



**Universidad de Buenos Aires**

Facultad de Ciencias Exactas y Naturales  
Departamento de Ciencias de la Computación

**Aplicación de Modelos de Lenguaje para la  
Identificación de Emociones en Twitter durante  
las Elecciones Presidenciales de 2022 en Colombia**

Tesis presentada para optar por el título de

**Magister en Explotación de Datos y  
Descubrimiento del Conocimiento**

*por*

**Juan Jose Iguaran Fernandez**

Director: Dr. Juan Manuel Pérez

Codirector: Dr. Germán Rosati

Buenos Aires, Septiembre del 2023



## AGRADECIMIENTOS

Quiero agradecer a la Maestría por todo el conocimiento y la formación brindada sin los cuales cual este trabajo no habría sido posible. Agradezco también a mis directores por todo su apoyo, paciencia y mentoría a lo largo de este proceso.

A Luis Felipe Nuñez, cuyas valiosa palabras "Tomate en serio" forjaron mi motivación. A Jaime Diaz, gracias a quien pude iniciar este viaje y estuvo presente de muchas formas. A Carolina Franco cuyo afecto, apoyo y acompañamiento fueron claves durante este proceso, y a Judith Fernandez, la autora de mi vida.

Todos los estados encuentran su origen en la mente. La mente es su fundamento y son creaciones de la mente. Si uno habla o actúa con un pensamiento impuro, entonces el sufrimiento le sigue de la misma manera que la rueda sigue la pezuña del buey...

Todos los estados encuentran su origen en la mente. La mente es su fundamento y son creaciones de la mente. Si uno habla o actúa con un pensamiento puro, entonces la felicidad le sigue como una sombra que jamás le abandona.

—*Dhammapada*

---

# Resumen

El estudio de Twitter como medio para el análisis de fenómenos sociales mediante técnicas de procesamiento del lenguaje natural (NLP), ha generado gran interés en los últimos años debido a la disponibilidad de grandes cantidades de datos en un ambiente relativamente espontáneo. Dentro de estas técnicas, la detección de emociones en el texto es especialmente relevante, ya que permite identificar la respuesta subjetiva de las personas ante los distintos eventos sociales presentados. Los modelos de lenguaje basados en redes neuronales, como los Transformers y específicamente BERT, han reemplazado las técnicas tradicionales de NLP debido a su capacidad para capturar el sentido y las relaciones entre las palabras en el texto. Sin embargo, en español hay pocos estudios de detección de emociones en redes sociales y ninguno que utilice modelos de lenguaje basados en Transformers en un contexto político.

Este trabajo tuvo como objetivo emplear modelos basados en BERT para detectar emociones en Twitter durante las elecciones presidenciales de Colombia en 2022. Para ello, se etiquetó manualmente un conjunto de datos constituido por tweets realizados durante dicho período y se realizaron experimentos de clasificación utilizando varios modelos de lenguaje pre-entrenados en español. A su vez, se utilizaron dichos modelos para analizar las respuestas emocionales de los usuarios, asociando los tweets a sectores políticos, así como su variación temporal.

Tanto para el proceso de etiquetado como para la predicción de los modelos, se observó en los tweets una presencia considerablemente mayor de las emocionesasco y alegría con respecto a tristeza y miedo. Esto a su vez provocó un desempeño mucho mejor de los modelo para los tweets etiquetados con estas primeras dos emociones. Así mismo, si bien el modelo seleccionado para la predicción, RoBERTuito, tuvo

un mejor desempeño respecto a los otros modelos estudiados en todas las etiquetas, en aquellas con poca presencia esta ventaja fue mas evidente. Esto se explica por que dicho modelo fue específicamente entrenado para la tarea de analizar tweets en español.

Por otro lado, a partir de los resultados de clasificación del modelo sobre la base de datos, se observó que la emoción de mayor presencia en los tweets asociados con la derecha fue el asco y para los tweets asociados con la izquierda fue alegría. Se observó también que los días en donde se presentó un numero relevante de etiquetas emocionales, fueron aquellos en donde hubo algún evento de relevancia política, como el día de las elecciones, o algún fenómeno mediático fue altamente discutido.

**Palabras Clave:** [BERT, Colombia, Elecciones, Detección de emociones]

# Índice general

|                                                                       |          |
|-----------------------------------------------------------------------|----------|
| Índice de figuras                                                     | 7        |
| Índice de cuadros                                                     | 9        |
| <b>1. Introduccion</b>                                                | <b>1</b> |
| 1.1. Extracción de opiniones en el texto . . . . .                    | 1        |
| 1.2. Avances técnicos recientes en análisis de sentimiento . . . . .  | 3        |
| 1.3. Análisis de sentimiento en Twitter en torno a política . . . . . | 4        |
| 1.4. Aporte de este trabajo . . . . .                                 | 5        |
| <b>2. Trabajo Previo</b>                                              | <b>7</b> |
| 2.1. Aprendizaje supervisado en la detección de emociones . . . . .   | 8        |
| 2.2. Redes neuronales para el análisis de texto . . . . .             | 9        |
| 2.3. Sentimientos y emociones en redes sociales . . . . .             | 14       |

|                                                                   |           |
|-------------------------------------------------------------------|-----------|
| <b>3. Metodología</b>                                             | <b>18</b> |
| 3.1. Datos . . . . .                                              | 18        |
| 3.1.1. Recolección y análisis exploratorio de datos . . . . .     | 19        |
| 3.1.2. Etiquetado . . . . .                                       | 23        |
| 3.2. Experimentos de clasificación . . . . .                      | 34        |
| 3.2.1. Modelos preentrenados . . . . .                            | 34        |
| 3.2.2. Entrenamiento de los modelos . . . . .                     | 35        |
| 3.2.3. Evaluación de los modelos . . . . .                        | 36        |
| <b>4. Resultados</b>                                              | <b>37</b> |
| 4.1. Desempeño de los modelos entrenados . . . . .                | 38        |
| 4.2. Distribución de emociones por orientación política . . . . . | 41        |
| 4.3. Emociones a lo largo del tiempo . . . . .                    | 43        |
| <b>5. Conclusiones</b>                                            | <b>49</b> |
| <b>Referencias</b>                                                | <b>52</b> |
| <b>A. Anexos</b>                                                  | <b>59</b> |

# Índice de figuras

|                                                                                                                 |    |
|-----------------------------------------------------------------------------------------------------------------|----|
| 2.1. Arquitectura de los Transformers . . . . .                                                                 | 12 |
| 2.2. Comparación entre mecanismos de redes. Fuentes Jurafsky (2000),<br>Wikipedia contributors (2022) . . . . . | 12 |
| 2.3. Arquitectura de BERT. Fuente Devlin et al. (2018) . . . . .                                                | 13 |
| 3.1. Porcentaje de Hashtags según la orientación política asignada . . . . .                                    | 21 |
| 3.2. Porcentaje de Tweets según la orientación política asignada . . . . .                                      | 22 |
| 3.3. Porcentaje de tweets según orientación política a lo largo del tiempo . . . . .                            | 23 |
| 3.4. Interfaz de etiquetado . . . . .                                                                           | 27 |
| 3.5. Porcentaje de correlación entre etiquetadores de las emociones asignadas a los tweets' . . . . .           | 29 |
| 3.6. Matriz de confusión entre etiquetas asignadas . . . . .                                                    | 32 |

|                                                                                                                                |    |
|--------------------------------------------------------------------------------------------------------------------------------|----|
| 4.1. Matriz de confusión entre etiquetas asignadas y predicciones del algoritmo para el dataset etiquetado . . . . .           | 40 |
| 4.2. Porcentaje de tweets clasificado según emociones por el algoritmo . .                                                     | 42 |
| 4.3. Porcentaje de tweets clasificados según emoción por el algoritmo, separado por orientación política del hashtag . . . . . | 43 |
| 4.4. Porcentaje diario del total de tweets clasificados en cada emoción a lo largo del tiempo . . . . .                        | 44 |
| 4.5. Porcentaje diario del total de tweets clasificados como Alegría, para cada orientación política . . . . .                 | 45 |
| 4.6. Porcentaje diario del total de tweets clasificados como Miedo, para cada orientación política . . . . .                   | 46 |
| 4.7. Porcentaje diario del total de tweets clasificados como Tristeza, para cada orientación política . . . . .                | 47 |
| 4.8. Porcentaje diario del total de tweets clasificados como Asco, para cada orientación política . . . . .                    | 48 |

# Índice de cuadros

|                                                                              |    |
|------------------------------------------------------------------------------|----|
| 3.1. Ejemplos de tweets clasificados . . . . .                               | 21 |
| 3.2. Descripción de las emociones usadas como etiquetas para los tweets .    | 26 |
| 3.3. Cantidad de tweets asignados a cada emocion por etiquetador . . . .     | 28 |
| 3.4. Cantidad de tweets etiquetados e indice de Fleiss Kappa en cada emoción | 30 |
| 3.5. Ejemplos de tweets con las etiquetas asignadas por cada etiquetador .   | 31 |
| 3.6. Cantidad de tweets por cantidad de etiquetas finales asignadas . . . .  | 33 |
| 3.7. Ejemplos de tweets con etiquetas multiples . . . . .                    | 33 |
| 3.8. Descripción de los modelos usados . . . . .                             | 35 |
| 4.1. Desempeño de los modelos entrenados medidos por F1 score . . . . .      | 39 |
| 4.2. Ejemplos de tweets con divergencias entre etiquetas y clasificaciones . | 41 |
| A.1. Hashtags . . . . .                                                      | 59 |

# Capítulo 1

## Introduccion

Este capítulo introduce el tema de detección emocional en el texto. Se aborda brevemente la evolución y el estado actual de este campo de estudio, así como las tecnologías recientes que han sido empleadas y que son aplicadas en este trabajo. Se resalta Twitter como una fuente valiosa para explorar los fenómenos sociales y se plantea el análisis de emociones en tweets sobre las elecciones presidenciales en Colombia en 2022. Finalmente, se destacan las contribuciones de este trabajo en ese contexto.

### 1.1. Extracción de opiniones en el texto

La detección automática de opiniones en el texto ha sido estudiada con creciente interés en los últimos años, donde se han utilizado técnicas de aprendizaje automático para identificar las respuestas subjetivas de las personas en dominios tan diversos co-

mo las reseñas de películas en línea, las noticias, los cuentos infantiles o los blogs. Este campo resulta de particular interés al estudiar la respuesta individual a fenómenos externos, ya que dentro del texto existe tanto información objetiva, que corresponde a hechos verificables, como información subjetiva, que corresponde a los procesos internos que los individuos experimentan y plasman en el texto. Poder identificar esta información subjetiva permite una aproximación a la respuesta interna del autor (VandenBos, 2007).

Inicialmente esta tarea fue reducida a la identificación de sentimientos positivos o negativos en el texto. Sin embargo, eventualmente surgió un interés por la detección de las emociones presentes en el texto. Según la definición de VandenBos (2007), una emoción es una respuesta que involucra elementos experienciales, conductuales y fisiológicos ante un evento personalmente significativo. En ese sentido, el análisis de emociones busca reconocer cómo son plasmadas dichas respuestas en el texto.

Históricamente, esto ha sido logrado con técnicas tradicionales de NLP, que involucran el entrenamiento de modelos de aprendizaje automático, específicamente de aprendizaje supervisado, que buscan predecir una variable objetivo en función de variables predictoras. En ese sentido, la variable a predecir es una etiqueta otorgada al texto, como un sentimiento negativo o positivo, y las variables predictoras se generan a partir del texto, usualmente proveniente de la construcción de n-gramas. En este contexto, contar con recursos etiquetados manualmente en donde ciertos términos fueran asociados a un sentimiento o emoción particular como el propuesto por Strapparava et al. (2004) fue determinante durante el desarrollo inicial del área, pues esto permitía asignar al texto una emoción o sentimiento en particular, basado en la asociación que existiera entre las palabras que lo constituían y dicho sentimiento.

## 1.2. Avances técnicos recientes en análisis de sentimiento

El estudio del análisis de sentimientos en general, y de emociones en particular, se ha realizado a través de técnicas tradicionales de NLP, cuyo enfoque se centra en la relación entre los términos y los estados emocionales. Recientemente, se ha avanzado hacia el uso de algoritmos basados en redes neuronales para capturar un sentido más general del texto basado en la relación que existe entre las palabras que lo componen. Inicialmente, para este propósito se utilizaron las RNN (Redes Neuronales Recurrentes), pero debido a su arquitectura, no era posible suministrar grandes cantidades de texto y captar la información contextual, ya que su entrenamiento se volvía demasiado pesado computacionalmente y la información del inicio del texto era difícilmente asociada a la información del final.

En ese contexto, surge una nueva arquitectura conocida como Transformers (Vaswani et al., 2017), que elimina la estructura recurrente. Este modelo utiliza únicamente múltiples capas de auto-atención. Al eliminar los pasos recurrentes, permitió la paralelización del cálculo y poder entrenar la red con grandes volúmenes de texto de manera eficiente. Esto permite captar información del sentido del texto de una manera más compleja y computacionalmente más eficiente.

BERT (Devlin et al., 2018), es un ejemplo de un modelo basado en Transformers que aprende el sentido del lenguaje de manera general a partir del entrenamiento con grandes volúmenes de datos provenientes de internet, como la Wikipedia. Este modelo pre-entrenado luego se utiliza para diversas tareas de procesamiento de lenguaje natural mediante el fine tuning de la última capa de la red, como por ejemplo el

análisis de sentimiento. En el caso del español, existen distintos modelos basados en BERT específicamente para este idioma, que han sido entrenados a partir de textos provenientes de datos abiertos como el propuesto por Canete et al. (2020).

### **1.3. Análisis de sentimiento en Twitter en torno a política**

En el ámbito de la extracción de opiniones en el texto, Twitter ha sido analizado como un medio particularmente interesante para el estudio indirecto de fenómenos sociales. Esto se debe a su capacidad para captar a una gran cantidad de usuarios que discuten sobre una amplia variedad de tópicos, empleando una cantidad limitada de palabras y un estilo propio. Este enfoque ha sido estudiado en el pasado por O'Connor et al. (2010), quienes analizaron cómo los textos provenientes de esta plataforma coinciden con otras aproximaciones de la realidad social, como las encuestas de opinión, en relación a eventos de relevancia popular. De manera similar, Tumasjan et al. (2010) y Mohammad et al. (2015) han utilizado esta red social para identificar las emociones presentes en los tweets relacionados con temas políticos.

No obstante, en el caso del español existen pocos estudios que se centren en la detección de emociones en redes sociales. Hasta ahora no se tiene conocimiento de ningún estudio que haya empleado modelos de lenguaje basados en Transformers con este propósito, especialmente en el contexto político

## 1.4. Aporte de este trabajo

El presente trabajo tiene como objetivo el empleo de modelos de lenguaje, específicamente modelos basados en BERT, para detectar emociones presentes en Twitter durante las elecciones presidenciales en Colombia en 2022.

Para obtener los datos de entrenamiento, se utilizó una base de datos obtenida a través de la descarga de tweets mediante la API, utilizando hashtags relacionados con el tema de las elecciones, entre el 22 de mayo y el 22 de junio de 2022, período que comprende la primera y segunda vuelta presidencial de las elecciones en Colombia. Estos hashtags y por lo tanto, los tweets relacionados con ellos, fueron asociados por el autor a los sectores políticos de izquierda, derecha o neutro, basándose en el contenido observado en los tweets y en el apoyo o rechazo que estos muestren. A partir de ahí, se procedió a tomar una muestra de 1200 tweets para ser etiquetados manualmente por el autor y los directores, con alguna de las emociones disponibles. Con los tweets etiquetados, se procedió a realizar el fine tuning de varios modelos de lenguaje preentrenados y a evaluar su desempeño para seleccionar el mejor. Este modelo fue utilizado para clasificar toda la base de datos y a partir de esta clasificación, se realizaron agrupaciones que permitieron determinar la respuesta emocional de los tweets asociados a cierto sector político, así como la variación temporal de esta respuesta.

Se observó que las emociones asco y alegría tuvieron una presencia mayor en los tweets etiquetados manualmente con respecto a las emociones tristeza y miedo. Además, usualmente los tweets a los que se les era asignado tristeza o miedo, se les asignaba también asco. Esto provocó un desempeño más pobre con respecto a asco y alegría. En cuanto a los resultados de clasificación obtenidos por el modelo de

mejor desempeño, robertuito, las etiquetas asignadas a los tweets de la base de datos, coincidieron con las respuestas emocionales esperadas en eventos clave durante las elecciones. Se observó además una mayor presencia de la etiqueta alegría en tweets cuya orientación política asignada fue la izquierda y más asco en los tweets a los que se les asignó orientación de derecha. Este fenómeno se explica en parte por el hecho de que el candidato victorioso es de tendencia izquierdista. El estudio pone a disposición el conjunto de datos y el modelo para futuras investigaciones sobre emociones en español en diferentes contextos. elecciones presidenciales en Colombia de 2022.

# Capítulo 2

## Trabajo Previo

El presente capítulo aborda los estudios más relevantes hasta la fecha relacionados con la detección de emociones en el texto, los modelos de aprendizaje automático utilizados en dicha tarea y su aplicación en el contexto de las redes sociales.

Se mencionan estudios que se centran en la detección de opiniones negativas o positivas en textos en distintos dominios, como las calificaciones de películas. Se destaca la importancia que ha tenido en este campo contar con una base de términos clasificados según su orientación emocional, lo cual permite realizar análisis automáticos. Además, se explora la evolución hacia el uso de redes neuronales, específicamente modelos basados en la arquitectura Transformer, como BERT, para capturar el sentido y las relaciones entre palabras en el análisis de texto. También se discute la aplicación de estos modelos en español. Por último, se aborda el análisis de sentimientos y emociones en redes sociales como blogs y Twitter, resaltando su utilidad en la comprensión de la percepción pública y en la identificación de estados emocionales en torno a eventos políticos y sociales.

## 2.1. Aprendizaje supervisado en la detección de emociones

La detección automática de las respuestas internas que las personas reflejan en el texto ante ciertos fenómenos ha sido estudiada con mucho interés en los últimos años. Así lo establecieron Pang et al. (2002), quienes destacaron la importancia de desarrollar sistemas capaces de identificar si existe opinión negativa o positiva ante textos que reflejen opiniones subjetivas. Para este propósito, se emplea el dominio de las reseñas en línea de películas, lo que facilita la tarea de construir algoritmos de aprendizaje supervisado, al contar con una clasificación negativa o positiva del texto por parte del usuario. Estos algoritmos buscan entrenar un modelo a partir de ejemplos previamente clasificados, que sea capaz de predecir una variable objetivo dadas unas variables predictoras. En ese sentido, las variables predictoras se construyen a partir del texto, principalmente mediante la elaboración de unigramas, que son elementos individuales del texto, como las palabras, y como variable objetivo se utiliza la clasificación asignada. Por otro lado, Turney (2002) intentó establecer mediante aprendizaje no supervisado si una determinada reseña en línea sobre diversos temas presenta un sentimiento negativo o positivo. Para ello, utiliza el concepto de orientación semántica presentado por Hatzivassiloglou and McKeown (1997) para determinar si una frase tiene una orientación negativa o positiva, y así poder decidir si la reseña en su conjunto es positiva o negativa. En general, la determinación de la orientación negativa o positiva dentro del texto constituye la detección de la polaridad.

Para llevar a cabo la detección automática de los estados emocionales presentes en el texto, ha sido importante contar con un corpus de términos que cuenten con una

clasificación previa. En este sentido, Strapparava et al. (2004) realizó una anotación manual de los estados emocionales usando las categorías desarrolladas por Ortony et al. (1987), de los términos encontrados en WordNet, una base de datos de términos en inglés agrupados en grupos de sinónimos y con relaciones semánticas entre grupos creada por Miller (1995). A partir de las anotaciones manuales de las emociones de ciertos términos presentes en la base, se estableció la categoría emocional de nuevos términos gracias a los sinónimos y las relaciones semánticas, construyendo así una base de datos de estados emocionales llamada WordNet-Affect. En esta misma línea, Wiebe et al. (2005) elaboró una anotación manual de los estados emocionales presentes en las oraciones de un gran volumen de noticias, teniendo en cuenta el contexto.

En el contexto de la clasificación automática de las emociones en el texto, Alm et al. (2005) utilizaron los cuentos infantiles para desarrollar un modelo de aprendizaje supervisado capaz de detectar emociones en las frases. Para ello, se elaboró una anotación manual de estas frases que constituyen la base de datos y, posteriormente, se generó un conjunto de features para dichas frases que se utilizaron para entrenar un clasificador lineal

## **2.2. Redes neuronales para el análisis de texto**

Las técnicas tradicionales de análisis de emociones en el texto han estado principalmente basadas en la relación que existe entre los términos que conforman el texto y un estado emocional determinado. Aunque este tipo de análisis ha sido de gran utilidad, no posee la capacidad de tener en cuenta el sentido, es decir, la relación

que existe entre las palabras que componen el texto, además del orden en que se presentan. Debido a esto, recientemente el campo de estudio ha migrado hacia el uso de algoritmos capaces de captar esta relación contextual, específicamente las redes neuronales.

Con el objetivo de resumir el estado del arte, Acheampong et al. (2021) realizaron un recuento de las distintas arquitecturas de redes neuronales utilizadas recientemente para esta tarea, incluidas las RNN (redes neuronales recurrentes), que fueron popularizadas entre otros, por Cho et al. (2014). Estas redes tienen una arquitectura adecuada para datos secuenciales, como las tareas de NLP. Son una estructura auto-regresiva, generando la salida del paso  $t+1$  con base en la entrada en el tiempo  $t+1$  y la salida en el tiempo  $t$ . Sin embargo, debido a su arquitectura, la optimización de los pesos que conforman la topología de las RNN se vuelve compleja para el cálculo del gradiente del error, ya que se debe pasar tantas veces por los pesos de la red como pasos en el tiempo existan (como el número de caracteres en una palabra), lo que hace difícil computar secuencias particularmente largas dada la carencia de paralelización del cálculo.

Aunque las RNN permiten preservar de alguna manera la relación contextual del texto al tener una estructura secuencial, esta impide la paralelización de su cómputo, ya que se requieren las salidas de ejemplos anteriores para realizar el siguiente paso en el tiempo. Además, debido a este funcionamiento secuencial, si hay un gran número de pasos en el tiempo, es poco probable que la red considere para un punto muy adelante, información presente al inicio. Para abordar estos inconvenientes, Vaswani et al. (2017) desarrollaron los Transformers, una arquitectura en la que los datos de entrada se introducen en la red de manera simultánea, como todas las palabras en una frase por ejemplo. Esto se logra a través de una arquitectura definida como auto-atención.

La auto-atención permite que una red utilice información de textos arbitrariamente largos de manera eficiente, ya que para cada elemento en la entrada, el modelo accede a todas las entradas hasta ese elemento. Además, el cálculo para cada elemento es independiente de los demás cálculos. Esto posibilita que la representación de cada dato, en este caso cada palabra, considere su importancia con respecto las demás palabras en la oración. Esta transformación, aplicada tanto a los datos de entrada como a los de salida, se emplea para entrenar los pesos de la red. De esta manera, la red puede procesar datos de manera simultánea y tener en cuenta toda la información disponible. Un esquema de esta arquitectura se puede apreciar en la Figura 2.1.

En la figura 2.2 se muestra una comparación entre los mecanismos de auto atención de los Transformers y la estructura recurrente de las RNN, en donde se aprecia la manera de calculo simultaneo de una y secuencial de la otra.

Gracias a la capacidad de paralelización y a la habilidad para conservar la relación entre palabras distantes en el texto que presentan los Transformers, es posible entrenar un modelo de lenguaje con grandes volúmenes de información para así contar con una red de gran poder predictivo. Es precisamente en este contexto que Devlin et al. (2018) desarrollaron un modelo basado en Transformers, capaz de aprender el sentido del lenguaje de manera general y luego utilizar lo aprendido para diversas tareas de procesamiento de lenguaje natural, denominado BERT (Bidirectional Transformers for Language Understanding). Para esto, se emplea una arquitectura donde las capas de entrada son las representaciones vectoriales de las palabras (Embeddings), y a partir de ahí se incluyen múltiples capas de Transformers. En su proceso de entrenamiento, se suministran dos oraciones consecutivas con palabras faltantes y se asignan dos tareas simultáneas: predecir qué palabra podría ser la faltante y determinar el orden de las oraciones. Esto le permite al modelo aprender del

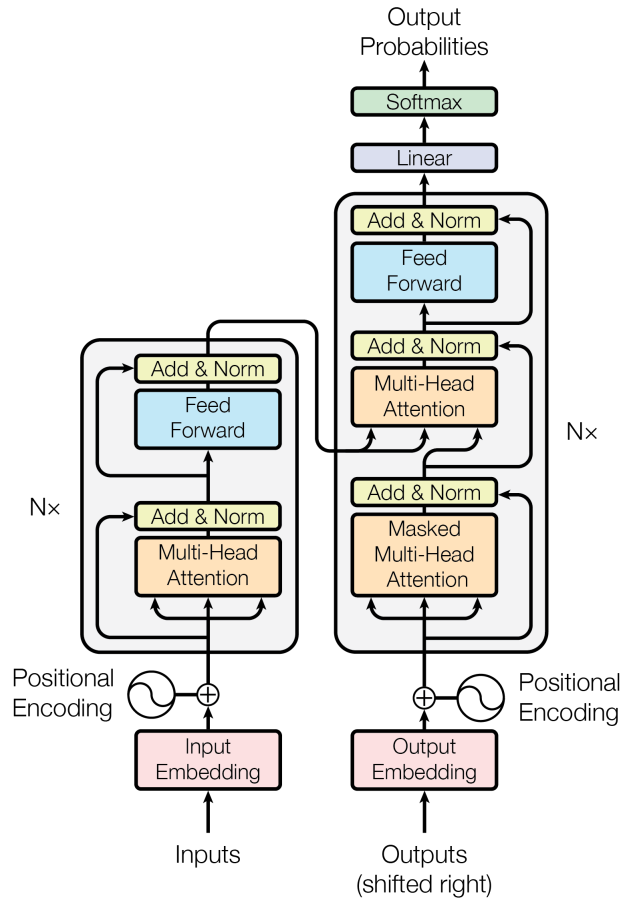


Figura 2.1: Arquitectura de los Transformers

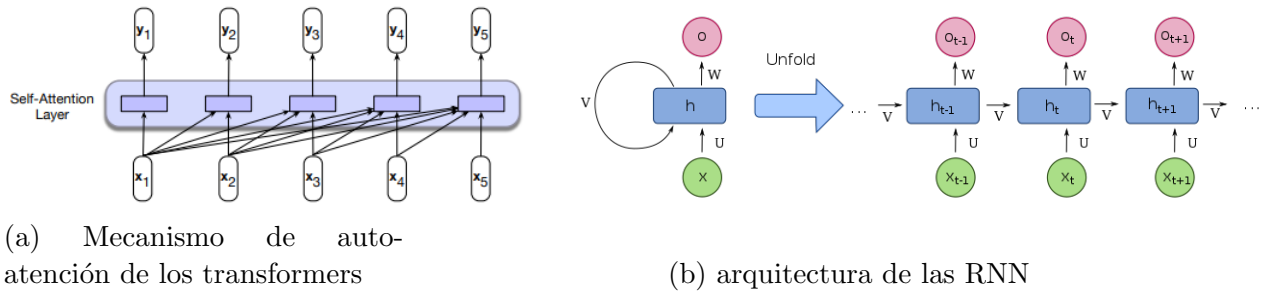


Figura 2.2: Comparación entre mecanismos de redes. Fuentes Jurafsky (2000), Wikipedia contributors (2022)

contexto del lenguaje a nivel de palabras, utilizando tanto las que vienen antes como las que vienen después como fuentes de información, y también aprender del contexto de las oraciones al identificar su secuencia. Para ello, se utilizó la totalidad de Wikipedia en su fase de entrenamiento. Una vez preentrenado, este modelo es capaz de utilizar su representación contextual del lenguaje para realizar diversas tareas de procesamiento de lenguaje natural, como análisis de sentimiento. Esto se logra agregando una capa final a la red que genere un posible resultado, por ejemplo, mediante una función softmax que permita asignar probabilidades a distintas clases y luego comparar este resultado con los valores esperados. El esquema de esta arquitectura se puede apreciar en la Figura 2.3.

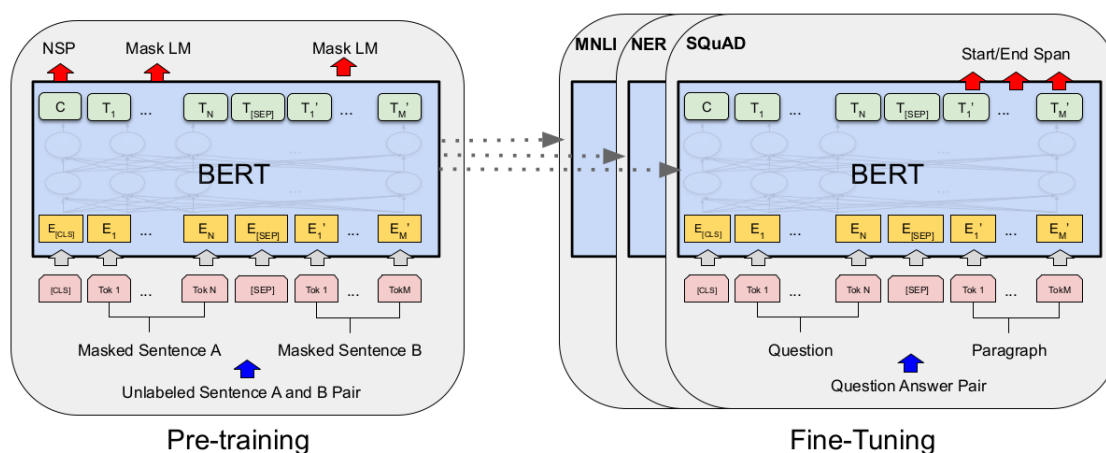


Figura 2.3: Arquitectura de BERT. Fuente Devlin et al. (2018)

En lo que respecta a la aplicación de modelos tipo BERT en español, Canete et al. (2020) destacan que hasta la fecha no existía un modelo de este tipo entrenado específicamente para este lenguaje, además del modelo BERT multilingüe. Por esta razón, se propusieron entrenar un modelo destinado al español, utilizando además de la Wikipedia en español, textos provenientes de publicaciones de las Naciones

Unidas, gobiernos y charlas TED. El resultado fue un modelo que supera al BERT multilingüe en múltiples tareas de evaluación para el español.

## 2.3. Sentimientos y emociones en redes sociales

El campo de estudio del análisis de sentimiento del texto en Internet ha cobrado cada vez más importancia, tanto para usuarios individuales como para la industria de la publicidad, el mercado financiero y la academia tal como lo describen Pang et al. (2008). Por lo tanto, los autores realizan un repaso de las distintas técnicas y aplicaciones que consideran relevantes hasta la fecha.

Un ámbito particular donde los usuarios individuales generan grandes volúmenes de información sobre diversos temas, y es por ende una fuente rica de datos para el análisis de sentimiento, es el de los blogs. Es por esta razón que Aman and Szpakowicz (2007) emplean textos provenientes de estos para realizar la detección de emociones presentes en las oraciones. Para lograr esto, primero realizan una anotación manual de las emociones presentes en las oraciones y luego construyen features para entrenar distintos modelos supervisados.

Twitter es un sitio de blogs en particular con un formato de microblogging, lo que significa publicaciones breves llamadas "tweets". Este sitio ha ganado gran relevancia en el análisis del texto debido a su popularidad entre usuarios de Internet de diversos tipos, desde marcas hasta individuos y políticos que abordan una amplia variedad de temas. En este contexto, Pak and Paroubek (2010) extraen tweets de diferentes usuarios que tratan sobre distintos temas para llevar a cabo un análisis de sentimientos. Para ello, identifican tweets que contienen emoticones felices y tristes,

a los cuales etiquetan como reflejo de sentimientos positivos o negativos, respectivamente. También consideran tweets provenientes de cuentas de medios de noticias, los cuales etiquetan como neutrales. Luego, construyen features utilizando n-gramas a partir de las palabras presentes en los tweets, y utilizan estos features para entrenar varios clasificadores.

Reconociendo la necesidad de contar con un corpus etiquetado como base para la identificación de emociones en Twitter, Roberts et al. (2012) seleccionan 14 temas con contenido emocionalmente fuerte y sus palabras clave asociadas, para usarlas como hashtags en la extracción de tweets. Luego, etiquetan manualmente los tweets con sus respectivas emociones. Esto proporciona una base de datos para poder entrenar un modelo de aprendizaje supervisado y evaluar así el poder predictivo de los modelos a partir de datos etiquetados.

La identificación del sentimiento en Twitter puede servir como una aproximación a la realidad social de los usuarios. Por esta razón, O'Connor et al. (2010) plantean la pregunta de si existe una correlación entre el sentimiento presente en Twitter y las encuestas de opinión. Para investigarlo, extraen mil millones de tweets entre 2008 y 2009, seleccionando aquellos que contienen palabras clave asociadas a los temas en estudio. Determinan el sentimiento de estos tweets basándose en la proporción de palabras con connotaciones negativas o positivas presentes en un tema particular en un día dado. Los resultados muestran una correlación significativa entre el sentimiento reflejado en los tweets y los resultados de las encuestas. Esta relación entre la realidad social y el sentimiento en los tweets también es explorada por Bollen et al. (2011), quienes miden el estado emocional de una muestra de tweets entre agosto y diciembre de 2008. Lo hacen evaluando la similitud entre las palabras de los tweets y ciertos términos clave asociados a estados emocionales. A través de esto, identifican

que ciertos eventos relevantes generan un impacto emocional significativo y duradero en los usuarios.

Dado que el ámbito político es un caso particular dentro de los fenómenos sociales y su impacto se refleja en el estado emocional de las personas, el sentimiento presente en los tweets relacionados con la política puede proporcionar un indicio de la percepción pública de ese ámbito. Con este enfoque, Tumasjan et al. (2010) proponen el uso de Twitter como plataforma para medir la percepción pública sobre la política durante las elecciones parlamentarias en Alemania en 2009. Una de sus preguntas de investigación se centra en los sentimientos expresados en los tweets que mencionan a los políticos en campaña. Utilizan un software capaz de identificar palabras clave asociadas a estados emocionales y cognitivos en el texto. El resultado es un perfil emocional para cada político que en términos generales, parece concordar con su discurso político.

Por otro lado, el análisis del contenido emocional presente en los tweets durante las elecciones presidenciales de 2012 en Estados Unidos fue abordado por Mohammad et al. (2015). En su estudio, se valieron de palabras clave específicas relacionadas con las elecciones para extraer los tweets pertinentes. Con el fin de analizar el contenido emocional de los tweets, los autores elaboraron un cuestionario detallado para etiquetar cada tweet con información precisa sobre sus sentimientos, emociones y la intención percibida del autor. Este cuestionario se diseñó de manera colaborativa y exhaustiva, considerando diferentes aspectos que podrían influir en la interpretación de los sentimientos, como la polaridad negativa o positiva, la emoción predominante y la intención subyacente detrás del tweet. Una vez diseñado el cuestionario, los autores lo sometieron a una plataforma de crowdsourcing en Amazon, donde participaron alrededor de 3000 etiquetadores en el proceso. Cada tweet fue evaluado

por varios etiquetadores, y la etiqueta final se determinó a partir de un consenso en los resultados de los etiquetadores. Este minucioso proceso de etiquetado permitió determinar la presencia de contenido emocional en aproximadamente 2000 tweets, y en caso de que existiera, identificar qué tipo de emoción estaba presente. Las opciones emocionales disponibles en el cuestionario fueron un total de 19, las cuales fueron posteriormente agrupadas en 8 categorías emocionales distintas. Las etiquetas finales derivadas del proceso de etiquetado se utilizaron para entrenar un modelo de aprendizaje supervisado, logrando un F1 score del 58.3%. El corpus etiquetado resultado de este proceso, fue puesto a disposición de la comunidad como recurso para investigaciones futuras. Dada la relevancia y proximidad de esta investigación al presente trabajo, este estudio se convirtió en una referencia crucial, inspirando en parte la metodología empleada.

# Capítulo 3

## Metodología

Este capítulo aborda la metodología empleada en el presente trabajo, la cual se centra en el análisis de tweets relacionados con las elecciones presidenciales en Colombia durante el período de mayo a junio de 2022. El capítulo comienza describiendo la sección de Datos, donde se detalla el proceso de obtención, análisis exploratorio y etiquetado de los datos recopilados. Posteriormente, en la sección de Modelos, se presentan los modelos de lenguaje utilizados y se expone el método correspondiente de entrenamiento y evaluación.

### 3.1. Datos

En esta sección se aborda la etapa inicial de recolección de datos, llevada a cabo entre la primera y segunda vuelta electoral, utilizando hashtags con contenido político. Luego se detalla el proceso de filtrado aplicado a esta base de datos para

retener únicamente los tweets relevantes, seguido de un análisis exploratorio de la información obtenida. También se describe el uso de los hashtags como herramienta para clasificar la orientación política los tweets, basándose en el contenido presente en los tweets que hacen uso de dichos hashtags.

Finalmente, se presenta el proceso de etiquetado manual de emociones en una muestra de los tweets recopilados. Se explica en detalle cómo se llevó a cabo este etiquetado, en el cual se eligieron 14 emociones posibles. Además, se establecen correlaciones entre las etiquetas asignadas por los etiquetadores, lo que resulta en la creación de etiquetas agrupadoras basadas en estas correlaciones. El proceso de llegada a las etiquetas finales para cada tweet se describe en profundidad, y se discute el nivel de acuerdo alcanzado entre los diferentes etiquetadores involucrados en el proceso.

### **3.1.1. Recolección y análisis exploratorio de datos**

El dataset inicialmente recopilado consta de 585,001 tweets obtenidos entre el 22 de mayo y el 22 de junio de 2022, período en el que tuvieron lugar la primera y la segunda vuelta de las elecciones presidenciales en Colombia, el 29 de mayo y el 19 de junio respectivamente. Para la extracción de los datos, se utilizaron 173 tendencias, es decir, hashtags por día, con contenido político que estuvieron presentes durante este lapso. Estos hashtags se obtuvieron de diferentes sitios web <sup>1 2 3</sup>, donde se detallan los hashtags que fueron tendencia en distintos países para fechas específicas. De allí,

---

<sup>1</sup><https://getdaytrends.com/>

<sup>2</sup><https://archive.twitter-trending.com/>

<sup>3</sup><https://www.exportdata.io/trends/worldwide>

se seleccionaron los hashtags que estuvieron en tendencia y estaban relacionados con el tema de las elecciones en Colombia durante el periodo de estudio.

El dataset recolectado fue sometido a un proceso de filtrado para eliminar aquellos tweets con menos de 5 palabras, los que tenían una proporción de menciones o hashtags superior al 20 % del total del texto y aquellos que contenían enlaces o provenían de usuarios con un número atípico de publicaciones. Esto redujo la base de datos a 193,348 tweets. A los hashtags utilizados les fue asignada alguna de tres orientaciones políticas: Izquierda, Derecha y Neutro. Esta asignación estuvo basada en la lectura de múltiples tweets que contuvieran el hashtag, el contenido asociado a ellos y la inclinación política percibida en términos generales. En el Cuadro A.1 en la sección de apéndices, se presentan los hashtags junto con su orientación política y la cantidad de tweets asociados a cada uno. A lo largo de este trabajo, cuando se hable de sector político, se hará referencia a esta clasificación.

Cabe destacar que la asignación de un sector político en particular a un hashtag se llevó a cabo a partir de la tendencia política observada en los tweets, y aunque en términos generales los tweets asociados al hashtag presentan cierta inclinación, esto no excluye la presencia de tweets cuya tendencia política sea contraria. En el Cuadro 3.1 se proporcionan ejemplos de tweets que muestran la relación del hashtag con la orientación política. Por ejemplo, el hashtag #EstallidoSocialEs fue catalogado como de derecha debido a que generalmente presenta tweets como el primer ejemplo. Sin embargo, también existen tweets que lo utilizan para expresar apoyo al candidato de izquierda, como se observa en el segundo ejemplo. De manera similar, el hashtag #YaEsSuficiente suele ser utilizado mayormente para respaldar a la izquierda, como se ve en el tercer ejemplo. No obstante, también se encuentran casos en los que se utiliza para apoyar a un candidato de derecha, como se muestra en el cuarto ejemplo.

Cuadro 3.1: Ejemplos de tweets clasificados

| Numero de Ejemplo | Tweet                                                                                                                                                                                                                                                                                  | Orientacion política |
|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| 1                 | @lcvelez @lafm #EstallidoSocialEs el arma de terror de Petro para obligar a votar por él.                                                                                                                                                                                              | Derecha              |
| 2                 | Colombia va por el cambio, a redoblar esfuerzos estos 4 días para derrotar a la corrupción. #PetroYFranciaSonElCambio #EstallidoSocialEs                                                                                                                                               | Derecha              |
| 3                 | #YaEsSuficiente Que los medios proclives al gbno, traten de darle aire boca a boca a un moribundo electoral, FICO. Ante su estancamiento en las encuestas y la distancia que le ha tomado Petro, pretenden en 1 acto de desesperacion, el insuflarlo de votantes de los cuales carece. | Izquierda            |
| 4                 | #YaEsSuficiente de mentir sobre @ingrodolfohdez , vayan a bucaramanga, vean lo que hizo y ahí si hablen..                                                                                                                                                                              | Izquierda            |

La distribución de los hashtags segun la orientación política asignada se refleja en la Figura 3.1, donde se puede apreciar que el sector neutro comprende más del 40 % del total de los hashtags, mientras que la izquierda y la derecha tienen una cantidad similar, alrededor de un 28 % cada uno.

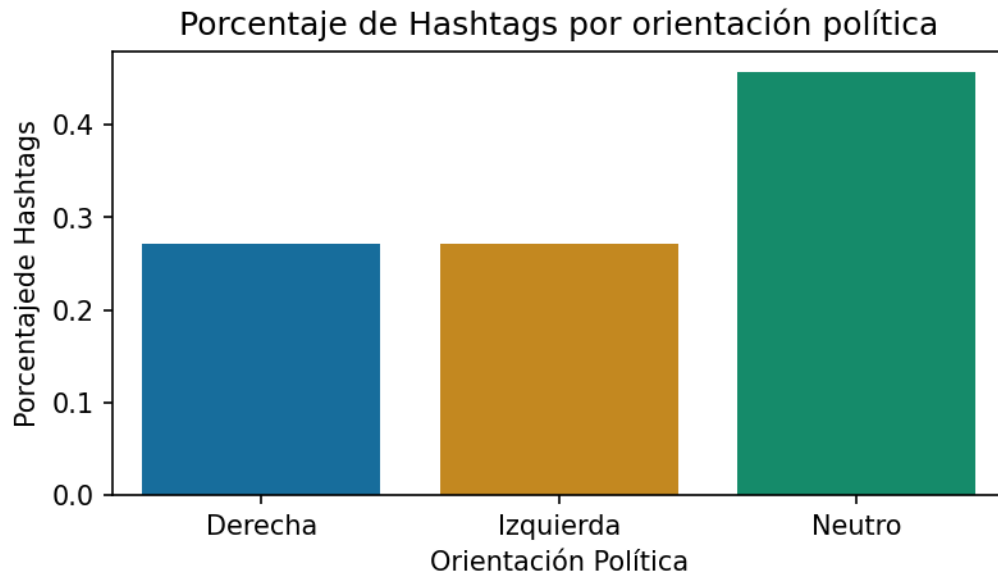


Figura 3.1: Porcentaje de Hashtags según la orientación política asignada

De manera similar, la Figura 3.2 muestra la distribución de los tweets según la orientación política asignada a los hashtags asociados. En esta Figura se aprecia que

el sector neutro concentra la mayoría de los tweets, con más del 46 %, seguido por la izquierda con un 33 % y finalmente la derecha con cerca de un 29

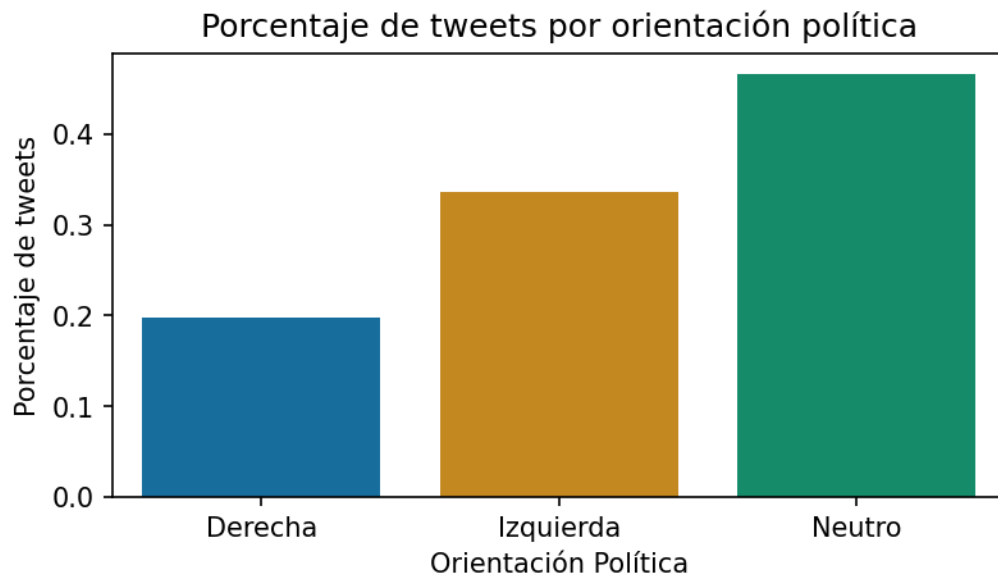


Figura 3.2: Porcentaje de Tweets según la orientación política asignada

Al analizar la cantidad de tweets por orientación política asignada a lo largo del tiempo, se obtienen los resultados observados en la Figura 3.3, que muestra el porcentaje de tweets de cada día para cada sector, en relación con el total de tweets dicho sector. Se puede notar que algunas fechas resultaron particularmente importantes: el 24 de mayo fue un día de debate y destaca el sector neutro, el 29 de mayo fue el día de la primera vuelta electoral, destacando los tres sectores, el 9 de junio fue cuando se divulgaron los llamados "Petro videos", unos videos filtrados que mostraban discusiones políticas del equipo de campaña de Petro, sobresaliendo la derecha. Además, las fechas cercanas al 19 de junio, día de la segunda vuelta, también presentan un aumento en la actividad y destacan los tres sectores.

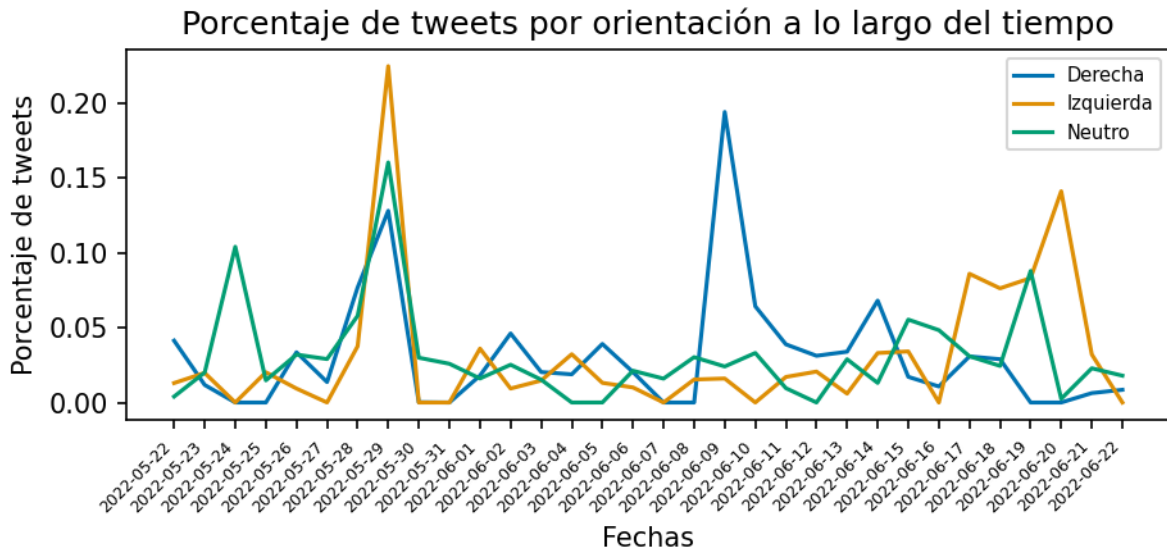


Figura 3.3: Porcentaje de tweets según orientación política a lo largo del tiempo

### 3.1.2. Etiquetado

Se procedió a etiquetar manualmente un conjunto de 1200 tweets seleccionados mediante una muestra aleatoria estratificada proporcional a la cantidad de tweets presentes en los hashtags. Las etiquetas asignadas a cada tweet correspondieron a las emociones identificadas en los mismos.

Una emoción, según la definición de la APA en su Diccionario de Psicología elaborado por VandenBos (2007), es un patrón complejo de reacción que involucra elementos experienciales, conductuales y fisiológicos. A través de estos elementos, un individuo intenta abordar un asunto o evento que tiene significado personal. La cualidad específica de la emoción, como el miedo o la ira, está determinada por el significado particular que el individuo atribuye al evento. Por ejemplo, si el evento implica una amenaza, es probable que se genere miedo. Las emociones consisten en

tres componentes distintos: una experiencia subjetiva, una respuesta fisiológica y una respuesta conductual o expresiva.

Dentro del contexto de este trabajo, los componentes de experiencia subjetiva y respuesta fisiológica no son accesibles. Por lo tanto, el enfoque principal durante el proceso de etiquetado se centró en la respuesta expresiva, específicamente en cómo el autor del tweet comunica a través del texto la respuesta emocional generada hacia la entidad a la cual va dirigido el tweet.

El proceso de etiquetado consistió en asignar emociones a un tweet a la vez, utilizando una plataforma de etiquetado con un esquema de selección múltiple. Se permitía asignar una o varias emociones al tweet, siguiendo las directrices establecidas en el manual de etiquetado <sup>4</sup>. Esta tarea fue llevada a cabo por el autor y los directores, generando así tres etiquetas independientes para cada tweet.

Para el esquema de etiquetado, se tomó como referencia el enfoque de Mohammad et al. (2015), en el cual el etiquetador responde varias preguntas y, al ser preguntado sobre la emoción presente en el tweet, puede seleccionar entre 19 emociones distintas. En este trabajo, después de una prueba iterativa para elegir las etiquetas más adecuadas, se llegó a la elección de 14 emociones posibles, además de la categoría ".otra". Estas emociones incluyen alegría, agrado, confianza, admiración, miedo, incertidumbre, sorpresa, asombro, tristeza, decepción, asco, desagrado, ira, odio y otra. En el Cuadro 3.2 se presenta una descripción detallada de las emociones seleccionadas, basada en definiciones encontradas en el Diccionario de Psicología de la APA,

---

<sup>4</sup>[https://docs.google.com/document/d/1hoUYKMaYHSeGe0Q2FqRVyahTin6T09Mu8wt1\\_HY6200/edit?usp=sharing](https://docs.google.com/document/d/1hoUYKMaYHSeGe0Q2FqRVyahTin6T09Mu8wt1_HY6200/edit?usp=sharing)

el Diccionario de Oxford <sup>5</sup>, el de la RAE <sup>6</sup> y Wikipedia <sup>7</sup>.

| Emoción       | Descripción                                                                                                                                                                                                                                                                                    |
|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Admiración    | La admiración es una emoción social que se siente al observar a personas de competencia, talento o habilidad que superan los estándares. La admiración facilita el aprendizaje social en grupos. La admiración motiva la superación personal a través del aprendizaje de los modelos a seguir. |
| Agrado        | Sensación moderada de felicidad o placer que siente una persona por algo que le gusta.                                                                                                                                                                                                         |
| Confianza     | La confianza implica que una parte se vuelve vulnerable ante otra, asumiendo que esta actuará en su beneficio. En una relación de confianza, el que confía no controla las acciones del otro.                                                                                                  |
| Alegría       | Es una emoción positiva que suele ir acompañada de bienestar. Se genera como resultado de un evento positivo.                                                                                                                                                                                  |
| Incertidumbre | La incertidumbre es la falta de seguridad, de confianza o de certeza sobre algo. Aparece en situaciones en las que no tenemos control total, en las que nos faltan respuestas e información, y nos puede generar inquietud, inseguridad, estrés, ansiedad e incluso miedo                      |
| Miedo         | El miedo surge ante amenazas reales o imaginarias de daño físico, emocional o psicológico. En textos, se muestra como amenazas hacia el autor del mensaje o lo que se menciona, cuando está vulnerable o en desventaja.                                                                        |
| Asombro       | La condición de estar asombrado; un estado de asombro abrumador, como por sorpresa o miedo repentino, horror o admiración.                                                                                                                                                                     |
| Sorpresa      | Se define como una reacción provocada por algo inesperado, extraño o novedoso para la persona. En el texto está principalmente asociada a resultados inesperados o descubrimientos singulares respecto al enunciado del tweet.                                                                 |
| Decepción     | La decepción es la insatisfacción que sigue al fracaso de expectativas o esperanzas. A diferencia del arrepentimiento que se centra en elecciones personales, la decepción se enfoca en el resultado en sí. Puede generar estrés psicológico.                                                  |
| Tristeza      | La tristeza es un dolor emocional causado por decadencia espiritual, manifestándose en llanto, abatimiento, falta de apetito, cansancio, etc. Ocurre cuando las expectativas no se cumplen o las circunstancias son dolorosas.                                                                 |
| Desagrado     | Una actitud o un sentimiento de disgusto o aversión.                                                                                                                                                                                                                                           |
| Asco          | Contiene una serie de estados con intensidades variables que van desde una leve aprensión hasta una intensa repulsión. Todos los estados de asco se desencadenan por la sensación de que algo es aversivo, repulsivo y/o tóxico.                                                               |

Cuadro 3.2: Descripción de las emociones usadas como etiquetas para los tweets

Continúa en la página siguiente

<sup>5</sup><https://www.oed.com>

<sup>6</sup><https://www.rae.es/>

<sup>7</sup><https://www.wikipedia.org/>

| Emoción | Descripción                                                                                                                                                                                                                                      |
|---------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Odio    | El odio es una intensa respuesta emocional negativa hacia ciertas personas, cosas o ideas, generalmente relacionadas con la oposición o repulsión hacia algo. El odio a menudo se asocia con intensos sentimientos de ira, desprecio y disgusto. |
| Ira     | La ira surge por objetivos no alcanzados o trato injusto, pudiendo ser peligrosa y relacionada con la violencia. En el texto, el autor reta o reclama por un derecho vulnerado, buscando justicia.                                               |

Cuadro 3.2: Descripción de las emociones usadas como etiquetas para los tweets

Este etiquetado se llevo a cabo usando la plataforma web Label Studio <sup>8</sup> que fue suministrada para cada uno de los etiquetadores, es decir el autor y los directores, a donde se cargaron los tweets a etiquetar. La plataforma permite definir una interfaz de etiquetado personalizada que es presentada en cada uno de los datos a etiquetar, en este caso, cada tweet. Para definir esta interfaz, se plantearon las preguntas necesarias para poder etiquetar adecuadamente los tweets, basándose en el cuestionario planteado por Mohammad et al. (2015). Esto se llevo a cabo mediante un proceso iterativo en donde se plantearon algunas preguntas, se etiquetaron algunos tweets a partir de ellas y basándose en la experiencia de etiquetado y el acuerdo obtenido, se replantaron las preguntas. El resultado final de este proceso fue la interfaz de etiquetado que se puede observar en la Figura 3.4. Cabe resaltar que para el caso de las emociones, se permitió al etiquetador asignar múltiples etiquetas a cada tweet. Por otro lado, aun cuando la categoría "Neutro" no se encuentra disponible para ser seleccionada, aquellos tweets para los cuales la pregunta "¿Existe contenido emocional en este tweet?" fue respondida con un "No", fueron catalogados con dicha etiqueta.

A través del uso de esta interfaz, cada uno de los etiquetadores logro proporcionar sus respuestas las preguntas planteadas para cada uno de los 1200 tweets tomados

---

<sup>8</sup><https://labelstud.io/>

En mi casa esta prohibido hablar de política, por múltiples diferencias, pero por aquí, QUE VIVA PETRO!!!  
#PetroPresidenteEnPrimera #PetroEsElCambio

**¿Es relevante el presente tweet para el analisis de emociones durante la campaña presidencial?**

☒ Si<sup>[1]</sup> ☐ No<sup>[2]</sup>

**¿Existe contenido emocional en este tweet?**

☒ Si<sup>[3]</sup> ☐ No<sup>[4]</sup>

**¿Esta el texto escrito de manera ironica?**

☐ Si<sup>[5]</sup> ☒ No<sup>[6]</sup>

**¿Que emocion describe mejor el texto?**

☐ Admiración<sup>[7]</sup> ☐ Asombro<sup>[8]</sup> ☐ Ira<sup>[9]</sup> ☐ Decepción<sup>[10]</sup> ☐ Asco<sup>[11]</sup> ☐ Desagrado<sup>[12]</sup> ☐ Miedo<sup>[13]</sup> ☐ Odio<sup>[14]</sup> ☐ Incertidumbre<sup>[15]</sup> ☒ Alegría<sup>[16]</sup>

☐ Agrado<sup>[17]</sup> ☐ Tristeza<sup>[18]</sup> ☐ Sorpresa<sup>[19]</sup> ☐ Confianza<sup>[20]</sup> ☐ Otra<sup>[x]</sup>

**¿Si en la pregunta anterior eligio Otra, con que emocion describiria mejor el tweet?**

Figura 3.4: Interfaz de etiquetado

como muestra. Este proceso produjo una base de datos por etiquetador que al ser consolidadas, permitió calcular la cantidad de tweets etiquetados según la emoción asignada que se observan en el Cuadro 3.3.

| Emocion       | Cantidad etiquetador 1 | Cantidad etiquetador 2 | Cantidad etiquetador 3 |
|---------------|------------------------|------------------------|------------------------|
| Alegría       | 61                     | 107                    | 125                    |
| Agrado        | 2                      | 36                     | 59                     |
| Confianza     | 417                    | 242                    | 395                    |
| Admiración    | 79                     | 25                     | 292                    |
| Miedo         | 84                     | 98                     | 76                     |
| Incertidumbre | 37                     | 28                     | 74                     |
| Sorpresa      | 6                      | 3                      | 4                      |
| Asombro       | 1                      | 15                     | 5                      |
| Tristeza      | 18                     | 10                     | 57                     |
| Decepción     | 137                    | 119                    | 74                     |
| Asco          | 82                     | 119                    | 374                    |
| Desagrado     | 494                    | 357                    | 463                    |
| Ira           | 26                     | 111                    | 161                    |
| Odio          | 23                     | 0                      | 115                    |
| Otra          | 4                      | 11                     | 21                     |
| Neutro        | 107                    | 111                    | 115                    |

Cuadro 3.3: Cantidad de tweets asignados a cada emocion por etiquetador

Al analizar las etiquetas asignadas por los etiquetadores, se encontró que estos asignaban etiquetas diferentes pero que estaban cercanas en un sentido semántico a algunos tweets. Para poder cuantificar esta situación, se procedió a calcular la correlación que tuvieron las etiquetas asignadas por los etiquetadores, obteniendo así los resultados observados en la Figura 3.5

Estos resultados llevaron a una conclusión similar a la planteada por Mohammad et al. (2015), donde resulta útil agrupar las etiquetas iniciales de manera que aquellas con cercanía semántica y, por ende, una correlación considerable en los resultados

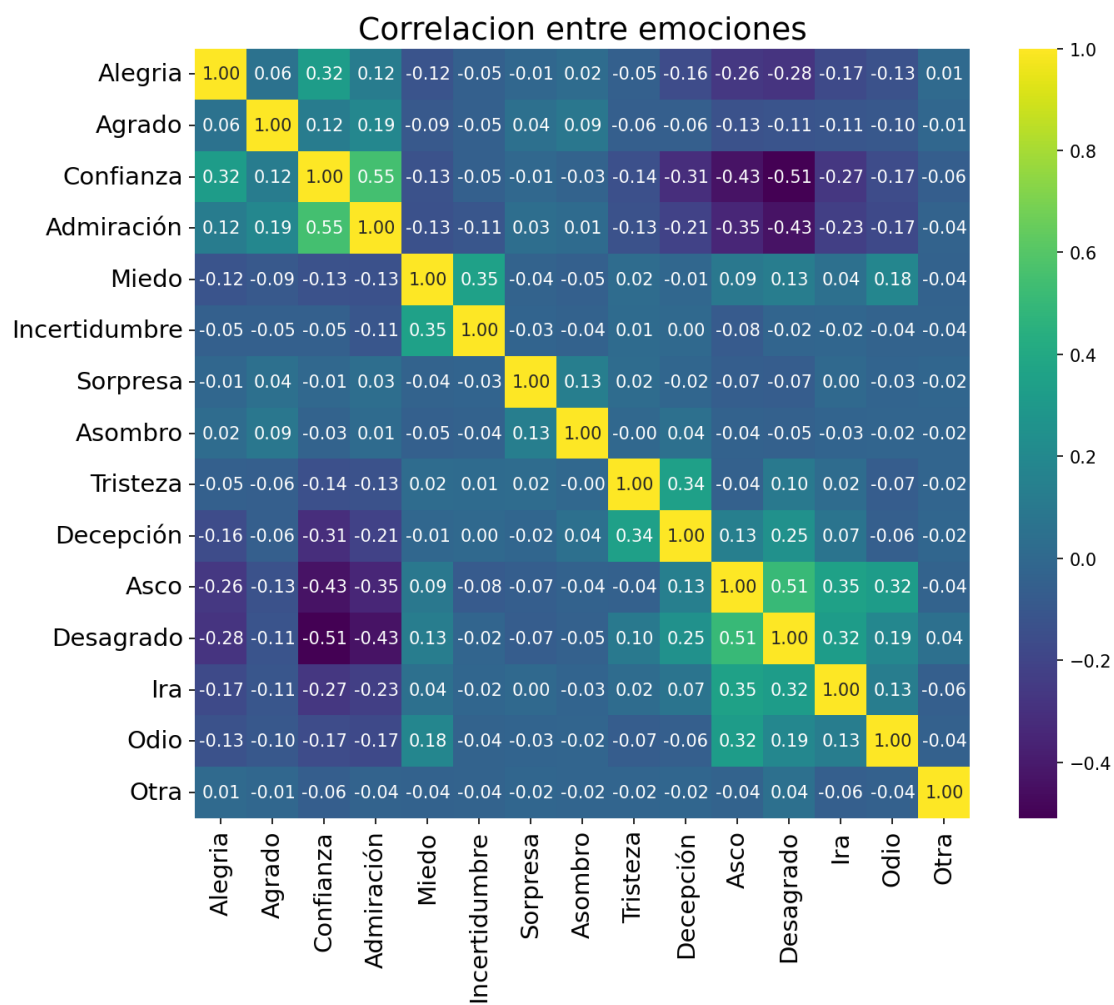


Figura 3.5: Porcentaje de correlación entre etiquetadores de las emociones asignadas a los tweets'

obtenidos, sean asignadas a un mismo grupo. En este sentido, se construyeron las siguientes 4 etiquetas agrupadoras: Alegría, que incluye las etiquetas Alegría, Agrado, Confianza y Admiración; Miedo, que incluye las etiquetas Miedo e Incertidumbre; Tristeza, que incluye las etiquetas Tristeza y Decepción; y Asco, que incluye las etiquetas Asco, Desagrado, Ira y Odio.

Para asignar las etiquetas finales a los tweets, se empleó el mismo sistema utilizado por Mohammad et al. (2015), donde más de la mitad de los etiquetadores debían estar de acuerdo en una etiqueta específica para que esta pudiera ser asignada. De esta manera, cada etiqueta agrupadora se asignó a un tweet si al menos dos de los etiquetadores le habían asignado alguna de las etiquetas que la componen. Además, debido a que se permitió a los etiquetadores asignar múltiples etiquetas a cada tweet, esto permitió la asignación de más de una etiqueta agrupadora.

El acuerdo observado entre los etiquetadores para cada una de estas 4 etiquetas agrupadoras se midió utilizando el índice de Fleiss Kappa. Los resultados se presentan en el Cuadro 3.4. Se puede observar que a la gran mayoría de los tweets se les asignó las etiquetas de Asco y Alegría con 464 y 580 tweets respectivamente.

|                       | alegria | miedo | tristeza | asco |
|-----------------------|---------|-------|----------|------|
| Cantidad de Tweets    | 464     | 98    | 103      | 580  |
| Indice de Fleiss Kapp | 0.69    | 0.47  | 0.4      | 0.62 |

Cuadro 3.4: Cantidad de tweets etiquetados e índice de Fleiss Kappa en cada emoción

Se aprecia así mismo cómo Alegría y Asco tuvieron puntajes relativamente altos en comparación con Tristeza y Miedo. Esto se explica en parte por la cercanía que estas dos últimas tienen con Asco, como se puede ver en el tweet número 1 del Cuadro 3.5. En él, dos de los etiquetadores coincidieron en asignar emociones relacionadas

con Tristeza, mientras que el tercero asignó emociones relacionadas con Asco. Este mismo fenómeno ocurre con Miedo, como se aprecia en el tweet número 2 del Cuadro. Allí también fue el caso en que uno de los etiquetadores asignó emociones relacionadas con Asco, mientras que los otros dos asignaron emociones relacionadas con Miedo.

| Numero de Ejemplo | Tweet                                                                                                                                                   | Etiquetador 1         | Etiquetador 2 | Etiquetador 3    |
|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|---------------|------------------|
| 1                 | #LoPeorDeEstasElecciones es la división de la gente de este país esta gente y cosas ya que pasan                                                        | Decepción             | Decepción     | Desagrado        |
| 2                 | @Zuletalleras Son 4 años O es la. Derecha va realizar un golpe de estado? No cree a sus comentarios son ofensivos e incendiarios.... #PetroEsPresidente | Miedo, Incerti-dumbre | Miedo         | Asco, Des-agrado |

Cuadro 3.5: Ejemplos de tweets con las etiquetas asignadas por cada etiquetador

La coincidencia en las etiquetas finales asignadas a los tweets , se puede observar en la Figura 3.6, donde se representa la co-ocurrencia de estas. En esta Figura, se destaca que la etiqueta de asco fue la que se asignó conjuntamente con mayor frecuencia a todas las demás etiquetas.

En este contexto, el Cuadro 3.6 presenta la cantidad de tweets etiquetados con diferentes números de etiquetas. Los tweets con 0 etiquetas corresponden a casos en los que los etiquetadores coincidieron en la ausencia de contenido emocional o hubo falta de acuerdo en la asignación de etiquetas. Es notable que la mayoría de los tweets se etiquetaron con una sola etiqueta, aunque también hay una cantidad significativa con múltiples etiquetas.

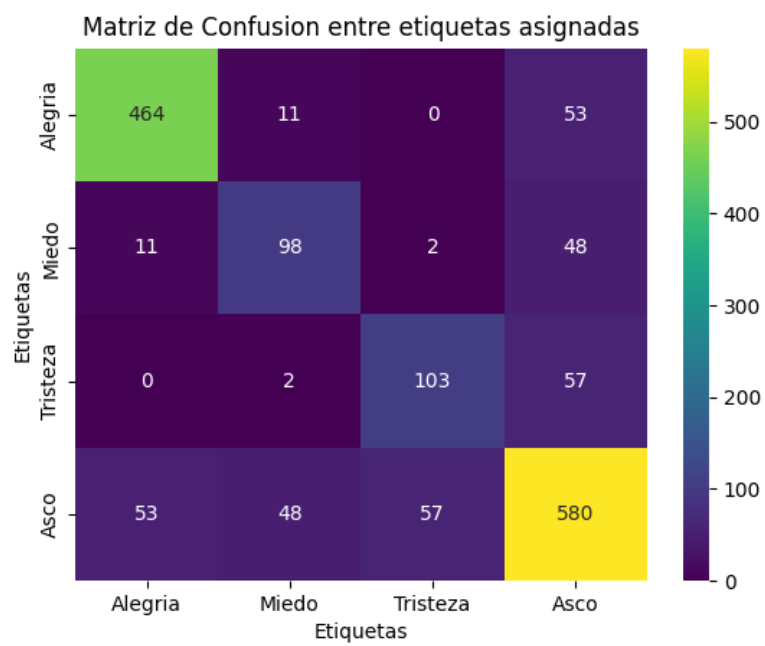


Figura 3.6: Matriz de confusión entre etiquetas asignadas

| Cantidad de etiquetas | Cantidad de Tweets |
|-----------------------|--------------------|
| 0                     | 123                |
| 1                     | 909                |
| 2                     | 165                |
| 3                     | 2                  |

Cuadro 3.6: Cantidad de tweets por cantidad de etiquetas finales asignadas

En el Cuadro 3.7 se presentan algunos ejemplos de tweets con múltiples etiquetas en la asignación final. Por ejemplo, el primer tweet comparte las etiquetas de alegría y asco debido a que el autor refleja diferentes emociones en frases distintas del tweet. Además, hay casos en los que una misma frase o todo el contenido del tweet refleja múltiples emociones para los etiquetadores. Tal es el caso de los ejemplos 2 y 3, que comparten las etiquetas de tristeza y asco, así como el ejemplo 4, que comparte las etiquetas de miedo y asco.

| Numero de Ejemplo | Tweet                                                                                                                                                                                                                                                                              | Etiquetas      |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| 1                 | En #ElDebateDefinitivo Fajardo con sus preguntas hizo subir de votos a @petrogustavo en los votos de los indecisos. La ausencia del ingeniero ayudó a Petro, Fico pierde sólo. #PetroGanaDebate, es el único candidato que gana con argumentos.                                    | alegria, asco  |
| 2                 | Comenzo el entrampamiento público en el debate presidencial. Ya ni disimulan. #ElDebateDefinitivo                                                                                                                                                                                  | tristeza, asco |
| 3                 | Fajardo quedo tan peinado que ya se quiere ir #ElDebateDefinitivo                                                                                                                                                                                                                  | tristeza, asco |
| 4                 | En conclusión #ElDebateDefinitivo deja claro que Petro es un peligro. Mentiroso, arrogante, tramposo, encubridor, manipulador, violento, pasivo agresivo, resentido. En resumen, una porquería de ser humano, qué digo humano... un demonio. Petro es el Pol Pot de Ciénaga de Oro | miedo, asco    |

Cuadro 3.7: Ejemplos de tweets con etiquetas multiples

## 3.2. Experimentos de clasificación

En esta sección, se aborda la descripción, entrenamiento y evaluación de modelos de lenguaje preentrenados. Dado que la tarea principal de este trabajo es analizar las emociones presentes en tweets escritos en español, se han seleccionado seis modelos diferentes debido a su relevancia para dicha tarea. Cada modelo se describe de manera concisa, y se explican las herramientas utilizadas tanto para el entrenamiento como para el almacenamiento de los resultados. Además, se detalla la división de los datos realizada y se mencionan las métricas elegidas para evaluar el entrenamiento y el rendimiento de los modelos en el conjunto de test.

### 3.2.1. Modelos preentrenados

Los modelos de lenguaje preentrenados se encuentran disponibles para su uso en la plataforma Hugging Face <sup>9</sup>, utilizando la librería Transformers <sup>10</sup>. A partir de esta plataforma, y considerando la pertinencia de los modelos para la tarea en cuestión, se han seleccionado seis modelos distintos para ser evaluados y escoger aquel que muestre el mejor desempeño. Los modelos elegidos son RoBERTuito, BERTIN, BETO, ELECTRICIDAD, RoBERTa y twitter-XLM, propuestos por Pérez et al. (2022), la Rosa et al. (2022), Cañete et al. (2020), Romero (2020), Fandiño et al. (2022a) y Barbieri et al. (2022) respectivamente. En el Cuadro 3.8, se proporciona una breve descripción de cada uno de estos modelos.

---

<sup>9</sup><https://huggingface.co/>

<sup>10</sup><https://huggingface.co/docs/transformers/index>

| Modelo       | Descripcion                                                                                                                                           |
|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| robertuito   | Modelo de lenguaje preentrenado para contenido generado por usuarios en español, entrenado siguiendo las pautas de RoBERTa en 500 millones de tweets. |
| bertin       | Modelo basados en RoBERTa entrenados desde cero en la parte española de mC4 usando Flax                                                               |
| electricidad | Electricidad-base-discriminator es un modelo base tipo Electra entrenado en el corpus BETO                                                            |
| beto         | BETO es un modelo BERT entrenado con un gran corpus español.                                                                                          |
| roberta      | El roberta-base-bne se basa en el modelo base de RoBERTa y ha sido preentrenado utilizando un corpus en español de 570 GB de texto limpio             |
| twitter-xlm  | Este es un modelo basado en XLM-roBERTa multilingüe entrenado en 198 millones de tweets y ajustado para el análisis de sentimientos.                  |

Cuadro 3.8: Descripción de los modelos usados

### 3.2.2. Entrenamiento de los modelos

Debido a que el fine-tuning de estos modelos requiere una gran capacidad de cómputo, se decidió utilizar el servicio de Google Colab <sup>11</sup>, donde es posible desarrollar notebooks haciendo uso gratuito de GPUs. Allí, se empleó la librería Transformers de Hugging Face para acceder a los modelos.

Para entrenar los modelos, se realiza una transformación de los datos de entrada en vectores tokenizados que puedan ser interpretados por el modelo. Estos datos transformados se suministran al modelo junto con los hiperparámetros y la métrica de evaluación a utilizar.

Los principales hiperparámetros empleados incluyen la implementación de AdamW, un algoritmo de optimización propuesto por Loshchilov and Hutter (2017). Se utilizó así mismo un learning rate de  $5e-05$ . Además, se realizaron 3 épocas de entrenamiento en el conjunto de entrenamiento y se utilizó un batch size de 8.

<sup>11</sup><https://colab.research.google.com/>

### 3.2.3. Evaluación de los modelos

Los modelos fueron entrenados y evaluados utilizando el promedio Micro de la métrica F1 score. F1 score esta definida de la siguiente manera:

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Donde TP es el número de verdaderos positivos, FP es el número de falsos positivos y FN es el número de falsos negativos en todas las clases. El Micro F1 score se calcula de la misma, pero teniendo en cuenta todas las clases presentes. Esta métrica fue elegida debido a que el modelo permite la clasificación múltiple, y con ella se puede evaluar el desempeño de las distintas clases simultáneamente.

Para entrenar y evaluar los modelos, los datos etiquetados se dividieron en conjuntos de train y test. Se optó por no realizar una tercera partición debido a que no se llevó a cabo una optimización de los hiperparámetros, además del tamaño del conjunto de datos etiquetados. Una vez que los datos se dividieron, se procedió a entrenar cada uno de los seis modelos en diez ocasiones, y se registraron los resultados de su desempeño. Esto se hizo para obtener una medición más robusta, dada la naturaleza aleatoria de las redes neuronales. Los resultados de entrenamiento de cada modelo se registraron en la plataforma Wandb <sup>12</sup>, una herramienta de seguimiento de modelos. Los detalles pueden observarse en el informe del proyecto creado en dicho sitio <sup>13</sup>.

---

<sup>12</sup><https://wandb.ai/>

<sup>13</sup><https://api.wandb.ai/links/juanjose.if3/mtzaowe2>

# Capítulo 4

## Resultados

En este capítulo, se presentan los resultados de la clasificación de emociones en tweets utilizando los modelos de lenguaje preentrenados. Se seleccionó el modelo robertuito” para la tarea de predicción debido a su excelente desempeño, especialmente en las emociones miedo y tristeza. Se examina la distribución de emociones en diferentes orientaciones políticas, observando que el sentimiento de asco es predominante en los tweets asociados a la derecha, mientras que aquellos relacionados con la izquierda tienden a mostrar más alegría. Además, se analiza cómo las emociones fluctúan en momentos clave, como días de elecciones o eventos políticos relevantes durante el periodo estudiado.

## 4.1. Desempeño de los modelos entrenados

En el Cuadro 4.1, se presentan los resultados de la métrica F1, tanto para el promedio micro como para las diferentes emociones en cada modelo. Estos valores se muestran como la media y la desviación estándar de los experimentos realizados. Es importante mencionar que debido al limitado volumen de datos disponibles, el modelo "bertin" no logró converger en uno de los entrenamientos, por lo que este resultado se excluyó. Se destaca que el modelo robertuito" tuvo el mejor desempeño con un F1 micro de 76.6 %, convirtiéndose en la elección preferida. El desempeño de este modelo se debe a que tiene un ámbito de entrenamiento más cercano al objetivo de este trabajo, ya que se entrenó con tweets en español. A diferencia de otros modelos como "beto", que estaba entrenado en español pero no específicamente para Twitter, y "twitter-xlm", diseñado para varios idiomas pero no en específico para español. La tabla también presenta los resultados de la métrica F1 para cada emoción en particular, mostrando que todos los modelos tienen un buen desempeño en alegría y asco, a excepción de "bertin". Sin embargo, los puntajes son significativamente más bajos para miedo y tristeza en todos los modelos. No obstante, robertuito.ºbtuvo el mejor desempeño en ambas emociones.

| Model        | Promedio Micro | Alegria   | Miedo     | Tristeza  | Asco      |
|--------------|----------------|-----------|-----------|-----------|-----------|
| bertin       | 46.4±20.5      | 40.6±35.8 | 1.8±4.3   | 3.3±8.0   | 57.7±17.5 |
| beto         | 71.5±2.0       | 79.5±3.1  | 27.4±13.3 | 36.3±11.1 | 79.4±2.2  |
| electricidad | 67.7±2.5       | 75.6±3.5  | 0.0±0.0   | 8.5±10.1  | 74.1±2.8  |
| roberta      | 72.3±2.8       | 80.5±3.4  | 28.6±9.9  | 42.3±15.1 | 79.6±3.8  |
| robertuito   | 76.6±2.3       | 83.6±2.4  | 36.6±12.6 | 47.7±7.9  | 83.1±2.5  |
| twitter-xlm  | 72.8±2.7       | 80.8±3.5  | 27.1±9.7  | 43.5±10.6 | 80.8±2.0  |

Cuadro 4.1: Desempeño de los modelos entrenados medidos por F1 score

En la Figura 4.1, se muestra una matriz de confusión con los resultados de la clasificación en el conjunto de test. Esta matriz ilustra el número de coincidencias entre las etiquetas y las predicciones. Se observa que varios de los tweets etiquetados con miedo fueron predichos como asco.

Esta situación se explica en parte porque algunos de estos tweets efectivamente contenían la emoción de miedo junto con asco, pero el modelo solo logró identificar esta ultima, como se ejemplifica en el tweet número 1 en el Cuadro 4.2. Esto también ocurrió en el caso de la tristeza, como se aprecia en el tweet número 2. Para la alegría, se observa que también hubo casos en los que tweets fueron etiquetados con esta emoción pero en las predicciones les fue asignado asco. En muchos casos, esto se debió a que un mismo tweet expresaba dos emociones diferentes, una en cada frase, como se ilustra en el tweet número 3 del cuadro. En relación a los tweets que no tenían etiquetas y, por ende, se representan como neutrales en la Figura 4.1, se presentó un fenómeno en el cual los etiquetadores no percibían suficiente carga emocional como para asignar una etiqueta. Sin embargo, el modelo los clasificó con la etiqueta más

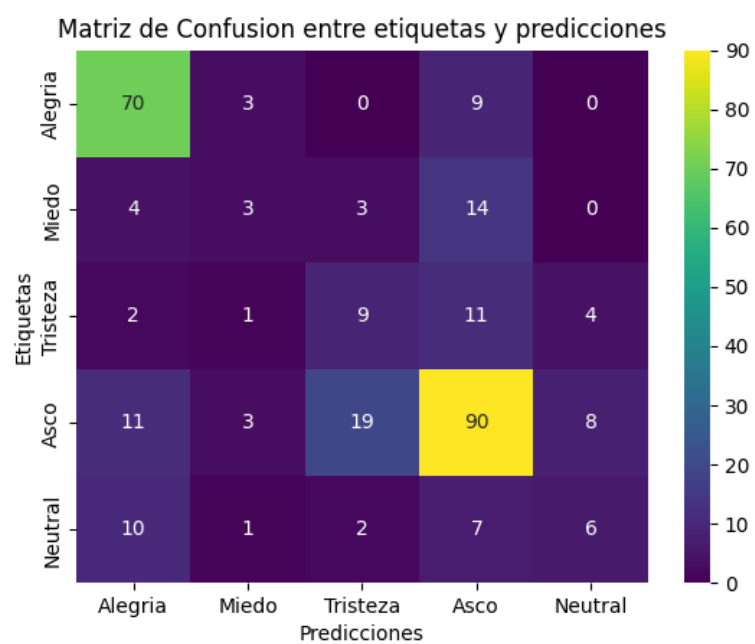


Figura 4.1: Matriz de confusión entre etiquetas asignadas y predicciones del algoritmo para el dataset etiquetado

adecuada según su contenido, como se ilustra en el ejemplo número 4.

| Numero de Ejemplo | Tweet                                                                                                                                                                                                                                                                              | Etiquetas      | Predicciones |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|--------------|
| 1                 | En conclusión #ElDebateDefinitivo deja claro que Petro es un peligro. Mentiroso, arrogante, tramposo, encubridor, manipulador, violento, pasivo agresivo, resentido. En resumen, una porquería de ser humano, qué digo humano... un demonio. Petro es el Pol Pot de Ciénaga de Oro | Miedo, Asco    | Asco         |
| 2                 | #ElDebateDefinitivo @sergio_fajardo fue a vender ilusiones y convencer no a debatir no propone sólo dice redundancias, me sorprende profe sera que de proyectos no sabe tanto como de licenciatura...                                                                              | Tristeza, Asco | Asco         |
| 3                 | Juan Roberto Vargas es una persona decente. Néstor Morales es una porquería de persona. #DebateFinal                                                                                                                                                                               | Alegria, Asco  | Asco         |
| 4                 | #FedericoEsColombia Me uno Federico Gutiérrez es Colombia.                                                                                                                                                                                                                         | Neutral        | Alegria      |

Cuadro 4.2: Ejemplos de tweets con divergencias entre etiquetas y clasificaciones

## 4.2. Distribución de emociones por orientación política

A partir del despliegue el modelo y la obtención de las etiquetas emocionales para todos los tweets del conjunto de datos, se obtuvieron los resultados observados en la Figura 4.2, donde se muestra el porcentaje de tweets que recibió cada una de las distintas etiquetas emocionales. En dicha figura, se puede observar que la emoción más preponderante fue el asco, abarcando más del 50 % de los tweets. Luego, le sigue la alegría con aproximadamente el 40 %. Por último, tanto la tristeza como el miedo tuvieron una presencia menor, rondando alrededor del 12 % y el 9 %, respectivamente.

Al examinar la distribución de cada emoción según la orientación política asigna-

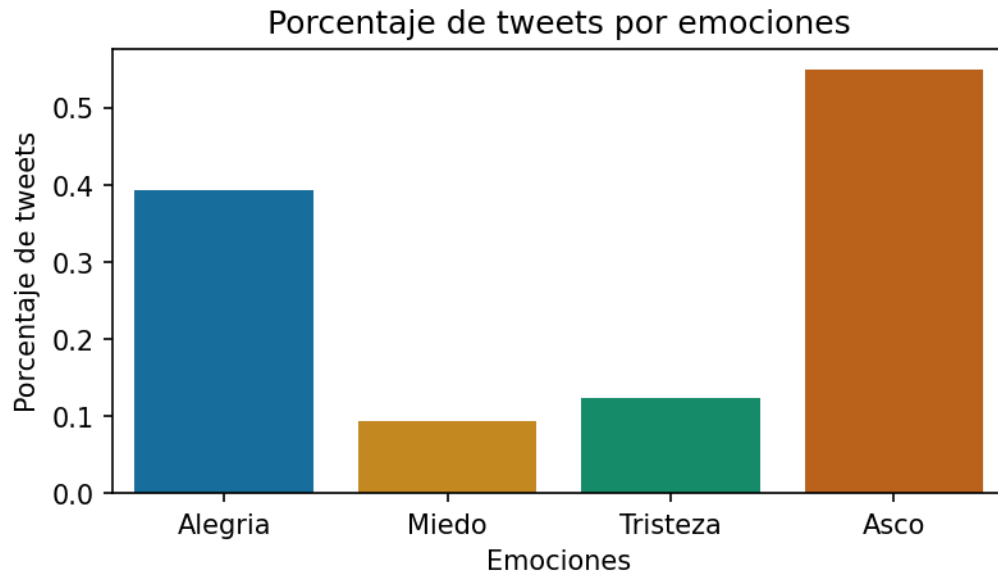


Figura 4.2: Porcentaje de tweets clasificado según emociones por el algoritmo

da al tweet a partir del hashtag empleado, se obtienen los resultados que se presentan en la Figura 4.3. En este gráfico, se destaca que el sector neutral fue el que mostró una mayor cantidad de tweets etiquetados con miedo, representando aproximadamente el 12 % de sus tweets. Por otro lado, tanto las orientaciones derecha e izquierda tuvieron una presencia menor, alrededor del 8 % y 6 %, respectivamente. En cuanto a la alegría, se revela que los tweets con orientación izquierda asignada, fueron los que presentaron una mayor frecuencia de esta emoción, alcanzando el 58 %. Luego, las orientaciones neutral y derecha tuvieron alrededor de un 30 % cada uno. Respecto al asco, se destaca la preeminencia de los tweets con la orientación de derecha asignada, representando un 69 % del total. La orientación neutral se ubica en segundo lugar con un 60 % de los tweets, mientras que los tweets con la orientación izquierda asignada tienen el menor porcentaje de los tres, un 40 %. Finalmente, en relación a la tristeza, la orientación neutral fue la que mostró una mayor presencia de esta emoción con

un 22 % de su total, cifra significativamente mayor que las orientaciones derecha e izquierda, que tuvieron un 6 % y 3 %, respectivamente.

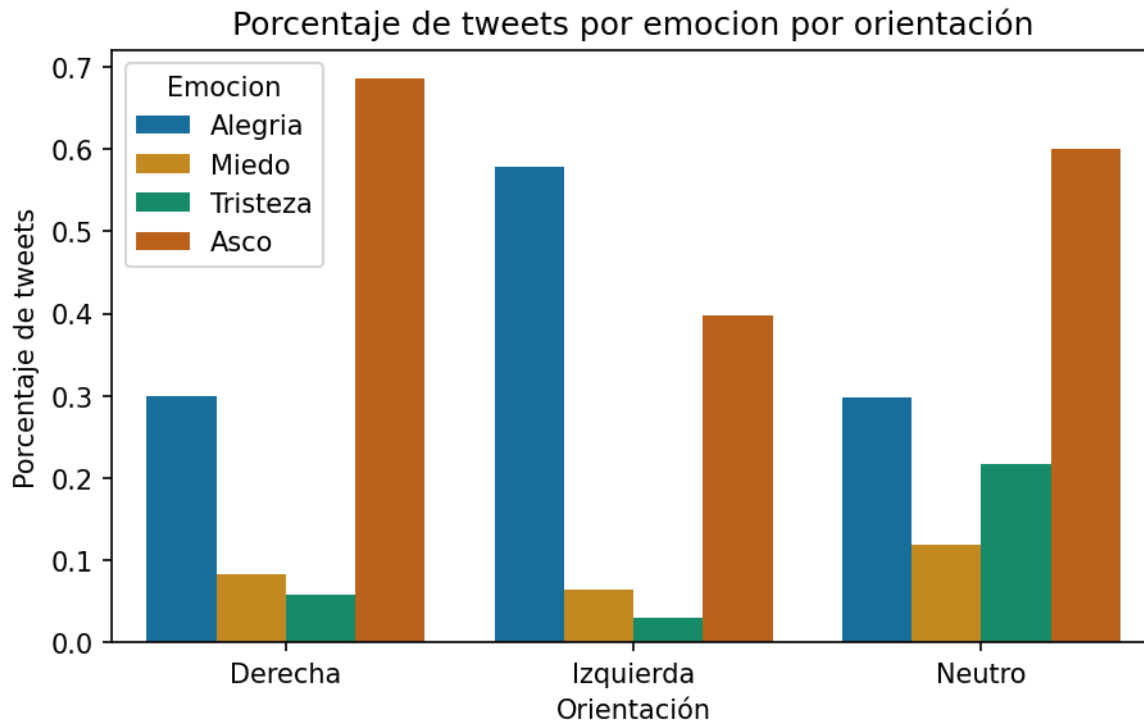


Figura 4.3: Porcentaje de tweets clasificados según emoción por el algoritmo, separado por orientación política del hashtag

### 4.3. Emociones a lo largo del tiempo

Para determinar el porcentaje de tweets que presentó una emoción en un día específico, se dividió el número de tweets con esa emoción en dicho día entre el total de tweets con esa misma emoción. Esto se ilustra en la Figura 4.4, donde se puede observar que el 29 de mayo fue un día particularmente activo en términos

emocionales. Durante ese día, alrededor del 27 % de los tweets estuvieron asociados con las emociones de alegría y miedo, un 20 % con tristeza y un 9 % con asco. Este día coincidió con la primera vuelta presidencial. De manera similar, entre los días 9 y 10 de junio hubo un repunte en las emociones de asco, tristeza y miedo, con porcentajes del 9 %, 6 % y 5 % respectivamente. Estos días estuvieron marcados por el evento conocido como Petro videos, que se refiere a la filtración de videos grabados durante una reunión del equipo del candidato Petro, en la que se discutían estrategias de campaña<sup>1</sup>. Posteriormente, se observa otro aumento en las emociones de asco, miedo y tristeza alrededor del 16 de junio, fecha en la que se habló del debate final al cual el candidato Rodolfo Hernández se negó a participar. Finalmente, el 19 de junio, que corresponde a la segunda vuelta, se observa un incremento en las emociones de alegría, tristeza y miedo.

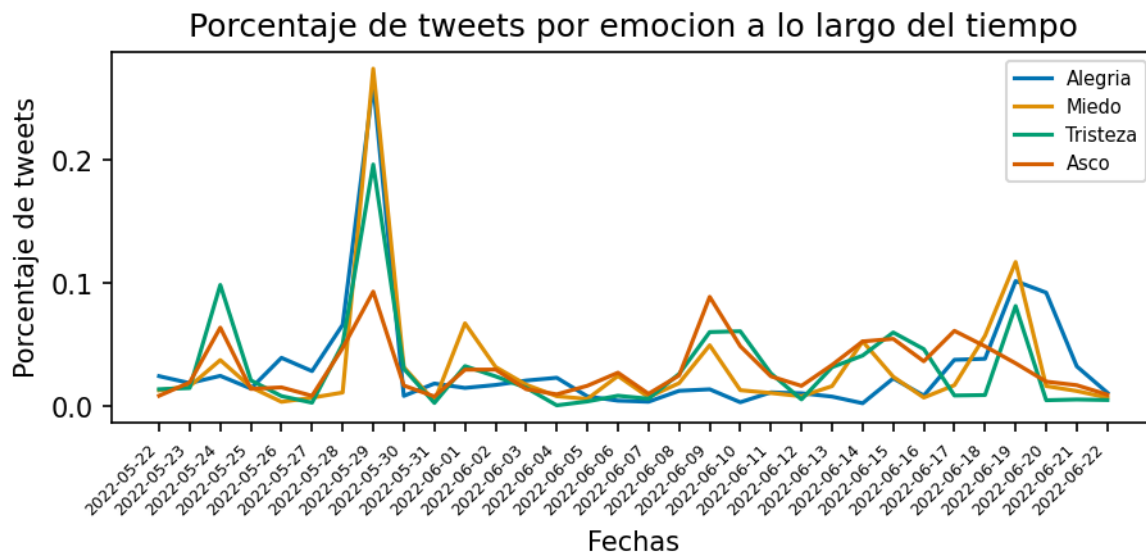


Figura 4.4: Porcentaje diario del total de tweets clasificados en cada emoción a lo largo del tiempo

<sup>1</sup><https://es.wikipedia.org/wiki/Petrovideos>

En relación a cada emoción y orientación, se obtuvo el porcentaje correspondiente que tuvieron los tweets asignados a cada orientación en un día particular. Esto se logró dividiendo el número de tweets de una orientación, con una emoción específica en dicho día entre el total de tweets de esa misma orientación con esa misma emoción. Esto se puede apreciar en la Figura 4.5 para la emoción de alegría, en la cual se observa que el 29 de mayo todas las orientaciones experimentaron un repunte en esta emoción. Posteriormente, cerca del 19 de junio, las tres orientaciones muestran un crecimiento en la emoción de alegría, que luego disminuye, primero en la derecha, seguida por la orientación neutral y finalmente en la izquierda.

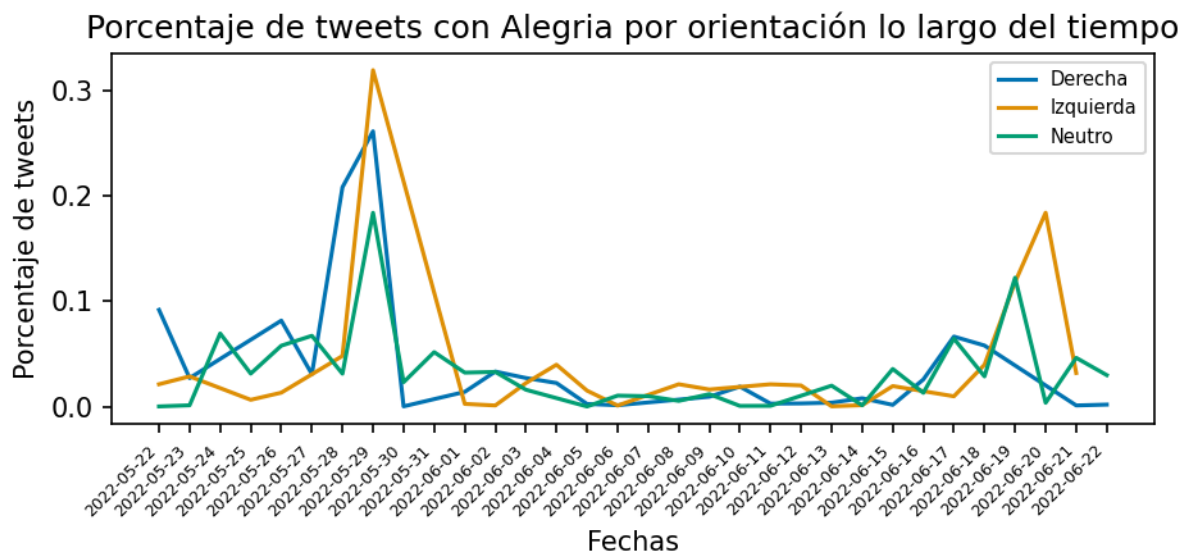


Figura 4.5: Porcentaje diario del total de tweets clasificados como Alegría, para cada orientación política

En el caso del miedo, se puede observar en la Figura 4.6 un notable pico en las tres orientaciones entre el 29 de mayo y el 1 de junio, como respuesta a la primera vuelta. Posteriormente, el 9 de junio hubo un repunte del miedo en la derecha, como consecuencia del fenómeno conocido como los "Petro videos". El 14 de junio también

se evidencia un aumento en el miedo en la derecha, debido a rumores de un posible estallido social. Finalmente, cerca del 19 de junio se observa un pico de miedo en las orientaciones neutro e izquierda.

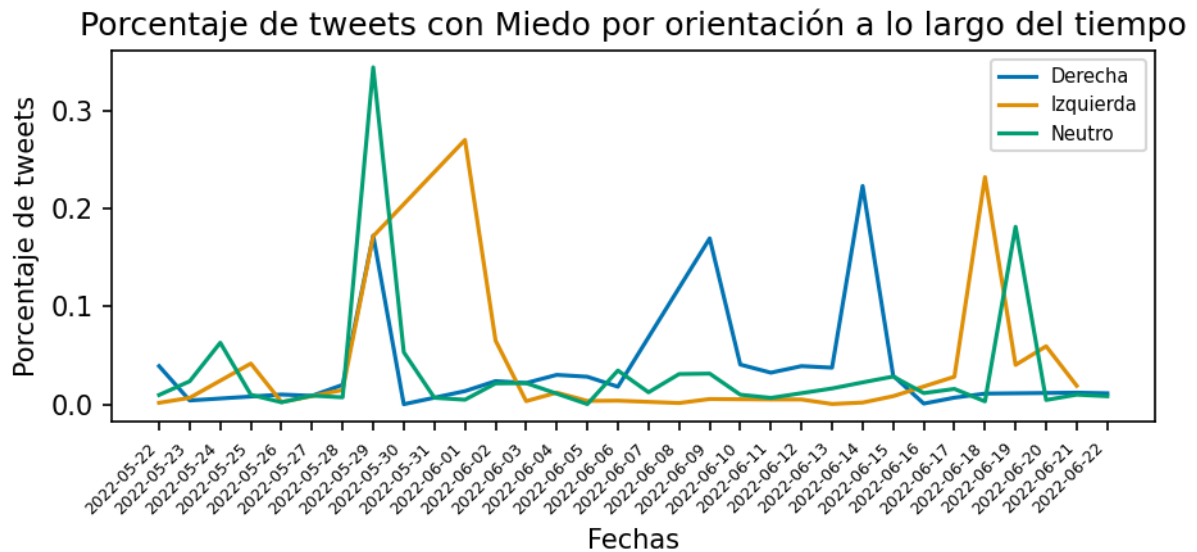


Figura 4.6: Porcentaje diario del total de tweets clasificados como Miedo, para cada orientación política

En cuanto a la tristeza, se aprecia en la Figura 4.7 que las tres orientaciones tuvieron un repunte inicial cerca del 24 de mayo, como resultado de un debate. Entre el 29 de mayo y el 1 de junio, la izquierda y la orientación neutral experimentaron otro incremento, alcanzando un 36 % y un 22 % respectivamente. El 9 de junio se observa un aumento en la tristeza en la derecha, relacionado con los "Petro videos", y el 10 de junio hubo un repunte en la orientación neutral, en el que se discutieron temas decepcionantes relacionados con las elecciones. El 14 de junio, la derecha experimentó nuevamente un aumento en la tristeza, también relacionado con la posibilidad de un estallido social. Finalmente, cerca del 19 de junio, todos los sectores mostraron un incremento en la tristeza.

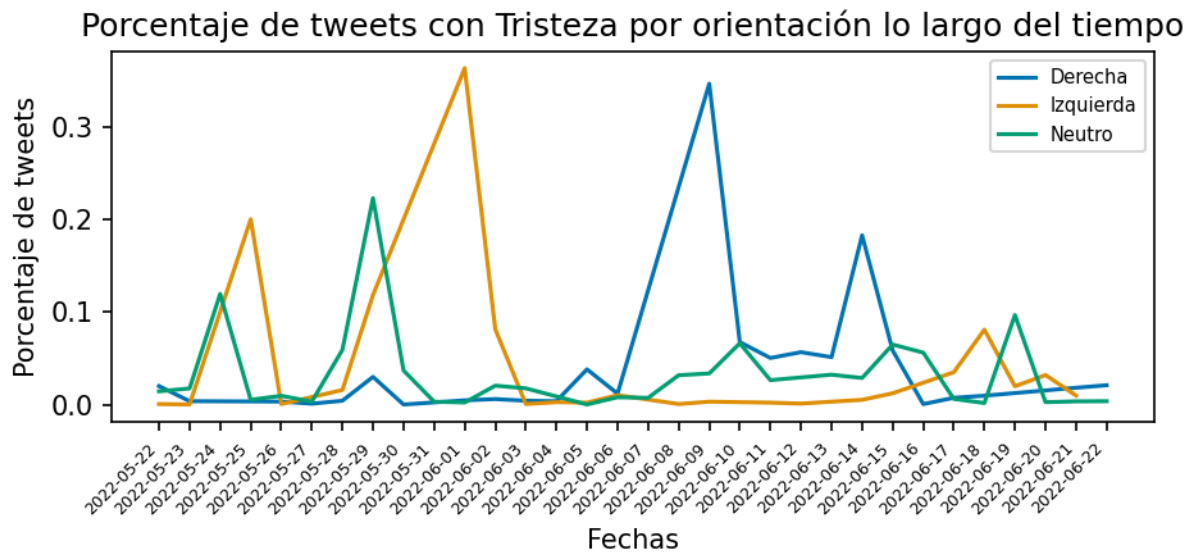


Figura 4.7: Porcentaje diario del total de tweets clasificados como Tristeza, para cada orientación política

En relación al asco, se puede observar en la Figura 4.8 que al inicio hay un pico en la orientación neutral el 24 de mayo, como reacción al debate. Luego, todas las orientaciones experimentaron un repunte el 29 de mayo. Es notable el gran pico que tuvo la derecha el 9 de junio, coincidiendo con los "Petro videos", con más del 27%. De manera similar, la izquierda alcanzó su punto máximo el 17 de junio, llegando a cerca del 20%. En esta fecha se habló de la negativa del candidato Rodolfo Hernández a participar en el debate final.

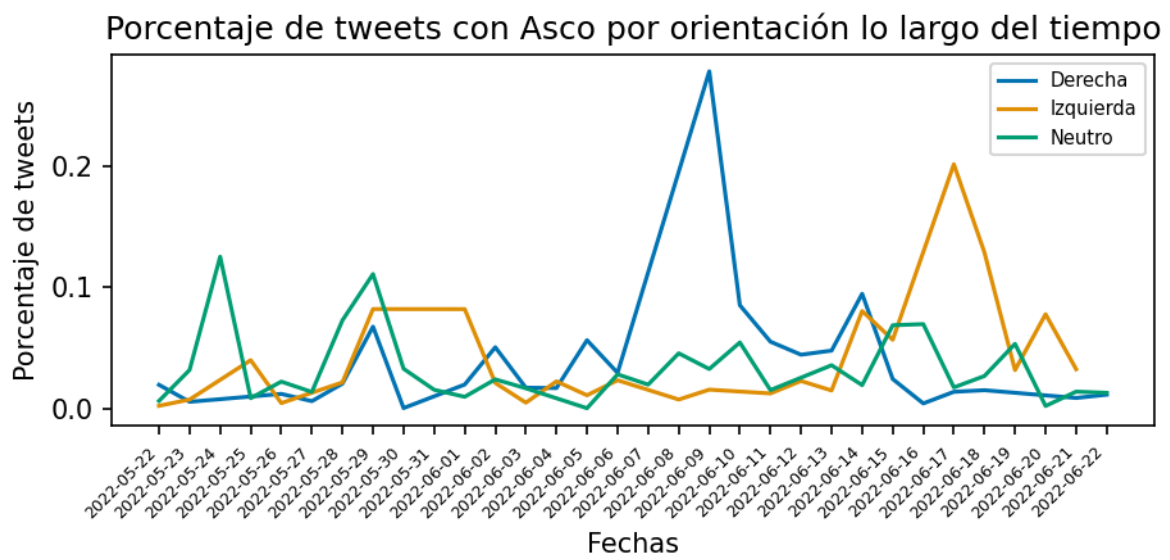


Figura 4.8: Porcentaje diario del total de tweets clasificados como Asco, para cada orientación política

## Capítulo 5

### Conclusiones

El presente trabajo logró, a través del fine tuning de un modelo de lenguaje pre-entrenado, en este caso robertuito-base-uncased”, identificar las emociones presentes en los tweets relacionados con las elecciones presidenciales en Colombia del 2022. Estas predicciones permitieron poder asociar estas emociones a distintas orientaciones políticas, a partir de la asignación de una orientación a un tweet en particular, a partir del hashtag usado y la tendencia observada en el mismo. Así mismo , se pudo observar la fluctuación por emoción y por sector de los tweets a lo largo del periodo estudiado.

Para poder seleccionar el modelo para esta tarea, se procedió realizar un fine tuning y evaluación de distintos modelos pre entrenados que fueron considerados relevantes debido a la cercanía de los mismos con el presente trabajo. La evaluación de estos modelos fue llevada a cabo a partir del calculo del promedio Micro del F1 score en las predicciones realizadas.

El proceso de fine tuning fue posible gracias a la creación de un recurso propio, un conjunto de datos de 1200 tweets etiquetados manualmente por el autor y los directores. Esta tarea de etiquetado se realizó utilizando una interfaz web y siguiendo un manual interno que permitió asignar una o varias de las 14 emociones disponibles a cada tweet. Para que una emoción fuera asignada al tweet, al menos dos etiquetadores debían coincidir en ella. Finalmente, cada tweet se clasificó con una de las cuatro etiquetas resultantes de la agrupación de las etiquetas originales, basada en su correlación. Es importante señalar la dificultad inherente en la tarea de etiquetado, ya que intenta lograr una clasificación objetiva de una actividad subjetiva, lo que requirió un proceso iterativo para desarrollar un manual y una interfaz de etiquetado que se acercara a este propósito. Además, aunque todos los etiquetadores son hispanohablantes nativos, solo el autor es originario de Colombia, lo que llevó a que ciertos usos del lenguaje o situaciones contextuales específicas fueran más claros para él que para los otros etiquetadores.

Las predicciones realizadas por el modelo coinciden con las respuestas emocionales esperadas en ciertos eventos importantes durante las elecciones. Los resultados del modelo muestran una presencia mucho mayor de alegría y asco que de miedo y tristeza, lo cual concuerda con lo observado durante el proceso de etiquetado, ya que estas dos emociones fueron menos frecuentes y a menudo acompañadas por la emoción del asco. Estas razones también explican el rendimiento relativamente más bajo en términos de capacidad predictiva para las emociones menos prevalentes, en comparación con la alegría y el asco. Es importante destacar que estos resultados se obtuvieron con una muestra de 1200 tweets etiquetados, por lo que aumentar el tamaño de la muestra podría haber mejorado el rendimiento del algoritmo, especialmente en las emociones con menor participación.

En términos generales, se observó una tendencia mayor hacia la alegría en los tweets a los que les fue asignada la orientación política de izquierda y hacia el asco por parte de la derecha. Esto se debe en parte a que el candidato victorioso fue de tendencia izquierdista, lo que generó estas respuestas emocionales en los respectivos sectores políticos, especialmente en fechas clave como el día de las elecciones.

Este trabajo pone a disposición el conjunto de datos etiquetados y el modelo utilizado para futuras investigaciones, como la identificación de emociones en español en contextos distintos al de este estudio. Asimismo, muestra cómo, para el contexto particular de las elecciones presidenciales en Colombia en 2022, los usuarios de Twitter que utilizaron los hashtags incluidos expresaron sus emociones.

# Referencias

- Acheampong, F. A., Nunoo-Mensah, H., and Chen, W. (2021). Transformer models for text-based emotion detection: a review of bert-based approaches. *Artificial Intelligence Review*, 54(8):5789–5829.
- Alm, C. O., Roth, D., and Sproat, R. (2005). Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pages 579–586.
- Aman, S. and Szpakowicz, S. (2007). Identifying expressions of emotion in text. In *International Conference on Text, Speech and Dialogue*, pages 196–205. Springer.
- Barbieri, F., Anke, L. E., and Camacho-Collados, J. (2022). Xlm-t: Multilingual language models in twitter for sentiment analysis and beyond. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 258–266.
- Bollen, J., Mao, H., and Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 450–453.

- Canete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., and Pérez, J. (2020). Spanish pre-trained bert model and evaluation data. *Pml4dc at iclr*, 2020:1–10.
- Cañete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., and Pérez, J. (2020). Spanish pre-trained bert model and evaluation data. In *PML4DC at ICLR 2020*.
- Cerón-Guzmán, J. A. and León-Guzmán, E. (2016). A sentiment analysis system of spanish tweets and its application in colombia 2014 presidential election. In *2016 IEEE international conferences on big data and cloud computing (BDCloud), social computing and networking (socialcom), sustainable computing and communications (sustaincom)(BDCloud-socialcom-sustaincom)*, pages 250–257. IEEE.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- Davidov, D., Tsur, O., and Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. In *Coling 2010: Posters*, pages 241–249.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Ekman, P. (1993). Facial expression and emotion. *American psychologist*, 48(4):384.
- Fandiño, A. G., Estapé, J. A., Pàmies, M., Palao, J. L., Ocampo, J. S., Carrino, C. P., Oller, C. A., Penagos, C. R., Agirre, A. G., and Villegas, M. (2022a). Maria: Spanish language models. *Procesamiento del Lenguaje Natural*, 68.

- Fandiño, A. G., Estapé, J. A., Pàmies, M., Palao, J. L., Ocampo, J. S., Carrino, C. P., Oller, C. A., Penagos, C. R., Agirre, A. G., and Villegas, M. (2022b). Maria: Spanish language models. *Procesamiento del Lenguaje Natural*, 68.
- Gil, G. B., Jesús, A. B. d., and Lopéz, J. M. M. (2013). Combining machine learning techniques and natural language processing to infer emotions using spanish twitter corpus. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 149–157. Springer.
- Go, A., Bhayani, R., and Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12):2009.
- Gonzalez, J. A., Hurtado, L.-F., and Pla, F. (2021). Twilbert: Pre-trained deep bidirectional transformers for spanish twitter. *Neurocomputing*, 426:58–69.
- Hatzivassiloglou, V. and McKeown, K. (1997). Predicting the semantic orientation of adjectives. In *35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics*, pages 174–181.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Jurafsky, D. (2000). *Speech & language processing*. Pearson Education India.
- Kouloumpis, E., Wilson, T., and Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg! In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 538–541.
- la Rosa, J. D., Ponferrada, E. G., Romero, M., Villegas, P., de Prado Salas, P. G., and Grandury, M. (2022). Bertin: Efficient pre-training of a spanish language model using perplexity sampling. *Procesamiento del Lenguaje Natural*, 68(0):13–23.
- Loshchilov, I. and Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

- Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- Mohammad, S. (2012). # emotional tweets. In \* *SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 246–255.
- Mohammad, S., Bravo-Marquez, F., Salameh, M., and Kiritchenko, S. (2018). Semeval-2018 task 1: Affect in tweets. In *Proceedings of the 12th international workshop on semantic evaluation*, pages 1–17.
- Mohammad, S. M., Zhu, X., Kiritchenko, S., and Martin, J. (2015). Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management*, 51(4):480–499.
- O’Connor, B., Balasubramanyan, R., Routledge, B. R., and Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth international AAAI conference on weblogs and social media*.
- Ortony, A., Clore, G. L., and Foss, M. A. (1987). hatzivassiloglou1997predicting. *Cognitive science*, 11(3):341–364.
- Pak, A. and Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*.
- Pang, B., Lee, L., et al. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2):1–135.
- Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.
- Pérez, J. M., Furman, D. A., Alonso Alemany, L., and Luque, F. M. (2022). Ro-

- BERTuito: a pre-trained language model for social media text in Spanish. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 7235–7243, Marseille, France. European Language Resources Association.
- Picard, R. W. (2000). *Affective computing*. MIT press.
- Plaza-del Arco, F. M., Jiménez Zafra, S. M., Montejo Ráez, A., Molina González, M. D., Ureña López, L. A., and Martín Valdivia, M. T. (2021). Overview of the emoeval task on emotion detection for spanish at iberlef 2021.
- Plaza-del Arco, F. M., Martín-Valdivia, M. T., Ureña-López, L. A., and Mitkov, R. (2020a). Improved emotion recognition in spanish social media through incorporation of lexical knowledge. *Future Generation Computer Systems*, 110:1000–1008.
- Plaza-del Arco, F. M., Strapparava, C., Lopez, L. A. U., and Martín-Valdivia, M. T. (2020b). Emoevent: A multilingual emotion corpus based on different events. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1492–1498.
- Roberts, K., Roach, M. A., Johnson, J., Guthrie, J., and Harabagiu, S. (2012). Empatweet: Annotating and detecting emotions on twitter. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 3806–3813.
- Romero, M. (2020). Spanish electra by manuel romero. <https://huggingface.co/mrm8488/electricidad-base-discriminator/>.
- Sidorov, G., Haro, S. N. G., and Vázquez, V. A. C. (2016). Construcción de un corpus marcado con emociones para el análisis de sentimientos en twitter en español. *Revista Escritos BUAP*, 1(1).
- Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez,

- N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Trevino, A., and Gordon, J. (2012). Empirical study of machine learning based approach for opinion mining in tweets. In *Mexican international conference on Artificial intelligence*, pages 1–14. Springer.
- Strapparava, C., Valitutti, A., et al. (2004). Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, page 40. Lisbon, Portugal.
- Tumasjan, A., Sprenger, T., Sandner, P., and Welpe, I. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 4, pages 178–185.
- Turney, P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*.
- VandenBos, G. R. (2007). *APA dictionary of psychology*. American Psychological Association.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Wang, W., Chen, L., Thirunarayan, K., and Sheth, A. P. (2012). Harnessing twitter”big data” for automatic emotion identification. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, pages 587–592. IEEE.
- Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2):165–210.
- Wiebe, J. M. (1994). Tracking point of view in narrative. *arXiv preprint cmp-lg/9407019*.
- Wikipedia contributors (2022). Recurrent neural network — Wikipedia, the free

encyclopedia. [https://en.wikipedia.org/w/index.php?title=Recurrent\\_neural\\_network&oldid=1109264340](https://en.wikipedia.org/w/index.php?title=Recurrent_neural_network&oldid=1109264340). [Online; accessed 26-September-2022].

Yu, H. and Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136.

# Apéndice A

## Anexos

Cuadro A.1: Hashtags

| date       | trending                | Sector    | Count |
|------------|-------------------------|-----------|-------|
| 2022-05-22 | NosUnimosONosJodemos    | Derecha   | 427   |
| 2022-05-22 | FedericoEsColombia      | Derecha   | 1597  |
| 2022-05-22 | LoPeorDeEstasElecciones | Neutro    | 357   |
| 2022-05-22 | MiVotoEsPetroYFrancia   | Izquierda | 2131  |
| 2022-05-23 | MiVotoEsPetroYFrancia   | Izquierda | 2131  |
| 2022-05-23 | UnTramposoEs            | Neutro    | 1255  |
| 2022-05-23 | FedericoEsColombia      | Derecha   | 1597  |
| 2022-05-23 | MeInquieta              | Neutro    | 696   |
| 2022-05-24 | ElDebateDefinitivo      | Neutro    | 7546  |
| 2022-05-24 | AColombiaLePreocupa     | Neutro    | 1088  |
| 2022-05-24 | LeApuestoA              | Neutro    | 326   |
| 2022-05-24 | UnTramposoEs            | Neutro    | 1255  |
| 2022-05-24 | GrandezaEs              | Neutro    | 498   |
| 2022-05-25 | VotoPor                 | Neutro    | 1093  |
| 2022-05-25 | YaEsSuficiente          | Izquierda | 816   |
| 2022-05-25 | ElDebateDefinitivo      | Neutro    | 7546  |

Continua en la siguiente pagina

Cuadro A.1: Hashtags

| date       | trending                 | Sector    | Count |
|------------|--------------------------|-----------|-------|
| 2022-05-25 | vanessapregúnteleafico   | Izquierda | 323   |
| 2022-05-25 | AColombiaLePreocupa      | Neutro    | 1088  |
| 2022-05-25 | ConvenceUnoMásPorPetro   | Izquierda | 174   |
| 2022-05-26 | FicoElGanador            | Derecha   | 2578  |
| 2022-05-26 | ElMejorCandidatoEs       | Neutro    | 1681  |
| 2022-05-26 | ElDebateDecisivo         | Neutro    | 1305  |
| 2022-05-26 | PetroEsElCambio          | Izquierda | 457   |
| 2022-05-26 | MagdalenaVotaPetro       | Izquierda | 150   |
| 2022-05-27 | PetroEs                  | Neutro    | 877   |
| 2022-05-27 | EsElMomentoDe            | Neutro    | 1334  |
| 2022-05-27 | FicoElGanador            | Derecha   | 2578  |
| 2022-05-27 | ElDomingoVoyA            | Neutro    | 298   |
| 2022-05-27 | ElMejorCandidatoEs       | Neutro    | 1681  |
| 2022-05-28 | LaVidaEnUnVoto           | Izquierda | 221   |
| 2022-05-28 | VoyConFico               | Derecha   | 78    |
| 2022-05-28 | YoVotoPetroPresidente    | Izquierda | 3376  |
| 2022-05-28 | DebateFinal              | Neutro    | 5217  |
| 2022-05-28 | VoyConFederico           | Derecha   | 4148  |
| 2022-05-28 | FicoElGanador            | Derecha   | 2578  |
| 2022-05-29 | EleccionesColombia       | Neutro    | 7088  |
| 2022-05-29 | CualquieraMenosPetro     | Derecha   | 957   |
| 2022-05-29 | Elecciones2022           | Neutro    | 5978  |
| 2022-05-29 | PetroPresidenteEnPrimera | Izquierda | 4032  |
| 2022-05-29 | PetroPresidente2022      | Izquierda | 1712  |
| 2022-05-29 | ElCambioEnPrimera        | Izquierda | 3077  |
| 2022-05-29 | SegundaVuelta            | Neutro    | 1752  |
| 2022-05-29 | MiVotoEsSecretosky       | Izquierda | 734   |
| 2022-05-29 | VotoPorFicoYLara         | Derecha   | 846   |
| 2022-05-29 | PetroNuncaSeráPresidente | Derecha   | 461   |
| 2022-05-29 | HoyMeIdentiFICO          | Derecha   | 92    |
| 2022-05-29 | YoVotoPetroPresidente    | Izquierda | 3376  |
| 2022-05-29 | ColombiaConPetro         | Izquierda | 3263  |
| 2022-05-29 | VoyConFederico           | Derecha   | 4148  |
| 2022-05-29 | ColombiaDecide           | Neutro    | 3906  |

Continúa en la siguiente página

Cuadro A.1: Hashtags

| date       | trending                 | Sector    | Count |
|------------|--------------------------|-----------|-------|
| 2022-05-29 | HoyGana                  | Neutro    | 687   |
| 2022-05-29 | PetroYFranciaSonElCambio | Izquierda | 637   |
| 2022-05-29 | PetroNUNCAseraPresidente | Derecha   | 461   |
| 2022-05-30 | SegundaVuelta            | Neutro    | 1752  |
| 2022-05-30 | LaPrimeraVueltaFue       | Neutro    | 693   |
| 2022-05-30 | Elecciones2022           | Neutro    | 5978  |
| 2022-05-30 | VotoPorFicoYLara         | Derecha   | 846   |
| 2022-05-30 | HoyGana                  | Neutro    | 687   |
| 2022-05-31 | SegundaVuelta            | Neutro    | 1752  |
| 2022-05-31 | MiCandidatoEs            | Neutro    | 2123  |
| 2022-06-01 | FraudeElectoral          | Izquierda | 2814  |
| 2022-06-01 | VelezPregunteleAPetro    | Derecha   | 192   |
| 2022-06-01 | pactosabroso             | Izquierda | 79    |
| 2022-06-01 | ElVerdaderoCambioEs      | Neutro    | 787   |
| 2022-06-01 | RodolfoPresidente2022    | Derecha   | 2720  |
| 2022-06-01 | CreoEnColombiaPor        | Neutro    | 658   |
| 2022-06-01 | Antipetrismo             | Derecha   | 176   |
| 2022-06-02 | MeSumoA                  | Neutro    | 307   |
| 2022-06-02 | ColombiaNecesita         | Neutro    | 2381  |
| 2022-06-02 | RodolfoPresidente2022    | Derecha   | 2720  |
| 2022-06-02 | UnaPreguntaParaPetroFm   | Derecha   | 850   |
| 2022-06-02 | FraudeElectoral          | Izquierda | 2814  |
| 2022-06-02 | ElDebateFeminista        | Izquierda | 56    |
| 2022-06-03 | PetroElCambioSeguro      | Izquierda | 3145  |
| 2022-06-03 | RodolfoPresidente2022    | Derecha   | 2720  |
| 2022-06-03 | ColombiaQuiere           | Neutro    | 600   |
| 2022-06-03 | ColombiaNecesita         | Neutro    | 2381  |
| 2022-06-03 | UnDesastre               | Neutro    | 357   |
| 2022-06-04 | PetroElCambioSeguro      | Izquierda | 3145  |
| 2022-06-04 | YoPactoConPetro          | Izquierda | 761   |
| 2022-06-04 | RodolfoPresidente2022    | Derecha   | 2720  |
| 2022-06-05 | LaCorrupcionPetrista     | Derecha   | 1492  |
| 2022-06-05 | VotarEnBlancoEs          | Neutro    | 1466  |
| 2022-06-05 | GraciasPero              | Neutro    | 719   |

Continúa en la siguiente página

Cuadro A.1: Hashtags

| date       | trending                | Sector    | Count |
|------------|-------------------------|-----------|-------|
| 2022-06-05 | PetroElCambioSeguro     | Izquierda | 3145  |
| 2022-06-06 | VotarEnBlancoEs         | Neutro    | 1466  |
| 2022-06-06 | MeLimpioELCuloConEsaLey | Izquierda | 149   |
| 2022-06-06 | LaCorrupciónPetrista    | Derecha   | 775   |
| 2022-06-06 | GraciasPero             | Neutro    | 719   |
| 2022-06-06 | AliasElIngeniero        | Izquierda | 505   |
| 2022-06-07 | UnaVerdadEs             | Neutro    | 1504  |
| 2022-06-07 | VotarEnBlancoEs         | Neutro    | 1466  |
| 2022-06-08 | UnaBurlaEs              | Neutro    | 1467  |
| 2022-06-08 | YoContratoAPetro        | Izquierda | 1511  |
| 2022-06-08 | NoQuiero                | Neutro    | 921   |
| 2022-06-08 | UnaVerdadEs             | Neutro    | 1504  |
| 2022-06-08 | ConPetroGanasTú         | Izquierda | 259   |
| 2022-06-09 | PetrismoSucio           | Derecha   | 5447  |
| 2022-06-09 | RumboALaVictoria        | Izquierda | 271   |
| 2022-06-09 | PetroVideos             | Derecha   | 2011  |
| 2022-06-09 | MeDaVergüenza           | Neutro    | 195   |
| 2022-06-09 | PetroNoMateaRodolfo     | Derecha   | 37    |
| 2022-06-09 | NoTodoVale              | Neutro    | 257   |
| 2022-06-09 | PetroMentiroso          | Derecha   | 524   |
| 2022-06-09 | LosPetroVideosSon       | Derecha   | 3034  |
| 2022-06-09 | TodoVale                | Neutro    | 796   |
| 2022-06-09 | YoContratoAPetro        | Izquierda | 1511  |
| 2022-06-09 | ColombiaEspera          | Neutro    | 918   |
| 2022-06-10 | HablaMalDeColombia      | Neutro    | 2231  |
| 2022-06-10 | PetrismoSucio           | Derecha   | 5447  |
| 2022-06-10 | PetroVideos             | Derecha   | 2011  |
| 2022-06-10 | CampañaSuciaEs          | Neutro    | 1149  |
| 2022-06-10 | MeVoyConElIngeniero     | Derecha   | 278   |
| 2022-06-11 | SoyPetroSoyPaz          | Izquierda | 2125  |
| 2022-06-11 | PetroVideos             | Derecha   | 2011  |
| 2022-06-11 | HablaMalDeColombia      | Neutro    | 2231  |
| 2022-06-11 | PetrismoSucio           | Derecha   | 5447  |
| 2022-06-11 | QuéCansancio            | Neutro    | 459   |

Continúa en la siguiente página

Cuadro A.1: Hashtags

| date       | trending                     | Sector    | Count |
|------------|------------------------------|-----------|-------|
| 2022-06-12 | LaMaldadPetrista             | Derecha   | 2152  |
| 2022-06-12 | SoyPetroSoyPaz               | Izquierda | 2125  |
| 2022-06-12 | ArizabaletaLameCulos         | Izquierda | 328   |
| 2022-06-13 | ArizabaletaSicarioMoral      | Izquierda | 386   |
| 2022-06-13 | PetroRespondaleAlosNiños     | Derecha   | 332   |
| 2022-06-13 | VanessaEnSegundaVotoPor      | Neutro    | 199   |
| 2022-06-13 | UnGolpeBajoEs                | Neutro    | 676   |
| 2022-06-13 | EsUnaBajeza                  | Neutro    | 501   |
| 2022-06-13 | LaGranEncuesta               | Neutro    | 70    |
| 2022-06-13 | LaMaldadPetrista             | Derecha   | 2152  |
| 2022-06-13 | NosFalta                     | Neutro    | 1155  |
| 2022-06-14 | PetrismoAcosaALaPrensa       | Derecha   | 461   |
| 2022-06-14 | YoBloqueoASemana             | Izquierda | 3508  |
| 2022-06-14 | LaCuotaDeBenedetti           | Derecha   | 416   |
| 2022-06-14 | EstallidoSocialEs            | Derecha   | 1964  |
| 2022-06-14 | NoMeDejaDormir               | Neutro    | 1178  |
| 2022-06-15 | DebateObligadoEs             | Neutro    | 3491  |
| 2022-06-15 | YoBloqueoASemana             | Izquierda | 3508  |
| 2022-06-15 | LoCorrectoEs                 | Neutro    | 1495  |
| 2022-06-15 | CaracolPetrista              | Derecha   | 409   |
| 2022-06-15 | presidentepetro              | Izquierda | 859   |
| 2022-06-15 | EstallidoSocialEs            | Derecha   | 1964  |
| 2022-06-16 | IrrespetuosoEs               | Neutro    | 2125  |
| 2022-06-16 | debate                       | Neutro    | 791   |
| 2022-06-16 | NosHaceFalta                 | Neutro    | 985   |
| 2022-06-16 | RodolfoElPresidentedelPueblo | Derecha   | 408   |
| 2022-06-16 | MeDecepciona                 | Neutro    | 452   |
| 2022-06-17 | RHGanador                    | Derecha   | 2277  |
| 2022-06-17 | RodolfoCobarde               | Izquierda | 4843  |
| 2022-06-17 | YoVotoPor                    | Neutro    | 1748  |
| 2022-06-17 | EsteDomingo                  | Neutro    | 1033  |
| 2022-06-17 | PetroPresidenteEnsegunda     | Izquierda | 1359  |
| 2022-06-18 | ExguerrilleroOIngeniero      | Neutro    | 2760  |
| 2022-06-18 | RegistraduriaResponda        | Izquierda | 3107  |

Continúa en la siguiente página

Cuadro A.1: Hashtags

| date       | trending                     | Sector    | Count |
|------------|------------------------------|-----------|-------|
| 2022-06-18 | MiVotoEsAntiFraude           | Izquierda | 1592  |
| 2022-06-18 | RHGanador                    | Derecha   | 2277  |
| 2022-06-18 | PetroPresidenteEnsegunda     | Izquierda | 1359  |
| 2022-06-19 | PetroEsPresidente            | Izquierda | 15521 |
| 2022-06-19 | Elecciones2022               | Neutro    | 5978  |
| 2022-06-19 | EleccionesPresidenciales2022 | Neutro    | 327   |
| 2022-06-19 | ExguerrilleroOIngeniero      | Neutro    | 2760  |
| 2022-06-19 | EleccionesColombia2022       | Neutro    | 4484  |
| 2022-06-19 | RegistraduriaResponda        | Izquierda | 3107  |
| 2022-06-19 | VivirSabroso                 | Izquierda | 790   |
| 2022-06-20 | PetroEsPresidente            | Izquierda | 15521 |
| 2022-06-20 | ExguerrilleroOIngeniero      | Neutro    | 2760  |
| 2022-06-20 | Elecciones2022               | Neutro    | 5978  |
| 2022-06-21 | PetroEsPresidente            | Izquierda | 15521 |
| 2022-06-21 | NuestraPrioridadEs           | Neutro    | 1253  |
| 2022-06-21 | PetroBastaYa                 | Derecha   | 564   |
| 2022-06-21 | DePetroEspero                | Neutro    | 429   |
| 2022-06-21 | PetroNecesita                | Neutro    | 377   |
| 2022-06-22 | EsMuyImportante              | Neutro    | 1255  |
| 2022-06-22 | PetroBastaYa                 | Derecha   | 564   |
| 2022-06-22 | LosImpuestosDePetro          | Neutro    | 359   |