



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

**Maestría en Ciencia de Datos**

**Universidad Autónoma de Nuevo León**

## **Tarea 1 - Análisis estadístico de comentarios del primer debate presidencial**

Por

Juan Jesús Torres Solano

Profesor: José Alberto Benavides Vázquez

# Índice general

<b>List of Figures</b>	<b>2</b>
<b>1 Introducción</b>	<b>2</b>
1.1 Primer objetivo .....	2
1.2 Segundo objetivo .....	2
<b>2 descripción de los datos</b>	<b>3</b>
<b>3 Metodología</b>	<b>4</b>
3.1 método de recopilación .....	4
<b>4 Resultados</b>	<b>5</b>
<b>References</b>	<b>6</b>

# **1. INTRODUCCIÓN**

En el presente trabajo se realiza el análisis estadístico de texto, empleando como fuente a los comentarios de la transmisión en vivo del primer debate presidencial, realizado el día 7 de abril del 2024; los comentarios presentan una estructura lingüística mal estructurada, de igual forma, se identifican Emojis entre palabras o comentarios completos a partir de los mismos.

## **1.1. PRIMER OBJETIVO**

Generar información estadística que nos permita visualizar los elementos característicos, que forman parte de los comentarios, como son, la frecuencia palabras, votos a los comentarios (Top 10 de votos), frecuencias de participación en los comentarios (Top 10 de Usuarios) y la relación entre los votos y las replicas.

## **1.2. SEGUNDO OBJETIVO**

Realizar una comparativa entre el numero de menciones y la estadística de intención de voto previo a este debate, a modo de comparación estadística entre la expresión del nombre de los candidatos en los comentarios y la intención de voto documentada hasta ese momento.

## 2. DESCRIPCIÓN DE LOS DATOS

Se identifican un total de 4307 comentarios de los cuales 1955 presentan algún tipo de comentario textuales ademas de Emojis, en el cuadro 2.1 podemos observar una muestra representativa del conjunto de datos, de este conjunto emplearemos las columnas siguientes:

- text : Comentario del Usuario.
- author : Nombre de usuario que realiza el comentario.
- votes : numero de votos recibidos
- replies : nuemero de replicas en los comentarios.

ID	cid	text	time	author	channel	votes	replies	photo	heart	reply	time_parsed
0	UgzVmyrPVgEBr9gxD2N4AaABAg	El reloj debió verse en todo momento, para ten...	hace 1 mes	@celinaramirez6363	UC-oQImqSgqEG8633yeZy4A	394	16.0	https://yt3.ggpht.com/ytciAldro_i9HnH5acjeSTDn...	False	False	1.713569e+09
1	UgyC-sngH2JL9TgI7m84AaABAg	Me uno a exigir que pongan los relojes de los ...	hace 1 mes	@puellacodicum8569	UC4SVhkvZdpk5wZoa2sB3A	555	19.0	https://yt3.ggpht.com/ytciAldro_k8pN9LRX3SzxX...	False	False	1.713569e+09
2	UgxLz2HxYmRxBu90od4AaABAg	Me parece muy bien cómo contestó Maynes, fue d...	hace 1 mes	@alondrareal8518	UCwXDEdHak4qo3kGzUlnJA	79	NaN	https://yt3.ggpht.com/ytciAldro_n2uIdu2oU52m0X...	False	False	1.713569e+09
...	...	...	...	...	...	...	...	...	...	...	...
1953	Ugd9_f5xw-zYV8wT3x4AaABAg	Maynes, palero de Claudia #CrimesAgainstHumani...	hace 2 semanas	@yamerosolitario2	UCLD8LO3LOJeD4QITudTJxw	0	NaN	https://yt3.ggpht.com/ytciAldro_mk4-V9s1KN7G-W...	False	False	1.714952e+09
1954	UgyGU_RvOWnj6CpK2M24AaABAg	Xóchil Presidente	hace 1 mes	@marcoantonioarroyo952	UCM9S9x9pCM0H2F5u_7yKA	0	NaN	https://yt3.ggpht.com/ytciAldro_nQnx-VeJ6E8B4...	False	False	1.713569e+09
1955	UgjiN0jvRu1pB44x4AaABAg	Cuánto dinero les pagaste por esos reconocimie...	hace 1 mes	@maru641	UC4uT9ULvwtSo29SmD2e8g	0	NaN	https://yt3.ggpht.com/POohj0B-XE9089HWJOPG7uzF...	False	False	1.713569e+09

**Cuadro 2.1:** estructura y contenido de los datos descargados

## **3. METODOLOGÍA**

### **3.1. MÉTODO DE RECOPIACIÓN**

## **4. RESULTADOS**

## BIBLIOGRAFÍA

- [1] A. Savitzky, M. J. E. Golay, Smoothing and differentiation of data by simplified least squares procedures., *Analytical Chemistry* 36 (8) (1964) 1627–1639. doi: 10.1021/ac60214a047.
- [2] C. J. Harvey, J. M. Pilcher, R. J. Eckersley, M. J. Blomley, D. O. Cosgrove, Advances in ultrasound, *Clinical Radiology* 57 (3) (2002) 157–177. doi:10.1053/crad.2001.0918.