



UNIVERSIDAD DISTRITAL  
FRANCISCO JOSÉ DE CALDAS

**BBVA**  
Data & Analytics

# Extracción de Insights Clave para la Toma de Decisiones a partir de Comentarios Negativos de Detractores del Banco BBVA

Trabajo de Grado Modalidad Pasantía

Wilson Eduardo Jerez Hernández

**Director:** Luis Alejandro Masmela Caíta

Universidad Distrital Francisco José de Caldas

Diciembre 2024

# Contenido

# Resumen

Este trabajo presenta el desarrollo de un modelo de análisis de sentimientos basado en técnicas de **Machine Learning**, con el objetivo de clasificar y obtener *insights* a partir de comentarios negativos emitidos por los detractores del Banco BBVA. El propósito principal es identificar áreas de mejora en los servicios del banco para aumentar la satisfacción del cliente y reducir el número de detractores. Para lograrlo, se emplea **procesamiento de lenguaje natural (NLP)** sobre datos textuales obtenidos de redes sociales y encuestas internas.

# Palabras Clave

- Machine Learning
- Análisis de Sentimiento
- NLP
- BBVA
- Clientes Detractores

# Agradecimientos

Quiero agradecer profundamente a mi familia por su amor, paciencia y apoyo incondicional en cada paso de este camino.

A mis docentes, especialmente a mi director Alejandro Masmela, por su orientación y confianza durante todo este proceso.

Al semillero IPREA, gracias por darme la oportunidad de aprender y crecer en lo que más me apasiona.

A mi Sangha, por ser siempre una fuente de tranquilidad, sabiduría y luz en mi vida.

A John Calvo, gracias por tu guía, paciencia y enseñanzas. Tu apoyo fue clave para lograr el equilibrio entre mi crecimiento académico y personal.

Este trabajo es fruto del esfuerzo de todos, y les estaré siempre agradecido.

# Introducción

El sector bancario ha experimentado transformaciones profundas debido a la creciente digitalización, la evolución en las expectativas de los clientes y una competencia cada vez más feroz. En este entorno, la capacidad de captar y analizar la retroalimentación de los clientes se ha vuelto esencial para las entidades bancarias que desean adaptarse rápidamente a las demandas del mercado.

Los comentarios negativos, en particular, representan una fuente valiosa de información, ya que pueden revelar problemas recurrentes o áreas de oportunidad que, si se abordan correctamente, podrían traducirse en una mayor satisfacción y fidelización de los clientes.

## Introducción (cont.)

En este contexto, **BBVA Colombia** ha tomado la iniciativa de implementar un modelo de análisis de sentimientos utilizando técnicas avanzadas de *Machine Learning*. Este modelo se enfoca en la clasificación y análisis de los comentarios emitidos por los **detractores**, aquellos clientes que han expresado insatisfacción o frustración con los servicios ofrecidos.

El objetivo principal de este enfoque es identificar de manera eficiente las áreas críticas de mejora, permitiendo a BBVA Colombia realizar ajustes estratégicos que mejoren la experiencia del cliente y reduzcan el índice de detractores.

## Introducción (cont.)

Para lograrlo, se emplea el **procesamiento de lenguaje natural (NLP)**, una rama de la inteligencia artificial que permite analizar y comprender grandes volúmenes de datos textuales de manera automatizada.

Las fuentes de estos datos incluyen **redes sociales**, donde los clientes expresan sus opiniones de forma pública, y **encuestas internas** del banco, que proporcionan una visión más detallada y directa de las experiencias de los usuarios.

A través de este análisis, BBVA Colombia busca convertir los comentarios negativos en oportunidades de mejora, fortaleciendo así su oferta de valor y consolidando su posición en un mercado altamente competitivo.



# Objetivo General

Desarrollar un modelo de análisis de sentimiento basado en técnicas de *Machine Learning* que permita extraer **insights clave** a partir de los comentarios negativos emitidos por los detractores del Banco BBVA Colombia, con el fin de proporcionar información estratégica que facilite la toma de decisiones en la alta gerencia y contribuya a la mejora continua de los servicios ofrecidos.

# Objetivos Específicos

- ➊ **Transformación de comentarios en representaciones vectoriales:** Implementar el modelo **TfidfVectorizer** para convertir los comentarios negativos en representaciones vectoriales, identificando palabras y términos clave relacionados con áreas de insatisfacción.
- ➋ **Identificación de temas mediante Topic Modeling (LDA):** Aplicar técnicas de **Latent Dirichlet Allocation (LDA)** para agrupar comentarios en tópicos relevantes.

## Objetivos Específicos (cont.)

- 3 **Evaluación de la coherencia del modelo:** Calcular valores de coherencia para distintos números de temas, garantizando que los temas extraídos sean consistentes.
- 4 **Análisis de áreas críticas de insatisfacción:** Realizar un análisis profundo para determinar áreas con mayores niveles de insatisfacción y proporcionar recomendaciones claras.

## Objetivos Específicos (cont.)

- 5 **Visualización de resultados:** Generar visualizaciones que muestren la distribución de los temas, facilitando la comprensión y comunicación de los resultados.
- 6 **Soporte para la toma de decisiones estratégicas:** Proporcionar información clave para orientar acciones correctivas y mejorar la experiencia del cliente.

# BBVA Colombia: Una Entidad Financiera Líder

BBVA Colombia es una destacada institución bancaria que forma parte del **Grupo BBVA**, uno de los conglomerados financieros más grandes del mundo. A lo largo de su historia, la entidad ha experimentado una notable evolución, marcada por adquisiciones estratégicas, innovaciones tecnológicas y un fuerte compromiso con el desarrollo sostenible [2].

# Historia y Fundación

- Fundada en 1956 como **Banco Ganadero**, enfocada en el desarrollo agropecuario en Colombia.
- En 1996, el **Banco Bilbao Vizcaya (BBV)** adquirió el 34,7 % de las acciones, iniciando su expansión en el país [2].

# Evolución y Fusión

- En 1998, BBV incrementó su participación adquiriendo un 15 % adicional de acciones.
- La entidad pasó a llamarse **BBV Banco Ganadero**.
- En 1999, la fusión entre BBV y Argentaria dio lugar a **BBVA**.
- En 2004, se renombró formalmente como **BBVA Colombia** [2].

# Adquisición de Granahorrar y Consolidación

- En 2005, BBVA Colombia adquirió **Granahorrar**, entidad enfocada en el mercado hipotecario.
- En 2006, se fusionaron bajo la marca BBVA Colombia.
- Consolidó su liderazgo en el sector hipotecario y amplió su presencia en el país [2].



# Innovación y Transformación Digital

- Pionera en la transformación digital del sector bancario en Colombia.
- Promueve el uso de canales digitales para mejorar la experiencia del cliente.
- Campaña **Uga Uga** para impulsar el cambio hacia la banca digital [3].

# Inversión en Tecnología y Desarrollo Sostenible

- En 2023, inversión récord de más de \$235.000 millones en tecnología.
- Renovación de aplicaciones móviles y expansión de productos digitales.
- Crecimiento del 98 % en financiación sostenible, movilizando \$6,7 billones [4].

# Análisis de Texto y Modelado de Tópicos

El análisis de texto y el modelado de tópicos son herramientas clave para comprender las preocupaciones de los clientes detractores. Este marco teórico describe los conceptos principales utilizados en este estudio, enfocándose en cómo estas técnicas permiten identificar áreas de insatisfacción y mejorar la toma de decisiones estratégicas.

# Técnicas de Análisis de Texto

- **TF-IDF Vectorizer:** Transforma comentarios en representaciones numéricas, ponderando la importancia de las palabras.
- **Stemming:** Reduce las palabras a su raíz o forma base.
- **Lematización:** Transforma palabras en su forma base considerando contexto y gramática [8].

# Modelado de Tópicos con LDA

**Latent Dirichlet Allocation (LDA)** es una técnica que:

- Identifica temas más comunes en un conjunto de documentos.
- Asume que cada documento es una mezcla de varios temas.
- Cada tema está compuesto por un conjunto de palabras clave [9].

# Evaluación de Coherencia de Modelos

- Garantiza que los temas identificados sean útiles y consistentes.
- Mide la interpretabilidad de los temas.
- Asegura que las palabras agrupadas tengan significado coherente [10].

# Visualización de Resultados y Toma de Decisiones

- Genera visualizaciones que muestran la distribución de temas.
- Facilita la comprensión de la prevalencia de cada tema.
- Ayuda en la toma de decisiones estratégicas para mejorar la experiencia del cliente.

# Descripción de los Datos

- Conjunto de **398 registros** de quejas de clientes de una entidad bancaria ficticia [7].
- Cada registro incluye detalles de la interacción, queja específica, tiempo de resolución y medida de satisfacción.
- Quejas asociadas a productos o servicios específicos.
- Indicadores clave como *Days\_To\_Resolve* y *NPS Response* (Net Promoter Score) [5].
- Descripciones textuales utilizadas para análisis de texto.



# Descripción del Modelo

El desarrollo del modelo se realizó utilizando **Python** y herramientas de procesamiento de lenguaje natural como:

- **NLTK**: Biblioteca para trabajar con texto en Python.
- **spaCy**: Biblioteca para procesamiento avanzado de lenguaje natural.
- **scikit-learn**: Biblioteca para aprendizaje automático en Python.
- **Gensim**: Biblioteca para modelado de temas y similitud de documentos.

# Proceso de Implementación

## 1 Limpieza de Datos:

- ▶ Corrección ortográfica y gramatical.
- ▶ Unificación de idioma.
- ▶ Eliminación de ruido (hashtags, emoticonos, etc.).
- ▶ Normalización del texto.

# Proceso de Implementación (cont.)

## 2 Preprocesamiento del Texto:

- ▶ Eliminación de *stopwords* (palabras comunes sin significado relevante).
- ▶ Lematización (reducir palabras a su forma base).
- ▶ Stemming (reducir palabras a su raíz).

# Proceso de Implementación (cont.)

## 3 Creación del Corpus:

- ▶ Tokenización de comentarios (dividir en palabras individuales).
- ▶ Generación del diccionario de términos.

## 4 Aplicación de TF-IDF:

- ▶ Transformación de comentarios en representaciones vectoriales.
- ▶ Ponderación de palabras según su importancia.

# Proceso de Implementación (cont.)

## 5 Modelado de Tópicos con LDA:

- ▶ Entrenamiento del modelo LDA con diferentes números de temas.
- ▶ Evaluación de la coherencia para seleccionar el número óptimo de temas.

## 6 Visualización de Resultados:

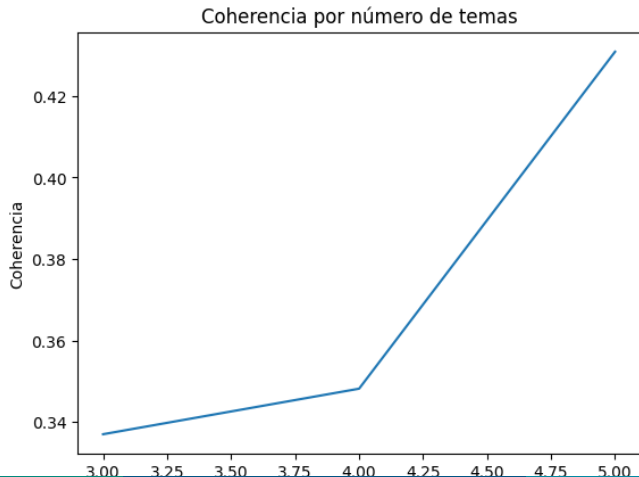
- ▶ Generación de gráficos de coherencia.
- ▶ Visualización de la distribución de temas.

# Pseudocódigo del Modelo

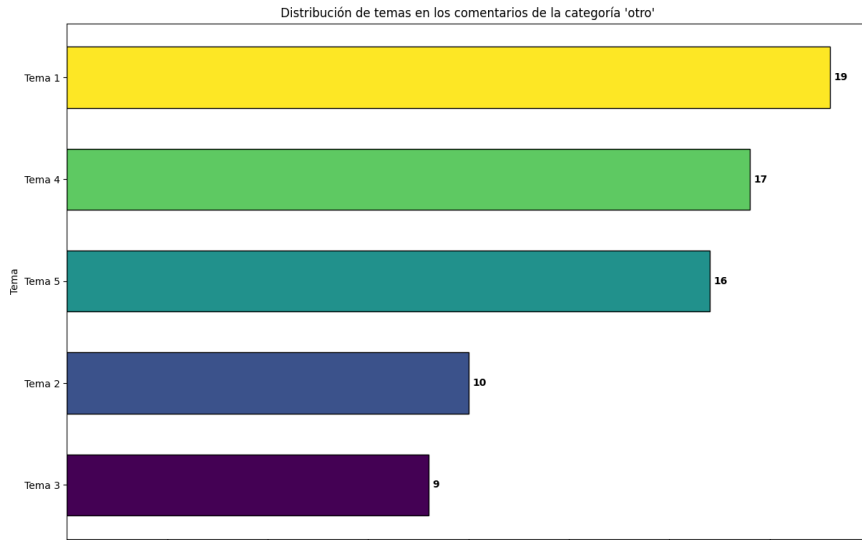
1. Configurar herramientas de procesamiento de texto.
2. Preparar y traducir los datos.
3. Preprocesar los comentarios:
  - Lematización y stemming.
4. Crear diccionario y corpus para LDA.
5. Aplicar TF-IDF.
6. Desarrollar el modelo LDA:
  - Entrenar modelos con diferentes números de temas.
  - Calcular coherencia.
7. Visualizar resultados.

# Número Óptimo de Tópicos

La evaluación de coherencia del modelo arrojó que el **número óptimo de temas es 5**, ya que maximiza la coherencia de los temas identificados.



# Distribución de los Temas Identificados





# Interpretación de los Tópicos

- **Tema 1:** Insatisfacción con el servicio al cliente.
  - ▶ Palabras clave: decepción, inaceptable, respuesta.
- **Tema 2:** Actitud y comportamiento del personal.
  - ▶ Palabras clave: grosería, relación, insatisfactorio.

# Interpretación de los Tópicos (cont.)

- **Tema 3:** Tiempo de espera y rapidez en la atención.
  - ▶ Palabras clave: espera, minutos, rápido.
- **Tema 4:** Problemas en sucursales y personal.
  - ▶ Palabras clave: personal, sucursal, inquietud.
- **Tema 5:** Problemas técnicos con aplicaciones y transacciones móviles.
  - ▶ Palabras clave: aplicación, transacción, fallo.

# Recomendaciones

- **Expansión del análisis** a otros canales de retroalimentación (chats, llamadas, correos electrónicos).
- **Integración de encuestas** de satisfacción instantáneas tras interacciones clave.
- **Desarrollo de modelos predictivos** basados en NPS para anticipar comportamiento de clientes.

## Recomendaciones (cont.)

- **Exploración de técnicas más avanzadas** de análisis de texto (e.g., BERT, Word2Vec) para mejorar la comprensión semántica.
- **Automatización de la retroalimentación** mediante sistemas que analicen comentarios de forma continua.
- **Futuros estudios académicos y colaboraciones** para mejorar la robustez de los modelos desarrollados.

# Anexos

Los anexos de este trabajo incluyen el código desarrollado en **Google Colab**, así como las bases de datos. Puedes acceder al código en el siguiente enlace:

[https://colab.research.google.com/drive/1AVMlirkqufPE-BpA9\\_ZLStWFVA-XSetJ?usp=sharing](https://colab.research.google.com/drive/1AVMlirkqufPE-BpA9_ZLStWFVA-XSetJ?usp=sharing)

# Referencias