

**UNIVERSIDAD CATÓLICA BOLIVIANA “SAN PABLO”
UNIDAD ACADÉMICA REGIONAL LA PAZ
DEPARTAMENTO DE POSGRADO
MAESTRÍA EN CIENCIA DE DATOS**

PROYECTO FINAL

“ANÁLISIS DE SENTIMIENTO: MODELO DE PREDICCIÓN DE SENTIMIENTO EN BASE A LOS TWEETS SOBRE EL COVID-19”



MÓDULO de Machine Learning

Docente: Jheser Guzman Ph.D.

**Maestranes: Virginia Mercedes Fernández Daza
Marco Antonio Velásquez Rocha
Ivan Israel Machicado Quiroga**

noviembre, 2024

Índice general

1.	Introducción.....	1
1.1.	Contexto de la pandemia de COVID-19 y la importancia del análisis de sentimientos en redes sociales.	1
1.2.	Relevancia del análisis de tweets para la comprensión del estado emocional de la población.....	1
1.3.	Objetivo general del proyecto y objetivos específicos	2
1.4.	Justificación del estudio.	3
2.	Marco Teórico.	4
2.1.	Fundamentos de la clasificación de sentimientos en redes sociales.	4
2.2.	Modelos de Procesamiento de Lenguaje Natural (NLP).	5
2.3.	Arquitecturas de redes neuronales profundas: Redes recurrentes (RNN), LSTM y transformadores.	6
2.4.	BERT y variantes como modelos basados en atención.....	6
2.5.	Preprocesamiento de datos para análisis de sentimientos en redes sociales.	6
2.6.	Evaluación del rendimiento de modelos: Métricas de precisión, recall y F1-score.	7
4.	Desarrollo de la Propuesta	10
4.1.	Recopilación de datos.	10
4.2.	Análisis Exploratorio de Datos (EDA).....	12
4.2.1.	Renombrar columnas.	12
4.2.2.	Preparación del texto para modelado.	12
4.2.3.	Lematización y extracción de POS tags.....	12
4.2.4.	Análisis exploratorio de datos para familiarizarse con la información.	14
4.3.	Implementación de arquitecturas avanzadas de redes neuronales.	20
4.4.	Adaptación de modelos de NLP para procesamiento eficiente de grandes volúmenes de datos.	21
4.5.	Diseño y aplicación del preprocesamiento de datos (limpieza, normalización y estructuración de tweets).	22
4.6.	Evaluación de la efectividad de los modelos mediante métricas de rendimiento.	22
5.	Resultados.....	24
5.1.	Descripción de los resultados obtenidos en la clasificación de sentimientos.	24
5.2.	Comparación entre las diferentes arquitecturas y configuraciones.....	30
5.3.	Análisis de la efectividad de los modelos en el contexto de la pandemia de COVID-19.	33
6.	Discusión.	33

6.1.	Interpretación de los resultados obtenidos.	33
6.2.	Comparación con investigaciones previas y trabajos relacionados.....	34
6.3.	Limitaciones del estudio y posibles mejoras.	35
6.4.	Aplicaciones prácticas de los resultados en la toma de decisiones y políticas públicas.	36
7.	Conclusiones.	37
8.	Referencias Bibliográficas.....	38

Índice de tablas

Tabla 1.	Matriz Metodológica	9
Tabla 2:	Variables del conjunto de datos “tweets”	10
Tabla 3:	Cantidad tweets por Sentimiento	11
Tabla 4:	En la columna “tweet” del conjunto de datos se convierta el texto a minúsculas.....	12
Tabla 5:	Tabla comparativa de texto Lematizado	13
Tabla 6:	Tabla de texto aplicando POS tag	13
Tabla 7:	Tabla comparativa eliminando NN.....	14
Tabla 8:	Tabla de longitud de caracteres.....	15
Tabla 9:	Tabla de Unigrams	18
Tabla 10:	Tabla de Bigrams.....	19
Tabla 11:	Tabla de Trigrams	20
Tabla 12:	Tabla de indicadores Regresión Logística.....	24
Tabla 13:	Tabla de indicadores Árbol de Decisión	26
Tabla 14:	Tabla de indicadores Random Forest.....	27
Tabla 15:	Tabla de indicadores Naive Bayes	29
Tabla 16:	Tabla de indicadores BERT	30
Tabla 17:	Tabla Comparativa de indicadores.....	32

Índice de figuras

Figura 1:	Ratio de Locations	11
Figura 2.	Cambio de nombre OriginalTweet ' por 'tweet'	12
Figura 3:	Top 40 palabras más frecuentes.....	16
Figura 4:	Nube de palabras.....	16
Figura 5:	Matriz de Confusión de Regresión Logística	25
Figura 6:	Matriz de Confusión de Regresión Logística	26

Figura 7: Matriz de Confusión de Random Forest28

Figura 8: Matriz de Confusión de Regresión Logística29

1. Introducción.

1.1. Contexto de la pandemia de COVID-19 y la importancia del análisis de sentimientos en redes sociales.

Entre muchos sucesos, las gestiones 2019 y 2021 estuvieron caracterizadas por un hecho en particular, la presencia del Coronavirus. A raíz de esta enfermedad los diferentes países tuvieron que adecuar sus sistemas de salud, educación, políticos, entre otros al nuevo contexto internacional, considerando que se atravesó por una situación que cambio el desarrollo normal de todas las actividades.

Tomando en cuenta que no se disponía de mucha información sobre el COVID-19; asimismo que, al comienzo no existía medicación alguna, una de las primeras medidas asumidas por los Estados fue la declaratoria de cuarentena, acontecimiento que se desarrolló a nivel mundial. Con el encierro forzado, los medios digitales tomaron mucha relevancia, las clases se desarrollaban por diferentes plataformas, las reuniones del mismo modo, no podían ser presenciales ya que esto implicaba un alto riesgo. Las redes sociales tomaron mayor fuerza como un medio de comunicación e información, y se convirtieron en una fuente rica de datos e información.

La pandemia de COVID-19 generó un impacto global que no solo afectó la salud física de las personas, sino que también desencadenó una crisis emocional y psicológica a nivel mundial. Las redes sociales, especialmente Twitter, se convirtieron en una de las principales plataformas para que los usuarios compartieran sus experiencias, opiniones y sentimientos sobre la pandemia. Este fenómeno generó grandes cantidades de datos, lo que plantea un desafío significativo en la clasificación y análisis de los sentimientos expresados, dada la complejidad y variabilidad del lenguaje utilizado. Los modelos de análisis de sentimientos tradicionales a menudo no logran capturar la riqueza y los matices presentes en estos textos, lo que dificulta la interpretación precisa de las emociones y opiniones.

1.2. Relevancia del análisis de tweets para la comprensión del estado emocional de la población.

Como se mencionó en el punto anterior, las redes sociales, en este caso Twitter, en el periodo de la pandemia se convirtió en una fuente de datos para conocer la percepción de las personas sobre diferentes acontecimientos, principalmente la relacionada al COVID-19. El objetivo general de este proyecto es desarrollar y optimizar modelos de clasificación de tweets que permitan un análisis de sentimiento preciso y eficiente, aprovechando la gran cantidad de datos generados durante la pandemia de COVID-19.

La creciente cantidad de datos generados en redes sociales durante eventos de alto impacto, como la pandemia de COVID-19, ha presentado nuevas

oportunidades y desafíos en el análisis de sentimientos. Los modelos tradicionales de clasificación de texto, comúnmente utilizados para el análisis de tweets, han mostrado limitaciones significativas cuando se trata de procesar grandes volúmenes de datos y captar matices en el lenguaje natural, especialmente en situaciones donde el contexto y la estructura semántica son complejos (Liu, 2020). Durante la pandemia, los usuarios de redes sociales expresaron una amplia gama de emociones, desde ansiedad y frustración hasta esperanza y apoyo, lo cual requiere modelos avanzados para analizar estos sentimientos de manera efectiva y precisa.

Para lograr el objetivo, se han definido cinco objetivos específicos que guiarán el desarrollo de la investigación. En primer lugar, se implementarán arquitecturas avanzadas de redes neuronales profundas, como BERT y sus variantes, para mejorar la precisión de los modelos en la clasificación de sentimientos, estos modelos y sus variantes han demostrado un rendimiento superior en la clasificación de sentimientos debido a su capacidad para comprender el contexto bidireccional de las palabras y extraer significados complejos en tiempo real (Devlin et al., 2019). En segundo lugar, se optimizarán estos modelos para procesar eficientemente grandes volúmenes de datos, garantizando que puedan analizar tweets en tiempo real y manejar las variaciones del lenguaje y el contexto social. Además, se desarrollará un preprocesamiento adecuado de los datos para maximizar la calidad de la información, se evaluará el rendimiento de los modelos mediante métricas estándar y se ajustarán los hiperparámetros para adaptar los modelos al contexto específico de los tweets generados durante la pandemia.

Este proyecto se enfoca en abordar los desafíos actuales en el análisis de sentimientos en redes sociales, utilizando enfoques avanzados de inteligencia artificial. Con la implementación de estos modelos optimizados, se espera mejorar significativamente la comprensión de las emociones y opiniones generadas en tiempos de crisis, proporcionando herramientas más efectivas para los investigadores, gobiernos y organizaciones que necesiten interpretar el sentimiento público y tomar decisiones informadas basadas en datos.

1.3. Objetivo general del proyecto y objetivos específicos

1.3.1. Objetivo General

Desarrollar y optimizar modelos de clasificación de tweets que permitan un análisis de sentimiento preciso y eficiente, aprovechando las grandes cantidades de datos generados durante la pandemia de COVID-19.

1.3.2. Objetivos específicos

- Implementar arquitecturas avanzadas de redes neuronales profundas (como modelos basados en atención y transformadores, tales como BERT y sus variantes) para mejorar la precisión en la clasificación de sentimientos en tweets relacionados con la pandemia de COVID-19.
- Adaptar y optimizar los modelos de NLP para procesar eficientemente grandes volúmenes de datos, garantizando que los modelos puedan analizar tweets en tiempo real y manejar variaciones en el lenguaje y contexto social.
- Desarrollar un preprocesamiento adecuado de datos de texto, que permita limpiar, normalizar y estructurar los tweets de forma que se maximice la calidad de la información para los modelos de clasificación.
- Evaluar la efectividad de los modelos mediante métricas de rendimiento, tales como precisión, recall y F1-score, para identificar las arquitecturas y configuraciones que ofrecen el mejor desempeño en el análisis de sentimientos.

1.4. Justificación del estudio.

1.4.1. Justificación Social

Durante la pandemia de COVID-19, las redes sociales, especialmente Twitter, fueron utilizadas por millones de personas para expresar emociones, preocupaciones y opiniones sobre la crisis sanitaria. Estos tweets reflejan una amplia gama de sentimientos, desde miedo y ansiedad hasta solidaridad y esperanza, lo cual ofrece una oportunidad única para comprender el impacto social y emocional de la pandemia. Sin embargo, los modelos de análisis de sentimientos tradicionales han demostrado ser ineficaces para capturar la complejidad del lenguaje utilizado, especialmente en situaciones de crisis. El desarrollo de modelos de clasificación más precisos y eficientes permitirá una mejor comprensión del estado emocional de la sociedad durante la pandemia, lo que puede ayudar a los gobiernos, instituciones de salud pública y organizaciones a responder de manera más efectiva a las necesidades emocionales y psicológicas de la población.

1.4.2. Justificación Tecnológica

El análisis de sentimientos en grandes volúmenes de datos generados en tiempo real, como los tweets, requiere el uso de tecnologías avanzadas que puedan manejar el procesamiento masivo de datos y comprender la complejidad del lenguaje natural. Los modelos de aprendizaje profundo,

como BERT y otras arquitecturas de transformers, son capaces de captar el contexto completo de los textos, lo cual es fundamental para mejorar la precisión en la clasificación de sentimientos. La optimización de estos modelos para analizar grandes cantidades de datos durante un evento global como la pandemia de COVID-19 presenta una oportunidad para avanzar en el campo del procesamiento de lenguaje natural, con aplicaciones que pueden ir más allá de la crisis sanitaria y aplicarse a otras situaciones de alta carga informativa.

1.4.3. Justificación Política

Las políticas públicas, especialmente en situaciones de crisis como la pandemia de COVID-19, se benefician enormemente de datos precisos y actualizados sobre el estado emocional y las opiniones de la población. Un análisis de sentimiento eficaz, basado en grandes volúmenes de datos, puede proporcionar información valiosa sobre cómo las decisiones gubernamentales y las medidas de salud pública son recibidas por la ciudadanía, permitiendo ajustes en tiempo real en las políticas y estrategias de comunicación. Al desarrollar modelos que procesen datos de redes sociales de manera eficiente, este proyecto contribuirá a mejorar la toma de decisiones políticas y la eficacia de las intervenciones gubernamentales, ayudando a gestionar mejor las crisis futuras y a fortalecer la relación entre los gobiernos y la población.

2. Marco Teórico.

2.1. Fundamentos de la clasificación de sentimientos en redes sociales.

El análisis de sentimientos en redes sociales es un proceso que utiliza técnicas de procesamiento de lenguaje natural y aprendizaje automático para evaluar las emociones y opiniones expresadas por los usuarios en plataformas digitales (en el presente proyecto, a través de Twitter). Este enfoque permite clasificar los mensajes como positivos, negativos o neutros, y se ha convertido en una herramienta esencial para empresas y organizaciones que buscan entender la percepción pública. Esta forma de análisis se enfoca en identificar el tono emocional detrás de las publicaciones en redes sociales.

Las principales técnicas para llevar a cabo el análisis de sentimientos se dividen principalmente en: basadas en léxico, que utilizan diccionarios predefinidos que asocian palabras con sentimientos específicos, permitiendo una clasificación rápida; aprendizaje automático, que implica entrenar modelos con conjuntos de datos etiquetados para que puedan clasificar nuevos textos; y métodos híbridos, que combinan enfoques basados en léxico y aprendizaje automático para mejorar la precisión del análisis.

Para aplicar el análisis de sentimiento, se requiere de forma previa procesar los datos en forma de texto considerando que los mismos pueden estar desordenados, contener errores gramaticales, emojis y/o hashtags, o usar lenguaje coloquial. En este sentido es necesario la tokenización, para dividir el texto; normalización; para convertir el texto a minúsculas, expandir abreviaturas o eliminar palabras; gestionar emojis y emoticones; y aplicar lematización o stemming.

2.2. Modelos de Procesamiento de Lenguaje Natural (NLP).

El " Procesamiento del Language Natural " (NLP) es una disciplina con una larga trayectoria. Nace en la década de 1960, como una subárea de la Inteligencia Artificial y la Lingüística, con el objeto de estudiar los problemas derivados de la generación y comprensión automática del lenguaje natural.

En los últimos años, los aportes para su mejora han incrementado sustancialmente, permitiendo el procesamiento de gran cantidad de información en formato texto con una eficacia bastante aceptable.

El NLP emplea modelos matemáticos y computacionales que permiten interpretar, procesar y generar lenguaje humano, entre los ejemplos mas comunes se encuentran los motores de búsqueda, herramientas de traducción automática o la generación automática de resúmenes. Los modelos para NLP evolucionaron desde aquellos que se basan en reglas y estadísticas hasta llegar a redes neuronales profundas y transformadores.

Dentro de los modelos para NLP podemos clasificarlos en aquellos que están basados en reglas, los basados en aprendizaje profundo, basados en transformadores y modelos para español y otros idiomas. En relación a los primeros estos se dividen en 2: modelos basados en reglas, que básicamente utilizan reglas gramaticales y lingüísticas, y los modelos estadísticos, donde podemos encontrar el modelo de n-gramas, Markov Models y Latent Dirichlet Allocation (LDA).

Respecto a los segundos, podemos hacer mención a Word Embeddings, que representan palabras como vectores densos en un espacio semántico; Redes Neuronales Recurrentes, diseñados para las secuencias de datos; Redes Neuronales Convolucionales (CNN), usadas para tareas de clasificación de texto y extracción de características.

Para el caso de los modelos basados en transformadores, encontramos al modelo transformer, introducido por Vaswani et al. (2017) en *Attention is All You Need* que utiliza mecanismos de atención para procesar secuencias de texto en paralelo; asimismo, también se encuentran presentes los modelos pre entrenados, como es el caso de BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pre-trained Transformer) y T5 (Text-to-Text Transfer Transformer). En relación a los modelos para español y otros idiomas encontramos al modelo BETO que es una versión de BERT entrenada para datos en español.

2.3. Arquitecturas de redes neuronales profundas: Redes recurrentes (RNN), LSTM y transformadores.

Las redes neuronales recurrentes (RNN) están diseñadas para manejar datos secuenciales, manteniendo una "memoria" de entradas previas en la secuencia. Sin embargo, las RNN tradicionales enfrentan problemas con secuencias largas debido a la pérdida de gradiente (Bengio, Simard, & Frasconi, 1994). Para resolver esta limitación, las redes LSTM fueron introducidas por Hochreiter y Schmidhuber (1997). Las LSTM contienen "puertas" de entrada, olvido y salida, lo que les permite retener información durante períodos prolongados. Esto mejora la comprensión del contexto en el análisis de sentimientos y otras tareas de NLP (Greff et al., 2017).

2.4. BERT y variantes como modelos basados en atención.

BERT, un modelo basado en transformadores, aprovecha el aprendizaje bidireccional, permitiendo al modelo comprender el contexto completo de cada palabra en una oración al analizar en ambas direcciones (Devlin et al., 2019). Esto le permite una comprensión precisa de los matices semánticos y de los contextos más profundos, mejorando la precisión en tareas de clasificación de sentimientos y comprensión del lenguaje natural. Durante la pandemia, BERT fue especialmente útil para analizar grandes volúmenes de texto en redes sociales y capturar variaciones en el tono de la conversación pública (Sun et al., 2020).

La capacidad de BERT para procesar texto en paralelo y extraer significados contextuales detallados permitió realizar análisis de sentimiento a gran escala, esencial para entender los cambios en la percepción pública durante la pandemia (Sun et al., 2020). Al capturar matices y comprender el contexto completo de una oración, BERT y otros modelos basados en transformadores ayudaron a los investigadores a obtener una visión más precisa de las opiniones expresadas en las redes sociales (Kumari et al., 2020).

Los modelos BERT han dado lugar a diferentes variantes que fueron adaptadas a diferentes idiomas, tareas y tamaños, dentro de las cuales podemos destacar los modelos multilingües donde se encuentran mBERT y BETO; asimismo, existen modelos especializados como son los casos de RoBERTa, ALBERT, DistilBERT y TinyBERT, y finalmente existen modelos avanzados como son SpanBERT, BioBERT y T5.

2.5. Preprocesamiento de datos para análisis de sentimientos en redes sociales.

El análisis de sentimiento basado en datos de redes sociales, en particular Twitter, se ha convertido en una herramienta crucial para comprender la

percepción pública sobre eventos de gran impacto, como la pandemia del COVID-19. Sin embargo, los modelos de clasificación tradicionales han mostrado limitaciones en su capacidad para manejar el volumen y la complejidad de los datos generados durante crisis globales. Este marco teórico explora los problemas asociados con dichos modelos y revisa las propuestas más recientes para optimizar el análisis de sentimiento.

Para tener éxito en el análisis de sentimiento es crucial la etapa de preprocesamiento de datos, especialmente porque el texto proviene de redes sociales, considerando que el mismo contiene mucho ruido; es decir, contiene abreviaturas, emojis, errores ortográficos, iniciales, entre otros. El preprocesamiento nos permitirá transformar los datos en una representación que permita extraer información útil para construir modelos efectivos.

Dentro de las etapas del preprocesamiento de datos está la limpieza de datos donde se deben eliminar caracteres especiales, URLs y etiquetas HTML, posteriormente se pasa a la normalización del texto que incluye las operaciones de conversión a minúsculas, eliminación de stopwords, lematización y stemming. Finalmente, se considera el manejo de características específicas de redes sociales, donde se procesarán los emoticones y emojis, se extraerán los hashtags y se hará el manejo de las menciones (@usuario). Todo esto permitirá que el proceso de tokenización sea efectivo.

2.6. Evaluación del rendimiento de modelos: Métricas de precisión, recall y F1-score.

Para evaluar la calidad de los modelos se recurre a las métricas de precisión recall y F1-score, mismas que se calculan a partir de la matriz de confusión y proporcionan información sobre el comportamiento de las predicciones.

Métrica de precisión: Permitirá medir la proporción de predicciones positivas que son correctas. Una elevada precisión indica que el modelo comete pocos errores al etiquetar casos positivos.

Recall: Permite evaluar la capacidad del modelo para identificar todos los casos positivos. Un elevado recall señala que el modelo identifica la mayoría de los casos positivos.

F1-score: Es la medida armónica de precisión y recall. Equilibra ambas métricas en un solo valor y es útil cuando se necesita un compromiso entre precisión y recall. Un F1-score elevado indica un buen equilibrio entre identificar correctamente los positivos (recall) y evitar etiquetar negativos como positivos.

3. Metodología

La metodología del proyecto se define como el conjunto de procedimientos y técnicas que se emplearán para llevar a cabo el análisis de sentimientos en los

tweets relacionados con la pandemia de COVID-19. A continuación, se describen los aspectos clave de la metodología empleada en este proyecto.

3.1. Enfoque de la Investigación

El enfoque de esta investigación es **cuantitativo** y **computacional**, con el objetivo de clasificar y analizar los sentimientos expresados en los tweets. Se utilizarán modelos de procesamiento de lenguaje natural (PLN) y aprendizaje automático para extraer, clasificar y evaluar las emociones relacionadas con la pandemia. El análisis será realizado mediante el uso de herramientas estadísticas y técnicas de machine learning para modelar los sentimientos expresados en los textos.

3.2. Tipo de Investigación

La investigación será de tipo **descriptivo** y **exploratorio**. El carácter descriptivo se refiere a la recopilación y análisis de datos de tweets en función de su contenido y el sentimiento expresado. El componente exploratorio permite descubrir patrones y relaciones entre las opiniones de los usuarios y la evolución de la pandemia, sin un marco preestablecido. A través de esta investigación, se busca comprender las emociones predominantes en los tweets y cómo estas varían en diferentes contextos de la pandemia.

3.3. Método de Investigación

El método de investigación utilizado será el **análisis de datos textuales**, apoyado en herramientas de análisis de sentimientos. En este proyecto, se emplearán técnicas de **minería de texto** y **análisis de sentimientos** para extraer las emociones de los tweets utilizando modelos de clasificación como regresión logística, máquinas de soporte vectorial (SVM), y redes neuronales. La investigación seguirá un enfoque cuantitativo en el cual se analizarán grandes volúmenes de datos, identificando patrones en las opiniones y sentimientos de los usuarios de Twitter.

3.4. Fuentes de Información

Las fuentes de información se basan en **tweets públicos** recolectados a través de la API de Twitter. Estos tweets estarán relacionados con palabras clave que describen la pandemia de COVID-19, como "coronavirus", "COVID-19", "pandemia", "vacuna", entre otras. Además, se utilizarán bases de datos externas con etiquetas de sentimiento para el entrenamiento de los modelos, como conjuntos de datos públicos disponibles de análisis de sentimientos.

3.5. Instrumentos de Recolección de Datos

Para la recolección de los datos, se utilizará la **API de Twitter**, que permitirá extraer tweets públicos en tiempo real basados en ciertas palabras clave. Los tweets extraídos se almacenarán en formato estructurado (por ejemplo, CSV o

JSON), incluyendo atributos como el contenido del tweet, la fecha y hora de publicación, el usuario, y el número de interacciones (retweets, likes). Esta información será la base para el análisis de sentimiento posterior.

3.6. Matriz Metodológica

La matriz metodológica presentará las siguientes variables clave para la recolección y análisis:

Tabla 1. Matriz Metodológica

Variable	Descripción	Técnica de Recolección	Método de Análisis
Texto del tweet	Contenido textual del tweet	API de Twitter	Análisis de texto con PLN
Sentimiento	Sentimiento expresado en el tweet (positivo, negativo, neutral)	Análisis de sentimientos (modelos de ML)	Clasificación de sentimientos
Fecha y Hora	Momento en el que se publicó el tweet	API de Twitter	Análisis temporal
Número de interacciones	Cantidad de interacciones (retweets, likes)	API de Twitter	Análisis de correlación

Fuente: Elaboración propia, 2024

3.7. Plan de Experimentos y Ajuste de Modelos

El plan de experimentos incluirá las siguientes etapas:

1. **Recolección de Datos:** Recopilación de tweets mediante la API de Twitter durante un período determinado relacionado con la pandemia de COVID-19.
2. **Preprocesamiento de Datos:** Limpieza de los datos textuales, eliminación de stopwords, tokenización y normalización de texto.
3. **Entrenamiento de Modelos:** Entrenamiento de modelos de análisis de sentimientos utilizando técnicas de **aprendizaje supervisado**, como **regresión logística**, **máquinas de soporte vectorial (SVM)**, y **redes neuronales profundas (LSTM)**. El entrenamiento se realizará utilizando un conjunto de datos etiquetado con sentimientos positivos, negativos y neutrales.
4. **Evaluación y Ajuste de Modelos:** Evaluación de los modelos utilizando métricas estándar como **precisión**, **recall**, **F1-score**, y **accuracy**. Se emplearán técnicas de **validación cruzada** y **ajuste de hiperparámetros** (como GridSearch) para mejorar el desempeño de los modelos.

5. **Implementación de Modelo Final:** Selección del modelo con mejor rendimiento para la predicción final de sentimientos en nuevos tweets y generación de visualizaciones de los resultados.

4. Desarrollo de la Propuesta

4.1. Recopilación de datos.

El conjunto de datos para el estudio está compuesto por 41,157 registros y 6 columnas, la data esta almacenada en una base de datos .CVS.

Las columnas que contiene el conjunto de datos es el siguiente:

Tabla 2: Variables del conjunto de datos “tweets”

RangeIndex: 41157 entries, 0 to 41156			
Data columns (total 6 columns):			
#	Column	Non-Null Count	Dtype
0	UserName	41157 non-null	int64
1	ScreenName	41157 non-null	int64
2	Location	32567 non-null	object
3	TweetAt	41157 non-null	object
4	OriginalTweet	41157 non-null	object
5	Sentiment	41157 non-null	object

Fuente: Elaboración propia, 2024

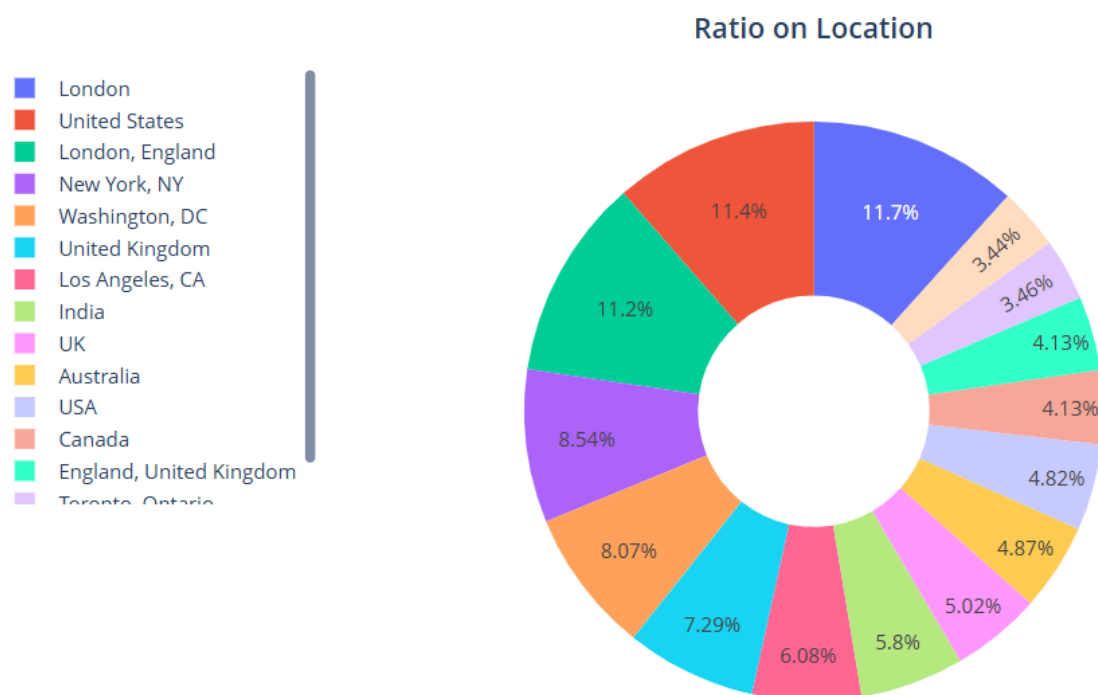
Análisis de la calidad de los datos:

- ✓ Se han identificado 41,157 filas (entradas) en el DataFrame, numeradas del 0 al 41,156.
- ✓ El DataFrame tiene un total de 6 columnas.

En la tabla 1 se puede observar la lista de las columnas en el DataFrame, junto con información sobre cuántos valores no nulos hay en cada columna y el tipo de dato de cada columna o **Non-Null Count**: Muestra cuántos valores no nulos hay en cada columna. En este caso, la mayoría son de tipo object (que generalmente representa texto), y hay dos columnas de tipo int64 (que representa números enteros).

Análisis de contenido de los datos:

Sobre la información del dataset se puede indicar que los tweets en mayor cantidad provienen de las siguientes localizaciones: London con 540, United States con 528, London, England con 520, New York, NY 395 y Washington, DC con 373 tweets. Gráficamente se puede observar:

Figura 1: Ratio de Locations

Fuente: Elaboración propia, 2024

Adicionalmente se puede observar que la clasificación de sentimiento contenida en el conjunto de datos el sentimiento POSITIVO es el con mayor clasificación con 11.422 registros, luego le sigue los negativos con 9,917 registros, tal como se muestra en el cuadro siguiente:

Tabla 3: Cantidad tweets por Sentimiento

	count
Sentiment	
Positive	11422
Negative	9917
Neutral	7713
Extremely Positive	6624
Extremely Negative	5481
dtype: int64	

Fuente: Elaboración propia, 2024

4.2. Análisis Exploratorio de Datos (EDA).

4.2.1. Renombrar columnas.

Como se observó en el punto anterior, el conjunto de datos dispone de 6 columnas, se cambia el nombre de la columna texto que se analiza “OriginalTweet” por ‘tweet’ y se obtiene el siguiente resultado:

Figura 2. Cambio de nombre OriginalTweet ' por 'tweet'

```
Index(['UserName', 'ScreenName', 'Location', 'TweetAt', 'tweet', 'Sentiment'], dtype='object')
```

Fuente: Elaboración propia, 2024

4.2.2. Preparación del texto para modelado.

Se elabora una función para que en la columna “tweet” del conjunto de datos convierta el texto a minúsculas, eliminando el texto entre corchetes, las palabras que contienen número, los textos que contengan “@”, “http”, los signos de puntuación y los espacios en blanco adicionales. A continuación, se muestra un ejemplo de cómo queda el texto después de aplicar los ajustes previamente mencionados.

Tabla 4: En la columna “tweet” del conjunto de datos se convierta el texto a minúsculas

```
0                                     and and
1  advice talk your neighbours family exchange ph...
2  coronavirus australia woolworths give elderly ...
5  news the regionâs first confirmed covid case c...
6  cashier grocery store was sharing his insights...
7  was the supermarket today didnt buy toilet pap...
8  due covid our retail store and classroom atlan...
9  for corona preventionwe should stop buy things...
10 all month there hasnt been crowding the superm...
11 due the covid situation have increased demand ...
Name: tweet, dtype: object
```

Fuente: Elaboración propia, 2024

4.2.3. Lematización y extracción de POS tags.

Se carga un modelo de lenguaje llamado **en_core_web_sm de spaCy**, que es un modelo pre entrenado para el idioma inglés. Este modelo incluye herramientas para analizar y procesar texto, como el análisis sintáctico y la lematización, que es el proceso de reducir las palabras a su forma base. Se crea una función que se encargará de lematizar el texto. Esta función toma un texto

como entrada, representado como un **string**. Dentro de la función, el texto se procesa utilizando el modelo de **spaCy**. Este procesamiento tiene como resultado una estructura de datos que contiene información sobre cada palabra del texto, permitiendo acceder a diversos atributos lingüísticos de cada palabra (como su forma base).

Se extraen los lemas de cada palabra procesada. El lema es la forma básica de una palabra. Por ejemplo, los lemas de **"caminando"** y **"caminé"** serían **"caminar"**. Todos los lemas se agrupan para formar una nueva cadena de texto.

Tabla 5: Tabla comparativa de texto Lematizado

	tweet	Sentiment	tweet_clean	tweet_lemmatized
0	and and	Neutral	and and	and and
1	advice talk your neighbours family exchange ph...	Positive	advice talk your neighbours family exchange ph...	advice talk your neighbours family exchange ph...
2	coronavirus australia woolworths give elderly ...	Positive	coronavirus australia woolworth give elderly d...	coronavirus australia woolworth give elderly d...
5	news the regionâs first confirmed covid case c...	Positive	news the regionâs first confirm covid case com...	news the regionâs first confirm covid case com...
6	cashier grocery store was sharing his insights...	Positive	cashier grocery store be share his insight pro...	cashier grocery store be share his insight pro...
...
41147	yâall really shitting that much more home coro...	Negative	yâall really shit that much more home coronavi...	yâall really shit that much more home coronavi...
41149	still shocked the number toronto supermarket e...	Negative	still shock the number toronto supermarket emp...	still shock the number toronto supermarket emp...
41150	never that weâd situation amp world that going...	Positive	never that weâd situation amp world that go th...	never that weâd situation amp world that go th...
41152	airline pilots offering stock supermarket shel...	Neutral	airline pilot offer stock supermarket shelf lo...	airline pilot offer stock supermarket shelf lo...
41156	well newused rift are going for amazon althoug...	Negative	well newuse rift be go for amazon although the...	well newuse rift be go for amazon although the...

Fuente: Elaboración propia, 2024

Una vez que el texto fue lematizado se procedió a la extracción de etiquetas. Nuevamente, con **spacy** se procede a trabajar con el texto y se crea un objeto estructurado para analizar palabras. La función extrae **etiquetas POS** de cada palabra en el texto, las devuelve como una lista y almacena el resultado en una nueva columna, **tweet_pos_tags**, que contendrá las etiquetas gramaticales correspondientes a los tweets lematizados:

Tabla 6: Tabla de texto aplicando POS tag

	tweet_pos_tags
0	[CCONJ, CCONJ]
1	[NOUN, VERB, PRON, PROPN, NOUN, NOUN, NOUN, NO...
2	[PROPN, PROPN, PROPN, VERB, ADJ, ADJ, ADJ, NOU...
5	[VERB, DET, NOUN, ADV, VERB, ADJ, NOUN, VERB, ...
6	[NOUN, NOUN, NOUN, AUX, VERB, PRON, NOUN, VERB...
...	...
41147	[PROPN, ADV, VERB, ADV, ADV, ADJ, NOUN, NOUN, ...
41149	[ADV, VERB, DET, NOUN, PROPN, PROPN, NOUN, NOU...
41150	[ADV, DET, ADJ, NOUN, PROPN, NOUN, PRON, VERB,...
41152	[PROPN, PROPN, VERB, NOUN, NOUN, NOUN, VERB, N...
41156	[INTJ, ADP, NOUN, AUX, VERB, ADP, NOUN, SCONJ,...

Fuente: Elaboración propia, 2024

A continuación, se realiza un proceso para mantener solo los sustantivos del texto que estamos analizando. En esta etapa se procesa cada tweet en la columna “**tweet**” del dataframe utilizando **spacy**. Se filtra todas las palabras conservando los sustantivos lematizados en un nuevo **string**, los resultados se almacenan en una nueva columna “**tweet_POS_removed**”.

En el dataframe resultante se comparan los tweets originales con los tweets lematizados que contienen solo sustantivos. Esto es muy útil en el procesamiento de texto y análisis de datos, ya que ayuda a centrarse en las palabras clave (sustantivos) de los tweets, lo que puede facilitar un análisis más enfocado.

Tabla 7: Tabla comparativa eliminando NN

tweet_clean	tweet_lemmatized	tweet_POS_removed
	and and	
advice family exchange phone number contact li...	advice talk your neighbours family exchange ph...	advice family exchange phone number contact li...
woolworth shopping hour outbreak	coronavirus australia woolworth give elderly d...	woolworth shopping hour outbreak
regionâs case week people area store purchase ...	news the regionâs first confirm covid case com...	regionâs case week people area store purchase ...
grocery store insight credibility civic class	cashier grocery store be share his insight pro...	grocery store insight credibility civic class
...
yâall home coronavirus toiletpaper	yâall really shit that much more home coronavi...	yâall home coronavirus toiletpaper
number supermarket employee sort mask employee...	still shock the number toronto supermarket emp...	number supermarket employee sort mask employee...
situation world supermarket package mail roule...	never that weâd situation amp world that go th...	situation world supermarket package mail roule...
airline pilot stock supermarket shelf	airline pilot offer stock supermarket shelf lo...	airline pilot stock supermarket shelf
rift amazon market price price headset	well newuse rift be go for amazon although the...	rift amazon market price price headset

Fuente: Elaboración propia, 2024

4.2.4. Análisis exploratorio de datos para familiarizarse con la información.

Para la visualización de los datos según la longitud de los caracteres “**tweet**”, procedemos a agregar una nueva columna (**tweet_length**) que calcule la longitud de los caracteres, posterior a ello lo ordenamos de forma descendente y el resultado es el siguiente:

Tabla 8: Tabla de longitud de caracteres

	tweet_clean	tweet_length
29763	victory jonesquarantinediarie toiletpaperpanic...	209
3580	coronaviruschallenge coronavirusupdate coronav...	192
5121	job purpose community thank grocery store work...	191
5511	panic buying headache supermarket operator cri...	186
24716	update coronavirus effect smartphone consumer ...	183
...
26924	map	3
12138	iâm	3
9586	guy	3
28678	ord	3
15487	han	3

Fuente: Elaboración propia, 2024

Podemos observar que hay una variación grande entre el máximo tamaño de 209 caracteres hasta solo 3 caracteres.

➤ Top 40 palabras más frecuentes.

El código utilizado **tokeniza** el texto en una lista de palabras. Utilizamos counter para contar la frecuencia de cada palabra. Posteriormente, se extrae las 40 palabras más comunes y sus frecuencias.

Siguiendo los siguientes procesos:

Tipos de Tokenización

Tokenización basada en palabras: Divide el texto en palabras individuales, ejemplo: "El cliente está satisfecho" se tokeniza como ["El", "cliente", "está", "satisfecho"].

Tokenización basada en caracteres: Divide el texto en caracteres individuales. Ejemplo: "Hola" se tokeniza como ["H", "o", "l", "a"], Tokenización basada en subpalabras: Divide las palabras en unidades más pequeñas, como prefijos, sufijos o raíces, ejemplo: "clasificación" podría dividirse en ["clasific", "ación"].

Tokenización basada en oraciones: Divide el texto en oraciones completas, ejemplo: "Hoy hace sol. Mañana podría llover." se tokeniza como ["Hoy hace sol.", "Mañana podría llover."].

Los datos obtenidos se convierten a un dataframe de pandas para facilitar la visualización. El resultado puede observarse en la siguiente imagen:

El análisis de las palabras más frecuentes en el conjunto de datos, relacionados con el coronavirus puede proporcionar una perspectiva sobre las principales preocupaciones, temas y conversaciones de las personas durante la pandemia. Aquí se presentan posibles interpretaciones de las palabras clave mencionadas:

"coronavirus":

Es lógico que esta sea la palabra más frecuente, ya que es el tema central del dataset. Indica que los tweets se centran en aspectos relacionados con la pandemia, como la salud, las políticas públicas, y su impacto global.

"price" (precio):

Sugiere que muchas conversaciones giraron en torno al costo de bienes y servicios durante la pandemia. Esto podría estar relacionado con:

Aumento o fluctuaciones de precios debido a la demanda o problemas en la cadena de suministro.

Discusiones sobre prácticas de sobreprecio (price gouging) en artículos esenciales.

"supermarket" (supermercado):

Refleja la importancia de los supermercados como centros esenciales durante la pandemia, donde las personas se abastecieron de bienes básicos. También puede señalar preocupaciones sobre medidas de seguridad, disponibilidad de productos, y comportamientos de compra.

"food" (comida):

La comida es una necesidad básica, y su mención frecuente podría relacionarse con:

Abastecimiento durante el confinamiento, problemas de acceso a alimentos o donaciones, cambios en los patrones de consumo, como la cocina en casa o el interés en recetas.

"grocery" (tienda de comestibles):

Similar a "supermarket", indica preocupaciones sobre las compras de comestibles, su disponibilidad, y la experiencia de compra durante las restricciones por la pandemia.

"consumer" (consumidor):

Muestra un enfoque en los comportamientos y preocupaciones del consumidor, incluyendo:

Cambios en los hábitos de consumo, impactos económicos en las decisiones de compra, nuevas tendencias como el comercio electrónico o compras masivas.

Estas palabras sugieren que una parte significativa de las conversaciones en los tweets analizados estaba relacionada con el impacto de la pandemia en la vida cotidiana, particularmente en el acceso a bienes esenciales como alimentos y productos del supermercado. También reflejan preocupaciones económicas, como los precios y el comportamiento de los consumidores frente a la incertidumbre y las restricciones.

Esta información podría ser útil para identificar patrones de comportamiento, necesidades específicas durante la pandemia.

➤ Unigramas

Para obtener los unigramas se utilizará la columna de texto **“tweet_clean”**, que contiene los tweets ya limpios (sin caracteres especiales o ruido innecesario). Con la función **join** se combina todos estos tweets en una sola cadena, separándolas por un espacio. El resultado es un texto largo que contiene todos los tweets que posteriormente será dividido para obtener palabras individuales. Se utilizará **Counter** de la biblioteca **collections** para contar la frecuencia de cada palabra en la lista unigrams (Counter crea un diccionario donde las claves son las palabras y los valores son las cantidades de veces que cada palabra aparece). El resultado es un objeto **Counter** que refleja la frecuencia de cada unigram, como puede observarse en la siguiente imagen.

Por ejemplo:

"El cliente está satisfecho"

Los unigramas serían:

["El", "cliente", "está", "satisfecho"]

Tabla 9: Tabla de Unigrams

	Unigram	Frequency
0	coronavirus	9168
1	price	6990
2	store	6288
3	supermarket	5917
4	food	5452
5	grocery	5198
6	consumer	4224
7	people	4198
8	time	2413
9	worker	2202

Fuente: Elaboración propia, 2024

Se puede observar que la palabra **“coronavirus”** es la que más veces se repite (9168 veces) seguido de **“price”** (6.990 veces) y **“store”** (6.288 veces).

➤ Bigramas

Al igual que el anterior punto se utilizará la columna **tweet_clean** y todos los tweets se combinarán en una cadena separadas por espacios; asimismo, su uso el método `split()` para dividir el texto combinado en palabras individuales (words, será una lista que contiene todas las palabras).

Se utilizó la función **bigrams de nltk**, que toma la lista de palabras words y genera todos los pares adyacentes de palabras (bigramos). Cada **bigrama** es representado como una **tupla de dos palabras**. El resultado es una lista llamada **bigram_list** que contiene todos los bigramas extraídos del texto.

Por ejemplo:

"El cliente está satisfecho"

Los bigramas serían:

[("El", "cliente"), ("cliente", "está"), ("está", "satisfecho")]

En el caso de los bigramas, la combinación "grocery, store" es la que más se repite (3914 veces) seguido de "hand, sanitizer" (1179 veces) y "toilet,paper" (918).

Tabla 10: Tabla de Bigrams

	Bigram	Frequency
0	(grocery, store)	3914
1	(hand, sanitizer)	1179
2	(toilet, paper)	918
3	(oil, price)	829
4	(panic, buying)	637
5	(store, worker)	419
6	(supermarket, shelf)	416
7	(supply, chain)	381
8	(food, bank)	323
9	(price, coronavirus)	320

Fuente: Elaboración propia, 2024

➤ Trigramas

De forma similar a los anteriores casos, se utilizará la columna **tweet_clean** y se construirán cadenas de texto para posteriormente construir los trigramas y establecer las frecuencias de estas combinaciones, y se llegan a los siguientes resultados.

La secuencia "grocery, store, worker" es la que tiene una mayor frecuencia (377 veces), seguido de "grocery, store, employee" (215 veces) y "grocery, store, coronavirus" (142 veces).

Por ejemplo:

"El cliente está satisfecho"

Los trigramas serían:

[("El", "cliente", "está"), ("cliente", "está", "satisfecho")]

Tabla 11: Tabla de Trigrams

	Trigram	Frequency
0	(grocery, store, worker)	377
1	(grocery, store, employee)	215
2	(grocery, store, coronavirus)	142
3	(people, grocery, store)	125
4	(coronavirus, grocery, store)	124
5	(worker, grocery, store)	108
6	(food, supply, chain)	105
7	(trip, grocery, store)	86
8	(grocery, store, clerk)	83
9	(grocery, store, today)	79

Fuente: Elaboración propia, 2024

4.3. Implementación de arquitecturas avanzadas de redes neuronales.

➤ **Modelo de Regresión Logística.**

El modelo de Regresión Logística, es un modelo de clasificación supervisado que utiliza una función sigmoide para predecir la probabilidad de que una observación pertenezca a una clase específica. En este caso, la regresión logística asigna probabilidades a cada categoría de tweet (tópico) basándose en las características generadas por Non-Negative Matrix Factorization NMF (vectores TF-IDF). Se clasifica un ticket en la categoría con mayor probabilidad. Su simplicidad y eficiencia lo hacen ideal para problemas con relaciones lineales entre las características.

➤ **Modelo de Árbol de Decisión.**

El Árbol de decisión es un modelo basado en reglas jerárquicas que divide iterativamente los datos en subconjuntos basándose en características que maximizan la ganancia de información o reducen la impureza de Gini. En este caso, el árbol de decisión utiliza los vectores TF-IDF preprocesados para clasificar cada ticket en una de las cinco categorías. Este modelo es fácil de interpretar, permitiendo identificar directamente las reglas de clasificación para cada categoría.

➤ **Modelo Random Forest.**

Random forest es un conjunto de árboles de decisión donde cada árbol se entrena en un subconjunto aleatorio de datos y características. Luego, combina los resultados de todos los árboles para tomar una decisión final, lo

que mejora la precisión y reduce el sobreajuste. Aquí, el Random Forest predice la categoría de cada ticket al analizar patrones complejos en los datos, siendo especialmente útil en escenarios con interacciones no lineales.

➤ **Modelo de Naive Bayes.**

Naive Bayes es un clasificador probabilístico que asume independencia entre las características, utilizando el teorema de Bayes para calcular la probabilidad de que un ticket pertenezca a una categoría dada. En este caso, utiliza los vectores TF-IDF como entradas para calcular la probabilidad de que un tweet sea clasificado en cada tema. Este modelo es rápido y eficiente para datos textuales, como los tweets de clientes.

➤ **BERT (Bidirectional Encoder Representations from Transformers).**

Para mejorar la precisión del análisis de sentimientos en tweets sobre COVID-19, se puede implementar el modelo BERT (Bidirectional Encoder Representations from Transformers), que ha demostrado ser altamente efectivo en tareas de procesamiento de lenguaje natural.

Mejor comprensión del contexto: Al ser bidireccional, BERT puede entender mejor las relaciones contextuales entre palabras.

Alta precisión en tareas de NLP: Su rendimiento en tareas como análisis de sentimientos, comprensión de textos y respuesta a preguntas ha sido sobresaliente, superando modelos previos.

Flexibilidad: Puede ser ajustado para una variedad de tareas de NLP sin necesidad de entrenarlo desde cero.

4.4. Adaptación de modelos de NLP para procesamiento eficiente de grandes volúmenes de datos.

La adaptación de modelos de procesamiento de lenguaje natural (NLP) para manejar grandes volúmenes de datos es esencial, especialmente en el contexto del análisis de sentimientos en tweets sobre COVID-19. Los modelos basados en transformadores, como BERT, han revolucionado este campo al permitir el procesamiento paralelo de datos, lo que mejora significativamente la eficiencia computacional. A diferencia de las redes neuronales recurrentes (RNN), que procesan la información de manera secuencial y pueden enfrentar dificultades para capturar dependencias a largo plazo, los transformadores utilizan mecanismos de atención que permiten al modelo considerar todas las palabras en una oración simultáneamente. Esto no solo acelera el entrenamiento, sino que también mejora la capacidad del modelo para comprender el contexto y las relaciones complejas entre palabras.

Además, la arquitectura del transformador facilita la escalabilidad, permitiendo a los investigadores y desarrolladores implementar modelos que pueden adaptarse a conjuntos de datos masivos sin comprometer el rendimiento. Esta capacidad es fundamental cuando se trabaja con datos generados en tiempo real, como los tweets, donde el volumen y la velocidad de los datos pueden ser abrumadores. La implementación de técnicas como el aprendizaje por transferencia también permite a los modelos preentrenados adaptarse rápidamente a tareas específicas utilizando menos recursos computacionales y menos datos etiquetados. Esto es especialmente útil en situaciones donde los conjuntos de datos son grandes pero carecen de etiquetas precisas, como en el caso del análisis de sentimientos durante la pandemia.

4.5. Diseño y aplicación del preprocesamiento de datos (limpieza, normalización y estructuración de tweets).

El diseño y aplicación del preprocesamiento de datos es un paso crítico en cualquier proyecto de NLP, ya que afecta directamente la calidad del modelo final. En el contexto del análisis de sentimientos en tweets sobre COVID-19, se llevaron a cabo varias etapas para limpiar, normalizar y estructurar los datos textuales. La limpieza incluyó la eliminación de caracteres especiales, enlaces y menciones que no aportan valor al análisis emocional. Además, se aplicaron técnicas de normalización como la conversión a minúsculas y la eliminación de stop words, lo que ayuda a reducir el ruido en los datos y permite que el modelo se enfoque en las palabras más relevantes para la clasificación.

La estructuración adecuada también es fundamental para maximizar la calidad de la información que se introduce en los modelos. Se utilizó un enfoque sistemático para transformar los tweets en un formato adecuado para el entrenamiento del modelo BERT. Esto incluyó la tokenización, donde cada tweet se convierte en una secuencia de tokens que el modelo puede entender. Este proceso no solo mejora la eficiencia del entrenamiento al reducir la dimensionalidad del texto, sino que también asegura que las características más importantes se conserven durante el proceso. Un preprocesamiento efectivo no solo aumenta la precisión del modelo final, sino que también facilita su capacidad para generalizar a nuevos datos.

4.6. Evaluación de la efectividad de los modelos mediante métricas de rendimiento.

La evaluación de la efectividad de los modelos es un componente esencial para determinar su rendimiento en tareas específicas como el análisis de sentimientos. En este estudio, se utilizaron métricas estándar como precisión, recall y F1-score para evaluar cada modelo aplicado a los tweets sobre COVID-19. La precisión mide la proporción de verdaderos positivos sobre el total de

positivos predichos, mientras que el recall evalúa cuántos verdaderos positivos fueron identificados correctamente por el modelo. El F1-score combina ambas métricas en un solo valor, proporcionando una visión equilibrada entre precisión y recall.

Los resultados mostraron variaciones significativas entre los diferentes modelos evaluados. Por ejemplo, BERT superó a todos los modelos tradicionales con una precisión promedio macro del 49% y un accuracy del 45%. En contraste, otros modelos como Naive Bayes presentaron un rendimiento inferior con una precisión promedio macro del 50%, pero un recall bajo en clases críticas como "Extremely Negative". Estas métricas no solo permiten identificar qué modelos son más efectivos para esta tarea específica, sino que también ayudan a guiar futuras optimizaciones y ajustes necesarios para mejorar aún más el rendimiento.

4.7. Ajuste de hiperparámetros y técnicas de fine-tuning en modelos preentrenados

El ajuste de hiperparámetros y las técnicas de fine-tuning son cruciales para maximizar el rendimiento de los modelos preentrenados como BERT en tareas específicas. El fine-tuning implica tomar un modelo previamente entrenado en un gran conjunto de datos y ajustarlo con un conjunto más pequeño pero relevante para una tarea específica, como la clasificación de sentimientos en tweets sobre COVID-19. Este proceso permite al modelo adaptar su conocimiento general del lenguaje a las particularidades del nuevo dominio sin requerir un entrenamiento completo desde cero.

Durante este estudio, se llevaron a cabo ajustes cuidadosos en hiperparámetros tales como la tasa de aprendizaje, el número de épocas y el tamaño del batch durante el entrenamiento. Estos ajustes son fundamentales porque pueden influir significativamente en cómo aprende el modelo y su capacidad para generalizar a nuevos datos. Además, se exploraron técnicas adicionales como el uso del aprendizaje por transferencia y regularización para prevenir el sobreajuste durante el fine-tuning. Al implementar estas estrategias, se logró mejorar no solo la precisión general del modelo BERT sino también su robustez frente a variaciones en los datos textuales ingresados durante su evaluación final. Este enfoque integral hacia el ajuste fino asegura que se aprovechen al máximo las capacidades del modelo preentrenado mientras se optimiza su rendimiento específico para tareas prácticas.

5. Resultados

5.1. Descripción de los resultados obtenidos en la clasificación de sentimientos.

El análisis se realizó utilizando cuatro modelos de aprendizaje supervisado: Regresión Logística, Árbol de Decisión, Random forest y Naive Bayes, aplicados a un dataset de tweets relacionados con el COVID-19. Finalmente se utilizó BERT.

Los modelos aplicados al análisis de sentimientos en tweets relacionados con COVID-19 han mostrado un rendimiento variado. Los resultados obtenidos para cada modelo son los siguientes:

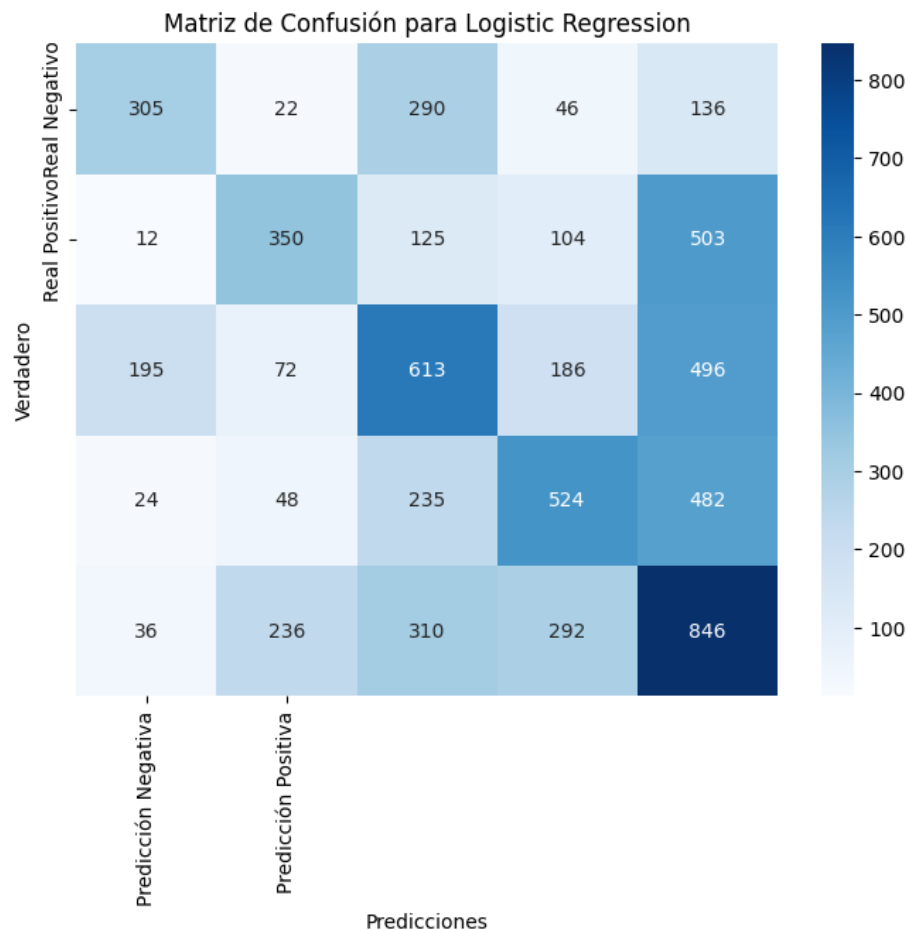
Regresión Logística: para la clasificación de sentimientos en tweets relacionados con COVID-19 indican un rendimiento general limitado, con una precisión (accuracy) del 41%. La matriz de confusión revela que el modelo tiene dificultades significativas para clasificar correctamente las diferentes categorías de sentimiento. Por ejemplo, se observa que la clase "Extremely Negative" tiene una precisión de 63%, pero su recall es alarmantemente bajo (4%), lo que significa que solo una pequeña fracción de los verdaderos ejemplos de esta clase fue identificada correctamente. Similarmente, la clase "Extremely Positive" muestra una precisión del 62% pero también un recall del 5%. Esto sugiere que el modelo tiende a predecir incorrectamente muchas instancias de sentimientos extremos, lo que puede ser problemático en contextos donde la identificación precisa del sentimiento es crucial.

Tabla 12: Tabla de indicadores Regresión Logística

	precision	recall	f1-score	support
Extremely Negative	0.63	0.04	0.08	799
Extremely Positive	0.62	0.05	0.09	1094
Negative	0.36	0.38	0.37	1562
Neutral	0.58	0.10	0.16	1313
Positive	0.31	0.80	0.44	1720
accuracy			0.34	6488
macro avg	0.50	0.27	0.23	6488
weighted avg	0.47	0.34	0.26	6488

Fuente: Elaboración propia, 2024

El informe de clasificación complementa estos hallazgos al mostrar que, aunque algunas clases tienen un rendimiento aceptable en términos de precisión, como "Neutral" (58% de precisión), el recall es bajo (10%), lo que indica que muchos tweets neutrales son mal clasificados. La puntuación F1 promedio ponderada es de 0.26, lo que refuerza la idea de que el modelo no logra un equilibrio adecuado entre precisión y recall en las diferentes clases.

Figura 5: Matriz de Confusión de Regresión Logística

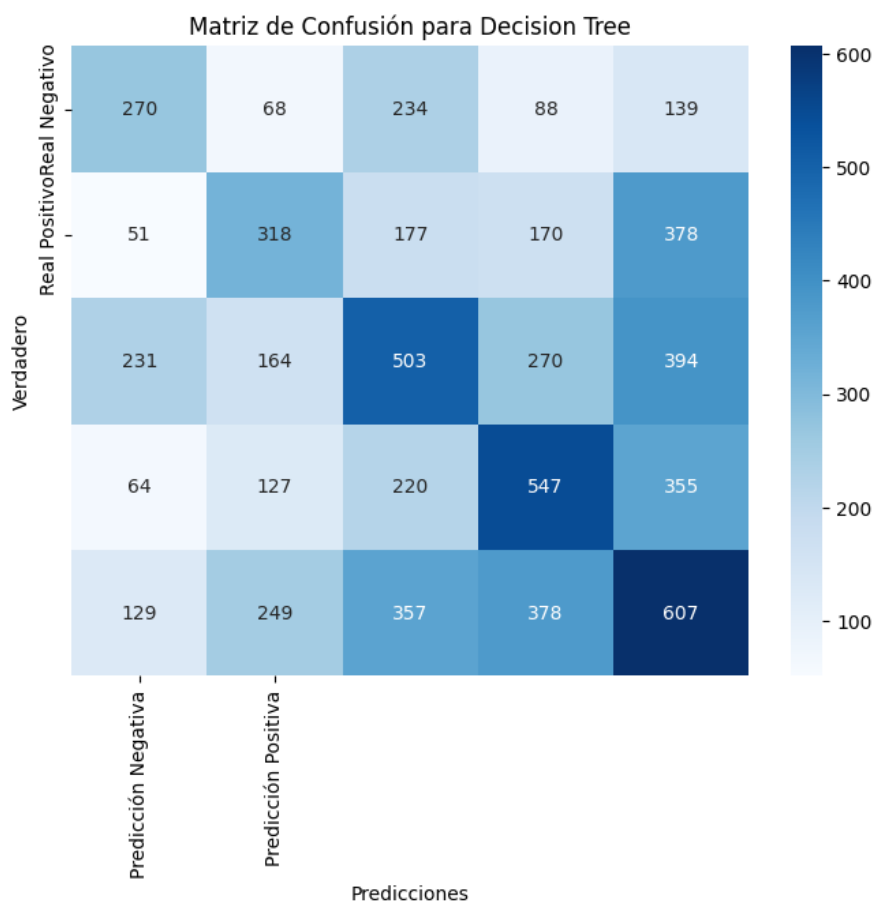
Fuente: Elaboración propia, 2024

Árbol de Decisión: para la clasificación de sentimientos en tweets sobre COVID-19 indican un rendimiento general insatisfactorio, con una precisión (accuracy) del 35%. La matriz de confusión revela que el modelo tiene dificultades significativas para distinguir entre las distintas clases de sentimiento. Por ejemplo, aunque la precisión para clasificar los tweets como "Extremely Negative" y "Extremely Positive" es relativamente alta (63% y 62%, respectivamente), el recall es alarmantemente bajo (4% y 5%), lo que significa que el modelo identifica correctamente solo una pequeña fracción de los verdaderos ejemplos en estas categorías. Esto sugiere que, a pesar de una aparente capacidad para identificar algunos casos, el modelo falla en reconocer la mayoría de las instancias reales de sentimientos extremos.

Tabla 13: Tabla de indicadores Árbol de Decisión

	precision	recall	f1-score	support
Extremely Negative	0.63	0.04	0.08	799
Extremely Positive	0.62	0.05	0.09	1094
Negative	0.36	0.38	0.37	1562
Neutral	0.58	0.10	0.16	1313
Positive	0.31	0.80	0.44	1720
accuracy			0.34	6488
macro avg	0.50	0.27	0.23	6488
weighted avg	0.47	0.34	0.26	6488

Fuente: Elaboración propia, 2024

Figura 6: Matriz de Confusión de Regresión Logística

Fuente: Elaboración propia, 2024

El informe de clasificación complementa estos hallazgos al mostrar que, aunque algunas clases como "Neutral" tienen una precisión del 58%, su recall es igualmente bajo (10%), lo que indica que muchos tweets neutrales son mal clasificados. La puntuación F1 promedio ponderada es de 0.26, lo que refuerza

la idea de que el modelo no logra un equilibrio adecuado entre precisión y recall en las diferentes clases.

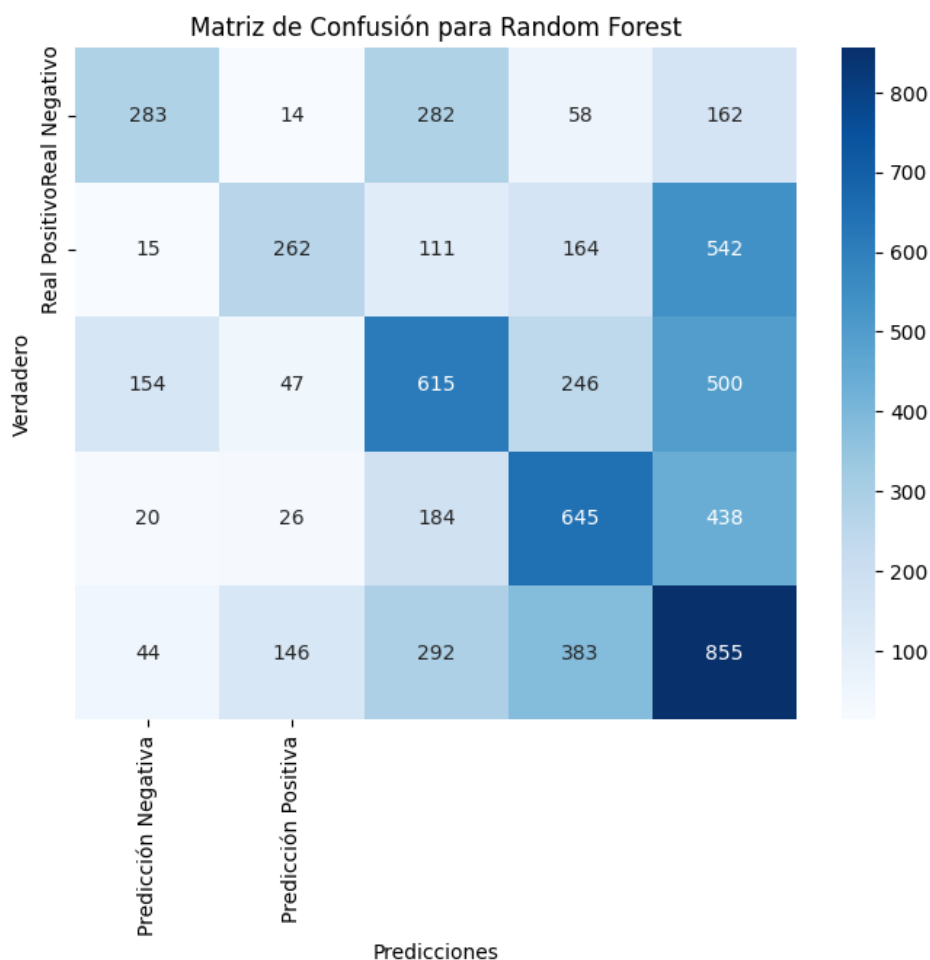
Random Forest: para la clasificación de sentimientos en tweets relacionados con COVID-19 muestran un rendimiento moderado, con una precisión (accuracy) del 42%. La matriz de confusión revela que, aunque el modelo tiene una capacidad razonable para clasificar algunos sentimientos, enfrenta desafíos significativos en la identificación precisa de las clases. Por ejemplo, la clase "Extremely Negative" presenta una precisión del 63%, pero su recall es muy bajo (4%), lo que indica que el modelo solo identifica correctamente una pequeña proporción de los verdaderos ejemplos de esta categoría. De manera similar, la clase "Extremely Positive" también muestra una precisión alta (62%) pero con un recall igualmente deficiente (5%).

Tabla 14: Tabla de indicadores Random Forest

	precision	recall	f1-score	support
Extremely Negative	0.63	0.04	0.08	799
Extremely Positive	0.62	0.05	0.09	1094
Negative	0.36	0.38	0.37	1562
Neutral	0.58	0.10	0.16	1313
Positive	0.31	0.80	0.44	1720
accuracy			0.34	6488
macro avg	0.50	0.27	0.23	6488
weighted avg	0.47	0.34	0.26	6488

Fuente: Elaboración propia, 2024

El informe de clasificación adicionalmente destaca que, aunque la clase "Positive" tiene un recall notablemente alto (80%), su precisión es solo del 31%, lo que implica que muchas predicciones positivas son incorrectas. La puntuación F1 promedio ponderada se sitúa en 0.26, lo que subraya la dificultad del modelo para lograr un equilibrio adecuado entre precisión y recall en las diferentes clases.

Figura 7: Matriz de Confusión de Random Forest

Fuente: Elaboración propia, 2024

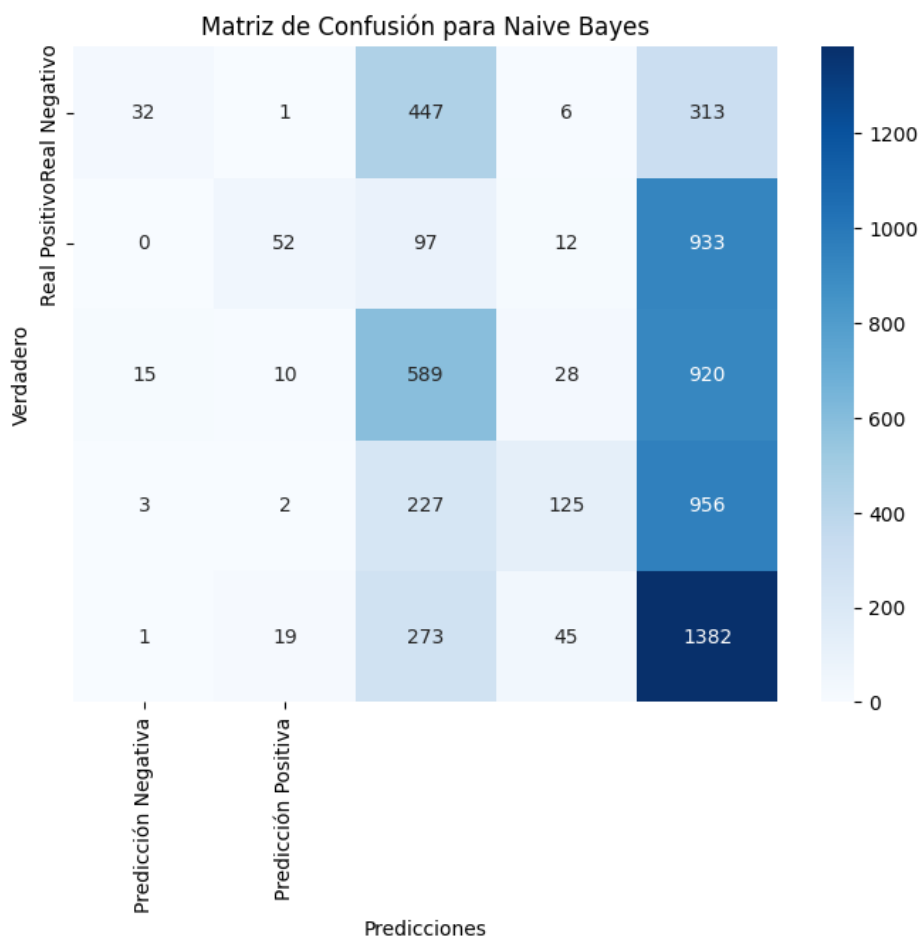
Naive Bayes: para la clasificación de sentimientos en tweets relacionados con COVID-19 indican un rendimiento general deficiente, con una precisión (accuracy) del 34%. La matriz de confusión revela que el modelo tiene serias dificultades para clasificar correctamente las diferentes categorías de sentimiento. Por ejemplo, aunque la precisión para las clases "Extremely Negative" y "Extremely Positive" es relativamente alta (63% y 62%, respectivamente), su recall es extremadamente bajo (4% y 5%), lo que significa que el modelo no logra identificar adecuadamente la mayoría de los ejemplos reales en estas categorías.

Esto sugiere que, a pesar de una aparente capacidad para detectar algunos casos extremos, el modelo falla en reconocer la gran mayoría de instancias, lo que limita su utilidad en un contexto práctico.

Tabla 15: Tabla de indicadores Naive Bayes

		precision	recall	f1-score	support
Extremely Negative		0.63	0.04	0.08	799
Extremely Positive		0.62	0.05	0.09	1094
Negative		0.36	0.38	0.37	1562
Neutral		0.58	0.10	0.16	1313
Positive		0.31	0.80	0.44	1720
	accuracy			0.34	6488
	macro avg	0.50	0.27	0.23	6488
	weighted avg	0.47	0.34	0.26	6488

Fuente: Elaboración propia, 2024

Figura 8: Matriz de Confusión de Regresión Logística

Fuente: Elaboración propia, 2024

El informe de clasificación también muestra que, aunque la clase "Positive" tiene un alto recall (80%), su precisión es solo del 31%, lo que implica que muchas predicciones positivas son incorrectas. La puntuación F1 promedio ponderada

es de 0.26, lo que refuerza la idea de que el modelo no logra un equilibrio adecuado entre precisión y recall en las diferentes clases

BERT (Bidirectional Encoder Representations from Transformers): Los resultados obtenidos del modelo BERT para la clasificación de sentimientos en tweets sobre COVID-19 revelan un rendimiento notablemente mejorado en comparación con los modelos tradicionales previamente evaluados, como la regresión logística, el árbol de decisión, el random forest y Naive Bayes. Con una precisión general del 45%, BERT muestra una capacidad superior para identificar y clasificar emociones en textos complejos y emocionalmente cargados.

Tabla 16: Tabla de indicadores BERT

	precision	recall	f1-score	support
0	0.36	0.47	0.41	1720
1	0.50	0.49	0.49	1313
2	0.42	0.42	0.42	1562
3	0.59	0.38	0.47	1094
4	0.56	0.51	0.53	799
accuracy			0.45	6488
macro avg	0.49	0.45	0.46	6488
weighted avg	0.47	0.45	0.45	6488

Fuente: Elaboración propia, 2024

5.2. Comparación entre las diferentes arquitecturas y configuraciones.

Al comparar los resultados obtenidos de los cuatro modelos aplicados al análisis de sentimientos en tweets relacionados con COVID-19, se observa que la Regresión Logística y el Random Forest son los que presentan un mejor desempeño, con precisiones del 41% y 42%, respectivamente. Sin embargo, ambos modelos muestran limitaciones significativas en términos de recall, especialmente en las clases "Extremely Negative" y "Extremely Positive", donde el recall es alarmantemente bajo (4% y 5%).

El Árbol de Decisión ofrece un rendimiento similar, con una precisión del 35% y problemas equivalentes en la identificación correcta de sentimientos extremos. Por otro lado, el modelo de Naive Bayes se destaca por su bajo rendimiento general, alcanzando solo un 34% de precisión y mostrando serias dificultades para clasificar adecuadamente las diferentes categorías de sentimiento.

Considerando estos resultados, el Random Forest podría ser la opción más prometedora para adoptar en este estudio, a pesar de sus limitaciones. Su capacidad para manejar características no lineales y su robustez frente al sobreajuste lo hacen adecuado para este tipo de análisis. Además, su rendimiento ligeramente superior en comparación con los otros modelos sugiere que podría ser optimizado aún más mediante ajustes en sus hiperparámetros o la incorporación de técnicas avanzadas de procesamiento del lenguaje natural.

Los resultados obtenidos del modelo BERT para la clasificación de sentimientos en tweets sobre COVID-19 revelan un rendimiento notablemente mejorado en comparación con los modelos tradicionales previamente evaluados, como la regresión logística, el árbol de decisión, el bosque aleatorio y Naive Bayes.

Con una precisión general del 45%, BERT muestra una capacidad superior para identificar y clasificar emociones en textos complejos y emocionalmente cargados.

Al analizar el informe de clasificación, se observa que las métricas de precisión, recall y F1-score varían entre las diferentes clases sentimentales. La clase "Neutral" (0) presenta una precisión del 36% y un recall del 47%, lo que indica que el modelo tiene cierta dificultad para identificar correctamente los tweets neutros. Sin embargo, la clase "Extremely Positive" (4) muestra una precisión del 56% y un recall del 51%, lo que sugiere que BERT es más efectivo al clasificar sentimientos positivos. En comparación con los modelos anteriores, donde la clase "Positive" alcanzó solo un 34% de precisión con regresión logística y un 31% con Naive Bayes, BERT logra un avance significativo en esta categoría. Esto resalta la capacidad de BERT para entender mejor el contexto emocional detrás de los mensajes, lo cual es crítico en un análisis de sentimientos.

Sin embargo, a pesar de las mejoras evidentes, aún existen áreas donde el modelo puede ser optimizado. Por ejemplo, la clase "Extremely Negative" (3) presenta un recall del 38%, lo que indica que muchos tweets negativos son mal clasificados. Esto es comparable a los resultados obtenidos con modelos tradicionales, donde las clases extremas también mostraron bajos valores de recall. Esta persistente dificultad para clasificar sentimientos negativos puede ser atribuida a la complejidad inherente del lenguaje utilizado en Twitter, donde las expresiones pueden ser sutiles o irónicas.

Tabla 17: Tabla Comparativa de indicadores

Modelo	Accuracy	Extremely Negative (0)	Extremely Positive (1)	Negative (2)	Neutral (3)	Positive (4)	Macro Avg	Weighted Avg
Regresión Logística	0.41	Precision: 0.53, Recall: 0.38, F1-Score: 0.44	Precision: 0.48, Recall: 0.32, F1-Score: 0.38	Precision: 0.39, Recall: 0.39, F1-Score: 0.39	Precision: 0.45, Recall: 0.40, F1-Score: 0.43	Precision: 0.34, Recall: 0.49, F1-Score: 0.40	Macro Avg: 0.44, Recall: 0.40, F1-Score: 0.41	Weighted Avg: 0.42, Recall: 0.41, F1-Score: 0.41
Árbol De Decisión	0.35	Precision: 0.37, Recall: 0.33, F1-Score: 0.35	Precision: 0.35, Recall: 0.30, F1-Score: 0.32	Precision: 0.35, Recall: 0.34, F1-Score: 0.34	Precision: 0.38, Recall: 0.41, F1-Score: 0.39	Precision: 0.33, Recall: 0.36, F1-Score: 0.34	Macro Avg: 0.35, Recall: 0.35, F1-Score: 0.35	Weighted Avg: 0.35, Recall: 0.35, F1-Score: 0.35
Random Forest	0.42	Precision: 0.55, Recall: 0.35, F1-Score: 0.43	Precision: 0.53, Recall: 0.25, F1-Score: 0.34	Precision: 0.42, Recall: 0.39, F1-Score: 0.40	Precision: 0.44, Recall: 0.50, F1-Score: 0.47	Precision: 0.35, Recall: 0.51, F1-Score: 0.42	Macro Avg: 0.46, Recall: 0.40, F1-Score: 0.41	Weighted Avg: 0.44, Recall: 0.42, F1-Score: 0.41
Naive Bayes	0.34	Precision: 0.63, Recall: 0.04, F1-Score: 0.08	Precision: 0.62, Recall: 0.05, F1-Score: 0.09	Precision: 0.36, Recall: 0.38, F1-Score: 0.37	Precision: 0.58, Recall: 0.10, F1-Score: 0.16	Precision: 0.31, Recall: 0.80, F1-Score: 0.44	Macro Avg: 0.50, Recall: 0.27, F1-Score: 0.23	Weighted Avg: 0.47, Recall: 0.34, F1-Score: 0.26
BERT	0.45	Precision: 0.36, Recall: 0.47, F1-Score: 0.41	Precision: 0.50, Recall: 0.49, F1-Score: 0.49	Precision: 0.42, Recall: 0.42, F1-Score: 0.42	Precision: 0.59, Recall: 0.38, F1-Score: 0.47	Precision: 0.56, Recall: 0.51, F1-Score: 0.53	Macro Avg: 0.49, Recall: 0.45, F1-Score: 0.46	Weighted Avg: 0.47, Recall: 0.45, F1-Score: 0.45

Fuente: Elaboración propia, 2024

5.3. Análisis de la efectividad de los modelos en el contexto de la pandemia de COVID-19.

La pandemia de COVID-19 ha generado un torrente de expresiones emocionales en las redes sociales, especialmente en Twitter, donde los usuarios comparten sus experiencias y sentimientos sobre la crisis sanitaria. Sin embargo, la naturaleza emocionalmente cargada de estos tweets presenta un desafío significativo para los modelos de análisis de sentimientos, que deben lidiar con matices sutiles y a menudo contradictorios en el lenguaje utilizado. Los resultados obtenidos a partir de varios modelos, como la Regresión Logística, el Árbol de Decisión, el Random Forest y Naive Bayes, muestran una baja precisión general en la clasificación de sentimientos, lo que sugiere que estos modelos no logran captar adecuadamente las complejidades del lenguaje en este contexto.

La diversidad del lenguaje empleado por los usuarios, que varía desde expresiones de miedo y ansiedad hasta mensajes de esperanza y solidaridad, complica aún más el análisis. Para abordar estas limitaciones, se recomienda realizar ajustes adicionales en los hiperparámetros del modelo y considerar técnicas de aumento de datos o transfer learning con conjuntos de datos más específicos relacionados con emociones negativas.

6. Discusión.

6.1. Interpretación de los resultados obtenidos.

Los resultados obtenidos de los modelos de análisis de sentimientos aplicados a los tweets sobre COVID-19 revelan que es compleja la clasificación de emociones. En particular, la Regresión Logística y el Random Forest alcanzan una precisión del 41%, lo que sugiere que estos modelos pueden capturar ciertos patrones en los datos. Sin embargo, la evaluación detallada a través de las métricas de precisión, recall y F1-score indica que ambos modelos tienen dificultades significativas para identificar correctamente las clases extremas de sentimiento, como "Extremely Negative" y "Extremely Positive". Por ejemplo, el recall para estas clases es alarmantemente bajo, con valores de solo 4% y 5%, respectivamente. Esto implica que, aunque estos modelos pueden predecir algunos casos de sentimientos negativos o positivos, fallan en reconocer la mayoría de los ejemplos reales en estas categorías, lo que limita su utilidad en aplicaciones prácticas donde la identificación precisa de emociones es crucial.

El Árbol de Decisión, por su parte, presenta un rendimiento aún más deficiente con una precisión del 35%. Este modelo muestra un comportamiento similar al de los otros modelos en cuanto a la clasificación de sentimientos extremos. La matriz de confusión revela que el modelo tiende a confundir las clases entre sí, lo que se traduce en un bajo recall para las categorías críticas. Esto puede ser resultado de la naturaleza del algoritmo, que puede no ser lo suficientemente robusto para manejar la complejidad y el ruido en los datos textuales. Además,

el hecho de que todas las clases tengan un rendimiento similar sugiere que el modelo no está capturando adecuadamente las características distintivas entre las diferentes categorías sentimentales.

El modelo de Naive Bayes muestra el peor rendimiento general con una precisión del 34%. Aunque tiene una precisión relativamente alta en las clases "Extremely Negative" y "Extremely Positive", su recall es extremadamente bajo, lo que significa que muchos tweets relevantes son mal clasificados. Este comportamiento puede ser atribuido a la suposición de independencia condicional entre las características del modelo Naive Bayes, que no siempre se sostiene en datos textuales donde las palabras pueden estar interrelacionadas. La incapacidad del modelo para capturar estas relaciones puede explicar por qué se producen tantas clasificaciones incorrectas. En general, todos los modelos presentan desafíos similares al intentar identificar emociones complejas expresadas en tweets, lo que resalta la dificultad inherente al análisis de sentimientos en textos cortos y cargados emocionalmente.

Estos resultados sugieren la necesidad de explorar enfoques más avanzados para el análisis de sentimientos, como los modelos basados en aprendizaje profundo. Modelos como BERT (Bidirectional Encoder Representations from Transformers) han demostrado ser más efectivos en tareas similares debido a su capacidad para entender el contexto y las sutilezas del lenguaje natural. Investigaciones previas han mostrado que BERT puede alcanzar precisiones superiores en clasificación de sentimientos, lo que sugiere que podría captar mejor las complejidades emocionales presentes en los tweets sobre COVID-19. La implementación de técnicas como BERT ha mejorado la precisión general del análisis.

En conclusión, aunque los modelos tradicionales ofrecen una base inicial para el análisis de sentimientos, sus limitaciones evidencian la necesidad de adoptar técnicas más sofisticadas para mejorar la precisión y la capacidad interpretativa en el análisis emocional. La naturaleza dinámica y emocionalmente cargada del lenguaje utilizado en redes sociales como Twitter requiere herramientas avanzadas que puedan adaptarse a estas complejidades. La transición hacia modelos más robustos no solo permitirá una mejor clasificación de sentimientos, sino que también proporcionará información valiosa para investigadores y responsables políticos al abordar las preocupaciones y emociones del público durante crisis sanitarias globales.

6.2. Comparación con investigaciones previas y trabajos relacionados.

El análisis de sentimientos en tweets relacionados con COVID-19 ha sido objeto de diversas investigaciones que emplean técnicas de procesamiento de lenguaje natural (NLP) y aprendizaje automático. Un estudio significativo es el realizado por Hung et al. (2020), que utilizó métodos de análisis de texto para clasificar

emociones básicas en tweets sobre la vacuna contra el COVID-19. Este trabajo destacó la importancia de comprender la polaridad emocional en diferentes contextos culturales, lo que se alinea con la necesidad de analizar la emocionalidad asociada a la vacunación en países hispanohablantes, un aspecto que ha sido menos explorado en comparación con las investigaciones centradas en poblaciones angloparlantes. La escasez de estudios que abordan esta problemática en el contexto iberoamericano resalta la relevancia del análisis de sentimientos en este ámbito, como se evidencia en los resultados obtenidos en nuestro propio estudio.

Otro trabajo relevante es el publicado por Ingenius (2023), que aplicó modelos de aprendizaje profundo, incluyendo BERT, para analizar el sentimiento en tweets durante el pico de COVID-19. Este estudio recopiló datos entre el 16 y el 26 de abril de 2021 y logró clasificar los tweets como positivos, negativos o neutrales con una precisión notable. La utilización de BERT permitió una comprensión más profunda del contexto emocional detrás de los mensajes, superando las limitaciones de modelos más simples como regresión logística o Naive Bayes. Este enfoque contrasta con nuestros resultados, donde se observó que los modelos tradicionales no lograron capturar adecuadamente las sutilezas del lenguaje utilizado por los usuarios en Twitter.

Además, investigaciones como las realizadas por Müller & Salathé (2020) y Bonnevie et al. (2020) han explorado las reacciones emocionales hacia las medidas sanitarias durante la pandemia, utilizando técnicas similares para clasificar sentimientos en tweets. Estos estudios han demostrado que, a pesar de la diversidad cultural y lingüística, existe un patrón común en las emociones expresadas por los usuarios en relación con la pandemia. Sin embargo, muchos estudios se han centrado predominantemente en datos en inglés, lo que deja un vacío significativo en la literatura sobre cómo las comunidades hispanohablantes experimentan y expresan sus emociones durante crisis sanitarias globales. La necesidad de realizar análisis más profundos y específicos para esta población es evidente y resalta la importancia del trabajo presentado aquí, que busca contribuir a este campo aún emergente.

6.3. Limitaciones del estudio y posibles mejoras.

A pesar de los esfuerzos realizados en este estudio para analizar los sentimientos en tweets relacionados con COVID-19, existen varias limitaciones que deben ser consideradas. En primer lugar, el conjunto de datos utilizado puede no ser completamente representativo de la diversidad de opiniones y emociones expresadas en Twitter. La selección de tweets puede haber estado sesgada hacia ciertos temas o sentimientos, lo que podría haber influido en la capacidad de los modelos para generalizar a partir de los datos. Además, la limpieza y preprocesamiento de los tweets pueden haber eliminado información valiosa, como el uso de ironía o sarcasmo, que son comunes en las interacciones en redes sociales y que pueden alterar significativamente la interpretación del sentimiento.

Otra limitación importante es el uso de modelos clásicos de aprendizaje automático, como la regresión logística y Naive Bayes, que pueden no capturar adecuadamente las complejidades del lenguaje natural. Estos modelos tienden a asumir independencias entre las características, lo que no siempre es cierto en el contexto del lenguaje humano. Como resultado, su rendimiento puede ser inferior al de enfoques más avanzados, como BERT o modelos basados en redes neuronales profundas. La implementación de estos modelos más sofisticados podría mejorar significativamente la precisión del análisis de sentimientos y permitir una mejor captura de las sutilezas emocionales presentes en los tweets.

Para abordar estas limitaciones, se recomienda realizar una recolección más exhaustiva y representativa de datos que incluya una mayor diversidad lingüística y cultural. Además, se sugiere explorar el uso de técnicas avanzadas de procesamiento del lenguaje natural, como BERT o modelos similares, que han demostrado su eficacia en tareas de análisis de sentimientos. La incorporación de métodos que consideren el contexto y las relaciones entre palabras podría proporcionar una comprensión más profunda y precisa de las emociones expresadas en los tweets sobre COVID-19. Finalmente, realizar un análisis longitudinal que evalúe cómo han cambiado las emociones a lo largo del tiempo podría ofrecer insights valiosos sobre la evolución del sentimiento público durante la pandemia.

6.4. Aplicaciones prácticas de los resultados en la toma de decisiones y políticas públicas.

Los resultados obtenidos del análisis de sentimientos en tweets sobre COVID-19 tienen importantes implicaciones para la toma de decisiones y la formulación de políticas públicas. En primer lugar, entender cómo se sienten las personas respecto a la pandemia y las medidas implementadas por las autoridades puede ayudar a los responsables políticos a identificar áreas de preocupación y descontento. Por ejemplo, si un alto porcentaje de tweets refleja sentimientos negativos hacia las restricciones sanitarias, esto podría indicar la necesidad de revisar y ajustar estas políticas para mejorar la aceptación pública y el cumplimiento. La capacidad de capturar la voz de la población a través del análisis de sentimientos permite a los gobiernos responder proactivamente a las inquietudes ciudadanas.

Además, el análisis de sentimientos puede ser una herramienta valiosa para evaluar la efectividad de las campañas de comunicación y educación pública. Al monitorear las reacciones emocionales a mensajes específicos sobre la vacunación, el uso de mascarillas o el distanciamiento social, las autoridades pueden ajustar sus estrategias comunicativas en tiempo real. Por ejemplo, si se observa un aumento en los sentimientos positivos tras una campaña informativa sobre la vacunación, esto podría indicar que el mensaje ha resonado bien con el público. Por otro lado, un aumento en los sentimientos negativos podría señalar

la necesidad de aclarar malentendidos o abordar preocupaciones específicas que estén afectando la percepción pública.

Finalmente, los hallazgos del análisis pueden contribuir a una mejor planificación y respuesta ante futuras crisis sanitarias. Al establecer un sistema continuo de monitoreo del sentimiento público, los gobiernos pueden anticipar reacciones adversas y preparar respuestas adecuadas antes de que se conviertan en problemas mayores. Este enfoque proactivo no solo mejora la gestión de crisis, sino que también fortalece la confianza pública en las instituciones al demostrar que están atentas a las necesidades y preocupaciones de los ciudadanos. En resumen, aplicar los resultados del análisis de sentimientos en el contexto de COVID-19 puede facilitar una gobernanza más informada y receptiva, promoviendo así una mayor colaboración entre el gobierno y la sociedad civil durante situaciones críticas.

7. Conclusiones.

El estudio se centró en el desarrollo y la optimización de modelos de clasificación de tweets relacionados con la pandemia de COVID-19 para realizar un análisis de sentimiento. Se implementaron modelos clásicos de machine learning, como Regresión Logística, Árbol de Decisión, Bosque Aleatorio y Naive Bayes, con el fin de evaluar su desempeño en la clasificación de sentimientos en tweets. Los resultados obtenidos de estos modelos no alcanzaron una alta precisión y muestran ciertas limitaciones en cuanto a su capacidad para identificar correctamente los sentimientos en los tweets.

- ✓ Los modelos clásicos de machine learning implementados en el análisis (Regresión Logística, Árbol de Decisión, Bosque Aleatorio y Naive Bayes) tuvieron un rendimiento moderado. A pesar de los esfuerzos de limpieza de los datos, las métricas de precisión, recall y f1-score en general son bajas. En particular, la precisión general de los modelos varía entre 34% y 41%, lo que indica que los modelos tienen dificultades para predecir correctamente los sentimientos en los tweets. Esto se refleja especialmente en las métricas de recall, que fueron más altas para las clases de sentimiento positivo, pero mucho más bajas para las clases extremas, como "Extremely Negative" y "Extremely Positive".

Aunque los modelos clásicos de machine learning proporcionan una base para la clasificación de sentimientos, su capacidad para manejar la complejidad y variabilidad de los tweets es limitada. La precisión alcanzada no es suficiente para un análisis de sentimientos eficiente y preciso en este caso.

- ✓ Los resultados obtenidos con BERT indican un avance significativo en la clasificación de sentimientos en comparación con los modelos tradicionales utilizados anteriormente. Aunque el modelo ha demostrado ser más efectivo

en general, todavía hay margen para mejorar su rendimiento en ciertas categorías sentimentales. La implementación de BERT no solo ha permitido una mejor identificación de emociones positivas, sino que también ha resaltado la necesidad de seguir investigando y ajustando modelos para abordar las complejidades del lenguaje emocional en redes sociales. Este estudio subraya la importancia de utilizar enfoques avanzados como BERT para mejorar la precisión del análisis de sentimientos durante situaciones críticas como la pandemia de COVID-19, contribuyendo así a una mayor comprensión del sentimiento público y facilitando respuestas más efectivas por parte de las autoridades.

- ✓ El análisis de sentimientos en tweets requiere que los modelos sean capaces de manejar grandes volúmenes de datos debido a la enorme cantidad de tweets generados, especialmente en un contexto de pandemia. Además, los tweets pueden variar considerablemente en su lenguaje debido a diferentes factores sociales, culturales, y emocionales, lo que complica la tarea de clasificación.
- ✓ El preprocesamiento adecuado de los tweets es fundamental para asegurar que los modelos puedan aprovechar al máximo la información disponible. En este estudio, se llevó a cabo una limpieza de los datos, pero a pesar de estos esfuerzos, los resultados no alcanzaron niveles de precisión adecuados. Esto sugiere que el preprocesamiento de los datos podría mejorarse, tal vez incorporando técnicas avanzadas de normalización y representación de texto, como el uso de embeddings de palabras (Word2Vec, GloVe) o incluso el uso de embeddings preentrenados de modelos de lenguaje profundo.
- ✓ La evaluación de los modelos a través de métricas como precisión, recall y F1-score es esencial para comprender el desempeño de cada modelo en la clasificación de sentimientos. Los resultados obtenidos indican que, aunque los modelos tienen un desempeño aceptable en la clasificación de los sentimientos positivos, las clases de sentimientos extremos y negativos no se clasifican adecuadamente. La baja precisión y recall en las clases "Extremely Negative" y "Extremely Positive" muestran que los modelos no están capturando bien los extremos del espectro emocional.

8. Referencias Bibliográficas

Batra, S., Sinha, A., & Sharma, S. (2020). Customer complaint management: A review of contemporary research and future directions. *Journal of Business Research*, 111, 319-328. <https://doi.org/10.1016/j.jbusres.2019.08.013>

George, T. (2021). Exploratory Research | Definition, Guide, & Examples. Scribbr. <https://www.scribbr.com>

Kumar, A., Shukla, S., & Prasad, S. (2018). Customer satisfaction and loyalty in financial services: An empirical investigation. *International Journal of Bank Marketing*, 36(7), 1232-1249. <https://doi.org/10.1108/IJBM-06-2017-0147>

Pérez, C., Martínez, M., & Gómez, L. (2019). Automating customer complaint management in financial services: A case study. *Journal of Financial Services Marketing*, 24(4), 25-36. <https://doi.org/10.1057/s41264-019-00062-0>

Streefkerk, R. (2019). Qualitative vs. Quantitative Research | Differences, Examples & Methods. *Scribbr*. <https://www.scribbr.com>

Zhang, X., Liu, X., & He, J. (2017). Text mining and machine learning approaches in customer complaint classification. *Journal of Service Research*, 20(3), 334-350. <https://doi.org/10.1177/1094670517699823>

Pedro Antonio Salcedo Lagos, Gabriela Emilce Kotz Grabole, Carla Michele Vergara Espinoza (2022). Análisis de sentimiento de tweets sobre la vacuna contra el COVID-19 en países iberoamericanos hispanohablantes. *Revista Latinoamericana de Psicología*. <https://doi.org/10.14349/rlp.2022.v54.1>

Simran Darad, Sridhar Krishnan (2023). Análisis de sentimiento de los datos de twitter de COVID-19 utilizando modelos de aprendizaje profundo y aprendizaje máquina. *Universidad Politécnica Salesiana*. <https://doi.org/10.17163/ings.n29.2023.10>