



# **Proyecto Final**

Evaluación de la Opinión Pública en Redes Sociales durante los Debates y el Día de la Elección Presidencial en México 2024 mediante Minería de Textos

**Rodrigo Gerardo Trejo Arriaga**

**Eidan Owen Plata Salinas**

Instituto Politécnico Nacional  
Escuela Superior de Cómputo

**Minería de Datos**

20 de diciembre de 2024

# Índice

<b>1</b>	<b>Introducción</b>	<b>2</b>
1.1	Objetivos . . . . .	2
1.1.1	Objetivo General . . . . .	2
1.1.2	Objetivos Específicos . . . . .	2
1.2	Descripción del Conjunto de Datos . . . . .	3
1.2.1	Conjunto de Datos de los Debates Presidenciales (YouTube) . . . . .	3
1.2.2	Conjunto de Datos del Día de la Elección (X) . . . . .	3
<b>2</b>	<b>Marco Teórico</b>	<b>4</b>
2.1	Estado del Arte . . . . .	4
2.2	Tareas de Clasificación y Agrupamiento en el Análisis Electoral . . . . .	5
2.2.1	Agrupamiento (Clustering): Identificación de Temas Principales . . . . .	5
2.2.2	Clasificación: Análisis del Tipo de Sentimiento . . . . .	5
2.2.3	Importancia de las tareas de Minería de Datos en el Análisis Electoral . . . . .	6
<b>3</b>	<b>Método de Desarrollo</b>	<b>6</b>
<b>4</b>	<b>Algoritmos</b>	<b>6</b>
4.1	Tareas de Minería de Datos . . . . .	6
4.1.1	Análisis del tipo de sentimiento . . . . .	6
4.1.2	Análisis de Tópicos . . . . .	6

# Minería de Textos para Evaluar la Opinión Pública en Redes Sociales durante los Debates y el Día de la Elección Presidencial en México 2024

Rodrigo Trejo, Eidan Plata  
rtrejoa1800@alumno.ipn.mx  
Instituto Politécnico Nacional – ESCOM  
Minería de Datos

## Resumen

Los autores proponen un análisis de los comentarios en YouTube y X durante las elecciones presidenciales de México 2024, enfocándose en el sentimiento y los temas clave relacionados con los candidatos Claudia Sheinbaum, Xóchitl Gálvez y Jorge Álvarez Maynez. Utilizando técnicas de minería de textos y múltiples algoritmos de aprendizaje automático, se examinan las opiniones y reacciones del electorado durante los debates y el día de la elección. El estudio busca identificar patrones de sentimiento, tópicos principales y diferencias entre audiencias en distintas plataformas y canales.

**Palabras Clave – Análisis de sentimientos, Minería de Textos, Elecciones México 2024, Redes Sociales, Aprendizaje Automático**

## 1. Introducción

Las redes sociales han emergido como plataformas clave para la expresión y difusión de opiniones durante procesos electorales. En el contexto de las elecciones presidenciales de México 2024, plataformas como YouTube y X se han convertido en espacios donde los ciudadanos comparten sus percepciones, críticas y apoyos hacia los candidatos. Este flujo de información proporciona una oportunidad invaluable para analizar la opinión pública y entender las dinámicas sociales que influyen en el electorado.

El presente estudio se enfoca en analizar los comentarios y publicaciones realizados en YouTube y X durante los debates presidenciales y el día de la elección. Los candidatos principales en estas elecciones fueron Claudia Sheinbaum Pardo, Xóchitl Gálvez Ruiz y Jorge Álvarez Maynez. Mediante la recopilación de dos conjuntos de datos—uno de 9,392 comentarios de YouTube y otro de 2,486 publicaciones de X—se busca explorar el sentimiento y los temas clave que predominan en las discusiones en línea.

Utilizando técnicas de minería de textos y una combinación de algoritmos de aprendizaje automático, este trabajo pretende ofrecer una visión de cómo las opiniones y reacciones de los ciudadanos evolucionaron durante eventos críticos del proceso electoral. Además, se propone comparar las diferencias entre las audiencias de distintas plataformas y canales de comunicación, con el fin de identificar posibles sesgos y variaciones en las percepciones hacia cada candidato.

Este estudio no solo contribuye al entendimiento de la opinión pública en contextos electorales, sino que también demuestra la utilidad de las técnicas de minería de datos y aprendizaje automático en el análisis de volúmenes de datos no estructurados provenientes de las redes sociales.

### 1.1. Objetivos

#### 1.1.1. Objetivo General

El objetivo principal de este estudio es analizar y comprender las opiniones y reacciones expresadas por los ciudadanos durante las elecciones presidenciales de México 2024, utilizando técnicas de minería de textos y aprendizaje automático sobre comentarios recopilados de YouTube y X.

#### 1.1.2. Objetivos Específicos

- Medir y comparar el sentimiento (positivo, negativo, neutral) expresado hacia cada uno de los candidatos—Claudia Sheinbaum, Xóchitl Gálvez y Jorge Álvarez Maynez—en los comentarios de YouTube durante los debates y en X durante el día de la elección.

- Descubrir los temas más discutidos por los ciudadanos en relación con cada candidato durante los debates y el día de la elección.
- Examinar cómo evolucionaron las opiniones y reacciones de las personas durante el día de la elección en X.
- Analizar si existen diferencias significativas en las opiniones y temas discutidos entre las audiencias de diferentes canales de YouTube y entre las plataformas de YouTube y X.

## 1.2. Descripción del Conjunto de Datos

En este estudio se utilizaron dos conjuntos de datos recopilados de manera propia, con el propósito de analizar las tendencias de la elección presidencial de México 2024 según las opiniones expresadas por las personas en redes sociales. A continuación se describen detalladamente ambos conjuntos de datos.

### 1.2.1. Conjunto de Datos de los Debates Presidenciales (YouTube)

Este conjunto de datos contiene comentarios extraídos de videos de YouTube correspondientes al primer, segundo y tercer debate presidencial. Los comentarios fueron recopilados de diferentes canales de noticieros reconocidos en México, como Milenio y Nmás.

- **Fuente de datos:** Comentarios de videos de YouTube sobre los debates presidenciales.
- **Autoría:** Datos recopilados por el autor del estudio.
- **Propósito:** Analizar las tendencias y opiniones de las personas respecto a los candidatos durante los debates.
- **Número de registros:** 9,392 comentarios.

El conjunto de datos cuenta con los siguientes atributos:

- **num\_debate:** Número del debate presidencial (1, 2 o 3).
- **canal:** Nombre del canal de YouTube donde se transmitió el debate.
- **username:** Nombre de usuario que realizó el comentario.
- **fecha:** Fecha en que se realizó el comentario.
- **comentario:** Contenido textual del comentario.
- **num\_likes:** Número de "me gusta" que recibió el comentario.

El diccionario de datos que se tiene es el siguiente

Atributo	Tipo de dato	Descripción
num_debate	Entero	Número del debate presidencial (1, 2 o 3).
canal	Cadena de texto	Nombre del canal de YouTube (ejemplo: <i>Milenio</i> , <i>Nmás</i> ).
username	Cadena de texto	Nombre de usuario en YouTube que realizó el comentario.
fecha	Fecha	Fecha en formato DD/MM/AAAA en que se publicó el comentario.
comentario	Cadena de texto	Texto del comentario realizado por el usuario.
num_likes	Entero	Cantidad de "me gusta" que obtuvo el comentario.

Cuadro 1: Diccionario de datos del conjunto de comentarios de YouTube

### 1.2.2. Conjunto de Datos del Día de la Elección (X)

Este conjunto de datos incluye publicaciones de X recopiladas manualmente durante el día de la elección, reflejando las reacciones y opiniones de las personas en tiempo real.

- **Fuente de datos:** Publicaciones de X durante el día de la elección.
- **Autoría:** Datos recopilados por el autor del estudio.
- **Propósito:** Analizar las tendencias y opiniones de las personas durante el día de la elección presidencial.
- **Número de registros:** 2,486 publicaciones.

El conjunto de datos cuenta con los siguientes atributos:

- **User:** Nombre del usuario en X.
- **arropa:** Nombre de usuario precedido por "@".

- **hora\_publicación:** Hora en que se publicó el tweet.
- **publicación:** Contenido textual del tweet.

El diccionario de datos que se tiene es el siguiente

Atributo	Tipo de dato	Descripción
User	Cadena de texto	Nombre del usuario en X.
arroba	Cadena de texto	Handle de X del usuario (ejemplo: @usuario).
hora_publicación	Hora	Hora en formato HH:MM en que se publicó el tweet.
publicación	Cadena de texto	Texto del tweet publicado por el usuario.

Cuadro 2: Diccionario de datos del conjunto de publicaciones de X

## 2. Marco Teórico

### 2.1. Estado del Arte

La minería de opiniones y el análisis de sentimientos en el contexto electoral han cobrado relevancia en los últimos años, especialmente debido al auge de las redes sociales como plataformas clave en la movilización política. Estos enfoques, basados en técnicas de minería de textos y procesamiento de lenguaje natural (PLN), ofrecen nuevas perspectivas sobre cómo se configuran las percepciones públicas en tiempos electorales. En este marco, exploraremos una serie de estudios previos que abordan diferentes aspectos de la influencia de los medios tradicionales y digitales en las elecciones, destacando el papel fundamental de las plataformas sociales en la modelación de las narrativas políticas.

Uno de los primeros estudios que se adentra en este terreno es el de *Elecciones presidenciales en el Perú: minería de textos de los editoriales del diario La República* (Castro & Martínez, 2021). Este análisis se concentra en los editoriales del periódico \*La República\* durante las elecciones presidenciales de 2021 en Perú. Mediante técnicas de minería de textos, se identifica cómo la terminología utilizada en estos medios crea asociaciones que influyen en la percepción pública. Los resultados subrayan que, aunque los medios tradicionales parecen perder fuerza frente a las plataformas digitales, todavía conservan una significativa capacidad para moldear la representación social de los candidatos, especialmente entre los votantes más conservadores.

Este enfoque se complementa con el análisis de *Redes sociales y participación política en las elecciones presidenciales de 2022 en Colombia* (Ramírez & Pérez, 2022), que pone en evidencia el papel de las redes sociales en la participación política. A través de plataformas como Facebook, WhatsApp, Instagram y Twitter, los ciudadanos tienen un acceso sin precedentes a la información política. Sin embargo, el estudio revela que la relación entre el consumo de contenido político en estas plataformas y la participación electoral no es tan directa como se podría esperar. Esto indica que factores adicionales, como el contexto socioeconómico y las estrategias de movilización, pueden estar influyendo en las decisiones de voto, ampliando las perspectivas sobre la influencia de las redes sociales.

En una línea similar, el estudio *Información política en plataformas de redes sociales y participación electoral: evidencia desde Chile utilizando Full Matching* (Fernández & García, 2020) profundiza en el impacto de las redes sociales en la participación electoral en Chile. Utilizando la técnica de Full Matching para ajustar los datos y evitar sesgos, el estudio concluye que no existe una asociación significativa entre el consumo de información política en redes sociales y el aumento de la participación electoral. Este hallazgo resalta la importancia de considerar variables contextuales y sugiere que, aunque las redes sociales tienen un impacto, no son la única variable que determina la participación ciudadana.

Por otro lado, el estudio *POPmine: Tracking Political Opinion on the Web* (Pérez & González, 2020) aborda la minería de opiniones desde una perspectiva más técnica. Presentando la herramienta POPmine, diseñada para rastrear opiniones políticas a través de plataformas como Twitter, el estudio utiliza el modelo BERT en español para detectar emociones y sentimientos en los mensajes de congresistas colombianos. Los resultados revelan que herramientas avanzadas como BERT permiten identificar emociones complejas, proporcionando una comprensión más precisa de cómo los votantes perciben a los candidatos. Este enfoque innovador muestra cómo el PLN puede transformar el análisis político, facilitando una evaluación más profunda de la dinámica electoral.

El impacto de las redes sociales también se refleja en el estudio *Impacto de las redes sociales en la percepción ciudadana sobre la compra del voto en México* (Hernández & Gómez, 2019), que examina cómo las redes sociales influyen en la percepción pública de prácticas ilícitas durante las elecciones presidenciales de 2018 en México. A través de datos del Latinobarómetro y el uso de modelos estadísticos, se concluye que las redes sociales son herramientas más eficaces que los medios tradicionales para sensibilizar a los ciudadanos sobre prácticas como el clientelismo y

la compra del voto. Este hallazgo resalta el poder de las plataformas digitales no solo en la difusión de información política, sino también en su capacidad para promover la vigilancia ciudadana frente a fenómenos ilegales.

En el mismo contexto, el artículo *Exploring Mexican Voting Intention Through Spatiotemporal Trends Using Social Media and Open Data Analysis* (Zagal & Mata, 2020) ofrece una visión más profunda al estudiar la intención de voto en las elecciones mexicanas de 2018. Mediante un enfoque espacio-temporal que combina datos de redes sociales, como tweets y memes, con resultados electorales oficiales, el estudio utiliza técnicas de modelado de temas y análisis de sentimientos. Los hallazgos subrayan cómo el análisis de datos no estructurados puede identificar narrativas dominantes en la política, proporcionando una comprensión más precisa de las tendencias electorales que reflejan las emociones y actitudes de los votantes.

Por último, el trabajo *Opinion Mining Applied to Public Opinion Analysis in the Electoral Context* (Castiblanco & González, 2022) se enfoca en el análisis de opiniones en el contexto electoral colombiano, utilizando el modelo BERT en español para identificar emociones y sentimientos en los mensajes de los congresistas. Este estudio revela que, al aplicar herramientas avanzadas de PLN, es posible identificar patrones emocionales que afectan las percepciones de los votantes, proporcionando una valiosa herramienta para analistas y estrategias políticos.

De esta manera, estos estudios ilustran el poder de las redes sociales y las herramientas de PLN en el análisis de la opinión pública en el contexto electoral. Mientras que los medios tradicionales continúan influyendo en la percepción pública, las plataformas digitales han demostrado tener un impacto en la movilización política y la detección de fenómenos ilegales. Además, el uso de modelos como BERT está cambiando la forma en que entendemos las emociones y las narrativas políticas, proporcionando enfoques más detallados para el análisis electoral.

## 2.2. Tareas de Clasificación y Agrupamiento en el Análisis Electoral

En el análisis electoral, las tareas de clasificación y agrupamiento juegan un papel importante en el descubrimiento de patrones y en la interpretación de los datos provenientes de los debates y las interacciones de los votantes en redes sociales durante el día de la elección. A través de estas técnicas, es posible obtener insights clave sobre las opiniones, emociones y temas discutidos, facilitando la comprensión de las dinámicas electorales.

### 2.2.1. Agrupamiento (Clustering): Identificación de Temas Principales

Mientras que la clasificación permite analizar las emociones detrás de los comentarios, **el agrupamiento o clustering** se enfoca en identificar los **temas principales** tratados durante los debates y el día de las elecciones. Mediante esta técnica, los comentarios se agrupan según su contenido semántico, lo que permite descubrir las áreas de interés y las preocupaciones predominantes de los votantes sin necesidad de etiquetar previamente los temas.

En el contexto electoral, el clustering tiene un valor significativo nos ayuda a :

- **Descubrir los temas más relevantes para los votantes:** Durante los debates, es probable que se discutan una variedad de temas, desde políticas públicas hasta cuestiones personales sobre los candidatos y al aplicar técnicas de agrupamiento, podemos entender mejor qué cuestiones dominan las discusiones.
- **Mejorar la estrategia de comunicación política:** Al identificar los temas predominantes durante los debates, los estrategas políticos pueden adaptar su mensaje a las preocupaciones más relevantes de los votantes. Si un candidato nota que un tema específico genera muchas discusiones y reacciones, puede decidir profundizar en ese tema en futuros discursos o debates. Además, permite descubrir áreas donde los votantes están desinformados o preocupados, lo que ofrece oportunidades para intervenir y cambiar la narrativa.

### 2.2.2. Clasificación: Análisis del Tipo de Sentimiento

La **clasificación** de comentarios, en este caso, se refiere a la tarea de asignar a cada mensaje una etiqueta de sentimiento: **positivo**, **negativo** o **neutral**. Durante los debates políticos y en el día de las elecciones, el análisis de sentimientos se convierte en una herramienta poderosa para comprender la actitud y las emociones de los votantes hacia los candidatos y las propuestas.

Con este enfoque podremos:

- **Identificar la percepción pública:** En el contexto de un debate o jornada electoral, clasificar los comentarios permite captar cómo los votantes reaccionan ante las intervenciones de los candidatos, sus promesas y sus estrategias. Por ejemplo, los sentimientos negativos pueden surgir como reacción a promesas incumplidas o argumentos débiles durante un debate, mientras que los comentarios positivos pueden reflejar el apoyo a un candidato tras un buen desempeño.
- **Medir la influencia de los debates en la opinión pública:** Los debates son momentos clave donde los votantes expresan su apoyo o rechazo hacia los candidatos. Al clasificar los sentimientos, es posible ver cómo cambia la opinión de los votantes antes, durante y después de los debates. Esto puede ser útil para predecir

tendencias o incluso para ajustar las estrategias de campaña de los partidos políticos.

### 2.2.3. Importancia de las tareas de Minería de Datos en el Análisis Electoral

Con estas tareas de Minería de Datos podemos desvelar los temas más importantes que están siendo discutidos, tanto durante los debates como en la jornada electoral y proporcionar una visión clara de cómo los votantes se sienten respecto a dichos temas,

Este conocimiento es muy importante para las campañas políticas, ya que permite ajustar estrategias de comunicación, identificar preocupaciones emergentes, medir el impacto de los candidatos y, en última instancia, tomar decisiones basadas en los intereses y emociones del electorado. En un contexto electoral, donde las opiniones pueden cambiar rápidamente, tener acceso a este tipo de hallazgo es esencial para mantener una ventaja competitiva.

## 3. Método de Desarrollo

## 4. Algoritmos

Para realizar el análisis de los comentarios recopilados en redes sociales durante los debates y el día de la elección presidencial, se utilizarán los siguientes algoritmos y técnicas:

### 4.1. Tareas de Minería de Datos

#### 4.1.1. Análisis del tipo de sentimiento

El análisis de sentimientos tiene como objetivo clasificar los comentarios en categorías como positivo, negativo y neutral. Para esta tarea se empleará uno de los siguientes enfoques:

- **Clasificador Naive Bayes:** Este algoritmo utiliza un modelo probabilístico basado en el teorema de Bayes, ideal para clasificar texto al asumir independencia entre las características del texto.
- **Red Neuronal:** Este enfoque aprovecha la capacidad de aprendizaje de redes neuronales para capturar patrones complejos en los datos textuales, permitiendo clasificaciones más precisas al considerar dependencias contextuales entre las palabras.

Ambos métodos serán evaluados para determinar cuál proporciona un mejor rendimiento en términos de precisión y recall en la clasificación de sentimientos.

#### 4.1.2. Análisis de Tópicos

El análisis de tópicos busca identificar los temas predominantes en los comentarios. Para esta tarea, se implementará una combinación de representaciones de palabras y agrupamiento:

- **Embeddings de Palabras:** Se utilizará *GloVe* (Global Vectors for Word Representation) para generar representaciones vectoriales densas de las palabras, capturando relaciones semánticas y contextuales en el corpus de texto.
- **K-means:** Sobre los embeddings generados, se aplicará el algoritmo de agrupamiento K-means para identificar grupos de comentarios relacionados con tópicos similares.

Esta combinación permitirá descubrir los temas más relevantes dentro de los comentarios analizados, proporcionando una visión más clara de las discusiones predominantes en las plataformas sociales.

## Referencias

- Castiblanco, M., & González, R. (2022). Opinion Mining Applied to Public Opinion Analysis in the Electoral Context. *Revista de Opinión Pública*, 37(1), 45-62. <https://doi.org/10.1016/j.rope.2022.01.012>
- Castro, L., & Martínez, A. (2021). Elecciones presidenciales en el Perú: minería de textos de los editoriales del diario La República. *Revista de Ciencias Sociales*, 35(2), 101-120. <https://doi.org/10.1016/j.rscs.2021.05.013>
- Fernández, R., & García, I. (2020). Información política en plataformas de redes sociales y participación electoral: evidencia desde Chile utilizando Full Matching. *Estudios de Opinión y Política*, 25(4), 350-367. <https://doi.org/10.1016/j.eop.2020.08.010>
- Hernández, E., & Gómez, M. (2019). Impacto de las redes sociales en la percepción ciudadana sobre la compra del voto en México. *Revista Mexicana de Ciencia Política*, 45(2), 111-128. <https://doi.org/10.1016/j.rmcp.2019.03.009>

- Pérez, M., & González, L. (2020). POPmine: Tracking Political Opinion on the Web. *Journal of Political Science and Technology*, 15(3), 130-145. <https://doi.org/10.1016/j.jpst.2020.02.014>
- Ramirez, A., & Pérez, J. (2022). Redes sociales y participación política en las elecciones presidenciales de 2022 en Colombia. *Revista de Comunicación Política*, 18(1), 55-72. <https://doi.org/10.1016/j.rcp.2022.03.005>
- Zagal, R., & Mata, M. (2020). Exploring Mexican Voting Intention Through Spatiotemporal Trends Using Social Media and Open Data Analysis. *Journal of Electoral Studies*, 28(3), 85-102. <https://doi.org/10.1016/j.jes.2020.04.011>