Presentado por Yilmer Palacios, Betina Cortés, Lida Jimena Rivera, Nelson Fabián López

Problem Set 4: Predicting Tweets

I. Introducción

En los últimos años, se ha generado un especial interés por parte de la academia para realizar análisis técnicos sobre las redes sociales que conduzcan a generar evidencia, principalmente en época de elecciones, para la toma de decisiones de los partidos políticos, candidatos y ciudadanos. Caso puntual objeto de análisis es la red social Twitter, la cual se ha convertido en una de las más empleadas en el mundo al ser un espacio que permite a las personas comunicar y estar en contacto a través de mensajes rápidos y frecuentes (Eskibel, 2022), y especialmente por las siguientes características (Gomila, 2020):

1) Favorece la viralidad de la información publicada en comparación con otras redes sociales, dado su diseño, interfaz y disponibilidad de recursos (hashtags, retuits, etiquetas o trendic topics)
2) Es la red social más utilizada por el sistema político y mediático, es decir, a pesar de no ser la plataforma más empleada, es la que tiene el público más influyente
3) Se ha convertido en una herramienta fundamental para el diseño y la ejecución de la estrategia comunicativa en campaña electoral de los partidos políticos
4) Su estructura facilita a los investigadores un acceso más directo a los datos, simplificando su extracción, procesamiento y análisis

En los últimos años, Twitter se ha convertido en un espacio fundamental para la difusión de ideas, información y sentimientos por parte de los usuarios. Aunque no es la plataforma más utilizada a nivel mundial, sí es la más influyente en el discurso público. Por esta razón, muchos políticos de todo el mundo la utilizan para comunicarse con su audiencia y dar a conocer sus iniciativas de política pública y de gobierno. Un ejemplo destacado de esta situación son las elecciones presidenciales de Estados Unidos, en las cuales la campaña electoral de Barack Obama en 2008 marcó un antes y un después. En esa ocasión, Twitter fue una de las herramientas más novedosas utilizadas por Obama, y para el día de la elección ya contaba con 100.000 seguidores. Cuatro años después, en 2012, Obama llegó al día de la elección con 20 millones de seguidores en Twitter, lo que demuestra la importancia que esta red social adquirió para la comunicación política en un corto período de tiempo (Eskibel, 2022).

Con esto en mente, el presente documento tiene como objetivo presentar un modelo predictivo que contribuye a determinar a quién pertenece cada tuit basado en el contenido de este por medio del análisis de sentimiento, considerando que el lenguaje refleja los valores e ideales de las personas que publican. La principal motivación de este estudio surge con la expresión "A rose by any other name would smell as sweet" (en español, "Una rosa con otro nombre olería igual de dulce") de la obra de William Shakespeare "Romeo y Julieta".

El conjunto de datos de entrenamiento contiene tuits de las cuentas de tres destacados políticos colombianos: Claudia López, Gustavo Petro y Álvaro Uribe. Se consideran cinco (5) modelos predictivos, sin embargo, los resultados no son los esperados al obtener un bajo rendimiento fuera de muestra. Se evalúan Redes Neuronales, Random Forest, XGBoost, Naive Bayes y Regresión Logística Multinomial.

Nota: La base de datos usada, al igual que el script de R y el presente documento están disponibles en el repositorio de GitHub en el siguiente enlace: https://github.com/Yilap/Repositorio_Taller4.git

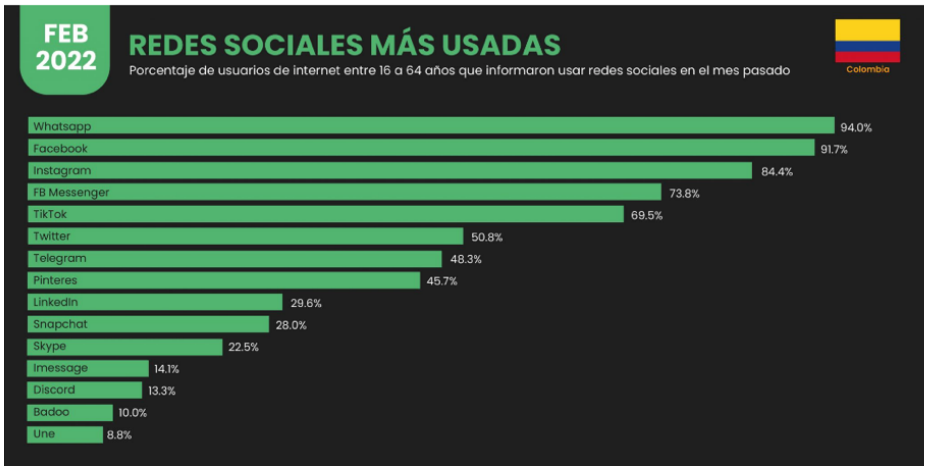
Contexto

Twitter se destaca por sus características únicas en comparación con otras redes sociales. Entre ellas, se encuentran (Eskibel, 2022):

1) Impacto: líderes de opinión, políticos, gobernantes, candidatos, periodistas, medios de comunicación, referentes sociales y culturales, formadores de opinión, entre otros
2) Noticia: una vía más directa para aparecer en las noticias que las ruedas de prensa tradicionales
3) Contacto: una forma rápida de contactar con personas relevantes para el político
4) Brevedad: un tweet son solo 140 caracteres, lo que permite una síntesis efectiva del pensamiento
5) Velocidad: una herramienta ideal para difundir o seguir en tiempo real las novedades de un evento que está desarrollándose
6) Interacción: permite intercambiar ideas, dialogar, discutir, defender posiciones y responder
7) Síntesis: se publican frases breves que sirvan para un título periodístico y que sean fácilmente recordadas
8) Receptividad: el público de Twitter recibe con mucha mayor naturalidad los mensajes vinculados a la política
9) Movilidad: Twitter es perfecto para usar desde los smartphones

En Colombia, el uso de las redes sociales ha experimentado un crecimiento significativo en los últimos años. Según Rosgaby (2022), el 81% de la población colombiana es usuaria activa de las redes sociales, lo que equivale a 41,8 millones de personas. El 52% de los usuarios son mujeres y el 48% son hombres, y el 37% de las mujeres y el 35% de los hombres se encuentran entre los 18 y 44 años. En promedio, los colombianos dedican alrededor de 3 horas y 46 minutos al día a conectarse a redes sociales, utilizando alrededor de 8 redes diferentes. Para el 2020, el 95% de los usuarios colombianos usaba Facebook, mientras que el 77% utilizaba Instagram. En el caso de Twitter, esta red social cuenta con una audiencia de 4,3 millones de usuarios, lo que representa el 8,4% de la población del país y el 12% de los usuarios de Internet en Colombia.

Como se muestra en la **Gráfica 1,** en la actualidad WhatsApp es la plataforma más utilizada por los colombianos, sin embargo, Twitter sigue siendo una herramienta importante para políticos, líderes de opinión y medios de comunicación debido a sus características únicas y su capacidad para difundir noticias y opiniones de forma rápida y directa. En este caso, es fundamental analizar el comportamiento de tres de los políticos más relevantes hoy en día: Claudia López, Gustavo Petro y Álvaro Uribe.



Gráfica 1 - Uso de redes sociales en Colombia
Fuente: Branch (2022)

II. Datos

a. Descripción de las fuentes de datos

Para el desarrollo de este Problem Set se utilizará un conjunto de datos de prueba que contiene 500 tweets sin etiquetar, el cual es extraído de a través de acceso programático a los datos de Twitter mediante las API (interfaces de programación de aplicaciones). De esta manera, Twitter permite acceder a partes del servicio mediante las API para permitir la creación de software que se integre con Twitter.

Los datos de Twitter son únicos y se extraen a partir de datos de la mayoría de las redes sociales. API ofrece acceso amplio a estos datos que los usuarios han decidido compartir de manera pública, en este caso los políticos colombianos objeto de estudio. Estos datos están conformados por dos bases de datos: test (1.500 observaciones) y la base train (9.349 observaciones).

b. Análisis descriptivo de los datos (estadísticas descriptivas)

Para el análisis de los Tweets se utilizó el paquete "tidytext", el cual incluye la función "unnest_tokens" para separar los Tweets en tokens (en este caso, palabras) para cada uno de los documentos. Con esto se construyó una matriz utilizando el criterio TFIDF, que permite determinar cuál es la palabra más representativa de cada uno de los personajes estudiados.

Para lograr esto, se unieron los datos de entrenamiento y prueba, con el fin de tener un inventario completo de las palabras que podrían aparecer en el análisis. Además, se aplicó la función anti_join para remover las stopwords, las cuales no aportan significado y solo ocupan espacio en la matriz. Luego, se filtraron las palabras con una longitud menor a tres caracteres, ya que tampoco aportan significado y pueden ser ruido en los análisis posteriores. Una vez limpiado el corpus de palabras, se utilizó la función wordStem para lematizar las palabras en español, es decir, reducirlas a su forma base o raíz. Esto se hizo para evitar redundancias y contar las palabras de manera más precisa. Adicionalmente, se aplicó la función filter_ng para crear nuevas columnas, como el número de comas, el número de punto y comas, el número de dos puntos, el número de espacios en blanco, el número de caracteres que no son letras ni números, el número de palabras que empiezan con mayúscula y el número de letras mayúsculas que aparecen en el tweet. Estas variables pueden ser importantes para reconocer la forma de redactar de cada político. Se calcularon los puntajes y se añadieron como columnas.

Además, se construyó la matriz término-documento (DTM) utilizando la función cast_dtm, la cual asigna un valor numérico que representa la frecuencia de aparición de cada palabra en cada uno de los documentos (tweets). Se utilizó el criterio TF-IDF para ponderar la importancia de cada término en cada documento. Por último, se eliminaron los términos que aparecían en menos del 0.03% de los documentos, para reducir la dimensionalidad de la matriz y mejorar la eficiencia de los análisis posteriores. Se obtuvieron dos matrices: una para los datos de entrenamiento (9349 observaciones y 858 variables) y otra para los datos de prueba. Estas matrices se utilizaron para construir modelos de clasificación de sentimiento basados en los tweets de los personajes políticos objeto de estudio.

Según los datos de la base de datos de entrenamiento, se observa que Claudia López es de quien más publicaciones se tienen, con un total de 3.470 tweets, seguida de Álvaro Uribe con 3.002 y Gustavo Petro con 2.877. En cuanto al número de caracteres, se evidencia que, en promedio, López quien hace tweets más largos, con un total de 242 caracteres, mientas que Petro usa alrededor de 194 y Uribe 160, teniendo en cuenta que la red social permite un máximo de 280 caracteres por publicación. Además, se destaca que López es quien

más utiliza hashtags en sus publicaciones, con un porcentaje del 46%, mientras que Uribe y Petro utilizan hashtags en un 12% y 8% de sus tweets, respectivamente. Los hashtags son importantes porque permiten etiquetar contenido con una palabra clave o tema específico, lo que facilita la búsqueda y agrupación de publicaciones relacionadas. Los resultados se resumen en la Tabla 1.

Tabla 1 - Estadísticas descriptivas variables numéricas

Descripción de los tweets Base de datos de entrenamiento			
Autor	**No.**	**Núm. Caracteres**	**% que usa hashtags**
López	3.470	242,62	46%
Petro	2.877	193,58	8%
Uribe	3.002	160,05	12%

Fuente: Elaboración propia

Para profundizar en el análisis, se crean nubes de palabras (wordclouds) basadas en el conjunto de datos de entrenamiento (train dataset) para cada uno de los tres políticos considerados en el estudio: Claudia López, Gustavo Petro y Álvaro Uribe. Las nubes de palabras permiten visualizar los términos más utilizados por cada uno de los políticos en su cuenta de Twitter. Esto puede ser útil para identificar patrones de lenguaje y estilo de comunicación que puedan ayudar a predecir quién escribió cada tuit, además de identificar temas y problemas comunes que son importantes para cada político. El análisis de nubes de palabras es una herramienta clave para identificar patrones y temas en grandes conjuntos de datos de texto.

En el caso de **Claudia López**, las 10 palabras más relevantes son: Bogotá, hoy, ciudad, gracias, vacunación, jóvenes, cuidado, seguridad, vida y mujeres.

Estas palabras están relacionadas con su cargo actual como alcaldesa de la capital del país. La palabra "Bogotá" es la más frecuente, lo cual es de esperarse, ya que se trata de la ciudad que gobierna. La frecuencia de la palabra "hoy" sugiere que la alcaldesa está comunicando acciones y decisiones en tiempo real. La presencia de palabras como "ciudad", "seguridad" y "vida" indica que López está intentando comunicar que está enfocando su gestión en mejorar la calidad de vida de los ciudadanos y en garantizar su seguridad. La aparición de la palabra "vacunación" en la tabla sugiere que la alcaldesa implementó medidas significativas para combatir la pandemia del COVID-19 y proteger a la población. La palabra "jóvenes" indica que la alcaldesa está

Gráfica 2 - Nube de palabras: Claudia López
Fuente: R Studio

prestando atención a la juventud y está trabajando en programas y políticas para su bienestar, al igual que "mujeres" sugiere que la alcaldesa está prestando atención a temas de equidad de género y derechos de las mujeres en su gestión. De esta forma, la frecuencia de estas palabras indica que la alcaldesa está comunicando a la población su trabajo en mejorar la calidad de vida de los ciudadanos y en promover la equidad y el bienestar social.

Por otra parte, las palabras más relevantes en los tweets del actual presidente de Colombia, **Gustavo Petro**, son: Colombia, Gobierno, hoy, país, Bogotá, debe, salud, pacto, ser y paz. Si bien se esperaban palabras como Colombia, Gobierno o país, resalta la presencia de "Bogotá" en la lista, aunque es coherente con el hecho de que Petro ha sido alcalde de la capital colombiana y que su carrera política ha estado ligada a esa ciudad. "Salud" es otra palabra importante en el wordcloud, considerando que ha sido una de las principales reformas que ha impulsado dentro de su proyecto político y que en los últimos años hemos estado inmersos en el contexto del COVID-19. Igualmente, la frecuencia de la palabra "pacto" tiene sentido por su pertenencia al partido del Pacto Histórico, así como la palabra "paz", pues su proyecto de la paz total ha sido una de las banderas de su agenda política. A grandes rasgos, se puede notar que su propuesta política se enfoca en la transformación social y económica del país.

Gráfica 3 - Nube de palabras: Gustavo Petro
Fuente: R Studio

Finalmente, en el caso del ex-presidente **Álvaro Uribe**, las palabras con mayor protagonismo son: usd, onza, Colombia, familia, solidaridad, país, Medellín, tonelada, social, hoy, violencia, democracia y centro. Las palabras USD, onza y tonelada aparecen con gran frecuencia porque es común que el mandatario comparta en su cuenta de Twitter información sobre actualidad macroeconómica y comercio internacional con indicadores con la tasa de cambio, el precio del petróleo, del café, entre otros. Por otra parte, destacan los conceptos de familia y solidaridad son propios del discurso conservador que promueve valores tradicionales, al igual que la palabra violencia, pues Uribe ha sido históricamente uno de los principales persecutores de grupos al margen de la ley como las FARC.

Gráfica 4 - Nube de palabras: Álvaro Uribe
Fuente: R Studio

Adicionalmente, aparecen palabras como Medellín, teniendo en cuenta que la mayoría de sus simpatizantes son del departamento de Antioquia, su ciudad natal y donde tiene más fuerza su ideología política. Para terminar, en cuanto a las palabras relacionadas con la política, es importante notar que democracia y centro son términos asociados al partido Centro Democrático, del cual Uribe es líder. En general, el análisis de las

palabras más frecuentes en los tweets del ex-presidente parece indicar una mayor preocupación por temas económicos y financieros, y menos atención a los problemas políticos y sociales del país. Esto es coherente con su ideología conservadora y su enfoque en mantener la estabilidad del sistema político y económico del país.

Ahora bien, en cuanto a la mediana de la longitud de caracteres para cada uno de los políticos analizados, se observa que la mediana para Uribe es de 157, para Petro 220 y para López de 264 (ver Gráfica 5). Se muestra también en la Gráfica 6 la distribución de los datos centrados en la cola de la derecha para los tweets de Claudia López y Gustavo Petro, y para el caso de Uribe los datos tienen una distribución sobre ambas colas. En las Gráficas 9 a la 13 (biagramas y triagramas) se muestran las palabras no por separado sino palabras en parejas para cada uno de los autores, algunos ejemplos significativos de estas palabras son: covid 19, Gobierno nacional, Colombia Humana, Pacto Histórico, bolsa ny, Centro Democrático, entre otros.

Finalmente, se realiza un análisis de sentimientos para cada uno de los autores, en donde se identifica los sentimientos negativos y positivos para cada uno de ellos. En la Gráfica 7 se evidencia el puntaje promedio de sentimientos por autor. Claudia López es la candidata que, en promedio, tiene más sentimientos positivos, seguida por Álvaro Uribe y, por último, Gustavo Petro. Entre las palabras positivas de López se encuentran: fiesta, atractivo, agradable y felicidad; las de Petro, mejorar y fiesta; y las de Uribe, conocimiento, bueno, brillante y acuerdo. Por el contrario, las palabras negativas para cada uno de ellos, respectivamente son: delincuente, choque, asesinato; retraso, perder, mal, basura, angustia y; esclavitud, delincuente, culpable, cruel.

En conclusión, se observa que los tweets de Claudia López están altamente relacionados con Bogotá y temas de "ciudad" como la "seguridad" y con enfoque de género. En el caso de Gustavo Petro, se evidencian palabras como "paz" y "salud", demostrando que se enfoca en temas sociales y de justicia. Para Álvaro Uribe, se presentan palabras como "violencia" y "democracia", lo que sugiere que su enfoque político está en temas relacionados con la seguridad y la democracia. Con estos hallazgos en mente, la siguiente sección apunta a la búsqueda del mejor modelo de predicción del autor de los tweets.

III. Modelos y resultados

En el ejercicio de predicción del autor de cada tweet, se entrenaron varios modelos para elegir el mejor desempeño. Se probaron cuatro opciones: XGBoost, Random Forest, Naive Bayes y Logit.

Se inició con Random Forest, un método de aprendizaje por conjuntos que combina múltiples árboles de decisión para mejorar la precisión y estabilidad de las predicciones. Para ajustar el modelo, se usó la implementación "ranger" de Random Forest y se creó una rejilla de ajuste con los hiperparámetros mtry, splitrule y min.node.size, que incluyó varios valores candidatos para cada hiperparámetro. Además, se configuró la función trainControl para utilizar la validación cruzada triple y las probabilidades de clase. Los valores finales utilizados para el modelo fueron mtry = 10, splitrule = extratrees y min.node.size = 1. El resultado mostró un accuracy de 0,807 dentro de la muestra, pero el envío a Kaggle para validar la capacidad de predicción del modelo tuvo un resultado de 0,323.

Luego, se utilizó XGBoost, un algoritmo de aumento de gradiente que construye un conjunto de árboles de decisión débiles. Se ajustaron los hiperparámetros, incluyendo el número de árboles a crecer, la profundidad máxima de los árboles, la tasa de aprendizaje y los parámetros de regularización. Los valores finales utilizados para el modelo fueron nrounds = 150, max_depth = 3, eta = 0,4, gamma = 0, colsample_bytree = 0,8, min_child_weight = 1 y subsample = 0,75. La precisión dentro de muestra fue de 0,814, pero el resultado en Kaggle fue de 0,316.

El modelo de Naive Bayes también se probó, alcanzando una precisión de 0,654 y un coeficiente kappa de 0,484 utilizando la validación cruzada sin preprocesamiento. Sin embargo, en Kaggle, el resultado de la capacidad de predicción del modelo fue de solo 0,327. Luego, se utilizó la regresión logística multinomial con glmnet, definiendo una cuadrícula de hiperparámetros sobre la que buscar, , con valores de alfa que van de 0 a 1 en incrementos de 0,1 y valores de lambda que van de 0 a 1 en incrementos logarítmicos. Los valores finales utilizados para el modelo fueron alfa = 0,3 y lambda = 0,01. La precisión dentro de la muestra fue de 0,805, pero el resultado en Kaggle fue de 0,31.

Finalmente, se usó el modelo de redes neuronales, este se ejecutó en Google Colab, puede ser revisado en el siguiente link:
https://colab.research.google.com/drive/1aK9h3COyZqSQB_TWR13fKDR_hNDjpwKP?usp=sharing

Para este modelo, se usaron las mismas bases de datos de training y test de los modelos anteriores, importando desde R un RData. La estructura de la red implementada es:

- Capa oculta con función RELU: Seleccionada porqué arrojo mejor resultado que la sigmoidea.
    - Se usó un dropout de 0,5 buscado reducir el Overfitting
- Capa Softmax, dado que la variable a predecir es una categórica (nombres de los políticos)

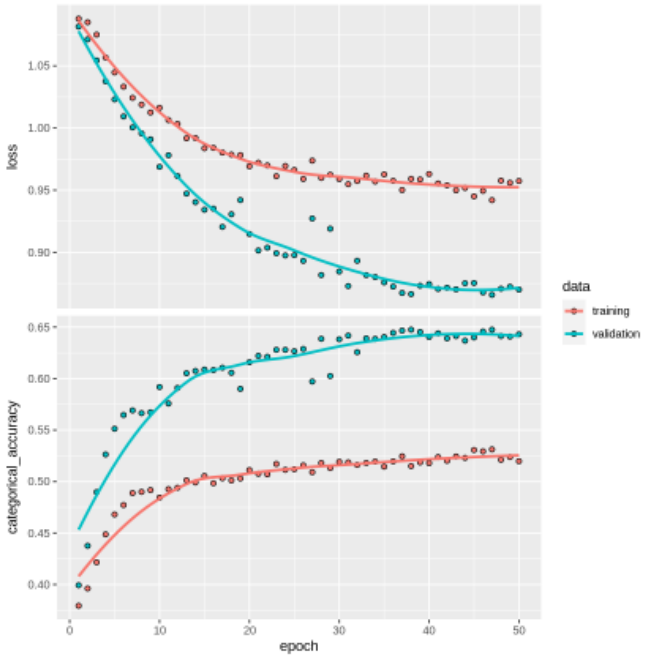
Respecto a la cantidad de neuronas usadas para la capa oculta, es importante señalar que inicialmente se usó la formula del Nh que se describe a continuación:

$$Nh = \frac{Ns}{\alpha(Ni + No)}$$

Usando un valor de alfa de 2, el Nh dio cercano a 4, por lo que este fue nuestro punto de partida para definir hiperparámetros. Finalmente, a prueba y error, los hiperparámetros definidos fueron los siguientes.

Parámetro	Valor	Observaciones
Cantidad de Neuronas	4	Buenos resultados iniciales, tomado a partir de la fórmula del Nh
Dropout rate	0.5	Intentamos usar mayores valores, pero empeoraron el modelo
Epoch	50	Tomado a prueba y error, aunque demorado, los mayores valores de Epoch arrojaron mejores resultados, elegido cuidando el Overfitting
Batch Size	2^6	Usado dado su buen performance, como era de esperarse, tuvo mayor costo computacional

Luego de varios ensayos de los hiperparámetros, los resultados obtenidos fueron:



Loss	0.86
Accuracy	0.65

Como se puede apreciar, la precisión no es la mejor, siendo un modelo más bien deficiente para la predicción de la data, cabe aclarar, que usamos otros hiperparametros los cuales aumentaron el accuracy, este efecto era especialmente causado por el aumento de neuronas en la capa oculta, pero como bien sabemos, esto trae consigo over fitting por lo que decidimos quedarnos con el modelo descrito anteriormente. Finalmente exportamos la data de Colab y regresamos al software R para hacer ajustes finales.

IV. Conclusiones y recomendaciones

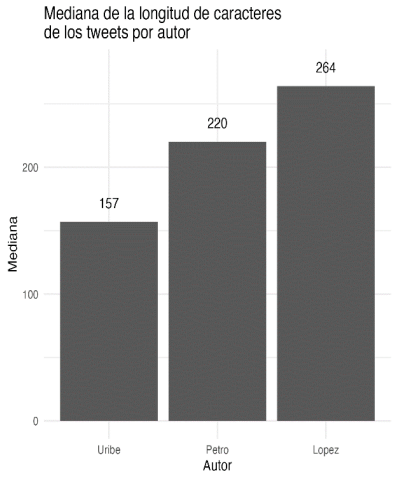
En el ejercicio de predecir el autor de cada tweet, se utilizaron diferentes algoritmos de aprendizaje automático como Redes Neuronales, Random Forest, XGBoost, Naive Bayes y Regresión Logística Multinomial. Sin embargo, a pesar de haber obtenido altas precisión dentro de la muestra, los modelos no lograron una buena capacidad de predicción fuera de la muestra, obteniendo una precisión significativamente baja en Kaggle.

Una posible explicación de por qué los modelos tienen un mal rendimiento fuera de la muestra podría ser el sobreajuste. Es posible que los modelos hayan aprendido patrones específicos de los datos de entrenamiento que no se presentan en los datos de prueba en Kaggle. Para evitar el sobreajuste, una recomendación sería aumentar el tamaño del conjunto de datos de entrenamiento y prueba, o utilizar técnicas de validación cruzada para evaluar el rendimiento del modelo en diferentes subconjuntos de datos. También se puede utilizar técnicas de regularización para reducir la complejidad del modelo y evitar el sobreajuste (si bien lo hicimos, en este caso no fuimos exitosos). Además, sería beneficioso realizar un análisis detallado de los datos y evaluar si los modelos están capturando adecuadamente la complejidad del problema. Es necesario validar el manejo que se le dio a la base de datos, si los datos son ruidosos o hay variables irrelevantes, podría ser necesario re procesarlos antes de ajustar los modelos.

En conclusión, es importante recordar que un modelo con una alta precisión dentro de la muestra no siempre se traducirá en un buen rendimiento fuera de la muestra. Se deben tomar medidas para evitar el sobreajuste y realizar un análisis detallado de los datos antes de ajustar los modelos.

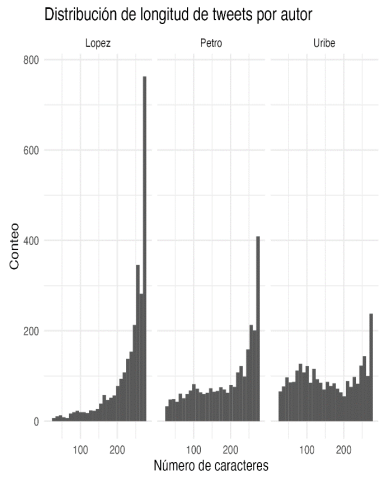
V. Anexos

Gráfica 5 – Mediana de longitud de caracteres



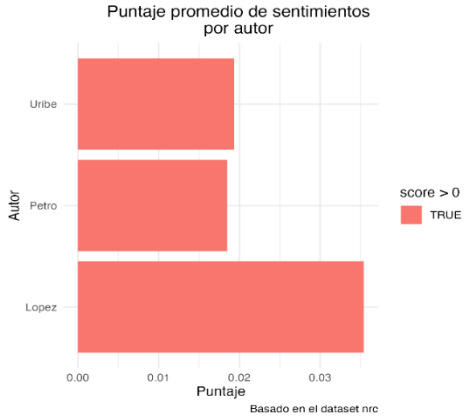
Fuente: R Studio

Gráfica 6 – Distribución de longitud



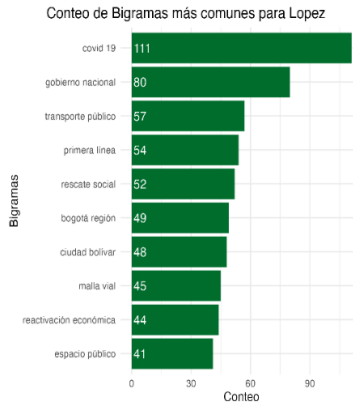
Fuente: R Studio

Gráfica 7 – Puntaje promedio de sentimientos



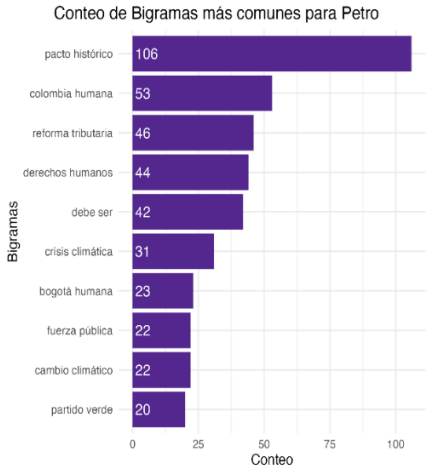
Fuente: R Studio

Gráfica 8 – Bigrama más común para Lopéz



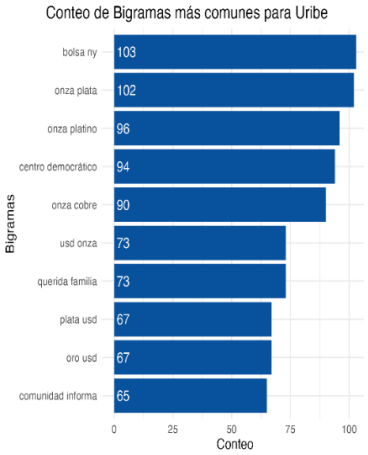
Fuente: R Studio

Gráfica 9 – Bígama más común para Petro



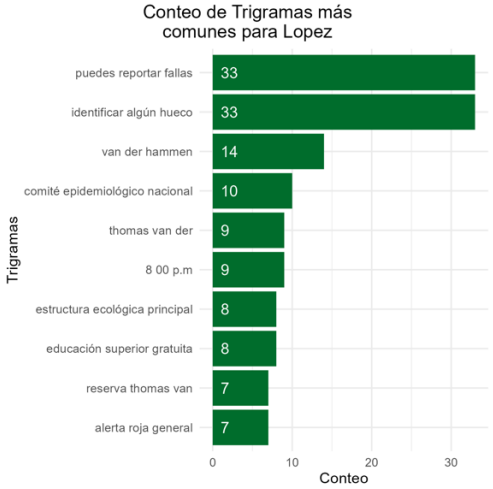
Fuente: R Studio

Gráfica 10 – Bigrama más común para Uribe



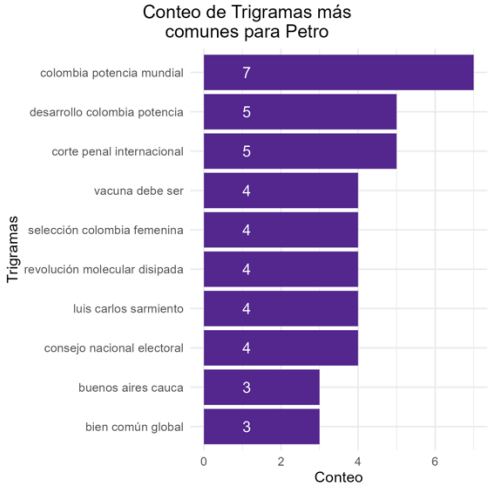
Fuente: R Studi

Gráfica 11 – Trigrama más común para López



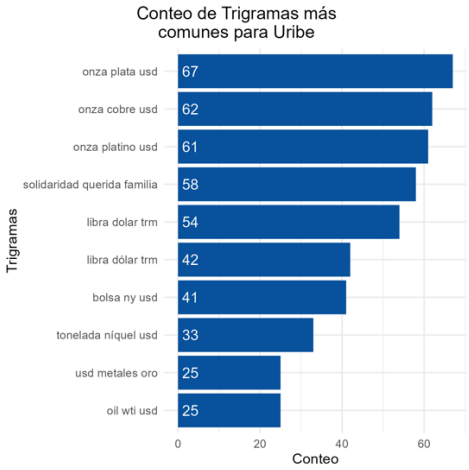
Fuente: R Studio

Gráfica 12 – Trigrama más común para Petro



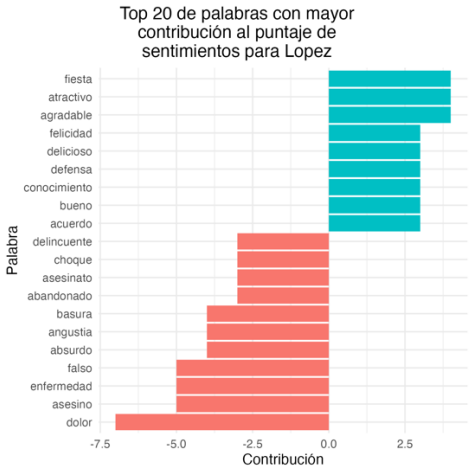
Fuente: R Studio

Gráfica 13 – Trigrama más común para Petro



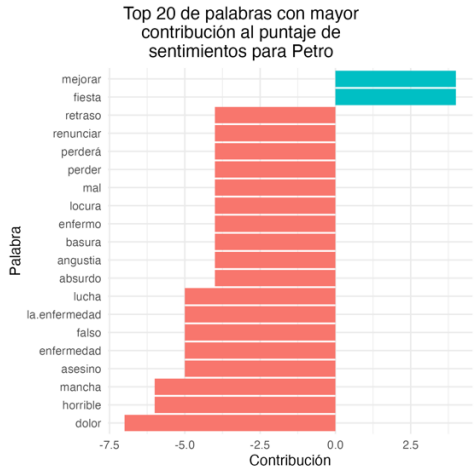
Fuente: R Studio

Gráfica 14 – Contribuciones sentimientos para López



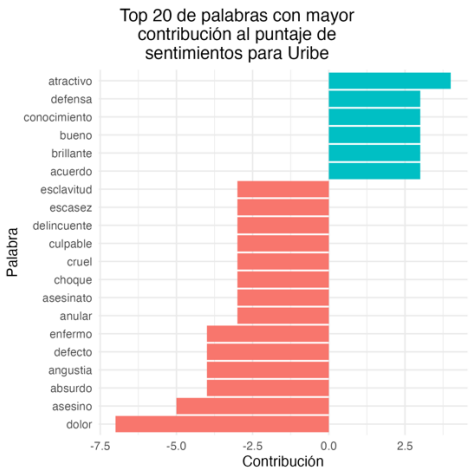
Fuente: R Studio

Gráfica 15 – Contribuciones sentimientos para Petro



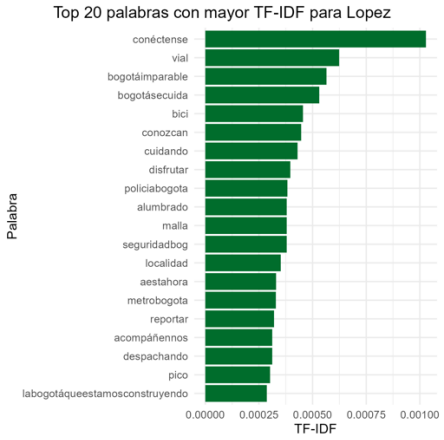
Fuente: R Studio

Gráfica 16 – Contribuciones sentimientos para Uribe



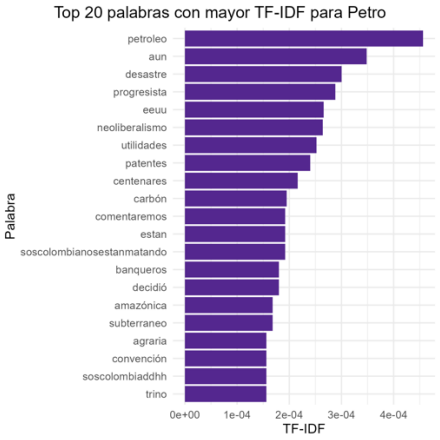
Fuente: R Studio

Gráfica 17 – TFIDF para López



Fuente: R Studio

Gráfica 18 – TFIDF para Petro

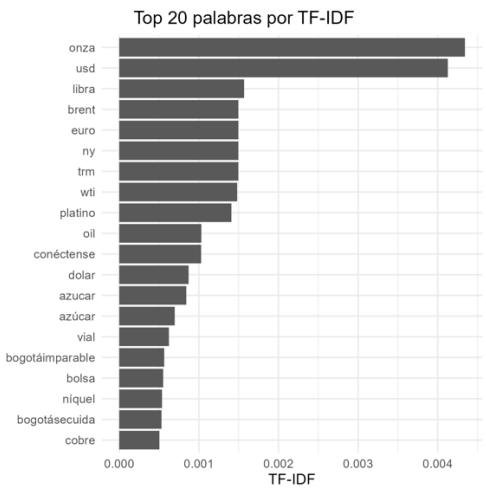


Fuente: R Studio

Gráfica 19 – TFIDF para Uribe

Fuente: R Studio

Gráfica 20 – Top 20 TFIDF



Fuente: R Studio

VI. Bibliografía

Eskibel, D., (2022). 10 razones por las que los políticos prefieren Twitter. Recuperado de: https://danieleskibel.com/twitter10/.

Gomila, G., (2020). ¿Para qué usan Twitter los partidos en campaña?. Recuperado de: https://agendapublica.elpais.com/noticia/13745/qu-usan-twitter-partidos-campana.

Rosgaby, M., (2022). Estadísticas de la situación digital de Colombia en el 2021-2022. Branch – Marketing digital. Recuperado de: https://branch.com.co/marketing-digital/estadisticas-de-la-situacion-digital-de-colombia-en-el-2021-2022/

Twitter., (s.f.). Información sobre las API de Twitter. Recuperado de: https://help.twitter.com/es/rules-and-policies/twitter-api

Twitter., (s.f.). Preguntas frecuentes para usuarios nuevos. Centro de ayuda. Recuperado de: https://help.twitter.com/es/resources/new-user-faq