



Proyecto Final

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 Gerardo Trejo Arriaga

Eidan Owen Plata Salinas

Instituto Politécnico Nacional
Escuela Superior de Cómputo

Minería de Datos

21 de noviembre de 2024

Índice

1	Introducción	2
2	Objetivos	2
2.1	Objetivo General	2
2.2	Objetivos Específicos	2
3	Material Utilizado	3
3.1	Conjunto de Datos de los Debates Presidenciales (YouTube)	3
3.1.1	Atributos del Conjunto de Datos	3
3.1.2	Diccionario de Datos	3
3.2	Conjunto de Datos del Día de la Elección (Twitter)	3
3.2.1	Atributos del Conjunto de Datos	4
3.2.2	Diccionario de Datos	4

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.

2. Objetivos

2.1. 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.

2.2. 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.

3. Material Utilizado

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.

3.1. 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.

3.1.1. Atributos del Conjunto de Datos

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.

3.1.2. Diccionario de Datos

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

3.2. 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.

3.2.1. Atributos del Conjunto de Datos

El conjunto de datos cuenta con los siguientes atributos:

- **User:** Nombre del usuario en X.
- **arroba:** Nombre de usuario precedido por “@”.
- **hora_publicación:** Hora en que se publicó el tweet.
- **publicación:** Contenido textual del tweet.

3.2.2. Diccionario de Datos

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