



Pontificia Universidad  
Católica del Ecuador

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR**

**FACULTAD DE INGENIERÍA**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL GRADO DE  
MAGISTER EN SISTEMAS DE INFORMACIÓN MENCIÓN CIENCIA DE DATOS**

**TEMA:**

**Análisis de sentimientos utilizando la red social “X” Twitter para medir el nivel de aceptación del nuevo presidente del Ecuador, Daniel Noboa (noviembre 2023 - abril 2024)**

**AUTOR:**

**Roly Steeven Cedeño Menéndez**

**DIRECTOR:**

**Eduardo José Montero Bermúdez Msc.**

**QUITO – ECUADOR**

**2024**

## ÍNDICE DE CONTENIDO

<b>1</b>	<b>CAPÍTULO I: INTRODUCCIÓN</b>	<b>10</b>
1.1	Planteamiento del Problema	10
1.2	Justificación	10
1.3	Antecedentes	11
1.4	Objetivos	11
1.4.1	Objetivo General	11
1.4.2	Objetivos Específicos	12
<b>2</b>	<b>CAPÍTULO II: MARCO TEÓRICO</b>	<b>13</b>
2.1	Análisis de Sentimientos	13
2.1.1	Concepto y Definición	13
2.1.2	Métodos y Herramientas para el Análisis de Sentimientos	13
2.1.3	Procesamiento del Lenguaje Natural	15
2.1.3.1	Minería De Datos	15
2.1.3.2	Procesamiento del lenguaje natural (PLN)	16
2.1.3.3	Aplicaciones del Procesamiento del Lenguaje Natural	17
2.2	Redes Sociales	18
2.2.1	Definición de Redes Sociales	18
2.2.2	Aplicaciones del Análisis de Sentimientos en Redes Sociales	18
2.2.3	<i>Twitter</i>	19
2.2.3.1	Características de <i>Twitter</i>	20
2.2.3.2	Uso de <i>Twitter</i> en Investigaciones Sociales	20
2.2.3.3	Limitaciones y Consideraciones Éticas	21
2.3	Opinión Pública	22
2.3.1	Concepto y Definición	22
2.3.2	Nivel de Aceptación	22

2.4	Ecuador.....	23
2.5	Daniel Noboa.....	23
<b>3</b>	<b>CAPÍTULO III: METODOLOGÍA .....</b>	<b>25</b>
3.1	Desarrollo.....	26
3.1.1	Obtención de los datos .....	26
3.1.2	Limpieza de los datos.....	27
3.1.3	Análisis de sentimientos .....	28
<b>4</b>	<b>CAPÍTULO IV: RESULTADOS .....</b>	<b>29</b>
4.1	Análisis General.....	29
4.1.1	Distribución de Cantidad por Categoría .....	29
4.1.2	Nube de Palabras .....	29
4.1.2.1	Unigramas .....	30
4.1.2.2	Bigramas.....	32
4.1.2.3	Trigramas.....	35
4.1.3	Distribución de Sentimientos por Mes .....	38
4.1.4	Distribución de Sentimientos por Ubicación .....	41
4.2	Entrenamiento de Modelos .....	42
4.2.1	<i>Naive Bayes</i> .....	43
4.2.2	<i>Support Vector Machine (SVM)</i> .....	43
4.2.3	<i>Logistic Regression (Regresión Logística)</i> .....	43
4.2.4	<i>Decision Tree (Árbol de Decisión)</i> .....	44
4.2.5	<i>Random Forest</i> .....	44
4.2.6	<i>XGBoost</i> .....	44
4.3	Evaluación de Modelos .....	44
4.3.1	Clases Balanceadas .....	45
4.3.2	Clases Moderadamente Desbalanceadas.....	48

4.3.3	Clases Totalmente Desbalanceadas.....	51
<b>5</b>	<b>CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES.....</b>	<b>55</b>
5.1	Conclusiones.....	55
5.2	Recomendaciones .....	55
	<b>REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>57</b>
	<b>ANEXOS.....</b>	<b>60</b>

## ÍNDICE DE FIGURAS

Figura 1 Esquema general del enfoque metodológico propuesto .....	26
Figura 2 Nube de Palabras de Unigramas Negativos.....	30
Figura 3 Nube de Palabras de Unigramas Positivos .....	31
Figura 4 Nube de Palabras de Unigramas Neutrales .....	32
Figura 5 Nube de Palabras de Bigramas Negativos.....	33
Figura 6 Nube de Palabras de Bigramas Positivos .....	34
Figura 7 Nube de Palabras de Bigramas Neutrales.....	35
Figura 8 Nube de Palabras de Trigramas Negativos.....	36
Figura 9 Nube de Palabras de Trigramas Positivos .....	37
Figura 10 Nube de Palabras de Trigramas Neutrales.....	38
Figura 11 Grafico de barras Distribución de sentimientos por mes.....	39
Figura 12 Nube de Palabras de Bigramas Negativos diciembre 2023.....	39
Figura 13 Nube de Palabras de Bigramas Positivos enero 2024 .....	40
Figura 14 Nube de Palabras de Bigramas Neutrales abril 2024 .....	41
Figura 15 Grafico de barras Distribución de Tweets por Ubicación .....	41
Figura 16 Grafico de barras Distribución de sentimientos por Ubicación.....	42
Figura 17 Matriz de Confusión de Modelos (clases balanceadas) datos de entrenamiento.....	46
Figura 18 Matriz de Confusión de Modelos (clases balanceadas) datos de prueba.....	46
Figura 19 Matriz de Confusión de Modelos (clases moderadamente desbalanceadas) datos de entrenamiento.....	49
Figura 20 Matriz de Confusión de Modelos (clases moderadamente desbalanceadas) datos de prueba.....	49
Figura 21 Matriz de Confusión de Modelos (clases totalmente desbalanceadas) datos de entrenamiento.....	52
Figura 22 Matriz de Confusión de Modelos (clases totalmente desbalanceadas) datos de prueba .....	52

Figura 23 Planes de la API de Twitter .....	60
Figura 24 Plataforma APIFY .....	60
Figura 25 Sección Store de la plataforma APIFY .....	61
Figura 26 Twitter Scraper 3.50/1000 .....	61
Figura 27 Twitter URL Scraper 2.50/1000 .....	61
Figura 28 Twitter Scraper 1.50/1000 .....	62
Figura 29 Twitter Scraper V2 0.30/1000 .....	62
Figura 30 Configuración términos de búsqueda .....	62
Figura 31 Resultado de la búsqueda .....	63
Figura 32 Tabla de datos filtrada .....	63
Figura 33 Función para limpiar los tweets.....	64
Figura 34 Función para dar un nuevo formato a la fecha .....	64
Figura 35 Tabla de Datos Después de la Limpieza.....	65
Figura 36 Ejemplo de Clasificación de Sentimientos Incorrecta por pysentiment.....	65
Figura 37 Ejemplo de Clasificación de Sentimientos Incorrecta por Textblob .....	65
Figura 38 Gráfico de Barras de la Distribución de Cantidad por Categoría .....	66

## ÍNDICE DE TABLAS

Tabla 1 Distribución de Cantidad por Categoría .....	29
Tabla 2 Matriz de confusión .....	45
Tabla 3 Exactitud de los Modelos (clases balanceadas) .....	47
Tabla 4 Precisión y Exhaustividad (Recall) de los Modelos (clases balanceadas).....	48
Tabla 5 F1-Score (Recall) de los Modelos (clases balanceadas).....	48
Tabla 6 Exactitud de los Modelos (clases moderadamente desbalanceadas) .....	50
Tabla 7 Precisión y Exhaustividad (Recall) de los Modelos (clases moderadamente desbalanceadas).....	50
Tabla 8 F1-Score (Recall) de los Modelos (clases moderadamente desbalanceadas).....	51
Tabla 9 Exactitud de los Modelos (clases totalmente desbalanceadas).....	53
Tabla 10 Precisión y Exhaustividad (Recall) de los Modelos (clases totalmente desbalanceadas) .....	53
Tabla 11 F1-Score (Recall) de los Modelos (clases totalmente desbalanceadas).....	54

## RESUMEN

El presente trabajo se centra en analizar los sentimientos expresados en la red social “X” *Twitter* con respecto al presidente Daniel Noboa, con el objetivo de determinar el nivel de aceptación por parte de la población ecuatoriana durante el período comprendido desde el 23 de noviembre de 2023 hasta el 30 de abril de 2024. Para ello, se identificaron y recopilaron 3177 *tweets* relevantes utilizando técnicas de *web scraping*, aplicando filtros específicos para excluir *tweets* con imágenes, videos y *retweets*.

La cuantificación y categorización de las respuestas hacia el presidente revelaron una predominancia de *tweets* neutrales del 79.7%, indicando una percepción mayormente neutral. Sin embargo, se observó una notable cantidad de críticas, representadas por el 16.6% de los *tweets* negativos, y relativamente pocos elogios, con solo el 3.7% de los *tweets* siendo positivos.

Los modelos de análisis de sentimientos utilizados incluyeron *Naive Bayes*, *Support Vector Machine (SVM)*, *Logistic Regression*, *Decision Tree*, *Random Forest* y *XGBoost*. Cada modelo fue evaluado utilizando métricas como exactitud, precisión, *recall* y *F1-Score*, tanto en los datos de entrenamiento como en los de prueba.

Los resultados mostraron variaciones significativas en el rendimiento de los modelos, destacándose que el modelo de *Naive Bayes* presentó un buen equilibrio en las métricas de exactitud y *F1-Score*, siendo el más adecuado para los datos desbalanceados. Las conclusiones sugieren que, aunque se obtuvieron resultados satisfactorios, es recomendable extender el período de estudio, incorporar otras redes sociales, incluir *retweets* y respuestas, y explorar modelos de análisis más avanzados para mejorar la precisión.

**Palabras clave:** Análisis de Sentimientos, Redes Sociales, “X” *Twitter*, Nivel de Aceptación, Ecuador, Daniel Noboa.

## **ABSTRACT**

*This study focuses on analyzing the sentiments expressed on the social network "X" Twitter regarding President Daniel Noboa, to determine the level of acceptance among the Ecuadorian population during the period from November 23, 2023, to April 30, 2024. To achieve this, 3,177 relevant tweets were identified and collected using web scraping techniques, applying specific filters to exclude tweets with images, videos, and retweets.*

*The quantification and categorization of the responses toward the president revealed a predominance of neutral tweets at 79.7%, indicating a largely neutral perception. However, a notable amount of criticism was observed, represented by 16.6% of negative tweets, and relatively few praises, with only 3.7% of tweets being positive.*

*The sentiment analysis models used included Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Decision Tree, Random Forest, and XGBoost. Each model was evaluated using metrics such as accuracy, precision, recall, and F1-Score, both on training and test data.*

*The results showed significant variations in the performance of the models, with the Naive Bayes model standing out for its good balance in accuracy and F1-Score metrics, making it the most suitable for unbalanced data. The conclusions suggest that, although satisfactory results were obtained, it is recommended to extend the study period, incorporate other social networks, including retweets and replies, and explore more advanced analysis models to improve accuracy.*

**Keywords:** *Sentiment Analysis, Social Networks, "X" Twitter, Acceptance Level, Ecuador, Daniel Noboa.*

# 1 CAPÍTULO I: INTRODUCCIÓN

## 1.1 Planteamiento del Problema.

En la era digital, la opinión pública se entretene en las redes sociales, dando voz a una diversidad de perspectivas y convirtiéndose en un reflejo dinámico de la sociedad. En este contexto, el papel de plataformas como "X" (*Twitter*) es crucial, siendo un espacio donde las voces de la población se expresan en tiempo real sobre eventos políticos y figuras relevantes (González, 2019).

La era de la información nos ha colocado en un punto de inflexión, donde las percepciones y opiniones sobre líderes políticos se ven moldeadas en gran medida por el ruido digital. Ante esta realidad, surge la imperiosa interrogante: ¿Cómo medir de manera precisa el nivel de aceptación del presidente Daniel Noboa en un entorno tan dinámico como *Twitter*? En el periodo que abarca desde noviembre de 2023 hasta abril de 2024, se plantea la necesidad de explorar las profundidades de las interacciones digitales para extraer *insights* que ayuden a comprender la evolución de la opinión pública en tiempo real.

En este contexto, el presente estudio se enfrenta al desafío de descifrar las emociones y percepciones expresadas en *Twitter* con respecto al presidente Daniel Noboa. Este fenómeno, de naturaleza práctica, requiere de herramientas tecnológicas avanzadas para traducir el clamor digital en una comprensión significativa del nivel de aceptación del nuevo mandatario.

## 1.2 Justificación

La propuesta de este proyecto de investigación adquiere una relevancia excepcional en el ámbito contemporáneo, fusionando el análisis de sentimientos y las redes sociales para desentrañar la compleja dinámica de la aceptación presidencial en el contexto ecuatoriano. La originalidad de este enfoque radica en su capacidad para ofrecer una mirada inmediata y profunda a la opinión pública, un componente crítico en la toma de decisiones políticas.

Desde el punto de vista teórico, este proyecto contribuye al campo de la ciencia de datos y análisis de sentimientos al aplicar estas disciplinas de manera específica a la evaluación de la aceptación presidencial. Al abordar la intersección de la política y la tecnología, se aporta al entendimiento de cómo las plataformas digitales modelan la percepción pública en tiempo real.

La metodología propuesta, centrada en el análisis de sentimientos en *Twitter*, demuestra su utilidad al ofrecer una herramienta efectiva para evaluar la percepción pública en tiempo real. Esta metodología puede ser replicada y adaptada en contextos similares, proporcionando un marco sólido para investigaciones futuras sobre la intersección de redes sociales y política.

Los resultados obtenidos de este análisis pueden ser aplicados directamente en la toma de decisiones políticas y estratégicas. La comprensión detallada de la aceptación presidencial y sus fluctuaciones permite a los líderes políticos ajustar sus enfoques y políticas de manera informada y receptiva a la voz ciudadana.

Este proyecto aspira a aportar una visión profunda y actualizada de la percepción pública hacia el presidente Daniel Noboa. Los beneficios prácticos incluyen la capacidad de anticipar tendencias, identificar áreas de mejora y, en última instancia, fortalecer la conexión entre la administración presidencial y la ciudadanía

### **1.3 Antecedentes**

En la última década, el aumento exponencial del uso de plataformas de redes sociales ha transformado la forma en que la sociedad se expresa y se relaciona con los eventos políticos (Segura, 2023). El análisis de sentimientos, específicamente en plataformas como *Twitter*, ha emergido como una herramienta poderosa para comprender la percepción pública en tiempo real (Navarro Bellido, 2023).

Tal es el caso de Donald Trump, que durante su campaña presidencial aprovechó las redes sociales de manera innovadora y polémica. Utilizó plataformas como *Twitter* para comunicarse directamente con sus seguidores y difundir mensajes controvertidos y provocativos que generaron un gran impacto mediático.

Trump utilizó su cuenta de *Twitter* como una herramienta para expresar sus opiniones, atacar a sus oponentes políticos y desafiar a los medios de comunicación tradicionales. Su estrategia en redes sociales fue fundamental para construir y movilizar su base de seguidores, convirtiendo su presencia en línea en una parte integral de su campaña presidencial (Ortiz Medina, 2021).

Este caso destaca la importancia de profundizar en la capacidad persuasiva de las redes sociales en términos de consecución de votos. Aunque otros factores también influyeron en la victoria de Trump, las redes le permitieron convencer a un número significativo de votantes.

La tecnología y las herramientas de minería de datos proporcionan ahora la capacidad de realizar análisis de sentimientos a gran escala, ofreciendo una visión más completa y dinámica de la opinión pública (Manosalvas, 2022).

Se observa una brecha de conocimiento en la aplicación específica del análisis de sentimientos en *Twitter* para evaluar la aceptación de un presidente en un entorno político específico, como es el caso del Ecuador (Tupiza, 2022). Esta brecha destaca la necesidad de investigaciones que adapten las técnicas de análisis de sentimientos a contextos políticos particulares, aprovechando la riqueza de datos generada por las interacciones en tiempo real en redes sociales.

### **1.4 Objetivos**

#### **1.4.1 Objetivo General**

Analizar los sentimientos expresados en la red social “X” *Twitter* con respecto al presidente Daniel Noboa para determinar el nivel de aceptación por parte de la población ecuatoriana, durante el período comprendido desde el 23 de noviembre del 2023 hasta el 30 de abril del

2024, utilizando técnicas de análisis de sentimientos.

#### **1.4.2 Objetivos Específicos**

- Identificar y recopilar datos relevantes de *Twitter* relacionados con el presidente Daniel Noboa en el periodo comprendido desde el 23 de noviembre del 2023 hasta el 30 de abril del 2024.
- Aplicar técnicas de análisis de sentimientos para evaluar las opiniones expresadas en los *tweets*.
- Cuantificar y categorizar las respuestas positivas y negativas hacia el presidente en la red social.
- Interpretar los resultados para proporcionar una evaluación del nivel de aceptación.

## 2 CAPÍTULO II: MARCO TEÓRICO

### 2.1 Análisis de Sentimientos

#### 2.1.1 Concepto y Definición

El análisis de sentimientos (abreviatura AS o SA en inglés *Sentiment Analysis*) es un campo de investigación de la PLN (Procesamiento de lenguaje natural) que intenta obtener automáticamente información subjetiva sobre un problema específico expresado en el texto de un documento determinado. De esta forma, mediante el análisis de sentimientos, podemos saber si el texto tiene un significado positivo o negativo (Pauli, 2019).

En otras palabras, el análisis de sentimientos es un conjunto de técnicas computacionales que se utilizan para extraer, clasificar, comprender y evaluar opiniones expresadas en fuentes de internet, portales de comentarios y otro contenido generado por los usuarios.

El objetivo principal del análisis de sentimientos es obtener términos semánticos que expresen emociones específicas, comprender las opiniones, actitudes y expectativas de los usuarios sobre temas específicos y analizar el comportamiento de los usuarios cuando se exponen a mensajes para determinar su impacto o poder predecir sus reacciones. Las emociones se clasifican en positivas, negativas o neutras (Moreno, 2020).

Este campo de estudio se ha vuelto crucial en la era de las redes sociales y el *big data*, ya que permite comprender y cuantificar las opiniones y emociones expresadas en grandes volúmenes de datos no estructurados.

#### 2.1.2 Métodos y Herramientas para el Análisis de Sentimientos

El análisis de sentimientos es un campo interdisciplinario que utiliza una variedad de métodos y herramientas para extraer, cuantificar y comprender las emociones y opiniones expresadas en diferentes fuentes de datos, como textos, redes sociales, reseñas de productos, entre otros. A continuación, se describen algunos de los métodos y herramientas comúnmente utilizados en este campo:

**Análisis de Diccionario:** Este método implica el uso de diccionarios de palabras con etiquetas de polaridad (positiva, negativa, neutral) para asignar puntajes a los términos encontrados en el texto. Se suman los puntajes de las palabras para obtener una puntuación general que indica el tono emocional del texto.

El empleo de metodologías basadas en diccionarios simplifica el desarrollo de sistemas de clasificación independientes del contexto, aunque esta aproximación conlleva desafíos inherentes. Uno de estos inconvenientes radica en la pérdida de precisión, ya que las palabras pueden adquirir distintas connotaciones según el entorno en que se empleen.

Por ejemplo, el término "silencioso" puede denotar aspectos positivos al referirse al funcionamiento silencioso de un lavarropas, pero adopta connotaciones negativas al asociarse al sistema de sonido de un televisor. Estos dilemas pueden abordarse mediante la creación de

diccionarios elaborados a partir de vocabularios específicos centrados en el ámbito de estudio deseado (Pauli, 2019).

**Aprendizaje Automático:** Los algoritmos de aprendizaje automático, se utilizan para entrenar modelos que pueden clasificar automáticamente los textos en categorías de sentimientos. Estos modelos aprenden a reconocer patrones en los datos de entrenamiento y luego pueden aplicarse a nuevos textos para predecir sus sentimientos. Por ejemplo:

- **Máquinas de Vectores de Soporte (SVM):** Las SVM son algoritmos de aprendizaje supervisado empleados para la clasificación y regresión. En el análisis de sentimientos, estas SVM se entrenan con muestras de textos previamente categorizados según su polaridad (positiva, negativa, neutra), con el fin de discernir distintas categorías de sentimientos en un espacio multidimensional. Posteriormente, estas SVM tienen la capacidad de asignar automáticamente nuevos textos a una de estas categorías (Moreno, 2020).
- **Árboles de Decisión:** Los árboles de decisión son algoritmos de aprendizaje supervisado que construyen un modelo en forma de estructura de árbol. Cada nodo interno del árbol representa una característica del texto, y las ramas representan las posibles decisiones basadas en esa característica (Ordóñez, 2021). En el análisis de sentimientos, los árboles de decisión pueden aprender patrones en los datos de entrenamiento para clasificar automáticamente nuevos textos en categorías de sentimientos.
- **Redes Neuronales:** Las redes neuronales, basadas en el funcionamiento del cerebro humano, se aplican en el análisis de sentimientos mediante arquitecturas como las redes neuronales convolucionales (CNN) o las redes neuronales recurrentes (RNN). Estas arquitecturas permiten capturar la complejidad y las relaciones contextuales presentes en el texto. Las redes neuronales pueden aprender de grandes conjuntos de datos de entrenamiento y generar representaciones semánticas profundas, las cuales son aprovechadas para la clasificación de sentimientos (Montañés Salas, 2020).

**Enfoques Híbridos:** Muchas veces, se combinan varios métodos para mejorar la precisión del análisis de sentimientos. Por ejemplo, se pueden utilizar métodos de diccionario para la polaridad de las palabras y luego aplicar técnicas de aprendizaje automático para tener en cuenta el contexto y la estructura del texto.

**Análisis de Redes Sociales:** es un campo de estudio que se enfoca en comprender las relaciones sociales y los patrones de interacción que se manifiestan a través de plataformas de redes sociales. El análisis de redes sociales implica el estudio y la interpretación del contenido generado por los usuarios en estas plataformas, con el objetivo de extraer información relevante sobre las opiniones, actitudes y comportamientos de la comunidad (Llivicota Buñay, 2021).

Algunas técnicas específicas utilizadas en el análisis de redes sociales son:

- **Seguimiento de Tendencias:** Esta técnica implica monitorear y analizar los temas que están siendo discutidos con mayor frecuencia en las redes sociales en un período de

tiempo específico. El seguimiento de tendencias permite identificar los temas más relevantes y populares entre los usuarios, lo que puede proporcionar información sobre los temas de interés y las preocupaciones predominantes en la comunidad.

- **Identificación de *Influencers*:** Los *influencers* son usuarios con una gran cantidad de seguidores y un alto grado de influencia en las redes sociales. Identificar y analizar a los *influencers* relevantes en una determinada comunidad puede ser útil para comprender quiénes están impulsando las conversaciones y cómo están afectando las opiniones y actitudes de los demás usuarios.
- **Análisis de la Difusión de Información:** Esta técnica se centra en estudiar cómo se propaga la información a través de las redes sociales. Se pueden utilizar métodos como el análisis de *retweets*, compartidos y menciones para rastrear cómo se comparten y difunden los mensajes en la red. Esto puede proporcionar información sobre la *viralidad* de ciertos temas y la influencia de ciertos usuarios en la difusión de información en la comunidad.
- **Detección de Opiniones Dominantes:** Esta técnica implica identificar las opiniones más prevalentes o dominantes dentro de una comunidad en particular. Se pueden utilizar métodos de análisis de sentimientos para categorizar los mensajes en positivos, negativos o neutrales, y luego determinar cuáles son las opiniones más comunes entre los usuarios. Esto puede proporcionar una comprensión más profunda de las actitudes y percepciones de la comunidad hacia un tema específico.

**Herramientas de *Software* Especializadas:** estas herramientas suelen ofrecer una amplia gama de funcionalidades para procesar, analizar y visualizar datos textuales. Estas funcionalidades pueden incluir la extracción de texto de diversas fuentes (como redes sociales, sitios *web*, documentos), el análisis semántico para comprender el significado de las palabras en su contexto, la visualización de datos y la generación de informes que resuman los hallazgos del análisis (Nazareno Gómez, 2021).

Algunas de las herramientas de *software* más populares para el análisis de sentimientos incluyen NLTK (*Natural Language Toolkit*), *TextBlob*, VADER (*Valence Aware Dictionary and sEntiment Reasoner*), *IBM Watson Natural Language Understanding*, *Google Cloud Natural Language API*, entre otras. Estas herramientas pueden ser utilizadas por investigadores, analistas de datos, empresas y organizaciones para obtener información valiosa sobre las opiniones y sentimientos de los usuarios expresados en texto.

## 2.1.3 Procesamiento del Lenguaje Natural

### 2.1.3.1 Minería De Datos

La minería de datos es un campo interdisciplinario que combina técnicas de estadística, aprendizaje automático, inteligencia artificial y bases de datos para descubrir patrones ocultos, relaciones y conocimientos valiosos en grandes conjuntos de datos. Estos datos pueden ser estructurados, semiestructurados o no estructurados, como texto, imágenes, audio, etc (Pérez-Suasnavas, 2020).

En el contexto del análisis de sentimientos, la minería de datos desempeña un papel crucial al permitir a los investigadores y analistas extraer información relevante de grandes volúmenes de texto no estructurado, como aquellos provenientes de redes sociales, blogs, noticias en línea, foros de discusión y otros medios digitales (Steiner Corrêa, 2019).

Algunas características importantes de la minería de datos son:

- **Exploración de Datos:** La minería de datos permite explorar grandes volúmenes de datos de texto procedentes de diversas fuentes, como redes sociales, *blogs*, foros en línea, noticias y más. Este proceso de exploración es fundamental para comprender la naturaleza y la calidad de los datos disponibles para el análisis de sentimientos.
- **Preprocesamiento de Texto:** Antes de aplicar técnicas de análisis de sentimientos, es necesario realizar un preprocesamiento de los datos de texto. Esto puede incluir la eliminación de caracteres especiales, la *tokenización*, la eliminación de palabras que no aportan significado (*stopwords*), la lematización o el *stemming*, entre otros procesos. La minería de datos proporciona herramientas y técnicas para realizar este preprocesamiento de manera eficiente.
- **Identificación de Patrones:** Una vez que los datos de texto están preprocesados, la minería de datos se utiliza para identificar patrones y tendencias significativas en los textos. Esto puede implicar el uso de algoritmos de aprendizaje automático, como la clasificación, la agrupación (*clustering*) o la asociación, para descubrir relaciones entre palabras, temas recurrentes, y otros aspectos relevantes para el análisis de sentimientos.
- **Visualización de Datos:** La visualización de datos es una parte importante de la minería de datos, ya que permite representar gráficamente los resultados del análisis. Esto facilita la interpretación de los patrones identificados y ayuda a comunicar los hallazgos de manera efectiva a otras partes interesadas.
- **Análisis de Sentimientos:** La minería de datos proporciona herramientas y técnicas para el análisis de sentimientos, permitiendo la evaluación automatizada de las emociones y opiniones expresadas en texto. Esto puede incluir la clasificación de textos en categorías de sentimientos (positivo, negativo, neutro), la identificación de temas o aspectos específicos de interés, y la detección de opiniones dominantes en la comunidad.
- **Modelado Predictivo:** Además de analizar datos históricos, la minería de datos también puede emplearse para construir modelos predictivos que ayuden a predecir tendencias futuras en el análisis de sentimientos. Estos modelos pueden basarse en técnicas de aprendizaje supervisado o no supervisado para clasificar automáticamente nuevos datos textuales en categorías de sentimientos.

### 2.1.3.2 Procesamiento del lenguaje natural (PLN)

El Procesamiento del Lenguaje Natural (PLN) es el área de investigación que se dedica a comprender los mecanismos del lenguaje, incluyendo su estructura, generación y las diferentes operaciones relacionadas con su manipulación. Entre estas operaciones se encuentran la creación de texto, la traducción entre idiomas, la elaboración de preguntas y respuestas, la

síntesis de resúmenes y la creación de sistemas de conversación automatizados, conocidos como *chatbots*, entre otros (Néstor Camilo Beltrán, 2021).

El objetivo principal del PLN es habilitar a las computadoras para comprender y procesar el lenguaje humano de manera eficaz. Este proceso implica analizar la semántica, la gramática, el contexto y otros aspectos del lenguaje (Manosalvas Porras, 2022). Para lograr esto, se utilizan diversas técnicas y métodos, como el aprendizaje automático, la lingüística computacional, la estadística, la teoría de la información, entre otros.

Una de las aplicaciones más destacadas del PLN es el análisis de sentimientos, que consiste en evaluar y comprender las emociones y opiniones expresadas en el texto. Esto implica la identificación de palabras clave, la detección de tono emocional, la clasificación de sentimientos y más.

Aunque el PLN enfrenta varios desafíos, como la ambigüedad del lenguaje, el reconocimiento de contexto, la comprensión de la ironía y el sarcasmo, entre otros. Sin embargo, los avances en técnicas de aprendizaje automático, el aumento en el poder computacional y el acceso a grandes conjuntos de datos han impulsado significativamente el progreso en este campo en las últimas décadas (López Hernández, 2022).

### 2.1.3.3 Aplicaciones del Procesamiento del Lenguaje Natural

**Detección de *Spam*:** las empresas proveedoras de servicios de correo electrónico emplean el PLN para identificar y filtrar mensajes no deseados o maliciosos. Este procedimiento implica analizar los correos electrónicos en busca de patrones lingüísticos que sugieran *spam* o intentos de *phishing*.

Estos patrones suelen manifestarse a través de la utilización excesiva de términos financieros, errores gramaticales, expresiones intimidantes, solicitudes de acción urgente, nombres de empresas importantes mal escritas, entre otros indicadores.

**Traducción automática:** herramientas como *Google Translate*, ejemplifica el uso extendido del PLN Una traducción automática efectiva que va más allá de simplemente sustituir palabras de un idioma por palabras en otro. Implica capturar de manera precisa tanto el significado como el tono del texto original y transmitirlos de manera fiel y con el mismo efecto en el idioma de destino.

**Agentes virtuales y *chatbots*:** son sistemas tecnológicos que han ganado prominencia en diversos ámbitos. Desde sistemas GPS que responden a comandos de voz hasta *software* de dictado de voz a texto, pasando por asistentes virtuales como Siri de *Apple* y *Alexa* de *Amazon*.

Estos sistemas hacen uso del reconocimiento de voz para interpretar los comandos recibidos y emplean técnicas de generación de lenguaje natural para ofrecer respuestas apropiadas o proporcionar información útil en respuesta a las consultas realizadas.

**Resumen de texto:** emplea técnicas de PLN para analizar grandes cantidades de texto digital

y generar resúmenes concisos y sinopsis. Estos resúmenes son útiles para crear índices, alimentar bases de datos de investigación o satisfacer a lectores que disponen de poco tiempo para leer el texto completo.

Las aplicaciones más avanzadas de resumen de texto aplican razonamiento semántico y generación de lenguaje natural (GLN) para añadir contexto relevante y extraer conclusiones significativas en los resúmenes.

**El análisis del sentimiento de las redes sociales** ha emergido como una herramienta esencial en el ámbito comercial para desentrañar información clave de los canales de redes sociales. Mediante el PLN, esta herramienta examina el lenguaje natural empleado por los usuarios en diversas plataformas sociales, extrayendo datos relacionados con las reacciones ante estímulos como productos, promociones y eventos. Estos datos pueden ser aprovechados por las empresas en una variedad de operaciones, incluyendo diseño y estrategias de *marketing*.

## **2.2 Redes Sociales**

### **2.2.1 Definición de Redes Sociales**

Las redes sociales son plataformas digitales que facilitan la conexión y comunicación entre individuos a través de la creación y compartición de contenido. Estas plataformas permiten la interacción en tiempo real, la difusión masiva de información y la formación de comunidades virtuales. Desempeñan un papel central en la construcción de la opinión pública y la influencia en la esfera política al proporcionar un espacio para la expresión individual y la participación colectiva (Redondo Romano, 2022).

Las redes sociales suelen ofrecer una variedad de funciones y características, que pueden incluir perfiles de usuario, noticias o *feeds* de contenido, herramientas de mensajería, opciones de compartir, funciones de etiquetado, grupos de interés, eventos, y más. Estas características fomentan la conexión y la participación de los usuarios, creando comunidades virtuales en torno a intereses compartidos, relaciones personales o actividades específicas.

Además de su uso personal y social, las redes sociales también tienen un impacto significativo en los negocios, el *marketing*, la política, la educación y otros ámbitos. Las empresas y organizaciones utilizan estas plataformas para promocionar productos y servicios, interactuar con clientes, reclutar talento, realizar investigaciones de mercado y gestionar la reputación de la marca (Valero Herrero, 2023).

Las redes sociales son un fenómeno cultural y tecnológico que ha transformado la forma en que nos comunicamos, colaboramos y nos relacionamos en la era digital, con implicaciones significativas en diversos aspectos de la vida cotidiana y la sociedad en general (Lozano-Blasco, 2023).

### **2.2.2 Aplicaciones del Análisis de Sentimientos en Redes Sociales**

Las aplicaciones del análisis de sentimientos en redes sociales son diversas y abarcan varios

campos:

**Gestión de la Reputación Online:** El análisis de sentimientos ayuda a las organizaciones a evaluar y gestionar su reputación en línea. Al monitorear las conversaciones en redes sociales, las empresas pueden identificar y abordar rápidamente comentarios negativos o crisis de reputación, así como también destacar y amplificar los comentarios positivos.

**Análisis de Opinión Política:** En el ámbito político, el análisis de sentimientos en redes sociales se utiliza para medir la percepción pública sobre candidatos, partidos políticos y cuestiones políticas. Esto proporciona a los políticos y estrategas una visión profunda de las preferencias y opiniones de los votantes, lo que les permite adaptar sus mensajes y estrategias de campaña de manera más efectiva.

**Análisis de Campañas de Marketing:** El análisis de sentimientos proporciona información valiosa sobre la eficacia de las campañas de *marketing* en redes sociales. Al evaluar las respuestas emocionales de los usuarios ante diferentes mensajes y contenido promocional, las organizaciones pueden optimizar sus estrategias y mejorar el rendimiento de sus campañas.

**Servicio al Cliente:** Las empresas pueden emplear el análisis de sentimientos para mejorar su servicio al cliente al identificar rápidamente las quejas y preocupaciones de los usuarios en redes sociales. Esto les permite responder de manera oportuna y resolver problemas antes de que afecten la satisfacción del cliente.

Ejemplos:

En un estudio de mercado utilizando la red social *Twitter* para los Hoteles Decameron se encontró un alto grado de aceptación de los servicios e instalaciones ofrecidos por el Hotel Decameron Mompiche, especialmente durante los años 2012 al 2015. Sin embargo, se observó una disminución en el número de opiniones positivas a partir del año 2016 (Manosalvas Porras, 2022). Estos resultados demuestran la utilidad del análisis de sentimientos en la comprensión de la satisfacción del cliente y la percepción de la marca en redes sociales.

Otro estudio realizado en el 2022 en *Twitter* de 395 *tweets* evaluados reveló que el 68% de los *tweets* relacionados con el Sr. Guillermo Lasso fueron neutrales, lo que indica que la mayoría de los mensajes carecían de sentimientos explícitos o eran irrelevantes. Un 21% de los *tweets* expresaron opiniones negativas, destacando términos como "renuncia" y "delincuencia". En contraste, el 12% de los *tweets* fueron positivos, con palabras como "bien" y "adelante" (Maldonado Ramones, 2022). Estos resultados ofrecen una visión clara de la percepción de los usuarios sobre el Sr. Guillermo Lasso en *Twitter*.

### 2.2.3 *Twitter*

*Twitter* recientemente renombrado como "X" es una plataforma de redes sociales que permite a los usuarios enviar y leer mensajes cortos, conocidos como "*tweets*". Fundada en 2006, se ha convertido en una de las redes sociales más populares y de rápido crecimiento en todo el mundo, con millones de usuarios activos. Su diseño simple y su límite de caracteres por

mensaje han contribuido a su singularidad y popularidad (Molano, 2021).

Los usuarios pueden seguir a otras personas, empresas y organizaciones para mantenerse al tanto de sus actualizaciones y publicaciones. Además, *Twitter* se ha convertido en una fuente importante de información en tiempo real, utilizada para noticias de última hora, debates políticos, campañas de *marketing* y mucho más.

### 2.2.3.1 Características de *Twitter*

**Limitación de caracteres:** Los *tweets* están limitados a 280 caracteres, lo que promueve la brevedad y la concisión en la comunicación.

**Seguimiento:** Los usuarios pueden seguir a otros usuarios para ver sus *tweets* en su línea de tiempo. Esto permite a las personas mantenerse actualizadas sobre las últimas noticias, eventos y publicaciones de interés.

**Retweets:** Los usuarios pueden compartir los *tweets* de otros en su propia línea de tiempo, lo que facilita la difusión de información y la participación en conversaciones.

**Menciones:** Los usuarios pueden mencionar a otros usuarios en sus *tweets* utilizando el símbolo "@" seguido del nombre de usuario. Esto permite iniciar conversaciones y dirigirse a personas específicas.

**Hashtags:** Los *hashtags* son palabras clave o frases precedidas por el símbolo "#". Se utilizan para etiquetar y categorizar los *tweets* por tema, lo que facilita la búsqueda y la participación en conversaciones sobre temas específicos.

**Tendencias:** *Twitter* muestra los temas más populares y discutidos en tiempo real en su sección de "Tendencias". Esto permite a los usuarios estar al tanto de lo que es relevante y popular en ese momento.

**API:** *Twitter* proporciona una *API* que permite a los desarrolladores acceder a datos en tiempo real y crear aplicaciones externas que interactúen con la plataforma de diversas formas, lo que ha impulsado la creación de herramientas de análisis, monitoreo y gestión de contenido.

### 2.2.3.2 Uso de *Twitter* en Investigaciones Sociales

En el estudio de (Cebral-Loureda, 2021), se examinó el papel de *Twitter* en los primeros meses de la pandemia de *COVID-19*. Se recopilaron 231,375 *tweets* de España y América Latina durante dos meses. El análisis de sentimientos reveló una clara predominancia de emociones negativas. El miedo fue la emoción más destacada, seguido de la confianza y la tristeza.

Particularmente hacia la segunda mitad de abril, la sensación de tristeza superó levemente a la de confianza. Esto coincidió con un aumento significativo de las muertes a nivel mundial, lo que generó una amplia gama de emociones negativas como dolor, hastío y frustración. Estos hallazgos destacan la importancia del análisis de texto computacional en la comprensión de la evolución de las percepciones sobre *COVID-19*, especialmente en una plataforma global como

*Twitter*.

En otro estudio realizado por (Cortés Díaz, 2021), se examinó un conjunto de datos que consistía en 50 *tweets* para cada marca, extraídos de las cuentas oficiales de *PlayStation* y *Xbox*. Los resultados revelaron que *PlayStation* obtuvo 33 comentarios positivos, 7 negativos y 10 neutrales, con una puntuación total de 102. Por otro lado, *Xbox* registró 21 comentarios positivos, 13 negativos y 16 neutrales, con una puntuación total de 66.

Aunque ambas marcas recibieron una cantidad significativa de comentarios positivos, también se observaron algunos comentarios negativos. Cada comentario positivo sumó 3 puntos, mientras que los neutrales sumaron 1 punto y los negativos restan 1 punto. Esto permitió evaluar la percepción general de los usuarios hacia las marcas. Se concluyó que *PlayStation* obtuvo una puntuación más alta que *Xbox*, lo que sugiere una mejor valoración por parte del público y una posición más sólida en el mercado de los videojuegos.

### **2.2.3.3 Limitaciones y Consideraciones Éticas**

El uso de *Twitter* como fuente de datos para investigaciones sociales y análisis de sentimientos presenta algunas limitaciones importantes que deben ser consideradas. Una de estas limitaciones radica en la representatividad de la muestra. Dado que no todos los usuarios de *Twitter* representan una muestra demográfica equitativa, los resultados pueden sesgarse hacia ciertos grupos demográficos o regiones geográficas, lo que podría afectar la validez de los hallazgos (Badiola Ramos, 2019).

Además, la calidad de los datos en *Twitter* puede variar considerablemente. Los *tweets* pueden contener errores ortográficos, lenguaje informal o incluso información falsa, lo que puede dificultar el análisis preciso de los sentimientos o percepciones expresadas. Además, algunos usuarios pueden tener múltiples cuentas o utilizar *bots* para amplificar artificialmente ciertas opiniones, lo que puede distorsionar la percepción de la opinión pública.

Desde una perspectiva ética, es crucial considerar la privacidad y el consentimiento de los usuarios al utilizar sus datos en investigaciones. Aunque los *tweets* son en su mayoría datos públicos, los investigadores deben ser cuidadosos al manejar la información personal y proteger la identidad de los usuarios cuando sea necesario. Además, es importante garantizar que el análisis de sentimientos no cause daño o discriminación a ningún grupo o individuo identificable (Butron, 2023).

Otra consideración ética importante es la transparencia y la honestidad en la presentación de los resultados. Los investigadores deben ser transparentes sobre las metodologías utilizadas en el análisis de sentimientos y las limitaciones de los datos, evitando exagerar o distorsionar los hallazgos para obtener conclusiones sensacionalistas o engañosas.

Si bien *Twitter* ofrece una valiosa fuente de datos para investigaciones sociales y análisis de sentimientos, es fundamental abordar adecuadamente las limitaciones y consideraciones éticas asociadas con su uso. Esto garantizará que las investigaciones sean válidas, éticas y contribuyan de manera significativa al conocimiento científico y social.

## **2.3 Opinión Pública**

### **2.3.1 Concepto y Definición**

La opinión pública se define como el conjunto de creencias, actitudes y percepciones compartidas por un grupo de individuos en relación con un tema específico en un momento determinado. Esta noción refleja la manifestación colectiva de la voluntad y los juicios de una sociedad sobre cuestiones políticas, sociales, culturales o económicas (Sánchez Lorenzo, 2019).

También puede definirse como el resultado de la interacción de diversos factores, como la información disponible, las experiencias personales, los valores culturales y las influencias sociales. Se trata de un fenómeno dinámico y cambiante, que puede ser moldeado por la acción de líderes políticos, medios de comunicación, grupos de interés y otros agentes sociales (Contreras Ruiz, 2022).

La comprensión de la opinión pública es esencial para los procesos democráticos, ya que proporciona información sobre las preferencias, demandas y percepciones de la ciudadanía, lo que permite a los líderes políticos y tomadores de decisiones adaptar sus políticas y estrategias en función de las necesidades y expectativas del pueblo.

El surgimiento de las redes sociales y las nuevas tecnologías de la comunicación ha transformado la dinámica de la opinión pública, permitiendo una mayor participación y expresión de las opiniones de los individuos. Plataformas como *Twitter*, *Facebook* y *YouTube* han ampliado el alcance y la velocidad con la que se difunden las ideas y las percepciones, creando nuevas oportunidades y desafíos para el análisis de la opinión pública.

### **2.3.2 Nivel de Aceptación**

El nivel de aceptación se refiere a la medida en que una determinada persona, idea, producto o evento es aceptada o aprobada por un grupo o una comunidad en particular. En el contexto del análisis de la opinión pública, el nivel de aceptación se utiliza para evaluar cómo es percibido un líder político, una política gubernamental, una empresa, un producto o cualquier otro aspecto relevante para la sociedad (Zuniga, 2023).

El nivel de aceptación puede variar considerablemente según diversos factores, como el contexto histórico, cultural, político y socioeconómico, así como también en función de la información disponible, las experiencias personales y figuras públicas.

Para evaluar el nivel de aceptación, se pueden utilizar diferentes metodologías, como encuestas de opinión, análisis de redes sociales, seguimiento de medios de comunicación, entre otros. Estas herramientas permiten recopilar y analizar datos sobre las actitudes y percepciones de la población en relación con un tema específico (Bárcenas Reséndiz, 2023).

Por otra parte, la aceptación política representa la medida en que un líder político o una figura de autoridad es respaldada y apoyada por la población. La aceptación política se ve influenciada

por factores como el desempeño gubernamental, la comunicación efectiva y la gestión de crisis, y se manifiesta a través de diversas formas de expresión, incluyendo las interacciones en redes sociales (Rojas Beltrán, 2023).

## **2.4 Ecuador**

Ecuador es una república democrática ubicada en la región noroeste de América del Sur. Limita al norte con Colombia, al sur y al este con Perú, y al oeste con el Océano Pacífico. El país se divide en 24 provincias y su sistema político se basa en la democracia representativa, con un presidente elegido por voto popular cada cuatro años.

Ecuador tiene una economía mixta, con sectores como la agricultura, la industria petrolera, el turismo y los servicios financieros desempeñando roles importantes. El país es conocido por su biodiversidad, que incluye la selva amazónica, las Islas Galápagos y la sierra andina. A pesar de los recursos naturales, Ecuador enfrenta desafíos económicos, incluida la desigualdad de ingresos y la pobreza (Quintero, 2021).

Es un país multicultural y multiétnico, con una población diversa que incluye comunidades indígenas, mestizas, afroecuatorianas y otras. La identidad nacional ecuatoriana se ha formado a partir de una mezcla de influencias indígenas, europeas y africanas, reflejadas en su arte, música, comida y tradiciones (Rodríguez, 2021). Los ecuatorianos tienen un fuerte sentido de la identidad nacional y una historia de activismo social y político.

Ecuador ha experimentado varios movimientos sociales y protestas a lo largo de su historia reciente, incluida la Revolución Ciudadana liderada por el expresidente Rafael Correa. La participación ciudadana en la política es importante en Ecuador, con la sociedad civil y los grupos comunitarios desempeñando un papel activo en la promoción de cambios políticos y sociales (Hurtado, 2019).

El uso de redes sociales, como *Twitter*, ha aumentado en Ecuador en los últimos años, brindando a los ciudadanos una plataforma para expresar sus opiniones políticas y sociales. El análisis de sentimientos en *Twitter* se ha convertido en una herramienta útil para medir la percepción pública sobre temas políticos y para entender las tendencias de opinión en tiempo real.

## **2.5 Daniel Noboa**

Daniel Roy Gilchrist Noboa Azín es un administrador, empresario y político ecuatoriano-estadounidense, nacido el 30 de noviembre de 1987 en Miami, Estados Unidos. Antes de su ascenso a la presidencia, fue asambleísta nacional del Ecuador entre 2021 y 2023, representando al movimiento Ecuatoriano Unido. Es miembro del movimiento político Acción Democrática Nacional.

El 15 de octubre de 2023, ganó las elecciones anticipadas para el período 2023-2025, convirtiéndose en el presidente constitucional de la República del Ecuador. Con 35 años de

edad en el momento de su elección, Daniel Noboa es el segundo presidente más joven en la historia de Ecuador y el más joven electo por voto popular (Herrera, 2023).

El 23 de noviembre de 2023, asumió la presidencia de Ecuador para completar el mandato inconcluso del presidente saliente Guillermo Lasso. Su mandato está programado para durar 18 meses, según lo establecido por las elecciones anticipadas. La presidencia de Noboa se produce en un momento crucial para Ecuador, con desafíos políticos, económicos y sociales significativos que enfrentar.

Durante su presidencia, se espera que Daniel Noboa implemente políticas y programas para abordar los problemas prioritarios del país y cumplir con las expectativas de la ciudadanía. Dada la naturaleza joven y dinámica de la presidencia de Daniel Noboa, es crucial comprender la percepción pública de su liderazgo.

*Twitter*, como una plataforma de redes sociales ampliamente utilizada en Ecuador, ofrece una ventana única para analizar el nivel de aceptación y las opiniones del público sobre las políticas y acciones de Noboa.

### 3 CAPÍTULO III: METODOLOGÍA

El diseño de la investigación se caracterizó por ser no experimental y longitudinal, lo que implica el seguimiento de los usuarios de *Twitter* que mencionan al presidente Daniel Noboa durante el período de noviembre de 2023 a abril de 2024. Se empleó un enfoque cuantitativo para analizar datos recopilados de una muestra representativa de usuarios que fueron seleccionados aleatoriamente en la plataforma.

La recopilación de datos se llevó a cabo mediante la obtención de *tweets* relacionados con el presidente Daniel Noboa utilizando herramientas de extracción de datos como la *API* de *Twitter* y herramientas de minería de datos. Esta recopilación se realizó de manera continua durante el periodo mencionado, con el objetivo de capturar una amplia gama de opiniones y tendencias en relación con la aceptación del presidente.

El análisis de sentimientos de los *tweets* recopilados implicó un preprocesamiento de texto que incluyó *tokenización*, eliminación de *stop words* y normalización. Se utilizó herramientas de procesamiento de lenguaje natural para determinar la polaridad (positiva, negativa, neutral) de cada *tweet*, lo que permitió evaluar la percepción pública del presidente Daniel Noboa a lo largo del tiempo.

Se desarrollaron indicadores específicos para cuantificar la polaridad y el volumen de *tweets* relacionados con la aceptación del presidente, lo que facilitó la medición del impacto de su gestión en la opinión pública. Se evaluó la precisión del análisis de sentimientos y se registraron métricas como la frecuencia de menciones y las tendencias temporales para obtener una comprensión más profunda de los resultados.

El análisis estadístico se centró en un enfoque descriptivo para describir la distribución de opiniones a lo largo del tiempo. Se utilizó herramientas y técnicas estadísticas para visualizar y analizar los datos de manera clara y concisa, lo que facilitó la interpretación de los resultados y la identificación de patrones significativos.

Se procedió a dividir el *dataset* en conjuntos de entrenamiento y prueba (*train* y *test*), con los cuales se entrenó un modelo de análisis de sentimientos utilizando un algoritmo de *machine learning*, que fue seleccionado en función de su desempeño en tareas similares.

Una vez entrenado, se evaluó el desempeño del modelo mediante métricas como la exactitud, precisión, *recall* (sensibilidad) y *F1-Score*. Se utilizó una matriz de confusión para analizar la distribución de sentimientos y la eficacia del modelo en la clasificación de los *tweets*.

Las conclusiones y recomendaciones se basaron en una interpretación de los hallazgos en términos de la aceptación del presidente Daniel Noboa.

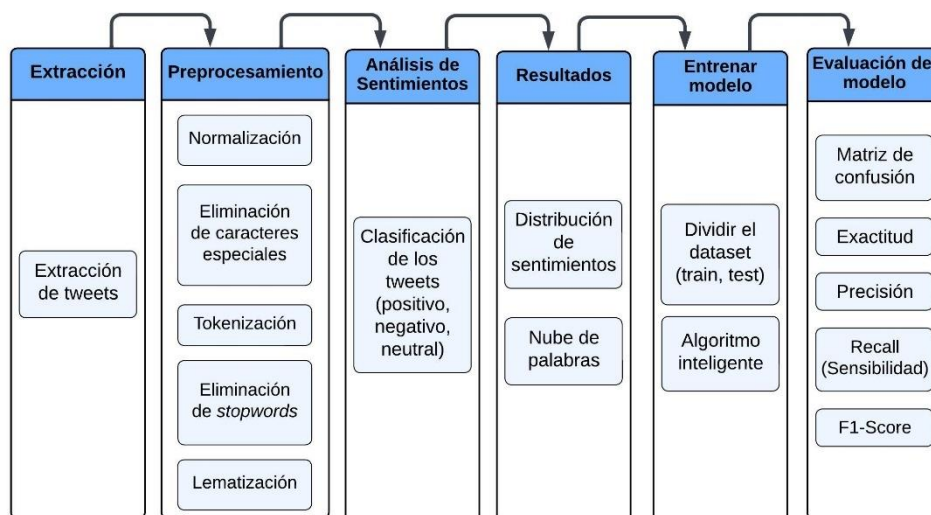


Figura 1 Esquema general del enfoque metodológico propuesto  
Fuente: Autor

### 3.1 Desarrollo

#### 3.1.1 Obtención de los datos

Para recopilar datos, inicialmente se planeaba utilizar la API de Twitter. Sin embargo, debido a que no permite la extracción de tweets (Figura 23), se exploraron otras opciones.

Como alternativa, se decidió recurrir al *scraping*, una técnica para extraer datos de sitios web. Básicamente, un programa de *scraping* navega por las páginas web, analiza su estructura y extrae la información relevante de acuerdo con los criterios establecidos. Por ejemplo, en la plataforma APIFY, el plan gratuito permite realizar la extracción de tweets con un límite de hasta \$5 y 8 GB de memoria (Figura 24).

En la sección de Store (Figura 25), se pueden encontrar "actores", que son scripts diseñados para realizar el *scraping* en diversas redes sociales. En nuestro caso, nos enfocamos en buscar actores específicamente para Twitter.

Cada actor tiene un precio diferente por cada conjunto de 1000 tweets. Por ejemplo, hay opciones que van desde \$3.50 hasta \$1.50 por cada 1000 tweets (Figura 26, 27 y 28).

Luego de una búsqueda exhaustiva, se identificó un actor que ofrece la misma cantidad de extracción por tan solo \$0.30 por cada 1000 tweets (Figura 29).

Después de elegir el actor adecuado, procedimos a configurar los términos de búsqueda.

Donde (Figura 30):

- (1) Daniel Noboa, es el término principal de búsqueda.
- (2) El margen de fechas de extracción de los tweets es desde el 23/11/2023 hasta el 30/04/2024

(3) Filtra y excluye todos los *tweets* que tengan imágenes o videos.

(4) Filtra y excluye todos los *retweets*.

Tras ejecutar la búsqueda, el actor nos proporcionó un total de 3177 *tweets*, cada uno con 87 atributos (*Figura 31*).

La mayoría de los atributos presentaban datos faltantes o carecían de relevancia para la investigación. Por lo tanto, se procedió a reducir la cantidad de atributos, dejando un total de 3 que resultaron ser más pertinentes para el análisis (*Figura 32*).

- **text:** el *tweet*
- **createdAt:** la fecha de publicación del *tweet*
- **author/location:** localidad de la cuenta

### 3.1.2 Limpieza de los datos

La limpieza de datos es un proceso fundamental que implica la eliminación o sustitución de palabras que no son relevantes para el análisis. Por ejemplo, esto puede incluir la normalización de letras con acentos, transformando todo el texto a minúsculas para evitar duplicados debido a la diferencia de capitalización, la eliminación de *emojis* que no aportan significado al texto, la eliminación de marcas de *retweet* (RT), números y palabras comunes que no aportan información semántica, conocidas como "*stop words*"

La limpieza se realizó mediante la función (*Figura 33*).

**Normalización de texto:** Este proceso implica la conversión del texto a minúsculas, asegurando coherencia en el análisis al tratar todas las palabras como equivalentes sin importar la capitalización.

```
tweet = tweet.lower()
```

**Eliminación de caracteres especiales y puntuación:** Los *tweets* suelen contener caracteres especiales, como *emojis*, menciones, *hashtags* y signos de puntuación. Estos elementos pueden no ser relevantes para el análisis de sentimientos y pueden afectar la precisión de los modelos.

```
tweet = re.sub(r'@[a-zA-Z0-9_]+', '', tweet)
```

```
tweet = re.sub(r'^[a-zA-Z\s]', '', tweet)
```

**Tokenización:** La *tokenización* implica dividir el texto en unidades más pequeñas, como palabras o términos. En el caso de los *tweets*, se divide el texto en palabras individuales o *tokens*. Esto facilita el procesamiento posterior del texto y permite realizar análisis a nivel de palabra.

```
tokens = word_tokenize(tweet)
```

**Eliminación de stopwords:** Las *stopwords* son palabras comunes que no aportan mucho significado al análisis, como "el", "la", "y", "o". Estas palabras suelen eliminarse durante el preprocesamiento para reducir el ruido en los datos y mejorar la eficacia del análisis.

```
stop_words = set(stopwords.words('spanish'))
```

```
tokens = [word for word in tokens if word not in stop_words]
```

**Lematización o stemming:** Estas técnicas se utilizan para reducir las palabras a su forma base o raíz. La lematización busca el lema de una palabra en función de su significado, mientras que el *stemming* simplemente elimina los sufijos para obtener la raíz de la palabra. Ambas técnicas ayudan a normalizar el texto y a reducir la variabilidad en los datos.

```
tokens = [lemmatizer.lemmatize(word) for word in tokens]
```

Con la función (Figura 34), se convierten los datos de la fecha de un formato “Sat Nov 25 23:57:01 +0000 2023” a otro más comprensible “Noviembre 2023”

Mediante el siguiente código los datos faltantes en el atributo “*author/location*” se llenaron con el dato “Desconocido” en otro atributo nombrado “*ubicacion*” para así garantizar que no exista datos faltantes.

```
df_tweets['ubicacion'] = df_tweets['author/location'].fillna('Desconocido')
```

Finalmente, los datos quedaron limpios y organizados como se muestra en la (Figura 35).

### 3.1.3 Análisis de sentimientos

Para el análisis de sentimientos, inicialmente se utilizaron las librerías *pysentiment* y *textblob*. Estas herramientas fueron empleadas con el objetivo de automatizar la clasificación de los sentimientos en los *tweets*. Sin embargo, tras varias pruebas y evaluaciones, se observó que los resultados obtenidos no eran óptimos ni confiables. Las clasificaciones automáticas no reflejaban con precisión el contexto y la semántica de los *tweets* en español, presentando inconsistencias significativas (Figura 36, 37).

Debido a la necesidad de contar con datos más precisos y representativos, se tomó la decisión de clasificar manualmente los sentimientos de los *tweets*. Este enfoque, aunque más laborioso, permitió una clasificación más precisa y adecuada de los sentimientos expresados. Sin embargo, es importante reconocer que este método introduce el sesgo del investigador. Aun así, este sesgo es preferible a no contar con datos adecuados para el análisis, ya que permite obtener una visión más fiel de la realidad observada.

## 4 CAPÍTULO IV: RESULTADOS

### 4.1 Análisis General

#### 4.1.1 Distribución de Cantidad por Categoría

En el análisis de sentimientos realizado sobre un total de 3,170 *tweets*, se obtuvo la siguiente distribución de categorías:

Sentimiento	N° Tweets	Porcentaje
Neutral	2,526	79.7%
Negativo	527	16.6%
Positivo	117	3.7%
<b>Total</b>	<b>3170</b>	<b>100%</b>

*Tabla 1 Distribución de Cantidad por Categoría*

*Fuente: Autor*

Estos resultados se visualizan mediante un gráfico de barras que muestra la cantidad de *tweets* en cada categoría (*Figura 38*).

La alta proporción de *tweets* neutrales (79.7%) sugiere que la mayoría de las menciones en *Twitter* sobre el presidente Daniel Noboa no expresan un sentimiento fuerte, ni positivo ni negativo. Esto puede indicar que muchas personas están simplemente informando o comentando sobre acciones y eventos relacionados con su presidencia sin emitir juicios de valor.

El 16.6% de *tweets* negativos refleja una presencia considerable de críticas y desaprobación hacia el presidente. Esto podría estar relacionado con ciertas políticas, decisiones o situaciones recientes que han generado descontento entre los usuarios de *Twitter*.

Por otro lado, solo el 3.7% de los *tweets* son positivos, lo que indica que hay relativamente pocas menciones de apoyo o aprobación explícita hacia el presidente Daniel Noboa. Esto quiere dar a entender que las acciones o decisiones del presidente no están generando una respuesta significativamente positiva en la plataforma.

En general, estos resultados indican que la percepción pública en *Twitter* sobre el presidente Daniel Noboa tiende a ser mayormente neutral, con una notable cantidad de críticas y relativamente pocos elogios.

#### 4.1.2 Nube de Palabras

La nube de palabras proporciona una representación visual de las palabras más frecuentes en los *tweets* categorizados como negativos, positivos y neutrales. Este enfoque permite identificar rápidamente los términos más comunes asociados con cada tipo de sentimiento y, por ende, entender mejor las temáticas y preocupaciones dominantes en la conversación.















**Tweets Positivos (Figura 9):**



*Figura 9 Nube de Palabras de Trigramas Positivos  
Fuente: Autor*

**“Derogacion tabla consumo”, “tabla consumo drogas”, y “prohibia detener narcotraficantes”** indican un enfoque positivo hacia cambios en la legislación relacionada con las drogas. La derogación de la tabla de consumo y la capacidad de detener narcotraficantes son vistas favorablemente.

**“Propuesta nueva ley”** sugiere apoyo hacia nuevas propuestas legislativas.

**“Destacar loable labor”, “loable labor papel”, y “labor papel canciller”** elogian el trabajo del canciller y su papel en varios eventos internacionales.

**“Iva materiales construcción”** muestra apoyo hacia medidas que afectan positivamente al sector de la construcción.

## Tweets Neutrales (Figura 10):

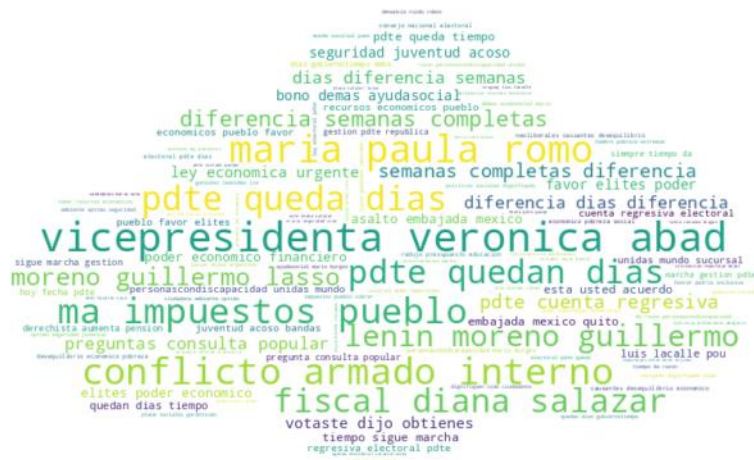


Figura 10 Nube de Palabras de Trigramas Neutrales

Fuente: Autor

“**Vicepresidenta veronica abad**” y “**Maria paula romo**” son mencionadas frecuentemente, lo que sugiere que estas figuras públicas son temas comunes en la conversación, aunque no se les asocia necesariamente con emociones fuertes.

“**Fiscal diana Salazar**” aparece con cierta frecuencia, destacando su relevancia en el contexto actual.

“**Conflicto armado interno**” y “**Asalto embajada mexico**” son eventos significativos mencionados de manera neutral, probablemente como reportes o discusiones sin juicio de valor.

### 4.1.3 Distribución de Sentimientos por Mes

En esta sección se presenta la distribución de los sentimientos expresados en los *tweets* a lo largo de los diferentes meses del período de estudio. El análisis revela cómo varían los sentimientos (positivos, negativos y neutrales) a lo largo del tiempo, proporcionando una visión temporal de las opiniones expresadas en “X” Twitter (Figura 11).

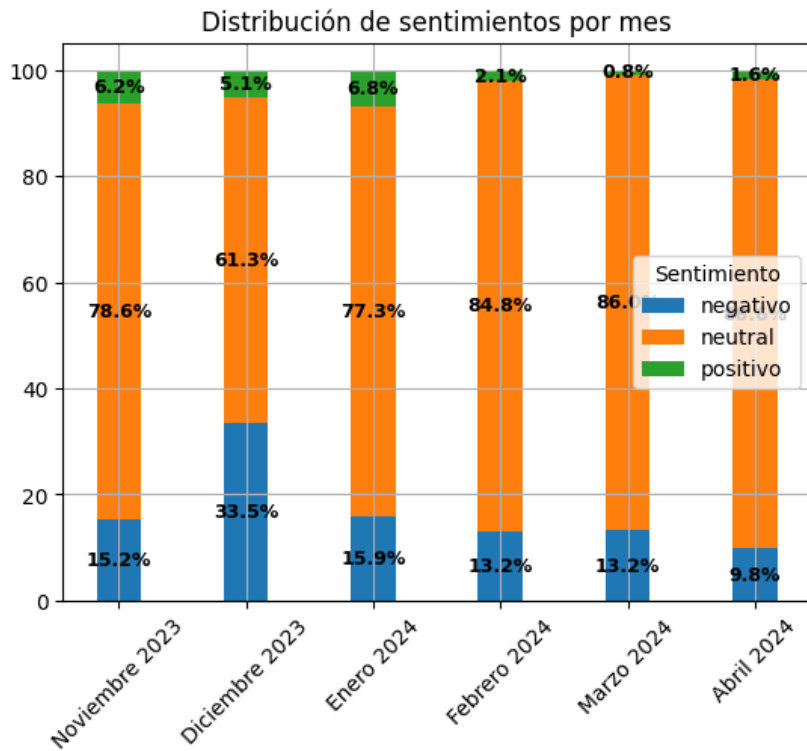


Figura 11 Grafico de barras Distribución de sentimientos por mes  
Fuente: Autor

### Tendencias de Sentimientos Negativos:

- El porcentaje de tweets negativos alcanzó su punto máximo en diciembre de 2023 con un 33.5%. Mediante la siguiente nube de palabras podemos saber los temas relacionados a estos tweets (Figura 12).
- Después de diciembre, los tweets negativos disminuyeron significativamente y se estabilizaron en torno al 13-16% en los meses siguientes.



Figura 12 Nube de Palabras de Bigramas Negativos diciembre 2023  
Fuente: Autor





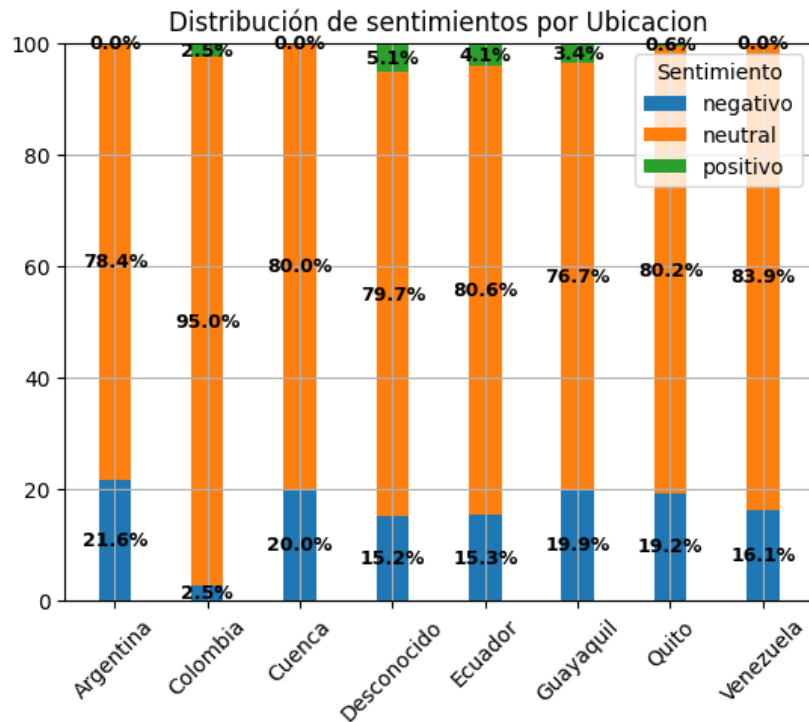


Figura 16 Gráfico de barras Distribución de sentimientos por Ubicación  
Fuente: Autor

La mayoría de los *tweets* provienen de ubicaciones desconocidas (858 *tweets*), seguidos por Guayaquil (643 *tweets*), Quito (531 *tweets*) y Ecuador (443 *tweets*).

En general, los *tweets* de todas las ubicaciones tienden a ser predominantemente neutrales indicando una actitud de espera y observación hacia la gestión del presidente Daniel Noboa.

Argentina, Cuenca, Quito y Venezuela muestran una notable ausencia de sentimientos positivos, esto sugiere una aceptación limitada o una falta de entusiasmo generalizado. Colombia es un caso peculiar con un 95% de *tweets* neutrales y un balance del 2.5% entre negativos y positivos.

Las ubicaciones específicas como Guayaquil y Quito tienen un mayor porcentaje de *tweets* negativos en comparación a la cantidad de *tweets*, lo que refleja críticas hacia políticas específicas o circunstancias locales.

## 4.2 Entrenamiento de Modelos

Se procedió a entrenar un modelo de clasificación usando los datos etiquetados. Para esto, se utilizó solo los datos etiquetados como positivos y negativos. El objetivo fue desarrollar un modelo que sea capaz de diferenciar eficazmente entre estos dos tipos de sentimientos.

Para entrenar un modelo de *machine learning*, es común utilizar una cantidad balanceada entre las clases (positivo y negativo). En este caso, contamos con 527 *tweets* negativos y solo 117 *tweets* positivos. Para abordar este desequilibrio y garantizar que el modelo cuente con los suficientes casos positivos y negativos para poder clasificar los *tweets* adecuadamente, se

propuso lo siguiente:

- **Clases Balanceadas:** Entrenar el modelo utilizando una cantidad igual de *tweets* positivos y negativos (117 ambas clases).
- **Clases Moderadamente Desbalanceadas:** Entrenar el modelo con 117 *tweets* positivos y 200 negativos.
- **Clases Totalmente Desbalanceadas:** Entrenar el modelo con todos los datos disponibles (117 positivos y 527 negativos).

Para llevar a cabo el entrenamiento del modelo para la primera propuesta (Clases Balanceadas), se extrajo aleatoriamente una porción de 117 *tweets* negativos de los 527 disponibles. Esta selección aleatoria garantiza la representatividad de los datos.

De manera similar, para el entrenamiento del modelo en la segunda propuesta (Clases Moderadamente Desbalanceadas), se extrajo aleatoriamente una porción de 200 *tweets* negativos de los 527 disponibles. Esta extracción aleatoria también buscó asegurar que los datos sean representativos para el modelo.

Una vez que se determinó la cantidad de datos a utilizar para cada caso, estos se dividieron en dos conjuntos: uno para entrenamiento (*train*) y otro para prueba (*test*). Esta división es esencial para evaluar el desempeño del modelo en datos no vistos previamente y suele realizarse en una proporción de 80% para entrenamiento y 20% para prueba.

Se evaluaron varios algoritmos de *machine learning* para determinar cuál era el más adecuado para el análisis de sentimientos. Los algoritmos considerados fueron los siguientes:

#### 4.2.1 *Naive Bayes*

El algoritmo *Naive Bayes* se basa en el teorema de *Bayes* y asume que las características son independientes entre sí. Es especialmente eficaz para problemas de clasificación de texto debido a su simplicidad y eficiencia. *Naive Bayes* es capaz de manejar grandes volúmenes de datos y es muy rápido en términos de entrenamiento y predicción (Kudryavtseva, 2024).

#### 4.2.2 *Support Vector Machine (SVM)*

*Support Vector Machine* es un algoritmo de clasificación que busca encontrar el hiperplano que mejor separa las diferentes clases en el espacio de características. *SVM* es conocido por su eficacia en espacios de alta dimensionalidad y es robusto ante datos no lineales mediante el uso de *kernels*. Es ideal para problemas de clasificación binaria como el análisis de sentimientos (Rodríguez Gómez, 2023).

#### 4.2.3 *Logistic Regression (Regresión Logística)*

La regresión logística es un método estadístico que modela la probabilidad de una clase como una función logística. A pesar de su simplicidad, es muy potente para problemas de clasificación binaria y es fácil de interpretar. Funciona bien con conjuntos de datos linealmente separables y proporciona una buena base para la clasificación (Navarro Bellido, 2023).

#### 4.2.4 *Decision Tree* (Árbol de Decisión)

Un árbol de decisión es un modelo predictivo que utiliza una estructura de árbol para tomar decisiones basadas en las características de los datos. Cada nodo del árbol representa una característica, y cada rama representa un posible valor de esa característica. Es fácil de entender y visualizar, aunque puede ser propenso al sobreajuste si no se podan adecuadamente (Von Matuschka, 2021).

#### 4.2.5 *Random Forest*

*Random Forest* es un algoritmo de *ensemble learning* que construye múltiples árboles de decisión durante el entrenamiento y emite la clase que es la moda de las clases (clasificación) o la media de las predicciones (regresión) de los árboles individuales. Este enfoque mejora la precisión y controla el sobreajuste, lo que lo hace robusto y eficaz para una variedad de tareas de clasificación (Rosenbrock, 2021).

#### 4.2.6 *XGBoost*

*XGBoost* (*Extreme Gradient Boosting*) es un algoritmo de *boosting* basado en árboles que utiliza una serie de árboles de decisión débilmente acoplados para mejorar iterativamente las predicciones. Es altamente eficiente y ofrece un rendimiento superior en comparación con otros algoritmos de aprendizaje automático debido a su capacidad para manejar datos desbalanceados y su regularización incorporada para evitar el sobreajuste (Carrión García, 2021).

### 4.3 Evaluación de Modelos

Para determinar el rendimiento de los algoritmos, es fundamental disponer de una serie de métricas que permitan evaluar de manera objetiva su eficacia en la clasificación de ejemplos proporcionados. No solo es crucial considerar las muestras clasificadas correctamente e incorrectamente, sino también aquellas que, aunque clasificadas de manera errónea, podrían haber sido etiquetadas correctamente. Existen cuatro posibles casos (Borja-Robalino, 2020):

- **Verdaderos Positivos (*True Positives* o *TP*):** Ejemplos que han sido correctamente clasificados como positivos.
- **Falsos Positivos (*False Positives* o *FP*):** Ejemplos clasificados incorrectamente como positivos, y que en realidad son negativos.
- **Verdaderos Negativos (*True Negatives* o *TN*):** Ejemplos que han sido correctamente clasificados como negativos.
- **Falsos Negativos (*False Negatives* o *FN*):** Ejemplos clasificados incorrectamente como negativos, y que en realidad son positivos.

Basándonos en estos casos y mediante la **matriz de confusión** (*Tabla 2*), podemos calcular diversas métricas para evaluar el rendimiento del modelo. Las métricas más comunes utilizadas son:

Matriz de confusión		Estimado por el modelo	
		Negativo (N)	Positivo (P)
Real	Negativo	TN	FP
	Positivo	FN	TP

Tabla 2 Matriz de confusión

Fuente: Autor

- **Exactitud (Accuracy):** Es el porcentaje de ejemplos correctamente **clasificados** sobre el total de ejemplos.

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precisión (Precision):** Es el porcentaje de ejemplos correctamente clasificados como positivos sobre el total de ejemplos clasificados como positivos.

$$Precisión = \frac{TP}{TP + FP}$$

- **Exhaustividad (Recall) o Sensibilidad (Sensitivity):** Es el porcentaje de ejemplos correctamente clasificados como positivos sobre el total de ejemplos que son **realmente** positivos.

$$Exhaustividad = \frac{TP}{TP + FN}$$

- **Valor-F (F1-score):** Es la media armónica de la precisión y la exhaustividad. Proporciona una medida equilibrada que toma en cuenta tanto la precisión como la exhaustividad, especialmente útil cuando se necesita un balance entre ambas métricas.

$$F1 - Score = 2 * \frac{Precisión * Exhaustividad}{Precisión + Exhaustividad}$$

#### 4.3.1 Clases Balanceadas

Para realizar la evaluación del desempeño de los modelos entrenados, es esencial considerar tanto los datos de entrenamiento como los datos de prueba. Esta doble evaluación permite asegurar que los modelos no tengan sobreajuste (*overfitting*), es decir, que no funcionen bien únicamente con los datos con los que fueron entrenados, sino que también sean capaces de generalizar y dar buenos resultados con nuevos datos no vistos previamente.

A través de la matriz de confusión de los datos de entrenamiento de cada modelo (*Figura 17*), observamos que la mayoría de los algoritmos se ajustan bien a los datos de entrenamiento, lo cual es esperado ya que estos datos fueron utilizados para entrenar los modelos. Sin embargo, notamos que algoritmos como *XGBoost* y *Naive Bayes* no alcanzaron un 100% de precisión en los datos de entrenamiento, lo cual puede ser indicativo de una menor tendencia al sobreajuste.

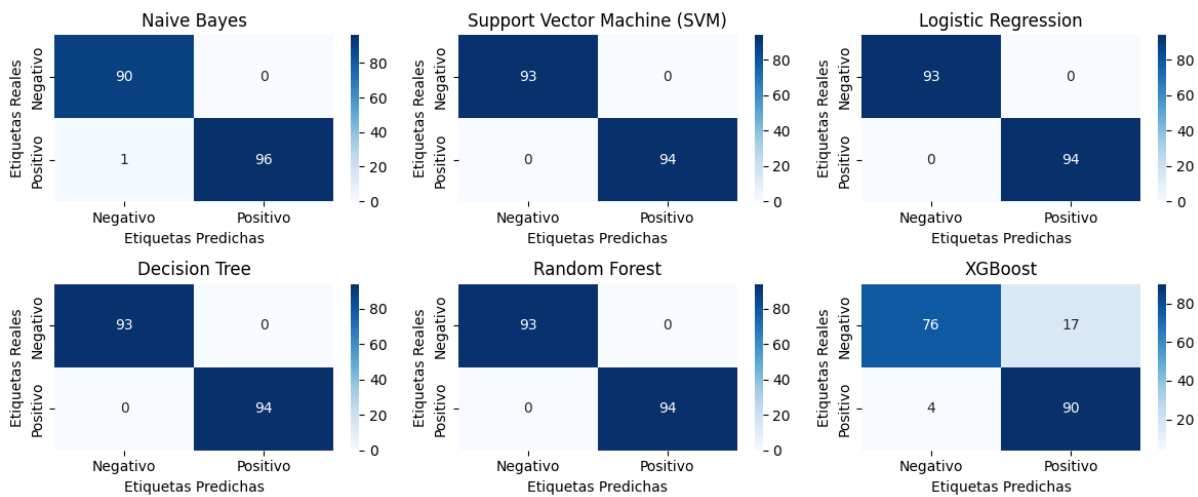


Figura 17 Matriz de Confusión de Modelos (clases balanceadas) datos de entrenamiento  
Fuente: Autor

Para asegurar que los modelos no tengan sobreajuste, evaluamos su desempeño utilizando la matriz de confusión con los datos de prueba (Figura 18). En esta evaluación, observamos que los algoritmos de *Support Vector Machine (SVM)* y *Logistic Regression* muestran una mayor precisión en la clasificación, logrando una alta tasa de *tweets* correctamente clasificados en comparación con el resto de los modelos.

Este análisis sugiere que mientras algunos modelos, como *SVM* y *Logistic Regression*, mantienen un buen desempeño en los datos de prueba, otros podrían tener sobreajuste, evidenciado por su disminución en precisión cuando se evalúan con datos no vistos previamente.

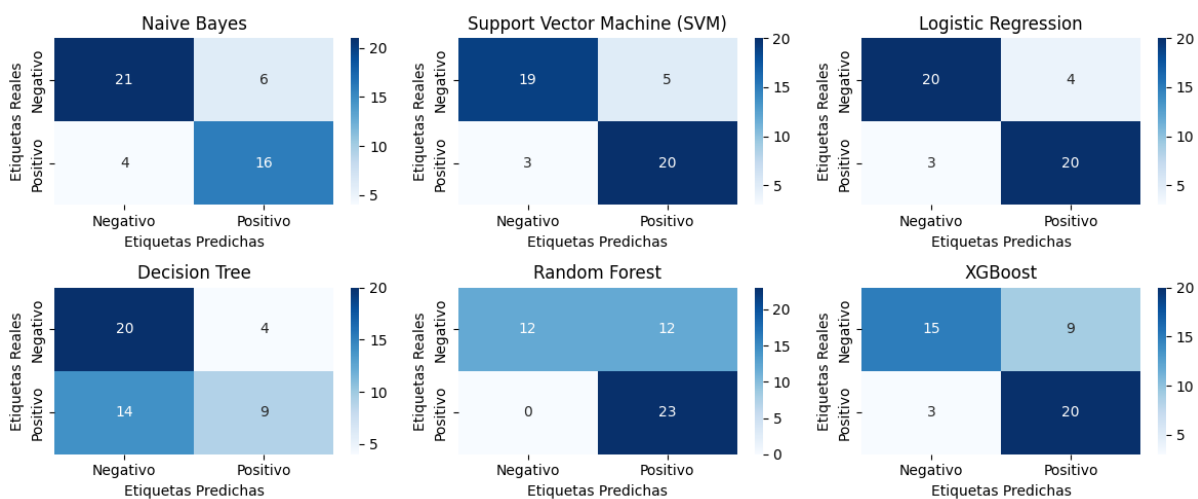


Figura 18 Matriz de Confusión de Modelos (clases balanceadas) datos de prueba  
Fuente: Autor

Los resultados obtenidos se pudieron verificar y complementar mediante las siguientes métricas de evaluación:

**Exactitud:** El modelo de *Logistic Regression* fue el mejor entre los evaluados, ya que tiene una alta exactitud tanto en los datos de entrenamiento (100%) como en los datos de prueba (85.1%). Aunque hay una ligera indicación de sobreajuste, su rendimiento en los datos de prueba fue superior al de los otros modelos. Esto indica que este modelo clasificó correctamente la mayor proporción de instancias.

Usar el modelo de *Naive Bayes* también puede ser una buena opción debido a su balance entre exactitud de entrenamiento y prueba, pero la *Logistic Regression* superó ligeramente en la exactitud en prueba, lo que la hace más confiable en general. Los otros modelos muestran signos de sobreajuste y un rendimiento inferior en los datos de prueba, lo que los hace menos adecuados para este conjunto de datos específico.

Modelo	Exactitud	
	train	test
<i>Naive Bayes</i>	99,5%	83,0%
<i>Support Vector Machine (SVM)</i>	100%	83,0%
<i>Logistic Regression</i>	100%	85,1%
<i>Decision Tree</i>	100%	61,7%
<i>Random Forest</i>	100%	74,5%
<i>XGBoost</i>	88,8%	74,5%

Tabla 3 Exactitud de los Modelos (clases balanceadas)

Fuente: Autor

**Precisión y Exhaustividad (Recall):** El modelo de *Logistic Regression* también mantuvo un buen equilibrio entre precisión (83.3%) y *recall* (87%). Una alta precisión asegura que la mayoría de las predicciones positivas sean correctas, esto quiere decir que se obtendrán menos Falsos Positivos (*FP*) y un alto *recall* garantiza que la mayoría de las instancias positivas reales sean capturadas por el modelo.

También cabe destacar que el modelo *Random Forest* tiene un *recall* de 100%, lo que significa que este modelo logró identificar correctamente todas las instancias positivas. Sin embargo, cuenta con un menor valor de precisión (65.7%) y una exactitud general más baja (74.5%). Esto sugiere que, aunque el modelo es muy eficaz para no omitir instancias positivas (altamente sensible), también clasifica incorrectamente una mayor cantidad de instancias negativas como positivas.

El modelo de *SVM* también presentó un buen rendimiento, pero la *Logistic Regression* tiene una ligera ventaja en precisión y *recall* en prueba, lo que la hace más confiable para este caso.

Modelo	Precision		Recall	
	train	test	train	test
<i>Naive Bayes</i>	100%	80,0%	99,0%	80,0%
<i>Support Vector Machine (SVM)</i>	100%	80,0%	100%	87,0%
<i>Logistic Regression</i>	100%	83,3%	100%	87,0%
<i>Decision Tree</i>	100%	69,2%	100%	39,1%
<i>Random Forest</i>	100%	65,7%	100%	100,0%

XGBoost	84,1%	69,0%	95,7%	87,0%
---------	-------	-------	-------	-------

Tabla 4 Precisión y Exhaustividad (Recall) de los Modelos (clases balanceadas)  
Fuente: Autor

**F1-Score:** El modelo de *Logistic Regression* tiene el *F1-Score* más alto en prueba (85.1%), lo que indica un excelente equilibrio entre precisión y *recall*, esto sugiere que el modelo es eficaz en términos de rendimiento general. La diferencia entre el *F1-Score* en entrenamiento y prueba es mínima, lo que indica que no hay un problema significativo de sobreajuste.

Los modelos de *Support Vector Machine (SVM)*, *Random Forest* y *Naive Bayes* también son muy destacables con un *F1-Score* alto en prueba, lo que los hace una buena opción, aunque ligeramente inferior al de *Logistic Regression*.

El modelo de *Decision Tree* muestra un sobreajuste significativo y un rendimiento muy inferior en prueba, por lo que no es recomendable.

Modelo	F1-Score	
	train	test
<i>Naive Bayes</i>	99,5%	80,0%
<i>Support Vector Machine (SVM)</i>	100%	83,3%
<i>Logistic Regression</i>	100%	85,1%
<i>Decision Tree</i>	100%	50,0%
<i>Random Forest</i>	100%	79,3%
XGBoost	89,6%	76,9%

Tabla 5 F1-Score (Recall) de los Modelos (clases balanceadas)  
Fuente: Autor

Para este caso de clases balanceadas, el mejor modelo fue *Logistic Regression*, debido a su rendimiento superior y equilibrado en todas las métricas clave (Exactitud, Precisión, *Recall* y *F1-Score*), y su buena generalización sin problemas significativos de sobreajuste. Por otro lado, el peor modelo fue *Decision Tree*, debido a su bajo rendimiento en las métricas clave y su tendencia al sobreajuste, lo que lo hace menos confiable para la clasificación en datos no vistos.

#### 4.3.2 Clases Moderadamente Desbalanceadas

Ahora veremos los resultados obtenidos para el caso desbalanceado donde se utilizó un *dataset* de 117 *tweets* positivos y 200 negativos.

A través de la matriz de confusión de los datos de entrenamiento de cada modelo (*Figura 19*), observamos que la mitad de los algoritmos se ajustan bien a los datos de entrenamiento, lo cual es esperado ya que estos datos fueron utilizados para entrenar los modelos. Sin embargo, notamos que algoritmos como *XGBoost*, *Naive Bayes* y *Logistic Regression* no alcanzaron un 100% de precisión en los datos de entrenamiento, lo cual puede ser indicativo de una menor tendencia al sobreajuste.

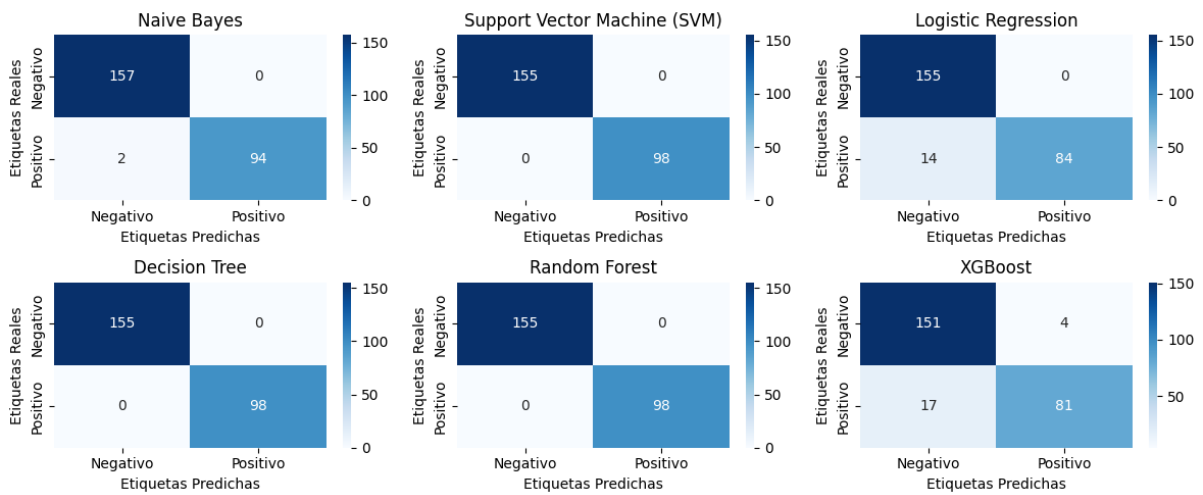


Figura 19 Matriz de Confusión de Modelos (clases moderadamente desbalanceadas) datos de entrenamiento  
Fuente: Autor

Para asegurar que los modelos no tengan sobreajuste, evaluamos su desempeño utilizando la matriz de confusión con los datos de prueba (Figura 20). En esta evaluación, observamos que los algoritmos de *Naive Bayes* y *Support Vector Machine (SVM)* muestran una mayor exactitud en la clasificación, logrando una alta tasa de *tweets* correctamente clasificados en comparación con el resto de los modelos.

Este análisis sugiere que el resto de modelos podrían tener sobreajuste a los datos de entrenamiento, evidenciado por su disminución en precisión cuando se evalúan con datos no vistos previamente.

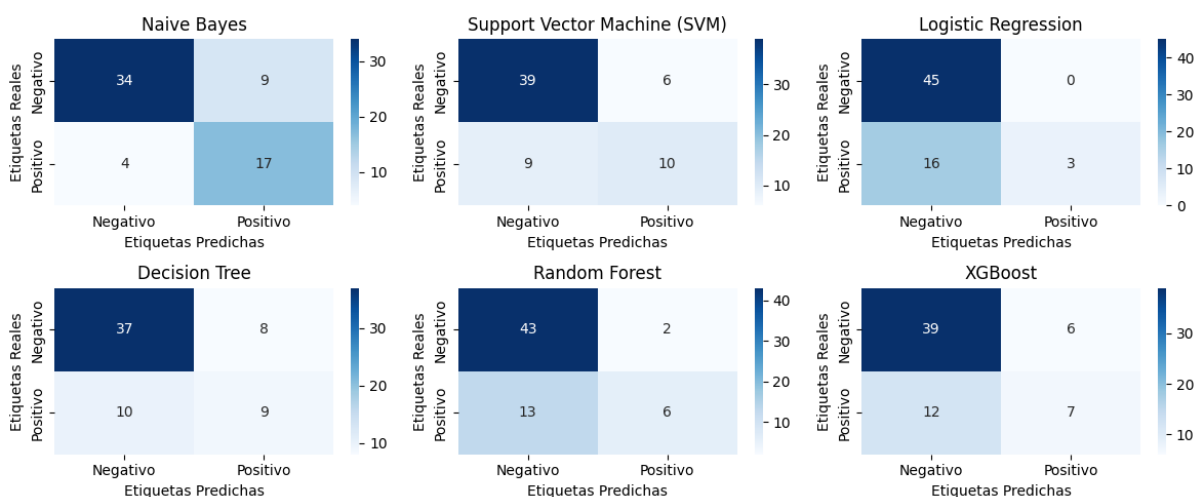


Figura 20 Matriz de Confusión de Modelos (clases moderadamente desbalanceadas) datos de prueba  
Fuente: Autor

Los resultados obtenidos se pudieron verificar y complementar mediante las siguientes métricas de evaluación:

**Exactitud:** El modelo de *Naive Bayes* muestra un buen equilibrio entre alta exactitud en la fase de prueba (79,7%) y una alta exactitud en la fase de entrenamiento (99,2%), sin caer en un sobreajuste extremo como los modelos que muestran una perfecta exactitud en la fase de entrenamiento. Este equilibrio indica que *Naive Bayes* generaliza mejor a datos no vistos. Esto indica que este modelo clasifica correctamente la mayor proporción de instancias.

El modelo de *Decision Tree*, a pesar de tener una perfecta exactitud en la fase de entrenamiento, tiene la menor exactitud en la fase de prueba (71,9%), indicando un fuerte sobreajuste. Esto lo convierte en el modelo menos confiable, ya que no generaliza bien a datos no vistos.

Modelo	Exactitud	
	train	test
<i>Naive Bayes</i>	99,2%	79,7%
<i>Support Vector Machine (SVM)</i>	100%	76,6%
<i>Logistic Regression</i>	94%	75,0%
<i>Decision Tree</i>	100%	71,9%
<i>Random Forest</i>	100%	76,6%
<i>XGBoost</i>	91,7%	71,9%

Tabla 6 Exactitud de los Modelos (clases moderadamente desbalanceadas)  
Fuente: Autor

**Precisión y Exhaustividad (Recall):** El *Naive Bayes* también mantiene un buen equilibrio entre precisión y *recall* en los datos de prueba, con un *recall* notablemente alto (81,0%) y una precisión moderada (65,4%). Esto sugiere que el modelo es capaz de capturar la mayoría de las instancias positivas correctamente, a pesar de una moderada precisión.

Es importante destacar que el modelo de *Logistic Regression* tiene una precisión de (100%), lo que significa que todos los *tweets* clasificados como positivos por el modelo son efectivamente positivos. Sin embargo, esto viene con una compensación significativa en el *recall* (15,8%). Una precisión de 100% indica que no hay Falsos Positivos (FP), pero el bajo *recall* sugiere que el modelo no está identificando correctamente muchos de los Verdaderos Positivos (VP), lo que lo hace inadecuado para una clasificación efectiva en este caso.

Modelo	Precisión		Recall	
	train	test	train	test
<i>Naive Bayes</i>	100%	65,4%	97,9%	81,0%
<i>Support Vector Machine (SVM)</i>	100%	62,5%	100%	52,6%
<i>Logistic Regression</i>	100%	100,0%	86%	15,8%
<i>Decision Tree</i>	100%	52,9%	100%	47,4%
<i>Random Forest</i>	100%	75,0%	100%	31,6%
<i>XGBoost</i>	95,3%	53,8%	82,7%	36,8%

Tabla 7 Precisión y Exhaustividad (Recall) de los Modelos (clases moderadamente desbalanceadas)  
Fuente: Autor

**F1-Score:** El *Naive Bayes* tuvo el *F1-Score* más alto en los datos de prueba (72,3%), lo que indica que mantiene un buen equilibrio entre precisión y *recall*. Esto sugiere que es el modelo

más efectivo para manejar la clasificación de datos nuevos sin caer en problemas de sobreajuste significativo.

Por otro lado, el modelo de *Logistic Regression* presenta el *F1-Score* más bajo en los datos de prueba (27,3%), indicando un rendimiento pobre en la clasificación de nuevos datos y una incapacidad significativa para generalizar desde los datos de entrenamiento.

Modelo	F1-Score	
	train	test
Naive Bayes	98,9%	72,3%
Support Vector Machine (SVM)	100%	57,1%
Logistic Regression	92%	27,3%
Decision Tree	100%	50,0%
Random Forest	100%	44,4%
XGBoost	88,5%	43,8%

Tabla 8 *F1-Score (Recall)* de los Modelos (clases moderadamente desbalanceadas)  
Fuente: Autor

Para este caso de clases moderadamente desbalanceadas, el modelo de *Naive Bayes* fue el más eficaz debido a su alto *F1-Score* y buen balance entre precisión y *recall* en los datos de prueba. Por otro lado, *Logistic Regression* fue el peor modelo, con un rendimiento significativamente inferior en *recall* y *F1-Score*, lo que lo hace inadecuado para este caso.

### 4.3.3 Clases Totalmente Desbalanceadas

Por último, veremos los resultados obtenidos para el caso desbalanceado con el total de los datos (117 *tweets* positivos y 527 negativos).

A través de la matriz de confusión de los datos de entrenamiento de cada modelo (*Figura 21*), observamos que los algoritmos de *Decisión Tree* y *Random Forest* se ajustan bien a los datos de entrenamiento. Sin embargo, notamos que algoritmos como *Naive Bayes*, *Support Vector Machine (SVM)* y *XGBoost* no alcanzaron un 100% de precisión en los datos de entrenamiento, lo cual puede ser indicativo de una menor tendencia al sobreajuste.

Por otro lado, el modelo de *Logistic Regression* falla completamente en la identificación de positivos.

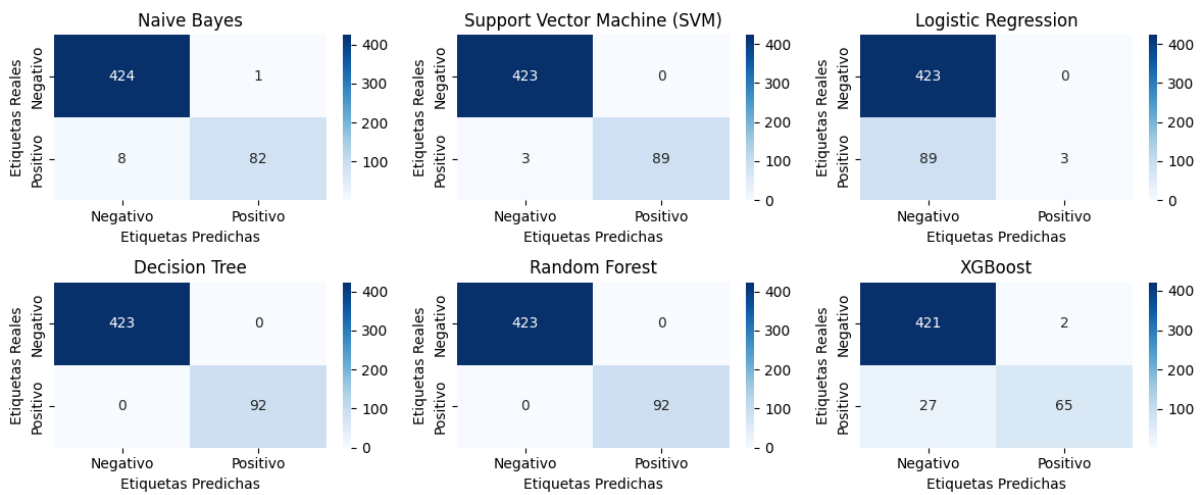


Figura 21 Matriz de Confusión de Modelos (clases totalmente desbalanceadas) datos de entrenamiento  
Fuente: Autor

Para asegurar que los modelos no tengan sobreajuste, evaluamos su desempeño utilizando la matriz de confusión con los datos de prueba (Figura 22). En esta evaluación, observamos que los algoritmos tienen una mejor precisión para clasificar los *tweets* negativos, pero para el caso de los positivos se podría decir que el *Naive Bayes* sobresale un poco del resto. Aunque no es perfecto, tiene la menor cantidad de Falsos Negativos (FN) y Falsos Positivos (FP) en comparación con otros modelos en los datos de prueba.

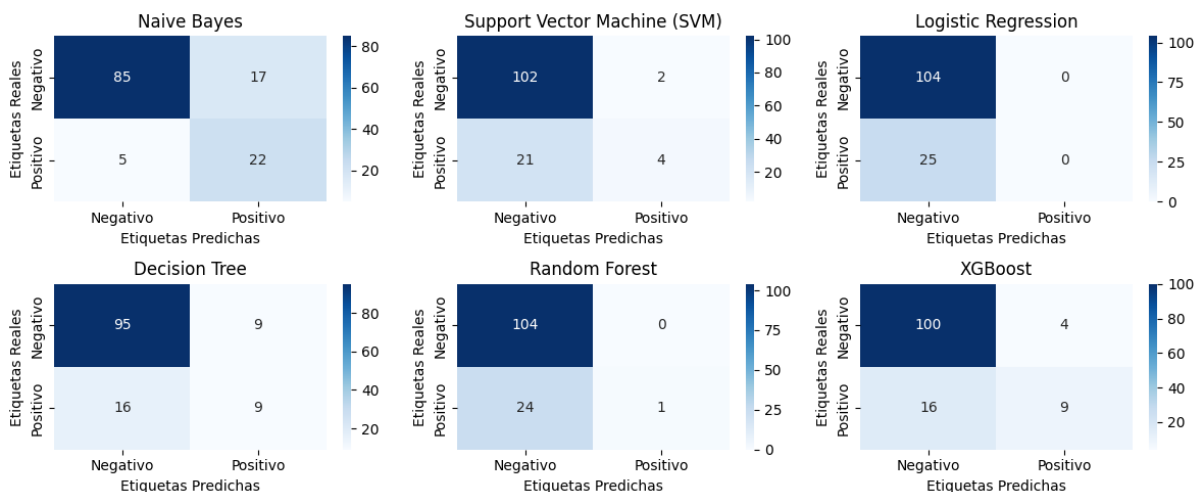


Figura 22 Matriz de Confusión de Modelos (clases totalmente desbalanceadas) datos de prueba  
Fuente: Autor

Los resultados obtenidos se pudieron verificar y complementar mediante las siguientes métricas de evaluación:

**Exactitud:** El modelo de *XGBoost* tuvo la mayor exactitud en el conjunto de prueba (84,5%), lo que indica que generaliza mejor a datos no vistos en comparación con los otros modelos.

Además, su exactitud en el conjunto de entrenamiento es alta pero no perfecta, lo que sugiere que evita el sobreajuste mejor que modelos como *Decision Tree* y *Random Forest*.

Por otro lado, el modelo de *Logistic Regression* tuvo la menor exactitud tanto en el conjunto de entrenamiento (83%) como en el de prueba (80,6%), lo que indica que no está completamente ajustado y no generaliza tan bien como los otros modelos.

Modelo	Exactitud	
	train	test
<i>Naive Bayes</i>	98,3%	82,9%
<i>Support Vector Machine (SVM)</i>	99%	82,2%
<i>Logistic Regression</i>	83%	80,6%
<i>Decision Tree</i>	100%	80,6%
<i>Random Forest</i>	100%	81,4%
<i>XGBoost</i>	94,4%	84,5%

Tabla 9 Exactitud de los Modelos (clases totalmente desbalanceadas)  
Fuente: Autor

**Precisión y Exhaustividad (*Recall*):** El *Naive Bayes* ofrece un buen equilibrio entre precisión y *recall* en el conjunto de prueba. Aunque la precisión (56.4%) en el conjunto de prueba es más baja que otros modelos, su *recall* (81.5%) es significativamente más alto, lo que es crucial en problemas con clases desbalanceadas, ya que indica que el modelo está capturando más Verdaderos Positivos (VP).

Es importante destacar que el modelo de *Random Forest* tuvo una precisión de 100%, lo que significa que todos los *tweets* clasificados como positivos por el modelo son efectivamente positivos. Sin embargo, esto viene con una compensación significativa en el *recall* (4%). Una precisión de 100% indica que no hay Falsos Positivos (FP), pero el bajo *recall* sugiere que el modelo no está identificando correctamente muchos de los Verdaderos Positivos (VP), lo que lo hace inadecuado para una clasificación efectiva en este caso.

El modelo de *Logistic Regression* tiene un rendimiento desastroso en el conjunto de prueba con una precisión y *recall* del 0,0%. Esto indica que el modelo no está haciendo ninguna predicción positiva correcta, siendo ineficaz para este problema específico.

Modelo	Precision		Recall	
	train	test	train	test
<i>Naive Bayes</i>	99%	56,4%	91,1%	81,5%
<i>Support Vector Machine (SVM)</i>	100%	66,7%	97%	16,0%
<i>Logistic Regression</i>	100%	0,0%	3%	0,0%
<i>Decision Tree</i>	100%	50,0%	100%	36,0%
<i>Random Forest</i>	100%	100,0%	100%	4,0%
<i>XGBoost</i>	97,0%	69,2%	70,7%	36,0%

Tabla 10 Precisión y Exhaustividad (*Recall*) de los Modelos (clases totalmente desbalanceadas)  
Fuente: Autor

***F1-Score*:** El *Naive Bayes* muestra el *F1-Score* más alto en el conjunto de prueba (66,7%),

indicando que tiene un mejor equilibrio entre precisión y *recall* en datos no vistos. Esto es crucial en problemas con clases desbalanceadas, ya que captura mejor los verdaderos positivos mientras mantiene un buen rendimiento general.

El modelo de *Logistic Regression* tiene un *F1-Score* extremadamente bajo en ambos conjuntos de datos (6% en entrenamiento y 0,0% en prueba), lo que indica que no está haciendo predicciones correctas y es ineficaz para este problema específico.

Modelo	F1-Score	
	train	test
<i>Naive Bayes</i>	94,8%	66,7%
<i>Support Vector Machine (SVM)</i>	98%	25,8%
<i>Logistic Regression</i>	6%	0,0%
<i>Decision Tree</i>	100%	41,9%
<i>Random Forest</i>	100%	7,7%
<i>XGBoost</i>	81,8%	47,4%

Tabla 11 *F1-Score (Recall) de los Modelos (clases totalmente desbalanceadas)*

Fuente: Autor

Para este caso de clases totalmente desbalanceadas, el modelo de *Naive Bayes* parece ser el más adecuado, ya que presenta una exactitud (82,9%) y el *F1-Score* más alto (66,7%). No obstante, estos resultados indican que el modelo tiene un rendimiento aceptable, aún requiere mejoras para aumentar tanto su precisión como su *recall*. Por otro lado, el modelo de *Logistic Regression* muestra una exactitud de 80,6% pero un *F1-Score* de 0,0% en el conjunto de prueba, indicando que no puede manejar el desbalance de clases en absoluto y no es efectivo en este escenario.

## 5 CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES

### 5.1 Conclusiones

1. El *web scraping* demostró ser un método efectivo y económico para obtener *tweets* relevantes. Esto resultó en la obtención de una muestra representativa de 3,177 *tweets*, proporcionando una base sólida para el análisis de sentimientos y la evaluación del nivel de aceptación del presidente Daniel Noboa.
2. La clasificación automática de los *tweets* mediante las librerías *pysentiment* y *textblob* mostró resultados subóptimos e inconsistentes, ya que no reflejaban con precisión el contexto y la semántica de los *tweets* en español (*Figura 42, 43*).
3. La clasificación manual de los sentimientos en los *tweets* proporcionó una categorización más precisa y adecuada, ya que permite capturar de manera más fiel las emociones y percepciones reales expresadas en el contenido.
4. La cuantificación y categorización de las respuestas hacia el presidente Daniel Noboa en “X” *Twitter* revelaron una predominancia de *tweets* neutrales (79.7%), indicando que la mayoría de las menciones no expresan un sentimiento fuerte, ya sea positivo o negativo. Sin embargo, el 16.6% de los *tweets* negativos refleja una presencia considerable de críticas y desaprobación. En contraste, solo el 3.7% de los *tweets* son positivos, lo que indica un apoyo explícito relativamente escaso hacia el presidente.
5. En la evaluación de los modelos entrenados en tres escenarios distintos (clases balanceadas, clases moderadamente desbalanceadas y clases totalmente desbalanceadas), se concluyó que *Naive Bayes* se destacó como el mejor modelo debido a su predominancia en los tres escenarios propuestos. Aunque en el primer escenario no fue el mejor modelo, se situó en tercer lugar con solo una diferencia del 3% en exactitud y del 5% en *recall* en comparación con *Logistic Regression*, lo que lo hace una opción fiable para manejar la variabilidad en los datos.

### 5.2 Recomendaciones

1. Ampliar el período de estudio podría proporcionar una visión más completa y detallada de la evolución de la percepción pública hacia el presidente Daniel Noboa. Un análisis a largo plazo permitiría identificar tendencias y cambios en los sentimientos de la población, ofreciendo una evaluación más precisa del impacto de sus políticas y acciones a lo largo del tiempo.
2. Incorporar otras redes sociales como *Facebook* e *Instagram* podría ofrecer una perspectiva más amplia y diversa sobre la aceptación del presidente. Estas plataformas tienen diferentes tipos de usuarios y formas de interacción, lo que enriquecería el análisis y permitiría captar una variedad más amplia de opiniones y sentimientos.
3. Incluir los *retweets* y las respuestas en futuros análisis para captar una imagen más

completa de la interacción y el impacto de los *tweets*, permitiendo entender mejor la dinámica de la conversación y la influencia de los usuarios.

4. Recopilar una cantidad mayor de datos para el análisis. Un volumen de datos más grande mejora la precisión de los modelos de análisis de sentimientos. Adicionalmente, contar con más volumen de datos nos permitiría probar técnicas de *Deep learning* más robustas como redes neuronales.
5. Se recomienda explorar *transformer models* tales como *BERT* o *GPT* como una alternativa para para clasificar los *tweets* dado que son modelos preentrenados con una cantidad masiva de textos.

## REFERENCIAS BIBLIOGRÁFICAS

- Badiola Ramos, J. (2019). *¿ Podemos comerciar Bitcoin usando análisis de sentimiento sobre Twitter?*
- Bárcenas Reséndiz, I. A. (2023). *Estrategias de mercadotecnia para la empresa GYBO mediante análisis de opinión pública.*
- Borja-Robalino, R. M.-G. (2020). Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning. *Revista Ibérica de Sistemas e Tecnologías de Informação*, 184-196.
- Butron, J. B. (2023). *Deteccion de Ideas Suicidas en Twitter.*
- Carrión García, J. A. (2021). Análisis de Sentimientos de Tweets en Español Basado en Técnicas Aprendizaje Supervisado.
- Cebral-Loureda, M. &.-P. (2021). *Los inicios de la pandemia de COVID19 en Twitter. Análisis computacional de la conversación pública en lengua española.*
- Contreras Ruiz, M. N. (2022). *Análisis del contenido del medio digital Aldia y su influencia en la opinión pública, Babahoyo 2021.*
- Cortés Díaz, N. S. (2021). *Uso del análisis de sentimientos para medir el posicionamiento de una marca a través de la información extraída de la plataforma Twitter.*
- González, G. F. (2019). *Comunicación política y redes sociales.*
- Herrera, S. &. (2023). *Daniel Noboa, el joven presidente del Ecuador.*
- Hurtado, O. (2019). *El poder político en el Ecuador.*
- Kudryavtseva, L. (2024). Explorando algoritmos de clasificación de textos: estudio comparativo en análisis de sentimientos y clasificación temática.
- Llivicota Buñay, W. A. (2021). *Análisis de sentimientos en comunidades digitales utilizando técnicas de big data para Determinar patrones de comportamiento orientado a fenómenos sociales.*
- López Hernández, J. (2022). *Análisis y tipificación de errores lingüísticos para una propuesta de mejora de informes médicos en español.*
- Lozano-Blasco, R. M.-A.-L. (2023). *Redes sociales y su influencia en los jóvenes y niños: Análisis en Instagram, Twitter y YouTube.*
- Maldonado Ramones, E. S. (2022). *Análisis de sentimientos en la red social Twitter mediante el procesamiento de lenguaje natural.*

- Manosalvas Porras, L. V. (2022). *Estudio de mercado utilizando la red social Twitter, para Hoteles Decameron*. PUCE-Quito.
- Manosalvas, L. (2022). *Estudio De Mercado Utilizando La Red Social Twitter, Para Hoteles Decameron*. PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR.
- Molano, E. C. (2021). *La influencia de las redes sociales en el ámbito de la comunicación*. Educación universitaria en ecosistemas de aprendizaje.
- Montañés Salas, R. M. (2020). *Análisis de sentimiento con aprendizaje profundo en un entorno de Big Data*.
- Moreno, F. Y. (2020). *Técnicas para la Clasificación de Sentimientos en Redes Sociales como Apoyo en el Marketing Digital*. Revista Ibérica de Sistemas e Tecnologías de Información.
- Navarro Bellido, J. (2023). *Modelos de aprendizaje automático en análisis de sentimiento: comparativa de rendimiento*. Universitat Politècnica de València.
- Nazareno Gómez, M. D. (2021). *Análisis de herramientas de extracción de comentarios de la red social twitter y creación de un corpus no relacional para el posterior uso en el procesamiento de lenguaje natural (PLN), y minería de opiniones (MO)*.
- Néstor Camilo Beltrán, E. C. (2021). *Procesamiento del lenguaje natural (PLN) -GPT-3, y su aplicación en la Ingeniería de Software*. Bogotá-Colombia.
- Ordóñez, C. A. (2021). *Minería de opinión: un análisis en tiempo real de tweets para Zacatecas*.
- Ortiz Medina, A. C. (2021). *La influencia de las redes sociales en la campaña política de Trump en las elecciones del 2016*.
- Pauli, P. A. (2019). *Análisis de sentimiento: comparación de algoritmos predictivos y métodos utilizando un lexicon español*.
- Pérez-Suasnavas, A. L. (2020). *Beneficios del uso de técnicas de minería de datos para extraer y analizar datos de twitter aplicados en la educación superior: una revisión sistemática de la literatura*.
- Quintero, F. A. (2021). *Comportamiento del consumidor de destinos turísticos de Ecuador*.
- Redondo Romano, S. (2022). *Influencia de las redes sociales en la construcción de opinión pública de los jóvenes de la 5. ta cohorte de Comunicación Social-Periodismo de la UdeA Seccional Urabá*.
- Rodríguez Gómez, D. (2023). *Análisis de datos y Machine Learning*.

- Rodríguez, J. G. (2021). *La interculturalidad en el Ecuador: un análisis crítico de sus fundamentos, aportes y desafíos en la educación.*
- Rojas Beltrán, M. F. (2023). *Impacto en la opinión pública desde la presidencia de Colombia por medio de la red social Twitter.*
- Rosenbrock, G. T. (2021). Técnicas de análisis de sentimientos aplicadas a la valoración de opiniones en el lenguaje español.
- Sánchez Lorenzo, J. (2019). *Opinión pública libre, opinión publicada. La libertad de comunicación pública y los medios de comunicación.*
- Segura, J. A. (2023). *Twitter y el derecho de libertad de expresión ya la identidad: Un análisis jurídico.* Magazine de las Ciencias: Revista de Investigación e Innovación.
- Steiner Corrêa, A. F. (2019). *Diseño, construcción y validación de LOGOS: una herramienta basada en el análisis de sentimientos multilingüe como apoyo a la toma de decisiones de marketing.*
- Tupiza, M. (2022). *Análisis de sentimientos en redes sociales con aplicaciones en política ecuatoriana.* PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR.
- Valero Herrero, S. (2023). *Panorama general de las redes sociales en el marketing de las organizaciones de empresa.*
- Von Matuschka, C. (2021). Análisis de sentimientos usando la red social Twitter; ¿qué sintieron los turistas que volaron en 2020 con seleccionadas aerolíneas sudamericanas?. *Revista de Turismo e Identidad*, 55-71.
- Zuniga, D. M. (2023). *Influencia de las redes sociales en los nuevos votantes y como se puede utilizar para fomentar la participación.*

## ANEXOS

### Encuentre el acceso adecuado para usted

Gratis	Básico	Pro	Empresa
<p>Para casos de uso de solo escritura y para probar la API X</p> <ul style="list-style-type: none"> <li>• Califique el acceso limitado a los puntos finales de publicación de publicaciones y carga de medios v2</li> <li>• 1500 publicaciones por mes: límite de publicaciones a nivel de aplicación</li> <li>• 1 ID de aplicación</li> <li>• Inicia sesión con X</li> <li>• Gratis</li> </ul> <p><b>Empezar</b></p>	<p>Para aficionados o prototipos.</p> <ul style="list-style-type: none"> <li>• Califique el acceso limitado al conjunto de puntos finales v2</li> <li>• 3000 publicaciones por mes: límite de publicaciones a nivel de usuario</li> <li>• 50.000 publicaciones por mes: límite de publicaciones a nivel de aplicación</li> <li>• 10,000 publicaciones por mes: límite de tasa de límite de lectura</li> <li>• 2 ID de aplicación</li> <li>• Inicia sesión con X</li> <li>• \$100 por mes</li> </ul> <p><b>Suscríbese ahora</b></p>	<p>Para startups que amplían su negocio</p> <ul style="list-style-type: none"> <li>• Acceso con velocidad limitada al conjunto de terminales v2, incluida la búsqueda y el flujo filtrado</li> <li>• 1.000.000 de publicaciones por mes: OBTÉN a nivel de aplicación</li> <li>• 300.000 publicaciones por mes: límite de publicaciones a nivel de aplicación</li> <li>• 3 ID de aplicaciones</li> <li>• Inicia sesión con X</li> <li>• \$5,000 por mes</li> </ul> <p><b>Suscríbese ahora</b></p>	<p>Para empresas y proyectos comerciales a escala.</p> <ul style="list-style-type: none"> <li>• Acceso a nivel comercial que satisface sus necesidades específicas y las de sus clientes</li> <li>• Servicios administrados por un equipo de cuentas dedicado</li> <li>• Transmisiones completas: repetición, métricas de participación, reposición y más funciones</li> <li>• Niveles de suscripción mensual</li> </ul> <p><b>Aplica ya</b></p>

Figura 23 Planes de la API de Twitter

Fuente: Captura de pantalla de la página 'X Developer Platform', <https://developer.twitter.com/en/products/twitter-api>, tomada el 9 de mayo de 2024

The screenshot shows the APIFY Billing interface. On the left is a navigation sidebar with options: Home, Store, Actors, Schedules, Storage, Proxy, Settings, and Billing. The main content area is titled 'Billing' and has tabs for 'Current period', 'Historical usage', and 'Subscription'. Under 'Current subscription', there is an 'Upgrade' button and a section for the 'Free plan'. The plan details are as follows:

- Free platform usage: \$5
- Actor compute units (CU): \$0.400 / CU
- Actor RAM: 8 GB
- Max concurrent Actor runs: 25
- Datacenter proxies: 5
- Residential proxies: \$13.00 / GB
- Proxy SERPs: \$3.00 / 1,000 SERPs
- Support level: Discord

At the bottom left, there is a 'Memory' usage indicator showing 0 MB / 8 GB and a 'Free usage' indicator showing \$2.54 / \$5.00. A note states 'Limit resets on May 3'.

Figura 24 Plataforma APIFY

Fuente: Captura de pantalla de la página 'APIFY', <https://console.apify.com/billing/subscription>, tomada el 9 de mayo de 2024

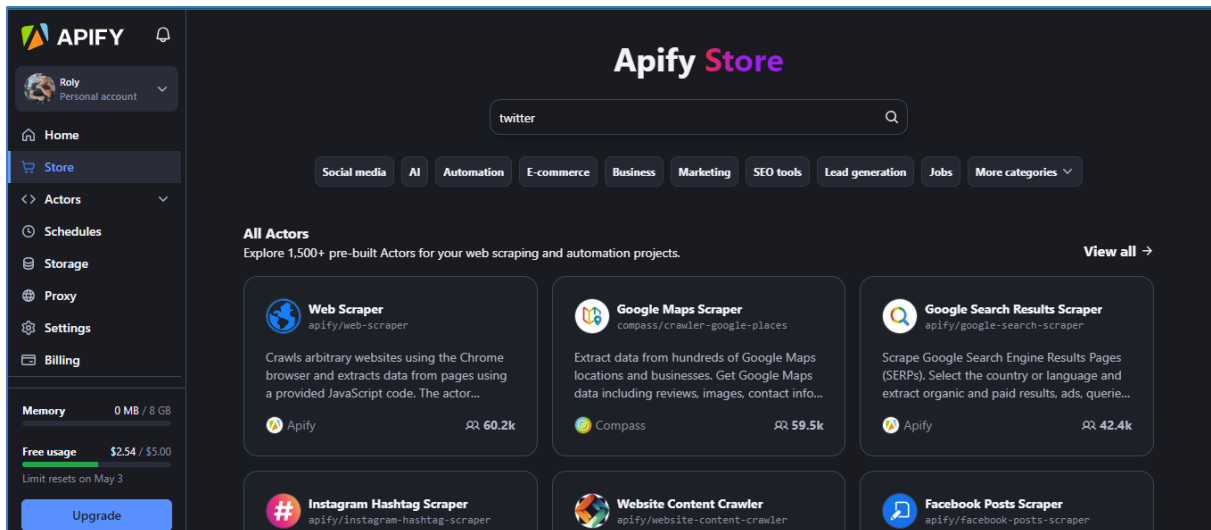


Figura 25 Sección Store de la plataforma APIFY

Fuente: Captura de pantalla de la página 'APIFY', <https://console.apify.com/store>, tomada el 9 de mayo de 2024

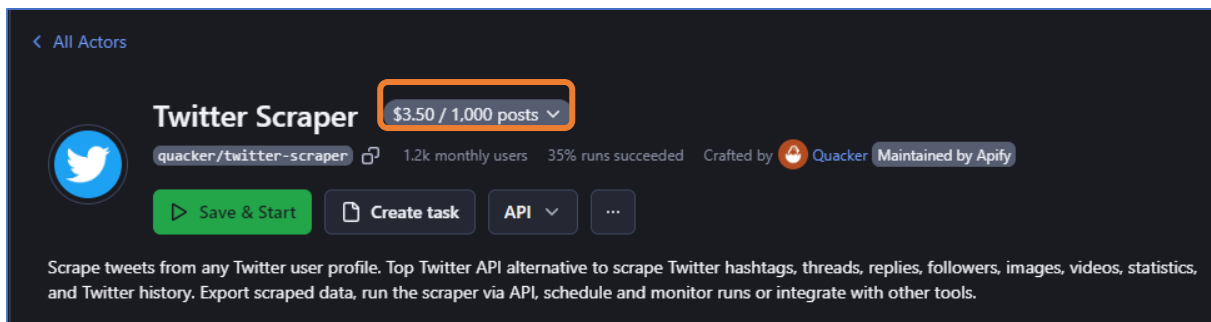


Figura 26 Twitter Scraper 3.50/1000

Fuente: Captura de pantalla de la página 'APIFY', <https://console.apify.com/actors/u6ppkMWax2E2MpEuF/input>, tomada el 9 de mayo de 2024

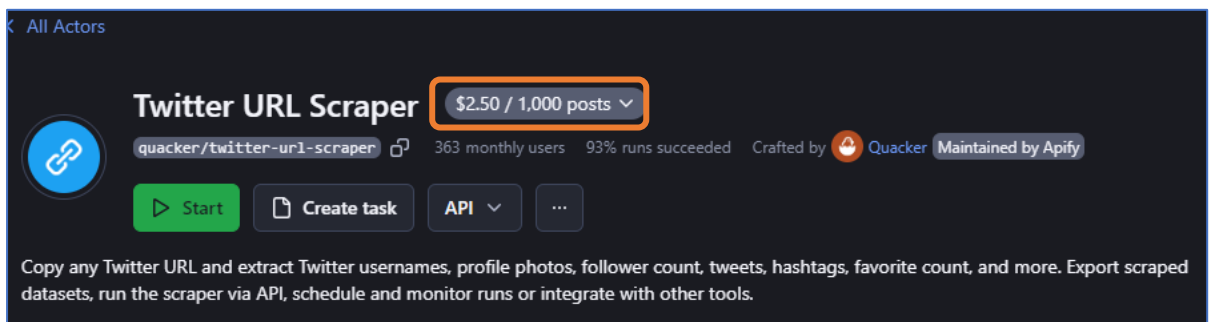


Figura 27 Twitter URL Scraper 2.50/1000

Fuente: Captura de pantalla de la página 'APIFY', <https://console.apify.com/actors/KVJr35xjTw2XyvMeK/input>, tomada el 9 de mayo de 2024

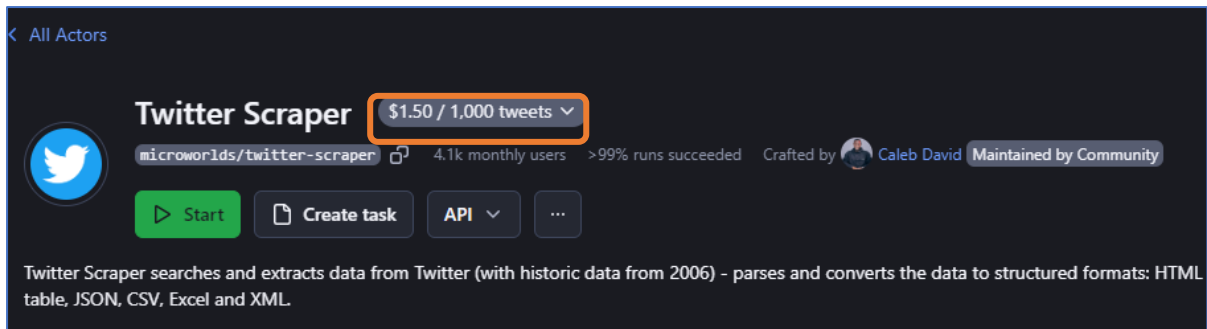


Figura 28 Twitter Scraper 1.50/1000

Fuente: Captura de pantalla de la página 'APIFY',  
<https://console.apify.com/actors/heLL6fUofdPgRXZie/input>, tomada el 9 de mayo de 2024

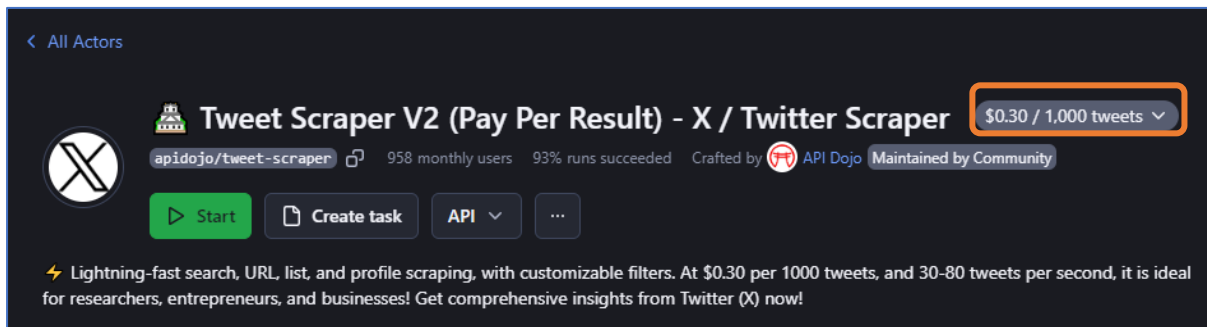


Figura 29 Twitter Scraper V2 0.30/1000

Fuente: Captura de pantalla de la página 'APIFY',  
<https://console.apify.com/actors/61RPP7dywgiy0JPD0/input>, tomada el 9 de mayo de 2024

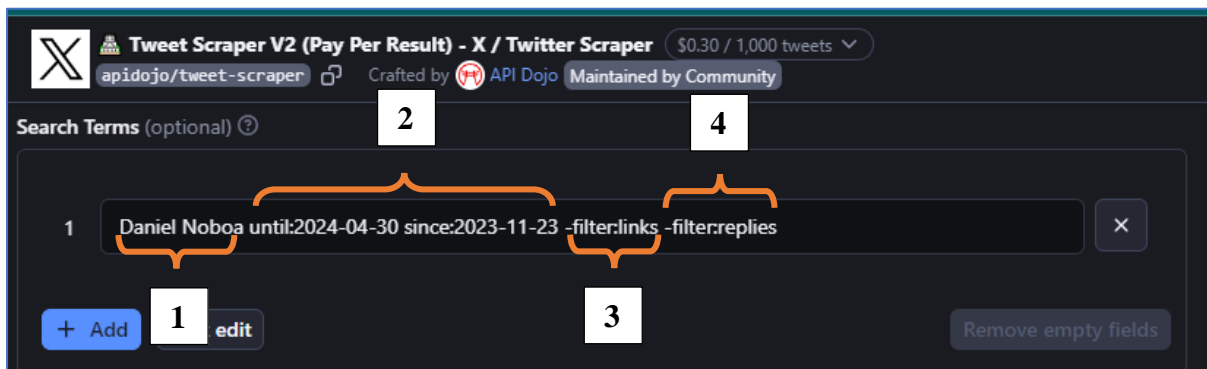


Figura 30 Configuración términos de búsqueda

Fuente: Captura de pantalla de la página 'APIFY',  
<https://console.apify.com/actors/61RPP7dywgiy0JPD0/input>, tomada el 9 de mayo de 2024

**Tweet Scraper V2 (Pay Per Result) - X / Twitter Scraper - Run**

✓ Succeeded Success! Everything went according to plan.

RESULTS	REQUESTS	PRICE	STARTED	DURATION
3177	0 of 0 handled	\$0.0063	2024-05-02 16:46	31 s

What next? Hide

Change input & start a new run | Schedule your actor | Integrate with other services | Show similar Actors

Output 3177 | Log | Input | Storage | Live view | Integrations 0

Summary Data (check Storage tab for all fields) | All fields | Table | JSON | Preview in new tab

Showing only 12 of 99 fields. Show all

#	URL	Twitter URL	ID	Text
1	https://x.com/santiagopenaope/status/1785080202474774750	https://twitter.com/santiagopenaope/status/1785080202474774750	1785080202474774750	A críte Noboa el pelc

Export 3177 results | Go to Actor | Integrate | Delete | Resurrect

Figura 31 Resultado de la búsqueda

Fuente: Captura de pantalla de la página 'APIFY',  
<https://console.apify.com/actors/61RPP7dywgiy0JPD0/runs/91WUn1ATpySsv39qV#output>, tomada el 9 de mayo de 2024

	text	createdAt	author/location
0	Todos los que pasan a los ministerios de Danie...	Sat Nov 25 23:57:01 +0000 2023	Guayaquil
1	Tienen que cuidar al presidente Daniel Noboa L...	Sat Nov 25 23:54:03 +0000 2023	ecuador
2	ÚLTIMA HORA: En su primer día como presiden...	Sat Nov 25 23:52:21 +0000 2023	NaN
3	Si Daniel Noboa y Felix Sánchez no obtienen bu...	Sat Nov 25 23:22:00 +0000 2023	Guayaquil Solid State Society
4	Daniel Noboa @DanielNoboaOk debería inmediatam...	Sat Nov 25 23:21:37 +0000 2023	NaN
...	...	...	...
676	Señor Presidente Daniel Noboa, cuándo se ponen...	Mon Apr 29 11:25:48 +0000 2024	NaN
677	Abogados mexicanos presentaron este domingo 28...	Mon Apr 29 04:53:35 +0000 2024	Alemania..
678	Y si fueron al festival en Guayaquil el año pa...	Mon Apr 29 02:12:05 +0000 2024	NaN
679	Abogados mexicanos presentan ante la Fiscal...	Mon Apr 29 01:23:06 +0000 2024	NaN
680	"Abogados presentan ante FGR denuncia penal en...	Mon Apr 29 00:38:23 +0000 2024	Coyoacán, Distrito Federal

3177 rows x 3 columns

Figura 32 Tabla de datos filtrada

Fuente: Autor

```

# Función para limpiar los tweets
def limpiar_tweet(tweet):
    # Convertir el texto a minúsculas
    tweet = tweet.lower()
    # Eliminar menciones a usuarios (@usuario)
    tweet = re.sub(r'@[a-zA-Z0-9_]+', '', tweet)
    # Eliminar tildes
    tweet = unidecode(tweet)
    # Eliminar caracteres especiales y signos de puntuación
    tweet = re.sub(r'^a-zA-Z\s', '', tweet)
    # Tokenización
    tokens = word_tokenize(tweet)
    # Eliminar palabras vacías (stop words)
    stop_words = set(stopwords.words('spanish'))
    tokens = [word for word in tokens if word not in stop_words]
    # Lematización
    tokens = [lemmatizer.lemmatize(word) for word in tokens]
    # Unir tokens en un solo texto nuevamente
    tweet_limpio = ' '.join(tokens)
    return tweet_limpio

```

Figura 33 Función para limpiar los tweets  
Fuente: Autor

```

from datetime import datetime
# Diccionario con los nombres de los meses en español
meses_espanol = {
    1: 'Enero',
    2: 'Febrero',
    3: 'Marzo',
    4: 'Abril',
    5: 'Mayo',
    6: 'Junio',
    7: 'Julio',
    8: 'Agosto',
    9: 'Septiembre',
    10: 'Octubre',
    11: 'Noviembre',
    12: 'Diciembre'
}

# Función para convertir la fecha al nombre del mes en español
def convertir_mes.fecha_str):
    fecha = datetime.strptime(fecha_str, '%a %b %d %H:%M:%S %z %Y')
    nombre_mes_espanol = meses_espanol[fecha.month] + fecha.strftime(' %Y')
    return nombre_mes_espanol

```

Figura 34 Función para dar un nuevo formato a la fecha  
Fuente: Autor

	tweet	mes	ubicacion
0	pasan ministerios reciben orden suspender cuen...	Noviembre 2023	Guayaquil
1	cuidar quieren asesinar vicepresidenta asuma c...	Noviembre 2023	Ecuador
2	ultima hora primer funciones derogo ley dro...	Noviembre 2023	Desconocido
3	felix sanchez obtienen buenos resultados junio...	Noviembre 2023	Guayaquil
4	deberia inmediatamente convocar asamblea const...	Noviembre 2023	Desconocido
...	...	...	...
3165	senor cuando ponen devoluciones dinero sri	Abril 2024	Desconocido
3166	abogados mexicano presentaron domingo abril ...	Abril 2024	alemania
3167	festival guayaquil ano pasado parque samanes ...	Abril 2024	Desconocido
3168	abogados mexicano presentan fiscalia general ...	Abril 2024	Desconocido
3169	abogados presentan fgr denuncia penal res...	Abril 2024	coyoacan distrito federal

3170 rows × 3 columns

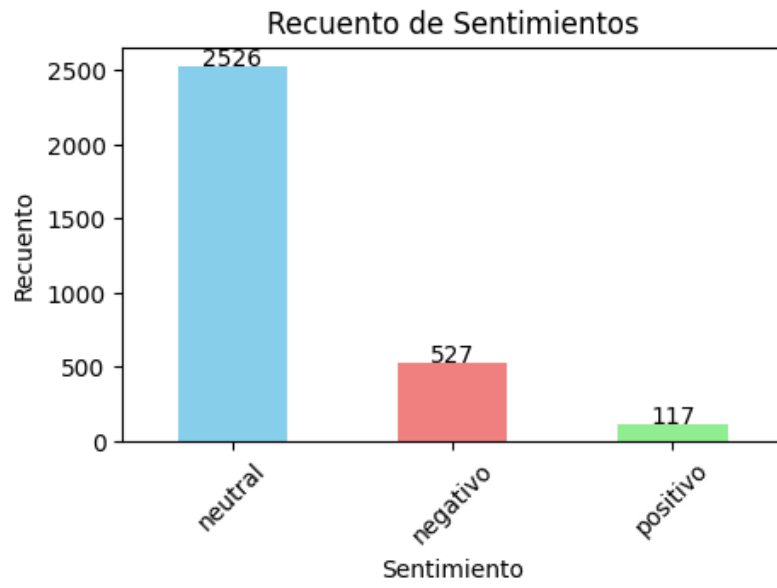
Figura 35 Tabla de Datos Después de la Limpieza  
Fuente: Autor

index	text	tweet	sentimiento
845	Entonces, Daniel Noboa, ¿el "plan bonito" era dejar escapar a Fito?	entonces el plan bonito dejar escapar fito	positivo
875	Y se voló #Fito como el gran ave Fénix, en analogía con el gran #Planfenix de nuestro magnífico presidente de cartón Daniel Noboa. Al final el país merece lo que tiene, por ello se votó.	volo fito gran ave fenix analogia gran planfenix magnifico carton final pais merece ello voto	positivo

Figura 36 Ejemplo de Clasificación de Sentimientos Incorrecta por pysentiment  
Fuente: Autor

index	text	tweet	sentiment
100	Daniel Noboa el Presidente ya en funciones q le sucede? me huele a dictador odiador o se cree omnipotente x favor mesura Sr nadie lo nombró una especie de dictador o algo parecido. pone a un tipo no acto para seguridad nacional. y lo peor a la Vice la manda a la guerra a Israel	funciones sucede huele dictador odiador cree omnipotente x favor mesura sr nadie nombro especie dictador parecido pone tipo acto seguridad nacional peor vice manda guerra israel	positive
875	Y se voló #Fito como el gran ave Fénix, en analogía con el gran #Planfenix de nuestro magnífico presidente de cartón Daniel Noboa. Al final el país merece lo que tiene, por ello se votó.	volo fito gran ave fenix analogia gran planfenix magnifico carton final pais merece ello voto	positive

Figura 37 Ejemplo de Clasificación de Sentimientos Incorrecta por Textblob  
Fuente: Autor



*Figura 38 Gráfico de Barras de la Distribución de Cantidad por Categoría  
Fuente: Autor*