

FICHA RESUMEN

PROYECTO APLICADO – MAESTRÍA EN CIENCIA DE DATOS

TÍTULO: APLICACIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA EL ANÁLISIS DE LOS COMENTARIOS DEL CLIENTE DEL SEGMENTO PERSONAS EN BANCALIBRE

1. **ÁREA DE TRABAJO:** Análisis de sentimientos, Inteligencia Artificial, Sector financiero
2. **TIPO DE PROYECTO (Aplicado, Innovación, Investigación):** Aplicado
3. **ESTUDIANTE(S):** Juliana Andrea Arias Vergara y Carlos Andrés Salcedo Vaca
4. **CORREO ELECTRÓNICO:** jandrea95@javerianacali.edu.co y salcedo20@javerianacali.edu.co
5. **DIRECCIÓN Y TELEFONO:** Calle 61 #56-51 apto 1119, Medellín | 3128781263
6. **DIRECTOR:** Diego Luis Linares Ospina
7. **VINCULACIÓN DEL DIRECTOR:** Docente Facultad de Ingeniería y Ciencias
8. **CORREO ELECTRÓNICO DEL DIRECTOR:** dlinares@javerianacali.edu.co
9. **CO-DIRECTOR (Si aplica):** Gloria Inés Álvarez Vargas
10. **GRUPO O EMPRESA QUE LO AVALA (Si aplica):** Bancolombia
11. **OTROS GRUPOS O EMPRESAS:** No aplica
12. **PALABRAS CLAVE (al menos 5):** Comentarios del cliente, NPS Relacional, Sector financiero, Procesamiento de Lenguaje Natural (NPL), Análisis de sentimientos, Inteligencia Artificial.
13. **FECHA DE INICIO:** Julio del 2024
14. **DURACIÓN ESTIMADA (En meses):** 12 meses
15. **RESUMEN:**

Actualmente, se capturan y almacenan grandes cantidades de datos en formato texto, lo cual representa un reto significativo para su procesamiento. El análisis manual de estos datos consume demasiado tiempo y es poco práctico, por lo que es necesario desarrollar estrategias para mejorar estos procesos y obtener los resultados con mayor rapidez. En este trabajo se implementan técnicas avanzadas de ciencia de datos para analizar los comentarios escritos de los clientes del segmento de Personas de Bancalibre. La importancia de este estudio radica en la creciente necesidad de las entidades financieras de escuchar y analizar las opiniones de sus clientes para mejorar su experiencia, fidelización y satisfacción. Se aplican técnicas como el análisis de sentimientos, enfocado en el NPS Relacional para conocer la satisfacción del cliente con la empresa en general, y la inteligencia artificial (IA), para identificar en detalle los temas de mejora más recurrentes. El proceso comienza con la preparación del conjunto de datos, la selección e implementación de la técnica de IA adecuada, la evaluación del desempeño del

modelo y finaliza con el desarrollo de una interfaz para la visualización de resultados. Se espera obtener un modelo que permita analizar de manera eficiente los comentarios de los clientes, identificar temas recurrentes y proporcionar una herramienta útil para apoyar la toma de decisiones orientadas a mejorar la experiencia del cliente. Las posibles aplicaciones de este trabajo incluyen áreas de la organización como Mercadeo, Estrategia comercial y Servicio al Cliente, que ven en esta información un beneficio para aplicar en sus procesos y su alcance en la escalabilidad con otros tipos de retroalimentación de clientes, ampliando así su impacto y utilidad.



Pontificia Universidad
JAVERIANA
Cali

**APLICACIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA EL ANÁLISIS DE LOS
COMENTARIOS DEL CLIENTE DEL SEGMENTO PERSONAS EN BANCALIBRE**

Juliana Andrea Arias Vergara Código: 8993154

Carlos Andrés Salcedo Vaca Código: 8993469

*Proyecto Aplicado para optar al título de
Magister en Ciencia de Datos*

Director

Diego Luis Linares Ospina

Codirectora

Gloria Inés Álvarez Vargas

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, ENERO 1 DE 2025

TABLA DE CONTENIDO

1. CONTEXTUALIZACIÓN DEL PROYECTO	2
1.1. Definición del problema	2
1.1.1. Planteamiento del problema	2
1.1.2. Formulación del problema.....	2
2. OBJETIVOS DEL PROYECTO	3
2.1. Objetivos.....	3
2.1.1. Objetivo general	3
2.1.2. Objetivos específicos	3
3. ALCANCE.....	4
4. JUSTIFICACIÓN	5
5. MARCO TEÓRICO Y ANTECEDENTES.....	6
3.1. Marco teórico	6
3.2. Antecedentes.....	10
4. PREPARACIÓN DE DATOS	13
4.1. Análisis descriptivo	13
4.2. Preprocesamiento de los comentarios.....	16
4.3. Vectorización y división de los comentarios.....	17
5. MODELADO Y ANÁLISIS DE RESULTADOS	18
5.1. Experimentación con modelos supervisados	18
5.2. Experimentación con modelos no supervisados	24
6. CONCLUSIONES Y TRABAJOS FUTUROS	40
6.1. Conclusiones.....	40
6.2. Trabajos futuros	41
7. REFERENCIAS BIBLIOGRÁFICAS	42
8. ANEXOS	44

Lista de Ilustraciones

Ilustración 1 - Histograma variable NPS (elaboración propia)	14
Ilustración 2 - Histograma variable nps_desc (elaboración propia).....	15
Ilustración 3 - Método del codo K-Means hasta 100 clusters (elaboración propia)	26
Ilustración 4 - Método del codo K-Means hasta 1000 clusters (elaboración propia)	26
Ilustración 5 - Método Silhouette Score para diferentes valores de k (elaboración propia)	27
Ilustración 6 - Método del codo DBscan hasta 1000 clusters (elaboración propia).....	29
Ilustración 7 - Muestra de nube de palabras para los tópicos identificados (elaboración propia).....	34
Ilustración 8 - Gráfico de burbujas con distribución de comentarios por tópico (elaboración propia).....	35
Ilustración 9 - Gráfico bidimensional t-SNE por tópico (elaboración propia)	36
Ilustración 10 - Representación pyLDAvis (elaboración propia)	37
Ilustración 11 - Visualización del tablero interactivo con los resultados obtenidos (elaboración propia)	39

Lista de tablas

Tabla 1 - Resultados de métricas modelos escenario 1 (elaboración propia).....	18
Tabla 2 - Resultados de métricas modelos escenario 2 (elaboración propia).....	19
Tabla 3 - Resultados de métricas modelos escenario 3 (elaboración propia).....	19
Tabla 4 - Resultados de métricas con NLP de los tres escenarios (elaboración propia)	20
Tabla 5 - Resultados de métricas con representaciones numéricas de texto de los tres escenarios (elaboración propia)	21
Tabla 6 - Resultados de métricas modelos balanceados con datos sintéticos (elaboración propia)	22
Tabla 7 - Métricas modelos con lematización y representación TF-IDF (elaboración propia)	23
Tabla 8 - Hiperparámetros considerados para cada modelo (elaboración propia)	23
Tabla 9 - Resultados modelos con optimización de hiperparámetros (elaboración propia)	24
Tabla 10 - Resultados métricas modelos factorización matricial (elaboración propia)	25
Tabla 11 - Resultados métricas modelos clusterización K-means (elaboración propia)	28
Tabla 12 - Resultados métricas modelo de clusterización DBScan con eps = 0.4 (elaboración propia)	29
Tabla 13 - Resultados métricas modelo de clusterización DBScan con eps = 0.6 (elaboración propia)	30
Tabla 14 – Comparación métricas mejores modelos no supervisados (elaboración propia)	30
Tabla 15 -Resumen nombres asignados por GPT o1 y sus palabras claves por tópico (elaboración propia) ..	32

INTRODUCCIÓN

La experiencia del cliente se refiere a las percepciones y emociones que estos tienen hacia un producto, servicio o empresa, y que expresan a través de diversos canales. En la actualidad se ha convertido en un factor clave y diferenciador para el éxito y en la competitividad de las organizaciones ya que tiene un impacto directo en los resultados, además que es fundamental para generar estrategias de fidelización, mejorar la satisfacción e impulsar el crecimiento de un negocio [1]. Sin embargo, las empresas enfrentan desafíos significativos para comprender y mejorar esta experiencia debido al gran volumen de datos no estructurados generados por los clientes que crecen exponencialmente a través de múltiples fuentes de información.

El reto de este proyecto radica en procesar estos datos no estructurados, específicamente los comentarios escritos de los clientes recopilados a través de una encuesta de lealtad en BancaLibre, con el fin de evaluar la satisfacción del cliente y posteriormente categorizarlas en los temas más frecuentes. Esto proporciona información valiosa como apoyo para diferentes equipos dentro de la organización en función de la toma de decisiones estratégicas o de negocio.

Para abordar este desafío, se aplican técnicas avanzadas de ciencia de datos, particularmente el análisis de sentimientos y la inteligencia artificial. El análisis de sentimientos, una técnica de procesamiento del lenguaje natural (NLP) se utilizó para clasificar los comentarios en categorías emocionales como positivas, negativas o neutrales. Por otro lado, la inteligencia artificial facilita la categorización de los comentarios negativos, lo que permite identificar áreas de mejora y el desarrollo de planes de acción específicos para la organización.

Como resultado, se obtiene un modelo efectivo de ciencia de datos que permite analizar los comentarios de los clientes, así como una interfaz básica para la visualización y consulta de los resultados obtenidos. Con esto se pretende, de manera indirecta, liberar recursos humanos para ejecutar tareas de mayor valor, lo que ayuda a la eficiencia operativa y la calidad de la experiencia del cliente en BancaLibre.

Con lo anterior, se busca demostrar cómo la aplicación de técnicas de ciencia de datos puede transformar la gestión de la experiencia del cliente, permitiendo a esta entidad financiera apoyarse en estos resultados para tomar decisiones informadas y centradas en el cliente, lo que promoverá una mayor satisfacción y lealtad entre sus usuarios.

1. CONTEXTUALIZACIÓN DEL PROYECTO

1.1. Definición del problema

1.1.1. Planteamiento del problema

En la actualidad, las entidades financieras reconocen la importancia de escuchar la voz del cliente como un recurso esencial para impulsar diversas estrategias orientadas a la mejora continua. En este contexto, BancaLibre se compromete a colocar al cliente en el centro de sus operaciones para enriquecer su experiencia, lo que se traduce en una mayor fidelización y satisfacción del cliente.

Para lograrlo, la entidad cuenta con equipos dedicados a la experiencia del cliente cuya misión es comprender sus necesidades, diseñar y optimizar sus interacciones, evaluar su satisfacción, entre otras funciones. Una responsabilidad destacada es la gestión del feedback de los clientes, que se recolecta y analiza por medio de los comentarios escritos a través de diversas herramientas, como grupos focales, entrevistas y el Net Promoter Score (NPS).

Aunque estas herramientas permiten recoger y cuantificar la opinión del cliente hasta cierto punto, aún existe una brecha en el análisis eficiente de grandes volúmenes de información no estructurada que resulte en una categorización de temas recurrentes. Esto se debe a que actualmente no se cuenta con los métodos adecuados, lo que implica procesos manuales y la asignación de una capacidad de talentos especializados representando una alta demanda de tiempo, lo cual sí se hiciera de otra forma facilitaría el análisis de los puntos de mejora en la entidad y se aprovecharía esta capacidad para trabajar en temas que generan mayor valor dentro de estas áreas.

A partir de esto, existe una necesidad de comprender detalladamente las opiniones de los clientes del segmento Personas, en especial para clasificar de manera efectiva los temas más recurrentes en las opiniones negativas. Esto apoyará a que el área de interés pueda generar planes de acción específicos y enfocados en mejorar aspectos concretos de la experiencia del cliente.

1.1.2. Formulación del problema

Dada la necesidad de comprender a profundidad los comentarios de los clientes de la entidad financiera se plantea como pregunta principal ¿Cómo analizar los comentarios de los clientes del segmento personas para apoyar la toma de decisiones en BancaLibre? y como preguntas de sistematización ¿Cómo seleccionar y preparar los datos requeridos para el análisis?, ¿Qué técnicas de Inteligencia Artificial deben ser seleccionadas para procesar los comentarios de los clientes?, ¿Cómo identificar las opiniones negativas y extraer los temas recurrentes a partir de éstas?, ¿Cómo medir el desempeño del modelo de acuerdo con los resultados obtenidos? y ¿De qué manera presentar los resultados obtenidos?.

2. OBJETIVOS DEL PROYECTO

2.1 Objetivos

2.1.1. Objetivo general

Construir un modelo de ciencia de datos basado en técnicas de Inteligencia Artificial para analizar los comentarios de los clientes del segmento personas para apoyar la toma de decisiones en Bancalibre.

2.1.2. Objetivos específicos

- Alistar el conjunto de datos recopilado para el análisis.
- Elegir la técnica de Inteligencia Artificial que mejor se adecue para procesar los datos.
- Implementar la técnica elegida para clasificar las opiniones y encontrar los temas más frecuentes.
- Evaluar el desempeño del modelo definido.
- Desarrollar una interfaz que permita presentar los resultados obtenidos.

3. ALCANCE

A continuación, se detallan los alcances específicos para cada uno de los objetivos del proyecto que permitirán abordarlo de manera exitosa, así como los elementos que quedan fuera del compromiso actual.

Para el primer objetivo que consiste en **alistar el conjunto de datos recopilado para el análisis**, se ejecutará la recopilación, limpieza e imputación necesaria de estos, garantizando que tengan las mejores condiciones para su posterior procesamiento. Para esto, se utilizará un conjunto robusto de datos que corresponden a las verbalizaciones del NPS Relacional en la entidad financiera de los clientes del segmento personas en el periodo del 2022-2023. La recopilación de datos adicionales o posteriores más allá del conjunto de datos definido no será considerada en este proyecto.

El segundo objetivo de **elegir la técnica de Inteligencia Artificial que mejor se adecue para procesar los datos**, se ejecutará un análisis comparativo de diversas técnicas de este tema, que permita evaluar y seleccionar la más adecuada según los criterios de precisión, eficiencia y aplicabilidad al conjunto de datos disponible. Si bien se testearán varias técnicas, no significa que se implementarán todas.

En el tercer objetivo de **implementar la técnica elegida para clasificar las opiniones y encontrar los temas más frecuentes**, se pretende aplicar la técnica seleccionada en el punto anterior para procesar y clasificar las opiniones recogidas en el conjunto de datos, en positivos, neutros o negativos para luego encontrar los temas más recurrentes de acuerdo algún tipo de clasificación, estos últimos proporcionarán insights valiosos para apoyar la definición de planes de acción en la entidad.

En el cuarto objetivo de **evaluar el desempeño del modelo definido**, se medirá el rendimiento y robustez del modelo implementado, utilizando las métricas más acertadas. Aunque es posible que se realicen ajustes de optimización al modelo, la implementación y mejora iterativa del mismo serán consideradas en trabajos futuros.

Por último, para el quinto objetivo de **desarrollar una interfaz que permita presentar los resultados obtenidos** se elaborará una interfaz básica que permita visualizar de manera clara y fácil solo los resultados obtenidos en el proyecto. Cabe aclarar que la creación de una plataforma robusta y escalable para la visualización interactiva de datos no está incluida en el proyecto.

Es importante aclarar que el proyecto se limita hasta la evaluación del desempeño del modelo desarrollado. La implementación o mejora de este en la entidad financiera no hace parte del alcance de este, por lo que no se presentará evidencia relacionada con estas actividades.

4. JUSTIFICACIÓN

Actualmente, el análisis de la voz del cliente es fundamental para identificar áreas de mejora y tomar decisiones estratégicas que aumenten la satisfacción y la fidelización del cliente en cualquier organización. Para el caso de Bancolombia, este proyecto proporcionará a diferentes áreas una herramienta poderosa, ya que permitirá automatizar el análisis de grandes volúmenes de datos no estructurados, lo que ayudará a reducir la carga de trabajo manual y contribuirá de manera indirecta a que los equipos especializados se enfoquen en tareas de mayor valor propias de su rol. Además, será un soporte para una mejor toma de decisiones estratégicas basadas en datos, fortaleciendo la posición de la organización como una entidad centrada en el cliente.

La fuente de datos que se utilizará es de origen privado y consiste en los resultados históricos del indicador NPS (Net Promoter Score) enfocados en la voz del cliente, descartando información privada y sensible del cliente para efectos del análisis. Estas bases de datos son confiables, ya que se generan de manera periódica (cuatrimestral) a través de muestreos de encuestas dirigidas a los grupos de interés definidos por el banco. Dichos datos son recopilados con el apoyo de empresas especializadas y custodiados por el área de experiencia del cliente. La entidad financiera proveerá estos datos de manera consensuada, actualmente se está gestionando la autorización escrita para el correcto uso y acceso a los datos.

La viabilidad del proyecto es alta debido a la gran relevancia de este tema en la organización, la disponibilidad de información y el apoyo del equipo humano dedicado a la experiencia del cliente, de quienes proporcionarán apoyo y retroalimentación valiosa. Asimismo, las herramientas y técnicas avanzadas de ciencia de datos e inteligencia artificial no solo contribuyen a la factibilidad técnica del proyecto, sino que garantiza un análisis eficiente y preciso.

Finalmente, la implementación exitosa de este proyecto tendrá un impacto significativo dada la actual necesidad de implementar procesos individuales y personalizados de manera rápida y fácil, lo que permitirá mejorar continuamente la experiencia del cliente, aumentando a su vez su satisfacción y su lealtad, aspectos fundamentales para el crecimiento y la competitividad del banco en el mercado financiero.

5. MARCO TEÓRICO Y ANTECEDENTES

3.1. Marco teórico

A continuación, se presentan los conceptos y técnicas más relevantes para el desarrollo del proyecto, las cuales permitirán dar un contexto introductorio a los diferentes temas que se abordarán.

- Net Promoter Score (NPS)

Es una métrica utilizada para medir la lealtad de los clientes hacia una empresa. Éste se basa en una pregunta clave: "¿Qué tan probable es que recomiende nuestra empresa/producto/servicio a un amigo o familiar?". Los encuestados responden en una escala de 0 a 10, y según esto se clasifican en tres categorías:

- **Promotores (9-10):** Clientes leales que continuarán comprando y referirán a otros, impulsando el crecimiento.
- **Neutros (7-8):** Clientes satisfechos, pero no entusiastas, que son susceptibles a las ofertas de la competencia.
- **Detractores (0-6):** Clientes insatisfechos que pueden dañar la marca con comentarios negativos.

El NPS se calcula restando el porcentaje de detractores del porcentaje de promotores, resultando en una puntuación que puede variar entre -100 y 100. [2]

- *NPS Relacional*

Es una variante del NPS, la cual enfoca en medir la lealtad del cliente a lo largo del tiempo, evaluando la relación general del cliente con la empresa/producto/servicio en lugar de una interacción única. Esta métrica se obtiene a partir de encuestas periódicas que preguntan a los clientes sobre su probabilidad de recomendación. [3]

- *Datos no estructurados*

Los datos no estructurados representan la mayor parte de la información recopilada actualmente y se caracterizan por no ajustarse a un formato predefinido, como tablas o bases de datos. Pueden presentarse en diversos tamaños y formas, incluyendo texto libre, imágenes, archivos de audio o videos, lo que dificulta su organización y análisis.

Debido a su naturaleza diversa, estos datos requieren técnicas avanzadas como indexación y extracción de términos para dotarlos de estructura y facilitar su procesamiento en entornos de Big Data. [4]

- *Procesamiento de Lenguaje Natural (NLP)*

El Procesamiento de Lenguaje Natural (PLN o NLP, por sus siglas en inglés) se refiere al uso de técnicas computacionales para analizar el lenguaje humano hablado y escrito. Esta disciplina combina la lingüística con la ciencia computacional para comprender, interpretar y generar lenguaje de forma automática. El NLP permite analizar oraciones completas dentro de un problema social, facilitando la identificación de relaciones semánticas entre conceptos como causas y consecuencias. Su aplicación es clave en contextos donde el lenguaje natural representa la unidad principal de análisis, como en árboles de problemas formulados bajo la Metodología del Marco Lógico. [5]

- *Análisis de sentimientos*

Es una rama del procesamiento de lenguaje natural que se centra en identificar y clasificar la carga emocional expresada en textos, determinando si es positiva, negativa o neutra. [6] En la actualidad, este análisis representa un gran desafío debido al alto volumen de información disponible y a la diversidad de fuentes de datos. Los sistemas de análisis de sentimientos permiten a las empresas comprender mejor a sus clientes, optimizar sus estrategias de atención y mejorar la experiencia del usuario, lo que impacta directamente en la reputación de la marca.

- *Inteligencia Artificial (IA)*

Es un campo de la informática que busca desarrollar sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, el razonamiento y la toma de decisiones. Se basa en algoritmos y modelos que permiten a las máquinas procesar información, reconocer patrones y adaptarse a distintos contextos.

Gracias a su evolución, la IA sigue transformando múltiples sectores, optimizando procesos y mejorando la eficiencia en diversas áreas. [7]

- *Machine Learning (ML)*

Es una disciplina de la inteligencia artificial que permite a los sistemas aprender de los datos y mejorar su rendimiento sin ser programados explícitamente. Su objetivo principal es desarrollar modelos predictivos capaces de identificar patrones y hacer pronósticos basados en la información disponible.

Estas técnicas han evolucionado a partir de métodos estadísticos tradicionales y se utilizan ampliamente en diversas áreas como el marketing, la inteligencia de negocios y la detección de fraudes. Su correcta implementación permite a las organizaciones optimizar procesos y mejorar la toma de decisiones. [8]

- *Preprocesamiento minería de texto*

Es una fase esencial en el análisis de datos textuales, ya que prepara el texto para su posterior análisis mediante técnicas de procesamiento de lenguaje natural (NLP). Entre las técnicas más comunes se encuentran el stemming y la lematización.

- **Stemming:** reduce las palabras a su raíz o forma base, eliminando sufijos; por ejemplo, las palabras "corriendo" y "corrió" se reducen a "corr".
- **Lematización:** transforma las palabras a su forma base o "lema" considerando el contexto y la gramática. Por ejemplo, "corriendo" se convierte en "correr". [9]

- *Representación numérica de texto*

Son necesarias para analizar y moldear textos en representaciones numéricas. Algunas de las técnicas más utilizadas incluyen Bag of Words (BoW), TF-IDF y Word2Vec.

- **Bag of Words (BoW):** representa el texto mediante la creación de un vocabulario de palabras únicas presentes en el corpus y cuenta la frecuencia de aparición de cada palabra en los documentos.
- **TF-IDF (Term Frequency-Inverse Document Frequency):** representación las palabras según su frecuencia en un documento y su presencia en el conjunto total de documentos. Destacando términos relevantes y reduciendo la influencia de palabras comunes. [9]
- **Sent2Vec:** genera representaciones vectoriales de tamaño fijo para palabras, frases o textos completos. Su enfoque consiste en combinar los embeddings de palabras dentro de una oración para obtener una representación significativa del conjunto. [10]
- *Aprendizaje supervisado y no supervisado*

El aprendizaje supervisado es una técnica de Machine Learning en la que los modelos se entrenan con datos etiquetados, es decir, con un conjunto de características y una variable respuesta conocida. Su objetivo es construir un modelo predictivo que permita estimar valores de nuevas observaciones basándose en los datos de entrenamiento. En contraste, el aprendizaje no supervisado trabaja con datos sin etiquetar, donde no hay una variable respuesta definida. Su finalidad es descubrir patrones y estructuras en los datos, agrupándolos de acuerdo con similitudes o características compartidas. [11]

Algunos algoritmos de aprendizaje supervisado:

- **Support Vector Machine (SVM):** Algoritmo de aprendizaje supervisado utilizado para clasificación y regresión. Su principio fundamental es encontrar un hiperplano óptimo que separe los datos en distintas categorías, maximizando la distancia entre los puntos más cercanos de cada grupo, conocidos como vectores de soporte. Este es ampliamente utilizado en reconocimiento de patrones y análisis de datos. [12]
- **Random Forest:** Algoritmo de aprendizaje supervisado que construye múltiples árboles de decisión durante el entrenamiento y combina sus resultados para mejorar la precisión y evitar el sobreajuste. Cada árbol se entrena con una muestra aleatoria del conjunto de datos y selecciona un subconjunto aleatorio de características en cada división, lo que introduce diversidad entre los árboles y fortalece el modelo final. [13]
- **Gradient Boosting:** Algoritmo de aprendizaje supervisado que construye modelos predictivos mediante la combinación de múltiples modelos débiles, como árboles de decisión, de manera secuencial. En cada iteración, se ajusta un nuevo modelo a los residuos o errores cometidos por el conjunto de modelos anteriores, con el objetivo de minimizar una función de pérdida específica. Este enfoque permite mejorar la precisión y robustez del modelo final, siendo especialmente útil en tareas de clasificación y regresión. [14]
- **XGBoost:** El Extreme Gradient Boosting es una implementación optimizada del Gradient Boosting destacada por su eficiencia y rendimiento en diversos contextos. [13]
- **Redes Neuronales:** Son modelos computacionales inspirados en el funcionamiento del cerebro humano, formadas por unidades básicas llamadas perceptrones, que trabajan en conjunto para procesar información. Estas unidades están organizadas en capas: una de entrada, capas ocultas intermedias y una de salida, permitiendo la detección de patrones complejos. Cada perceptrón realiza una combinación lineal de sus entradas y aplica una función de activación para producir una salida. A través de un proceso de entrenamiento, la red ajusta los pesos de sus conexiones para mejorar la precisión en tareas como clasificación, predicción y reconocimiento de patrones. [15]

Algunos algoritmos de aprendizaje no supervisado [16]:

- **Latent Semantic Indexing (LSI):** es una técnica estadística desarrollada para identificar términos semánticamente relacionados dentro de un corpus. Se basa en la técnica de Singular Value Decomposition (SVD) para descomponer matrices y encontrar patrones ocultos en la coocurrencia de palabras dentro de los textos. Su objetivo es revelar términos latentes que comparten un significado similar y agruparlos en temas.
- **Latent Dirichlet Allocation (LDA):** es una técnica de modelado de temas que identifica patrones ocultos en textos mediante un enfoque probabilístico. Asume que cada documento contiene una combinación de temas y que cada tema está compuesto por un grupo de palabras relacionadas. Esta técnica es ampliamente utilizada en la clasificación de documentos y la exploración de grandes volúmenes de texto.
- **Non-Negative Matrix Factorization (NMF):** es una técnica de descomposición matricial

similar a la Singular Value Decomposition (SVD), pero con la restricción de que solo trabaja con valores no negativos. Su objetivo es descomponer una matriz dada en dos matrices más pequeñas, cuya multiplicación aproxima a la matriz original.

3.2. Antecedentes

- *Análisis de sentimientos y emociones en redes sociales usando ML [17]*

El objetivo del trabajo es desarrollar y comparar sistemas de clasificación de emociones en textos en español extraídos de Twitter. Este sigue un enfoque sistemático que incluye la recopilación y etiquetado de datos, seguido del preprocesamiento del texto con tareas de tokenización, stemming, lematización y normalización, luego se usan técnicas de extracción de características como Bag of Words (BoW), Word2Vec y TF-IDF, y por último se aplican y evalúan varios algoritmos de clasificación supervisada como Support Vector Machine (SVM), Random Forest y Regresión Logística, empleando métricas como precisión, recall y F1 para seleccionar el modelo óptimo. Obteniendo como resultado, que el SVM es el más adecuado para la clasificación de emociones en este contexto.

Todas las herramientas y técnicas utilizadas en este trabajo podrían considerarse para el desarrollo del proyecto en curso y se encuentra una similitud parcial con el objetivo al querer identificar sentimientos y emociones, aunque varía en que se desarrolla en ámbitos diferentes y los resultados esperados en el proyecto van un poco más allá de la clasificación de emociones en este contexto. Por otra parte, el trabajo de referencia se centra en datos de redes sociales y en el proyecto se centran en comentarios de clientes recogidos a través de diversas herramientas como grupos focales, entrevistas y el Net Promoter Score (NPS).

- *El análisis de sentimientos para la gestión de redes sociales en el turismo [18]*

Este trabajo analiza la información subjetiva compartida en redes sociales en el contexto del turismo, enfocándose en la ciudad de Córdoba (España) y su influencia en la percepción turística. Su objetivo es integrar esta información en los sistemas de gestión de empresas turísticas para mejorar su rendimiento y toma de decisiones. A través de una metodología estructurada, primero definen el análisis de sentimientos y la relación con la reputación online, luego recopilan los datos de las webs de opinión y redes sociales, empleando herramientas como Lingmotif. Después a través de técnicas de preprocesamiento de datos y extracción de características, entrenan y validan modelos de clasificación supervisada como SVM y Random Forest, donde encuentran resultados que indican que las webs de opinión son más valiosas que las redes sociales para el análisis de sentimientos, resaltando la importancia de estas herramientas en la gestión y optimización de la experiencia turística y reputación online.

El trabajo proporciona un marco detallado para abordar en el análisis de sentimientos utilizando herramientas específicas como por ejemplo Lingmotif, que pueden llegar a ser aplicables al proyecto. Así mismo algunas de las técnicas utilizadas en este trabajo podrían adaptarse en el contexto del proyecto. Por otro lado, se identifica una similitud parcial al querer identificar sentimientos y emociones, aunque en contextos diferentes y buscan integrar estos resultados en estrategias de gestión para mejorar la experiencia del cliente para lanzar nuevos productos, mejorar ventas o fidelizarlos.

- *Técnicas de Análisis de Sentimientos Aplicadas a la Valoración de Opiniones en el Lenguaje Español [19]*

Este trabajo estudia el análisis de sentimientos sobre reseñas de películas en español, utilizando técnicas automatizadas de procesamiento y análisis de datos. Para esto utiliza un conjunto de más de 50,000 comentarios, evaluando varios modelos de procesamiento del lenguaje natural (PLN) para clasificar sentimientos de textos en español coloquial. Emplean algoritmos clásicos de aprendizaje automático como SVM, Naive Bayes, Regresión Logística y Random Forest, junto con métodos avanzados como embeddings con LSTM y transformers (BETO). Los cuales se evalúan con métricas estándar como exactitud, precisión y F1-Score, mostrando que BETO supera significativamente a las técnicas tradicionales en la clasificación de sentimientos.

Estos resultados obtenidos en este trabajo de referencia resaltan un punto muy llamativo donde las técnicas más actuales presentaron un mejor rendimiento para el ejercicio. Esto nos puede dar un indicio para ser consideradas de manera prioritaria en la aplicación del proyecto propuesto. Como en los trabajos referenciados hasta el momento se evidencia que la principal diferencia es con respecto a la profundidad de alcance de estos.

- *Aplicación de técnicas de Machine Learning para el Análisis de Interacciones en Twitter: Caso de estudio – Gestión de fallas operativas para el Banco Davivienda y el Neo Banco Daviplata [20]*

Este trabajo diseña e implementa un modelo de Machine Learning para analizar interacciones en Twitter y detectar fallas operativas en los servicios del Banco Davivienda y Daviplata, optimizando así los servicios bancarios y minimizando la insatisfacción del cliente. Siguiendo la metodología Cross-Industry Standard Process for Data Mining (CRISP-DM), se recopiló y preprocesó datos de Twitter de trece meses. Se utilizó K-means para segmentar las interacciones en categorías de problemas operativos, y se entrenaron modelos supervisados como SVM y árboles de decisión para categorizar futuras interacciones. El estudio concluye que SVM mostró mejores indicadores de desempeño, destacando la importancia de este modelo para gestionar proactivamente las fallas operativas y mejorar la experiencia del cliente.

El trabajo de referencia está muy relacionado con el proyecto, ya que se desarrolla en el contexto financiero, así mismo tiene un alcance similar donde adicional al análisis de sentimientos se profundiza en una categorización a partir de interacciones donde los usuarios presentan una disconformidad. Este se enfoca en la detección de fallas operativas en los servicios del Banco Davivienda y Daviplata, mientras que el proyecto busca identificar temas recurrentes relacionados con la experiencia del cliente en un segmento particular. Adicionalmente, se encuentra otra diferencia relacionada con los datos de entrada que se utilizan en el trabajo donde se parte de comentarios de clientes en Twitter mientras que en el proyecto se parte de los comentarios que se recolectan en la medición del NPS.

- *Análisis predictivo en Twitter para detectar patrones de personas con tendencia Hacktivista aplicando Big Data, Machine Learning y Deep Learning [21]*

El trabajo desarrolla un modelo predictivo para identificar y categorizar tweets con contenido hacktivista (uso no violento de herramientas digitales persiguiendo fines políticos), utilizando técnicas de Big Data, Machine Learning y Deep Learning, centrado en los seguidores de la cuenta de Twitter del expresidente de Colombia Iván Duque. Este sigue la metodología CRISP-DM, recopilan tweets mediante la API de Twitter, preprocesan los datos y crean un dataset específico para el análisis de hacktivismo. Implementan y entrenan modelos como regresión logística, árboles de decisión, SVM y redes neuronales, evaluándolos con métricas como precisión, recall y F1-Score. Los resultados destacan a SVM y redes neuronales por su precisión y confiabilidad, demostrando la viabilidad de predecir y categorizar tweets hacktivistas para prevenir ataques cibernéticos y mejorar la seguridad en redes sociales.

El trabajo se diferencia en que busca desarrollar un modelo enfocado en predecir y categorizar tweets con intenciones hacktivistas para prevenir ataques cibernéticos, mientras que el proyecto busca clasificar comentarios de clientes que apoyen la toma de decisiones estratégicas en BancaLibre. Así mismo, el contexto en que se desarrolla el trabajo de referencia es el campo político a partir de comentarios en redes sociales. Por otro lado, el trabajo aporta al proyecto en la consideración de incorporar la metodología CRISP – DM para el desarrollo de éste. También, se relacionan en que llegan a una categorización de comentarios y usan datos no estructurados, lo que puede ser de valor para el proyecto en el uso de estas técnicas de preparación y aplicación en este tipo de datos.

4. PREPARACIÓN DE DATOS

4.1. Análisis descriptivo

Para alistar el conjunto de datos recopilado y prepararlo para el análisis, primero se realizó una evaluación descriptiva de la base obtenida. Esta base de datos cuenta con 74,402 registros y 12 variables. De estas últimas, 4 son variables cuantitativas: el identificador (id), el año (a__o), el mes (mes) y la calificación del Net Promoter Score (nps) y las otras 8 son cualitativas: la técnica utilizada (técnica), la descripción del NPS (nps_desc), la verbalización de la emoción (verb_emocion), el tipo de encuesta (tipo_encuesta), el segmento (segmento), la región (region), el género (genero_cli) y la edad del cliente (edad_cli).

De estos, el atributo de entrada a los modelos son los comentarios de los clientes los cuales están registrados en la variable verb_emocion y el atributo de salida que se usara como etiquetas para los modelos supervisados están registrados en la variable nps_desc siendo estas Promotor, Neutro y Detractor.

Se comenzó con la exploración de cada uno de los campos verificando que no existen valores nulos en el conjunto de datos. Luego para comprender mejor la distribución de los datos, fue importante visualizar las variables clave como el nps y el nps_desc el cual hace referencia a los valores categóricos que se le asignan al NPS. Por eso se construyeron un histograma con los valores cuantitativos y un diagrama de barras con los valores cualitativos.

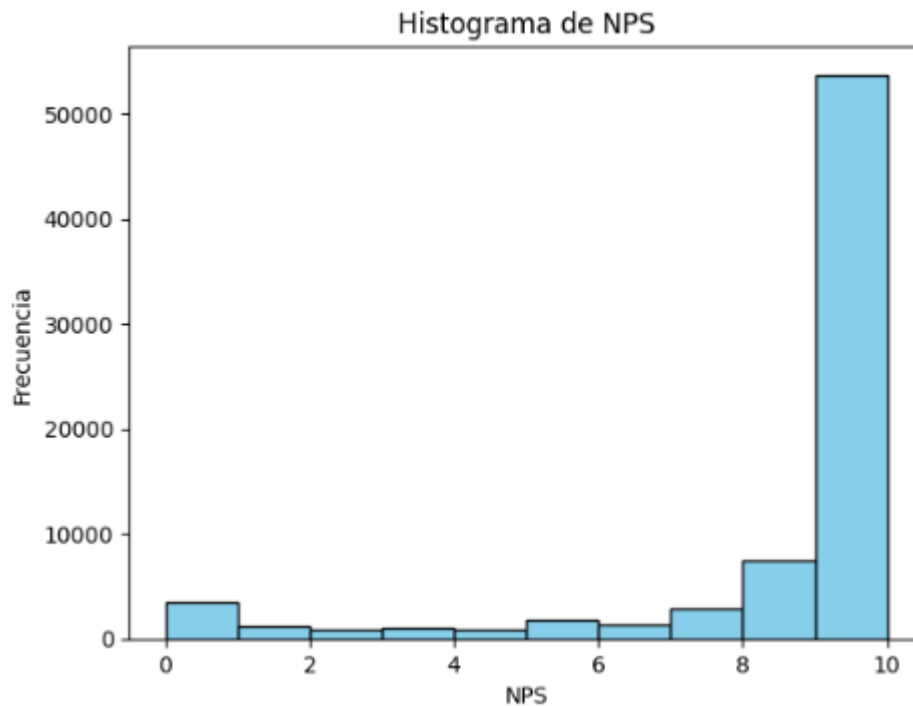


Ilustración 1 - Histograma variable NPS (elaboración propia)

En la **Ilustración 1 se observa el histograma de la variable nps**, donde se ilustra la distribución de las calificaciones otorgadas en el sistema de Net Promoter Score (NPS), que se clasifican de la siguiente manera:

- **Detractores:** Calificaciones de 0 a 6.
- **Neutros:** Calificaciones de 7 y 8.
- **Promotores:** Calificaciones de 9 y 10.

La distribución muestra una asimetría pronunciada hacia la derecha, lo que indica que un número significativo de clientes otorgaron la calificación máxima (10). Siendo la mayor frecuencia, con más de 50,000 observaciones, lo que sugiere una tendencia marcada de los clientes a calificar sus experiencias como muy positivas.

Por otro lado, se observa un número reducido de calificaciones entre 0 y 6, lo que corresponde a los detractores, lo que evidencia que los clientes insatisfechos son una minoría. Asimismo, un grupo pequeño pero significativo otorgó calificaciones de 7 u 8, es decir, clientes que no son ni promotores ni detractores, sino neutros representando una oportunidad para la entidad financiera de convertirlos en promotores mediante acciones específicas.

Este histograma evidencia una excelente satisfacción del cliente, donde una mayoría significativa da la máxima calificación. Sin embargo, el hecho de que haya pocos detractores podría llevar a una falsa sensación de seguridad, ya que un bajo número de comentarios negativos podría ocultar problemas puntuales en la experiencia del cliente que deben ser investigados en mayor profundidad.

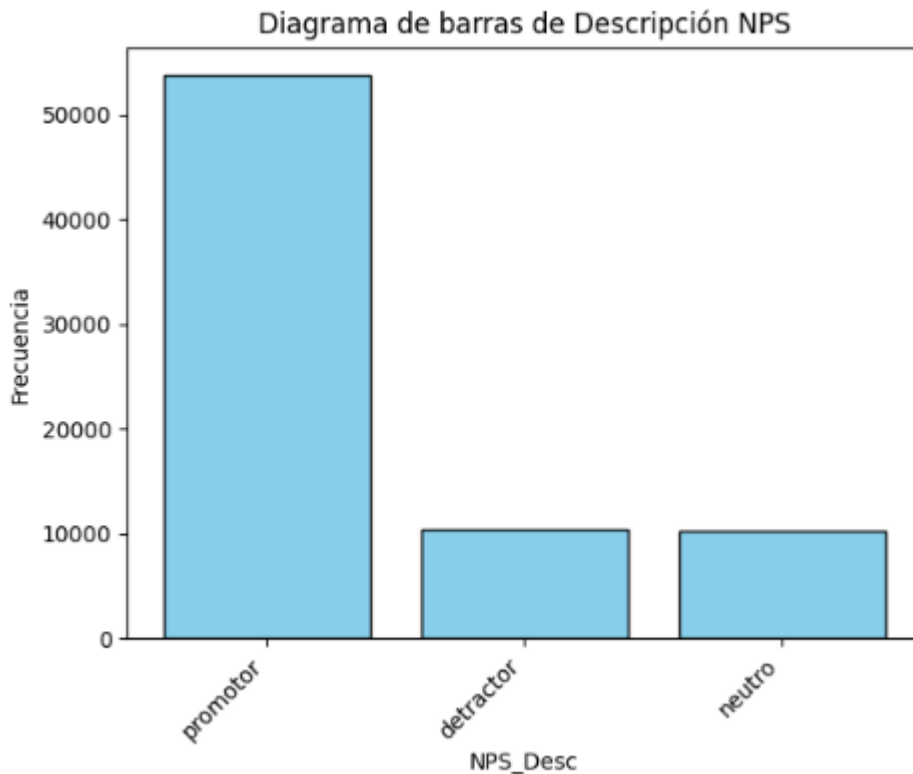


Ilustración 2 - Histograma variable nps_desc (elaboración propia)

La **Ilustración 2** presenta un diagrama de barras que clasifica a los clientes en promotores, detractores, y neutros, basada en la descripción derivada del puntaje NPS. La categoría de promotores es claramente dominante, con más de 50,000 registros, lo que respalda la tendencia observada en el histograma anterior. Esto significa que la mayoría de los clientes que interactuaron con la entidad financiera están dispuestos a recomendar sus servicios.

Por otro lado, las categorías de detractores y neutros están presentes en cantidades relativamente similares, pero significativamente menores que la de los promotores. Esto también es consistente con lo observado en el histograma, donde la mayoría de las calificaciones bajas y neutras eran pocas en comparación con las altas.

El diagrama de barras refuerza la conclusión de que la empresa cuenta con una base sólida de promotores. Sin embargo, es crucial no subestimar ni ignorar la proporción de detractores y clientes

neutrales, ya que representan un área clave para enfocar los esfuerzos de mejora. Abordar las inquietudes de los detractores y convertir a los neutros en promotores puede ser fundamental para optimizar la experiencia del cliente y fortalecer la lealtad hacia la entidad.

4.2. Preprocesamiento de los comentarios

Ahora, se describe el preprocesamiento de texto aplicado a los comentarios de los clientes que otorgaron su calificación en el NPS. Este proceso es fundamental para preparar los datos para un análisis más profundo y efectivo.

Se llevaron a cabo varias transformaciones en los comentarios (verb_emocion) para garantizar que estuvieran limpios y estandarizados. Primero, se eliminaron las tildes, lo que facilita el análisis, especialmente en contextos donde la acentuación puede variar. Luego, se convirtieron todas las letras a minúsculas para asegurar la uniformidad en la comparación de palabras.

Adicionalmente, se eliminaron los caracteres especiales y números que no aportan valor al análisis. Se llevó a cabo la eliminación de stopwords, es decir, palabras de parada que no aportan significado al análisis, con la excepción de la palabra "no", que se mantuvo debido a su importancia en el contexto de las opiniones. También se gestionaron las repeticiones de caracteres, permitiendo excepciones para las combinaciones "rr" y "ll", que son válidas en el español.

Otro aspecto importante del preprocesamiento fue la eliminación de palabras de un solo carácter o aquellas sin vocales, siempre que no pertenecieran a la lista de siglas definidas. Además, se reemplazaron las secuencias de dos o más espacios dentro de los registros por un único espacio. Finalmente, se eliminaron los registros que quedaron vacíos tras este proceso.

En resumen, las transformaciones realizadas a los comentarios son las siguientes:

1. Remoción de tildes.
2. Conversión a minúsculas.
3. Eliminación de caracteres especiales y números.
4. Eliminación de stopwords (excepto "no").
5. Reducción de repeticiones de caracteres (excepciones para "rr" y "ll").
6. Eliminación palabras de 1 carácter o palabras sin vocales.
7. Reemplazo de secuencias de 2 o más espacios por un solo espacio.
8. Eliminación de filas vacías luego del preprocesamiento anterior.

Como resultado de este preprocesamiento, se obtuvo un conjunto de datos compuesto por 71,906 registros y 12 variables, dividido para cada clase de la siguiente manera:

- Promotor: 51.971 registros
- Neutro: 9.760 registros
- Detractor: 10.175 registros

Evidenciando esto un desbalance significativo, con una predominancia de la clase promotor.

4.3. Vectorización y división de los comentarios

Una vez realizado el preprocesamiento de los comentarios, se procedió a transformar los textos en representaciones numéricas mediante la técnica TF-IDF (Term Frequency–Inverse Document Frequency). Esta técnica permite identificar no solo la frecuencia de las palabras en cada documento, sino también su relevancia dentro del conjunto total de textos, lo cual resulta especialmente útil para modelar el contenido semántico de los comentarios. El vectorizador fue entrenado sobre la base completa preprocesada, generando una matriz dispersa con un vocabulario final de 13.883 términos únicos, que sirvió como insumo para los modelos de clasificación y análisis de tópicos desarrollados posteriormente. Adicionalmente, se evaluaron otras dos técnicas de representación numérica Bag of Words, con un vocabulario de 16.197 términos, y embeddings Sent2Vec, el cual generó un vector de dimensión 384.

Por otro lado, antes de la división del conjunto en entrenamiento y prueba, se abordó el desbalance existente en la variable `nps_desc`, que clasifica a los clientes como Promotores, Neutros y Detractores. Inicialmente, se aplicó una técnica de submuestreo aleatorio de la clase mayoritaria, seleccionando un número igual de registros por clase según el tamaño de la clase minoritaria, lo que permitió obtener una base balanceada entre categorías.

Además, se implementó una técnica de balanceo más avanzada: SMOTE (Synthetic Minority Over-sampling Technique). Esta técnica genera ejemplos sintéticos de las clases minoritarias, pero solo fue aplicada al conjunto de entrenamiento, asegurando que el modelo aprendiera sobre una distribución equilibrada sin afectar la validez de la evaluación. El conjunto de prueba se mantuvo íntegramente compuesto por datos reales, garantizando así una medición justa y realista del desempeño del modelo en condiciones no manipuladas.

Posteriormente, estos conjuntos de datos fueron divididos en un 80% para entrenamiento (`train`) y 20% para prueba (`test`), utilizando una partición aleatoria estratificada para preservar la proporción entre clases.

5. MODELADO Y ANÁLISIS DE RESULTADOS

5.1. Experimentación con modelos supervisados

Para alcanzar este objetivo se inició la exploración con tres modelos supervisados, seleccionando el primero a partir de los antecedentes donde el modelo más exitoso fue el Support Vector Machine (SVM), la segunda elección partió de la gran utilidad que se le da al Random Forest y el tercero ya por decisión propia de evaluar si un árbol más robusto mejora el rendimiento se probó un Gradient Boosting. En estos tres modelos se utilizó la representación numérica de texto TF – IDF (Term Frequency-Inverse Document Frequency).

Se plantea un primer escenario teniendo en cuenta las tres clases Detractor, Neutro, y Promotor balanceando a la clase minoritaria (Neutro) con 9.760 registros. Al aplicar los modelos mencionados a este escenario se obtienen los siguientes resultados:

Modelo	Accuracy	Clase	f1-score
SVM	0.60	Detractor	0.69
		Neutro	0.42
		Promotor	0.65
Random forest	0.58	Detractor	0.69
		Neutro	0.37
		Promotor	0.64
Gradient Boosting	0.56	Detractor	0.66
		Neutro	0.40
		Promotor	0.58

Tabla 1 - Resultados de métricas modelos escenario 1 (elaboración propia)

De la **Tabla 1** se puede observar a primera vista que la exactitud de los tres modelos es baja con valores iguales y menores al 60% también se evidencia que es un desafío clasificar la clase Neutro mostrando esta un rendimiento promedio con un f1-score del 40% (0.40) para todos los modelos, mientras que las otras dos clases (detractores y promotores) superaron en promedio el 60% (0.60).

Debido al bajo rendimiento de la clase Neutro y buscando una solución para estos se toma la decisión de plantear un segundo escenario donde se tienen en cuenta solo dos clases Promotor y Detractor, agrupando en esta última los neutros con los detractores, aquí se balancea el conjunto de datos igualmente a la clase minoritaria (Detractor + Neutro) con 19.935 registros. Al aplicar los modelos mencionados a este segundo escenario se obtienen los siguientes resultados:

Modelo	Accuracy	Clase	f1-score
SVM	0.75	Detractor + Neutro	0.74
		Promotor	0.76
Random forest	0.75	Detractor + Neutro	0.74
		Promotor	0.76
Gradient Boosting	0.72	Detractor + Neutro	0.73
		Promotor	0.70

Tabla 2 - Resultados de métricas modelos escenario 2 (elaboración propia)

De la **Tabla 2** se evidencia una exactitud mayor de los tres modelos respecto al escenario anterior en promedio mayor al 70% (0.70) y un aumento en el f1-score promedio de Detractor + Neutro aquí del 74% (0.74) comparado con la clase Detractor del escenario anterior del 68% (0.68), obteniendo una mejoría de 6%.

Aunque estas métricas pueden considerarse “buenas” probaremos un tercer escenario donde se tendrán en cuenta dos clases Promotor y Detractor eliminando en su totalidad la clase Neutro ya que al analizarla es una clase compleja de clasificar ya que puede ser muy ambigua y esto puede afectar el rendimiento del modelo. En este último escenario también se balancean los datos hacia la clase minoritaria (Detractor) con 10.175 registros ya que solo se tienen en cuenta la clase Promotor y Detractor. Al aplicar los modelos mencionados a este tercer escenario se obtienen los siguientes resultados:

Modelo	Accuracy	Clase	f1-score
SVM	0.87	Detractor	0.87
		Promotor	0.86
Random forest	0.86	Detractor	0.87
		Promotor	0.86
Gradient Boosting	0.81	Detractor	0.83
		Promotor	0.79

Tabla 3 - Resultados de métricas modelos escenario 3 (elaboración propia)

De la **Tabla 3** se observa que la estrategia implementada en este escenario permite simplificar la tarea de clasificación y mejorar bastante el desempeño general del modelo. Al eliminar la clase neutra, los resultados mostraron un desempeño más equilibrado y sólido con un Accuracy y F1-score promedio superior a 84% (0.84) en ambas clases, consolidándose como la opción más eficiente hasta el momento.

De los resultados de los escenarios anteriores se opta por el **tercer escenario**, donde se obtuvieron las mejores métricas y teniendo en cuenta que parte de los objetivos del proyecto es identificar los temas recurrentes de mejora específicamente de los clientes detractores, no se genera impacto al eliminar la clase neutra.

Para cada uno de los tres escenarios mencionados, también se probaron combinaciones de técnicas de procesamiento del lenguaje natural (NLP) como Stemming, Lematización y una combinación de ambas, evaluando su impacto en la capacidad de los modelos para capturar la semántica de las opiniones.

Modelo	Escenario 1			Escenario 2			Escenario 3		
	Accuracy	Clase	f1-score	Accuracy	Clase	f1-score	Accuracy	Clase	f1-score
SVM + Stemming	0.61	Detractor	0.70	0.76	Detractor + Neutro	0.75	0.87	Detractor	0.87
		Neutro	0.43		Promotor	0.77		Promotor	0.86
		Promotor	0.67						
SVM + Lematización	0.60	Detractor	0.70	0.76	Detractor + Neutro	0.75	0.87	Detractor	0.88
		Neutro	0.42		Promotor	0.77		Promotor	0.87
		Promotor	0.65						
SVM + Stemming + Lematización	0.61	Detractor	0.71	0.77	Detractor + Neutro	0.76	0.87	Detractor	0.88
		Neutro	0.43						
		Promotor	0.67		Promotor	0.77		Promotor	0.87

Tabla 4 - Resultados de métricas con NLP de los tres escenarios (elaboración propia)

Según la **Tabla 4** el modelo más la lematización y el modelo más el stemming y la lematización mostraron los mejores resultados en el escenario 3, probablemente debe ser por la lematización ya que esta conserva la raíz de las palabras en su forma semántica más significativa. En consecuencia, se elige el modelo SVM más lematización ya que tiene el mismo rendimiento que el otro y con un paso menos.

Luego, aplicando a los mismos tres escenarios descritos anteriormente, se evaluaron otras técnicas de representación numérica del texto con el objetivo de comparar su desempeño frente a la ya utilizada TF-IDF y confirmar su eficiencia. Para esta experimentación se utilizó el modelo SVM, dado que hasta al momento ha sido el que mejor rendimiento ha mostrado en los distintos escenarios. Las técnicas probadas fueron Sent2Vec y Bag of Words (BoW), los cuales se aplicaron bajo las mismas condiciones para garantizar una comparación justa. A continuación, se presentan los resultados obtenidos:

Modelo	Escenario 1			Escenario 2			Escenario 3		
	Accuracy	Clase	f1-score	Accuracy	Clase	f1-score	Accuracy	Clase	f1-score
SVM + TF-IDF	0.60	Detractor	0.69	0.75	Detractor + Neutro	0.74	0.87	Detractor	0.87
		Neutro	0.42		Promotor	0.76		Promotor	0.86
		Promotor	0.65						
SVM + Sent2Vect	0.56	Detractor	0.66	0.73	Detractor + Neutro	0.72	0.81	Detractor	0.82
		Neutro	0.37		Promotor	0.73		Promotor	0.80
		Promotor	0.62						
SVM + Bag of Words	0.57	Detractor	0.67	0.74	Detractor + Neutro	0.72	0.85	Detractor	0.85
		Neutro	0.40		Promotor	0.76		Promotor	0.84
		Promotor	0.63						

Tabla 5 - Resultados de métricas con representaciones numéricas de texto de los tres escenarios (elaboración propia)

De los resultados de la **Tabla 5** se concluye que **los mejores resultados se mantienen con TF-IDF**, debido probablemente a su capacidad para captar el peso relativo de las palabras en el contexto del corpus, aunque bolsa de palabras también funcionó muy bien obteniendo exactitud de un 85% (0.85).

Escenarios de prueba frente a las técnicas avanzadas de balanceo de datos

Con la intención de evaluar la hipótesis de que la generación de datos sintéticos posterior a la representación numérica del texto mejoraría el rendimiento del modelo, se aplicaron técnicas avanzadas de balanceo de clases, como SMOTE a los mismos tres escenarios inicialmente planteados limitado al modelo de Support Vector Machine (SVM) dado que es el que mejor resultado ha generado hasta el momento de la siguiente manera:

- En el escenario 1:
 - Modelo 1: Promotor submuestreado a 30,000 y Detractor/Neutro sobremuestreados a 30,000.
 - Modelo 2: Promotor submuestreado a 30,000 y Detractor/Neutro sobremuestreados a 15,000.
- En el escenario 2:
 - Modelo 1: Promotor submuestreado a 30,000 y Detractor sobremuestreado a 30,000.
 - Modelo 2: Promotor submuestreado a 30,000 y Detractor sobremuestreado a 20,000.
- En el escenario 3:
 - Modelo 1: Promotor submuestreado a 30,000 y Detractor+Neutro sobremuestreados a 30,000.

- Modelo 2: Promotor submuestreado a 30,000 y Detractor+Neutro sobremuestreados a 15,000.

Modelo	Escenario 1			Escenario 2			Escenario 3		
	Accuracy	Clase	f1-score	Accuracy	Clase	f1-score	Accuracy	Clase	f1-score
Modelo 1	0.66	Detractor	0.55	0.78	Detractor + Neutro	0.63	0.87	Detractor	0.68
		Neutro	0.26		Promotor	0.84		Promotor	0.92
		Promotor	0.79		Detractor +Neutro	0.64		Detractor	0.69
Modelo 2	0.72	Detractor	0.60	0.80	Detractor +Neutro	0.64	0.88	Detractor	0.69
		Neutro	0.24		Promotor	0.86		Promotor	0.93
		Promotor	0.84						

Tabla 6 - Resultados de métricas modelos balanceados con datos sintéticos (elaboración propia)

La **Tabla 6** muestra que, a pesar de implementar este tipo de balanceo, las métricas de clasificación para la clase minoritaria no mostraron mejoras significativas. De hecho, al aumentar la cantidad de datos sintéticos, el rendimiento del modelo se deteriora un poco.

Estos hallazgos llevaron al rechazo de la hipótesis inicial, indicando que la generación de datos sintéticos, en este contexto, no contribuye a mejorar el desempeño del modelo.

Con estas experimentaciones, se **logró llegar a un modelo base utilizando un Support Vector Machine (SVM) aplicado al conjunto de datos balanceados a la clase minoritaria con lematización y representación TF-IDF**. Este modelo demostró un desempeño satisfactorio en el escenario de dos clases (promotor y detractor) eliminando la clase neutra, lo que lo convirtió en un punto de referencia ideal para compararlo con otros enfoques más modernos a juicio de expertos, como un Árbol de Decisión y un Perceptrón Multicapa, bajo las mismas condiciones de preprocesamiento y representación numérica de texto.

Por el lado del Árbol de Decisión, se optó por utilizar el XGBoost, un algoritmo robusto y ampliamente utilizado por su capacidad para manejar conjuntos de datos complejos y su eficiencia en tareas de clasificación. Asimismo, se experimentó con un Perceptrón Multicapa (MLP), una red neuronal básica que permite capturar relaciones no lineales en los datos. Ambos modelos se entrenaron utilizando el mismo conjunto de datos balanceados, con lematización y representación numérica de texto TF-IDF, para garantizar una comparación justa.

Posteriormente, se llevó a cabo un proceso de optimización de hiperparámetros para cada uno de los tres modelos (SVM, XGBoost y Perceptrón Multicapa) utilizando la técnica de búsqueda en cuadrícula (grid search).

Tras la experimentación descrita y la selección de los modelos a comparar, se ejecutaron los modelos utilizando un conjunto de datos balanceado hacia la clase minoritaria (Detractor), con un total de 10,175 registros. A continuación, se presentan los resultados obtenidos en términos de accuracy (exactitud) y f1-score para cada modelo:

Modelo	Accuracy	Clases	f1-score
SVM	0.87	Detractor	0.87
		Promotor	0.88
XGBoost	0.85	Detractor	0.85
		Promotor	0.86
Perceptrón Multicapa	0.83	Detractor	0.83
		Promotor	0.83

Tabla 7 - Métricas modelos con lematización y representación TF-IDF (elaboración propia)

De los resultados detallados en la **Tabla 7**, se observa que el modelo de Support Vector Machine (SVM) supera en exactitud y rendimiento tanto al XGBoost como al Perceptrón Multicapa, convirtiéndose en el modelo más efectivo para esta tarea.

Adicionalmente, se realizó el proceso de optimización de hiperparámetros mediante búsqueda en cuadrícula (*grid search*). A continuación, en la **Tabla 8** se describen los hiperparámetros considerados para cada modelo:

Modelo	Hiperparámetros
SVM	<ul style="list-style-type: none"> • C: Parámetro de regularización (0.5, 2, 50) • kernel: Función de transformación ('linear', 'poly', 'rbf', 'sigmoid') • gamma: Coeficiente de kernel {'scale', 'auto'}
XGBoost	<ul style="list-style-type: none"> • n_estimators: Número de árboles (50, 100, 200) • objective: Función objetivo ('binary:logistic', 'binary:logitraw') • eval_metric: Función de pérdida o error para evaluación ('logloss', 'auc')
Perceptrón Multicapa	<ul style="list-style-type: none"> • hidden_layer_sizes: Tamaño de las capas ocultas [(200,200), (100,100)] • learning_rate_init: Tasa de aprendizaje inicial [0.001, 0.01, 0.1] • activation: Función de activación ['tanh', 'relu'] • solver: Solucionar para la optimización del peso ['sgd', 'adam'] • early_stopping: Detener cuando no mejora el resultado [True]

Tabla 8 - Hiperparámetros considerados para cada modelo (elaboración propia)

Los resultados de la optimización se presentan a continuación en la **Tabla 9**:

Modelo	Mejores Hiperparámetros	Accuracy Optimizado	Clase	F1-score
SVM	C: 2, Kernel: 'rbf', gamma: 'scale'	0.87	Detractor	0.88
			Promotor	0.87
XGBoost	n_estimators: 200, Objective: 'binary:logistic', eval_metric: 'logloss'	0.85	Detractor	0.86
			Promotor	0.84
Perceptrón Multicapa	'activation': 'relu', 'early_stopping': True, 'hidden_layer_sizes': (200, 200), 'learning_rate_init': 0.001, 'solver': 'adam'	0.86	Detractor	0.87
			Promotor	0.86

Tabla 9 - Resultados modelos con optimización de hiperparámetros (elaboración propia)

Luego de la optimización de hiperparámetros, el modelo Support Vector Machine (SVM) mantuvo su superioridad, alcanzando una exactitud (*accuracy*) del 87% (0.87), seguido por el Perceptrón Multicapa con el 86% (0.86) y el XGBoost con el 85% (0.85).

5.2. Experimentación con modelos no supervisados

Ahora con el modelo que mejor clasifica los comentarios de los clientes en las clases promotor y detractor, el siguiente paso es aplicar técnicas de agrupamiento, con el propósito de identificar temas recurrentes dentro de estos comentarios con el fin de generar insights valiosos para la toma de decisiones estratégicas en el área de diseño de experiencia.

Para lograr este objetivo se usaron cinco modelos no supervisados. Los primeros tres se basan en la factorización de matrices elegidos a partir del libro "Text Analytics with Python" [16] y los dos restantes enfocados en clusterizar los temas recurrentes seleccionados a partir de la recomendación de expertos en el tema.

Los modelos basados en factorización de matrices son el Non-negative Matrix Factorization (NMF), el Latent Semantic Indexing (LSI) y el Latent Dirichlet Allocation (LDA) y los basados en clusterización son K-means y DBscan. Para todos los modelos se utilizó la representación numérica de texto TF – IDF (Term Frequency-Inverse Document Frequency) debido a que fue la mejor en los modelos supervisados.

- Para los modelos de factorización matricial, se definieron cuatro escenarios basados en diferentes cantidades de tópicos seleccionados con el fin de evaluar la calidad de la agrupación temática: 10, 20 y 100 tópicos. Asimismo, se establecieron tres métricas de

evaluación que permiten medir el desempeño de los modelos y realizar una comparación objetiva entre ellos para seleccionar el más adecuado. Las métricas utilizadas fueron el coeficiente de Silueta, el índice de Davies-Bouldin y el índice de Calinski-Harabasz.

Los modelos de factorización matricial se desarrollan teniendo en cuenta los cuatro escenarios planteados y se obtienen los siguientes resultados:

Escenario	# Tópicos	Modelo	Métricas		
			Silueta	Davies-Bouldin	Calinski-Harabasz
1	10	NMF	0.17	1.11	1306.99
		LSI	0.19	1.43	878.25
		LDA	0.75	0.39	10446.72
2	20	NMF	0.02	1.21	629.97
		LSI	0.03	1.92	326.07
		LDA	0.79	0.56	3452.5
3	100	NMF	-0.05	1.39	111.3
		LSI	-0.09	2.93	46.59
		LDA	0.63	0.57	596.87

Tabla 10 - Resultados métricas modelos factorización matricial (elaboración propia)

De la **Tabla 10 se observa notoriamente que el modelo LDA es el que obtiene las mejores métricas** en todos los escenarios, ya que la Silueta mientras más cercano a 1 da el entendimiento de tópicos claramente definidos y separados entre sí. Por otro lado, el índice de Davies Bouldin mientras más cercano a 0 señala una buena separación de temas y el índice de Calinski Harabasz con valores altos marca grupos compactos y bien separados.

En conjunto, estos resultados permiten concluir que el **Escenario 1 (10 tópicos con LDA)** ofrece el mejor equilibrio entre cohesión interna y separación entre tópicos, lo cual lo posiciona como la opción más robusta y coherente desde una perspectiva cuantitativa.

- Para los modelos de Clusterización, particularmente para K-Means se buscó el mejor número de clústers que podría ajustarse a partir de las gráficas del método del codo. Para esto se generó la visualización inicialmente con 100 clústers, como primera aproximación buscando el punto de inflexión que permitiera determinar el número de clusters ideal.

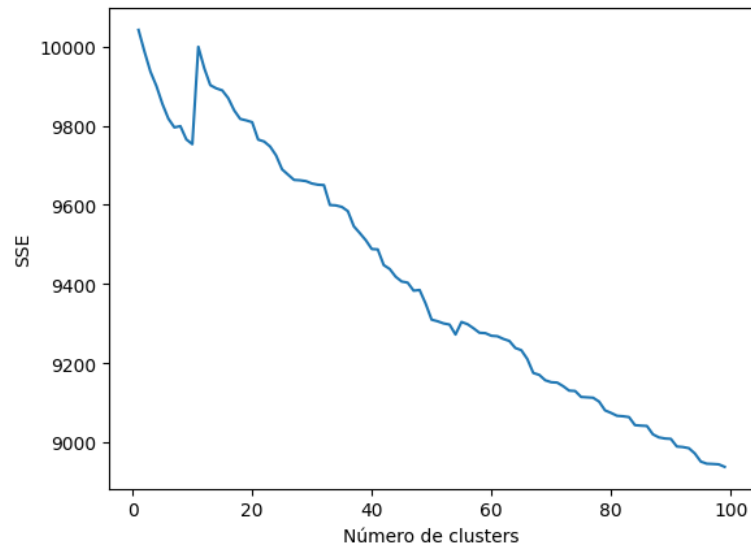


Ilustración 3 - Método del codo K-Means hasta 100 clusters (elaboración propia)

En la **Ilustración 3**, se observa una disminución progresiva del SSE (Suma de los Errores Cuadrados) a medida que se incrementa el número de clusters. Esta tendencia es esperada, ya que un mayor número de clusters permite una segmentación más ajustada de los datos. Sin embargo, el objetivo del método del codo es identificar un punto a partir del cual agregar más clusters no representa una mejora significativa en la reducción del error.

Aunque el gráfico no presenta un codo claramente definido, se identifica una posible inflexión en el rango entre 10 y 20 clusters, donde la pendiente de la curva comienza a suavizarse. A partir de este comportamiento en ese punto se podría inferir, que los beneficios de incrementar el número de clusters en términos de reducción del SSE se vuelven marginales.

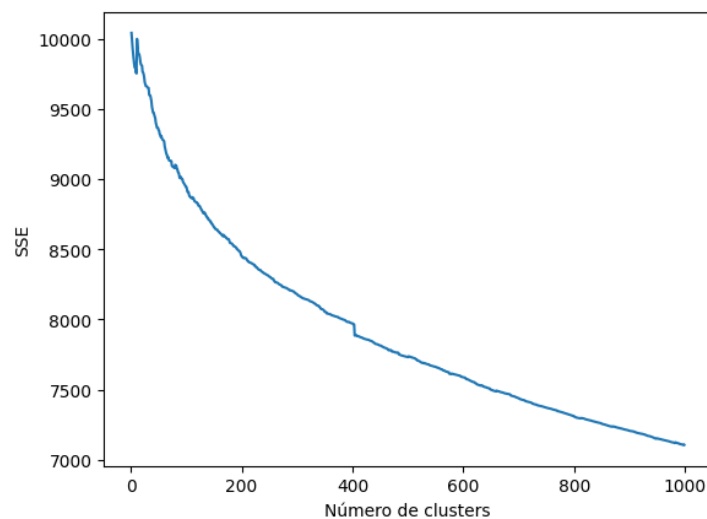


Ilustración 4 - Método del codo K-Means hasta 1000 clusters (elaboración propia)

No obstante, se amplió el análisis hasta 1000 clústers **Ilustración 4** para garantizar que no existiera un punto de inflexión más relevante en rangos superiores. En esta segunda gráfica, se reafirma que la curva desciende de forma cada vez más suave y continua, sin evidencia de un nuevo quiebre abrupto que indique un cambio significativo en la tasa de disminución del SSE.

Adicionalmente, se calculó el Silhouette Score para una amplia gama de valores de k como se muestra en la **Ilustración 5**, como métrica complementaria que evalúa la calidad de la segmentación en términos de cohesión separación entre clusters. Aunque los valores obtenidos fueron en su mayoría bajos o incluso negativos, este comportamiento es esperable en contextos donde los datos presentan alta dimensionalidad y dispersión, como ocurre con representaciones vectoriales tipo TF-IDF.

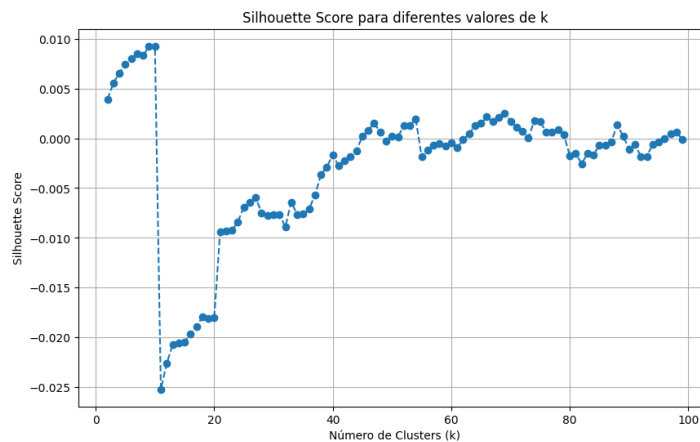


Ilustración 5 - Método Silhouette Score para diferentes valores de k (elaboración propia)

Lo relevante en este análisis fue el comportamiento de la curva, donde se observan **los valores más altos del Silhouette Score (incluso positivos) en el rango de 2 a 10 clusters**, indicando una mejor definición de los grupos en ese intervalo. A partir de $k \approx 10$, se identifica una caída abrupta en la calidad de la segmentación, con los valores más negativos entre $k \approx 10$ y 20 , seguida de una estabilización progresiva con puntajes bajos y sin mejoras significativas más allá de $k > 20$.

A partir de este análisis complementario, se refina la hipótesis planteada inicialmente con el método del codo. Aunque el SSE sugería una inflexión entre 10 y 20 clusters, el Silhouette Score aporta evidencia adicional de que el número óptimo de clusters podría estar más cerca del rango inferior, entre 5 y 10, donde se obtiene el mejor balance entre segmentación y coherencia interna. Esta convergencia metodológica aporta a la decisión de aplicar el modelo K-Means dentro del rango de los 10 y 20 clusters para evaluar la mejor opción de

segmentación bajo las condiciones evaluadas.

Basado en lo anterior, se ejecuta el modelo K-Means para estos dos valores de clústers que se determinaron en el análisis del método del codo.

Escenario	# Tópicos	Modelo	Silueta	Métricas	
				Davies-Bouldin	Calinski-Harabasz
1	10	K-means	0.0092	7.44	33.43
2	20	K-means	-0.018	2.97	12.70

Tabla 11 - Resultados métricas modelos clusterización K-means (elaboración propia)

De la **Tabla 11** se observa ninguno de los 2 escenarios planteados para el modelo K-means **logra formar agrupamientos bien definidos** ni internamente cohesionados, lo cual se ve reflejado en los bajos valores del índice de silueta, la pobre separación reflejada por el índice Davies-Bouldin, y los valores bajos del índice Calinski-Harabasz.

Por tanto, se concluye que **K-Means no es un modelo adecuado** para el objetivo buscado, y se reafirma la pertinencia de utilizar enfoques como **LDA**, que han mostrado un mejor desempeño tanto en métricas cuantitativas como en la calidad interpretativa de los tópicos.

Por otro lado, se experimentó con el modelo DBSCAN, realizando un proceso de búsqueda sistemática para identificar la configuración óptima de sus parámetros. Particularmente, se enfocó en encontrar los mejores valores para eps (radio de vecindad) y min_samples (mínimo de puntos requeridos para formar un clúster).

Para encontrar el valor adecuado del parámetro ϵ (epsilon) en el modelo DBscan, se utilizó el método del codo, que consiste en observar cómo varían las distancias entre los datos. En la **Ilustración 6**, se muestra cómo cambian estas distancias al ir ordenando los puntos. Lo que se busca es un punto donde la curva cambia bruscamente, formando una especie de "codo", ya que ese cambio indica el valor ideal de distancia para agrupar los datos.

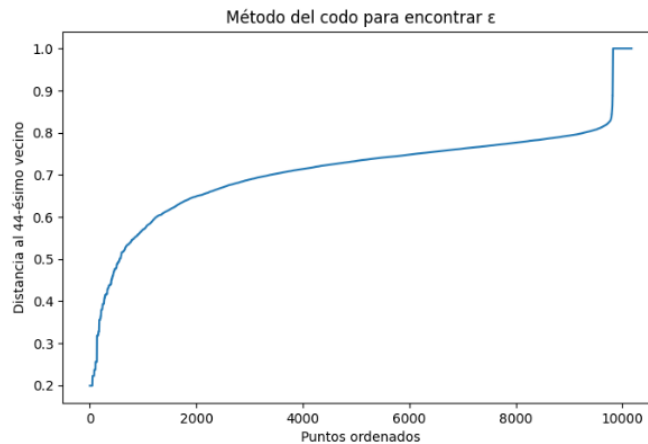


Ilustración 6 - Método del codo DBscan hasta 1000 clusters (elaboración propia)

En este caso, el punto de inflexión se encuentra entre **0.4 y 0.6**. Por lo cual de manera complementaria se buscó el parámetro de mínimo de ejemplos (min_samples) con ambos valores para obtener el mejor resultado. Para esto, para cada valor de eps, se aplicó una búsqueda sistemática sobre una grilla de valores de min_samples, evaluando valores entre 5 y 100, en incrementos de 5. Para cada combinación, se descartaron automáticamente aquellos casos en los que el modelo DBSCAN no lograba formar al menos dos clústeres válidos (por ejemplo, si todos los puntos eran etiquetados como ruido o se generaba un único grupo). Adicionalmente, solo se consideraron combinaciones donde las métricas internas de validación (Silhouette Score, índice de Davies-Bouldin e índice de Calinski-Harabasz) fueran computables y representativas.

A partir de la ejecución de la grilla, los resultados más destacados con eps = 0.4 se muestran en la **Tabla 12**:

Escenario	# Min Samples	Modelo	Métricas		
			Silueta	Davies-Bouldin	Calinski-Harabasz
1	40	DBscan	0.52	0.94	108.35
2	45	DBscan	0.64	0.69	163.53

Tabla 12 - Resultados métricas modelo de clusterización DBScan con eps = 0.4 (elaboración propia)

De la misma manera para $\text{eps}=0,6$, los mejores resultados se muestran en la **Tabla 13**:

Métricas					
Escenario	# Min Samples	Modelo	Silueta	Davies-Bouldin	Calinski-Harabasz
1	90	DBscan	0.16	2.37	52.24
2	95	DBscan	0.18	2.16	67.03

Tabla 13 - Resultados métricas modelo de clusterización DBScan con $\text{eps} = 0.6$ (elaboración propia)

Luego de evaluar los resultados obtenidos con dos configuraciones distintas del parámetro eps (0.4 y 0.6) en el modelo DBscan, se concluye que el mejor desempeño se logra con $\text{eps} = 0.4$ y $\text{min_samples} = 45$. Esta configuración ofrece un equilibrio superior entre la cohesión y separación de los grupos, como lo evidencian las métricas de Silhouette (0.64), Davies-Bouldin (0.69) y Calinski-Harabasz (163.53), que reflejan una estructura de clusters más definida y coherente. En comparación con el modelo K-means, DBscan demuestra un mejor desempeño en este caso.

Posterior a la experimentación descrita con los modelos no supervisados se comparan los dos mejores obtenidos por factorización de matrices y por clusterización en términos del coeficiente de Silueta, el índice de Davies-Bouldin y el índice de Calinski-Harabasz para cada modelo:

Métricas					
Escenario	# Tópicos # Min Samples	Modelo	Silueta	Davies-Bouldin	Calinski-Harabasz
1	10	LDA	0.75	0.39	10446.72
2	45	DBscan	0.64	0.69	163.53

Tabla 14 – Comparación métricas mejores modelos no supervisados (elaboración propia)

Al comparar los resultados de los modelos no supervisados, se observa que LDA presenta un desempeño significativamente superior al modelo DBscan, especialmente en las métricas de Silueta (0.75) y Calinski-Harabasz (10446.72). Esto sugiere que **LDA permite una mejor separación y coherencia interna de los temas identificados**, lo cual es crucial cuando se busca interpretar y etiquetar automáticamente grupos de comentarios con significado temático claro. **Por tanto, se concluye que LDA es el modelo más adecuado para el análisis de tópicos en este caso.**

Tras el ejercicio de experimentación con modelos no supervisados, cuyo objetivo fue identificar temas recurrentes que pudieran ofrecer insights valiosos para la toma de decisiones, se propuso usar un modelo de lenguaje largo (Large Language Model, LLM), con el propósito de asignar un nombre representativo a cada grupo temático mediante un prompt especialmente diseñado para

este fin. En este se proporciona al modelo el identificador del tema principal, la lista de comentarios agrupados, y se le solicita generar un nombre claro y conciso para cada grupo.

Para lograr esto, se utilizó ChatGPT a través del modelo GPT o1 entregando un prompt diseñado especialmente para esta tarea, el cual tiene las siguientes características y secciones clave:

1. **Rol:** Se indica que el sistema debe actuar como un experto en entidades y procesos del sector financiero en Colombia, con una visión integral del cliente.
2. **Contexto:** Se describe la importancia de la voz del cliente como insumo estratégico, el enfoque centrado en la experiencia del cliente y el uso de herramientas como NPS para capturar opiniones. También se menciona que los temas y palabras clave ya fueron identificados mediante análisis no supervisado y que se proporcionan columnas con comentarios, tema dominante y palabras clave.
3. **Tarea:** Se solicitan cinco acciones principales, incluyendo la propuesta de un nombre para cada tema, dos alternativas adicionales, y una justificación breve basada en las palabras clave y comentarios asociados.
4. **Condiciones:** Se definen los requisitos que debe cumplir cada nombre, tales como claridad, profesionalismo, evitar ambigüedades y mantener un tono formal, sin ser excesivamente técnico ni abstracto.
5. **Formato de respuesta esperado:** Se establece una estructura específica de salida con campos definidos para el nombre principal, dos alternativas y la justificación correspondiente.
6. **Criterios adicionales de evaluación:** Se detallan cuatro criterios para validar la calidad de los nombres propuestos: claridad, precisión, relevancia y consistencia en el estilo.
7. **Ejemplos:** Se proporcionan tanto ejemplos adecuados como inadecuados para ilustrar el tipo de nombres esperados.

En conjunto, este prompt permite transformar automáticamente los resultados del modelo LDA en **tópicos etiquetados de forma interpretativa y útil para equipos de análisis o gestión**, facilitando así la priorización y diseño de acciones de mejora sobre la experiencia del cliente.

El prompt construido y utilizado para este ejercicio se puede consultar en la sección Anexos de este documento junto con los resultados detallados. A partir de este ejercicio se llegó a los siguientes resultados:

Resumen de los nombres asignados por GPT o1 y sus palabras claves por t3pico:

T3pico	Palabras Clave del t3pico
0: "Cobros excesivos y tasas de inter3s elevadas"	manejo, alta, cuota, contactaron, tasa, interes, nunca, caro, cobro, ninguna
1: "Falta de contacto y seguimiento del asesor"	nunca, llamaron, respuesta, recib, conmigo, pesimo, comunicaron, dieron, asesor, contacto
2: "Atenci3n al cliente deficiente y sin soluci3n"	poca, lento, cobran, atendieron, interaccion, intuitivo, personal, porq, sirvio, echo
3: "Demoras y tr3mites engorrosos en los procesos"	costosa, mantiene, taza, caida, solucionaron, use, agilidad, aclara, aprueban, complejos
4: "Proceso de cr3dito lento y comunicaci3n insuficiente"	intereses, altos, mal, servicio, interes, alto, excelente, demasiado, costoso, gracias
5: "Desconfianza, inseguridad y percepci3n de fraude"	palabras, amabilidad, resolvio, eficiencia, entendible, feo, senal, mejorando, mintieron, flexibles
6: "Calidad de servicio inconsistente y falta de informaci3n"	funciona, bien, mala, veces, altas, lenta, atencion, app, sirve, falla
7: "Bloqueos y datos desactualizados en canales digitales"	cae, recarga, rechaza, sistema, caido, siempre, estafadores, lector, ultimamente, deficiente
8: "Fallos en la app y problemas con pagos digitales"	mas, credito, proceso, tarjeta, banco, cuenta, nunca, bancolombia, pagar, informacion
9: "Baja rentabilidad y limitada aceptaci3n de tarjetas"	ayudaron, comentarios, rentabilidad, tube, sirven, enganosa, resolvi, esperada, usd, solucionado

Tabla 15 -Resumen nombres asignados por GPT o1 y sus palabras claves por t3pico (elaboraci3n propia).

Una vez aplicado el prompt dise1ado para asignar nombres interpretables y relevantes a los temas identificados por el modelo LDA, se obtuvo un conjunto de 10 t3picos que se alinean con la cantidad de clusters definidos que reflejan con mayor claridad las principales 3reas de insatisfacci3n expresadas por los clientes. Los nombres asignados muestran una correspondencia con las palabras clave y comentarios originales, y cumplen con los criterios de claridad, precisi3n y relevancia exigidos por el contexto organizacional.

Los resultados muestran que los temas abordan problem3ticas recurrentes en distintos frentes del servicio financiero:

- **T3picos 0, 4 y 9** se relacionan directamente con aspectos financieros, como **cobros, tasas, rentabilidad** y **procesos crediticios**, que suelen impactar la percepci3n de valor y transparencia por parte del cliente.
- **T3picos 1, 2 y 6** reflejan **fallas en la interacci3n humana**, incluyendo la **falta de seguimiento por parte de los asesores**, la **ineficiencia en la atenci3n al cliente**, y la **inconsistencia en la calidad del servicio**.

- **Tópicos 3 y 5** agrupan percepciones negativas sobre **trámites operativos complejos y falta de confianza o percepción de riesgo**, factores que pueden incidir fuertemente en la decisión de continuar o abandonar la relación con la entidad.
- **Tópicos 7 y 8** se enfocan en problemas asociados a **canales digitales**, incluyendo **bloqueos, errores en la app y dificultades con pagos**, lo cual evidencia una necesidad de fortalecer la experiencia en los canales no presenciales.

En conjunto, la clasificación generada permite una visión estructurada de los temas críticos desde la perspectiva del cliente, lo que facilita la priorización de planes de mejora. Además, el lenguaje utilizado en los nombres es profesional, accesible y específico, lo que asegura su utilidad operativa para áreas de experiencia de cliente, calidad, riesgo o producto.

Este ejercicio valida la efectividad del prompt como herramienta de apoyo en la etiquetación semántica de tópicos generados por modelos de análisis no supervisado, promoviendo un enfoque más interpretativo y orientado a la acción siempre y cuando se defina de manera certera la condiciones de este.

Como se puede observar, los resultados obtenidos son claros, concretos y estrechamente relacionados con los grupos de comentarios analizados, lo que convierte a las herramientas de inteligencia artificial generativa en un recurso valioso para apoyar y optimizar tareas. Su capacidad para sintetizar y nombrar temas complejos de forma comprensible no solo facilita la interpretación de grandes volúmenes de datos, sino que también potencia la toma de decisiones estratégicas orientadas a la mejora continua.

Estrategia de visualización para el análisis de tópicos

Como parte de la etapa de análisis e interpretación de resultados, se implementó una estrategia de visualización complementaria orientada a facilitar la comprensión de los tópicos temáticos identificados mediante el modelo LDA. El objetivo fue brindar herramientas visuales que permitieran no solo validar los resultados desde una perspectiva analítica, sino también hacerlos accesibles a perfiles no técnicos, promoviendo así una lectura intuitiva y orientada a la toma de decisiones.

Las visualizaciones seleccionadas incluyen los siguientes tipos:

1. Nube de palabras

La **Ilustración 7** presenta gráficamente las palabras más frecuentes dentro de cada tema identificado. El tamaño de cada palabra es proporcional a su frecuencia de aparición en los comentarios clasificados bajo dicho tópico, lo que permite identificar de manera rápida los **conceptos clave y patrones semánticos predominantes**.

Tópico 0 - Cobros excesivos y tasas de interés elevadas



Ilustración 7 - Muestra de nube de palabras para los tópicos identificados (elaboración propia)

Cada nube de palabras está acompañada por la **descripción del tópico asignado mediante el modelo LLM**, lo que facilita una correspondencia directa entre las palabras representadas y el significado temático del grupo. Por ejemplo, en el **Tópico 0 – “Cobros excesivos y tasas de interés elevadas”**, sobresalen términos como *cuota*, *interés*, *manejo*, *tarjeta*, *alta* y *cobran*, los cuales refuerzan la validez del nombre asignado y evidencian la percepción de los clientes frente a **costos financieros considerados elevados**.

Estas visualizaciones resultan especialmente útiles para validar cualitativamente los resultados del modelo y apoyar la interpretación temática desde distintas áreas de la organización.

Las visualizaciones correspondientes a los demás tópicos se encuentran disponibles en la sección Anexos de este documento.

2. Gráfico de burbujas para la distribución de comentarios por tópico

Se utilizó un gráfico de burbujas para representar el **volumen de comentarios asociados a cada tema**. Esta visualización permite identificar los tópicos con mayor peso en términos de recurrencia dentro del conjunto de opiniones.

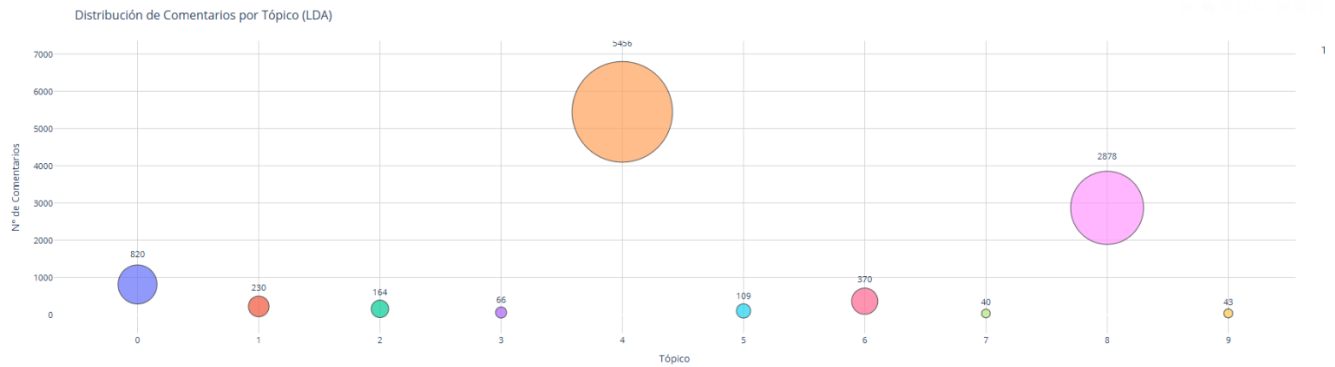


Ilustración 8 - Gráfico de burbujas con distribución de comentarios por tópico (elaboración propia)

En esta **Ilustración 8**, se observa que el **Tópico 4 – “Desinformación en crédito hipotecario”** concentra el mayor número de comentarios (5456), seguido por el **Tópico 8 – “Errores en app para pagos y transferencias”** (2878), lo cual permite priorizar la atención de temáticas con alto impacto para la experiencia del cliente.

3. Mapa de tópicos con agrupación semántica

La **Ilustración 9** presenta una **proyección bidimensional** de los comentarios analizados, mostrando cómo se agrupan en torno a los distintos tópicos identificados por el modelo LDA. Cada área coloreada representa un **clúster semántico diferenciado**, mientras que cada punto dentro de estas áreas corresponde a un **comentario individual** clasificado bajo un tópico específico.

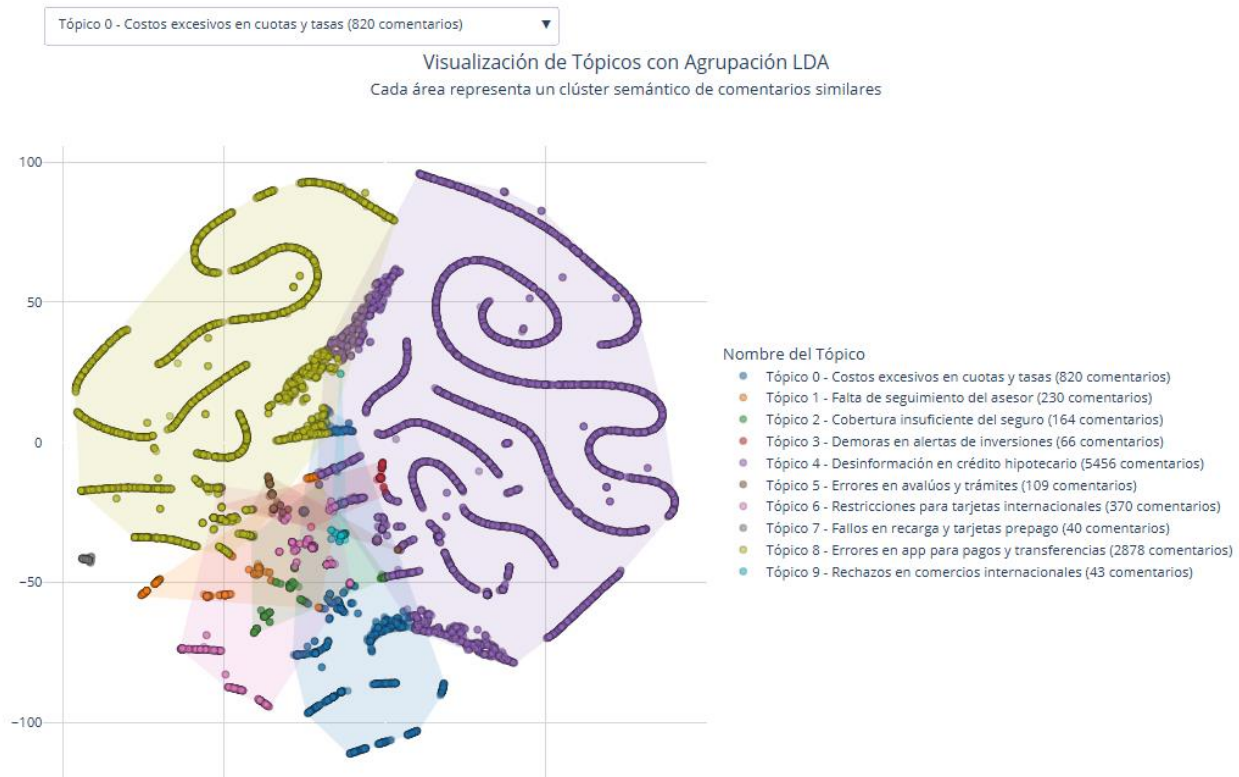


Ilustración 9 - Gráfico bidimensional t-SNE por tópico (elaboración propia)

Gracias a las funcionalidades interactivas ofrecidas por la biblioteca **Plotly** de Python, esta visualización permite **explorar de forma dinámica** los datos. Al posicionarse sobre un punto, es posible consultar el contenido exacto del comentario y verificar su asignación temática, lo cual facilita tanto la **validación del modelo** como la **revisión cualitativa** de los resultados. Adicionalmente, incorpora una **lista desplegable** que permite filtrar y visualizar un tópico en particular o todos en conjunto, favoreciendo el análisis enfocado.

Esta herramienta resulta especialmente útil para **identificar la cohesión interna** de los grupos, **detectar solapamientos o relaciones de cercanía entre tópicos**, y **comprender la distribución y densidad de los comentarios** dentro del espacio semántico proyectado. En conjunto, esta visualización contribuye de forma significativa a **mejorar la interpretabilidad de los resultados** y a **respaldar la toma de decisiones** mediante una comprensión más profunda y visual del contenido expresado por los clientes.

4. Mapa de distancias entre tópicos e identificación de términos salientes

Para complementar el análisis temático realizado mediante el modelo Latent Dirichlet Allocation (LDA), se empleó la librería **pyLDAvis**, una herramienta especializada en la **visualización interactiva**

de modelos de tópicos, ampliamente recomendada en la literatura de analítica de textos. En particular, su uso está fundamentado en las buenas prácticas propuestas por **Dipanjan Sarkar** en su libro *Text Analytics with Python* [16], donde se destaca pyLDavis como una de las herramientas más efectivas para **interpretar y validar de manera visual** los resultados de LDA. En la **Ilustración 10** se presenta este detalle:

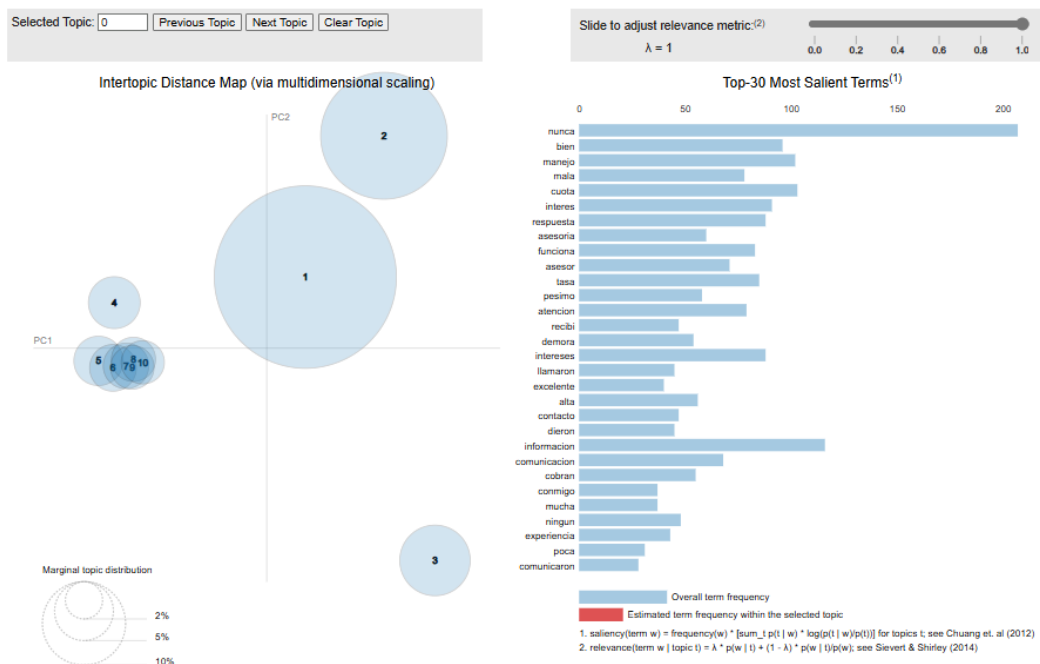


Ilustración 10 - Representación pyLDavis (elaboración propia)

El autor plantea que una de las limitaciones comunes en los modelos de tópicos es la dificultad para comprender la relación entre los temas generados y los términos que los componen. Por ello, recomienda el uso de pyLDavis como una solución práctica para representar gráficamente:

- Las **distancias entre tópicos** en un plano bidimensional, utilizando técnicas de escalamiento multidimensional (MDS), lo que permite observar qué tan separados o superpuestos están los temas (Visualización izquierda).
- Los **términos más relevantes o salientes** para cada tópico, calculados a partir de una métrica de relevancia ajustable que combina frecuencia y especificidad, permitiendo una exploración detallada del contenido semántico de cada clúster temático (Visualización derecha).

El script utilizado en este proyecto replica el flujo sugerido por Sarkar, donde se construyen los insumos requeridos para `pyLDavis.prepare()` —incluyendo las distribuciones de temas por

documento, la longitud de los textos, el vocabulario y la frecuencia de términos— y posteriormente se despliega la visualización de forma interactiva en un entorno como Google Colab, Jupiter Notebook o Streamlit.

La integración de esta herramienta permitió enriquecer el análisis al confirmar visualmente la **validez de la separación temática**, identificar relaciones o solapamientos entre tópicos y explorar los términos dominantes con mayor precisión semántica.

En línea con las recomendaciones del autor, la aplicación de pyLDavis en este proyecto no solo cumple una función técnica, sino que también facilita la **comunicación de hallazgos complejos a públicos no especializados**, consolidando así su pertinencia tanto desde el punto de vista metodológico como operativo.

Consolidación de visualizaciones y resultados en un tablero interactivo

Como resultado de la estrategia de análisis y visualización de tópicos descrita en los capítulos anteriores, se desarrolló un tablero interactivo que reúne en un solo entorno todas las visualizaciones clave generadas a partir del modelo LDA. Este tablero fue diseñado con el objetivo de facilitar el acceso, exploración y comprensión de los resultados por parte de usuarios no técnicos, evitando así la necesidad de navegar directamente por notebooks en Google Colab u otros entornos de programación, los cuales, si bien son útiles en la etapa de desarrollo, pueden resultar poco prácticos para el usuario final.

Previamente, se evaluaron distintas alternativas tecnológicas como Power BI y entornos web desarrollados desde cero con Python. Sin embargo, tras considerar criterios como flexibilidad, tiempo de desarrollo, facilidad de integración con librerías gráficas (como Plotly) y experiencia de usuario, se optó por utilizar Streamlit como solución central para la construcción del tablero. Esta herramienta permite crear aplicaciones web de forma rápida y sencilla, conservando el dinamismo e interactividad propios de las visualizaciones generadas en Python, y facilitando su despliegue tanto en un host local como en una plataforma online accesible desde distintos dispositivos y perfiles de usuario.

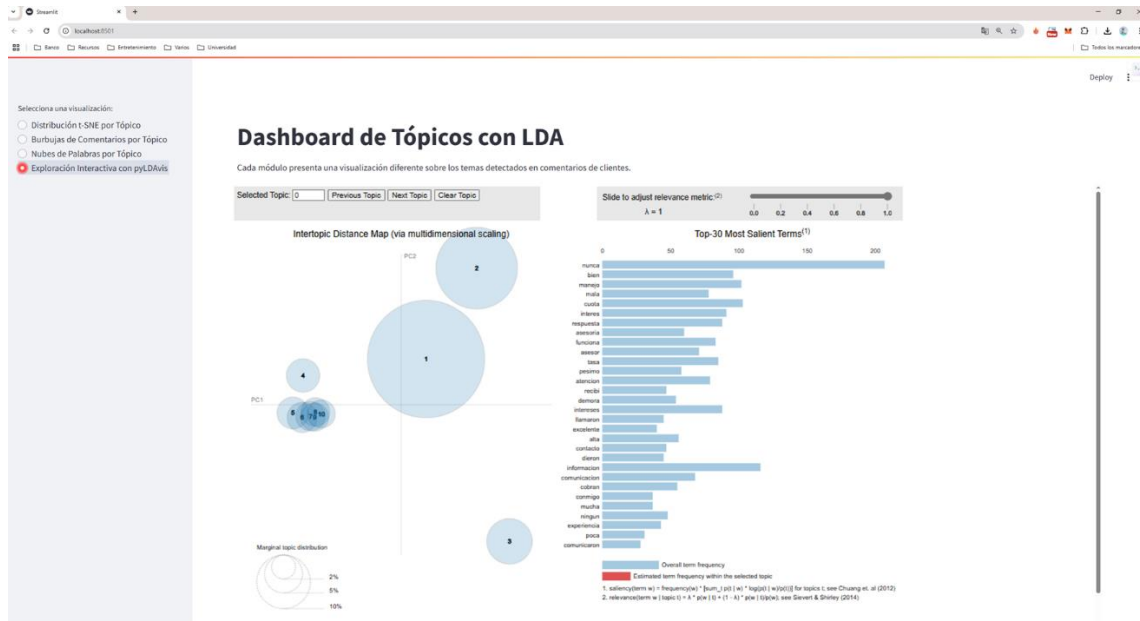


Ilustración 11 - Visualización del tablero interactivo con los resultados obtenidos (elaboración propia)

El tablero consolidado **Ilustración 11** incluye:

- Visualización de la distribución de comentarios por tópico mediante burbujas escaladas.
- Nubes de palabras que resumen de manera gráfica los términos más representativos por tema.
- Un mapa semántico que proyecta los comentarios agrupados por tópico en un espacio bidimensional.
- Y una sección de exploración avanzada mediante pyLDAvis, que combina análisis visual de distancias temáticas con identificación de términos salientes, ajustables mediante una métrica de relevancia.

Gracias a esta integración, se logró construir una interfaz limpia, dinámica y funcional, orientada a facilitar el entendimiento de los hallazgos sin sacrificar profundidad analítica. La implementación en Streamlit permitió también mantener el vínculo directo con los objetos del modelo y sus salidas, lo que facilita la actualización del tablero ante modificaciones y nuevos entrenamientos del modelo, promoviendo su uso práctico en procesos de seguimiento y toma de decisiones en tiempo real.

6. CONCLUSIONES Y TRABAJOS FUTUROS

6.1. Conclusiones

- Se logró construir un modelo de ciencia de datos basado en técnicas de aprendizaje supervisadas que permite clasificar automáticamente los comentarios de los clientes del segmento personas en BancaLibre. Esto representa un avance significativo en términos de eficiencia operativa, ya que reduce la carga manual sobre el área de experiencia del cliente y agiliza la gestión de solicitudes recurrentes. Esta automatización no solo optimiza recursos, sino que también mejora la capacidad de respuesta del banco frente a las necesidades de sus usuarios.
- Mediante la implementación de modelos no supervisados, se identificaron temas frecuentes asociados con la insatisfacción de los clientes (detractores), permitiendo al banco entender con mayor claridad los factores que afectan la experiencia del usuario. Este hallazgo habilita una toma de decisiones más informada y proactiva, orientada a resolver problemas estructurales y mejorar los servicios.
- La incorporación de herramientas de inteligencia artificial generativa, como ChatGPT aportó un valor significativo al proceso analítico del proyecto. Esta integración permitió traducir rápidamente los resultados del modelo LDA en tópicos claros, interpretativos y directamente accionables para las áreas de experiencia de cliente, producto y calidad. Lejos de reemplazar la intervención humana, esta tecnología potenció el análisis al ofrecer una visión más comprensible y estructurada de los temas críticos desde la voz del cliente. Esto facilitará la priorización de acciones de mejora, optimizando el tiempo de análisis y fortaleciendo la toma de decisiones estratégicas.
- La conversión de los resultados numéricos en visualizaciones interactivas permitió transformar datos complejos en información clara, accesible y accionable para los equipos de negocio. Herramientas como gráficos de burbujas, mapas semánticos y pyLDavis facilitaron la interpretación de los temas identificados, promoviendo decisiones más rápidas y fundamentadas. El desarrollo de un tablero interactivo en Streamlit consolidó estos insumos en una interfaz amigable, útil tanto para perfiles técnicos como no técnicos. Esta estrategia visual no solo mejora la comprensión de los hallazgos, sino que multiplica el valor del análisis al integrarlo directamente en los procesos de mejora continua de la organización.
- Este proyecto confirma que la ciencia de datos va más allá de la automatización, permite descubrir, estructurar y aprovechar información oculta en grandes volúmenes de datos no estructurados. Al combinar técnicas de análisis, visualización e inteligencia artificial generativa, se logró convertir los comentarios de los clientes en conocimiento operativo útil. El uso responsable y bien diseñado de estas herramientas se consolidó como un aliado clave para la toma de decisiones estratégicas, demostrando el impacto tangible que puede generar la ciencia de datos cuando se orienta a resolver desafíos reales de negocio.

6.2. Trabajos futuros

- Incorporar modelos avanzados de análisis de sentimientos en tiempo real, integrados directamente a los canales digitales del banco (app, página web, redes sociales). Esto permitiría no solo clasificar y entender los comentarios de los clientes, sino también detectar cambios emocionales en sus mensajes para activar alertas tempranas o recomendaciones automatizadas. De esta manera, BancaLibre podría fortalecer su capacidad de respuesta proactiva y elevar su nivel de personalización en la experiencia del cliente.
- Construir un sistema que no solo identifique los temas recurrentes en los comentarios de los clientes, sino que permita monitorear su evolución a lo largo del tiempo, detectando si ciertos tópicos crecen, disminuyen o se transforman. Esta línea de trabajo permitiría a BancaLibre anticiparse a posibles crisis o necesidades emergentes, midiendo el impacto de las acciones implementadas por la organización. Además, al identificar nuevas preocupaciones de los clientes de forma temprana, se fortalece la capacidad del banco para diseñar estrategias reactivas y preventivas más efectivas, alineadas con su compromiso de mejorar continuamente la experiencia del cliente.
- Replicar este proyecto a otras bases de datos de comentarios gestionadas por el área de experiencia del cliente, como formularios web, chats de atención o redes sociales. Esto con la intención de evaluar según la fuente de información, la aplicabilidad de un modelo supervisado de clasificación o si basta un enfoque no supervisado para extraer insights relevantes. Esta replicabilidad permitiría obtener una visión más completa del panorama de atención al cliente, identificar patrones comunes o diferenciados por canal, y priorizar acciones de mejora de forma más estratégica y segmentada.
- Integrar los resultados obtenidos a partir de modelos supervisados y no supervisados en tableros de control dinámicos o herramientas especializadas, con el fin de facilitar la trazabilidad de la experiencia individual del cliente (uno a uno) y habilitar la toma de decisiones anticipadas. Esta línea de trabajo permitiría a BancaLibre no solo visualizar patrones agregados, sino también monitorear interacciones específicas, identificar señales tempranas de insatisfacción y activar alertas que contribuyan a la retención de clientes.

7. REFERENCIAS BIBLIOGRÁFICAS

- [1] ORACLE, «¿Qué es la experiencia del cliente?,» ORACLE, [En línea]. Available: <https://www.oracle.com/co/cx/what-is-cx/>. [Último acceso: 2024 06 15].
- [2] A. M. Morte Serrano, *Net Promoter Score: Estudio, análisis y aplicación práctica*, Zaragoza, 2015.
- [3] A. Baquero, *Net Promoter Score (NPS) and Customer Satisfaction: Relationship and Efficient Management*, La Rioja, 2022.
- [4] J. Berman, *Principles and Practice of Big Data (Second Edition)*, Academic Press, 2018.
- [5] J. P. Pajaro Hernandez, *¿Qué es el Procesamiento de lenguaje natural (NLP)?*, Bogotá D.C., 2018.
- [6] S. P. Castillo Landinez y P. E. Caicedo Rodríguez, *Análisis de sentimientos, una herramienta para valorar la actitud del estudiante frente a un curso*, Bogotá D.C., 2019.
- [7] J. Cabanelas Omil, *Inteligencia artificial ¿Dr. Jekyll o Mr. Hyde?*, Guadalajara, 2019.
- [8] A. D. Jiménez Alfaro, «¿Qué es el machine learning?,» Medellín, 2022.
- [9] C. A. Herrera Rojas, *Análisis de textos utilizando técnicas de NLP*, Santiago, 2024.
- [10] M. Chajia y N. El Habib, *Customer Churn Prediction Approach Based on LLM Embeddings and Logistic Regression*, Fez, 2024.
- [11] G. Valenzuela González, *Aprendizaje Supervisado: Métodos, Propiedades y Aplicaciones*, Málaga, 2022.
- [12] B. C. Barbona Ivana, *Método de clasificación supervisada support vector machine: una aplicación a la clasificación automática de textos*, Rosario, 2018.
- [13] J. J. Espinosa-Zúñiga, *Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito*, Ciudad de México: Universidad Nacional Autónoma de México, Facultad de Ingeniería, 2020.
- [14] P. M. Castillo Castañeda, *Aplicación de aprendizaje automático para la predicción de clientes potenciales en procesos de mercadotecnia*, Zacatecas: Centro de Investigación en Matemáticas, A.C., 2017.
- [15] D. J. Matich, *Redes Neuronales: Conceptos Básicos y Aplicaciones*, Rosario: Universidad Tecnológica Nacional, Facultad Regional Rosario, 2001.
- [16] D. Sarkar, *Text Analytics with Python: A Practitioner's Guide to Natural Language Processing*, Berkeley: Apress, 2019.
- [17] A. Calvo Madurga, *Análisis de sentimientos y emociones en redes sociales usando ML*, Valladolid, 2020.
- [18] E. I. Cruces Cabezas, *El análisis de sentimientos para la gestión de redes sociales en el turismo*, Málaga, España, 2019.
- [19] A. Pascal, S. Trossero y G. Rosenbrock, *Técnicas de análisis de sentimientos aplicadas A la valoración de opiniones en el lenguaje español*, Salta, Argentina, 2021.

- [20] J. S. Arbeláez Trujillo, *Aplicación de técnicas de machine learning para análisis de interacciones en twitter: caso de estudio - gestión de fallas operativas para el Banco Davivienda y el neo Banco Daviplata*, Bogotá D.C., Colombia, 2023.
- [21] E. G. Gómez Hernández, *Análisis predictivo en Twitter para detectar patrones de personas con tendencia Hacktivista aplicando Big Data, Machine Learning y Deep Learning.*, Aguascalientes, México, 2022.

8. ANEXOS

Anexo 1 Detalle del Prompt y resultados

PROMPT:

Rol:

Actúa como un experto en entidades y procesos del sector financiero en Colombia, con una visión integral de cliente.

Contexto:

Para una entidad financiera líder en Colombia, es crucial escuchar la voz del cliente como un recurso esencial para impulsar estrategias de mejora continua. La organización adopta un enfoque centrado en el cliente para enriquecer su experiencia, lo que resulta en una mayor fidelización y satisfacción.

La entidad cuenta con equipos dedicados a la experiencia del cliente cuya misión es comprender necesidades, diseñar y optimizar interacciones, y evaluar la satisfacción. Una responsabilidad clave es la gestión del feedback de los clientes, recopilado a través de herramientas como grupos focales, entrevistas y Net Promoter Score (NPS).

Existe la necesidad de comprender detalladamente las opiniones de los clientes del segmento Personas, específicamente para clasificar y nombrar de manera efectiva los temas recurrentes en las opiniones negativas. Este análisis facilitará la generación de planes de acción específicos para mejorar la experiencia del cliente.

Los comentarios han sido analizados mediante modelos no supervisados, encontrando que el modelo Latent Dirichlet Allocation (LDA) ha identificado 10 temas clave con sus respectivas palabras relacionadas. Los resultados de este análisis se presentan en el archivo adjunto en donde se incluyen las siguientes columnas:

- Comentario – Lista de todos los comentarios reflejando la experiencia real del cliente.
- Tema dominante – Identificador del tema principal.
- Palabras clave del tema - Lista de palabras clave relacionadas con el tema.

Tarea:

1. A partir de los temas y palabras clave identificadas en la columna "Tema dominante" y los comentarios de la columna "Comentario", asigna un nombre claro, conciso y descriptivo para cada tema.
 2. El nombre debe reflejar el contenido del tema y relacionarse tanto con las palabras clave como con los comentarios asociados.
 3. El nombre debe ser fácilmente comprensible para un rol de analista o auxiliar dentro de la organización.
 4. Si el nombre propuesto no es satisfactorio, proporciona dos alternativas para cada tema.
 5. Incluye una breve justificación para cada nombre propuesto, explicando por qué es relevante y cómo refleja el contenido del tema.
-

Condiciones:

El nombre asignado al tema debe:

- Ser claro y conciso.
 - Reflejar el contenido del tema y relacionarse con las palabras clave y los comentarios.
 - No estar compuesto por una sola palabra o palabras aisladas.
 - Evitar términos ambiguos o vagos.
 - Tener un tono formal o profesional.
 - Ser fácil de entender para un analista o auxiliar sin necesidad de contexto adicional.
-

Formato de respuesta esperado:

1. Nombre propuesto: [Nombre principal]
 - Alternativa 1: [Nombre alternativo 1]
 - Alternativa 2: [Nombre alternativo 2]

- Justificación: Explica brevemente por qué el nombre refleja adecuadamente el contenido del tema y está alineado con las palabras clave y comentarios.

Criterios adicionales para evaluar el nombre:

- Claridad: El nombre debe ser fácil de entender para un analista o auxiliar.
- Precisión: El nombre debe describir el contenido del tema de manera directa y sin ambigüedades.
- Relevancia: El nombre debe estar alineado con las palabras clave y el contexto del feedback de los clientes.
- Consistencia: El estilo y estructura de los nombres deben mantenerse uniformes para facilitar su análisis y uso interno.

Ejemplos:

- ✓ Altas tasas de interés y costos financieros
- ✓ Problemas técnicos y demoras
- ✗ Interés (No es adecuado porque es ambiguo y poco claro)
- ✗ Incidencia por latencia en el canal transaccional (Uso excesivo de términos técnicos o jerga interna)
- ✗ Problemas generales (Uso de términos abstractos o genéricos)
- ✗ Problemas frecuentes relacionados con las tasas de interés y los tiempos de respuesta en las plataformas digitales de la entidad (Frasas demasiado largas o complejas).

RESPUESTA:

Tema dominante 0

1. **Nombre propuesto:** Cobros excesivos y tasas de interés elevadas
 - **Alternativa 1:** Altas tasas y costos financieros

- **Alternativa 2:** Intereses y cuotas superiores al mercado
- **Justificación:** Las palabras clave “interes, alta, tasa, cobran, cobro, costoso” y comentarios como “excesivo intereses” y “cobro de manejo” indican que los clientes perciben sobrecostos en intereses y cargos, por lo que el nombre resume con precisión esa preocupación.

Tema dominante 1

1. **Nombre propuesto:** Falta de contacto y seguimiento del asesor

- **Alternativa 1:** Ausencia de comunicación con el cliente
- **Alternativa 2:** Inexistencia de seguimiento postsolicitud
- **Justificación:** Términos como “nunca, llamaron, contacto, asesor” y comentarios que reiteran “nunca llamaron” reflejan carencia de acompañamiento; el nombre describe de forma clara esa brecha de servicio.

Tema dominante 2

1. **Nombre propuesto:** Atención al cliente deficiente y sin solución

- **Alternativa 1:** Respuestas tardías y falta de asesoría
- **Alternativa 2:** Servicio al cliente ineficiente
- **Justificación:** Palabras clave “pesimo, nunca respondieron, asesoría” y frases como “atendieron ignoraron” evidencian insatisfacción con la calidad y efectividad de la atención; el nombre destaca tanto la deficiencia como la ausencia de solución.

Tema dominante 3

1. **Nombre propuesto:** Demoras y trámites engorrosos en los procesos

- **Alternativa 1:** Lentitud y burocracia en la gestión
- **Alternativa 2:** Retrasos significativos en desembolsos y gestiones
- **Justificación:** El énfasis en “demora, mucha, burocracia” y comentarios sobre retrasos en desembolsos respaldan un nombre que combine la demora y la complejidad del trámite.

Tema dominante 4

1. **Nombre propuesto:** Proceso de crédito lento y comunicación insuficiente
 - **Alternativa 1:** Dificultades y retrasos en trámites de crédito
 - **Alternativa 2:** Ineficiencias en la gestión de productos de crédito
 - **Justificación:** Palabras como “credito, proceso, tiempo, respuesta” y quejas sobre falta de información evidencian que el núcleo del tema es la lentitud y poca claridad en la gestión de créditos.

Tema dominante 5

1. **Nombre propuesto:** Desconfianza, inseguridad y percepción de fraude
 - **Alternativa 1:** Sensación de engaño y falta de seguridad
 - **Alternativa 2:** Temor a fraude y baja credibilidad
 - **Justificación:** Términos “ladrones, inseguro, fraude” y comentarios que califican a la entidad de “ladrones” muestran una fuerte percepción de riesgo y falta de confianza; el nombre agrupa estas ideas de forma directa.

Tema dominante 6

1. **Nombre propuesto:** Calidad de servicio inconsistente y falta de información
 - **Alternativa 1:** Experiencia de atención irregular
 - **Alternativa 2:** Información insuficiente y servicio mejorable
 - **Justificación:** La mezcla de palabras “bien, mala, funciona, respuesta, información” refleja valoraciones dispares y carencia de datos claros; el nombre destaca la inconsistencia y la necesidad de mejorar la información al cliente.

Tema dominante 7

1. **Nombre propuesto:** Bloqueos y datos desactualizados en canales digitales
 - **Alternativa 1:** Información no confiable y bloqueos de tarjeta
 - **Alternativa 2:** Inexactitud de valores y restricciones en línea
 - **Justificación:** Palabras clave “reales, valores, viaje, bloquean, spam” y quejas sobre

bloqueos al viajar y valores incorrectos indican problemas de actualización y restricciones digitales; el nombre los resume claramente.

Tema dominante 8

1. **Nombre propuesto:** Fallos en la app y problemas con pagos digitales

- **Alternativa 1:** Errores frecuentes en la aplicación móvil
- **Alternativa 2:** Inestabilidad de la app y transacciones fallidas
- **Justificación:** Con términos como “app, aplicación, pagar, funciona, pagos” y comentarios sobre dificultades para usar la billetera móvil, el nombre refleja la raíz técnica y el impacto en pagos.

Tema dominante 9

1. **Nombre propuesto:** Baja rentabilidad y limitada aceptación de tarjetas

- **Alternativa 1:** Rendimientos insuficientes y problemas de aceptación
- **Alternativa 2:** Inversión poco rentable y cobertura limitada de tarjetas
- **Justificación:** La presencia de “rentabilidad, american express, descontento” junto con quejas sobre bajos rendimientos y rechazo de American Express señala dos frustraciones ligadas: poca ganancia y cobertura restringida; el nombre integra ambas.

Tópico 8 - Fallos en la app y problemas con pagos digitales



Tópico 9 - Baja rentabilidad y limitada aceptación de tarjetas

