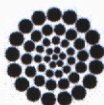




GOBIERNO DE
MÉXICO



CONAHCYT
CONSEJO NACIONAL DE HUMANIDADES
CIENCIAS Y TECNOLOGÍAS



**BIBLIOTECA INFOTEC
VISTO BUENO DE TRABAJO TERMINAL**

Maestría en Ciencia de Datos e Información
(MCDI)

Ciudad de México, a 26 de septiembre de 2024

**UNIDAD DE POSGRADOS
PRESENTE**

Por medio de la presente se hace constar que el trabajo de titulación:

"Análisis de sentimientos y emociones en Twitter, concordancia entre figuras públicas y sus seguidores o detractores"

Desarrollado por el alumno: **Ambrosio de Jesús Torres Mendoza**, bajo la asesoría del **Dr. Mario Graff Guerrero** cumple con el formato de Biblioteca, así mismo, se ha verificado la correcta citación para la prevención del plagio; por lo cual, se expide la presente autorización para entrega en digital del proyecto terminal al que se ha hecho mención. Se hace constar que el alumno no adeuda materiales de la biblioteca de INFOTEC.

No omito mencionar, que se deberá anexar la presente autorización al inicio de la versión digital del trabajo referido, con el fin de amparar la misma.

Sin más por el momento, aprovecho la ocasión para enviar un cordial saludo.

Mtro. Carlos Josué Lavandeira Portillo
Director Adjunto de Innovación y Conocimiento

jah
CJLP/jah

C.c.p. Felipe Alfonso Delgado Castillo.- Gerente de Capital Humano.- Para su conocimiento.

Ambrosio de Jesús Torres Mendoza.- Alumno de la Maestría en Ciencia de Datos e Información.- Para su conocimiento.



INFOTEC CENTRO DE INVESTIGACIÓN E
INNOVACIÓN EN TECNOLOGÍAS DE LA
INFORMACIÓN Y COMUNICACIÓN

DIRECCIÓN ADJUNTA DE INNOVACIÓN Y
CONOCIMIENTO
GERENCIA DE CAPITAL HUMANO
POSGRADOS

“Análisis de sentimientos y emociones en Twitter, concordancia entre figuras públicas y sus seguidores o detractores”

Implementación de un Proyecto Laboral,
Que para obtener el grado de MAESTRO EN
CIENCIA DE DATOS E INFORMACIÓN

Presenta:

Ambrosio de Jesús Torres Mendoza

Asesor:

Dr. Mario Graff Guerrero

Ciudad de México, septiembre, 2024.

Agradecimientos

Gracias al apoyo de mi familia, Ilse, Fer y Mony, al doctor Mario Graff por su paciencia y orientación, y al INFOTEC por darme la oportunidad de seguir aprendiendo.

Índice general

Índice de figuras	vii
Índice de tablas	viii
Abreviaturas y acrónimos	ix
Glosario	x
Resumen.....	xi
Introducción	1
Capítulo 1. Generalidades	4
1.1 Planteamiento del problema	4
1.2 Protocolo de investigación	8
1.3 Justificación.....	8
1.4 Límites y alcances	12
Capítulo 2. Base de datos.....	14
2.1 Construcción de la base de datos	14
2.1.1 Fuentes de información	14
2.1.2 Creación de la base de datos	16
2.1.3 Creación y definición de nuevas variables	16
2.2 Preprocesamiento de la base de datos	16
2.3 Análisis exploratorio de los datos.....	21
Capítulo 3. Diseño del estudio y ajuste de modelos	25
3.1 Marco teórico	25
3.2 Marco metodológico	31
3.3 Ajuste de los modelos	34
3.4 Análisis de los resultados	37
Conclusiones y recomendaciones	53
Fuentes de consulta	xii
ANEXO 1	xvii

Índice de términos	xviii
--------------------------	-------

Índice de figuras

Figura 1. Número de Tweets donde se menciona al aspirante.....	21
Figura 2. Número de tweets publicados por aspirante.....	21
Figura 3. Número de seguidores en el tiempo por aspirante.....	22
Figura 4. Porcentaje de tweets retweeteados por aspirante.....	22
Figura 5. Número de tweets publicados en el tiempo por aspirante.....	23
Figura 6. Fases de desarrollo del diseño de investigación.....	32
Figura 7. Porcentaje de tweets totales y tweets clasificados en sentimientos, por aspirante.....	35
Figura 8. Porcentaje de emociones expresadas en los tweets.....	39
Figura 9. Porcentaje de sentimientos expresados en los tweets.....	39
Figura 10. Porcentaje de emociones expresadas en los tweets por aspirante.....	40
Figura 11. Porcentaje de sentimientos expresados en los tweets por aspirante...	41
Figura 12. Porcentaje de emociones de los tweets publicados por los aspirantes por emociones y figura pública.....	42
Figura 13. Porcentaje de emociones de los tweets publicados por los usuarios por emociones y figura pública.....	43
Figura 14. Porcentaje de sentimientos de los tweets publicados por los aspirantes por sentimiento y figura pública.....	44
Figura 15. Porcentaje de sentimientos de los tweets publicados por los usuarios por sentimiento y figura pública.....	44
Figura 16. Palabras asociadas al sentimiento positivo de Evodio Vázquez y de los tweets.....	xx
Figura 17. Palabras asociadas al sentimiento negativo de Evodio Vázquez y de los tweets del público en general.....	xx
Figura 18. Palabras asociadas al sentimiento positivo de Félix Salgado y de los tweets del público en general.....	xxi
Figura 19. Palabras asociadas al sentimiento negativo de Félix Salgado y de los tweets del público en general.....	xxi

Figura 20. Palabras asociadas al sentimiento positivo de Marcos Parra y de los tweets del público en general.....	xxii
Figura 21. Palabras asociadas al sentimiento negativo de Marcos Parra y de los tweets del público en general.....	xxii
Figura 22. Palabras asociadas al sentimiento positivo de Mario Moreno y de los tweets del público en general.....	xxiii
Figura 23. Palabras asociadas al sentimiento negativo de Mario Moreno y de los tweets del público en general.....	xxiii

Índice de tablas

Tabla 1. Parámetros que se usaron en la búsqueda.....	16
Tabla 2. Pruebas de hipótesis de las proporciones de tweets clasificados contra el total de tweets.....	35
Tabla 3. Porcentaje y promedio de sentimientos de tweets publicados por figuras públicas por modelo y por aspirante.....	37
Tabla 4. Porcentaje y promedio de sentimientos de tweets publicados por público en general por modelo y por aspirante.....	38
Tabla 5. Comparativo de palabras más usadas entre figura pública y público en general para los sentimientos positivo y negativo.....	46
Tabla 6. Ejemplo de valores para calcular la correlación de Spearman.....	47
Tabla 7. Coeficientes de correlación de Spearman por sentimientos positivo y negativo.....	48
Tabla 8. Extracción de tópicos de los tweets de figuras públicas y usuarios.....	50
Tabla 9. Promedio de sentimiento por aspirante y tipo de población.....	51
Tabla 10. Extracción de tópicos.....	xxv

Abreviaturas y acrónimos

AI	Inteligencia artificial (del inglés "Artificial Intelligence").
API	Siglas en inglés de Interfaz de Programación de Aplicaciones
CSV	Siglas en inglés de valores separados por comas.
ML	Aprendizaje automático (del inglés "Machine Learning").
NCR	Siglas en inglés del Consejo de Investigación Nacional de Canadá.
ONTSI	Observatorio nacional de telecomunicaciones y de la SI
PLN	Procesamiento Natural del Lenguaje (del inglés "natural language processing NLP")
RAE	Real Academia Española
SAS	Sentiment-analysis-spanish
SVM	Máquinas de Vectores de Soporte (del inglés support vector machine)
TF-IDF	Frecuencia de términos – Frecuencia inversa del documento (del inglés "Term frequency – Inverse document frequency")

Glosario

“A”

Algoritmo: Un conjunto de instrucciones o reglas lógicas diseñadas para resolver un problema o realizar una tarea específica.

“B”

Big Data: Conjunto de datos tan grandes y complejos que resulta difícil procesarlos con herramientas convencionales de análisis de datos.

Blog: Sitio web que incluye, a modo de diario personal de su autor o autores, contenidos de su interés, actualizados con frecuencia y a menudo comentados por los lectores. Real Academia Española, RAE, (2022).

Bots: Programas automatizados que simulan interacción humana en las plataformas de redes sociales.

“E”

Endpoints: interfaz expuesta por un comunicante o un canal de comunicación o un punto de acceso a una API.

“F”

Followers o seguidores: Usuarios de Twitter que siguen a otro usuario.

Figura pública: un individuo o entidad que ha adquirido fama o notoriedad o ha participado en una controversia pública particular, Merriam-Webster, Dictionary, (2023).

H”

Hashtags: es un tipo de etiqueta o etiqueta de metadatos utilizado en los servicios de redes sociales y microblogging que hace que sea más fácil para los usuarios encontrar mensajes con un tema o contenido específico. (Arimetrics, 2024)

“K”

Keras: es una biblioteca de Redes Neuronales de Código abierto escrita en Python

“O”

OpenAI: es una empresa estadounidense de investigación y despliegue de inteligencia artificial fundada en 2015 e, inicialmente, sin ánimo de lucro. Su misión original era asegurar que la inteligencia artificial general fuese desarrollada como código libre y en beneficio de toda la humanidad. (Wikipedia, 2024).

“P”

Público en general: todas o la mayoría de las personas, especialmente aquellas que no forman parte de un grupo específico.

“S”

Stopwords: Palabras vacías las cuales carecen de sentido cuando se escriben solas.

“T”

TensorFlow: es una biblioteca de código abierto para aprendizaje automático a través de un rango de tareas, y desarrollado por Google (Wikipedia, 2024)..

Resumen

En la investigación de mercado y opinión pública tradicional, como son las entrevistas cara a cara, telefónicas o grupos de enfoque, se obtienen información precisa y de primera mano, sin embargo, tienen alto costos en recursos humanos, económicos y de tiempo, además de la complejidad logística. En contraparte, el auge de redes sociales como Twitter, ahora X, que es considerada la red de la opinión y de la política, permite analizar la aprobación, aceptación, o conexión de las figuras públicas y los usuarios de Twitter. De esta necesidad, surge este estudio el cual pretende analizar la concordancia o similitud de las emociones y sentimientos de lo publicado por las figuras públicas y lo publicado por los usuarios, ya sean seguidores o detractores.

Introducción

En las dos últimas décadas, se ha incrementado el uso de redes sociales. Estas permiten la comunicación directa entre marcas, empresas, gobiernos o figuras públicas, por un lado, y clientes, usuarios o público en general del otro lado. De las redes sociales, Twitter (ahora X) es una red que permite la expresión de la opinión y la política, y que da voz al ciudadano común para comunicar su punto de vista sobre los acontecimientos que le interesan o afectan, por esta razón es de interés de este estudio el análisis de los tweets publicados en español.

Esta investigación trata de analizar los sentimientos y emociones de las publicaciones en Twitter de una figura pública y de sus seguidores, sean simpatizante o detractores, en un periodo de tiempo específico. Además, hacer un análisis de concordancia entre los sentimientos, emociones y tópicos que se expresan en los tweets de dicha persona contra las publicaciones de sus seguidores o detractores, es decir, interesa conocer la aprobación o desaprobación del público en general, en relación a lo expresado por una figura pública.

En el **capítulo uno** se describe el planteamiento de problema, protocolo de investigación, justificación, límites y alcances. El desarrollo de estos conceptos permitirá definir una estructura adecuada al proceso de investigación. Según Samperi et al. (2014), se debe: “formular de manera lógica y coherente el problema de investigación cuantitativa con todos sus elementos; redactar objetivos y preguntas de investigación cuantitativa; y comprender los criterios para proponer y evaluar un problema de investigación cuantitativa”.

En el **capítulo dos** se describe la base de datos que se utilizó en este estudio, cómo se construyó, la fuente de información, cómo se usó el API de Twitter, las variables o características relevantes, la creación de nuevas variables, el procesamiento previo a la aplicación de algún modelo y por último un breve análisis exploratorio de los datos.

En el **capítulo 3** se define el marco teórico que es “una descripción detallada de cada uno de los elementos esenciales de la teoría, de tal manera que la formulación del problema y su solución sean una deducción lógica de ella” (Baena, 2017); el marco metodológico el cual tiene como propósito establecer cómo se llevará a cabo la investigación, esto es, la estrategia para obtener la información y las actividades para responder a los objetivos definidos (Lerma, 2009). Adicional, se mostrarán el análisis estadístico de los resultados y las conclusiones.

Capítulo 1

Generalidades

Capítulo 1. Generalidades

El objetivo de este capítulo es establecer los requisitos mínimos para un proceso de investigación cuantitativo como lo son: el **planteamiento del problema**, que de acuerdo con Samperi et al. (2014), es “afinar y estructurar más formalmente la idea de investigación” (p. 36); **protocolo de investigación** donde se establece la pregunta, el objetivo y la hipótesis de investigación; **la justificación** que indica los motivos y necesidades que llevan a seleccionar el tema para desarrollarlo, las cuales deben ser de origen externo u objetivo, y de origen interno o subjetivo, y **los límites y alcances** del estudio, Baena (2017).

1.1 Planteamiento del problema

En las dos últimas décadas el mundo ha cambiado mucho, y uno de los factores de este cambio se debe al auge del internet, y de éste el incremento del uso de las redes sociales, que según ONTSI (Observatorio nacional de telecomunicaciones y de la SI, 2011) son sitios “en la red cuya finalidad es permitir a los usuarios relacionarse, comunicarse, compartir contenido y crear comunidades [...]”; funcionan como una herramienta de democratización de la información que transforma a las personas en receptores y en productores de contenidos”. Las redes sociales generan todos los días una gran cantidad de datos, y forman parte de lo que llamamos Big Data. Actualmente, en México hay 81 millones de usuarios de redes sociales, con una tasa de penetración (porcentaje de usuarios) del 78.3%, Statista (2022). Los usuarios de Twitter en México son casi 14 millones, Statista (2022).

Según Gómez (2011) Twitter es:

Una herramienta que permite a los usuarios enviar y recibir de manera instantánea mensajes denominados tweets. Es un canal de comunicación rápido entre una persona y una comunidad, que en el lenguaje de Twitter se denominan: seguidores o “followers”.

Es una aplicación social útil que facilita enviar mensajes privados a nuestros seguidores, recibir mensajes de quienes seguimos y, lo más atractivo de todo, enviar mensajes públicos a cualquiera de los millones de usuarios registrados en el sistema. (p. 539).

Twitter es un micro blog donde millones de usuarios pueden expresar sus puntos de vista, y estar de acuerdo o en desacuerdo con lo que publican los demás usuarios. Es considerada como la red social de la política, donde figuras públicas y ciudadanos comunes expresan sus opiniones en temas públicos y sociales. Muchos de los tweets tienen una carga de sentimientos y emociones. En Twitter hay 4 formas más comunes de reaccionar a un tweet: responder, retweetear, me gusta y compartir; sin embargo, en este estudio solo se analizarán los tweets y retweets, es decir, los textos publicados en la plataforma.

Según la Real Academia Española, RAE, (2022) el sentimiento es “el hecho o efecto de sentir o sentirse” o “estado afectivo del ánimo”, mientras que la emoción es “la alteración del ánimo intensa y pasajera, agradable o penosa, que va acompañada de cierta conmoción somática” o “el interés, generalmente expectante, con que se participa en algo que está ocurriendo”. Los sentimientos y emociones son de gran importancia porque influyen en el comportamiento humano. De acuerdo con Wikipedia (2020), “se ha comprobado que el funcionamiento normal del cerebro cambia según el estado de ánimo subyacente y que, en ocasiones, incluso las decisiones racionales de las personas pueden verse notoriamente afectadas por los sentimientos”.

500 millones de tweets se publican al día en todo el mundo Websiterating (2022), y para analizar grandes volúmenes de datos y tan complejos como el lenguaje natural solo es posible hacerlo con la ciencia de datos.

Lui (2015) define la ciencia de datos como:

Un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados, lo cual es una continuación de algunos campos de análisis de datos como la estadística, la minería de datos, el aprendizaje automático, y la analítica predictiva, así como descubrimiento de conocimiento y minería de datos. La ciencia de datos se trata de convertir los datos en información. (p. 6)

Una de las ramas de la ciencia de datos es la inteligencia artificial (IA) y Rouhiainen (2018) define la inteligencia artificial como:

La IA es la capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano. Sin embargo, a diferencia de las personas, los dispositivos basados en IA no necesitan descansar y pueden analizar grandes volúmenes de información a la vez. Asimismo, la proporción de errores es significativamente menor en las máquinas que realizan las mismas tareas que sus contrapartes humanas. (p. 17)

Como parte de la inteligencia artificial hay un área que estudia el lenguaje natural, mejor conocido como procesamiento natural del lenguaje (PLN por sus siglas en inglés). Vásquez et al. (2009) lo define como:

La utilización de un lenguaje natural para comunicarnos con la computadora, debiendo ésta entender las oraciones que le sean

proporcionadas, el uso de estos lenguajes naturales facilita el desarrollo de programas que realicen tareas relacionadas con el lenguaje o bien, desarrollar modelos que ayuden a comprender los mecanismos humanos relacionados con el lenguaje. (p. 48)

Una vez que se procesan los tweets por medio de librerías de NLP se pueden utilizar algoritmos de machine learning para la obtención de sentimientos y emociones. De acuerdo con Russell, S. y Norvig, P. (2009) Machine Learning es:

El subcampo de las ciencias de la computación y una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan. Se dice que un agente aprende cuando su desempeño mejora con la experiencia y mediante el uso de datos; es decir, cuando la habilidad no estaba presente en su genotipo o rasgos de nacimiento. (p. 229)

1.2 Protocolo de investigación

En el protocolo de investigación se describen las partes para la implementación y organización de una investigación, estas son:

Objetivo general de investigación

Analizar la concordancia del discurso, de las emociones y los sentimientos expresados en los tweets de figuras públicas respecto a los tweets del público en general, ya sean seguidores o detractores.

Objetivos específicos de la investigación

Analizar los sentimientos y emociones de tweets publicados en español por figuras públicas y sus seguidores, en este caso, candidatos a la elección para gobernar el Estado de Guerrero, utilizando aplicaciones que permitan procesar textos en español.

Analizar la concordancia entre las figuras públicas y el público en general respecto al discurso, los tópicos y léxico que utilizan ambas poblaciones.

1.3 Justificación

En la investigación de mercados y opinión pública tradicional se utilizan métodos de investigación cuantitativa, como lo son las encuestas cara a cara, entrevistas telefónicas, y entrevistas en puntos de afluencia; así como métodos cualitativos, entre los que se encuentran los grupos de enfoque, entrevistas a profundidad o entrevistas contextuales, por mencionar algunos. Estos métodos de investigación, bien aplicados, son muy precisos y generan información de primera mano, pero son muy costosos en recursos económicos, tiempo y humanos. Una alternativa para analizar la aprobación, aceptación, o conexión de las figuras públicas y los usuarios es por medio de la escucha social.

El uso de redes sociales permite la comunicación directa entre marcas, empresas, gobiernos o figuras públicas, por un lado, y clientes, usuarios o público en general del otro lado. De las redes sociales, Twitter es la red de la opinión y de la política, que permite al ciudadano común expresar su punto de vista sobre los acontecimientos que le interesan o afectan, por esta razón es de interés de este estudio el análisis de los tweets publicados en español.

Una herramienta de la investigación de opinión pública y estudios de mercado es el marketing emocional, el cual (Guardiola, 2016) lo define como:

El marketing emocional es la disciplina del marketing que utiliza una marca dentro de una estrategia, con el objetivo de lograr un vínculo afectivo con usuarios, consumidores, clientes y futuros clientes, para que estos sientan la marca como algo propio y necesiten ser parte de ella.

Entonces, es de interés de esta investigación explorar, describir y analizar los sentimientos, emociones y tópicos que se expresan en Twitter, y a su vez, analizar la concordancia y similitud de lo que una figura pública expresa, respecto a lo que expresa el público en general. Por figura pública podemos mencionar a un candidato o candidata a gobernador, un diputado federal o un jefe de gobierno. El público en general son los ciudadanos comunes, periodistas, seguidores o detractores de la figura pública.

Para qué

Para la mayoría de las figuras públicas, como para las marcas reconocidas, es de interés conocer el grado de empatía, aceptación o aprobación que tienen sus acciones, decisiones, productos y servicios. Una forma de medir la aceptación o el rechazo, son las publicaciones en redes sociales, es decir, es de interés conocer el sentir de la gente y cómo se relacionan con sus deseos, necesidades y valores.

Beneficios que deriva de ella

Los principales benefactores son las marcas, empresas y figuras públicas que usan o son mencionados en Twitter. Porque se pueden generar de una forma relativamente fácil un análisis de sentimientos y emociones, es decir, pueden tener acceso a la opinión pública de esta red social. Es importante aclarar que el alcance de un análisis de sentimientos y emociones en Twitter se limita solo a eso, es decir, no tiene un mayor alcance que lo publicado en Twitter, a diferencia, por ejemplo, de una encuesta nacional para medir la intención de voto para las elecciones del presidente de la república.

Utilidad de metodología

No se pretende desarrollar ninguna nueva metodología de análisis de sentimientos, pero se tiene la intención de probar algunas herramientas que permitan medir la relación, concordancia o similitud entre los tweets de diferentes sujetos.

Varios estudios se han realizado entorno al análisis de sentimientos en Twitter utilizando herramientas de Machine Learning, por ejemplo:

Tellez et al. (2017) realizaron el estudio “A case study of Spanish text transformations for twitter sentiment analysis” el cual analiza la preparación y transformación de textos para el análisis de sentimientos. Algunos de los problemas del análisis de texto que mencionan son los errores ortográficos, el argot, repetición de letras, etc. Aplicaron transformaciones como: la lematización, derivación, eliminación de entidades, tokenizar, usar n-grams de palabras, todo esto con el objetivo de mejorar el clasificador. Usaron los clasificadores de Naïve Bayes, Árboles de decisión y la Máquina de Vector Soporte, SVM (máquinas de vectores de soporte por sus siglas en inglés). Un punto interesante en el artículo es la combinación de n-grams basados en palabras y q-grams basados en caracteres, y que utilizaron textos en español. El propósito de mi investigación es la aplicación del análisis de sentimientos, emociones y tópicos, y medir la

concordancia y similitud de textos entre dos poblaciones: figura pública y público en general.

El artículo “Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA” García y Berton (2010) toca el tema de análisis de sentimientos en Twitter, pero con un enfoque muy distinto al que se pretende en este estudio.

En otro artículo, “Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts” Nogueira y Gatti (2014) analizan tweets y comentarios cortos de películas, con el propósito de entrenar y predecir los sentimientos en positivo y negativo por medio de un modelo de red neuronal convolucional profunda.

Se han publicado muchos estudios de análisis de sentimientos utilizando textos de Twitter en inglés, y algunos estudios con textos en español, pero no se tienen estudios sobre la concordancia o similitud entre lo que twittea una figura pública y la relación con sus seguidores o detractores.

Se mencionaron tres casos de estudios de análisis de sentimientos y emociones en Twitter, como un simple ejemplo de lo que se ha publicado, y en la sección de referencias se pusieron algunos artículos consultados que tocan el tema de análisis de sentimientos en Twitter, con un enfoque diferente al que pretende este estudio.

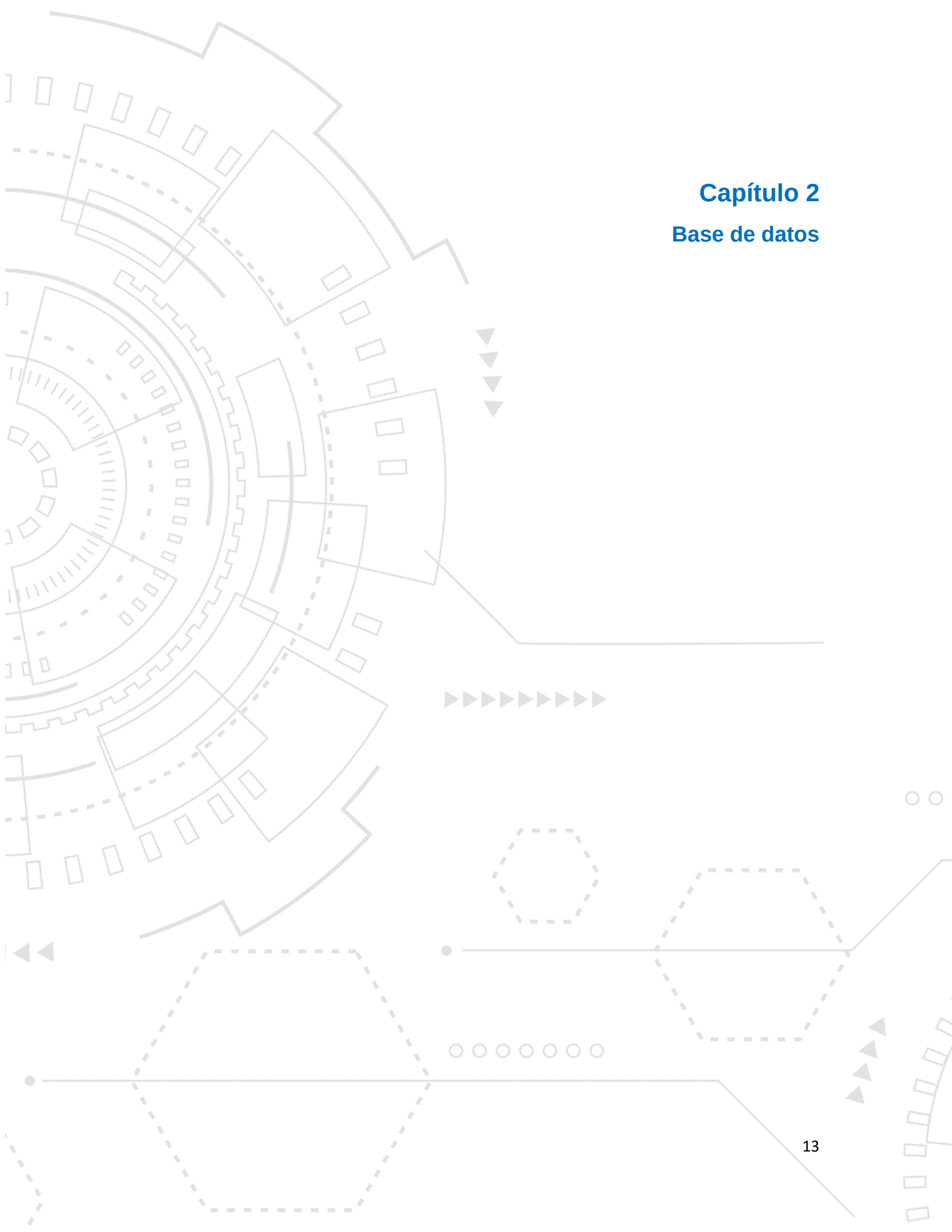
1.4 Límites y alcances

Límites

Se analizarán los sentimientos y emociones de tweets publicados por las figuras públicas, en este caso, los tweets de 4 aspirantes a gobernador por el estado de Guerrero:

- Mario Moreno Arcos (PRI),
- Evodio Velázquez Aguirre (PRD),
- Félix Salgado Macedonio, (MORENA) y
- Marcos Efrén Parra Gómez (PAN)

Y de los tweets publicados con relación a ellos por sus seguidores o detractores, en el periodo de enero del 2013 a febrero del 2021.



Capítulo 2

Base de datos

Capítulo 2. Base de datos

En este capítulo se explicará la construcción del dataset, la fuente de información, que en este caso es Twitter; las características que se usarán, de las cuales la más importante es el tweet o texto; el preprocesamiento que se aplicará previo a la aplicación de modelos de obtención de sentimientos y emociones y, por último, un análisis exploratorio de los datos.

2.1 Construcción de la base de datos

Como parte de la construcción del dataset se verá la fuente de información, es decir, de dónde y cómo se obtuvieron los datos y la creación del dataset.

Fuentes de información

Los Tweets se obtuvieron de publicaciones relacionadas con cada una de las personas de interés, aspirantes de gobernado por el estado de Guerrero. Para ello se requiere ser usuario de Twitter y tener las credenciales de acceso a la API, la solicitud se envía a dicha red social para obtener: **key**, **secret**, **access_token** y **access_secret**. Una vez que se tienen las claves de acceso, se requiere el usuario o la cuenta de Twitter de la figura pública, y realizar el web scraping para extraer los textos, estos se guardan en un dataset, para posteriormente realizar el análisis.

La extracción de datos (web scraping, en inglés) se realizó con R, con la librería **rtweet** (Package 'rtweet', 2023), y de ésta se usan dos funciones: **get_timeline** y **search_tweets**. Estas dos funciones facilitan la interacción y utilizan **endpoints** del API de Twitter.

La conexión a la API de Twitter se realizó por medio de los siguientes comandos:

```
apppname <- "XXXXXXXXXX"
key    <- "XXXXXXXXXXXXXXXXXXXX"
secret <- "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
```

```
access_token <- "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
access_secret <- "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
twitter_token <- create_token(app = appname, consumer_key = key,
                             consumer_secret = secret,
                             access_token = access_token,
                             access_secret = access_secret)
```

El **get_timeline** devuelve la línea de tiempo de cualquier usuario de Twitter (es decir, lo que han twitteado), Package 'rtweet' (2022). Por ejemplo:

```
GTL_MarioM <- get_timeline(user = "@ Mario_M_Arcos", n = 3000, parse = TRUE,
                           check = TRUE, include_rts = TRUE)
```

user	Nombre o nombres de usuarios para realizar la búsqueda.
n	Número de Tweets que regresa, estos son los últimos publicados.
parse	TRUE devuelve un data-frame.
include_rts	TRUE, incluye los re-tweets.

El **search_tweets** hace la búsqueda de tweets que cumplan un filtro específico (query), por ejemplo, uno o varios hashtags, usuarios, etc., y tiene un límite de 500 caracteres de longitud, Package 'rtweet' (2022).

```
ST_MarioM <- search_tweets("Guerrero OR MMA OR MarioMorenoArcos OR
UAGro", n = 100000, include_rts = F, parse = TRUE, retryonratelimit = TRUE, lang
= 'es')
```

La tabla 1 muestra los parámetros que se usaron para la obtención del dataset. En la primera columna *Aspirante* se describe por sí misma, la columna *get_timeline* es el nombre del usuario del aspirante y del cual el API hace la búsqueda, y por último, en *search_tweets* se agregaron palabras clave relacionadas con el aspirante para precisar la búsqueda. El dataset que devuelve R contiene en total 90 variables.

Tabla 1*Parámetros que se usaron en la búsqueda*

Aspirante	get_timeline	search_tweets
Marcos Parra	@marcoseparra	Marcos+Parra OR Marcos+Efren+Parra+Gomez OR marcoseparra OR #PiensaEnTaxco OR #TaxcoHistoriaConFuturo OR #HistoriaConFuturo
Mario Moreno	@Mario_M_Arcos	Mario+Moreno OR Mario+Moreno+Arcos OR Mario_M_Arcos OR #UnCuartoMas OR #GuerreroNosNecesitaTodos OR #MMA OR #MarioMorenoArcos
Evodio Velázquez	@EvodioVelazquez	Evodio+Velazquez OR Evodio+Velazquez+Aguirre OR EvodioVelazquez OR #SiQueremosPodemos OR #FADG OR #DalePlay
Félix Salgado	@FelixSalMac	Felix+Salgado OR Felix+Salgado+Macedonia OR FelixSalMac OR #Acapulco+CostaGrande+4T OR #Acapulco+4taTransformacion

Nota: tabla de elaboración propia.

Creación de la base de datos

Para cada candidato se generan dos data sets, uno del `get_timeline` y otro del `search_tweets`, esto es, en total tenemos 8 archivos con la misma estructura de datos los cuales se unieron en un solo dataset con 59,175 registros. El archivo se generó en formato CSV.

Creación y definición de nuevas variables

La variable más importante es **text**, que contiene el tweet publicado. El `screen_name` contiene el usuario quien publicó o re-tweeteo. La característica **Aspirante** contiene el nombre del candidato al que hace referencia el tweet. Se creó la variable *Población* para identificar los tweets del público en general de los tweets de la figura pública, con los códigos 0, 1 respectivamente.

2.2 Preprocesamiento de la base de datos

Como parte del preprocesamiento del dataset previo a la modelación se realizaron las siguientes actividades:

Normalización y limpieza de datos

Se aplicará la limpieza y la normalización de texto con:

- Solo dejar caracteres alfanuméricos y espacio.
- Aplicar NFD, Normal Form Decomposition, `unicodedata.normalize('NFD')`.
- Minúsculas y sin acentos.
- Quitar las URL, @screenname y Hashtags.

Identificación y eliminación de registro duplicados

```
df.duplicated().sum()
```

executed in 879ms, finished 19:36:18 2023-03-13

1533

```
#Eliminar duplicados
```

```
df.drop_duplicates(inplace=True)
```

executed in 927ms, finished 19:36:23 2023-03-13

```
df.duplicated().sum()
```

executed in 700ms, finished 19:37:26 2023-03-13

0

Se detectaron 1533 registros duplicados los cuales fueron eliminados.

Texto antes del procesamiento

```
df['text'].head(10)
```

executed in 24ms, finished 19:47:12 2023-03-13

```
0    Descanse en paz Don Humberto, un hombre íntegr...
1    @Mario_M_Arcos Estimado amigo agradezco las m...
2    Muchas gracias, Señor Gobernador @HectorAstudi...
3    Abrazo fraterno Lazaro para ti y toda tu famil...
4    @Mario_M_Arcos Muchas gracias estimado Mario. ...
5    Muchas felicidades a la ANAC, por estos 17 añ...
6    @HectorAprezaSI Te deseo mucha suerte estimado...
7    Buena referencia sobre @TaxcoGob @HectorAstudi...
8    @HectorAstudillo @TaxcoGob Muchas gracias Gobe...
9    @HectorAstudillo En medio del auto confinamien...
Name: text, dtype: object
```

Texto después del procesamiento

```
df['text_clean']=df['text'].map(lambda x:text_clean(str(x).lower()))|
```

executed in 1.19s, finished 19:47:45 2023-03-13

```
df['text_clean'].head(10)
```

executed in 16ms, finished 19:47:59 2023-03-13

```
0    descanse en paz don humberto un hombre integro...
1    mario m arcos estimado amigo agradezco las mue...
2    muchas gracias senor gobernador hectorastudill...
3    abrazo fraterno lazaro para ti y toda tu famil...
4    mario m arcos muchas gracias estimado mario ab...
5    muchas felicidades a la anac por estos 17 anos...
6    hectoraprezasi te deseo mucha suerte estimado ...
7    buena referencia sobre taxcogob hectorastudill...
8    hectorastudillo taxcogob muchas gracias gobern...
9    hectorastudillo en medio del auto confinamient...
Name: text_clean, dtype: object
```

Imputación de valores perdidos

Todos los tweets en el dataset tienen algún texto, pero se verificó que no haya valores nulos en esta variable.

```
df['text'].isna().sum()
```

executed in 25ms, finished 19:34:33 2023-03-13

```
0
```

Se verificó que hubiera tweets con al menos 6 caracteres

```
df[~df["text_clean"].map(lambda x:len(x)>5)].reset_index(drop=True)|
```

executed in 36ms, finished 12:14:37 2023-03-14

user_id	status_id	created_at	screen_name	text	source	display_text_width	reply_to_status_id
---------	-----------	------------	-------------	------	--------	--------------------	--------------------

0 rows × 92 columns

Identificación de ruido

Podemos identificar como ruido la repetición de letras: “bueeeenos días” estas se pueden reducir y dejar solo 2 caracteres repetidas.

```
aux = df['text_clean'].sample(10)
aux
```

executed in 17ms, finished 20:10:15 2023-03-13

```
43322          mario m arcos un saludoooo u 0001f495
31444    en verdad no entiendo como es que tenemos que ...
14972    azucenau los alcahuetes de la 4t son capaces d...
12467    traduccion felix salgado macedonio podra ser u...
47756    dos denuncias de felix salgado macedonio podra...
43988    mario delgado aseguro que con laydasansores mo...
2999     seguimos atentos las lluvias continuaran http ...
45620    u 0001f4cd merida es el escenario del eclipsed...
8614          felixsalmac lopezobrador violador
38013    felixsalgadomac eres un peligro para las mujer...
Name: text_clean, dtype: object
```

```
df['text_clean'][43322]
```

executed in 24ms, finished 23:02:05 2023-03-13

```
'mario m arcos un saludoooo u 0001f495'
```

```
df['text_clean']=df['text_clean'].map(lambda x:normalize_text(str(x)))
```

executed in 4.74s, finished 23:03:08 2023-03-13

```
df['text_clean'][43322]
```

executed in 24ms, finished 23:03:15 2023-03-13

```
'mario m arcos un saludoo u 001f495'
```

Stopwords

Se eliminaron Stopwords en español, porque uno de los paquetes que se utilizan para obtener los sentimientos y emociones utiliza un diccionario de léxico, y en este caso es más funcional quitar los Stopwords porque son palabras que no aportan a la identificación de sentimientos.


```
esp_stopwords = stopwords.words("spanish")
```

executed in 16ms, finished 12:15:18 2023-03-14

```
esp_stopwords = esp_stopwords+['http','https']
```

executed in 14ms, finished 12:15:24 2023-03-14

```
df["text_clean"] = df["text_clean"].map(lambda text: " ".join([x for x in text.split(" ") if x not in esp_stopwords]))
```

executed in 3.09s, finished 12:16:01 2023-03-14

```
df['text_clean'].sample(10)
```

executed in 18ms, finished 12:16:23 2023-03-14

```
49738    lgv22 brunogarza felixsalmac lopezobrador deci...
9086     laurabarranco felixsalmac lopezobrador chole s...
40526    presidente defiende violadores representa pres...
52096    felix salgado sera candidato guerrero morena p...
12137    cirogomezl felixsalmac lopezobrador deben reti...
8519     julioastillero felipecalderon lopezobrador fel...
18555    bustosdannyy vromog felixsalmac si supiera lee...
38249    morena niega cumplir lineamientos 3 3 vs viole...
8276     julioastillero alecasas123 lopezobrador felixs...
4973     llegara tri u 001f915 u 274c romario ibarra su...
Name: text_clean, dtype: object
```

Generación de tokens

Se generan los tokens de cada tweet como entrada de datos al paquete Syuzhet (Isasi, J., 2021)., más adelante se explican sus características.

```
df["text_clean"] = df["text_clean"].map(lambda x: x.split())
```

executed in 269ms, finished 12:19:01 2023-03-14

...

```
df['text_clean'].sample(10)
```

executed in 25ms, finished 12:19:25 2023-03-14

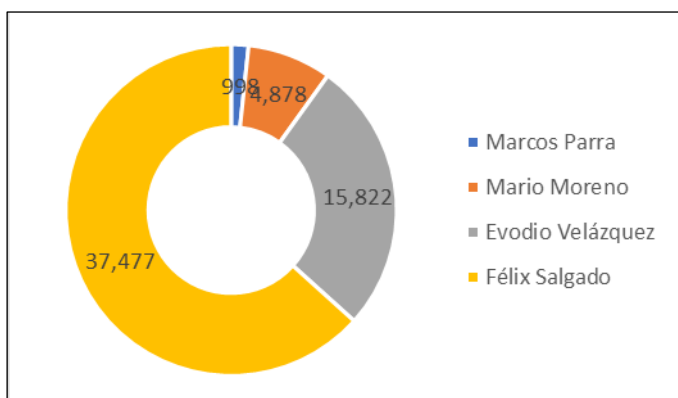
```
18222    [video, apagones, confirmacion, felix, salgado...
25188    [cria, fama, si, felix, salgado, macedonio, ll...
52187    [historiapedro, lalo1101, cnhj, morena, felixs...
50300    [marcela, feraud, felixsalmac, partidomorenamx...
19214    [77500ross, pornografia, infantil, martibatres...
29792    [solo, revise, deseche, intento, candidatura, ...
52364    [edfrancotweets, paolasaeb, ocasion, electores...
50430    [morena, respalda, felix, salgado, macedonio, ...
48167    [protestan, feministas, sede, nacional, morena...
52112    [felixsalmac, borracho, parrandero, violador, ...
Name: text_clean, dtype: object
```

2.3 Análisis exploratorio de los datos

Más del 60% de los 59 mil tweets hablan del aspirante Félix Salgado, podríamos pensar que es el más popular de los cuatro aspirantes como se observa en la figura 1. A principios de 2021 el señor Salgado fue acusado de abuso sexual, y fue muy sonado en los medios de comunicación.

Figura 1

Número de Tweets donde se menciona al aspirante.

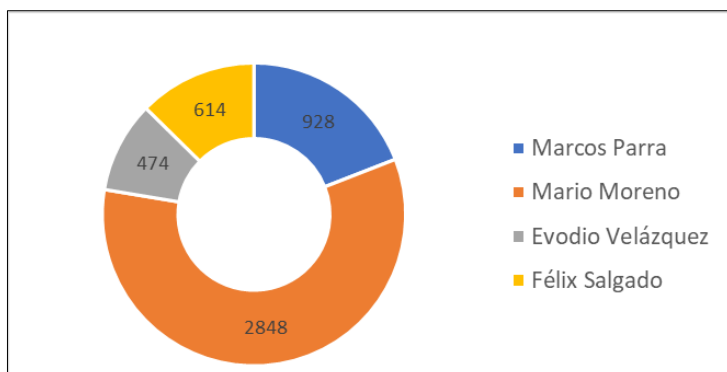


Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Considerando solo los tweets que publican los aspirantes se nota, en la figura 2, que el más activo en esta red social es Mario Moreno Arcos, y los otros tres aspirantes están muy por debajo de lo que publicó el señor Moreno.

Figura 2

Número de tweets publicados por aspirante.

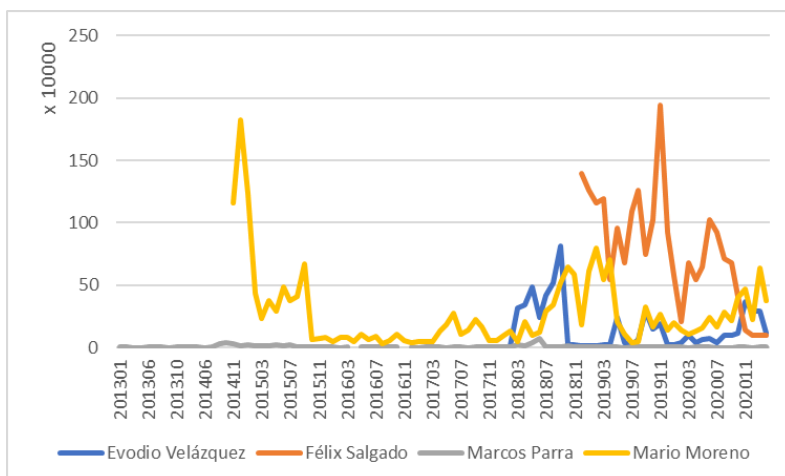


Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

En la figura 3 se muestra que en 2015 Mario Moreno es el aspirante con más seguidores, pero fue desbancado por Félix Salgado en 2019. En algún momento del 2019 Félix Salgado tuvo muchos seguidores, pero para finales de 2020 bajaron. Marcos Parra quien inició desde 2013 en esta red, es el que menos seguidores ha tenido.

Figura 3

Número de seguidores en el tiempo por aspirante.

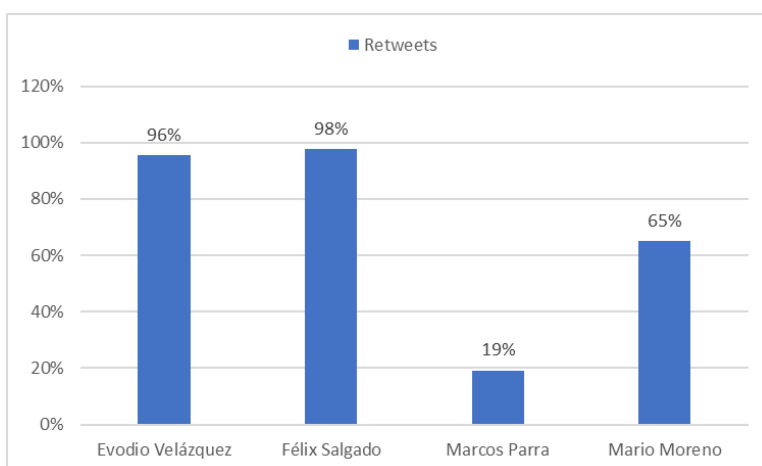


Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

De los tweets publicados por los aspirantes los de Mario Moreno son los más retwitteados, **al menos una vez**, como se muestra en la figura 4. Marcos Parra es el segundo aspirante que más twittea, pero no tiene mucho eco sus publicaciones.

Figura 4

Porcentaje de tweets retweeteados por aspirante.

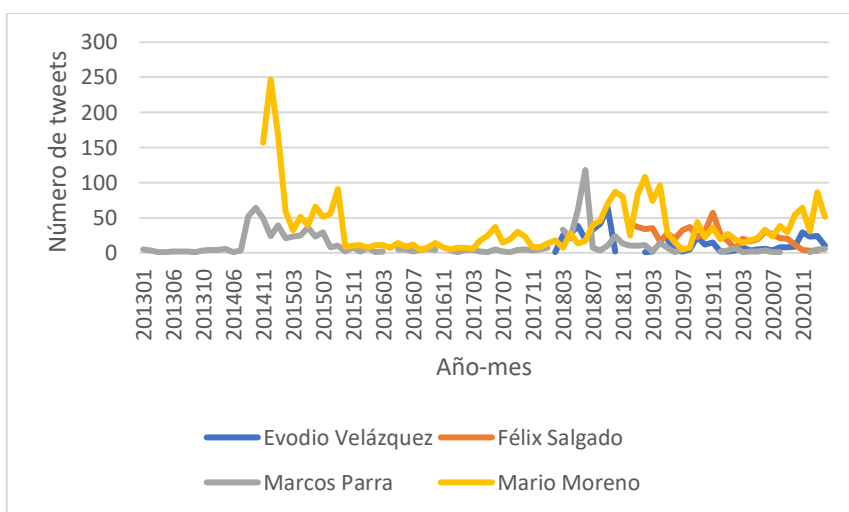


Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

En la figura 5 se observa que Marcos Parra es el aspirante que primero inicio el uso activo de Twitter, seguido de Mario Moreno. Al inicio de 2021, año de la elección de gobernador, no se observa un incremento en el uso de esta red social, solamente por Mario Moreno.

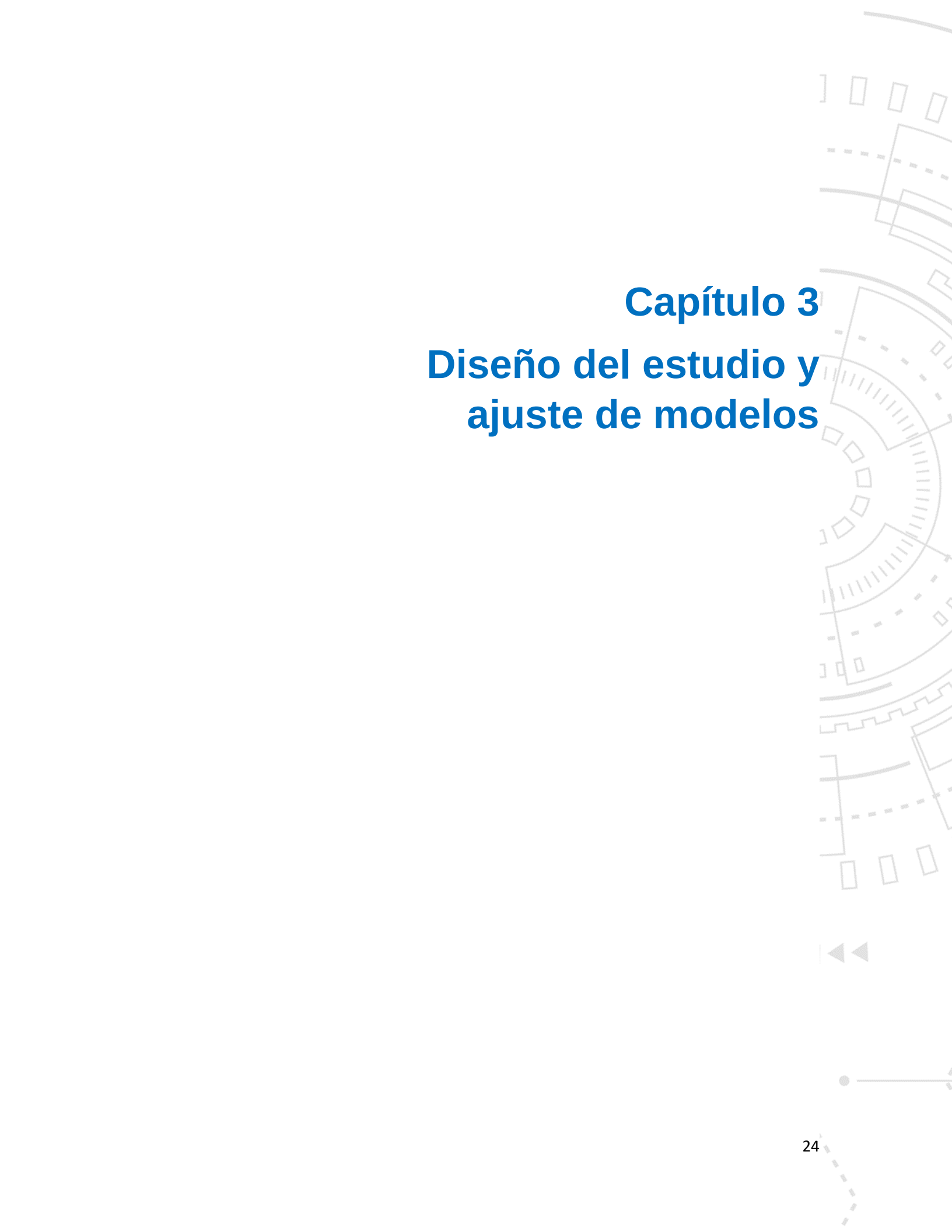
Figura 5

Número de tweets publicados en el tiempo por mes por aspirante.



Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

De las figuras anteriores se observa que hay grandes diferencias entre aspirantes, por ejemplo, en el número de tweets publicados por ellos, los publicados acerca de ellos y el número de seguidores.



Capítulo 3

Diseño del estudio y ajuste de modelos

Capítulo 3. Diseño del estudio y ajuste de modelos

El marco teórico de un proyecto de investigación proporciona la base conceptual y teórica sobre la cual subyace el estudio. La revisión de la literatura y la explicación de las teorías y conceptos relevantes ayudan a contextualizar y justificar la investigación, así como a definir las preguntas de investigación y las variables clave que se explorarán en el estudio. El marco metodológico el cual tiene como propósito establecer cómo se llevará a cabo la investigación, esto es, la estrategia para obtener la información y las actividades para responder a los objetivos definidos (Lerma, 2009).

En este capítulo se mostrarán los modelos utilizados y sus ajustes, y por último se mostrarán el análisis de los resultados.

3.1 Marco teórico

Redes sociales y Twitter

Redes sociales como plataforma de comunicación

Inicios de las redes sociales. Antes de las redes sociales solo había correos electrónicos, páginas web y foros. Las primeras redes fueron: SixDegrees, que afirmaba que es posible conectar con cualquier otra persona del mundo en tan solo 6 pasos. Entre 2002 y 2003 surgieron algunas redes como: Friendster, MySpace y LinkedIn (marketing4ecommerce, 2022).

Dos de las principales redes sociales actualmente son Facebook y YouTube. En 2004 Mark Zuckerberg, un estudiante de la Universidad de Harvard, creó la red social que hasta este momento es la más importante, Facebook. Actualmente, tiene más de 2,958 millones de usuarios activos al mes. En 2005 se creó YouTube, y es una red social con 2,514 millones de usuarios activos al mes.

Twitter surgió en 2006 y es la red social del micro blog. Hoy en día es tal su impacto que medios de comunicación como noticieros, televisión y radio dedican tiempo para comentar las tendencias o el impacto de algún tweet. Esta red tiene aproximadamente 556 millones de usuarios activos al mes.

Muchas otras redes siguen apareciendo y con mucha fuerza como Tik Tok, Instagram, Pinterest, WhatsApp, etc.

Twitter es una plataforma de microblogging, es decir, la comunicación es por medio de mensajes cortos. (Webempresa, 2018), Developer Platform, Twitter (2022). Algunas de las características de Twitter son: mensajes o tweets de 280 caracteres, esta limitación fomenta la concisión y la comunicación directa. Seguir y ser seguido esto crea una red de seguidores y permite a los usuarios personalizar su flujo de contenido. Los usuarios pueden reeditar (compartir) los tweets de otros en su propia línea de tiempo y, como muchas otras redes sociales, marcar el tweet con "Me gusta" para mostrar su aprecio o acuerdo.

Twitter hizo famoso el uso de hashtags (#), para etiquetar palabras clave o temas en los tweets. Los hashtags permiten a los usuarios encontrar y unirse a conversaciones específicas. Tik Tok, Instagram, Facebook, LinkedIn son redes que usan el hashtag. Twitter se actualiza constantemente, lo que es útil para obtener noticias y actualizaciones sobre eventos, tendencias y temas de interés en corto tiempo.

Una función que se hizo popular en las redes sociales, Facebook, YouTube, Tik Tok y Twitter, es la transmisión en vivo. En Twitter se llama "Periscope", lo que brinda a los usuarios la posibilidad de transmitir eventos en tiempo real y participar en discusiones en vivo. Por otro lado, los usuarios pueden enviar mensajes directos privados a otros usuarios, lo que permite la comunicación privada fuera del ámbito público de los tweets.

Los usuarios pueden crear listas personalizadas para organizar a las personas a las que siguen en grupos específicos, lo que facilita el seguimiento de diferentes temas o intereses.

Twitter cobra por la verificación de las cuentas con el propósito de ayudar a los usuarios a distinguir las cuentas auténticas de las falsas. Los “bots” son programas automatizados que simulan interacción humana en las plataformas de redes sociales.

Twitter, es considerada como la red de las celebridades y las figuras públicas que pueden interactuar directamente con sus seguidores o detractores, y permite una mayor cercanía entre estas personalidades y sus audiencias. Además, muestra las tendencias actuales y populares a nivel mundial y local, lo que permite a los usuarios mantenerse al tanto de los temas relevantes en cualquier momento.

Importancia de Twitter como fuente de datos para el análisis de sentimientos, (QuestionPro, 2023). Twitter con millones de usuarios activos, que generan una gran cantidad de datos constantemente de opiniones y emociones de personas y figuras públicas. Estos datos permiten seguir reacciones instantáneas a eventos como los electorales, lanzamientos de productos, crisis o desastres naturales.

No hay límite para los temas a discutir, por lo que los usuarios de Twitter comentan una amplia variedad de temas, desde política y deportes hasta entretenimiento y tecnología. Esto favorece el análisis de sentimientos en una amplia gama de dominios y sectores. Y aunque hay un límite en la longitud de los tweets, los usuarios aún pueden expresar sus opiniones y emociones de manera efectiva. Además, el uso de emojis, hashtags y menciones agrega niveles de información emocional que pueden analizarse.

La mayoría de los tweets son públicos y accesibles, lo que facilita la recopilación y el análisis de datos, sin embargo, hay un límite de consultas diarias y la consulta

masiva tienen un costo de extracción. También, es importante considerar la privacidad y ética al utilizar datos de Twitter.

Twitter ofrece una API que permite a los analistas de datos acceder a tweets en tiempo real y a datos históricos. Esto facilita la recopilación y el análisis de datos para proyectos de análisis de sentimientos. Pero hay límites de extracción de tweets por día. Twitter se ha utilizado en estudios con fines académicos y también con propósitos de marketing para comprender las tendencias de opinión, evaluar la satisfacción del cliente, realizar análisis políticos y mucho más.

Actualmente existen varias aplicaciones para la obtención de análisis de sentimientos (la mayoría para textos en inglés) que pueden procesar automáticamente grandes cantidades de tweets para extraer información emocional y de opinión.

Análisis de sentimientos

Fundamentos del análisis de sentimientos

El análisis de sentimientos (Sarduy et al., 2020), en el enfoque de la minería de texto y el procesamiento de lenguaje natural (PLN), pretende comprender y determinar las emociones, sentimientos y opiniones expresados en un texto, y tiene como objetivo identificar si el texto transmite sentimientos positivos, negativos o neutros, y la posible clasificación de las emociones como la felicidad, enojo, tristeza o miedo. Algunos conceptos clave relacionados con el análisis de sentimientos se describen a continuación:

El texto es el elemento principal de entrada y se puede obtener de redes sociales, reseñas en el internet de productos y servicios y de mensajes de aplicaciones. La polaridad se define como el sentimiento del texto en positivo, negativo o neutro, para determinar si el sentimiento expresado es favorable, desfavorable o neutral. La emoción a diferencia de la polaridad, se puede especificar las emociones,

como la alegría, la tristeza, el miedo, el enojo, la sorpresa, entre otras. Esto permite un análisis más detallado de las emociones expresadas en el texto.

Palabras que expresan sentimientos son previamente clasificadas para crear un diccionario de sentimientos y son recursos léxicos que asignan valores de polaridad y emociones a las palabras. Estos diccionarios se utilizan para el análisis de sentimientos automatizado, permitiendo evaluar cómo las palabras contribuyen al sentimiento general de un texto. Syuzhet, una de aplicaciones utilizadas en este proyecto, utiliza 4 diccionarios de sentimientos.

La opinión pública se puede medir sobre varios temas, productos o servicios vía análisis de sentimientos. Esta es una aplicación por parte de empresas que desean comprender la satisfacción del cliente o para los políticos que buscan evaluar la percepción pública sobre campañas electorales o sobre la gestión de su administración. En la minería de redes sociales se explotan los datos en estas plataformas para medir las opiniones, sentimientos y emociones expresadas, con el propósito de comprender tendencias y opiniones de usuarios.

El análisis de sentimientos de reseñas y comentarios (reviews, en inglés) permite identificar áreas de mejora y a tomar decisiones basadas en la retroalimentación del cliente. Amazon y Mercado Libre usan reseñas para evaluar los productos y servicios que ofrecen en sus plataformas.

Técnicas de análisis de sentimientos

Las técnicas de análisis de sentimientos las podemos clasificar en dos: estrategias basadas en **aprendizaje automático** (machine learning) y estrategias basadas en **diccionarios** (lexicon-based) (Gonzalo, S., 2020).

En las estrategias de machine learning podemos clasificar con modelación supervisada, o no supervisada. En el primer caso tenemos un conjunto de datos

con sentimientos previamente clasificados y deseamos entrenar un modelo que permita predecir, con cierta precisión, futuros casos. En la modelación no supervisada no se tiene una clasificación previa de los textos y solo podemos agruparlos en función de sus características e identificar patrones entre grupos similares.

Enfoques basados en diccionarios. En esta técnica se utilizan diccionarios de sentimientos previamente clasificados, ésta se realiza manualmente y en forma masiva, por ejemplo, el paquete **Syuzhet** trabaja con cuatro diccionarios de sentimientos: Bing, Afinn, Stanford y NRC.

Figuras Públicas en Twitter

Twitter es usado para compartir noticias, anuncios y eventos en tiempo real, permitiendo a sus seguidores estar al tanto de su carrera, proyectos, causas o campañas políticas. Además, Twitter permite la interacción directa con la audiencia, mejorando la cercanía y confianza con sus seguidores (Gestion.Org, 2019). Por ejemplo, contestando preguntas sobre temas personales, profesionales o políticos.

Las figuras públicas pueden utilizar Twitter para dar explicaciones o abordar una situación directamente con el objetivo de gestionar la percepción pública. Algunas figuras públicas utilizan la interacción con detractores como una estrategia arriesgada para defender su punto de vista. Otra herramienta es el uso de hashtags y unirse a tendencias populares para generar una mayor visibilidad y participación de las figuras públicas.

Muchos políticos y líderes utilizan Twitter para comunicarse con sus seguidores políticos y el público en general. Por este medio comparten sus ideas, puntos de vista y su discurso político, también, pueden conectarse con otros líderes,

celebridades y expertos en su campo, enriqueciendo las ideas y el discurso político.

3.2 Marco metodológico

El marco metodológico tiene como propósito establecer cómo se llevará a cabo la investigación, esto es, la estrategia para obtener la información y las actividades para responder a los objetivos definidos (Lerma, 2009).

Tipo de investigación

El propósito de este estudio es descriptivo en el sentido de analizar los sentimientos y emociones de los tweets publicados de dos poblaciones: figuras públicas y público en general, y la concordancia entre ambos. En el conjunto de datos, no se tiene información de quiénes son seguidores o detractores con respecto a la figura pública, a lo más, se tiene una variable que identifica a qué figura pública se refiere el tweet y otra variable que indica quién publicó el tweet. Adicional, se tiene la fecha de publicación de los tweets de 2013 a febrero 2021.

Unidad de análisis

La unidad de análisis son los tweets publicados en esta plataforma relacionados con cuatro figuras públicas, que a inicios del 2021 eran aspirantes a la candidatura de gobernador por el estado de Guerrero: Mario Moreno Arcos, Evodio Velázquez Aguirre, Félix Salgado Macedonio y Marcos Efrén Parra Gómez.

Muestra

Por medio de la API de Twitter se extrajeron más de 59 mil tweets relacionados con los cuatro aspirantes previamente mencionados. El conjunto de datos contiene información de tweets publicados de 2013 a febrero de 2021.

Variables

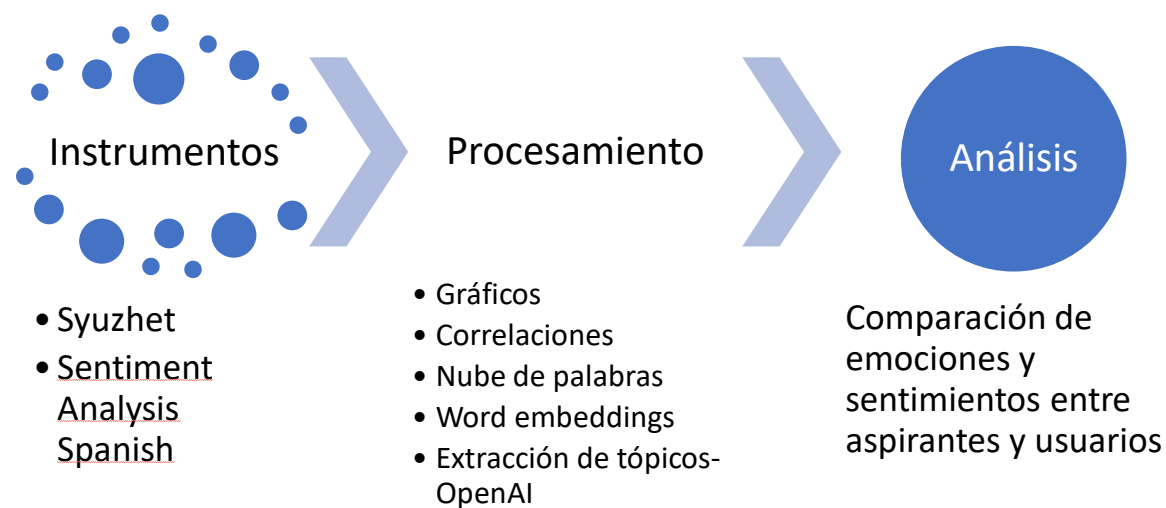
El conjunto de datos de los tweets contiene una variable/característica denominada “text”. Otra variable denominada “screen_name” contiene el usuario que publicó o re-tweeteo. La característica “Aspirante” contiene el nombre del candidato al que hace referencia el tweet. Otras variables de interés son: followers_count, que es el número de seguidores del aspirante; retweet_count, que es número de veces que el tweet fue re- retweet; location, que es el lugar donde se originó el tweet y fecha en que se publicó.

Diseño de investigación

Una vez que se procesó el conjunto de datos, como se describió en el capítulo 2, las fases de desarrollo del diseño de investigación que se siguieron iniciaron con aplicación de instrumentos como Syuzhet y sentiment análisis spanish (pypi.org, 2021); posteriormente el procesamiento de datos cuyos resultados se exponen en gráficos, correlaciones, nubes de palabras, word embedding y extracción de tópicos-OpenAI; y finalmente el análisis de la información, ver figura 6.

Figura 6

Fases de desarrollo del diseño de investigación



Nota: figura de elaboración propia.

Instrumentos

Se usaron dos paquetes para la extracción de sentimientos y emociones que **procesan textos en español**: Syuzhet y Sentiment-analysis-spanish (SAS). El primero genera la polaridad de sentimientos positivo y negativo, y 8 emociones: enfado, anticipación, asco, miedo, alegría, tristeza, sorpresa y confianza, Syuzhet (2021). La segunda aplicación, Sentiment-analysis-spanish, (pypi.org, 2021), proporciona una polaridad de sentimientos en un rango de 0 a 1, donde valores cercanos a 1 significa más positivo, y próximos a cero negativos. Para entrenar este modelo, el desarrollador, utilizó librerías de Tensor Flow y Keras, aplicando redes neuronales convolucionales.

Procesamiento

De las emociones y sentimientos se generaron gráficos, correlaciones entre las poblaciones y las emociones, nube de palabras y word embeddings, estos últimos para la extracción de tópicos con OpenAI, y se compararon los tweets de las figuras públicas contra los tweets de los usuarios.

Para la obtención de word embeddings se usó las siguientes características de OpenAI (OpenAI Embeddings, 2023):

- Model: text-embedding-ada-002
- Encoding: cl100k_base
- Dimensión del vector: 1024

Y para para la extracción de los tópicos (OpenAI, Create Chat Completion, 2024):

- `openai.Completion.create()`
- Engine: text-davinci-003

Análisis

Con los resultados obtenidos del procesamiento se procederá a realizar el análisis con el propósito de comprobar si existe o no existe concordancia entre los tweets de las figuras públicas y lo publicado por los usuarios de Twitter.

3.3 Ajuste de los modelos

Obtención de las emociones y sentimientos por aplicación

Syuzhet

Este paquete trabaja con cuatro diccionarios de sentimientos: Bing, Afinn, Satnford y **NRC**. Este último, es el que se utilizó en este proyecto, porque es el único disponible en varios idiomas, **incluido el español**. El vocabulario genera valores de sentimiento negativo o positivo y ocho emociones, es decir, para cada categoría se genera una variable con valores de 0 y hasta 4. “Si hay un valor mayor a 0 quiere decir, primero, que dicho término existe en el diccionario NRC y, segundo, que tiene un valor asignado para alguna emoción y/o sentimiento” (Isasi, 2021). Este sistema ha sido desarrollado por Saif M. Mohammad, científico del Consejo de Investigación Nacional de Canadá (NRC, por sus siglas en inglés).

El léxico de Syuzhet ha sido construido manualmente mediante la técnica de Escalamiento por Máxima Diferencia (MaxDiff), que evalúa la preferencia en una serie de alternativas (Mohammad y Turney). El léxico cuenta con 14,182 unigramas (tokens) con las categorías de sentimientos positivo y negativo y las emociones de enfado, anticipación, disgusto, miedo, alegría, tristeza, sorpresa y confianza, y está disponible en más de cien idiomas (previa traducción automática) (Isasi, 2023). Para cada sentimiento y emoción se genera una característica en el conjunto de datos.

Isasi (2021) hace algunas advertencias del uso de este paquete:

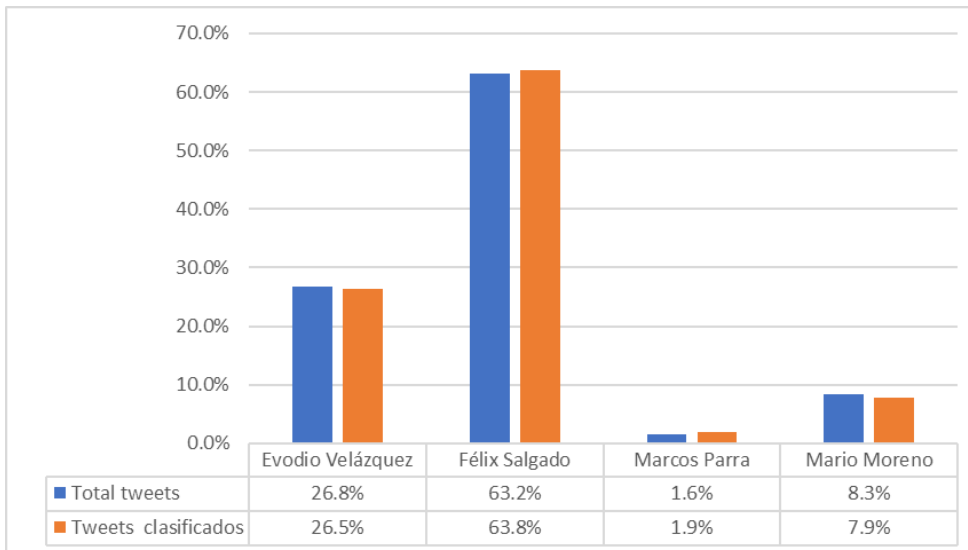
El léxico en español es una traducción directa realizada por traducción automática (estos sistemas son ya muy fiables entre inglés y español, pero no así en otras lenguas que NRC dice poder analizar como, por ejemplo, el euskera). La puntuación de cada palabra, es decir, la valencia sentimental y emocional, tiene un sesgo cultural y temporal que debes tener en cuenta y
3un término que a los participantes en la encuesta les pareciera positivo a ti

puede parecerle negativo. No se recomienda el uso de esta metodología en textos con gran carga metafórica y simbólica. El método no captará la negación de un sentimiento positivo como, por ejemplo, la frase “no estoy feliz”

El conjunto de datos de este proyecto contiene 59,175 tweets de los cuales sólo 11,506 pudieron ser clasificados por Syuzhet. La figura 7 muestra el porcentaje de tweets clasificados por Syuzhet y el porcentaje del total de tweets por aspirante.

Figura 7

Porcentaje de tweets totales y tweets clasificados en sentimientos, por aspirante.



Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

La tabla 2 muestra la prueba estadística de diferencia de proporciones, donde la hipótesis nula (H_0) es que las proporciones de los *tweets clasificados* con respecto a la proporción del *total de tweets* **son iguales** contra la alternativa (H_A) de que son diferentes por aspirante. De acuerdo a los valores p no se rechazan H_0 .

Tabla 2

Pruebas de hipótesis de las proporciones de tweets clasificados contra el total de tweets.

Aspirante	Total tweets	Tweets clasificados	Z	p-valor
Marcos Parra	1.6%	1.9%	- 0.230	0.818
Mario Moreno	8.3%	7.9%	0.427	0.669
Evodio Velázquez	26.8%	26.5%	0.393	0.694
Félix Salgado	63.2%	63.8%	- 0.861	0.389

Nota: tabla de elaboración propia.

Sentiment-analysis-spanish

Análisis de sentimientos en español es “una librería de Python que utiliza redes neuronales convolucionales para predecir el sentimiento de las oraciones en español. El modelo se entrenó a partir de más de 800,000 reseñas de usuarios de las páginas eltener, decathlon, tripadvisor, filmaffinity y ebay. Estas reseñas se extrajeron mediante web scraping con el proyecto opinion-reviews-scraper()” (sentiment-analysis-spanish, 2021).

El modelo se entrenó utilizando las bibliotecas Keras y Tensorflow. Se logró una exactitud (accuracy) de validación con datos no vistos del 88%.

El resultado de esta librería es muy sencillo en comparación con Syuzhet, porque solo devuelve la polaridad del sentimiento, es decir, un valor entre 0 y 1, donde 1 es la probabilidad de que el tweet sea positivo, y las probabilidades cercanas a cero serán tweets negativos, esto implica que valores alrededor de 0.5 se refieren a tweets neutrales.

A diferencia Syuzhet, en este caso se obtuvieron valores para los 59 mil tweets del conjunto de datos.

3.4 Análisis de los resultados

Comparativo entre modelos

El objetivo de este estudio es comparar la concordancia de los sentimientos y emociones entre las figuras públicas y el público en general, pero resulta interesante comparar los resultados entre los dos paquetes utilizados: Syuzhet y SAS. La comparación no es tan directa porque con Syuzhet se pueden calcular los porcentajes de tweets positivos y negativos, y con SAS solo podemos calcular el promedio de la polaridad de los tweets, es decir, el promedio indica que tan positivo o negativo son los tweets.

Las tablas 3 y 4 muestran el porcentaje de sentimientos y el promedio de la polaridad de tweets publicados por **figuras públicas** y **público en general**, respectivamente, por modelo y por aspirante. En la tabla 3, al comparar resultados entre modelos se observa que Marco Parra y Mario Moreno expresan más sentimientos positivos en sus tweets, 75% en promedio, y coincide con un resultado mayor de la polaridad, 0.364 y 0.382. En el caso de Félix Salgado el porcentaje de tweets positivos (64.9%) es menor y también el promedio de la polaridad, es decir, en estos tres casos hay cierta concordancia entre modelos, pero para Evodio Velázquez no ocurre esto, porque el porcentaje de tweets positivos es alto (74.8%) y el promedio de la polaridad es el más bajo de los 4 aspirantes, 0.261. En las gráficas y en las tablas, los porcentajes se refieren a resultados de Syuzhet y los promedios a los de SAS.

Tabla 3

*Porcentaje y promedio de sentimientos de tweets publicados por **figuras públicas** por modelo y por aspirante.*

Aspirante	Syuzhet		Sentiment Analysis Spanish
	Negativo	Positivo	Promedio Polaridad
Evodio Velázquez	25.2%	74.8%	0.261
Félix Salgado	35.1%	64.9%	0.284
Marcos Parra	24.0%	76.0%	0.364
Mario Moreno	25.1%	74.9%	0.382

Nota: tabla de elaboración propia.

En la tabla 4, se observan porcentajes más bajos del sentimiento positivo en comparación con lo observado en la tabla 3, y lo mismo se observa con el promedio de la polaridad, esto implica que las figuras públicas expresan tweets

más positivos (en promedio 18%) que el público en general. Al comparar los dos modelos se observa que un promedio 50% de sentimiento positivo está relacionado con una polaridad promedio 0.260, y un aumento del sentimiento positivo de 66.9% se refleja en un aumento de la polaridad de 0.361. En este caso se observa una concordancia entre los resultados de los dos modelos.

Tabla 4

*Porcentaje y promedio de sentimientos de tweets publicados por **público en general** por modelo y por aspirante.*

Aspirante	Syuzhet		Sentiment Analysis Spanish
	Negativo	Positivo	Promedio Polaridad
Evodio			
Velázquez	50.0%	50.0%	0.268
Félix Salgado	45.9%	54.1%	0.236
Marcos Parra	50.0%	50.0%	0.274
Mario Moreno	33.1%	66.9%	0.361

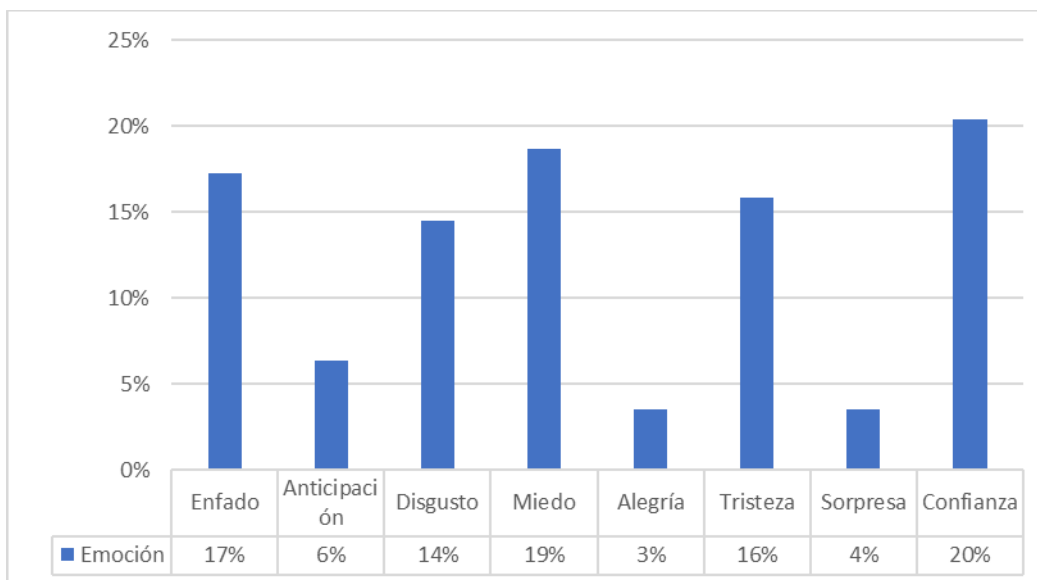
Nota: tabla de elaboración propia,

La siguiente figuras, del 8 al 11, muestra las emociones tomando en cuenta **todos los tweets** clasificados y a todos los aspirantes.

En la figura 8 se observa que las tres emociones que más resaltan son confianza, enfado y miedo, en este orden de importancia, y los sentimientos menos expresados en los tweets son anticipación, alegría y sorpresa.

Figura 8

Porcentaje de emociones expresadas en los tweets.

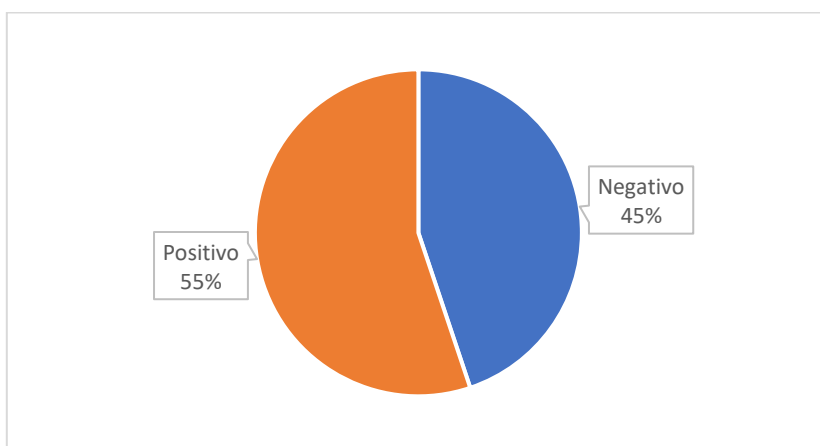


Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

En La figura 9 muestra que el 55% de todos los tweets son positivos, ligeramente mayor que el porcentaje de los tweets clasificados como negativos, podríamos decir que están más o menos balanceados.

Figura 9

Porcentaje de sentimientos expresados en los tweets.

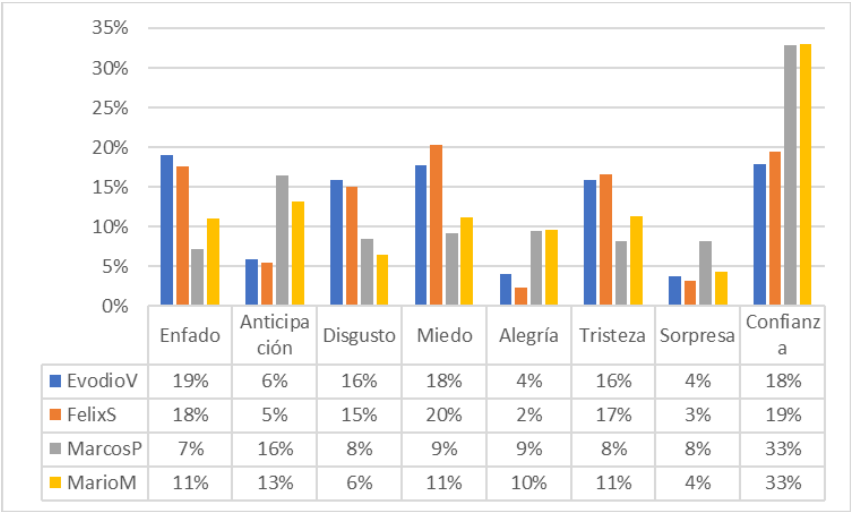


Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Análisis por aspirante

Las figuras 10 y 11 muestran las emociones incluyendo los tweets de los aspirantes y del público en general por el aspirante. Los tweets donde se menciona a Marco Parra y a Mario Moreno expresan más confianza y en segundo lugar anticipación. Las emociones que resaltan en los tweets donde se menciona a Félix Salgado son miedo, confianza y enfado, en contraste con un 2% de Alegría. En el caso de Evodio Velázquez las principales emociones son enfado, confianza y miedo, las mismas que Félix Salgado, como se muestra en la figura 10.

Figura 10
Porcentaje de emociones expresadas en los tweets por aspirante.

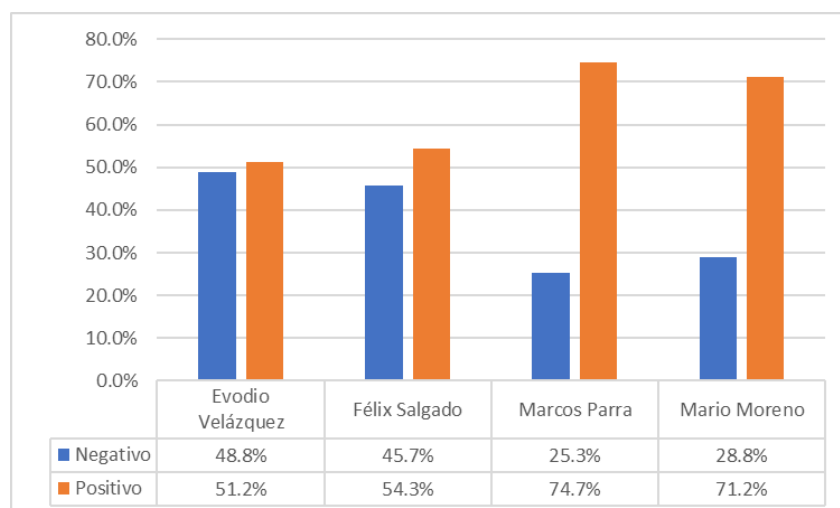


Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

En la figura 11 se aprecia que los tweets donde se mencionan a Evodio V. y a Félix S. tienen casi la misma proporción de sentimientos positivos y negativos, y en los tweets donde se mencionan a Marco P. y a Mario M. los tweets son en proporción más positivos en una relación 3 a 1 con respecto a los negativos.

Figura 11

*Porcentaje de **sentimientos** expresados en los tweets por aspirante.*



Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Análisis de emociones de figuras públicas versus usuarios

La figura 12 muestra los tweets publicados **por los aspirantes** agrupados por emociones y figura pública, y la figura 13 es el mismo caso, pero contiene los tweets expresados por el **público en general**.

Comparación de las emociones expresadas por figuras pública y usuarios

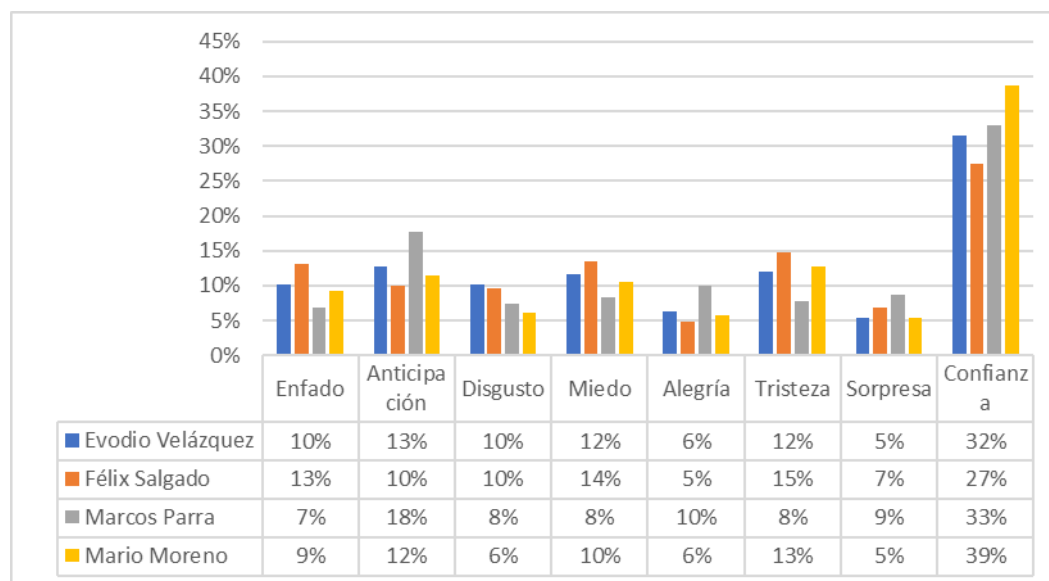
La confianza es la emoción que expresan más las figuras públicas y los usuarios, pero las figuras públicas en mayor porcentaje. La mayor discordancia en la confianza entre ambos la tiene Evodio Vázquez con un 32% contra 17% de los usuarios. En la figura 12, se nota como predomina la emoción de la confianza respecto a las otras emociones, 33% de confianza contra 10% de las otras emociones, mientras que las emociones de los usuarios, figura 13, observa un patrón diferente 23% de confianza contra 11% de las otras emociones.

El comportamiento de las emociones de los tweets es muy parecido entre Evodio V. y Félix S., estos aspirantes muestran más confianza en sus tweets (29.5% en promedio) que lo expresado por los usuarios (18% en promedio).

Los tweets del público en general y los de Marcos Parra tienen casi el mismo nivel de confianza, sin embargo, las emociones de anticipación, disgusto y miedo son muy diferentes, esto es, hay concordancia en la confianza, pero no en las otras emociones. Mario Moreno es el aspirante que más usa esta red social de los 4 aspirantes, y posiblemente se benefició porque el porcentaje de las emociones negativas como enfado, disgusto y miedo del público en general son más bajas.

Figura 12

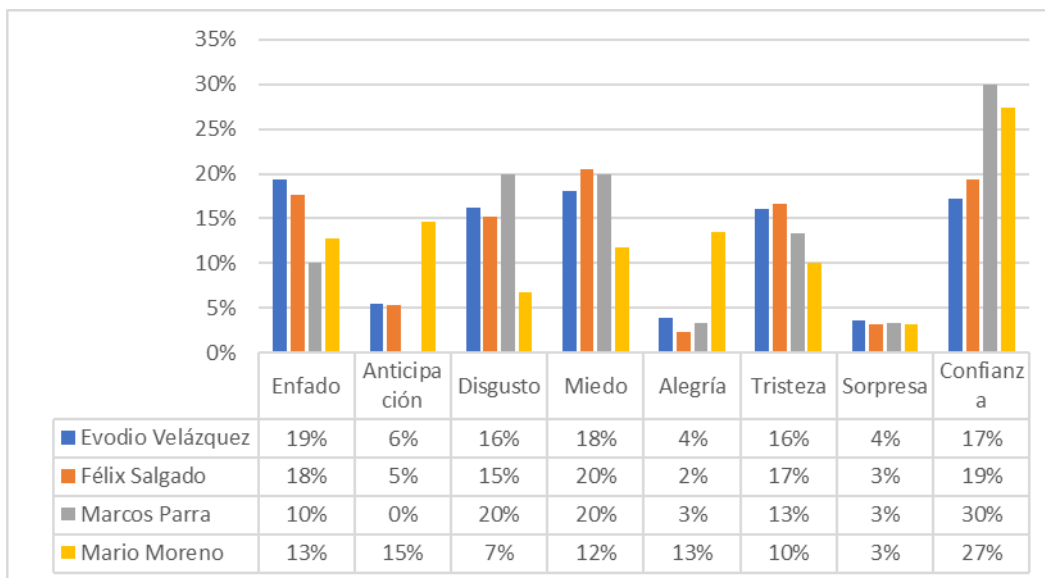
*Porcentaje de **emociones** de los tweets publicados **por los aspirantes** por emociones y figura pública.*



Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Figura 13

Porcentaje de **emociones** de los tweets publicados **por los usuarios** por emociones y figura pública.



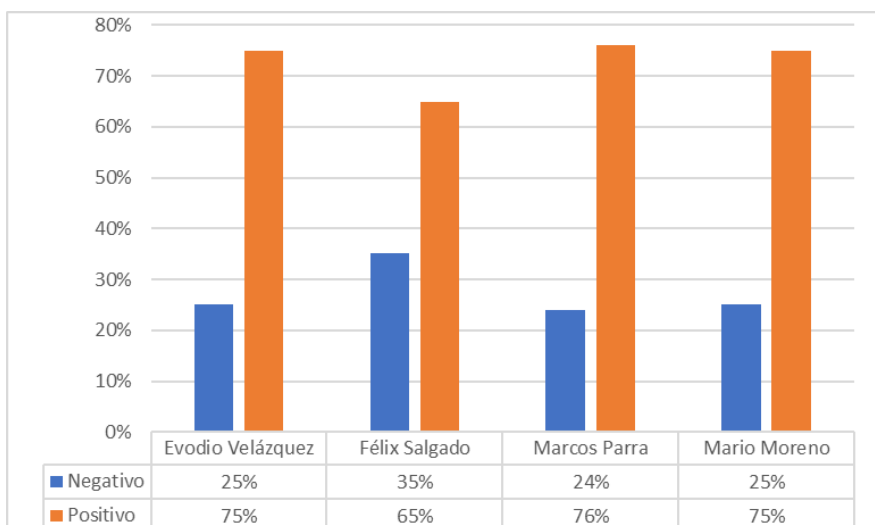
Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Comparación de los sentimientos expresados por figuras pública y usuarios

Al analizar los sentimientos entre aspirantes y usuarios se observa el mismo patrón que en las emociones, es decir, los tweets de las figuras públicas son más positivos que lo que expresa el público en general, figuras 14 y 15, a excepción de Mario Moreno, donde los usuarios expresan un 67% de tweets positivos, en comparación con un 50% de los demás aspirantes, figura 15.

Figura 14

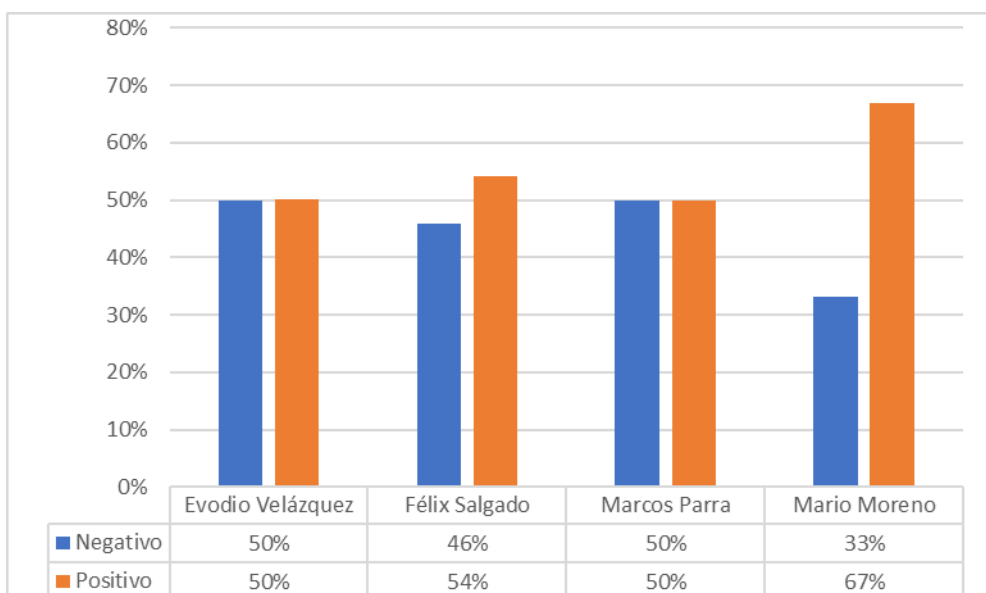
Porcentaje de **sentimientos** de los tweets publicados **por los aspirantes** por sentimiento y figura pública.



Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Figura 15

Porcentaje de **sentimientos** de los tweets publicados **por los usuarios** por sentimiento y figura pública.



Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Análisis de palabras

Como parte del análisis de sentimientos y emociones se analizaron las palabras más comunes entre lo expresado por los aspirantes contra lo mencionado por los usuarios con el objetivo de encontrar concordancia en el discurso entre ambas poblaciones. La tabla 5 muestra las 5 palabras más mencionadas para los sentimientos positivo y negativo de cada aspirante, por figura pública y público en general, con su frecuencia relativa.

Par Evodio Vázquez se observa que sólo la palabra Guerrero es común entre ambas poblaciones, es decir, no se observa que hablen del mismo discurso o tema. Por otro lado, se observa que hay varias palabras en común en lo expresado en ambos sentimientos, positivo y negativo.

Son pocas las palabras que tienen en común entre el Félix Salgado y el público en general y no se observa concordancia en el diálogo de ambas poblaciones. En el caso del sentimiento negativo la palabra *lopezobrador* que es común en ambos, tiene un peso relativo mayor en los usuarios, 0.075, que el de Félix Salgado, 0.044.

En el caso de Marcos Parra, en el sentimiento positivo no hay palabras en común, y para el sentimiento negativo tienen en común Taxco y presidente.

Con Mario Moreno no se observan palabras en común en el sentimiento positivo y, en el caso del sentimiento negativo solo tienen en común la palabra Guerrero.

Tabla 5

Comparativo de palabras más usadas entre figura pública y público en general para los sentimientos positivo y negativo

Positivo				Negativo			
Figura pública	Wh	Público en general	Wh	Figura pública	Wh	Público en general	Wh
Evodio Vázquez				Evodio Vázquez			
hoy	0.065	felixsalmac	0.134	hoy	0.053	felixsalmac	0.129
acapulco	0.050	salgado	0.055	duda	0.041	mario	0.065
civil	0.043	felix	0.051	prd	0.041	salgado	0.062
proteccion	0.040	mario	0.049	guerrero	0.036	felix	0.056
guerrero	0.035	guerrero	0.036	gracias	0.036	guerrero	0.031
Félix Salgado				Félix Salgado			
presidente	0.050	hoy	0.053	presidente	0.063	felixsalmac	0.111
mexico	0.047	duda	0.041	pueblo	0.049	lopezobrador	0.075
abrazos	0.044	prd	0.041	lopezobrador	0.044	salgado	0.057
amlo	0.038	guerrero	0.036	amlo	0.044	felix	0.048
lopezobrador	0.032	vamos	0.036	hermano	0.024	unvioladornosera	0.035
Marcos Parra				Marcos Parra			
gobernador	0.054	mario	0.120	taxco	0.073	grupo	0.050
guerrero	0.053	moreno	0.070	dia	0.047	parra	0.038
hoy	0.052	arcos	0.064	guerrero	0.036	taxco	0.038
hector	0.039	delgado	0.026	duda	0.031	presidente	0.038
astudillo	0.038	radio	0.026	presidente	0.031	municipal	0.038
Mario Moreno				Mario Moreno			
hoy	0.051	mario	0.117	hoy	0.046	mario	0.127
gobernador	0.048	arcos	0.066	gobernador	0.046	moreno	0.106
general	0.045	moreno	0.059	hector	0.043	fight	0.040
director	0.043	radio	0.031	guerrero	0.038	decision	0.031
guerrero	0.038	formula	0.031	gracias	0.036	guerrero	0.021

Nota: tabla de elaboración propia, datos extraídos de X (antes Twitter, 2021).

El anexo 2 contiene la nube de palabras de los sentimientos positivo y negativo de lo expresado por figuras públicas y por el público en general para cada uno de los aspirantes.

Correlación entre emociones y tipo de población, coeficientes de correlación Spearman

Una forma de medir una asociación entre variables es por medio de la correlación de **Spearman** (Siegel, 1970). Esta prueba es **no paramétrica**, y podemos aplicarla para nuestro caso, ya que las emociones que genera Syuzhet tienen un grado de intensidad, que va de 0 al 4, es una variable ordinal, y la variable **tipo de población**, también la podemos considerar como ordinal, donde 0 es público en general y 1 figura pública. La tabla 6 muestra un ejemplo de cómo son los datos para generar las correlaciones entre tipo de **población** y la emoción **positiva**.

Tabla 6

Ejemplo de valores para calcular la correlación de Spearman.

Aspirante	Sentimiento	Población	Positivo
Evodio V.	Positivo	0	4
Evodio V.	Positivo	0	2
Evodio V.	Positivo	1	4
Evodio V.	Positivo	0	1
Evodio V.	Positivo	1	2
Evodio V.	Positivo	0	3
Evodio V.	Positivo	0	3
Evodio V.	Positivo	1	4
Evodio V.	Positivo	0	2
Evodio V.	Positivo	0	2
Evodio V.	Positivo	0	1
...	...	...	...

Nota: tabla de elaboración propia, datos extraídos de X (antes Twitter, 2021).

La hipótesis nula establece que son independientes la emoción o sentimiento con respecto al tipo de población, público en general y figura pública. Si se rechaza la hipótesis nula (para $p\text{-valor} < 0.05$) implica que para esa emoción hay una correlación entre las dos variables: población y emoción.

La tabla 7 muestra que las emociones *positivo* y *confianza* son significativas y con coeficiente positivo. Al ser significativo el coeficiente de correlación implica que si pasamos de usuario (0) a aspirante (1), entonces hay un grado mayor en la emoción, en conclusión, implica discordancia entre ambos, Evodio V. y el público en general.

En la misma tabla se muestra que la emoción *negativo* y *miedo* son significativas, en el primer caso el coeficiente es negativo, es decir, hay una correlación negativa de Félix Salgado con respecto al público en general.

En el caso de Marco Parra no se observan coeficientes significativos, esto implica concordancia entre el aspirante y los usuarios.

En la tabla 13 se muestra que el sentimiento *negativo*, *confianza* y *disgusto* son significativos. Los valores del estadístico para los sentimientos negativo y disgusto son negativos implica una correlación inversa entre el aspirante y el público en general.

Tabla 7

Coeficientes de correlación de Spearman por sentimientos positivo y negativo.

Aspirante	Sentimiento	r	p-valor
Evodio V.	Confianza	0.101	0.003
Evodio V.	Disgusto	-0.048	0.149
Evodio V.	Miedo	-0.052	0.127
Evodio V.	Negativo	-0.047	0.073
Evodio V.	Positivo	0.065	0.010
Felix S.	Confianza	0.039	0.059
Felix S.	Disgusto	-0.001	0.952
Felix S.	Miedo	0.045	0.028
Felix S.	Negativo	-0.035	0.046
Felix S.	Positivo	0.002	0.923
Marcos P.	Confianza	0.099	0.290
Marcos P.	Negativo	0.084	0.542
Marcos P.	Positivo	0.057	0.477
Mario M.	Confianza	0.135	0.004
Mario M.	Disgusto	-0.216	0.008
Mario M.	Miedo	0.005	0.952
Mario M.	Negativo	-0.151	0.017
Mario M.	Positivo	-0.071	0.079

Nota: tabla de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Extracción de tópicos

En la tabla 8 se observan algunos ejemplos de la extracción de tópicos con algunos de sus tweets. El anexo 3 contiene una lista más completa de tópicos con sus respectivos tweets.

En general, no se observa concordancia entre los tópicos que expresa el aspirante con respecto a los tópicos mencionados por el público, por ejemplo, Evodio Vázquez habla sobre Félix Salgado y los usuarios sobre el tema del agua y del turismo en Acapulco. En el caso de tópicos de Félix Salgado, él habla de apoyo al presidente, y los usuarios sobre la acusación de agresión sexual que recae sobre el señor Félix.

Lo usuarios de Twitter expresan comentarios sobre el mismo Marcos parra, mientras él, habla de temas como una sesión del congreso. En el caso de Mario Moreno, los usuarios expresan temas como el diálogo y respeto por la opinión mayoritaria, y Mario Moreno habla de la infraestructura en diferentes comunidades.

Tabla 8

Extracción de tópicos de los tweets de figuras públicas y usuarios

Tópicos y tweets de usuarios en relación a Evodio Vázquez	
Tema 1	Críticas a Félix Salgado Macedonio, candidato a gobernador de Guerrero, por presuntos abusos sexuales y vínculos con el crimen organizado.
Tweet 1	@EstefaniaVeloz Mas respeto para el candidato violador, te puedes quedar sin chambaaa o en una de esas te busca, ya ves que
Tweet 2	Un grupo de mujeres vestidas de negro se manifestaron este jueves frente a la sede nacional del @PartidoMorenaMx, contra la designación
Tópicos y tweets expresados por Evodio Vázquez	
Tema 1	Agua y turismo en Acapulco, México.
Tweet 1	El agua es la fuerza motriz de toda la naturaleza, los invita a cuidar nuestro líquido vital y a no desperdiciarlo. ¡Cuidémosla!
Tweet 2	Acapulco es la ciudad que más reditúa económicamente al estado. El turismo es la actividad principal, dejando más de la mitad de...
Tópicos y tweets de usuarios en relación a Félix Salgado	
Tema 1	Todos los tweets tratan sobre la misma persona, Felix Salmac, y la necesidad de que se haga justicia con respecto a las acusaciones de agresión sexual.
Tweet 1	Dije que FelixS, @MelecioEric @ChumelTorres No es ninguna maroma. Ni apoyo a Morena, ni a López ni a Félix Salgado. Simplemente así es la ley y sea
Tweet 2	FelixS, @AristeguiOnline sin pruebas ataca a @FelixSalMac deberia usted demandar a la señora señor senador, por que esta usando redes sociales...
Tópicos y tweets expresados por Félix Salgado	
Tema 1	Apoyo al Presidente López Obrador y sus políticas.
Tweet 1	Fox dijo: "Le vamos de la madre a la 4T". "Les vamos a dar en la madre". Vocabulario bien aplaudido por las feministas del P.
Tweet 2	Sigo insistiendo: no habrá Cuarta Transformación, sí AMLO no destituye a los Ministros viejos y mañosos que encabeza Eduardo Medina
Tópicos y tweets de usuarios en relación a Marcos Parra	
Tema 1	Tweets relacionados con Marcos Parra y su trabajo.
Tweet 1	@bigail_Parra La verdad un horror de la portera se comió el gol.
Tweet 2	Lamento mucho esta pérdida irreparable, amigo @marcoseparra
Tópicos y tweets expresados por Marcos parra	
Tema 1	Todos los tweets están relacionadas con una sesión del Congreso.
Tweet 1	El Consejo sustituye a la Comisión de Derechos Humanos, ya que esta muy a menudo fue criticada por miembros que ni siquiera ...
Tweet 2	En la misma sesión se eligió por unanimidad (41 votos de 41 diputados) a Ramón Navarrete Magdaleno como nuevo... http://t.co/xpJJE
Tópicos y tweets de usuarios en relación a Mario Moreno	
Tema 1	Los tweets expresan apoyo a Mario Moreno y Roberto Santos y abogan por el diálogo y el respeto por la opinión mayoritaria.
Tweet 1	El discurso de Moreno. "El diálogo es la solución". El diálogo jamás puede ser la solución, aquí se impone la voluntad de la mayoría
Tweet 2	ZONA CERO Mario Moreno, construye una nueva historia para Guerrero
Tópicos y tweets expresados por Mario Moreno	
Tema 1	Los tweets mencionan proyectos de infraestructura en diferentes comunidades
Tweet 1	Calle Principal de la colonia Bella Vista, la cual sirve de acceso para colonias como la PRD, y Amelitos, entre... http://t.co/05g
Tweet 2	Recorrido por el interior de la Comisaría Municipal de El Ocotito que acabamos de inaugurar, y que incluye una... http://t.co/qXQN

Nota: tabla de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Resultados con Sentiment-analysis-spanish (SAS)

Sentiment-analysis-spanish genera solo un valor para cada tweet, donde valores cercanos a 1 son tweets positivos y valores cercanos a 0 negativos. En la tabla 9 se observa que los más positivos o, dicho de otra forma, los menos negativos son Mario Moreno y Marco Parra, y los más negativos son Félix Salgado, seguido de Evodio.

De acuerdo con la tabla 9 hay concordancia entre lo que publica Evodio de lo que expresa el público en general, en este caso ambos tienen un valor de 0.26. Los tweets publicados por Félix Salgado son ligeramente más positivos que lo expresado por el público en general, 0.236 contra 0.284.

De acuerdo a la tabla 9 hay poca concordancia entre los publicados por Marcos Parra y el público en general, este último es más negativo. Los tweets de Mario Moreno son ligeramente más negativos que los del público en general. Podríamos decir que sí hay concordancia entre ambas figuras públicas.

Tabla 9

Promedio de sentimiento por aspirante y tipo de población

Aspirante	Público en general	Figura Pública	Total general
Evodio Velázquez	0.268	0.261	0.268
Félix Salgado	0.236	0.284	0.236
Marcos Parra	0.274	0.364	0.357
Mario Moreno	0.361	0.382	0.373
Total general	0.249	0.354	0.258

Nota: tabla de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Conclusiones y recomendaciones

Conclusiones y recomendaciones

En general, es difícil encontrar librerías o aplicaciones para la obtención de emociones y sentimientos **de textos en español**. Una opción es traducir los textos al inglés para después usar librerías como NLTK, TextBlob o Vader. Pero la traducción en inglés en forma masiva tiene un costo, y en este caso se tienen más de 50 mil tweets.

El conjunto de datos con el que se trabajó este proyecto se extrajo en febrero del 2021, actualmente, la extracción de datos de esta plataforma de Twitter está más restringida, tanto por el número de tweets que se pueden bajar por día como por la cantidad de características.

Un propósito de este estudio es la obtención de emociones y sentimientos de tweets publicados en español, y se generaron por medio de dos aplicaciones, Syuzhet y Sentiment-Analysis-Spanish, en relación a dos poblaciones: figuras públicas y público en general. Adicionalmente, interesa hacer un análisis **descriptivo** de la concordancia entre el discurso de los aspirantes y los usuarios de Twitter.

Las figuras 10 y 11, porcentaje de emociones/sentimientos expresados en los tweets por aspirante, muestran una clara diferencia entre las emociones y sentimientos entre cada uno de los aspirantes. Por ejemplo, los tweets relacionados con Mario Moreno y Marco Parra, en general, son más positivos, en comparación con los tweets de Evodio Vázquez y Félix Salgado. Estas dos figuras muestran los tweets de los aspirantes y de los usuarios en conjunto.

Al comparar la concordancia de las emociones y sentimientos por aspirante en relación con lo expresado por el público en general (PG) se observa que:

Evodio Vázquez expresa más confianza que el PG, y éste expresa enfado y miedo como las 2 emociones más recurrentes, por lo que hay una discordancia en las emociones, figuras 12 y 13. Lo mismo se observa en la figura 14 y 15, donde el Sr. Vázquez es más positivo, 75%, contra un 50% de los usuarios. Analizando la tabla 5 de palabras, se observa una discordancia en las palabras. Para Evodio Vázquez las palabras positivas son hoy, Acapulco, civil, y para el PG son felixsalmac, Salgado, Félix. En las palabras negativas el Sr. Vázquez hace referencia a hoy y duda, y el PG menciona a felixsalmac, Mario y Salgado. Estas palabras negativas hacen referencia a que en ese momento al Sr. Salgado se le acusaba de un delito.

En el caso del Sr. **Félix Salgado** se observa algo parecido al aspirante anterior, hay una discordancia entre él y el PG en las emociones *confianza*, *miedo* y *Enfado*, figuras 12 y 13. En el discurso positivo el Sr. Salgado menciona a presidente, México y abrazos, y en positivo, el PG menciona a hoy, más, duda. En el sentimiento negativo Félix Salgado expresa con más frecuencia presidente y pueblo, y los usuarios felixsalmac, lopezobrador y salgado.

Marcos Parra parece tener más concordancia con el público en general, por ejemplo, confianza es la emoción más presente en ambos casos, aspirante y PG, alrededor de un 30%, figuras 12 y 13. Sin embargo, en los sentimientos, positivo y negativo, no se observa mucha coincidencia, 76% contra 50%, figuras 14 y 15. En el discurso, en el sentimiento positivo Marcos Parra menciona con más frecuencia a gobernador, guerrero, hoy, y los usuarios menciona a Mario Moreno y arcos, tabla 5.

Por último, el Sr. **Mario Moreno** parece tener concordancia en la emoción de *confianza*, en ambos casos (aspirante y usuarios) es la emoción más expresada, figuras 12 y 13. Mario Moreno menciona como positivo hoy, gobernado y general y los usuarios Mario, arcos y Moreno.

Las nubes de palabras (Anexo 2) asociadas a emociones positivas pueden compartir palabras con emociones negativas. Puede haber algunos casos en que un solo tweet expresa ambos sentimientos, si este es el caso, Syuzhet calificará al tweet como positivo y como negativo, es el caso de este tweet:

“Todos hemos padecido la pérdida de un ser querido sin duda las personas se van, pero su recuerdo nos permite que sigan viviendo en nuestra mente corazón que está festividad de día de muertos sea para honrarlos recordarlos en familia”

La parte negativa es la pérdida de un ser querido, y la positiva, el recuerdo que permanece en nuestro corazón.

Hay palabras que se pueden usar en ambos sentimientos, positivos y negativos, porque tienen un contexto general, como felixsalmac, Guerrero, morena, etc. La nube de palabras, como análisis exploratorio, tiene el inconveniente de que la clasificación del sentimiento del tweet se realiza a nivel de unigramas (tokens), y no implica que todas las palabras de ese tweet son positivas o negativas.

Después de analizar las emociones y sentimientos de los tweets publicados, por medio de gráficas, nubes de palabras, correlaciones entre sentimientos y tipo de población y extracción de tópicos, se llega a la conclusión de que **hay muy poca concordancia entre lo que expresa una figura pública y lo que transmiten los usuarios**. Se observa que las figuras públicas expresen muy pocas emociones y sentimientos relacionados con lo que expresa el público en general y viceversa. Creo que hay una ventana de oportunidad para que las figuras públicas empaticen y conecten con los usuarios.

Para estudios futuros se podría desarrollar el perfilamiento de los usuarios de Twitter en relación a los aspirantes y hacer el análisis respectivo, por ejemplo, identificar a los usuarios como seguidores, detractores o neutros.

Fuentes de consulta

- Alifio Dieksona, Z., Bagas Prakosoa, M.R., Qalby Putraa, M.S., Fadel Syaputraa, M.S., Achmada, S., Sutoyo, R. (2022). 7th International Conference on Computer Science and Computational Intelligence. Procedia Computer Science.
- Ankita, Saleena, N. (2018). An Ensemble Classification System for Twitter Sentiment Analysis. International Conference on Computational Intelligence and Data Science (ICCIDS 2018). Procedia Computer Science.
- Arimetrics (2024). <https://www.arimetrics.com/glosario-digital/hashtag>.
- Baena, G. (2017). Metodología de la investigación. Grupo Editorial Patria.
- Bird S. Klein E. y Loper E. (2009), *Natural Language Processing with Python*, O'REILLY.
- Cortez Vázquez A., Vega Huerta, H., Pariona Quispe, J. (2009). Procesamiento de lenguaje natural. *Revista de Ingeniería de Sistemas e Informática*, 6(2), 48.
- Developer Platform, (2022). Twitter. <https://developer.twitter.com/en/docs>
- Dini, L. y Bittar, A., (2016). *Emotion Analysis on Twitter: The Hidden Challenge*. Holmes Semantic Solutions.
- Garcia, K. y Berton, L. (2010). Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied Soft Computing Journal*. ELSEVIER.
- Gestion.Org (2019). Personajes públicos y redes sociales. <https://www.gestion.org/personajes-publicos-y-redes-sociales/>
- Gómez Luz María, García Torres Cecilia, (2011), Twitter, *Comité de Educación Departamento de Educación, SCARE*, 38(4), 539-540.
- Gonzalo, S. (2020). Introducción a las técnicas de análisis de sentimientos. Universidad de Granada.

- Guardiola, E. (2016). MARKETING EMOCIONAL. El poder de las emociones. Obtenido de <https://www.inboundcycle.com/blog-de-inbound-marketing/marketingemocional-estrategia-marca>.
- Grupo Bit, ¿Cuántos datos se producen en un minuto?, <https://business-intelligence.grupobit.net/blog/cuantos-datos-se-producen-en-un-minuto>
- How to analyze the sentiment of your own Tweets, (2023). Twitter. <https://developer.twitter.com/en/docs/tutorials/how-to-analyze-the-sentiment-of-your-own-tweets>
- Isasi, J. (2021). Análisis de sentimientos en R con 'syuzhet'. Programming Historian. <https://programminghistorian.org/es/lecciones/analisis-de-sentimientos-r>
- Lerma, H. (2009). Metodología de la investigación: propuesta, anteproyecto y Proyecto. Ecoe Ediciones.
- Liu, Alex, (2015), Data Science and Data Scientist. IBM Analytics.
- Manning, C. D., y Schütze, H., (1999), Foundations of Statistical Natural Language Processing, The MIT Press.
- Marketing4ecommerce, (2022). <https://marketing4ecommerce.net/historia-de-las-redes-sociales-evolucion/#Los%20inicios>
- Merriam-Webster, Dictionary, (2023), <https://www.merriam-webster.com/>
- Nogueira dos Santos, C., y Gatti M., (2014), *Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts*. Brazilian Research Lab. IBM Research.
- NRC Word-Emotion Association Lexicon. National Research Council Canada. <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>
- ONTSI, Las Redes Sociales en Internet, Observatorio nacional de telecomunicaciones y de la SI.
- OpenAI, Embeddings (2023). <https://platform.openai.com/docs/guides/embeddings/what-are-embeddings>

OpenAI, Create Chat Completion, (2023). <https://platform.openai.com/docs/api-reference/chat>

Diez, D.M., Barr C.D. y Cetinkaya, M. R. (2016). *OpenIntro Statistics*. OpenIntro.

Package 'rtweet', (2023).

pypi.org, (2021). sentiment-analysis-spanish 0.0.25
<https://pypi.org/project/sentiment-analysis-spanish/>

QuestionPro, (2023). Análisis de sentimientos en Twitter: Qué es y cómo hacerlo.
<https://www.questionpro.com/blog/es/analisis-de-sentimientos-en-twitter/>

REAL ACADEMIA ESPAÑOLA, (2022), Diccionario de la lengua española, 23.^a ed., <https://dle.rae.es>

Russell, S. y Norvig, P., (2009). *Inteligencia Artificial: Un Enfoque Moderno* (3 ed). p. 229.

Sahayak, V., Shete, V., Pathan, A., (2015). Sentiment Analysis on Twitter Data. *International Journal of Innovative Research in Advanced Engineer.*

Sampieri H. R., Fernández C. C. y Baptista L. P., (2014), *Metodología de la Investigación*, Mc Graw Hill, (6ta ed.)

Sarduy, L., Milián, V. y Batista, R. (2020). Método para el análisis de sentimientos en la red de mensajería instantánea de la Universidad de las Ciencias Informáticas. Universidad de las Ciencias Informáticas.

Sentimiento. (2022, 7 de septiembre). Wikipedia, La enciclopedia libre. Fecha de consulta: 20:25, octubre 29, 2022 desde https://es.wikipedia.org/wiki/Sentimiento#cite_note-3

Siegel, S. (1970). *Estadística no paramétrica, aplicada a las ciencias de la conducta*. Trillas.

Statista, Las Redes Sociales en México, (2022), https://es.statista.com/temas/7392/las-redes-sociales-en-mexico/#topicHeader_wrapper

- Tellez, E. S., Miranda J. S., Graff M., Moctezuma D., Siordia O. S. y Villaseñor E. A., (2017). *A case study of Spanish text transformations for twitter sentiment analysis*. Expert Systems with Applications.
- Uthirapathy, S. E., Sandaman, D. (2023). Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model. *International Conference on Machine Learning and Data Engineering. Procedia Computer Science*
- Valderrama, B., Emociones: una taxonomía para el Desarrollo Emocional, *ARANDU-UTIC – Revista Científica Internacional*, 2(1) 19.
- Webempresa, (2018). https://www.webempresa.com/blog/que-es-twitter-como-funciona-2.html#Que_es_twitter.
- Wikipedia, (2024). La enciclopedia libre.
https://es.wikipedia.org/wiki/TensorFlow#cite_note-YoutubeClip-1.
<https://es.wikipedia.org/wiki/OpenAI>.
- Websiterating, (2022), 50 + ESTADÍSTICAS Y HECHOS DE TWITTER PARA 2022, <https://www.websiterating.com/es/research/twitter-statistics/>
- Zeeshan Ansaria, M., Aziza, M. B., Siddiquib, M. O., Mehraa, H., Singha, K. P. (2019). Analysis of Political Sentiment Orientations on Twitter. *Procedia Computer Science. International Conference on Computational Intelligence and Data Science*.

ANEXOS

ANEXO 1

Índice de términos

“A”

Algoritmo.....18

“B”

Big Data16

Blog17, 39

Bots.....40

“E”

Endpoint28

“F”

Followers o seguidores.....17,
45

“H”

Hashtags.....27,29,39, 40, 43

“K”

Keras.....46, 50

“O”

OpenAI.....45, 46

“P”

Público en general.....11

“S”

Stopwords.....31

“T”

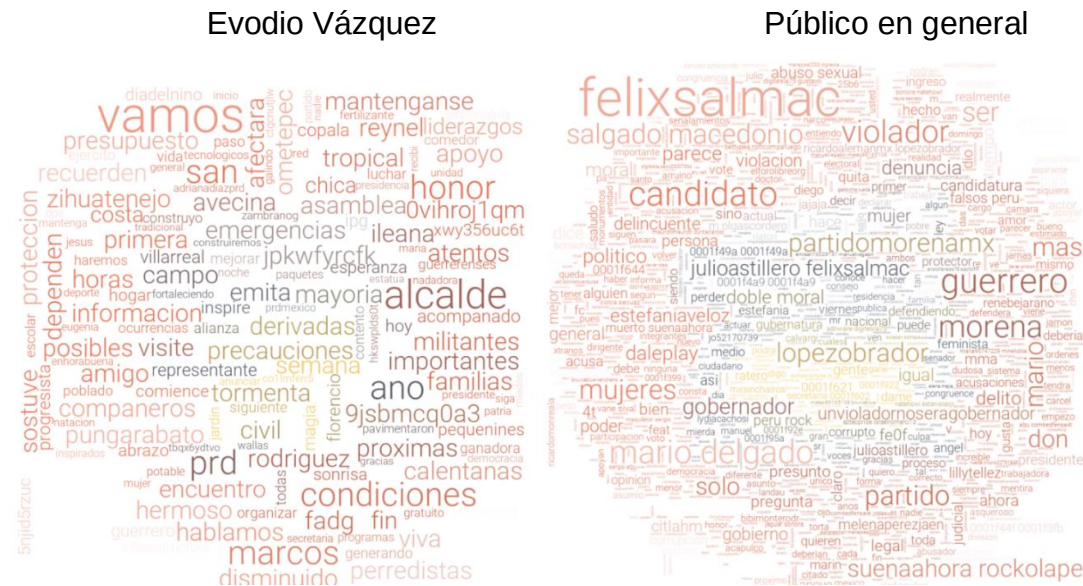
TensorFlow.....50

ANEXO 2

Nube de palabras

Figura 16

Palabras asociadas al **sentimiento positivo** de Evodio Vázquez y de los tweets del público en general.



Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Figura 17

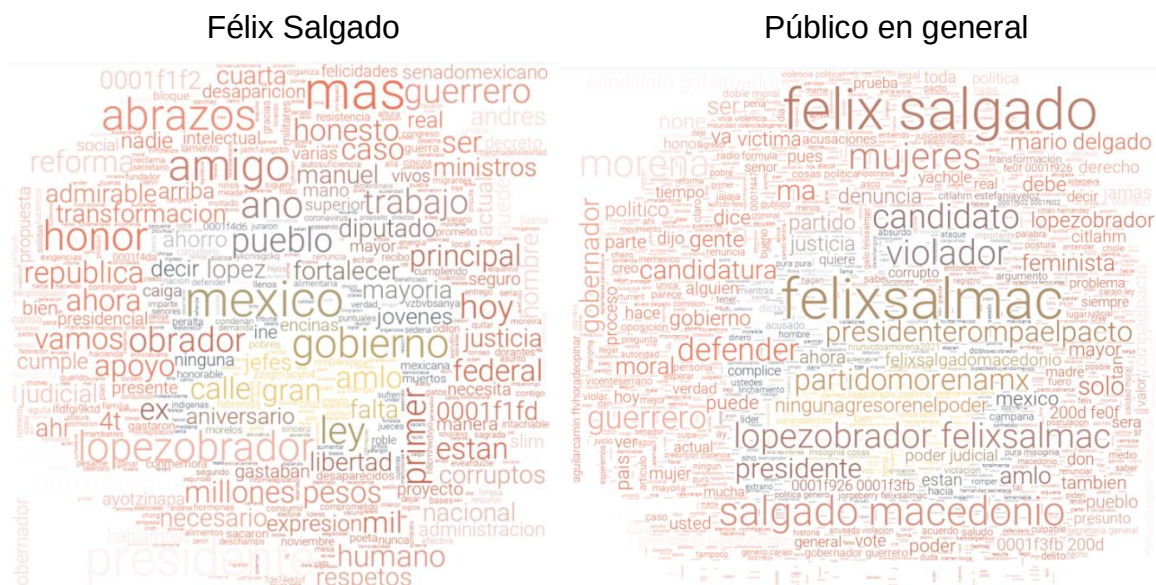
Palabras asociadas al **sentimiento negativo** de Evodio Vázquez y de los tweets del público en general.



Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Figura 18

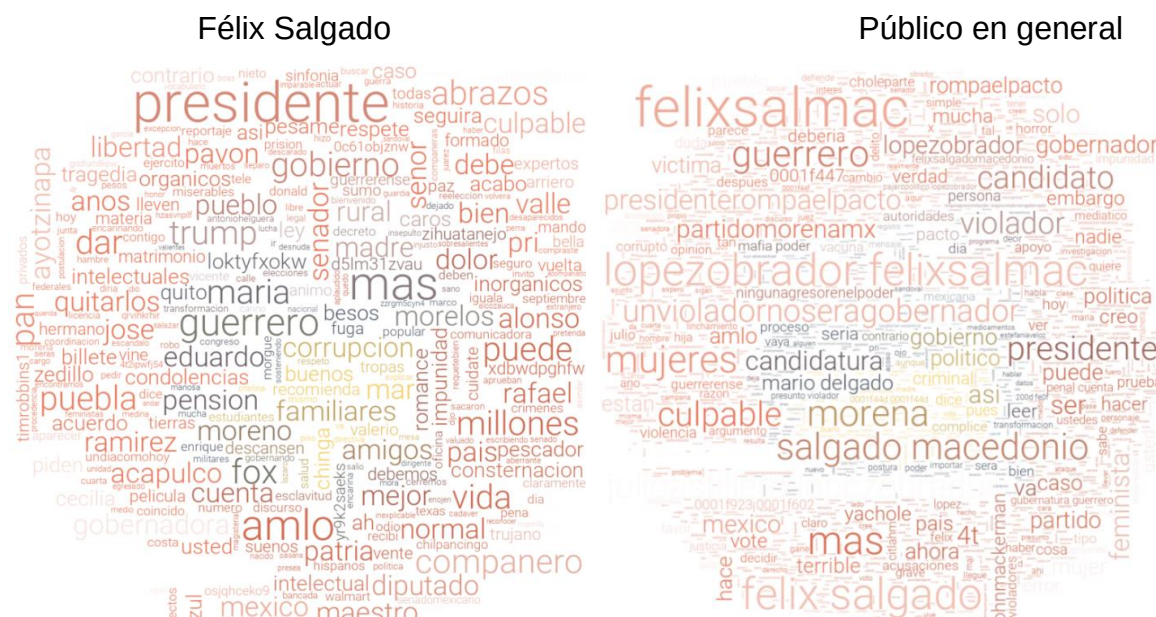
Palabras asociadas al **sentimiento positivo** de Félix Salgado y de los tweets del público en general.



Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Figura 19

Palabras asociadas al **sentimiento negativo** de Félix Salgado y de los tweets del público en general.

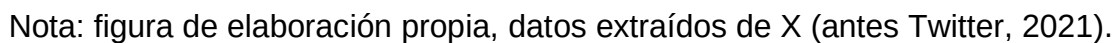


Nota: figura de elaboración propia, datos extraídos de X (antes Twitter, 2021).

Palabras asociadas al **sentimiento positivo** de Marcos Parra y de los tweets del público en general.



Palabras asociadas al **sentimiento negativo** de Marcos Parra y de los tweets del público en general.



Palabras asociadas al **sentimiento positivo** de Mario Moreno y de los tweets del público en general.



Palabras asociadas al **sentimiento negativo** de Mario Moreno y de los tweets del público en general.



ANEXO 3

Extracción de tópicos

Tabla 10

Tópicos y tweets de usuarios en relación a Evodio Vázquez	
Tema 1	Críticas a Félix Salgado Macedonio, candidato a gobernador de Guerrero, por presuntos abusos sexuales y vínculos con el crimen organizado.
Tweet 1	@EstefaniaVeloz Mas respeto para el candidato violador, te puedes quedar sin chambaaa o en una de esas te busca, ya ves que
Tweet 2	Un grupo de mujeres vestidas de negro se manifestaron este jueves frente a la sede nacional del @PartidoMorenaMx, contra la designación
Tweet 3	@lozaime @FelixSalMac No, soy pro AMLO 100% y a mi en lo personal no me gusta que este señor después de lo que es acusado vaya a...
Tema 2	Críticas al gobierno actual y sus políticas.
Tweet 1	@PikoSarinana @julioastillero @FelixSalMac Ahora, sugiero, vengán a hacer un sondeo a GUERRERO, pero en campo, no desde la comodidad
Tweet 2	@pozole_con_col @julioastillero @PartidoMorenaMx @FelixSalMac A eso me refiero con la "izquierda" mojigata. El mismo discurso y ...
Tweet 3	@PartidoMorenaMx : Aprobamos a @FelixSalMac como candidato, pero ¿cómo minorizamos el escándalo?
Tópicos y tweets expresados por Evodio Vázquez	
Tema 1	Agua y turismo en Acapulco, México.
Tweet 1	El agua es la fuerza motriz de toda la naturaleza, los invita a cuidar nuestro líquido vital y a no desperdiciarlo. ¡Cuidémosla!
Tweet 2	Acapulco es la ciudad que más reedita económicamente al estado. El turismo es la actividad principal, dejando más de la mitad de...
Tweet 3	Inauguré la tercera etapa del sistema de agua potable del proyecto El Mirador en la colonia Alta Cuauhtémoc, beneficiando a 11 comunidades.
Tema 2	Todas las revisiones analizan experiencias positivas con iniciativas de gobiernos locales.
Tweet 1	El siguiente paso que daré, en breve, será un informe comparativo del primer año de gobierno de la alcaldesa Adela Román y del pri.
Tweet 2	Me siento muy contento porque en este poblado, en infraestructura, se construyó un comedor escolar en el jardín de niños y se pavi.
Tópicos y tweets de usuarios en relación a Félix Salgado	
Tema 1	Todos los tweets tratan sobre la misma persona, Felix Salmac, y la necesidad de que se haga justicia con respecto a las acusaciones de agresión sexual.
Tweet 1	Dije que FelixS, @MelecioEric @ChumelTorres No es ninguna maroma. Ni apoyo a Morena, ni a López ni a Félix Salgado. Simplemente así es la ley y sea
Tweet 2	FelixS, @AristeguiOnline sin pruebas ataca a @FelixSalMac debería usted demandar a la señora señor senador, por que esta usando redes sociales...
Tema 2	Crítica a Felix Salmac.

Tweet 1 @XimenaVald3 @FelixSalMac @claraluzflores Basura de persona ese tipejo, y @claraluzflores debería renunciar a MORENA ya que alberg

Tópicos y tweets expresados por Félix Salgado

Tema 1 Apoyo al Presidente López Obrador y sus políticas.

Tweet 1 Fox dijo: "Le vamos de la madre a la 4T". "Les vamos a dar en la madre". Vocabulario bien aplaudido por las feministas del P.

Tweet 2 Sigo insistiendo: no habrá Cuarta Transformación, sí AMLO no destituye a los Ministros viejos y mañosos que encabeza Eduardo Medina

Tema 2 Los tweets reflejan la necesidad de responsabilidad financiera y reforma en México, particularmente en lo que respecta a las pensiones de los funcionarios gubernamentales.

Tweet 1 Eso sin tomar en cuenta el dinero que gana por impartir conferencias. <https://t.co/yf4TN6vBAj>

Tweet 2 ¿Qué pasaría si liquidamos a los Ministros que no se quieren ajustar a la nueva ley de remuneraciones? Propongo retirarles la mil...

Tópicos y tweets de usuarios en relación a Marcos Parra

Tema 1 Pérdida y Urgencia.

Tweet 1 Respetuosamente nos unimos a la pena que embarcar al M.A. @marcoseparra, Presidente Municipal de Taxco de Alarcón, ante la irrepara

Tweet 2 #ServicioSocial Se requiere con urgencia donantes de sangre O+ para Marcos Andrés Ducon Parra, quien se encuentra en estado crítico

Tema 2 Tweets relacionados con Marcos Parra y su trabajo.

Tweet 1 @bigail_Parra La verdad un horror de la portera se comió el gol.

Tweet 2 Lamento mucho esta pérdida irreparable, amigo @marcoseparra

Tweet 3 Fotografía a principios de la década de 1950, bloques del Silencio..a la derecha de la gráfica la esquina de Aserradero hacia Marc

Tópicos y tweets expresados por Marcos parra

Tema 1 Todos los tweets están relacionadas con una sesión del Congreso.

Tweet 1 El Consejo sustituye a la Comisión de Derechos Humanos, ya que esta muy a menudo fue criticada por miembros que ni siquiera ...

Tweet 2 En la misma sesión se eligió por unanimidad (41 votos de 41 diputados) a Ramón Navarrete Magdaleno como nuevo... <http://t.co/xpJjE>

Tema 2 Todos los tweets mencionan políticas o acciones gubernamentales.

Tweet 1 @JoseCardenas1 ¿Qué nos extraña? ¿Recuerdan el caso de la niña Paulete? Otro ridículo.

Tweet 2 Descanse en paz Don Humberto, un hombre íntegro. <https://t.co/QKlpzcYh8q>

Tweet 3 Con FCH el precio internacional del petróleo se incrementó y el gobierno absorbió el costo para evitar mayor impacto #QueBajelaGas

Tópicos y tweets de usuarios en relación a Mario Moreno

Tema 1 Los tweets expresan apoyo a Mario Moreno y Roberto Santos y abogan por el diálogo y el respeto por la opinión mayoritaria.

Tweet 1	El discurso de Moreno. "El diálogo es la solución". El diálogo jamás puede ser la solución, aquí se impone la voluntad de la mayoría
Tweet 2	ZONA CERO Mario Moreno, construye una nueva historia para Guerrero
Tema 2	Mixed Martial Arts (MMA)
Tweet 1	Dana White confirma los planes para Title Eliminator entre Colby Covington y Leon Edwards https://t.co/HhkLIROnVg #UFC258 #MMA #Da
Tweet 2	Dana White defiende postura de Charles Oliveira por rechazar enfrentar a Michael Chandler a solo días de UFC 258 https://t.co/qBVJ
Tweet 3	Otra vez se cancela Chimaev vs Edwards.
Tópicos y tweets expresados por Mario Moreno	
Tema 1	Los tweets mencionan proyectos de infraestructura en diferentes comunidades
Tweet 1	Calle Principal de la colonia Bella Vista, la cual sirve de acceso para colonias como la PRD, y Amelitos, entre... http://t.co/05g
Tweet 2	Recorrido por el interior de la Comisaría Municipal de El Ocotito que acabamos de inaugurar, y que incluye una... http://t.co/qXQN
Tweet 3	El Gobierno de Héctor Astudillo Flores, realizó una importante obra en la comunidad de Mazatlán: la pavimentación de la calle de a...
Tema 2	El tema común de estos tweets es el trabajo del presidente Andrés Manuel López Obrador y sus esfuerzos para brindar asistencia directa a los grupos vulnerables.
Tweet 1	La coordinación, los consensos y la unidad, son indispensables para lograr lo que hoy también es el objetivo principal del Presidente.
Tweet 2	Les comparto el enlace de la entrevista de esta mañana con @janupi a través de @Radio_Formula hablamos del proceso político que esta
Tweet 3	Nuestro reconocimiento más amplio a los integrantes de Protección Civil.
Nota: tabla de elaboración propia, datos extraídos de X (antes Twitter, 2021).	