marca Motorola tiene un sentimiento de 0,58, etcétera.

**Tabla 34***Desglose de cálculo de valoración final de una alternativa en ejemplo*

Atributo	Valor del atributo	Sentimiento del atributo	Importancia técnica del atributo	Aporte a valoración final
precio	48738	0,70	27	19
Marca	Motorola	0,58	3	2
Resolución de la cámara trasera principal	51	0,51	6	3
Resolución de video de la cámara trasera	1920 px x 1080 px	0,42	2	1
Resolución de la cámara frontal principal	15	0,54	6	3
Cantidad de cámaras traseras	3	0,53	2	1
Resolución de video de la cámara frontal	1920 px x 1080 px	0,51	2	1
Peso	0,17	0,55	3	2
Modelo del procesador	MediaTek Helio G85	0,66	5	3
Cantidad de núcleos del procesador	8	0,48	15	7
Velocidad del procesador	2,1	0,51	5	3
Modelo de GPU	Mali-G52 MC2	0,56	5	3
Memoria interna	128	0,51	27	14
Memoria RAM	4	0,48	15	7
Tamaño de la pantalla	6,4	0,35	31	11
Tipo de resolución de la pantalla	Full HD+	0,47	12	6
Tecnología de la pantalla	OLED	0,47	12	6
Píxeles por pulgada de la pantalla	405	0,50	12	6
Capacidad de la batería	4970	0,46	36	17
Tipo de batería	Polímero de litio	0,41	4	2
Con carga rápida	1	0,41	4	2
sonido	-	0,03	3	0
sistema	-	0,57	2	1
diseño	-	0,39	1	0

- Importancia técnica del atributo: importancia técnica del atributo según los valores del paso (2).
- Aporte a valoración final: aporte que hace un atributo en la valoración final de la alternativa. Se calcula como el producto de las columnas "Sentimiento del atributo" y "Importancia técnica del atributo".

Finalmente, la valoración final de la alternativa es la sumatoria de los aportes de cada atributo, en este caso, 118.

**Alternativa 2.** Como se puede ver en la Tabla 35, las columnas "Atributo" e "Importancia técnica del atributo" son las mismas respecto de la primera alternativa (Tabla 34) dado que éstas son independientes de las alternativas. La actual alternativa es un celular marca Xiaomi de precio 117.000, de resolución de cámara trasera principal de 64 Mpx, etcétera.

Se repite el procedimiento de buscar el sentimiento de cada atributo según los resultados de la atribución. En este caso, el precio de 117.000 tiene asociado un sentimiento de 0,08, la marca Xiaomi un sentimiento de 0,57, etcétera.

A diferencia del caso anterior, hay atributos que no toman valor para esta alternativa como "Resolución de video de la cámara frontal" y "Velocidad del procesador". Tal como se explicó, se le asigna el peor sentimiento del atributo en cada caso. A su vez, dado que la alternativa no tiene opiniones, los atributos "sonido", "sistema" y "diseño" no tienen un sentimiento puesto que simplemente no se sabe qué opina el consumidor respecto del desempeño de la alternativa en estos aspectos.

**Tabla 35***Desglose de cálculo de valoración final de una alternativa en ejemplo*

Atributo	Valor del atributo	Sentimiento del atributo	Importancia técnica del atributo	Aporte a valoración final
precio	117000	0,08	27	2
Marca	Xiaomi	0,57	3	2
Resolución de la cámara trasera principal	64	0,65	6	4
Resolución de video de la cámara trasera	1920 px x 1080 px	0,42	2	1
Resolución de la cámara frontal principal	15	0,54	6	3
Cantidad de cámaras traseras	3	0,53	2	1
Resolución de video de la cámara frontal		0,51	2	1
Peso	0,19	0,45	3	1
Modelo del procesador	MediaTek Dimensity 1100	0,54	5	3
Cantidad de núcleos del procesador	8	0,48	15	7
Velocidad del procesador		0,01	5	0
Modelo de GPU	Mali-G77 MC9	0,64	5	3
Memoria interna	128	0,51	27	14
Memoria RAM	8	0,58	15	9
Tamaño de la pantalla	6,6	0,49	31	15
Tipo de resolución de la pantalla	Full HD+	0,47	12	6
Tecnología de la pantalla	IPS	0,26	12	3
Píxeles por pulgada de la pantalla	393	0,46	12	6
Capacidad de la batería	4970	0,46	36	17
Tipo de batería	Polímero de litio	0,41	4	2
Con carga rápida	1	0,41	4	2
sonido	-		3	0
sistema	-		2	0
diseño	-		1	0

De nuevo, se procede a calcular la valoración final de la alternativa sumando los aportes de cada atributo y se obtiene el valor de 100.

Se repite el proceso mostrado en ambos ejemplos para todas las alternativas y se muestran al consumidor las 10 alternativas con mejor valoración final tal como se ilustra en la Figura 10.

### Figura 10

*Las 10 alternativas más recomendadas al consumidor en ejemplo*

Las 10 alternativas que más te recomendamos									
	precio	Marca	Línea	Modelo	Año de lanzamiento	Nombre c	Versi	Capa origin	Resolució
1	44,999.0000	Motorola	Moto G	G41	<NA>	Android	11	<NA>	48 Mpx
2	40,599.0000	Motorola	Moto G	G31	<NA>	Android	11	<NA>	50 Mpx
3	61,499.0000	Xiaomi	Redmi	Note 11 (S	2,022.0000	Android	11	MIUI 13	50 Mpx
4	92,999.0000	Samsung	Galaxy A	A53 5G	2,022.0000	Android	12	One UI 4	64 Mpx
5	66,999.0000	Xiaomi	Redmi	Note 10	2,021.0000	Android	11	MIUI 12	48 Mpx
6	69,999.0000	Xiaomi	Redmi	Note 10S	2,021.0000	Android	11	MIUI 12	64 Mpx
7	49,999.0000	Xiaomi	Redmi	10	2,021.0000	Android	11	MIUI 12.5	50 Mpx
8	131,295.0000	Xiaomi	Mi	11T	2,021.0000	Android	11	MIUI 12.5	108 Mpx
9	54,999.0000	Motorola	Moto G	G60s	2,021.0000	Android	11	<NA>	64 Mpx
10	67,990.0000	Xiaomi	Redmi	Note 10	2,021.0000	Android	11	MIUI 12	48 Mpx

## 5. DISCUSIÓN

En un primer lugar, se lleva a cabo un análisis de los resultados del presente estudio y se comparan con los trabajos anteriormente publicados. Luego, se procede a describir las limitaciones encontradas durante el desarrollo del proyecto. Finalmente, se mencionan algunos posibles desafíos a futuro.

### 5.1. ANÁLISIS DE RESULTADOS

A la vista de los resultados, se demuestra que es posible desarrollar una herramienta capaz de evaluar las alternativas de un producto y, según las necesidades de cada cliente, recomendar la alternativa óptima a cada uno.

La metodología CRISP-DM resultó ser un marco general muy eficiente y efectivo para guiar el desarrollo del proyecto, en especial, la posibilidad de hacer varias iteraciones tal y como Wirth y Hipp (2000) recomiendan hacer. Los mismos autores hablan de que la metodología sirve para documentar la experiencia, para aumentar la rapidez y repetibilidad de los proyectos y eso fue comprobado en el presente estudio.

Los resultados del estudio muestran que es posible descubrir las necesidades del cliente a partir de contenido generado por el usuario (CGU) como las reseñas del producto y, de esa manera, evitar los métodos tradicionales de encuestas los cuales son complejos y costosos. Esto es acorde a la propuesta de Timoshenko y Hauser (2019) quienes aseguran que el contenido generado por el usuario es una fuente de necesidades de los clientes, como mínimo, comparable a la de las entrevistas vivenciales, probablemente una fuente mejor. A su vez, tal como los autores indican, se evidenció la necesidad de que el último paso del proceso del descubrimiento de necesidades del cliente deba ser llevado a cabo por un humano y no por la máquina.

Esta es una de las mayores diferencias con el estudio de Hung-Cheng y Shih-Wen (2004), los cuales proponen una herramienta de evaluación de alternativas pero descubriendo las necesidades del cliente a partir de encuestas.

Si bien el objetivo de la matriz de la calidad es que una organización descubra cómo diseñar un producto en base al gusto del cliente, fue de utilidad para poder relacionar necesidades del cliente y atributos del producto tal como afirma Timoshenko y Hauser (2019). Sin embargo, se tuvo que tener en cuenta una consideración muy importante: las necesidades de los clientes deben tener una suma de relaciones similar para que la evaluación de alternativas dependa exclusivamente de la importancia que asigna cada cliente a cada necesidad.

El tipo de análisis de sentimientos óptimo fue el basado en aspectos dado que en muchos casos las personas hablaban de varias necesidades de los clientes (aspectos) del celular (entidad) y tenían una opinión diferente sobre cada una de éstas. Esto guarda relación con los resultados de Feldman (2013) quien identifica que las reseñas de los productos suelen ser el ejemplo típico de los casos en donde es útil la aplicación del análisis de sentimientos basado en aspectos.

A pesar de que no hay una forma de evaluar los resultados del sentimiento de los atributos, estos siguen la lógica. Los “mejores” valores tienen un sentimiento positivo mientras que los “peores”, negativo.

- Mayor es el sentimiento respecto de la memoria interna, cuanto mayor es su capacidad.
- Mayor es el sentimiento respecto del tamaño de pantalla, cuanto más grande es la pantalla.
- Mayor es el sentimiento respecto de la memoria RAM, cuanto más GB tiene.

- Mayor es el sentimiento respecto de la capacidad de la batería, cuanto más capacidad tiene.
- Mayor es el sentimiento respecto de la tecnología de la pantalla, cuanto más reciente y avanzada es la tecnología.
- Mayor es el sentimiento respecto del precio, cuanto mayor es la calidad que se recibe a cambio.

Sin embargo, no hay una perfecta linealidad en los atributos, es decir, hay valores que, de acuerdo a la lógica, deberían tener un buen sentimiento y sin embargo no es así. Por mencionar algunos ejemplos, el valor de 12 GB en memoria RAM, el valor 477 ppi en Píxeles por pulgada de la pantalla, el valor de 105 Mpx en la resolución de la cámara trasera principal o el valor de 855 mAh en capacidad de la batería.

- En los dos primeros casos, se nota un bajo número de opiniones asociadas al valor probablemente porque al ser un valor extremo, pocas alternativas toman ese valor.

Si bien la ponderación es fundamental para evitar que valores con muy pocas opiniones alteren los resultados, también afecta a los valores como los señalados. Al tener pocas opiniones, es poco confiable su sentimiento y la herramienta lo atenúa considerablemente.

- El caso de la capacidad de la batería de 855 mAh podría explicarse en que el precio promedio de las alternativas con tal batería es de \$6500, el cual podría ser tan bajo que el consumidor espera menos calidad y se sorprende ante un rendimiento, en este caso de la batería, por encima de sus expectativas.
- En el caso de la resolución de la cámara trasera principal no se pudo identificar una posible explicación. Es evidente que hay muchos factores que podrían tener



influencia como identificar que la opinión habla de la cámara cuando no lo hace, determinar mal el sentimiento de la necesidad cámara, etcétera.

La comparación de alternativas tiene en cuenta toda la información relevante en la decisión y se escogen las alternativas óptimas para cada cliente. El consumidor asigna una importancia para cada necesidad del cliente, las cuales se traducen en una importancia técnica para cada atributo. Luego, se obtiene una valoración final para cada alternativa sumando para todos los atributos, el producto entre su importancia y su sentimiento. Esta es una de las formas de evaluación que identifican Payne et al (1991). Estos aseguran que a veces las personas toman decisiones de acuerdo con este procedimiento normativo, aunque lo más frecuente es que las personas tomen decisiones utilizando procesos de decisión más sencillos los cuales tienen en cuenta menos información. Sin embargo, en el caso de la herramienta el consumidor utiliza esta forma de evaluación sin esfuerzo gracias al procesamiento computacional.

La implementación de una página web es una forma muy útil para que la herramienta esté disponible para cualquier consumidor puesto que lo único que se requiere es algún dispositivo electrónico con acceso a internet. Esto concuerda con el despliegue realizado por Hung-Cheng y Shih-Wen (2004).

## 5.2. LIMITACIONES

Algunas de las limitaciones enfrentadas durante el desarrollo del proyecto fueron:

- Tiempos de extracción y procesamiento de datos. Ambos tiempos son relativamente largos, por lo que, se demoró mucho en probar e implementar nuevos cambios con el fin de mejorar la herramienta. La extracción tarda

aproximadamente entre 4 y 6 horas y algunas etapas de procesamiento cerca de 1 hora.

- Calidad de datos.
  - Varias publicaciones de Mercado libre se refieren a una misma alternativa. En otras palabras, una alternativa es vendida por varios vendedores o incluso, por el mismo vendedor a diferente precio.
  - Muchas publicaciones similares (ya sea que se refieren a la misma alternativa o con un leve cambio) tienen las mismas opiniones.
  - Como la carga de datos de una publicación depende en gran medida de cada vendedor, un gran porcentaje de los valores de los atributos están mal cargados, o bien, varias publicaciones se refieren a un mismo valor de forma diferente.

Por tales motivos, la etapa de preparación de los datos fue la que demandó más atención y tiempo.

- Inexistencia de librerías que implementan correctamente el procesamiento de texto en idioma español. En consecuencia, se debió recurrir al desarrollo de funciones propias para abordar dicha tarea.
- Inexistencia de librerías gratuitas para el análisis de sentimientos basado en aspectos en idioma español. Por tal motivo, se debió implementar una librería de análisis de sentimientos a nivel documento y mediante un modelo propio identificar los aspectos de la entidad y su sentimiento.

### 5.3. FUTURAS INVESTIGACIONES

Algunos desafíos a futuro son:

- Aumentar el número de alternativas y opiniones a partir de la extracción de datos de varias páginas como Amazon, EBay, Twitter y otras.
- Extraer datos de la gran mayoría de los productos del mercado.
- Automatizar el descubrimiento de las necesidades del cliente con el fin de dejar de requerir una revisión manual.
- Desarrollar un modelo de análisis de sentimientos basado en aspectos en idioma español a partir de redes neuronales para mejorar la eficacia de la herramienta.

## 6. CONCLUSIONES

Si bien la evaluación de alternativas podría llegar a ser omitida en casos particulares como una compra rutinaria, ésta es una tarea prácticamente omnipresente, es decir, se realiza en la gran mayoría de los casos independientemente del producto, consumidor o modelo de proceso de compra.

La dificultad de la evaluación de alternativas crece conforme aumenta el número de alternativas y su número de atributos, menor es la familiaridad del producto (funciones, usos, terminologías, etcétera) y mayor es la importancia de la elección según lo que cuesta ya sea en términos de tiempo o dinero. Dado la creciente globalización de los mercados y su constante cambio por las presiones competitivas, cada vez la elección es más difícil. Los consumidores tienen que escoger entre cientos de opciones muy similares y, en muchas ocasiones, no conoce cuales son ni sus características.

En consecuencia, el consumidor se ve obligado a invertir muchos recursos, principalmente tiempo, y a simplificar dicha evaluación a punto tal que, a pesar de todo el esfuerzo, es muy posible dejar fuera la alternativa óptima.

Así, una herramienta que automatice el proceso de evaluación y brinde aquella alternativa óptima para cada consumidor en cuestión de segundos es totalmente imprescindible. Si bien existieron algunos intentos de crearla, no se encontró ninguna que esté actualmente disponible al público. Por ello es el gran impacto que la herramienta desarrollada en el presente estudio puede tener. A pesar de la gran cantidad de limitaciones encontradas en el transcurso del proyecto, los resultados fueron muy satisfactorios y, sumado a que hay un gran margen para escalar el proyecto, la herramienta tiene mucho potencial.

De nada servirá si el consumidor no tiene acceso para poder utilizar la herramienta, por lo que, se desarrolló una página web<sup>4</sup> la cual trata de hacer posible su uso a pesar de que quien lo haga no tenga conocimiento ya sea del producto, o bien, de tecnología.

Finalmente, hay un sinfín de implicaciones prácticas proyectadas más allá del gran ahorro de tiempo.

1. Dado que las alternativas con poca información tienen poca probabilidad de ser recomendadas, esto podría fomentar al aporte de más y mejor información por parte de los vendedores.
2. La voz del cliente cada vez tomaría mayor relevancia. Por un lado, la evaluación se basa totalmente en la opinión de estos, por lo que, las empresas se preocuparán más por la satisfacción del cliente. Por otro, se podría lograr que los vendedores empiecen a brindar información de atributos importantes para la evaluación y que hoy no son ofrecidos.
3. Dado lo exhaustivo que es la herramienta, los consumidores posiblemente comprarán una alternativa que ni siquiera hubiesen considerado. Esto haría que se eleve la competencia lo cual favorece al consumidor.
4. En un futuro, las páginas que ofrecen productos al público podrían tener vinculada la herramienta. Esto no solo ayudaría al consumidor para que el proceso de compra sea rápido y fácil sino también a los vendedores puesto que posiblemente cuanto mayor es la facilidad para consumir, mayor serán las compras.
5. Tal vez, con más potencia para procesar los datos y logrando automatizar algunos procesos que aún son manuales, el cliente podría decir cualquier producto a relevar y la herramienta llevarlo a cabo en el momento.

---

<sup>4</sup> Link a la página: <https://nachomondino1-evaluacion-compra-automatizada-my-appinicio-leworg.streamlitapp.com/>

Por todo lo mencionado, a solo un *click* de distancia, los consumidores disponen de una herramienta que podría favorecerlos desde muchísimas perspectivas.

## 7. REFERENCIAS BIBLIOGRÁFICAS

Azevedo, A., & Santos, M. F. (2008). *KDD, SEMMA and CRISP-DM: a parallel overview.*

IADS-DM.

Butler, P., & Peppard, J. (1998). *Consumer purchasing on the Internet*, *European Management Journal*, 16(5), 600–610. doi:10.1016/s0263-2373(98)00036-x

Cerda, Hugo (2000). *Los elementos de la investigación. Cómo reconocerlos, diseñarlos y construirlos.* Bogotá: Editorial El Búho

Feldman, R. (2013). *Techniques and applications for sentiment analysis.* *Communications of the ACM*, 56(4), 82. doi:10.1145/2436256.2436274

Hauser, J. & Clausing, D. (1988). *The house of quality.* *Harvard Bus. Rev.* 66(3):63–73

Hauser, John & Griffin, Abbie & Klein, Robert & Katz, Gerry & Gaskin, Steven. (2010). *Quality Function Deployment (QFD).* 10.1002/9781444316568.wiem05023.

Hernández Sampieri, R., Fernández, C. y Baptista, P. (2010). *Metodología de la investigación* (5° Ed.). México, D.F., México: McGraw Hill Interamericana.

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). *DMME: Data mining methodology for engineering applications—a holistic extension to the CRISP-DM model.* *Procedia Cirp*, 79, 403-408.

- 
- Hung-Cheng Tsai, Shih-Wen Hsiao (2004), *Evaluation of alternatives for product customization using fuzzy logic*, Information Sciences, Volume 158, Pages 233-262, ISSN 0020-0255, <https://doi.org/10.1016/j.ins.2003.08.001>.
- Juan Manuel Pérez, Juan Carlos Giudici & Franco Luque (2021), *psentimiento: A Python Toolkit for Sentiment Analysis and SocialNLP tasks*.
- Kotler, P. (2017). *Principles of Marketing* [Fundamentos de marketing]. Pearson Education. Seventh European Edition.
- Madhavan M., Chandrasekar, K., (2015), *Consumer Buying Behavior – An overview of theory and models*, St. Theresa Journal of Humanities and Social Sciences
- Malhotra, N. K. (1982). *Information load and consumer decision making*. Journal of consumer research, 8(4), 419-430.
- Medhat, W., Hassan, A., & Korashy, H. (2014). *Sentiment analysis algorithms and applications: A survey*. Ain Shams engineering journal, 5(4), 1093-1113.
- Mitchell, R. (2018). *Web Scraping with Python*, Second Edition. 978-1-491-998557-1
- Moro, Sérgio & Cortez, Paulo & Laureano, Raul. (2011). *Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology*. Proceedings of the European Simulation and Modelling Conference.
- Payne, J., Bettman, J. R., & Johnson, E. J. (1991). *Consumer decision making* [Toma de decisiones del consumidor]. Handbook of consumer behaviour, 50-84.



- 
- Punj, G. N., & Stewart, D. W. (1983). *An interaction framework of consumer decision making*. Journal of Consumer Research, 10(2), 181-196.
- Prasad, R. K., & Jha, M. K. (2014). *Consumer buying decisions models: A descriptive study*. International journal of innovation and applied studies, 6(3), 335.
- Rojas, V. M. N. (2021). *Metodología de la Investigación: diseño, ejecución e informe*. Ediciones de la U.
- Timoshenko, A., & Hauser, J. R. (2019). *Identifying customer needs from user-generated content*. Marketing Science, 38(1), 1-20.
- Wirth, R. & Hipp, Jochen. (2000). *CRISP-DM: Towards a standard process model for data mining*. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining.
- Yacuzzi, Enrique; Martín, Fernando (2003), *QFD: conceptos, aplicaciones y nuevos desarrollos*, Serie Documentos de Trabajo, No. 234, Universidad del Centro de Estudios Macroeconómicos de Argentina (UCEMA), Buenos Aires

## 8. ANEXO

### Resultados de Modelado

**Tabla 36**

*Sentimiento asociado a cada valor del atributo "Marca"*

<b>Valor</b>	<b>Nºopis</b>	<b>Sentimiento</b>
TCL	138	0,66
Motorola	180	0,58
Alcatel	84	0,58
Xiaomi	244	0,57
ZTE	55	0,55
LG	36	0,52
BLU	32	0,49
Quantum	40	0,47
Kodak	22	0,36
Samsung	316	0,36
Nokia	30	0,36
iPro	27	0,34
Apple	40	0,34
Sansei	14	0,20
Panacom	17	0,17
Blackview	12	0,12
Noblex	15	0,06
Energizer	9	0,05
Kanji	15	0,05
Excess	4	
P&S Mobile	3	
Konka	7	
CAT	5	
Roadstar	0	
Philco	2	
Plum	0	
Sky Devices	6	
Cyrus	1	
celular adultos	0	
Ulefone	3	
RugGear	2	
Umidigi	0	
Realme	0	
Nubia	0	
BLU,IPRO, OTRAS MARCAS	4	

Valor	Nºopis	Sentimiento
Cubot	0	
Logic	0	
BLU - IPRO - NOKIA - GENERICO	2	
Doogee	1	
IPRO - BLU - EXC SEGUN STOCK	6	
BlackBerry	3	
Google	0	
Pcbox	7	
Sony	1	
Asus	0	
Infinix	0	

*Nota:* Atributo relacionado con necesidad del cliente “Relación precio calidad”

### Tabla 37

*Sentimiento asociado a cada valor del atributo “Resolución de video de la cámara trasera”*

Valor	Nºopis	Sentimiento
3840 px x 2160 px	433	0,65
1920 px x 1080 px	631	0,42
7680 px x 4320 px	97	0,29
1280 px x 720 px	16	
320 px x 240 px	3	
2048 px x 1152 px	12	
800 px x 480 px	1	
176 px x 144 px	0	
640 px x 480 px	0	
2160 px x 3840 px	5	
1080 px x 1920 px	7	

*Nota:* Atributo relacionado con necesidad del cliente “Calidad de fotos y videos”

**Tabla 38**

*Sentimiento asociado a cada valor del atributo “Resolución de la cámara frontal principal”*

Valor [Mpx]	Nºopis	Sentimiento
30	204	0,58
15	321	0,54
9	298	0,52
25	37	0,51
20	78	0,34
6	214	0,31
12	30	0,16
1	21	
3	1	
0	1	

*Nota:* Atributo relacionado con necesidad del cliente “Calidad de fotos y videos”

**Tabla 39**

*Sentimiento asociado a cada valor del atributo “Cantidad de cámaras traseras”*

Valor	Nºopis	Sentimiento
4	460	0,53
3	423	0,53
2	180	0,37
1	164	0,17
5	2	

*Nota:* Atributo relacionado con necesidad del cliente “Calidad de fotos y videos”

**Tabla 40**

*Sentimiento asociado a cada valor del atributo “Resolución de video de la cámara frontal”*

Valor	Nºopis	Sentimiento
3840 px x 2160 px	226	0,54
1920 px x 1080 px	762	0,51
1280 px x 720 px	28	
1080p a 30fps	0	
4K a 30/60 fps	0	
720 px x 1280 px	2	

*Nota:* Atributo relacionado con necesidad del cliente “Calidad de fotos y videos”

**Tabla 41***Sentimiento asociado a cada valor del atributo "Peso"*

Valor [kg]	Nºopis	Sentimiento
0,20	61	0,60
0,16	55	0,58
0,17	50	0,55
0,14	69	0,54
0,18	61	0,53
0,19	70	0,45
0,21	35	0,27
0,21	33	0,06
0,24	28	0,05

*Nota: Atributo relacionado con necesidad del cliente "Tamaño de pantalla"***Tabla 42***Sentimiento asociado a cada valor del atributo "Modelo del procesador"*

Valor	Nºopis	Sentimiento
MediaTek Helio P22	17	0,84
MediaTek Dimensity 1200-Ultra	17	0,80
Apple A13 Bionic	22	0,76
Snapdragon 865	17	0,75
Exynos 1280	37	0,74
Snapdragon 888+ 5G	19	0,71
Snapdragon 460	16	0,69
Snapdragon 778G 5G	21	0,68
MediaTek Helio G85	28	0,66
MediaTek MT6762 Helio P22	30	0,65
Snapdragon 695 5G	25	0,65
MediaTek MT6382 Helio G95	25	0,63
Snapdragon 778G	23	0,61
Snapdragon 662	18	0,60
Qualcomm SM8450 Snapdragon 8 Gen 1 (4 nm)	25	0,60
MediaTek Dimensity 720	18	0,58
MediaTek Helio G80	42	0,58
Snapdragon 678	12	0,57
Snapdragon 870	22	0,56
Apple A11 Bionic	10	0,55
MediaTek Dimensity 1100	12	0,54
MediaTek Helio G88	11	0,53
MediaTek Helio G95	14	0,52

<b>Valor</b>	<b>Nºopis</b>	<b>Sentimiento</b>
Snapdragon 765G	18	0,52
Exynos 9609	16	0,51
Spreadtrum SC9832E	14	0,48
Exynos 2100	13	0,47
MediaTek MTK6762D	67	0,46
Unisoc T700	16	0,46
Snapdragon 732G	15	0,45
Snapdragon 860	20	0,44
Snapdragon 720G	31	0,44
Apple A12 Bionic	7	0,43
MediaTek MT6765 Helio P35	49	0,43
MediaTek Dimensity 800U	10	0,42
Apple A9	7	0,41
Snapdragon 695	9	0,38
Snapdragon 680	14	0,38
Snapdragon 480 Plus	31	0,36
Apple A15 Bionic	6	0,35
Snapdragon 780G	18	0,35
Exynos 850	52	0,34
Unisoc SC9863A	13	0,33
Snapdragon 888 5G	6	0,31
Unisoc T606	16	0,30
MediaTek Dimensity 700	5	0,29
MediaTek Helio G96	5	0,26
MediaTek Helio G35	17	0,26
MediaTek MT6833 Dimensity 700	12	0,24
Qualcomm Snapdragon 750G 5G (8 nm)	9	0,23
Snapdragon 888	9	0,22
MediaTek MT6768 Helio P65	12	0,21
Apple A14 Bionic	4	0,17
Exynos 7570	11	0,15
Qualcomm QM215	57	0,14
Exynos 9611	4	0,12
MediaTek MT6261D	6	0,09
MediaTek MT6853 Dimensity 700	8	0,05
MediaTek MT6762V Helio A25	8	0,02
MediaTek MT6762G Helio G25	23	0,00
MediaTek MT6572	6	-0,01
MediaTek MT6779 Helio P90	3	-0,02
MediaTek Helio G25	9	-0,15
Spreadtrum SC6531E	6	-0,18
MediaTek MT6739	25	-0,31
Mediatek MT6765V/CB Helio G37	17	-0,34

Valor	Nºopis	Sentimiento
MTK6739 Quad Core	2	
Unisonic SC9863	2	
Spreadtrum SC6531F	2	
MediaTek MT6261M	2	
MediaTek MT6771T Helio P70	2	
Spreadtrum SC6531DA	0	
MediaTek MT6737M	0	
SPRD 6531E	2	
Exynos 850 (8nm)	0	
MediaTek MT6785 Helio G95	2	
Unisoc SC9863	0	
MediaTek MT6771V/CT Helio P70	0	
SC6531E	0	
MediaTek Helio P60	1	
Snapdragon 660	1	
MediaTek MT6762V Helio P22	2	
MediaTek MT6580	1	
SC9832E	1	
MediaTek MT6771 Helio P60	0	
Snapdragon 835	0	
MediaTek MT6762V/WD Helio A25	2	
Unisoc SC7731E	0	
Helio G95	0	
Spreadtrum SC9863A	0	
MediaTek Dimensity 810	2	
QUALCOMM SNAPDRAGON 695	0	
MediaTek MT6762V/WD	2	
Google Tensor	0	
MediaTek MT6572M	0	
Unisoc T618	0	
Snapdragon 730G	0	
SM8450	0	
MediaTek Dimensity 800	0	
Dimensity 700	0	
Exynos 2100 Octa-Core	0	
Snapdragon 750G	1	
MediaTek Dimensity 920 5G	0	
Snapdragon 630	0	
Snapdragon 855	0	
Snapdragon 210	0	
MediaTek MT6785T Helio G90T	0	
Unisoc T610	0	
Helio G25	0	

Nota: Atributo relacionado con necesidad del cliente “Velocidad de procesamiento”

**Tabla 43**

*Sentimiento asociado a cada valor del atributo "Cantidad de núcleos del procesador"*

Valor	Nºopis	Sentimiento
8	1132	0,48
4	156	-0,02
6	49	

*Nota: Atributo relacionado con necesidad del cliente "Velocidad de procesamiento"*

**Tabla 44**

*Sentimiento asociado a cada valor del atributo "Velocidad del procesador"*

Valor [GHz]	Nºopis	Sentimiento
2,4	173	0,62
2,1	293	0,51
3,1	40	0,50
1,7	233	0,43
2,3	274	0,38
2,7	54	0,23
2,9	52	0,18
1,3	131	0,01
1,2	0	

*Nota: Atributo relacionado con necesidad del cliente "Velocidad de procesamiento"*

**Tabla 45**

*Sentimiento asociado a cada valor del atributo "Modelo de GPU"*

Valor	Nºopis	Sentimiento
Mali-G68	37	0,74
Adreno 650	39	0,64
Adreno 642L	44	0,64
Mali-G76 MC4	41	0,64
Mali-G77 MC9	29	0,64
Adreno 660	34	0,63
Adreno 610	48	0,58
Mali-G52 MC2	82	0,56
Adreno 619	75	0,51
Apple GPU	21	0,50
Adreno 618	46	0,45
Mali-G57	22	0,44
Mali-G57 MC2	48	0,42



<b>Valor</b>	<b>Nºopis</b>	<b>Sentimiento</b>
IMG GE8320	51	0,42
Mali-G52	67	0,37
PowerVR GE8320	183	0,37
Mali-G72 MP3	22	0,34
Mali-G57 MC3	18	0,30
Adreno 620	18	0,27
Adreno 612	12	0,26
Adreno 640	20	0,25
Mali-T820	12	0,20
Apple GPU MP4	10	0,20
PowerVR GT7600	7	0,19
Adreno 642	18	0,18
IMG8322	13	0,15
Adreno 308	57	0,14
Mali-T720 MP2	11	0,07
IMG GE8100	9	-0,02
PowerVR 8320	9	-0,07
PowerVR GE8100	16	-0,20
Mali-G72	1	
Adreno 730	0	
Adreno 512	1	
Mali-T820 MP1	1	
Mali-G76	0	
Adreno 540	0	
Broadcom	0	
PowerVR GM9446	3	
PowerVR Series7XT Plus	0	
Mali-G78 MP20	0	
Mali-G57 MC4	0	
IMG 9XM-HP8	0	
Mali-G68 MC4	0	
Adreno 508	0	
Adreno 304	0	
Mali-G76 3EEMC4	0	
Mali-G52 MP2	0	

*Nota:* Atributo relacionado con necesidad del cliente “Velocidad de procesamiento”

**Tabla 46**

*Sentimiento asociado a cada valor del atributo “Tipo de resolución de la pantalla”*

Valor	Nºopis	Sentimiento
Full HD+	278	0,47
HD+	183	0,24
Quad HD+	20	0,23
HD	9	0,17
FHD+	10	0,11
QXGA+	11	0,02
FWVGA+	12	-0,08
FWVGA	10	-0,12
Full HD	5	
QQVGA	3	
HVGA	1	
QVGA	1	
Edge Quad HD+	0	
Quad HD	1	
WQHD+	0	

*Nota:* Atributo relacionado con necesidad del cliente “Calidad de imagen de pantalla”

**Tabla 47**

*Sentimiento asociado a cada valor del atributo “Píxeles por pulgada de la pantalla”*

Valor [ppi]	Nºopis	Sentimiento
405	118	0,50
393	91	0,46
421	52	0,33
341	60	0,31
267	150	0,23
477	15	0,10
284	46	0,09
161	27	-0,01
228	16	-0,07
101	11	

*Nota:* Atributo relacionado con necesidad del cliente “Calidad de imagen de pantalla”

**Tabla 48**

*Sentimiento asociado a cada valor del atributo "Tipo de batería"*

<b>Valor</b>	<b>Nºopis</b>	<b>Sentimiento</b>
Polímero de litio	544	0,41
Ion de litio	223	0,28
Polímero ion-litio	0	
Li-ion	6	

*Nota: Atributo relacionado con necesidad del cliente "Duración de la batería"*

**Tabla 49**

*Sentimiento asociado a cada valor del atributo "Con carga rápida"*

<b>Valor</b>	<b>Nºopis</b>	<b>Sentimiento</b>
1	571	0,41
0	354	0,27

*Nota: Atributo relacionado con necesidad del cliente "Duración de la batería"*