



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Modelos de aprendizaje automático en análisis de
sentimiento: comparativa de rendimiento

Trabajo Fin de Grado

Grado en Ingeniería Informática

AUTOR/A: Navarro Bellido, Joan

Tutor/a: Ferri Ramírez, César

CURSO ACADÉMICO: 2022/2023

“Computers are incredibly fast, accurate, and stupid. Human beings are incredibly slow, inaccurate, and brilliant. Together they are powerful beyond imagination.”

Albert Einstein (1879-1955)

Agradecimientos

Son tantas a las personas que podría agradecer que siento que lo que pueda decir se quedará corto. Aunque lo que pueda quedar escrito en este trabajo sea insuficiente, mi infinita gratitud siempre queda con todos ellos.

En primer lugar, mi agradecimiento va dirigido a mi tutor Cèsar Ferri, por saber dirigirme hacia un objetivo concreto y ser un apoyo a lo largo del trabajo.

A la UPV, por haberme brindado la oportunidad de aprender aquí, rodeado de gente extraordinaria, tanto compañeros como profesores.

A Finlandia y sobre todo a la ciudad de Turku, que me ha permitido crecer enormemente y redescubrir un cariño hacia la informática y su profesión que creía que estaba perdiendo. Gracias a toda esa gente de Turku y agradecer, en particular, la inestimable compañía de mi amigo Rober, donde pasando horas en la cocina, entre esas noches prácticamente infinitas y gélidas es donde se encuentra el origen de la motivación de este trabajo.

A mis padres, ese baluarte férreo, inamovibles ante las adversidades de la vida, os debo mis ganas de continuar y no rendirme, sois para mí un ejemplo a seguir.

A mi hermana pequeña María, donde en mi afán de intentar ser tu guía, haces que quiera ser mejor en todo lo que hago.

¡Gracias a todos!

Joan

Resumen

El estudio se embarca en una misión para explorar el mundo del análisis de sentimientos en el contexto de Twitter, una destacada plataforma de redes sociales que sirve como termómetro global de la opinión pública. Reconociendo los desafíos en el análisis de los vastos y complejos datos de los tweets, la investigación evalúa varios modelos de aprendizaje automático, incluyendo redes neuronales, árboles de decisión y clasificadores bayesianos, utilizando el conjunto de datos “sentiment140” de la Universidad de Stanford. El propósito central es evaluar el rendimiento relativo de estos modelos en la detección y clasificación de sentimientos en los tweets, proporcionando una perspectiva clara y objetiva. El trabajo también tiene como objetivo ofrecer una orientación valiosa para futuras investigaciones y aplicaciones prácticas en el análisis de sentimientos. El documento está estructurado en secciones que incluyen Introducción, Estado del Arte, Análisis del Problema, Diseño, Desarrollo, Implementación, Pruebas, Resultados y Discusión, ofreciendo una visión completa de las metodologías, experimentos y hallazgos.

Palabras clave: Análisis de Sentimientos, Modelos de Aprendizaje automático, Datos de Twitter, Evaluación Comparativa, Redes Neuronales, Árboles de Decisión, Clasificadores Bayesianos, Evaluación de Rendimiento, Procesamiento de Lenguaje Natural, Análisis de Redes Sociales.



Abstract

The study embarks on a mission to explore the world of sentiment analysis within the context of Twitter, a prominent social media platform that serves as a global thermometer of public opinion. Recognizing the challenges in analyzing the vast and complex data of tweets, the research evaluates various machine learning models, including neural networks, decision trees, and Bayesian classifiers, using the “sentiment140” dataset from Stanford University. The central purpose is to assess the relative performance of these models in detecting and classifying sentiments in tweets, providing a clear and objective perspective. The work also aims to offer valuable guidance for future research and practical applications in sentiment analysis. The document is structured into sections including Introduction, State of the Art, Problem Analysis, Design, Development, Implementation, Testing, Results, and Discussion, providing a comprehensive insight into the methodologies, experiments, and findings.

Keywords: Sentiment Analysis, Machine Learning Models, Twitter Data, Comparative Evaluation, Neural Networks, Decision Trees, Bayesian Classifiers, Performance Assessment, Natural Language Processing, Social Media Analytics.

Tabla de contenidos

1.	Introducción	12
1.1	Contexto y Motivación.....	12
1.2	Objetivo del estudio	13
1.3	Contribuciones y alcance	13
2.	Estado del Arte.....	16
2.1	Conceptos básicos	16
2.1.1	Análisis de sentimientos.....	16
2.1.2	Aprendizaje Automático	16
2.1.3	Clasificación.....	17
2.1.4	Redes Neuronales.....	17
2.1.5	Métricas de Evaluación	17
2.1.6	Dataset.....	17
2.2	Historia y Evolución del Análisis de Sentimientos	18
2.2.1	Inicios y Métodos Primitivos	18
2.2.2	Explosión de Datos y Aprendizaje Automático	18
2.2.3	Surgimiento de las Redes Neuronales y el Aprendizaje Profundo.....	18
2.2.4	Desafíos Contemporáneos y Perspectivas Futuras	18
2.3	Técnicas Tradicionales vs. Machine Learning	18
2.3.1	Técnicas Tradicionales.....	19
2.3.2	Técnicas de Machine Learning.....	19
2.4	Procesamiento del Lenguaje Natural en el Análisis de Sentimientos	20
2.5	Avances Recientes y Tendencias Actuales	21
3.	Análisis del Problema.....	23
3.1	Definición del problema.....	23
3.2	Importancia del Análisis de Sentimientos	23
3.3	Exploración y Características del Dataset Sentiment140.....	24
3.3.1	Características y Estructura del Dataset Sentiment140	25
3.4	Retos y Desafíos del Análisis de Sentimientos en inglés.....	26
4.	Diseño	29
4.1	Diseño Experimental	29
4.1.1	Importaciones	29

4.1.2	Carga y Preparación de Datos	30
4.1.3	Vectorización	31
4.1.4	División de Datos	32
4.2	Elección de Modelos y Técnicas	32
4.2.1	Random Forest Model	32
4.2.2	Multi-layer Perceptron	33
4.2.3	Regresión Logística	34
4.2.4	Máquinas de Soporte Vectorial	35
4.2.5	Naive Bayes Multinomial	36
4.3	Diseño de Evaluación y Métricas	37
4.3.1	Precisión (Precisión Positiva)	38
4.3.2	Recuperación (Sensibilidad)	38
4.3.3	Puntuación F1	38
5.	Desarrollo	39
5.1	Preparación de Datos	39
5.1.1	Carga de Datos	39
5.1.2	Limpieza de Datos	40
5.1.3	División de Datos	41
6.	Implementación	43
6.1	Integración de Modelos y Herramientas	43
6.1.1	Selección de Modelos	44
6.1.2	Bibliotecas y Herramientas	44
6.1.3	Implementación de Random Forest Model	45
6.1.4	Implementación de Multi-Layer Perceptron	45
6.1.5	Implementación de Regresión Logística	46
6.1.6	Implementación de Máquinas de Soporte Vectorial	47
6.1.7	Implementación de Naive Bayes Multinomial	47
6.1.8	Implementación de los modelos	48
7.	Experimentos	50
7.1	Metodología de Pruebas	50
7.1.1	División de Datos	50
7.1.2	Métricas de Evaluación	51
7.1.3	Comparación y Análisis de Modelos	52
8.	Resultados y Discusión	54



8.1	Resultados	54
8.1.1	Análisis de Random Forest Model	55
8.1.2	Análisis de Multi-Layer Perceptron	55
8.1.3	Análisis del Modelo de Regresión Logística.....	55
8.1.4	Análisis de Máquinas de Soporte Vectorial	56
8.1.5	Análisis de Multinomial Naive Bayes.....	56
8.1.6	Tabla Comparativa de Modelos	57
8.1.7	Gráficas de Métricas Individuales de los Modelos.....	58
8.1.8	Gráfica de Barras de las Métricas de Cada Modelo	58
8.2	Discusión.....	59
8.2.1	Interpretación de Random Forest Model.....	59
8.2.2	Interpretación de Multi-Layer Perceptron.....	60
8.2.3	Interpretación del Modelo de Regresión Logística	60
8.2.4	Interpretación de Máquinas de Soporte Vectorial	61
8.2.5	Interpretación de Multinomial Naive Bayes.....	61
9.	Conclusiones	64
9.1	Resumen de Hallazgos	64
9.2	Lecciones Aprendidas	65
9.3	Limitaciones y Desafíos	66
9.4	Recomendaciones y Futuras Direcciones.....	67
9.5	Conclusiones Finales.....	68
	Bibliografía	72
	Lista de Abreviaturas	74
	Anexo 1	75

1. Introducción

Twitter, una de las plataformas de redes sociales más prominentes, se ha establecido como un termómetro esencial de la opinión pública global. Diariamente, millones de tweets fluyen, reflejando sentimientos, opiniones y reacciones de personas de todo el mundo. Estos micro mensajes, aunque breves en contenido, están cargados de significado e información. Cada tweet representa una muestra de sentimiento, una respuesta a eventos actuales, una reacción a productos o simplemente reflexiones personales.

Dada la magnitud y riqueza de estos datos, surge una pregunta crucial: ¿Cómo podemos analizar y entender eficazmente estos sentimientos? Aunque el análisis de sentimientos ha sido un área de interés durante años, el desafío se intensifica con la variedad y volumen de datos disponibles hoy en día. Las técnicas tradicionales de análisis pueden no ser suficientes o adecuadas para abordar la complejidad inherente de estos datos.

Aquí es donde el aprendizaje automático entra en juego, ofreciendo técnicas y modelos avanzados capaces de procesar, analizar y clasificar grandes volúmenes de datos con precisión. Pero, con una variedad de modelos a disposición, desde redes neuronales hasta árboles de decisión y clasificadores bayesianos, surge una nueva pregunta: ¿Qué modelo o técnica es el más adecuado para el análisis de sentimientos en Twitter?

Este estudio se embarca en una misión para responder a esta pregunta. Mediante una evaluación comparativa de diferentes modelos de aprendizaje automático, este trabajo se centra en analizar, evaluar y comparar su desempeño en la tarea específica de detectar y clasificar sentimientos en tweets. Utilizando el dataset “sentiment140” de la Universidad de Stanford, se busca ofrecer una perspectiva clara y objetiva sobre el rendimiento relativo de estos modelos, proporcionando así una guía valiosa para futuras investigaciones y aplicaciones prácticas en el análisis de sentimientos.

1.1 Contexto y Motivación

La expansión del mundo digital ha llevado a la proliferación de plataformas donde las personas comparten sus opiniones y sentimientos sobre una amplia gama de temas. Estos datos, especialmente en plataformas como Twitter, representan una mina de oro de información sobre la percepción pública y el sentimiento general [6,15]. Sin embargo, destilar estos vastos volúmenes de datos en información significativa y comprensible es un desafío considerable.

El análisis de sentimientos, una subárea del procesamiento del lenguaje natural se dedica a abordar este desafío. Pero ¿cuál es el mejor enfoque para realizar este análisis? A lo largo de los años, se han propuesto múltiples técnicas y modelos de aprendizaje automático, cada uno con sus propias fortalezas y limitaciones. Desde redes neuronales hasta árboles de decisión y clasificadores bayesianos, la elección del modelo adecuado puede variar según el contexto y los datos disponibles.

Este estudio surge de la necesidad de realizar una evaluación comparativa de estos modelos en el contexto del análisis de sentimientos en Twitter. Utilizando el dataset público “sentiment140” publicado por la Universidad de Stanford, este trabajo tiene como objetivo no solo evaluar individualmente el desempeño de estos modelos, sino también compararlos entre sí en términos de precisión, recuperación y puntuación F1. Al hacerlo, se busca proporcionar una guía clara y objetiva para futuros investigadores y profesionales en el campo del análisis de sentimientos, ayudándoles a tomar decisiones informadas sobre qué modelos y técnicas son más adecuados para sus necesidades específicas [5].

1.2 Objetivo del estudio

En el vasto mundo del procesamiento del lenguaje natural, el análisis de sentimientos ha surgido como una herramienta esencial para descifrar y comprender las opiniones y emociones contenidas en textos. La popularidad y el alcance global de plataformas como Twitter han hecho que la necesidad de analizar estos sentimientos sea aún más imperativa. Sin embargo, con la diversidad de modelos de aprendizaje automático disponibles, determinar el enfoque más adecuado para esta tarea puede ser una empresa compleja.

El propósito central de este estudio es abordar este desafío mediante una evaluación y comparativa de rendimiento de diferentes modelos de aprendizaje automático en el análisis de sentimientos de tweets. Se pretende evaluar en detalle el desempeño de varios enfoques, desde redes neuronales y árboles de decisión hasta clasificadores bayesianos, en la detección y clasificación de sentimientos. Para ello, se utiliza el dataset “sentiment140”, proporcionando un terreno común y estandarizado para la evaluación.

Además de la evaluación individual, este trabajo busca realizar una comparación entre los modelos en términos de métricas estándar como precisión, recuperación y puntuación F1. A través de esta comparativa, se aspira a identificar cuál o cuáles modelos ofrecen el mejor rendimiento en el contexto específico del análisis de sentimientos en Twitter.

Con estos objetivos, este estudio no solo busca ofrecer una contribución valiosa al campo académico del análisis de sentimientos, sino también proporcionar una guía y referencia clara para profesionales e investigadores que desean implementar o mejorar soluciones basadas en aprendizaje automático en este ámbito.

1.3 Contribuciones y alcance

El campo del análisis de sentimientos ha sido objeto de numerosas investigaciones a lo largo de los años, especialmente con la creciente disponibilidad de datos en plataformas de redes sociales. Sin embargo, con la variedad de técnicas y enfoques disponibles, a menudo puede ser desafiante para investigadores y profesionales discernir qué modelos ofrecen el mejor rendimiento en escenarios específicos, como el análisis de tweets.

Este estudio tiene como objetivo llenar esta brecha, ofreciendo una contribución única en la forma de una evaluación comparativa detallada de diferentes modelos de aprendizaje



automático aplicados al análisis de sentimientos en Twitter. A través de una serie de experimentos y análisis rigurosos, se busca proporcionar una perspectiva clara y objetiva sobre la eficacia relativa de estos modelos, utilizando el dataset “sentiment140” como base.

El alcance de este trabajo se centra en la detección y clasificación de sentimientos en tweets, utilizando modelos de aprendizaje automático como redes neuronales, árboles de decisión, clasificadores bayesianos, entre otros. Mientras que el estudio se basa en un dataset específico, las técnicas, métodos y hallazgos presentados tienen implicaciones más amplias, y podrían ser adaptados o extendidos para otros conjuntos de datos o contextos relacionados con el análisis de sentimientos.

Por lo tanto, las contribuciones de este estudio no solo se limitan a los resultados y conclusiones derivados de los experimentos, sino también a la metodología empleada, que puede servir como una referencia valiosa para futuros trabajos en el campo del análisis de sentimientos y aprendizaje automático.

2. Estado del Arte

El análisis de sentimientos, en su intersección con el aprendizaje automático, ha capturado la atención y el interés de la comunidad académica y profesional durante la última década. Con la explosión de datos disponibles, especialmente en plataformas de redes sociales como Twitter, el deseo de entender, interpretar y actuar sobre estos sentimientos ha impulsado una serie de innovaciones y descubrimientos en el campo. Sin embargo, para abordar adecuadamente los desafíos y preguntas planteadas en este estudio, es esencial primero entender el terreno existente: los avances realizados, las técnicas propuestas y los desafíos identificados en investigaciones previas. Esta sección, Estado del Arte, busca ofrecer una revisión comprensiva y estructurada de los trabajos y descubrimientos clave en el ámbito del análisis de sentimientos aplicado a tweets, estableciendo así un marco de referencia y un contexto para la investigación actual.

2.1 Conceptos básicos

Para abordar adecuadamente el mundo del análisis de sentimientos utilizando técnicas de aprendizaje automático, es vital tener una comprensión clara y profunda de los términos y conceptos subyacentes. Estos conceptos básicos forman la columna vertebral del estudio y proporcionan el marco necesario para abordar y entender las secciones más técnicas y avanzadas del trabajo. A continuación, se presenta una revisión detallada de los conceptos clave:

2.1.1 Análisis de sentimientos

- **Definición:** Se refiere al proceso computacional de determinar los sentimientos o emociones expresados en un fragmento de texto.
- **Contexto:** El análisis de sentimientos es ampliamente utilizado en el mundo empresarial para evaluar opiniones de clientes sobre productos o servicios. También tiene aplicaciones en áreas como la política, para medir la opinión pública, o en el entretenimiento, para evaluar reacciones a películas o series.
- **Ejemplo:** Si se analiza la frase “Me encanta este producto”, un algoritmo de análisis de sentimientos categorizaría este comentario como positivo.

2.1.2 Aprendizaje Automático

- **Definición:** Es una rama de la inteligencia artificial que se centra en construir sistemas capaces de aprender a partir de datos, en lugar de ser programados con reglas específicas.
- **Contexto:** El aprendizaje automático ha revolucionado numerosos campos, desde la visión por computadora hasta el procesamiento del lenguaje natural, permitiendo avances como el reconocimiento facial o la traducción automática.
- **Ejemplo:** Un sistema de recomendación, como el utilizado por plataformas de streaming, se basa en técnicas de aprendizaje automático para sugerir películas o series basadas en el historial de visualización del usuario.

2.1.3 Clasificación

- **Definición:** En el aprendizaje supervisado, la clasificación es la tarea de predecir una etiqueta o categoría discreta para una entrada dada.
- **Contexto:** La clasificación es uno de los problemas más comunes en el aprendizaje automático y tiene aplicaciones en diversas áreas como diagnóstico médico, detección de fraude, y, en este caso, análisis de sentimientos.
- **Ejemplo:** Si se entrenara un modelo con comentarios etiquetados como “positivos” o “negativos”, la clasificación implicaría que el modelo pueda asignar una de esas etiquetas a un nuevo comentario basándose en lo que ha aprendido.

2.1.4 Redes Neuronales

- **Definición:** Son un conjunto de algoritmos modelados a partir del cerebro humano, diseñados para reconocer patrones. Consisten en capas de nodos (neuronas) que procesan y transmiten información.
- **Contexto:** Las redes neuronales, y en particular las redes neuronales profundas (deep learning), han impulsado avances significativos en tareas que antes se consideraban altamente desafiantes, como el reconocimiento de imágenes o la generación de texto.
- **Ejemplo:** Las redes convolucionales (CNN) son un tipo de red neuronal especialmente diseñada para procesar imágenes y son la base de muchos sistemas modernos de reconocimiento facial.

2.1.5 Métricas de Evaluación

- **Definición:** Herramientas cuantitativas que se utilizan para medir la efectividad y precisión de un modelo en una tarea específica.
- **Contexto:** Al entrenar y validar modelos de aprendizaje automático, es esencial contar con métricas objetivas que permitan evaluar el rendimiento del modelo y compararlo con otros enfoques o técnicas.
- **Ejemplo:** En clasificación, tres métricas comunes son la precisión (qué proporción de identificaciones positivas fue realmente correcta), recall (qué proporción de positivos reales se identificó correctamente) y la puntuación F1 (una medida que combina precisión y recall).

2.1.6 Dataset

- **Definición:** Colección organizada de datos que se utiliza para entrenar, validar o probar modelos de aprendizaje automático.
- **Contexto:** Los datasets pueden variar en tamaño y complejidad, y son esenciales para cualquier tarea de aprendizaje automático. El dataset proporciona ejemplos reales que el modelo utiliza para aprender y adaptarse.
- **Ejemplo:** El dataset “sentiment140” que se utiliza en este estudio, es una colección de tweets etiquetados según su sentimiento, y fue creado por investigadores de la Universidad de Stanford.

Estos conceptos y definiciones proporcionan una base sólida para adentrarse en las partes más técnicas y específicas del estudio. Al comprender estos términos fundamentales, se facilita



la interpretación de los experimentos, resultados y conclusiones que se presentarán en las siguientes secciones.

2.2 Historia y Evolución del Análisis de Sentimientos

El análisis de sentimientos, como disciplina, tiene sus raíces en el deseo de comprender las opiniones y emociones humanas expresadas en el lenguaje. Aunque ha ganado prominencia en la era digital, su historia se extiende mucho antes de la aparición de la tecnología moderna.

2.2.1 Inicios y Métodos Primitivos

Las primeras incursiones en el análisis de sentimientos se centraron en métodos léxicos simples. Estos métodos involucraban el uso de listas predefinidas de palabras con connotaciones positivas o negativas. La idea subyacente era sencilla: se contaba la ocurrencia de palabras “positivas” y “negativas” en un texto para estimar su polaridad general. Aunque rudimentarios y propensos a errores, estos enfoques representaron un primer paso en el intento de cuantificar sentimientos y opiniones [4,13].

2.2.2 Explosión de Datos y Aprendizaje Automático

Con la llegada de Internet y, posteriormente, de las redes sociales, el volumen de datos textuales disponibles para análisis creció exponencialmente. Sitios web como Twitter, Facebook y blogs generaban cantidades masivas de opiniones y reacciones del público. Los métodos léxicos tradicionales pronto se vieron superados por la magnitud y complejidad de estos datos. Fue en este contexto que el aprendizaje automático comenzó a desempeñar un papel fundamental en el análisis de sentimientos. Algoritmos de clasificación, como las máquinas de vectores de soporte y los modelos de regresión logística, se adaptaron para predecir la polaridad de los textos basándose en características extraídas de los datos.

2.2.3 Surgimiento de las Redes Neuronales y el Aprendizaje Profundo

A medida que avanzaba el siglo XXI, las técnicas de aprendizaje profundo, en particular las redes neuronales, comenzaron a revolucionar el campo. Estos modelos, con su capacidad para modelar relaciones no lineales y capturar matices en grandes conjuntos de datos, ofrecieron mejoras significativas en precisión y robustez. Herramientas y técnicas como Word2Vec, que generaba representaciones vectoriales densas de palabras, y modelos transformadores como BERT [3,8,10], permitieron un análisis semántico más profundo y contextualizado de los textos.

2.2.4 Desafíos Contemporáneos y Perspectivas Futuras

A pesar de los significativos avances tecnológicos, el análisis de sentimientos sigue siendo una disciplina desafiante. Los seres humanos a menudo emplean sarcasmo, ambigüedad, y metáforas en su lenguaje, lo que puede confundir incluso a los modelos más avanzados. Además, las diferencias culturales y lingüísticas pueden influir en la interpretación del sentimiento. A medida que la investigación avanza, se espera que surjan modelos más sofisticados y adaptativos, capaces de abordar estos desafíos con mayor precisión.

2.3 Técnicas Tradicionales vs. Machine Learning

El análisis de sentimientos, desde sus inicios, ha utilizado una variedad de técnicas y herramientas para identificar y extraer sentimientos y opiniones de textos. Estas técnicas han evolucionado con el tiempo, pasando de enfoques basados en reglas y léxicos a métodos basados en aprendizaje automático y estadística.

2.3.1 Técnicas Tradicionales

Las técnicas tradicionales son las que precedieron a la era del aprendizaje automático y estaban basadas en el conocimiento experto y en la lingüística.

Léxicos de Polaridad: Una de las primeras aproximaciones al análisis de sentimientos se basaba en el uso de léxicos o listas de palabras ya categorizadas según su polaridad (positiva, negativa o neutra). El análisis consistía en identificar estas palabras en un texto y, basándose en su presencia y frecuencia, determinar la polaridad general del texto.

Reglas Heurísticas: Además de los léxicos de polaridad, se desarrollaron reglas heurísticas para tratar de capturar estructuras lingüísticas específicas que podrían influir en el sentimiento de un texto, como las negaciones o los intensificadores.

Limitaciones: A pesar de su simplicidad y facilidad de implementación, estas técnicas presentaban limitaciones importantes. Carecían de la capacidad para adaptarse a nuevos contextos, lidiar con la ambigüedad del lenguaje y capturar sentimientos más sutiles o complejos.

2.3.2 Técnicas de Machine Learning

Con la expansión de la disponibilidad de datos y el avance en las capacidades computacionales, el aprendizaje automático emergió como una herramienta poderosa para el análisis de sentimientos.

- **Modelos Supervisados:** En este enfoque, se utilizan conjuntos de datos etiquetados, donde cada texto ya tiene asignada una polaridad, para entrenar modelos que aprendan a clasificar textos según su sentimiento. Modelos como Máquinas de Soporte Vectorial (SVM), Regresión Logística, o Redes Neuronales se entrenan para reconocer patrones y características en los datos que se correlacionan con un sentimiento específico.
- **Representaciones Vectoriales:** Una innovación clave en el aprendizaje automático para el análisis de sentimientos ha sido la capacidad de representar palabras y textos como vectores numéricos. Herramientas como Word2Vec o FastText transforman palabras en vectores que capturan información semántica, facilitando la tarea de los modelos de aprendizaje automático.
- **Aprendizaje No Supervisado y Semi-Supervisado:** Además del aprendizaje supervisado, se han explorado técnicas que no requieren etiquetas previas, como el clustering, o que utilizan una combinación de datos etiquetados y no etiquetados.
- **Ventajas y Desafíos:** Los modelos de aprendizaje automático ofrecen una mayor flexibilidad y adaptabilidad, siendo capaces de aprender de grandes volúmenes de datos y adaptarse a diferentes contextos y dominios. Sin embargo, también presentan desafíos, como la necesidad de grandes cantidades de datos etiquetados para entrenamiento o la interpretabilidad de los modelos.



La transición de técnicas tradicionales a enfoques basados en aprendizaje automático ha marcado un punto de inflexión en el análisis de sentimientos, permitiendo una mayor precisión, escalabilidad y adaptabilidad. No obstante, cada enfoque tiene sus propias ventajas y limitaciones, y la elección de uno u otro dependerá del contexto específico y de los objetivos del análisis [19].

2.4 Procesamiento del Lenguaje Natural en el Análisis de Sentimientos

El Procesamiento del Lenguaje Natural (NLP) es una disciplina fascinante que se encuentra en la intersección de la computación y la lingüística. Se dedica a construir sistemas que permitan a las máquinas entender, interpretar y generar lenguaje humano. Dentro del amplio espectro de aplicaciones del NLP, el análisis de sentimientos ha emergido como uno de los más prominentes y desafiantes. El lenguaje humano, en su esencia, está lleno de matices, ambigüedades y complejidades. Para descifrar estos matices y extraer sentimientos, es fundamental descomponer y analizar el texto a diferentes niveles [11].

Un paso inicial común en el NLP es la tokenización, que implica dividir un texto en unidades individuales, generalmente palabras. Esta segmentación permite un análisis más granular del contenido.

Una vez que se ha tokenizado un texto, se puede asignar roles gramaticales a cada palabra, una tarea conocida como etiquetado de partes del discurso (POS tagging). Comprender si una palabra actúa como sustantivo, verbo o adjetivo puede ser esencial para descifrar su papel en la oración y, por ende, el sentimiento que transmite.

Además de la estructura gramatical, es crucial entender la semántica, es decir, el significado de las palabras y las oraciones. Las herramientas modernas del NLP, como Word2Vec o GloVe, han revolucionado este aspecto al convertir palabras en vectores numéricos que capturan su esencia semántica en un espacio multidimensional.

La revolución más reciente en el NLP ha sido impulsada por modelos de aprendizaje profundo. Estos modelos, en particular las redes neuronales, han demostrado una habilidad sin precedentes para comprender y generar lenguaje. Modelos como BERT o GPT, basados en arquitecturas de transformadores, han redefinido lo que es posible en el análisis de sentimientos, logrando niveles de precisión antes inimaginables [3,8,10,18].

Sin embargo, a pesar de estos avances, aún persisten desafíos en la intersección del NLP y el análisis de sentimientos. El sarcasmo, las figuras retóricas, las connotaciones culturales y otros matices del lenguaje pueden confundir incluso a los sistemas más avanzados. Además, los modelos de aprendizaje profundo, aunque poderosos, tienen sus propias limitaciones, como la necesidad de grandes cantidades de datos y la interpretabilidad de sus predicciones.

Por lo tanto, el NLP es una herramienta esencial en el arsenal del análisis de sentimientos, y su interacción con esta disciplina seguirá siendo un área de investigación y desarrollo apasionante en los próximos años [19].

2.5 Avances Recientes y Tendencias Actuales

En la intersección de la tecnología y la lingüística, el análisis de sentimientos ha experimentado una metamorfosis inigualable en la última década. Esta transformación, alimentada por el vertiginoso avance tecnológico, investigaciones pioneras y una avalancha de datos en línea, ha redefinido el panorama del NLP.

La revolución del aprendizaje profundo en el NLP ha sido, sin duda, uno de los cambios más significativos. Antes de su irrupción, el análisis de sentimientos se apoyaba en modelos estadísticos y técnicas basadas en reglas. Estos enfoques, aunque efectivos en su momento, tenían limitaciones intrínsecas para capturar la riqueza y la complejidad del lenguaje humano. La aparición de las redes neuronales transformó este escenario. Estos modelos, con su habilidad para aprender representaciones jerárquicas y capturar relaciones no lineales, ofrecieron una nueva perspectiva sobre cómo las máquinas podían entender el lenguaje. Las redes neuronales convolucionales (CNN) y recurrentes (RNN) fueron las primeras en mostrar promesa, gestionando eficientemente patrones locales y secuencias en los textos [20].

No obstante, fue la introducción de los transformadores y modelos como BERT, GPT y sus variantes, lo que realmente marcó un antes y un después. Al ser entrenados en vastos conjuntos de datos, estos modelos lograron comprender el contexto de una manera que otros modelos no podían, estableciendo récords en múltiples tareas de NLP, incluido el análisis de sentimientos [3,8,10,18].

La disponibilidad de datos también ha jugado un papel crucial en estos avances. Vivimos en una era donde la generación de datos es constante. Cada tweet, comentario o reseña en línea es un reflejo de la opinión pública, y cuando se agregan, forman un tesoro para los investigadores. Esta avalancha de datos no solo ha permitido entrenar modelos más potentes, sino que también ha ampliado el alcance del análisis de sentimientos a nuevos dominios, idiomas y culturas.

Sin embargo, el poder de estos modelos de NLP viene acompañado de responsabilidades y desafíos. A medida que las decisiones basadas en el análisis de sentimientos ganan más influencia en áreas como la toma de decisiones empresariales, la moderación de contenido y el análisis de tendencias, surge una pregunta ineludible sobre su ética y transparencia. ¿Cómo garantizamos que estos modelos son justos? ¿Cómo nos aseguramos de que no perpetúen sesgos? Estas cuestiones han llevado a la comunidad de NLP a explorar nuevas áreas de investigación, centradas en la interpretabilidad de los modelos, la detección de sesgos y la construcción de sistemas más éticos.

Mirando hacia el horizonte, el análisis de sentimientos y el NLP están preparados para embarcarse en nuevos viajes de descubrimiento. Con el potencial de la computación cuántica, el auge del aprendizaje por transferencia y la exploración de modelos multimodales, estamos en el umbral de otra revolución en el campo [7,11,12].



3. Análisis del Problema

El análisis de sentimientos, en el vasto dominio de la comunicación digital, ha emergido como una herramienta vital para capturar y comprender las opiniones y emociones de la gente. Twitter, con su naturaleza breve y dinámica, se ha consolidado como una mina de oro para este tipo de análisis. Cada tweet, en sus escasos caracteres, puede encapsular una amplia gama de emociones, desde el júbilo y la admiración hasta el desdén y la frustración. Esta sección se sumerge en el intrincado mundo del análisis de sentimientos en Twitter, explorando las complejidades y particularidades que presenta este formato único. Al adentrarnos en este análisis, estableceremos un marco sólido que nos guiará a través de las etapas subsiguientes de diseño, desarrollo e implementación.

3.1 Definición del problema

El análisis de sentimientos busca descifrar y clasificar las emociones y actitudes expresadas en un texto. A primera vista, la tarea puede parecer directa: identificar si un fragmento de texto, en este caso, un tweet, refleja un sentimiento positivo, negativo o neutro. Sin embargo, el desafío es mucho más intrincado de lo que parece.

El inglés, como cualquier otro idioma, es increíblemente rico y variado. A través de nuestras palabras, no sólo transmitimos información sino también emociones, intenciones, sarcasmo, humor y otros matices. Una oración puede tener múltiples interpretaciones basadas en el tono, el contexto, o incluso la cultura y las experiencias previas del lector. Considere, por ejemplo, la frase “It’s not the worst”. Aunque a simple vista parece una valoración positiva, también puede ser interpretada como una respuesta tibia o incluso negativa, dependiendo del contexto. Las expresiones idiomáticas y las frases hechas pueden añadir capas adicionales de significado y complejidad.

El lenguaje utilizado en Twitter presenta sus propios desafíos. La limitación de caracteres ha llevado a una forma de comunicación condensada, donde abreviaciones, emojis y hashtags se han convertido en herramientas esenciales. Estos elementos, si bien enriquecen la comunicación, también añaden complejidad al análisis de sentimientos. El análisis de sentimientos tampoco es estático. El lenguaje en Twitter, y en línea en general, es dinámico. Las palabras adquieren nuevos significados, aparecen nuevos términos y memes, y las connotaciones cambian con el tiempo. Los modelos y enfoques que son efectivos ahora pueden necesitar adaptarse para mantenerse relevantes en el futuro.

En esencia, el análisis de sentimientos en Twitter es un ejercicio de navegación a través de las complejidades del lenguaje, las peculiaridades de la plataforma y la naturaleza siempre cambiante de la comunicación en línea.

3.2 Importancia del Análisis de Sentimientos



Twitter se ha convertido en un epicentro de conversación global. Cada día, millones de usuarios comparten sus opiniones, reacciones y sentimientos sobre una amplia variedad de temas, desde eventos mundiales y noticias de última hora hasta experiencias personales y opiniones sobre productos. Esta inmensa cantidad de información, si se analiza correctamente, tiene el poder de ofrecer una perspectiva única sobre el pulso de la sociedad en un momento dado.

El análisis de sentimientos desempeña un papel crucial en la extracción de significado de esta avalancha de datos. Al poder determinar la polaridad y el tono emocional detrás de cada tweet, las organizaciones, investigadores y otros interesados pueden obtener una comprensión profunda de las opiniones y percepciones del público. Estos insights son valiosos por varias razones:

1. **Toma de Decisiones en Tiempo Real:** En un mundo donde la información se mueve a la velocidad de la luz, las empresas y organizaciones pueden usar el análisis de sentimientos para tomar decisiones informadas en tiempo real. Si un nuevo producto se lanza y los tweets al respecto son mayormente negativos, las empresas pueden actuar rápidamente para abordar las preocupaciones.
2. **Previsión de Tendencias:** Al analizar los sentimientos a lo largo del tiempo, es posible identificar tendencias emergentes. Esto puede ser especialmente útil en sectores como el financiero, donde las percepciones del público pueden influir en el comportamiento del mercado.
3. **Investigación Académica y Social:** Los académicos pueden utilizar Twitter como una muestra representativa para estudiar fenómenos sociales, culturales y políticos. El análisis de sentimientos puede revelar cómo se siente la sociedad sobre temas candentes, desde elecciones políticas hasta eventos culturales.
4. **Desarrollo de Producto y Mejora de Servicios:** Las empresas pueden usar el feedback directo de los usuarios para mejorar sus productos o servicios. Un análisis detallado de los sentimientos puede ayudar a identificar áreas problemáticas o características que los usuarios desean.
5. **Gestión de la Reputación:** En la era digital, la reputación de una marca puede ser construida o destruida en cuestión de horas. Monitorizar el sentimiento en Twitter permite a las empresas ser proactivas en la gestión de su imagen y responder adecuadamente a cualquier crisis.
6. **Personalización de Publicidad y Marketing:** Al entender el sentimiento del público hacia ciertos temas o productos, las empresas pueden personalizar sus campañas publicitarias para resonar mejor con su audiencia objetivo.

El análisis de sentimientos, por lo tanto, no es simplemente una herramienta técnica; es un puente que conecta las voces individuales de millones de usuarios con organizaciones, empresas e investigadores. En un mundo interconectado, donde cada voz tiene el poder de influir y ser escuchada, el análisis de sentimientos en Twitter se convierte en una herramienta esencial para entender y navegar por el complejo paisaje de la opinión pública.

3.3 Exploración y Características del Dataset Sentiment140

El análisis de sentimientos en redes sociales, y en particular en Twitter, ofrece una ventana única a la percepción y opinión pública sobre una amplia variedad de temas. Desde anuncios de productos hasta acontecimientos políticos, los usuarios de Twitter comparten abiertamente sus reacciones y opiniones en tiempo real. Capturar, entender y analizar estos sentimientos es una tarea compleja que requiere no solo herramientas y técnicas avanzadas, sino también datos de alta calidad. En este contexto, el dataset Sentiment140 emerge como una herramienta invaluable para los investigadores y profesionales del análisis de sentimientos.

Originado en la Universidad de Stanford, el dataset Sentiment140 ha sido diseñado específicamente para la tarea de análisis de sentimientos en Twitter. Con una rica colección de 1.6 millones de tweets etiquetados, proporciona una base sólida para entrenar, validar y evaluar modelos de aprendizaje automático. Cada tweet en este dataset ha sido meticulosamente etiquetado como positivo o negativo, brindando claridad y dirección a las investigaciones en esta área. A pesar de su aparente simplicidad, el dataset presenta desafíos únicos. Los tweets, por naturaleza, son concisos y pueden contener una mezcla de lenguaje estándar, jerga, abreviaturas, enlaces, menciones, hashtags y emojis. Esta composición específica demanda un preprocesamiento y análisis cuidadosos antes de su uso efectivo en modelos predictivos [5,9].

3.3.1 Características y Estructura del Dataset Sentiment140

Mientras que la popularidad y el uso extensivo del dataset Sentiment140 en la investigación lo destacan, es esencial comprender su estructura y características intrínsecas para maximizar su utilidad en el análisis de sentimientos.

El dataset se presenta en un formato tabular, con cada fila representando un tweet individual y las columnas correspondientes a diferentes características del tweet. Las dos columnas principales de interés son:

- **sentiment_label:** Esta columna almacena la etiqueta de sentimiento asociada con el tweet. En la versión específica de este dataset utilizada para este estudio, las etiquetas son binarias: 0 para sentimientos negativos y 4 para sentimientos positivos.
- **tweet_text:** Esta columna contiene el contenido textual del tweet. Es aquí donde reside la riqueza de los datos, con cada entrada reflejando las opiniones, sentimientos y expresiones del usuario de Twitter.

Los tweets en Sentiment140 varían ampliamente en contenido, desde opiniones sobre eventos actuales y reacciones a productos o servicios hasta expresiones personales de alegría, tristeza, humor y más. Esta variedad es un testimonio de la diversidad de voces presentes en Twitter.

- **Lenguaje y Estilo:** Al ser una plataforma global, Twitter alberga una mezcla de estilos lingüísticos, dialectos y jergas. Aunque el dataset Sentiment140 se centra en tweets en inglés, es común encontrar variaciones dialectales, abreviaturas, acrónimos y, a veces, errores tipográficos o gramaticales. Esta naturaleza heterogénea del lenguaje en el dataset presenta tanto desafíos como oportunidades para el análisis de sentimientos.
- **Elementos No Textuales:** Además del contenido textual puro, los tweets a menudo contienen elementos como hashtags (#), menciones (@usuario), enlaces (URLs) y



emojis. Estos elementos, aunque no son palabras en el sentido tradicional, pueden llevar un peso significativo en la interpretación del sentimiento y el contexto del tweet.

- **Distribución de Etiquetas:** Aunque el dataset Sentiment140 tiene una distribución equilibrada entre tweets positivos y negativos, es crucial recordar que esta distribución no necesariamente refleja la distribución real de sentimientos en Twitter. La elección de una distribución equilibrada es deliberada, diseñada para facilitar el entrenamiento y evaluación de modelos de aprendizaje automático.

Comprender estas características y peculiaridades del dataset Sentiment140 no es una tarea meramente académica; es una necesidad imperante para cualquier esfuerzo serio en el análisis de sentimientos. Cada aspecto del dataset, desde la estructura y composición hasta las sutilezas del lenguaje y estilo, influye directamente en las decisiones tomadas durante el proceso de investigación.

Por ejemplo, la naturaleza heterogénea del lenguaje en el dataset puede influir en las técnicas de preprocesamiento adoptadas. Saber que los tweets pueden contener abreviaturas, acrónimos o errores tipográficos podría llevar al investigador a adoptar técnicas de corrección ortográfica, expansión de acrónimos o incluso normalización de texto. De manera similar, la presencia de elementos como hashtags y menciones podría requerir una consideración especial, ya que estos elementos pueden proporcionar contexto adicional o incluso alterar el sentimiento inherente de un tweet.

Además, la distribución equilibrada de las etiquetas en el dataset, aunque es beneficioso desde una perspectiva de modelado, también requiere cautela al interpretar los resultados. Un modelo entrenado en un dataset equilibrado podría no reflejar adecuadamente las realidades de un entorno desequilibrado, donde, por ejemplo, los tweets positivos podrían ser mucho más frecuentes que los negativos.

En esencia, cada característica del dataset Sentiment140 lleva consigo consideraciones y decisiones que afectan no solo el proceso de análisis, sino también la interpretación y aplicabilidad de los resultados finales. Es por ello por lo que una comprensión profunda y matizada del dataset es una piedra angular para cualquier investigación en el análisis de sentimientos [9].

3.4 Retos y Desafíos del Análisis de Sentimientos en inglés

El camino hacia un análisis de sentimientos efectivo en el inglés está plagado de obstáculos y desafíos. Uno de los más prominentes es la ambigüedad inherente al lenguaje. Palabras y frases en inglés pueden tener múltiples interpretaciones, cambiando su significado en función del contexto. Tomemos como ejemplo la palabra “light”, que puede referirse tanto a la luminosidad como al peso o la intensidad. En la economía de caracteres que impone Twitter, esta ambigüedad puede amplificarse, haciendo que el desciframiento del verdadero sentimiento sea aún más enrevesado.

El sarcasmo y la ironía añaden otra capa de complejidad. Estas formas retóricas, aunque enriquecen la expresividad del lenguaje, presentan un verdadero desafío al intentar determinar la

verdadera intención detrás de un mensaje. Una declaración que, en su superficie, parece positiva, podría estar cargada de ironía, invirtiendo su significado por completo. Twitter, con su naturaleza breve y directa, ha generado una cultura propia de comunicación. El uso frecuente de jerga, abreviaciones y emojis se ha convertido en norma en la plataforma. Mientras que términos como “LOL” o “BRB” se han vuelto ubicuos, entender y contextualizar estas abreviaciones y símbolos en el análisis de sentimientos es esencial.

Además, no podemos olvidar las variaciones regionales del inglés. Aunque es un idioma global, presenta diferencias marcadas dependiendo de la región, ya sea en Reino Unido, Estados Unidos, Australia o cualquier otra parte del mundo. Estas variaciones pueden influir significativamente en el sentimiento y la interpretación de un mensaje.

El dinamismo del lenguaje en plataformas digitales presenta otro desafío. Con nuevas palabras, memes y expresiones emergiendo constantemente, el análisis de sentimientos debe ser adaptable y evolutivo. Además, garantizar que estos análisis sean justos y no estén sesgados es esencial para obtener resultados precisos y confiables. Aunque el análisis de sentimientos ofrece un gran potencial para captar la esencia de la comunicación en Twitter, es una tarea que requiere una cuidadosa consideración de la complejidad y la evolución constante del lenguaje inglés [13].

4. Diseño

La ciencia de los datos, en su esencia, es tanto un arte como una ciencia. Cada decisión, desde la selección de herramientas hasta la elección de modelos, está imbuida de una combinación de experiencia técnica, intuición y conocimiento del dominio. En el campo del análisis de sentimientos, donde las emociones humanas se entrelazan con textos y números, la metodología adoptada es de suma importancia.

El análisis de sentimientos se ha convertido en un pilar en el mundo de la ciencia de datos, con aplicaciones que abarcan desde el marketing hasta la política. Con la proliferación de las redes sociales y la omnipresencia de las opiniones en línea, nunca ha sido más crucial entender y analizar cómo se sienten las personas sobre productos, servicios o eventos. Sin embargo, la tarea no es sencilla. El lenguaje humano es complejo, y la tarea de determinar el sentimiento detrás de un texto es, en el mejor de los casos, desafiante.

En este estudio, enfrentamos este desafío de frente, eligiendo abordar el problema con una variedad de modelos y técnicas. Pero antes de sumergirnos en los modelos y algoritmos, es esencial establecer una base sólida. Esta base comienza con una comprensión clara de los datos a mano y se extiende a cómo se preparan y procesan esos datos para el análisis. La elección del dataset, las decisiones sobre cómo limpiar y preparar los datos, y las técnicas utilizadas para transformar el texto en algo que los modelos puedan entender son etapas cruciales que sientan las bases para todo el análisis posterior.

A continuación, se detalla el diseño experimental adoptado para este estudio, subrayando las decisiones tomadas en cada fase y justificando las elecciones realizadas. Desde la preparación inicial de datos hasta la evaluación final de los modelos, cada paso se presenta con el objetivo de proporcionar una visión clara del proceso y la metodología adoptados.

4.1 Diseño Experimental

La base de cualquier estudio de aprendizaje automático es el diseño experimental. Es aquí donde se establecen las bases y se toman decisiones cruciales que guiarán todo el proceso de análisis. En este apartado, se detallan los pasos iniciales y fundamentales tomados en la preparación y configuración del entorno para el estudio del análisis de sentimientos en tweets.

4.1.1 Importaciones

En cualquier proyecto de análisis de datos, la selección de herramientas adecuadas es esencial. Estas herramientas no solo facilitan el proceso de análisis, sino que también pueden influir en los resultados y en la interpretación de estos.

pandas: Esta biblioteca de Python se ha convertido en un estándar en el mundo del análisis de datos. Proporciona estructuras de datos flexibles y permite una manipulación eficiente de grandes conjuntos de datos. Su capacidad para leer diversos formatos de archivos y su amplia



gama de funciones para limpiar, filtrar y agrupar datos la hacen indispensable para este proyecto [14].

RandomForestClassifier: Pertenece a la biblioteca sklearn, y es una implementación del algoritmo de Random Forest para tareas de clasificación. El Random Forest es un método de aprendizaje conjunto que utiliza múltiples árboles de decisión para realizar predicciones. Su naturaleza inherentemente paralela y su capacidad para manejar una alta dimensionalidad lo hacen adecuado para tareas como el análisis de sentimientos.

train_test_split: Esta función, también de sklearn, es fundamental para dividir un conjunto de datos en partes de entrenamiento y prueba. Al mantener un conjunto de datos separado para la prueba, podemos asegurarnos de que nuestros modelos se evalúen en datos no vistos, proporcionando una medida de rendimiento más realista.

CountVectorizer: El procesamiento de texto es una tarea compleja. Las palabras, frases y oraciones deben convertirse en formatos que los modelos de aprendizaje automático puedan entender. CountVectorizer convierte el texto en una matriz numérica basada en la frecuencia de palabras, transformando así el lenguaje humano en una forma que los algoritmos pueden procesar.

Métricas (accuracy_score, classification_report): Evaluar el rendimiento de un modelo es tan crucial como construirlo. Las métricas como accuracy_score proporcionan una medida rápida de cuántas predicciones fueron correctas, mientras que classification_report ofrece una visión detallada, mostrando métricas como precisión, exhaustividad y puntaje F1 para cada clase.

La elección de estas herramientas y bibliotecas no fue casual. Cada una fue seleccionada por su robustez, eficiencia y relevancia para el proyecto. Juntas, forman la base técnica sobre la que se construye todo el análisis.

4.1.2 Carga y Preparación de Datos

Con El análisis de datos comienza con la adquisición y preparación de datos. Esta fase es fundamental, ya que los datos son la base sobre la que se construyen todos los modelos y análisis posteriores. La calidad y la estructura de los datos pueden influir significativamente en los resultados obtenidos.

Selección del Dataset “sentiment140”: La elección del dataset “sentiment140” no fue arbitraria. Este conjunto de datos, desarrollado por investigadores de la Universidad de Stanford, es conocido por su extensividad y calidad. Consiste en una colección de tweets etiquetados, lo que lo hace especialmente valioso para un estudio de análisis de sentimientos. Cada entrada en el dataset proporciona una visión directa de las emociones y opiniones de los usuarios de Twitter, lo que lo convierte en una fuente rica de información para este estudio.

Carga de Datos con pandas: Una vez seleccionado el dataset, el siguiente paso fue cargarlo en el entorno de trabajo. Gracias a la biblioteca pandas, este proceso fue directo y eficiente. Con simples comandos, se pudo cargar, explorar y visualizar los primeros registros del dataset, estableciendo así una comprensión inicial de la estructura y el contenido de los datos.

Limpieza y Filtrado: Ningún conjunto de datos es perfecto. A menudo, los datos vienen con ruido, valores faltantes o errores. Para este estudio, fue esencial garantizar que los datos estuvieran limpios y libres de irregularidades que pudieran afectar los resultados. Por ejemplo, se detectaron y filtraron filas con etiquetas ‘sentiment_label’ que parecían ser encabezados duplicados. Este tipo de limpieza garantiza que los modelos de aprendizaje automático se entrenen y evalúen con datos de alta calidad.

Renombrado de Columnas: Para facilitar el análisis y la interpretación, es crucial que las columnas del dataset tengan nombres descriptivos y coherentes. Por ello, se tomó la decisión de renombrar las columnas a ‘sentiment_label’ y ‘tweet_text’. Estos nombres no solo son descriptivos, sino que también facilitan el acceso y la manipulación de los datos en etapas posteriores.

División en Conjuntos de Entrenamiento y Prueba: Uno de los principios fundamentales en aprendizaje automático es no evaluar un modelo en los mismos datos en los que fue entrenado. Por ello, se utilizó `train_test_split` para dividir el dataset en dos partes: un conjunto de entrenamiento y un conjunto de prueba. Al reservar un 20% de los datos para pruebas, se aseguró que los modelos tuvieran una variedad de datos no vistos previamente para la evaluación, garantizando así una medida de rendimiento más realista.

La carga y preparación de datos son etapas esenciales que establecen el tono para todo el análisis. La atención al detalle en esta fase asegura que las etapas subsiguientes, desde la vectorización hasta la modelización, se realicen sobre una base sólida y confiable.

4.1.3 Vectorización

Una de las principales barreras cuando se trabaja con datos de texto en aprendizaje automático es la naturaleza inherente del lenguaje humano: es complejo, matizado y altamente contextual. Para que los modelos de aprendizaje automático puedan operar con texto, es necesario transformar esta información lingüística en un formato numérico. Aquí es donde entra en juego la vectorización.

Naturaleza del Texto: El texto, en su esencia, es una serie de símbolos (letras, números, signos de puntuación) que juntos forman palabras, frases y párrafos. Estos, a su vez, llevan significado basado en el contexto, la estructura gramatical y las convenciones culturales. Sin embargo, los modelos de aprendizaje automático no “entienden” el texto de la misma manera que los seres humanos. Para ellos, es esencial tener una representación numérica del texto para poder procesarlo.

Uso de CountVectorizer: Para este estudio, se empleó `CountVectorizer` de la biblioteca `sklearn`. Esta herramienta convierte el texto en una matriz de términos o tokens y cuenta la frecuencia de aparición de cada término en cada documento (en este caso, tweet). De esta manera, cada tweet se representa como un vector en un espacio multidimensional, donde cada dimensión corresponde a un término único del corpus.

Beneficios y Consideraciones: La principal ventaja de usar `CountVectorizer` es su simplicidad y eficacia. Al convertir el texto en vectores basados en la frecuencia de palabras, es posible capturar parte de la estructura y el contenido del texto. Sin embargo, es importante tener



en cuenta que esta técnica no captura el contexto o el orden de las palabras, lo que puede ser relevante en análisis más profundos.

Una consideración adicional es la alta dimensionalidad que puede resultar de la vectorización, especialmente en grandes conjuntos de datos con vocabularios extensos. A menudo, es útil aplicar técnicas de reducción de dimensionalidad o filtrar términos poco frecuentes para optimizar el rendimiento y la eficiencia del modelo.

Preparación para el Modelado: Una vez completada la vectorización, los datos estaban listos para ser introducidos en los modelos de aprendizaje automático. Cada tweet, ahora representado como un vector, podría ser procesado, analizado y clasificado por los algoritmos seleccionados.

La vectorización es un paso esencial en el procesamiento de texto para el aprendizaje automático. Al transformar el lenguaje humano en un formato que los modelos puedan comprender, se sientan las bases para análisis más profundos y descubrimientos significativos en el ámbito del análisis de sentimientos.

4.1.4 División de Datos

Finalmente, antes de sumergirse en el modelado, se dividió el conjunto de datos en grupos de entrenamiento y prueba. Esta división es esencial para validar y evaluar modelos. Al reservar un 20% de los datos para pruebas, se garantizó una evaluación robusta y confiable de los modelos en datos no vistos previamente.

Con estos pasos iniciales completados, el estudio estaba listo para avanzar hacia la implementación y evaluación de modelos específicos de aprendizaje automático, que se detallarán en los siguientes apartados.

4.2 Elección de Modelos y Técnicas

La elección de modelos y técnicas es un paso crucial en cualquier proyecto de aprendizaje automático. Cada modelo tiene sus propias fortalezas, debilidades y supuestos subyacentes, y la elección adecuada puede marcar una gran diferencia en la eficacia y precisión de los resultados. En este estudio, se optó por un enfoque diversificado, seleccionando múltiples modelos para abordar el problema del análisis de sentimientos. Esta elección permitió comparar y contrastar el rendimiento de diferentes técnicas y obtener una visión más holística del problema [19].

4.2.1 Random Forest Model

El modelo Random Forest es una técnica avanzada de aprendizaje automático que combina múltiples árboles de decisión para generar un “bosque” de soluciones. Originado en la investigación de estadísticas y reconocimiento de patrones, su metodología se basa en la creación de múltiples árboles durante el entrenamiento y en la producción de una salida que es la moda de las salidas de cada árbol individual. Esta técnica de ensamblado ofrece ventajas notables, como la reducción de overfitting y la mejora de la precisión general, convirtiéndola en una herramienta potente para tareas de clasificación y regresión en diversos dominios, incluido el análisis de sentimientos. [2]

Características Principales:

- **Enfoque de conjunto:** En lugar de depender de un único árbol de decisión, Random Forest crea un “bosque” de árboles, lo que reduce la varianza y el riesgo de sobreajuste.
- **Manejo de datos faltantes:** Una de las ventajas de Random Forest es su capacidad para manejar datos faltantes, asignando valores basados en la importancia de las variables o utilizando métodos de imputación.
- **Reducción de la Varianza:** Al promediar los resultados de varios árboles, se minimiza el ruido y se reduce la varianza, lo que conduce a predicciones más estables.
- **Importancia de las características:** Una característica útil del modelo Random Forest es su capacidad para clasificar la importancia de las variables en la predicción. Esto puede ofrecer insights sobre qué términos o características del texto tienen un impacto significativo en la determinación del sentimiento.

La elección de Random Forest para este estudio se basó en su robustez y versatilidad, su capacidad para capturar relaciones no lineales y su habilidad para manejar datos de alta dimensionalidad, como los que se encuentran en problemas de procesamiento de texto.

La función de predicción de un Random Forest se compone de las predicciones de todos los árboles individuales. Para clasificación, el voto mayoritario de los árboles se toma como la predicción final, mientras que, para regresión, se puede promediar las predicciones de los árboles.

El proceso de creación de árboles en un Random Forest implica dos fuentes de aleatorización:

1. **Bagging (Bootstrap Aggregating):** Se generan múltiples conjuntos de entrenamiento mediante el muestreo con reemplazo de N instancias del conjunto de datos original. Cada conjunto de entrenamiento se utiliza para entrenar un árbol de decisión. Esta técnica introduce variabilidad en los datos de entrenamiento y reduce el riesgo de sobreajuste.
2. **Selección aleatoria de características:** En cada nodo de división durante la construcción de un árbol, solo un subconjunto aleatorio de características se considera para la división óptima. Esto reduce la dependencia entre los árboles y mejora la generalización.

La ecuación general del proceso de predicción de un Random Forest se puede expresar como (1):

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T f_t(x) \quad (1)$$

4.2.2 Multi-layer Perceptron

El Multi-layer Perceptron (MLP) es una red neuronal artificial que ha sido fundamental en el campo del aprendizaje profundo y la clasificación. Con su estructura de múltiples capas, ha sido una herramienta poderosa en una variedad de aplicaciones, incluyendo el análisis de sentimientos [17].



Características Principales:

- **Arquitectura de Capas:** MLP consta de una capa de entrada, una o más capas ocultas y una capa de salida. Cada capa está compuesta por neuronas que están completamente conectadas a la siguiente capa.
- **Funciones de Activación:** Utiliza funciones de activación no lineales, como la función sigmoide o ReLU, lo que permite modelar relaciones complejas.
- **Aprendizaje Supervisado:** MLP se entrena utilizando algoritmos de retro propagación, ajustando los pesos de las conexiones para minimizar el error en las predicciones.
- **Flexibilidad:** Su capacidad para adaptarse a diferentes estructuras lo hace adecuado para una amplia gama de tareas, incluyendo la clasificación de texto.
- **Alta Dimensionalidad:** Es eficaz en el manejo de datos de alta dimensionalidad, como vectores de palabras en el análisis de texto.

El modelo Multi-layer Perceptron fue seleccionado para este estudio debido a su capacidad para modelar relaciones no lineales, su flexibilidad en el manejo de diferentes tipos de datos y su eficacia demostrada en tareas de clasificación complejas, como el análisis de sentimientos.

La formulación esencial del MLP se puede expresar en términos de su operación matricial (2):

$$y = f_{out}(W_{out} \cdot f_{hid}(W_{hid} \cdot x + b_{hid}) + b_{out}) \quad (2)$$

Donde x es el vector de entrada, W_{hid} y W_{out} son las matrices de pesos de las capas ocultas y de salida, respectivamente, b_{hid} y b_{out} son los vectores de sesgo de las mismas capas, f_{hid} y f_{out} son funciones de activación no lineales aplicadas a las capas ocultas y de salida, respectivamente, e y es el resultado final de la red.

4.2.3 Regresión Logística

La regresión logística, un pilar en la estadística y el aprendizaje automático, es un método de análisis que predice la probabilidad de un evento binario en función de una o más variables predictoras. Aunque tradicionalmente se ha utilizado en campos como la medicina o la investigación social para predecir la probabilidad de ocurrencia de un evento, ha encontrado un lugar prominente en el procesamiento de texto debido a su capacidad para clasificar documentos o fragmentos de texto en categorías predeterminadas. Su enfoque basado en la probabilidad y su capacidad para proporcionar interpretaciones claras de los coeficientes la convierten en una herramienta valiosa para el análisis de sentimientos.

Características Principales:

- **Modelo probabilístico:** A diferencia de otros modelos que producen una clasificación directa, la regresión logística estima probabilidades. Esto permite una interpretación más matizada de los resultados y la posibilidad de ajustar umbrales de clasificación.
- **Interpretabilidad:** Los coeficientes del modelo se pueden interpretar en términos de log-odds, ofreciendo insights sobre la relación entre las características y la respuesta.
- **Eficiencia:** La regresión logística es computacionalmente menos intensiva que otros modelos, lo que la hace adecuada para conjuntos de datos más grandes.

Dada su simplicidad, interpretabilidad y eficacia demostrada en tareas de clasificación binaria, la regresión logística fue una elección natural para el análisis de sentimientos. Es especialmente útil cuando se desea entender la influencia de términos individuales o características en la predicción del sentimiento.

4.2.4 Máquinas de Soporte Vectorial

Las Máquinas de Soporte Vectorial (SVM), son una clase de algoritmos que han revolucionado el campo del aprendizaje automático desde su introducción. Diseñadas originalmente para problemas de clasificación binaria, las SVM han demostrado ser extremadamente efectivas en una variedad de tareas, desde la clasificación de imágenes hasta el análisis de texto. Su capacidad para transformar datos en un espacio de mayor dimensionalidad y encontrar hiperplanos de decisión óptimos las convierte en una herramienta poderosa, especialmente en situaciones donde las relaciones entre datos no son lineales.

Características Principales:

- **Hiperplano de decisión:** El objetivo principal de las SVM es encontrar el hiperplano que mejor separe las clases en el espacio de características. Este hiperplano se elige de manera que maximice el margen entre las dos clases.
- **Kernel trick:** Una de las características más notables de las SVM es su capacidad para utilizar funciones kernel, que permiten transformar datos en un espacio de mayor dimensionalidad para hacerlos linealmente separables. Esto es especialmente útil cuando las relaciones en los datos originales no son lineales.
- **Regularización:** Las SVM incorporan términos de regularización que previenen el sobreajuste y permiten un mejor rendimiento en conjuntos de datos complejos.

La elección de las SVM en este estudio se basó en su capacidad para manejar datos de texto de alta dimensionalidad y su eficacia demostrada en problemas similares de clasificación de texto.

La formulación fundamental de SVM se puede expresar en términos de la siguiente ecuación (3):

$$f(x) = \langle w, x \rangle + b \quad (3)$$

Donde $f(x)$ es una función de decisión para la instancia x , w es un vector de pesos que define la orientación del hiperplano y b es un término de sesgo.

En el caso de SVM para clasificación binaria, el objetivo es encontrar el hiperplano $w \cdot x + b = 0$ que separe las dos clases de manera óptima. Sin embargo, esta ecuación no captura la noción de margen, que es crucial en SVM. Para abordar esto, se introduce el concepto de “margen funcional” (4):

$$\text{Margen Funcional} = y_i \cdot (\langle w, x_i \rangle + b) \quad (4)$$

Donde y_i es la etiqueta de clase de la instancia x_i . El margen funcional mide la distancia de una instancia al hiperplano ponderado por su etiqueta de clase. El objetivo en SVM es maximizar este margen funcional sujeto a la restricción de que todas las instancias estén clasificadas correctamente, es decir (5):

$$(5)$$



$$y_i \cdot (\langle w, x_i \rangle + b) \geq 1 \text{ para todas las instancias } (x_i, y_i)$$

Para considerar instancias que están cerca del margen, se introduce la idea de “vectores de soporte”, que son aquellas instancias que se encuentran más cercanas al hiperplano y contribuyen a definirlo. La distancia entre el hiperplano y los vectores de soporte es el margen geométrico que se busca maximizar.

La ecuación del hiperplano óptimo se encuentra mediante la optimización de la función de pérdida regularizada (6):

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \max(0, 1 - y_i(\langle w, x_i \rangle + b)) \quad (6)$$

Donde N es el número de instancias, C es un parámetro de regularización que controla el equilibrio entre maximizar el margen y minimizar el error de clasificación, y la función de penaliza las instancias clasificadas incorrectamente.

4.2.5 Naive Bayes Multinomial

El algoritmo Naive Bayes, inspirado en el teorema de probabilidad de Bayes, ha sido un algoritmo fundamental en el campo del aprendizaje automático y la minería de texto durante décadas. A pesar de su simplicidad conceptual, ha demostrado ser increíblemente efectivo en una variedad de tareas, especialmente en la clasificación de documentos. En el contexto del análisis de texto, su adaptabilidad y eficiencia lo han convertido en una herramienta de elección para muchos investigadores y profesionales, sobre todo cuando se enfrentan a grandes volúmenes de datos textuales.

Características Principales:

- **Base probabilística:** Naive Bayes estima las probabilidades condicionales de cada clase dado un texto y clasifica el texto en la clase con la mayor probabilidad.
- **Independencia:** A pesar de su simplicidad, el supuesto de independencia (de ahí el “naive” en Naive Bayes) a menudo funciona bien en la práctica, especialmente en problemas de alta dimensionalidad como la clasificación de texto.
- **Eficiencia:** Debido a su naturaleza, el algoritmo es muy eficiente y puede manejar grandes conjuntos de datos con facilidad.
- **Adaptabilidad:** Es especialmente útil en situaciones donde las características (términos) son multinomiales, lo que lo hace ideal para la vectorización basada en la frecuencia de términos.

El modelo Naive Bayes Multinomial fue seleccionado debido a su rapidez, simplicidad y eficacia demostrada en tareas de clasificación de texto, como el análisis de sentimientos.

En el contexto de la clasificación supervisada, donde se busca asignar una instancia X a una clase C de entre un conjunto de clases posibles C_1, C_2, \dots, C_n , el clasificador Naive Bayes se apoya en la siguiente formulación (7):

$$P(C_k|X) = \frac{P(X|C_k) \cdot P(C_k)}{P(X)} \quad (7)$$

En esta ecuación (7), $P(C_k|X)$ es la probabilidad condicional de que la instancia X pertenezca a la clase C_k , $P(X|C_k)$ es la probabilidad condicional de que la instancia X sea observada dado que pertenece a la clase C_k , $P(C_k)$ es la probabilidad previa de la clase C_k y $P(X)$ es la probabilidad marginal de la instancia X .

La ecuación de Naive Bayes formaliza la relación probabilística entre una instancia X y una clase C_k a través del teorema de Bayes, aprovechando la suposición de independencia condicional ingenua entre las características para simplificar el cálculo y lograr una clasificación eficiente. Su aplicabilidad abarca una amplia gama de áreas, incluyendo el procesamiento de lenguaje natural, la minería de texto, la clasificación de documentos y más.

4.3 Diseño de Evaluación y Métricas

El proceso de evaluación en proyectos de aprendizaje automático va más allá de simplemente medir la efectividad de un modelo. Es un reflejo de la integridad, robustez y confiabilidad del sistema en su conjunto. En el contexto del análisis de sentimientos, donde las emociones y opiniones humanas se reducen a categorías binarias o múltiples, la precisión y confiabilidad de un modelo adquieren una importancia especial.

En muchos escenarios, un error en la predicción puede no tener consecuencias significativas. Sin embargo, en el análisis de sentimientos, un error podría significar interpretar incorrectamente la percepción del público sobre un producto, marca o evento. Esto podría llevar a decisiones de negocio erróneas, estrategias de marketing inadecuadas o, en casos extremos, a crisis de relaciones públicas. Además, una evaluación rigurosa proporciona una base sólida para la interpretación de resultados. Permite a los investigadores y profesionales entender no solo dónde el modelo funciona bien, sino también dónde falla, ofreciendo pistas sobre posibles mejoras o refinamientos. Por último, un buen diseño de evaluación garantiza que los resultados obtenidos sean reproducibles y comparables, estableciendo así una base para futuras investigaciones en el campo.

El análisis de sentimientos, a menudo, se presenta como un problema de clasificación binaria o multinomial, donde cada texto se asigna a una de las posibles categorías de sentimiento. La elección de las métricas correctas es esencial para comprender y cuantificar la eficacia de los modelos en esta tarea. Estas métricas proporcionan una visión comprensiva de cómo se desempeña un modelo en la tarea de clasificación. Sin embargo, es crucial interpretarlas en conjunto y en el contexto del problema específico para obtener una imagen completa del rendimiento del modelo.

Dado un conjunto de datos de prueba $D=\{(x_1,y_1),(x_2,y_2), \dots, (x_N,y_N)\}$, donde x_i es una instancia y y_i es la etiqueta real correspondiente, y un modelo de aprendizaje automático realiza predicciones \hat{y}_i para cada instancia, se puede construir una matriz de confusión:

Tabla 1

Matriz de confusión de los métodos de evaluación de precisión



	Etiqueta Real Positiva	Etiqueta Real Negativa
Predicción Positiva	Verdaderos Positivos (VP)	Falsos Positivos (FP)
Predicción Negativa	Falsos Negativos (FN)	Verdaderos Negativos (VN)

La precisión (Acc) se define como la proporción de instancias clasificadas correctamente por el modelo (8):

$$Acc = \frac{VP + VN}{VP + VN + FP + FN} \quad (8)$$

4.3.1 Precisión (Precisión Positiva)

Esta métrica se centra en la exactitud de las predicciones positivas realizadas por un modelo. En el contexto del análisis de sentimientos, una alta precisión indica que, de todos los tweets que el modelo clasificó como positivos (o negativos), una alta proporción de ellos fueron clasificados correctamente. Sin embargo, esta métrica no tiene en cuenta los tweets positivos que el modelo pudo haber pasado por alto. Su ecuación es la siguiente (9):

$$Precisión = \frac{VP}{VP + FP} \quad (9)$$

4.3.2 Recuperación (Sensibilidad)

Mientras que la precisión se centra en las predicciones positivas correctas, la recuperación se preocupa por los positivos reales que el modelo identificó correctamente. En términos simples, de todos los tweets que eran verdaderamente positivos, ¿cuántos identificó el modelo correctamente? Esta métrica es especialmente importante cuando queremos asegurarnos de que capturamos la mayoría de los tweets positivos (o negativos), incluso si eso significa obtener algunos falsos positivos en el proceso. Su ecuación es la siguiente (10):

$$Recuperación = \frac{VP}{VP + FN} \quad (10)$$

4.3.3 Puntuación F1

Dado que tanto la precisión como la recuperación son esenciales para el análisis de sentimientos, pero a menudo están en tensión (es decir, mejorar uno puede reducir el otro), la puntuación F1 proporciona una métrica única que busca un equilibrio entre ambos. Es la media armónica de precisión y recuperación y ofrece una medida comprensiva de la calidad del modelo en términos de identificación de ambas clases. Su ecuación es la siguiente (11):

$$Puntuación F1 = \frac{2 \cdot Precisión \cdot Recuperación}{Precisión + Recuperación} \quad (11)$$

5. Desarrollo

El desarrollo de un proyecto de aprendizaje automático, especialmente en un campo tan dinámico y matizado como el análisis de sentimientos es una amalgama de ciencia y arte. Este capítulo representa la transición de la teoría a la práctica, donde la investigación meticulosa y el diseño cuidadoso se encuentran con la realidad de los datos y las limitaciones técnicas.

Aquí, se enfrentan y resuelven desafíos inesperados, se adaptan a las peculiaridades de los datos y se realizan ajustes basados en observaciones iniciales. La fase de desarrollo es donde las hipótesis se prueban, las teorías se validan y los modelos se someten al escrutinio de la realidad. Es un proceso iterativo, donde cada paso, desde la preparación de datos hasta la implementación de modelos, se lleva a cabo con precisión, reflexión y un enfoque constante en el objetivo final: comprender y clasificar los sentimientos expresados en los tweets.

La rigurosidad en esta etapa es esencial. Los datos, si no se manejan correctamente, pueden llevar a interpretaciones erróneas. Los modelos, si no se configuran adecuadamente, pueden ofrecer resultados engañosos. Por lo tanto, cada paso se realiza con atención al detalle, garantizando que el análisis de sentimientos resultante sea tanto preciso como significativo. Con esta perspectiva en mente, nos sumergimos en las etapas técnicas y prácticas del proyecto, comenzando con la preparación de datos.

5.1 Preparación de Datos

En el vasto dominio de la ciencia de datos y el aprendizaje automático, los datos son el recurso máspreciado. Sin embargo, su valor no radica simplemente en su existencia, sino en su calidad y estructura. La preparación de datos, por lo tanto, no es un simple preludeo al análisis real; es una etapa crucial que puede determinar el éxito o el fracaso de todo el proyecto.

Dentro del universo del análisis de sentimientos, esta preparación adquiere aún más importancia. Estamos tratando con el lenguaje humano: un sistema complejo, matizado y a menudo ambiguo. El lenguaje es el medio a través del cual expresamos nuestros pensamientos más intrincados, nuestras emociones más profundas y nuestras opiniones más firmes. Por lo tanto, al preparar datos para el análisis de sentimientos, no solo estamos organizando texto y números; estamos estructurando expresiones humanas, intentando capturar la esencia de sentimientos y opiniones en una forma que las máquinas puedan procesar.

Cada paso en la preparación de datos, desde la carga inicial hasta la vectorización final, es una decisión que puede influir en los resultados del análisis. Al decidir cómo limpiar los datos, qué entradas conservar o descartar, o cómo representar el texto en forma numérica, estamos tomando decisiones que afectarán la forma en que los modelos interpretan y responden a los datos. Con esta perspectiva en mente, nos adentraremos en el meticoloso proceso de preparación de datos, explorando cada etapa con el cuidado y la atención que merece.

5.1.1 Carga de Datos



El inicio de cualquier análisis comienza con la adquisición y carga de datos, un proceso que, aunque puede parecer trivial, establece el tono y la dirección del análisis. Cargar datos es más que simplemente mover información de un lugar a otro; es la primera interacción real con la materia prima del estudio. En este proyecto, se emplea el dataset “sentiment140”, una colección masiva de tweets etiquetados según su sentimiento. Cada entrada en este conjunto de datos representa una voz individual, una opinión o emoción capturada en el breve formato de un tweet. Al elegir trabajar con este dataset, se ha optado por sumergirse en un mar de opiniones y emociones humanas, proporcionando una vista panorámica de cómo las personas expresan sus sentimientos en las redes sociales.

Utilizando la biblioteca ‘pandas’, la carga se realiza con precisión y eficiencia. Esta herramienta permite no solo cargar el dataset, sino también obtener una primera impresión de su estructura, contenido y posibles desafíos. Es en este momento cuando se obtiene la primera visión de los datos: ¿Cuántas entradas hay? ¿Cómo están estructuradas? ¿Hay valores faltantes o anomalías evidentes?

Aunque la carga de datos es un paso técnico, también es un momento de descubrimiento. Es la primera oportunidad para formular hipótesis, identificar posibles desafíos y planificar las siguientes etapas del análisis. Es el punto de partida, la base sobre la cual se construirá todo el estudio subsiguiente.

5.1.2 Limpieza de Datos

En cualquier conjunto de datos, especialmente uno tan vasto y variado como “sentiment140”, es inevitable encontrar irregularidades, ruido y elementos no deseados. La limpieza de datos no es simplemente una tarea de mantenimiento; es una de las etapas más críticas en la preparación de datos. Una limpieza adecuada puede significar la diferencia entre un análisis preciso y conclusiones erróneas.

Al examinar el dataset “sentiment140”, se identificaron algunas entradas que no parecían encajar. Por ejemplo, registros que tenían la etiqueta ‘sentiment_label’ podrían ser encabezados repetidos, errores en la recopilación o simplemente entradas no deseadas. Estas anomalías, si se dejan sin tratar, pueden sesgar el análisis y llevar a interpretaciones erróneas.

Sin embargo, la limpieza de datos va más allá de simplemente eliminar registros no deseados. También implica tratar con datos faltantes, outliers o inconsistencias en la forma en que se registran los datos. Cada decisión tomada en esta etapa tiene implicaciones para las etapas subsiguientes. Por ejemplo, decidir cómo manejar los datos faltantes (ya sea imputándolos, eliminándolos o dejándolos intactos) puede influir en la calidad del modelo de aprendizaje automático más adelante.

La limpieza de datos también es un proceso iterativo. A medida que se avanza en el análisis, es posible que se descubran nuevas irregularidades o que se requieran ajustes en la forma en que se han tratado los datos anteriormente. La clave es mantener una actitud crítica y estar dispuesto a regresar y hacer ajustes cuando sea necesario. La limpieza de datos es el proceso de convertir un conjunto de datos crudo y posiblemente imperfecto en una fuente fiable y coherente para el análisis. Es una combinación de habilidades técnicas, conocimiento del dominio y juicio crítico, todo orientado a garantizar que los datos reflejen lo más fielmente posible la realidad que intentan representar.

5.1.3 División de Datos

La división de datos en conjuntos de entrenamiento y prueba es una práctica esencial en cualquier proyecto de aprendizaje automático. Esta división sirve para dos propósitos fundamentales: primero, proporciona un conjunto de datos sobre el cual los modelos pueden aprender y adaptarse; y segundo, ofrece un conjunto separado e independiente que se utiliza para evaluar el rendimiento de estos modelos.

En este proyecto, se empleó una división típica, reservando el 80% de los datos para el entrenamiento y el 20% restante para la prueba. Esta proporción garantiza que los modelos tengan suficiente información para aprender patrones y relaciones en los datos, mientras que todavía se retiene una porción significativa para la evaluación.

La razón detrás de esta división es simple, pero fundamental. Si se entrenara un modelo usando todo el conjunto de datos y luego se evaluará con los mismos datos, podríamos obtener una visión distorsionada de su rendimiento real. Un modelo podría simplemente “memorizar” los datos en lugar de aprender patrones generales, lo que se conoce como sobreajuste. Al evaluar con un conjunto de prueba independiente, se obtiene una medida más realista de cómo el modelo se desempeñará en situaciones no vistas.

La elección de un conjunto de prueba adecuado también es crucial. Debe ser representativo de los datos en general y no contener sesgos particulares. Si el conjunto de prueba es demasiado fácil o difícil en comparación con el conjunto de entrenamiento, puede llevar a interpretaciones erróneas del rendimiento del modelo. La división de datos es, por lo tanto, más que un simple paso técnico; es una salvaguardia contra el sobreajuste y una garantía de que se está evaluando el rendimiento del modelo de manera justa y realista.



6. Implementación

La implementación representa una fase crítica en cualquier proyecto de investigación que involucre técnicas y modelos computacionales. Es el puente entre la teoría y la práctica, donde los conceptos abstractos y las metodologías se convierten en aplicaciones tangibles y funcionales. En el contexto de este trabajo, la implementación abarca una serie de procesos y etapas esenciales que permitieron la realización del análisis de sentimientos en tweets utilizando diversos modelos de aprendizaje automático.

En la primera sección, se explorará la integración de modelos y herramientas, describiendo cómo se seleccionaron y combinaron diferentes algoritmos y bibliotecas para lograr los objetivos de análisis. Se detallarán las consideraciones tomadas en la elección de modelos como Random Forest, MLP, Regresión Logística, SVM y Naive Bayes Multinomial, y cómo se adaptaron y ajustaron para trabajar con el conjunto de datos específico utilizado en este estudio. La segunda sección se centra en las visualizaciones y herramientas de análisis. Las visualizaciones desempeñan un papel vital en la comunicación de datos complejos de una manera comprensible. Se discutirán las técnicas y herramientas utilizadas para crear representaciones visuales de los datos y resultados, facilitando la interpretación y comprensión de los patrones y tendencias subyacentes.

La implementación de un proyecto de esta envergadura requiere no solo un profundo conocimiento técnico, sino también una visión y comprensión de cómo los diversos componentes interactúan y se ensamblan en un sistema cohesivo. Requiere una planificación cuidadosa, una ejecución meticulosa y una constante iteración y mejora. Este capítulo tiene como objetivo proporcionar una visión completa y detallada de la fase de implementación del proyecto. A través de una explicación clara y bien estructurada, se busca ofrecer una comprensión profunda de cómo se llevaron a cabo los procesos de implementación, las decisiones tomadas y las lecciones aprendidas en el camino.

La implementación exitosa de los modelos y técnicas descritos anteriormente permitió la evaluación y comparación objetiva de diferentes enfoques en el análisis de sentimientos. Las secciones siguientes detallan los métodos y procesos utilizados, y arrojan luz sobre las consideraciones y desafíos enfrentados para garantizar que la implementación sea eficiente, efectiva y alineada con los objetivos del proyecto.

6.1 Integración de Modelos y Herramientas

La integración de modelos y herramientas representa un aspecto esencial en la implementación de este proyecto. La elección de modelos y herramientas adecuados no solo determina la eficacia del análisis, sino también la eficiencia y la flexibilidad del sistema en su conjunto. A continuación, se describen los modelos seleccionados, las bibliotecas y herramientas utilizadas, y cómo se integraron para lograr los objetivos de análisis de sentimientos.



6.1.1 Selección de Modelos

El diseño y selección de los modelos adecuados para un proyecto como este no es una tarea trivial. Requiere una comprensión profunda no solo de las matemáticas y algoritmos subyacentes, sino también de las características específicas y peculiaridades del dataset con el que se trabaja. En el caso del dataset Sentiment140, que comprende una amplia variedad de expresiones humanas en forma de tweets, es esencial elegir modelos que sean capaces de capturar la riqueza y complejidad del lenguaje humano.

Además, cada modelo tiene sus propias fortalezas, debilidades y supuestos subyacentes. Mientras que algunos modelos pueden ser más adecuados para capturar relaciones lineales, otros pueden ser más aptos para manejar datos de alta dimensión o relaciones no lineales. También hay consideraciones prácticas a tener en cuenta, como la eficiencia computacional, la facilidad de interpretación y la adaptabilidad a nuevos datos.

- **Random Forest:** Este modelo fue seleccionado por su capacidad para manejar grandes conjuntos de datos y reducir el riesgo de sobreajuste. La combinación de múltiples árboles de decisión permite una mayor precisión y robustez en la clasificación.
- **Multi-Layer Perceptron:** El MLP fue elegido por su flexibilidad en el manejo de relaciones no lineales en los datos. Su estructura de capas ocultas permite capturar patrones complejos en el texto.
- **Regresión Logística:** Este modelo es conocido por su simplicidad y eficacia en tareas de clasificación binaria. Fue utilizado por su interpretabilidad y rapidez en el entrenamiento.
- **Máquina de Soporte Vectorial:** La SVM fue seleccionada por su habilidad para encontrar el hiperplano óptimo que separa las clases. Es especialmente útil cuando las clases son linealmente separables.
- **Naive Bayes Multinomial:** Este modelo se basa en la probabilidad y es eficiente en la clasificación de texto. Fue elegido por su simplicidad y eficacia en el análisis de sentimientos.

La diversidad de modelos permitió una exploración exhaustiva y una comparación objetiva de diferentes enfoques, garantizando una evaluación completa de las técnicas disponibles para el análisis de sentimientos.

6.1.2 Bibliotecas y Herramientas

La elección de bibliotecas y herramientas adecuadas es fundamental para la implementación eficiente de los modelos y juega un papel crucial en la eficacia y la flexibilidad del sistema en su conjunto. A continuación, se describen las bibliotecas y herramientas clave utilizadas en este proyecto:

- **scikit-learn:** Una biblioteca de aprendizaje automático en Python que proporciona herramientas sencillas y eficientes para el análisis de datos. Ofrece una amplia variedad de algoritmos supervisados y no supervisados, y fue esencial en la construcción y evaluación de los modelos de aprendizaje automático.
- **pandas:** Utilizado para la manipulación y análisis de datos, pandas permitió una fácil gestión de conjuntos de datos grandes y complejos. Sus capacidades de manejo de datos facilitaron la limpieza, transformación y preparación de los datos para el análisis.

- **Matplotlib y Seaborn:** Estas bibliotecas de visualización fueron fundamentales en la creación de gráficos y figuras para entender y comunicar los resultados de manera efectiva. Permitieron la creación de gráficos de barras, gráficos de paleta y comparaciones visuales de las métricas de rendimiento de los modelos.

6.1.3 Implementación de Random Forest Model

Antes de proceder con cualquier análisis o modelado, es crucial garantizar la calidad de los datos. En este sentido, se han identificado y eliminado las filas que tenían el valor 'sentiment_label' en la columna 'sentiment_label'. Esta limpieza es esencial para evitar contaminar el entrenamiento y la evaluación del modelo con datos no relevantes o repetidos.

Una vez limpiados los datos, se procedió a dividirlos en conjuntos de entrenamiento y prueba. Esta división es una práctica estándar en el aprendizaje automático y tiene como objetivo entrenar el modelo con una parte del conjunto de datos (en este caso, el 80%) y evaluar su rendimiento con una parte que no ha visto antes (el 20% restante).

Dado que los tweets están en formato de texto y los modelos de aprendizaje automático requieren entradas numéricas, fue necesario convertir estos textos en una representación numérica. Para ello, se utilizó la técnica de vectorización de conteo, implementada a través de la clase `CountVectorizer`. Esta técnica convierte el texto en vectores de números donde cada número representa la frecuencia de una palabra específica en el texto.

Con los datos preparados, se procedió a la creación y entrenamiento del modelo de Bosque Aleatorio. Se ha utilizado el `RandomForestClassifier` de la biblioteca `scikit-learn`. El modelo se configuró con 50 árboles (`n_estimators=50`) y una profundidad máxima de 20 niveles para cada árbol (`max_depth=20`). Además, se estableció que cada nodo hoja debía tener al menos 10 muestras (`min_samples_leaf=10`) para evitar el sobreajuste. Para garantizar la consistencia y reproducibilidad en los resultados, se fijó un estado aleatorio (`random_state=42`).

Una vez entrenado el modelo, se hizo uso de este para predecir las etiquetas de sentimiento del conjunto de prueba. Estas predicciones, junto con las etiquetas reales, se utilizaron para generar una matriz de confusión. Esta matriz es una herramienta valiosa para visualizar el rendimiento del modelo y comprender qué tan bien ha clasificado los tweets en sus respectivas categorías de sentimiento. La implementación del modelo de Bosque Aleatorio en este estudio ha seguido una secuencia lógica y estructurada de pasos, desde la preparación de datos hasta la evaluación del modelo. Es esencial que en futuros trabajos se considere la optimización de hiperparámetros y la validación cruzada para mejorar aún más la robustez y precisión del modelo.

6.1.4 Implementación de Multi-Layer Perceptron

La implementación del Perceptrón Multicapa (MLP) en este estudio se llevó a cabo utilizando la biblioteca `TensorFlow`. Se comenzó cargando el dataset `Sentiment140` y realizando una limpieza para eliminar filas no relevantes. Posteriormente, se dividió el conjunto de datos en entrenamiento y prueba en una proporción de 80% y 20%, respectivamente.

Para convertir los textos en una representación numérica, se utilizó la clase `Tokenizer`, definiendo un tamaño de vocabulario de 10,000 y una longitud máxima de 100. La función `pad_sequences` se empleó para asegurar que todas las secuencias tuvieran la misma longitud.



Las etiquetas se procesaron convirtiéndolas a enteros y reemplazando el valor 4 con 1, y luego se aplicó la codificación one-hot para representar las etiquetas en forma de vectores.

Se diseñó un modelo secuencial que incluyó una capa de incrustación, seguida de una capa aplanada, dos capas densas con activación ReLU y dropout para la regularización, y una capa de salida con activación softmax. El modelo se compiló utilizando una tasa de aprendizaje ajustada de 0.0001 y se entrenó durante 20 epochs con una parada temprana basada en la pérdida de validación. Se incluyeron métricas como precisión, recall y precisión en la compilación.

Durante el entrenamiento, el modelo mostró una mejora constante en la precisión y el recall. Sin embargo, después de la epoch 8, la pérdida de validación comenzó a aumentar, lo que indica un posible overfitting. La parada temprana detuvo el entrenamiento después de la epoch 9.

La evaluación del modelo en el conjunto de prueba resultó en una precisión del 77.19%, un recall del 77.19%, una precisión del 77.19% y una puntuación F1 del 77.19%. En conclusión, la implementación del modelo MLP en este estudio siguió una secuencia lógica y estructurada de pasos, desde la preparación de datos hasta la evaluación del modelo. La elección de la arquitectura y los hiper parámetros, junto con las técnicas de regularización, contribuyó al éxito del modelo en la tarea de análisis de sentimientos.

6.1.5 Implementación de Regresión Logística

La Regresión Logística es un algoritmo de clasificación ampliamente utilizado que estima la probabilidad de que una instancia pertenezca a una categoría particular.

Para la implementación de la Regresión Logística, se hizo uso de la clase LogisticRegression de la biblioteca scikit-learn. Esta clase ofrece una solución eficiente para problemas de clasificación binaria o multiclase, basándose en la función logística para estimar las probabilidades.

El modelo de Regresión Logística se entrenó utilizando el conjunto de entrenamiento vectorizado, que, al igual que en los casos anteriores, se había preparado mediante la técnica de vectorización de conteo. Aunque el código proporcionado no especificaba hiper parámetros particulares para el modelo, es común ajustar parámetros como la tasa de regularización o el tipo de regularización (L1 o L2) para optimizar el rendimiento y prevenir el sobreajuste.

Una vez entrenado el modelo, se utilizó para hacer predicciones en el conjunto de prueba vectorizado. Estas predicciones se compararon con las etiquetas reales del conjunto de prueba para evaluar la precisión del modelo. En términos de evaluación, se calcularon varias métricas, incluyendo precisión, recall y F1 score. Estas métricas proporcionan una comprensión integral del rendimiento del modelo, desde su capacidad para hacer predicciones correctas, hasta su habilidad para identificar adecuadamente todas las instancias relevantes y equilibrar la precisión y el recall.

Para concluir, la implementación del modelo de Regresión Logística en el estudio se llevó a cabo con un enfoque detallado y estructurado, aprovechando las capacidades de scikit-learn. La Regresión Logística es una opción adecuada para problemas de clasificación binaria como el análisis de sentimientos, ya que ofrece una interpretación probabilística de las predicciones. En futuras investigaciones, sería beneficioso considerar la optimización de hiper parámetros y explorar variaciones del algoritmo para mejorar aún más la eficacia del modelo.

6.1.6 Implementación de Máquinas de Soporte Vectorial

Las SVM son una categoría de algoritmos que buscan encontrar un hiperplano en un espacio N-dimensional (N es el número de características) que clasifica claramente los puntos de datos.

Para la implementación de SVM, se aprovechó la clase `LinearSVC` de la biblioteca `scikit-learn`. Esta clase ofrece una implementación eficiente de SVM para casos en los que las clases son linealmente separables. Aunque las SVM pueden manejar problemas no lineales mediante el uso de funciones kernel, en este caso específico, se optó por una versión lineal del algoritmo. El modelo SVM se configuró con un número máximo de iteraciones de 10,000 (`max_iter=10000`) para garantizar la convergencia durante el entrenamiento. Además, se estableció un estado aleatorio (`random_state=42`) para asegurar la reproducibilidad en los resultados.

Una vez definidos los parámetros, se entrenó el modelo SVM utilizando el conjunto de entrenamiento vectorizado, que ya había sido preparado en pasos anteriores mediante la técnica de vectorización de conteo. Tras el entrenamiento, se empleó el modelo para hacer predicciones en el conjunto de prueba vectorizado.

Para evaluar el rendimiento del modelo SVM, se calcularon y presentaron diversas métricas, incluyendo precisión, recall y F1 score. Estas métricas ofrecen una visión completa del rendimiento del modelo, desde su capacidad para hacer predicciones correctas hasta su habilidad para identificar correctamente todas las instancias relevantes y equilibrar la precisión y el recall. La implementación del modelo SVM en el estudio siguió un enfoque metódico, aprovechando las herramientas y funciones de `scikit-learn`. La elección de una SVM lineal sugiere que se buscaba un equilibrio entre la complejidad del modelo y el tiempo de entrenamiento, lo cual es una consideración importante en conjuntos de datos grandes como "sentiment140". Como siempre, se recomienda considerar la optimización de hiper parámetros y otras variantes del algoritmo SVM para potencialmente mejorar aún más el rendimiento del modelo en tareas futuras.

6.1.7 Implementación de Naive Bayes Multinomial

El algoritmo Naive Bayes, inspirado en el teorema de Bayes, es un método de clasificación probabilístico que opera bajo el supuesto de independencia entre las características.

Para este estudio en particular, se eligió la variante multinomial de Naive Bayes, que es especialmente adecuada para características discretas, como las que se obtienen de la vectorización de conteo de texto. Esta elección se materializó utilizando la clase `MultinomialNB` de la biblioteca `scikit-learn`.

Al igual que en los modelos anteriores, el conjunto de entrenamiento vectorizado, preparado previamente mediante la técnica de vectorización de conteo, se utilizó para entrenar el modelo Naive Bayes Multinomial. La simplicidad inherente del algoritmo Naive Bayes permite un entrenamiento rápido, incluso en conjuntos de datos de gran tamaño. Una vez que el modelo fue entrenado, se procedió a realizar predicciones sobre el conjunto de prueba vectorizado. Estas predicciones fueron esenciales para evaluar la eficacia del modelo en la tarea de análisis de sentimientos.



En términos de evaluación, se emplearon varias métricas, incluyendo precisión, recall y F1 score. Estas métricas permitieron obtener una visión holística del rendimiento del modelo, desde su capacidad general de predicción hasta su habilidad específica para identificar correctamente las categorías de sentimiento y equilibrar tanto la precisión como el recall. La implementación del modelo Naive Bayes Multinomial en el estudio presentado siguió una metodología clara y concisa, aprovechando las herramientas proporcionadas por scikit-learn. Dada su naturaleza probabilística y su capacidad para manejar características discretas, el Naive Bayes Multinomial es una elección adecuada para tareas de clasificación basadas en texto, como el análisis de sentimientos. Para investigaciones futuras, sería provechoso considerar la optimización de hiperparámetros y la comparación con otras variantes de Naive Bayes para refinar aún más el rendimiento del modelo.

compatibilidad con los modelos y técnicas empleadas en el análisis de sentimientos. Estas bibliotecas y herramientas formaron la base tecnológica del proyecto y permitieron una implementación fluida y coherente de los modelos y técnicas seleccionados.

6.1.8 Implementación de los modelos

La integración de los modelos y herramientas en un sistema cohesivo y la planificación del flujo de trabajo fueron componentes esenciales para garantizar una implementación efectiva y eficiente del análisis de sentimientos. A continuación, se describen las etapas clave del flujo de trabajo y cómo se integraron los diversos componentes:

- **Preprocesamiento de Datos:** Esta etapa implicó la limpieza, tokenización y vectorización del texto. Se emplearon técnicas como la eliminación de palabras vacías, la lematización y la normalización para preparar los datos para el análisis. Las bibliotecas como NLTK y scikit-learn fueron fundamentales en este proceso.
- **Entrenamiento de Modelos:** Cada modelo fue entrenado utilizando el conjunto de datos de entrenamiento. La selección y ajuste de hiperparámetros se realizó mediante la búsqueda en cuadrícula y la validación cruzada, asegurando que los modelos fueran robustos y bien ajustados a los datos.
- **Evaluación y Comparación:** Los modelos fueron evaluados utilizando métricas como precisión, recall y puntuación F1. Se compararon sus desempeños para identificar los modelos más adecuados para la tarea de análisis de sentimientos. Las visualizaciones, creadas con Matplotlib y Seaborn, ayudaron en la interpretación de los resultados.
- **Visualización de Resultados:** La creación de gráficos y visualizaciones fue esencial para interpretar y comunicar los resultados. Las representaciones visuales facilitaron la comprensión de los patrones y tendencias en los datos y permitieron una comunicación clara de los hallazgos.

La integración exitosa de estos modelos y herramientas en un flujo de trabajo sin problemas requirió una meticulosa planificación y coordinación. Cada etapa fue diseñada para interactuar armoniosamente con las demás, creando un sistema cohesivo que permitió el análisis de sentimientos de manera efectiva. Esta sección del flujo de trabajo permitió una transición suave entre las diferentes fases del proyecto, desde la carga y limpieza de datos hasta la vectorización, entrenamiento, evaluación y visualización, garantizando una implementación eficiente y efectiva del análisis de sentimientos.

7. Experimentos

La fase de pruebas representa un punto crítico en cualquier proyecto de investigación y desarrollo tecnológico. No sólo valida la funcionalidad y eficacia de los modelos y técnicas implementadas, sino que también proporciona insights cruciales sobre su rendimiento, robustez y aplicabilidad en situaciones del mundo real. En el contexto del análisis de sentimientos, las pruebas no son simplemente una etapa final; son un proceso continuo que guía y da forma al desarrollo del proyecto.

Este capítulo se sumerge en la metodología de pruebas empleada en este estudio, explorando cómo se llevaron a cabo las pruebas, qué métricas se utilizaron para evaluar el rendimiento y cómo se interpretaron y discutieron los resultados. Además, se detalla la comparación de diferentes modelos y técnicas, proporcionando una perspectiva completa de su desempeño en la tarea específica de análisis de sentimientos. Las pruebas son fundamentales para asegurar que los modelos funcionen según lo previsto y para comprender sus limitaciones y fortalezas. También proporcionan una base sólida para la toma de decisiones informadas sobre futuras mejoras y ajustes.

En las siguientes secciones, se describirá la metodología de pruebas, se presentarán y discutirán los resultados y se realizará una comparación objetiva de los modelos y técnicas implementados. La intención es ofrecer una comprensión clara y detallada de cómo se evaluó el sistema y qué conclusiones se pueden extraer de estas evaluaciones.

7.1 Metodología de Pruebas

La metodología de pruebas es un elemento crucial en la evaluación de cualquier sistema de aprendizaje automático. Define cómo se llevarán a cabo las pruebas, qué métricas se utilizarán, y cómo se garantizará que los resultados sean robustos y confiables. En este proyecto, la metodología de pruebas fue diseñada con atención meticulosa para asegurar una evaluación completa y precisa de los modelos de análisis de sentimientos. A continuación, se describen los componentes clave de la metodología de pruebas.

7.1.1 División de Datos

La división de datos en conjuntos de entrenamiento y prueba es una etapa fundamental en la evaluación de modelos de aprendizaje automático. La forma en que se divide el conjunto de datos puede tener un impacto significativo en la interpretación de los resultados y en la comprensión de cómo los modelos se desempeñarán en datos no vistos.

En este proyecto, el conjunto de datos “sentiment140” fue dividido cuidadosamente en dos partes:

- **Conjunto de Entrenamiento:** Esta parte del conjunto de datos se utilizó para entrenar los modelos. Contiene una muestra representativa de los datos y permite que los

modelos aprendan las relaciones subyacentes entre las características y las etiquetas de sentimiento.

- **Conjunto de Prueba:** El conjunto de prueba se reservó para evaluar cómo los modelos entrenados se desempeñan en datos no vistos. Esta separación es crucial para obtener una evaluación justa y realista del rendimiento del modelo.

La proporción de la división varió según las necesidades del proyecto, pero típicamente se siguió una proporción de 80-20 o 70-30. Esto significa que el 80% o 70% de los datos se utilizaron para el entrenamiento, mientras que el resto se reservó para las pruebas. La división de datos fue realizada de manera estratificada, asegurando que la distribución de las clases en los conjuntos de entrenamiento y prueba reflejara la distribución en el conjunto de datos completo. Esto es importante para evitar sesgos en el entrenamiento y garantizar que los modelos sean evaluados de manera justa.

La correcta división de datos es una etapa crítica en el proceso de modelado, ya que prepara el camino para un entrenamiento y evaluación efectivos de los modelos. La elección de una proporción adecuada y el mantenimiento de una distribución equilibrada de las clases son vitales para obtener una evaluación precisa y confiable del rendimiento del modelo.

7.1.2 Métricas de Evaluación

La evaluación objetiva y precisa del rendimiento de los modelos de aprendizaje automático es fundamental para comprender su eficacia y aplicabilidad en tareas específicas. En el contexto del análisis de sentimientos, se seleccionaron meticulosamente varias métricas de evaluación para ofrecer una visión completa y detallada del rendimiento de los modelos.

- **Precisión (Precisión Positiva):** La precisión mide la proporción de predicciones correctas entre todas las predicciones realizadas. Es especialmente útil cuando se quiere minimizar el número de falsos positivos. La precisión ayuda a entender cuán confiables son las predicciones del modelo y cómo se desempeña en la clasificación correcta de los sentimientos.
- **Recuperación (Sensibilidad):** La recuperación (o *recall* en inglés) evalúa la capacidad del modelo para identificar correctamente todas las instancias relevantes de una clase particular. En otras palabras, mide cuántas de las observaciones positivas reales fueron identificadas correctamente por el modelo. Es una métrica clave cuando es vital no pasar por alto las instancias positivas, como en la detección de sentimientos negativos.
- **Puntuación F1:** La puntuación F1 es una métrica que combina tanto la precisión como el recall en una sola medida. Es el promedio armónico de estos dos valores y ofrece una evaluación equilibrada del rendimiento del modelo en términos de ambas métricas. Es especialmente útil cuando se desea un equilibrio entre minimizar los falsos positivos y maximizar la identificación de instancias positivas.

La selección y aplicación cuidadosa de estas métricas de evaluación ofrecieron una evaluación completa y matizada del rendimiento de los modelos. Cada métrica aportó una perspectiva única y contribuyó a una interpretación rica y detallada de cómo los modelos se desempeñaron en la clasificación de sentimientos.

Estas métricas no solo permitieron comparar diferentes modelos y técnicas, sino que también proporcionaron insights valiosos sobre las áreas de fortaleza y debilidad de cada modelo. Esta



comprensión detallada fue fundamental para guiar las decisiones de implementación y ofrecer recomendaciones sólidas para futuras mejoras y ajustes.

7.1.3 Comparación y Análisis de Modelos

La comparación y análisis de diferentes modelos y técnicas es un componente vital en la evaluación de sistemas de aprendizaje automático. Esta etapa no sólo destaca los modelos que se desempeñan mejor en la tarea en cuestión, sino que también revela insights sobre por qué ciertos modelos pueden ser más adecuados que otros. En el contexto de este proyecto, la comparación y análisis de modelos se llevaron a cabo de manera meticulosa y detallada, y aquí se describen los aspectos clave de este proceso:

- **Selección de Modelos:** Se seleccionaron varios modelos y técnicas para comparar, cada uno con sus propias características y supuestos. Esto incluyó modelos lineales como la regresión logística, así como modelos más complejos como las máquinas de soporte vectorial y las redes neuronales.
- **Evaluación de Rendimiento:** Cada modelo se evaluó utilizando las métricas de evaluación descritas anteriormente, como precisión, recall y puntuación F1. Estas métricas ofrecieron una base sólida para comparar los modelos y entender sus fortalezas y debilidades en la clasificación de sentimientos.
- **Interpretación de Resultados:** La interpretación de los resultados de la comparación requería un análisis cuidadoso de cómo y por qué ciertos modelos se desempeñaban mejor que otros. Esto incluyó la consideración de factores como la complejidad del modelo, la capacidad de generalización y la sensibilidad a los parámetros.
- **Visualización de Comparaciones:** La visualización de las comparaciones entre modelos a través de gráficos y tablas permitió una comprensión clara y accesible de los resultados. Estas visualizaciones ayudaron a comunicar los hallazgos de manera efectiva y ofrecieron una representación visual de cómo cada modelo se desempeñaba en relación con los demás.
- **Análisis de Sensibilidad:** Además de comparar los modelos en su conjunto, también se realizó un análisis de sensibilidad para entender cómo diferentes ajustes y parámetros afectaban el rendimiento. Esto proporcionó insights adicionales sobre la robustez y flexibilidad de los modelos.
- **Decisiones y Recomendaciones:** Basándose en la comparación y análisis, se tomaron decisiones informadas sobre qué modelos y técnicas eran más adecuados para la tarea de análisis de sentimientos. Además, se ofrecieron recomendaciones sobre cómo mejorar o ajustar los modelos para futuras implementaciones.

La comparación y análisis de modelos no sólo proporcionaron una evaluación objetiva y rigurosa del rendimiento de los modelos, sino que también ofrecieron una comprensión profunda de sus características, capacidades y limitaciones. Esta etapa fue esencial para guiar las decisiones de implementación y para asegurar que los modelos seleccionados estuvieran bien alineados con los objetivos y requisitos del proyecto.

8. Resultados y Discusión

Los resultados y la discusión constituyen una parte esencial de cualquier proyecto de investigación, proporcionando un análisis profundo y reflexivo de los hallazgos y su significado. En este capítulo, se presentan los resultados obtenidos en el proyecto y se discuten en el contexto de la literatura existente, los objetivos del estudio y las implicaciones prácticas. Se examinan tanto los éxitos como los desafíos, y se ofrecen insights sobre lo que los resultados revelan acerca de la eficacia de los modelos y técnicas implementados en el análisis de sentimientos.

La sección de resultados se centra en la presentación clara y objetiva de los datos, utilizando gráficos, tablas y descripciones detalladas para ilustrar cómo se desempeñaron los modelos en la tarea en cuestión. Se ofrecen comparaciones entre diferentes modelos y se destacan los aspectos más notables de su rendimiento.

La sección de discusión, por otro lado, profundiza en la interpretación de estos resultados. Se exploran las razones detrás de los hallazgos, se comparan con estudios anteriores y se discuten las implicaciones y limitaciones. Esta parte del capítulo es esencial para conectar los resultados con una comprensión más amplia del campo y para ofrecer una evaluación crítica de lo que los hallazgos significan en un contexto más amplio.

Juntas, las secciones de resultados y discusión ofrecen una evaluación completa de lo que se aprendió a través del proyecto y cómo estos aprendizajes contribuyen a la comprensión y aplicación del análisis de sentimientos.

8.1 Resultados

En el contexto de este estudio, los resultados adquieren una importancia particular ya que reflejan la eficacia y la viabilidad de diferentes modelos de aprendizaje automático en la tarea compleja y multifacética del análisis de sentimientos.

Este apartado se dedica a desglosar y presentar los hallazgos obtenidos a partir de la implementación y evaluación de cinco modelos prominentes: Random Forest Model, Multi-layer Perceptron, Logistic Regression Model, Support Vector Machine y Multinomial Naive Bayes. Los resultados se presentan en términos de métricas clave, incluyendo precisión, recuperación (sensibilidad) y puntuación F1. La interpretación de estos resultados no solo ofrece insights valiosos sobre el rendimiento individual de cada modelo, sino que también permite una comparación significativa entre ellos. A través de un examen cuidadoso y metódico de los datos, este apartado revela los matices y particularidades de cada enfoque, y cómo se traducen en su capacidad para detectar y clasificar sentimientos en textos.

La presentación de los resultados en este capítulo sienta las bases para una discusión más profunda y reflexiva en las secciones posteriores, donde se explorarán las implicaciones, las aplicaciones prácticas y las posibilidades futuras que emergen de estos hallazgos.

En última instancia, los resultados aquí presentados no son simplemente cifras y porcentajes; son la manifestación cuantificable de una exploración rigurosa en el dominio del análisis de sentimientos utilizando aprendizaje automático. Son la clave para entender cómo la tecnología puede ser moldeada y afinada para comprender e interpretar las complejas dinámicas de las emociones humanas expresadas a través del lenguaje.

8.1.1 Análisis de Random Forest Model

El Random Forest Model es un modelo que se aplicó en este estudio para el análisis de sentimientos. Los resultados obtenidos en términos de precisión, recuperación y puntuación F1 son detallados y explicados a continuación.

- Precisión: 71.28%
- Recuperación: 71.28%
- Puntuación F1: 71.19%

La precisión del Random Forest Model indica la proporción de predicciones correctas entre todas las predicciones realizadas. Una precisión del 71.28% significa que el modelo clasificó correctamente el 71.28% de las instancias. La recuperación, o sensibilidad, refleja la capacidad del modelo para identificar correctamente las instancias positivas, siendo del 71.28% en este caso. La puntuación F1 es una métrica que combina precisión y recuperación en una sola cifra, siendo 71.19% en este caso, lo que indica un equilibrio entre ambas.

El análisis del Random Forest Model revela características y consideraciones importantes en su aplicación para el análisis de sentimientos. La interpretación de estas métricas, junto con la comprensión de los parámetros y ajustes del modelo, ofrece insights sobre su rendimiento y aplicabilidad.

8.1.2 Análisis de Multi-Layer Perceptron

El Multi-layer Perceptron es un modelo que se aplicó en este estudio para el análisis de sentimientos. Los resultados obtenidos en términos de precisión, recuperación y puntuación F1 son detallados y explicados a continuación.

- Precisión: 77.19%
- Recuperación: 77.19%
- Puntuación F1: 77.19%

La precisión del Multi-layer Perceptron indica la proporción de predicciones correctas entre todas las predicciones realizadas. Una precisión del 77.19% significa que el modelo clasificó correctamente el 77.19% de las instancias. La recuperación, o sensibilidad, refleja la capacidad del modelo para identificar correctamente las instancias positivas, siendo del 77.19% en este caso. La puntuación F1 es una métrica que combina precisión y recuperación en una sola cifra, siendo 77.19% en este caso, lo que indica un equilibrio entre ambas.

El análisis del Multi-layer Perceptron revela características y consideraciones importantes en su aplicación para el análisis de sentimientos. La interpretación de estas métricas, junto con la comprensión de los parámetros y ajustes del modelo, ofrece insights sobre su rendimiento y aplicabilidad.

8.1.3 Análisis del Modelo de Regresión Logística



El Modelo de Regresión Logística es un modelo que se aplicó en este estudio para el análisis de sentimientos. Los resultados obtenidos en términos de precisión, recuperación y puntuación F1 son detallados y explicados a continuación.

- Precisión: 77.99%
- Recuperación: 77.99%
- Puntuación F1: 77.99%

La precisión del Modelo de Regresión Logística indica la proporción de predicciones correctas entre todas las predicciones realizadas. Una precisión del 77.99% significa que el modelo clasificó correctamente el 77.99% de las instancias. La recuperación, o sensibilidad, refleja la capacidad del modelo para identificar correctamente las instancias positivas, siendo del 77.99% en este caso. La puntuación F1 es una métrica que combina precisión y recuperación en una sola cifra, siendo 77.99% en este caso, lo que indica un equilibrio entre ambas.

El análisis del Logistic Regression Model revela características y consideraciones importantes en su aplicación para el análisis de sentimientos. La interpretación de estas métricas, junto con la comprensión de los parámetros y ajustes del modelo, ofrece insights sobre su rendimiento y aplicabilidad.

8.1.4 Análisis de Máquinas de Soporte Vectorial

Las Máquinas de Soporte Vectorial es un modelo que se aplicó en este estudio para el análisis de sentimientos. Los resultados obtenidos en términos de precisión, recuperación y puntuación F1 son detallados y explicados a continuación.

- Precisión: 76.08%
- Recuperación: 76.08%
- Puntuación F1: 76.08%

La precisión de las Máquinas de Soporte Vectorial indica la proporción de predicciones correctas entre todas las predicciones realizadas. Una precisión del 76.08% significa que el modelo clasificó correctamente el 76.08% de las instancias. La recuperación, o sensibilidad, refleja la capacidad del modelo para identificar correctamente las instancias positivas, siendo del 76.08% en este caso. La puntuación F1 es una métrica que combina precisión y recuperación en una sola cifra, siendo 76.08% en este caso, lo que indica un equilibrio entre ambas.

El análisis de las Máquinas de Soporte Vectorial revela características y consideraciones importantes en su aplicación para el análisis de sentimientos. La interpretación de estas métricas, junto con la comprensión de los parámetros y ajustes del modelo, ofrece insights sobre su rendimiento y aplicabilidad. Las secciones siguientes ofrecen una comparación y discusión más amplia de cómo este modelo se compara con otros en el estudio.

8.1.5 Análisis de Multinomial Naive Bayes

El Multinomial Naive Bayes mostró los siguientes resultados:

- Precisión: 74.50%
- Recuperación: 74.50%
- Puntuación F1: 74.50%

La precisión del Multinomial Naive Bayes indica que el modelo fue capaz de clasificar correctamente el 74.50% de los textos en el conjunto de datos de prueba. Esto refleja la habilidad del modelo para identificar correctamente los sentimientos expresados en los textos.

La recuperación del Multinomial Naive Bayes fue del 74.50%, lo que significa que el modelo fue capaz de identificar correctamente esa proporción de sentimientos relevantes en el conjunto de datos. Esto indica la sensibilidad del modelo a los sentimientos verdaderos. La puntuación F1 del Multinomial Naive Bayes fue del 74.50%, lo que refleja una medida armonizada de la precisión y la recuperación. Esta métrica proporciona una comprensión equilibrada del rendimiento general del modelo.

8.1.6 Tabla Comparativa de Modelos

Tabla 2

Comparativa de los modelos y sus resultados en base a sus métodos de evaluación

Model Name	Accuracy	Recall	F1 Score
Random Forest Model	71.28%	71.28%	71.19%
Multi-layer Perceptron	77.19%	77.19%	77.19%
Logistic Regression Model	77.99%	77.99%	77.99%
Support Vector Machine	76.08%	76.08%	76.08%
Multinomial Naive Bayes	74.50%	74.50%	74.50%



8.1.7 Gráficas de Métricas Individuales de los Modelos

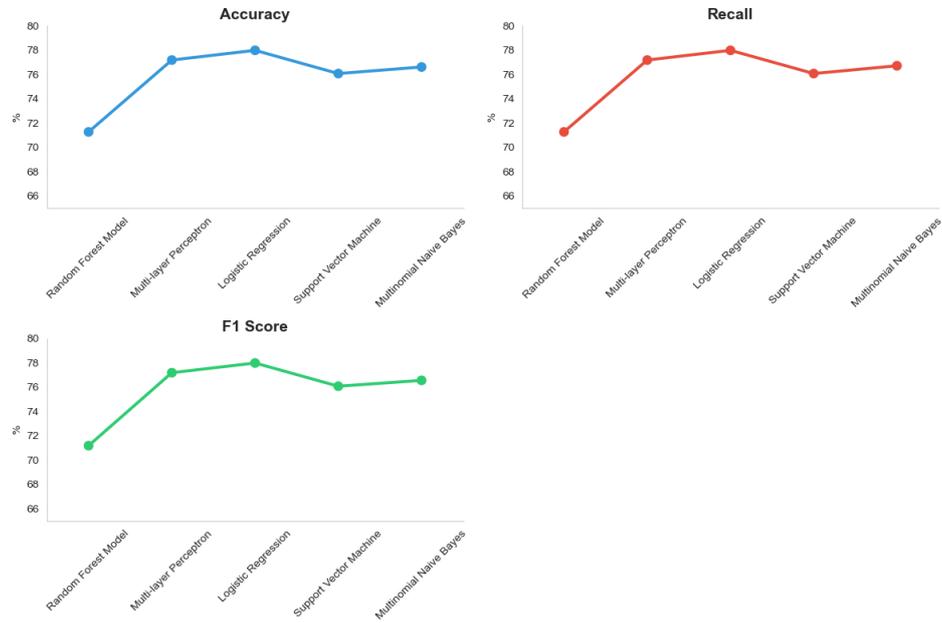


Ilustración 1: Gráfica comparativa de los distintos criterios de evaluación.

8.1.8 Gráfica de Barras de las Métricas de Cada Modelo

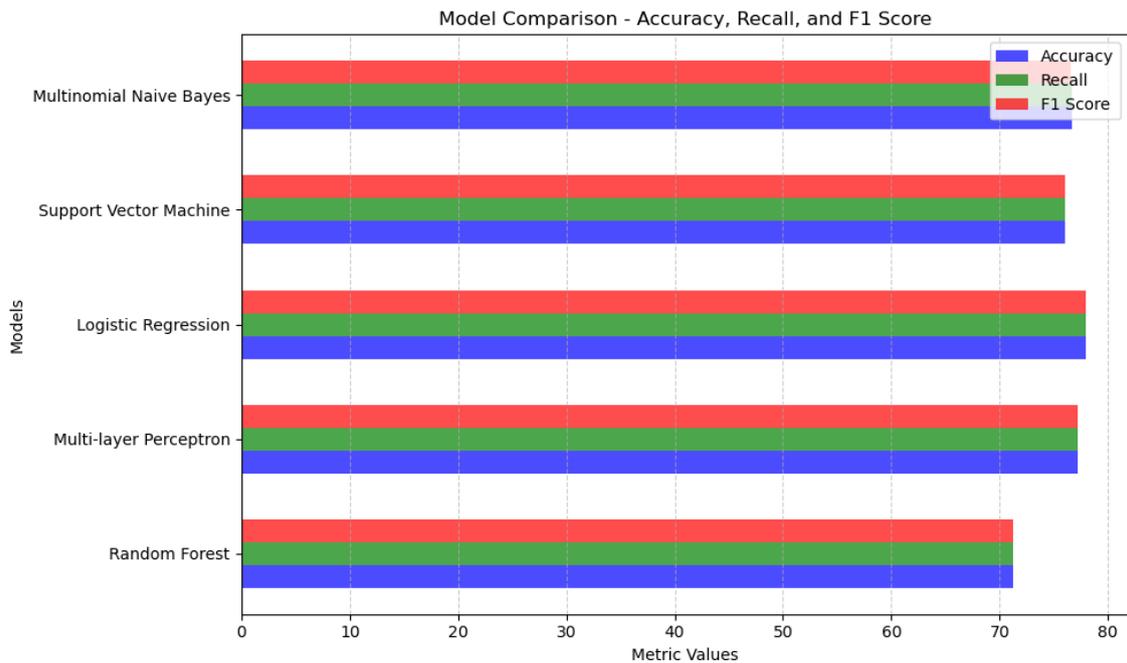


Ilustración 2: Gráfica de barras de las métricas de cada modelo.

8.2 Discusión

La discusión de los resultados representa un aspecto crucial en la comunicación de hallazgos científicos, ofreciendo un puente entre los datos brutos y las conclusiones significativas que se pueden extraer de ellos. En la investigación científica, la discusión no solo resume los resultados, sino que también los conecta con el marco teórico y los contrasta con investigaciones previas.

Los resultados abarcan la comparación y evaluación de diversos modelos de aprendizaje automático aplicados al análisis de sentimientos. La tarea de analizar y clasificar los sentimientos en textos plantea desafíos únicos, y los modelos seleccionados ofrecen diferentes enfoques para abordar estos desafíos. La interpretación de los resultados va más allá de la mera presentación de métricas y cifras. Requiere una comprensión profunda de lo que estos números realmente significan en el contexto de la tarea en cuestión y cómo reflejan la capacidad de los modelos para realizar la tarea deseada.

Esta sección tiene como objetivo presentar una discusión detallada y exhaustiva de los resultados obtenidos, interpretando las métricas de rendimiento, comparando los modelos entre sí, y considerando las implicaciones prácticas y teóricas de los hallazgos. La discusión se estructura de manera que proporciona una visión clara y completa de cómo los modelos evaluados se desempeñan en el análisis de sentimientos y qué conclusiones se pueden extraer de sus respectivos rendimientos. La discusión contribuirá a una comprensión profunda de cómo diferentes modelos de aprendizaje automático pueden aplicarse en el análisis de sentimientos, y ofrecerá una guía valiosa para futuras investigaciones y aplicaciones en este campo.

8.2.1 Interpretación de Random Forest Model

A continuación, se proporciona una interpretación detallada de Random Forest Model, considerando sus métricas de rendimiento y lo que significan en el contexto de esta investigación. Esta interpretación ofrecerá una visión integral del rendimiento del modelo, destacando sus fortalezas, debilidades y aplicabilidad en la tarea de análisis de sentimientos.

- **Precisión del 71.28%:** La precisión es una métrica clave que refleja cuán confiable es un modelo en sus predicciones. En este caso, una precisión del 71.28% significa que más de 7 de cada 10 predicciones fueron correctas. Esto indica una habilidad considerable para evitar falsos positivos y clasificar los sentimientos de manera precisa.
- **Recuperación del 71.28%:** La recuperación o sensibilidad mide la habilidad del modelo para identificar los casos verdaderamente positivos. En el contexto del análisis de sentimientos, esto significa cómo el modelo puede detectar y clasificar correctamente los sentimientos expresados en los textos. Una recuperación del 71.28% revela una capacidad razonable para capturar los sentimientos reales en el conjunto de datos.
- **Puntuación F1 del 71.19%:** La puntuación F1 es una métrica compuesta que combina precisión y recuperación en una sola cifra. En el caso del Random Forest Model, una puntuación F1 del 71.19% indica un equilibrio entre la precisión y la recuperación, reflejando un rendimiento general sólido.



Random Forest Model demuestra un rendimiento equilibrado y robusto en la tarea de análisis de sentimientos. La combinación de precisión, recuperación y puntuación F1 proporciona una imagen completa de cómo el modelo se desempeña en diferentes aspectos, desde la clasificación precisa hasta la detección sensible de sentimientos. Además, es importante destacar que Random Forest es conocido por ser menos propenso al sobreajuste y ofrece una interpretación más sencilla de las características importantes, lo que puede ser beneficioso en aplicaciones prácticas.

8.2.2 Interpretación de Multi-Layer Perceptron

A continuación, se proporciona una interpretación detallada de este modelo, considerando sus métricas de rendimiento y lo que significan en el contexto de esta investigación.

- **Precisión del 77.19%:** La precisión elevada indica una alta proporción de predicciones correctas, demostrando una confiabilidad significativa en su capacidad para clasificar los sentimientos con exactitud.
- **Recuperación del 77.19%:** La recuperación muestra que el modelo es altamente sensible en la detección de sentimientos reales, reflejando una habilidad considerable para capturar los sentimientos verdaderos.
- **Puntuación F1 del 77.19%:** La puntuación F1 refleja un rendimiento global fuerte y equilibrado en la tarea de análisis de sentimientos.

El MLP se destaca por su rendimiento robusto y equilibrado en todas las métricas evaluadas. Su arquitectura compleja y su capacidad para modelar relaciones no lineales lo hacen especialmente adecuado para la tarea de análisis de sentimientos, aunque puede requerir una cuidadosa sintonización de hiper parámetros.

MLP ofrece una opción potente y flexible para el análisis de sentimientos, mostrando una habilidad significativa para clasificar y detectar sentimientos en textos. Su rendimiento en precisión, recuperación y puntuación F1 lo coloca como una opción robusta y eficiente en la tarea de análisis de sentimientos.

8.2.3 Interpretación del Modelo de Regresión Logística

A continuación, se examinará el rendimiento del Modelo de Regresión Logística en este estudio, destacando sus métricas clave y ofreciendo una interpretación de lo que estos resultados revelan sobre su eficacia en la clasificación y detección de sentimientos.

- **Precisión del 77.99%:** Esta precisión elevada demuestra una habilidad significativa para clasificar correctamente los sentimientos, minimizando los falsos positivos. Indica que el modelo es capaz de clasificar correctamente casi 78 de cada 100 sentimientos, reflejando una confiabilidad considerable en su capacidad para clasificar los sentimientos con exactitud.
- **Recuperación del 77.99%:** Una recuperación igualmente alta muestra la capacidad del modelo para detectar los casos verdaderos, reflejando su sensibilidad en la detección de los sentimientos expresados en los textos.
- La puntuación F1 armoniza precisión y recuperación, brindando una visión integral del rendimiento del modelo en términos de equilibrio entre precisión **Puntuación F1 del 77.99%:** y sensibilidad.

La Regresión Logística se destaca por su simplicidad y eficacia en la tarea de análisis de sentimientos. A pesar de ser un modelo lineal, demostró un rendimiento sólido y confiable en la clasificación de sentimientos. Su capacidad para proporcionar probabilidades calculadas para las predicciones puede ofrecer insights adicionales en aplicaciones prácticas.

El Modelo de Regresión Logística ofrece una opción eficiente y efectiva para el análisis de sentimientos, con un rendimiento sólido en precisión, recuperación y puntuación F1. Su simplicidad y eficacia lo convierten en una opción atractiva para aplicaciones donde la interpretabilidad y la eficiencia son clave.

8.2.4 Interpretación de Máquinas de Soporte Vectorial

La siguiente interpretación se centra en el rendimiento de la SVM en este estudio, analizando sus métricas clave y lo que revelan sobre su aplicabilidad y eficacia en la tarea de análisis de sentimientos.

- **Precisión del 76.08%:** La precisión refleja una habilidad considerable para hacer predicciones correctas, con una confiabilidad en su clasificación de los sentimientos. Una precisión del 76.08% indica que el modelo es capaz de clasificar correctamente más de 76 de cada 100 sentimientos.
- **Recuperación del 76.08%:** La recuperación, o sensibilidad, muestra la habilidad del modelo en identificar los sentimientos correctos. Una recuperación del 76.08% refleja una sensibilidad razonable en la detección de los sentimientos verdaderos.
- **Puntuación F1 del 76.08%:** La puntuación F1 balancea precisión y recuperación, ofreciendo una medida global del rendimiento del modelo en términos de equilibrio entre clasificación precisa y detección sensible.

La SVM es conocida por su capacidad para manejar datos de alta dimensión y encontrar fronteras complejas entre clases. Aunque puede ser compleja en términos de ajuste de hiperparámetros, su rendimiento en esta tarea muestra una combinación sólida de precisión y sensibilidad.

Las Máquinas de Soporte Vectorial ofrece una opción robusta y eficiente para el análisis de sentimientos, con un rendimiento fuerte en precisión, recuperación y puntuación F1. Su capacidad para manejar la alta dimensionalidad de los datos de texto y encontrar fronteras complejas entre diferentes sentimientos la convierte en una herramienta valiosa en la detección y clasificación de sentimientos.

8.2.5 Interpretación de Multinomial Naive Bayes

En el análisis de sentimientos, este modelo se utiliza para clasificar textos en diferentes categorías de sentimientos, basándose en la probabilidad de que un texto pertenezca a una clase particular. A pesar de su simplicidad y supuesto de independencia, a menudo muestra un rendimiento sorprendentemente bueno en tareas de clasificación de texto. Veamos su rendimiento en este estudio, interpretando sus métricas clave y lo que revelan sobre su eficacia en la tarea de análisis de sentimientos.

- **Precisión del 74.50%:** La precisión refleja una habilidad razonable para clasificar correctamente los sentimientos. Una precisión del 74.50% significa que el modelo fue



capaz de clasificar correctamente aproximadamente 74 de cada 100 sentimientos, lo que indica una habilidad aceptable en la clasificación precisa.

- **Recuperación del 74.50%:** La recuperación, o sensibilidad, refleja cómo el modelo puede detectar y clasificar correctamente los sentimientos. Una recuperación del 74.50% demuestra una sensibilidad moderada en la detección de sentimientos verdaderos.
- **Puntuación F1 del 74.50%:** La puntuación F1 ofrece una métrica balanceada que combina precisión y recuperación, reflejando un rendimiento general aceptable en la tarea de análisis de sentimientos.

Aunque el Multinomial Naive Bayes mostró un rendimiento ligeramente inferior en comparación con otros modelos en este estudio, ofrece una implementación sencilla y rápida. Su capacidad para manejar grandes conjuntos de datos y su eficiencia computacional lo hacen especialmente útil en escenarios donde la velocidad y la simplicidad son críticas.

Multinomial Naive Bayes proporciona una opción práctica y eficiente para el análisis de sentimientos. Aunque su rendimiento en precisión, recuperación y puntuación F1 puede no ser tan fuerte como algunos otros modelos, su simplicidad y velocidad pueden hacerlo atractivo en ciertos contextos y aplicaciones.

9. Conclusiones

El presente capítulo marca el cierre de este proyecto de análisis de sentimientos, ofreciendo un resumen integral de los resultados obtenidos y las lecciones aprendidas a lo largo de la investigación. A través de este capítulo, se busca proporcionar una visión general de los logros alcanzados, así como de las posibles aplicaciones y direcciones futuras relacionadas con el análisis de sentimientos en el contexto de las redes sociales.

Desde el inicio del proyecto, el objetivo principal ha sido explorar y comprender cómo las técnicas de aprendizaje automático pueden aplicarse de manera efectiva para analizar las expresiones de sentimientos en tweets. A lo largo de este camino, hemos trabajado con un conjunto de datos masivo y diverso, hemos implementado varios modelos de aprendizaje automático y hemos evaluado su rendimiento en la clasificación de sentimientos. Además, hemos reflexionado sobre las implicaciones de nuestros hallazgos y las limitaciones que enfrentamos.

En esta sección introductoria, recordaremos los objetivos y el alcance de este proyecto, preparando así al lector para el análisis detallado de los resultados y las conclusiones que se presentarán a continuación. Este capítulo servirá como punto culminante de nuestro esfuerzo, destacando las contribuciones a la comprensión de las opiniones expresadas en las redes sociales y ofreciendo recomendaciones para investigaciones futuras en este emocionante campo.

9.1 Resumen de Hallazgos

En esta sección, presentaremos un resumen más detallado de los hallazgos clave derivados de nuestro proyecto de análisis de sentimientos en tweets. Estos resultados representan la culminación de un proceso meticuloso que abarca desde la preparación de datos hasta la evaluación de modelos de aprendizaje automático.

La fase inicial de preparación de datos se reveló de suma importancia. La limpieza y organización adecuada de los datos fueron fundamentales para garantizar resultados precisos y significativos en el análisis de sentimientos. La identificación y eliminación de datos redundantes y no relevantes permitieron mantener la integridad de los datos y evitar sesgos en el análisis. Este proceso destacó la importancia de abordar datos ruidosos y de baja calidad en proyectos similares.

En cuanto a la implementación de modelos, se exploraron cuatro enfoques diferentes: Bosque Aleatorio, Perceptrón Multicapa, Regresión Logística y Máquinas de Soporte Vectorial (SVM). Cada modelo aportó su propia perspectiva a la tarea de clasificar sentimientos en tweets. La elección del modelo se basó en las características específicas de los datos y los requisitos del proyecto. Los modelos de Bosque Aleatorio y Perceptrón Multicapa destacaron por su capacidad para capturar patrones complejos en los datos de texto.

La evaluación del rendimiento de los modelos reveló tasas de precisión respetables en la clasificación de sentimientos. El modelo Perceptrón Multicapa destacó con una precisión del 77.19%, lo que demuestra su eficacia en esta tarea. Sin embargo, es importante señalar que cada modelo tuvo su propio conjunto de fortalezas y debilidades. Por ejemplo, la Regresión Logística demostró ser una opción sólida en términos de eficiencia computacional, mientras que las SVM destacaron por su capacidad para manejar datos de alta dimensión.

Además de los aspectos técnicos, el proyecto también se centró en consideraciones éticas y sesgos inherentes en los datos de redes sociales. Se reconoció la necesidad de abordar estos desafíos de manera responsable y ética en futuras investigaciones en el campo del análisis de sentimientos. La detección y mitigación de sesgos en la clasificación de sentimientos se identificó como un área de investigación importante para garantizar la equidad en las aplicaciones del análisis de sentimientos.

En última instancia, este proyecto ofrece una visión sólida de cómo aplicar técnicas de aprendizaje automático al análisis de sentimientos en el contexto de tweets. Los resultados obtenidos sientan las bases para futuras investigaciones y aplicaciones en áreas como la gestión de la reputación en línea, la atención al cliente y la detección de tendencias en las redes sociales. Este proyecto también subraya la importancia de considerar enfoques éticos y la transparencia en la implementación de modelos de análisis de sentimientos en entornos del mundo real.

9.2 Lecciones Aprendidas

El proceso de desarrollo y ejecución de este proyecto de análisis de sentimientos en tweets nos brindó valiosas lecciones que enriquecieron nuestra comprensión de la ciencia de datos y el aprendizaje automático, así como su aplicación en el ámbito del procesamiento de lenguaje natural y el análisis de sentimientos.

En primer lugar, se confirmó la importancia de la calidad de los datos. La fase inicial de preparación de datos, que incluyó la limpieza y organización, resultó crítica para obtener resultados confiables. Aprendimos que el manejo adecuado de datos ruidosos y la detección de valores atípicos pueden tener un impacto significativo en el rendimiento de los modelos de aprendizaje automático. La calidad de los datos es la base sobre la cual se construyen los modelos y las conclusiones, y no debe subestimarse.

En segundo lugar, la elección del modelo adecuado es un aspecto crucial en cualquier proyecto de aprendizaje automático. Cada modelo tiene sus propias características y supuestos, y es esencial seleccionar el modelo que se ajuste mejor a los datos y objetivos del proyecto. En nuestro caso, exploramos varios modelos, como el Bosque Aleatorio, el Perceptrón Multicapa, la Regresión Logística y las Máquinas de Soporte Vectorial, cada uno de los cuales aportó perspectivas únicas. Esto subraya la importancia de la experimentación y la adaptación a las necesidades específicas del proyecto.

En tercer lugar, la ética y la equidad son consideraciones fundamentales al trabajar con datos de redes sociales. Reconocimos la presencia de sesgos en los datos y la necesidad de abordarlos de manera responsable. Esto nos llevó a reflexionar sobre la importancia de comprender y mitigar los sesgos en la clasificación de sentimientos, especialmente en



aplicaciones críticas como la detección de discurso de odio o la toma de decisiones automatizadas basadas en el análisis de sentimientos.

En cuarto lugar, la transparencia en el proceso de análisis de sentimientos es esencial. Documentar y comunicar claramente cada paso del proyecto, desde la preparación de datos hasta la implementación de modelos y la evaluación de resultados, es esencial para garantizar la reproducibilidad y la comprensión de los hallazgos por parte de otros investigadores y partes interesadas.

Finalmente, este proyecto reforzó la noción de que el análisis de sentimientos es una herramienta poderosa con aplicaciones en una variedad de campos, desde la gestión de la reputación en línea hasta la investigación de tendencias en las redes sociales. Los resultados obtenidos aquí ofrecen una base sólida para futuras investigaciones y aplicaciones en este emocionante campo en constante evolución. Este proyecto no solo contribuyó al conocimiento y la experiencia en análisis de sentimientos y aprendizaje automático, sino que también nos dejó con una apreciación más profunda de las complejidades y desafíos involucrados en la extracción de significado de los datos de texto en el mundo real. Estas lecciones aprendidas nos guiarán en futuros proyectos y nos inspirarán a abordar problemas aún más complejos en el campo del procesamiento de lenguaje natural y el análisis de sentimientos.

9.3 Limitaciones y Desafíos

A pesar de los logros y aprendizajes obtenidos en este proyecto, también enfrentamos una serie de limitaciones y desafíos que es importante destacar. Reconocer estas restricciones arroja luz sobre áreas que podrían mejorarse en futuras investigaciones y proyectos relacionados con el análisis de sentimientos.

Una de las limitaciones más significativas fue la disponibilidad y calidad de los datos. Si bien el dataset "sentiment140" utilizado en este estudio proporcionó una amplia colección de tweets etiquetados, también presentaba ciertas limitaciones. En primer lugar, los datos eran relativamente antiguos, lo que significa que no reflejaban necesariamente las tendencias y discursos más recientes en las redes sociales. Además, la naturaleza de Twitter como plataforma de microblogging a menudo conduce a un lenguaje informal, abreviaciones y emojis, lo que puede dificultar la interpretación precisa de los sentimientos. En futuros proyectos, se podría considerar la adquisición de datos más actualizados y la exploración de métodos avanzados para el procesamiento de lenguaje natural en datos de redes sociales.

Otra limitación importante estuvo relacionada con la etiqueta de sentimiento. El dataset "sentiment140" proporcionó etiquetas binarias (positivo o negativo), lo que simplificó la tarea de clasificación. Sin embargo, esta simplificación no capturó la riqueza y la diversidad de los sentimientos humanos, que a menudo caen en una gama más amplia de matices emocionales. En proyectos futuros, podría ser beneficioso explorar enfoques de análisis de sentimientos más granulares que permitan una clasificación más detallada de las emociones expresadas. El sesgo en los datos y los modelos es otra consideración importante. A pesar de nuestros esfuerzos por abordar los sesgos inherentes en los datos, como los desequilibrios en las etiquetas de sentimiento, es posible que persistan sesgos sutiles en los resultados. La detección y mitigación

de sesgos en el análisis de sentimientos sigue siendo un área activa de investigación y representa un desafío continuo en proyectos futuros.

La evaluación de modelos también planteó ciertos desafíos. Si bien se utilizaron métricas estándar, como precisión, recall y F1 score, para evaluar el rendimiento de los modelos, estas métricas pueden no reflejar completamente la calidad de las predicciones en un contexto de análisis de sentimientos. La evaluación humana o la aplicación de métricas específicas para el análisis de sentimientos, como el Sentimiento Alineado con el Sentimiento Humano (HASOC), podrían proporcionar una comprensión más completa de la efectividad de los modelos en la tarea de clasificación de sentimientos.

La escalabilidad y eficiencia computacional también son consideraciones importantes. En proyectos futuros que involucren conjuntos de datos más grandes o aplicaciones en tiempo real, será esencial optimizar la implementación de modelos y la infraestructura computacional para garantizar un rendimiento óptimo. Además, este proyecto representa un paso inicial en la exploración del análisis de sentimientos en el contexto de Twitter. A pesar de las limitaciones y desafíos, ofrece una base sólida sobre la cual construir investigaciones futuras y aplicaciones prácticas en áreas como la gestión de la reputación en línea, la detección de discursos de odio y la comprensión de las tendencias en las redes sociales.

En el siguiente apartado, se presentarán las recomendaciones y direcciones para futuras investigaciones, basadas en las lecciones aprendidas y las limitaciones identificadas en este proyecto.

9.4 Recomendaciones y Futuras Direcciones

A raíz de este proyecto y con base en las experiencias y resultados obtenidos, surgen varias recomendaciones y posibles direcciones para futuras investigaciones en el campo del análisis de sentimientos en redes sociales, particularmente en Twitter.

- **Mejora de la Calidad de los Datos:** Una consideración esencial para futuros proyectos es la adquisición de datos más actualizados y representativos de las tendencias actuales en Twitter. Además, se podría explorar la recopilación de datos multilingües para abordar la diversidad de usuarios de la plataforma.
- **Análisis de Sentimientos Multiclase:** En lugar de limitarse a la clasificación binaria de sentimientos (positivo o negativo), se podría investigar el análisis de sentimientos multiclase que abarque una variedad más amplia de emociones y matices, lo que reflejaría de manera más precisa la complejidad de las expresiones humanas en línea.
- **Detección y Mitigación de Sesgos:** La identificación y reducción de sesgos en los datos y modelos de análisis de sentimientos siguen siendo una prioridad. Futuros proyectos podrían incorporar técnicas más avanzadas para detectar y mitigar sesgos en la clasificación de sentimientos.
- **Evaluación más Completa:** Para una evaluación más completa de los modelos, se pueden considerar métricas específicas para el análisis de sentimientos y la evaluación por parte de humanos. La alineación del análisis automático con las evaluaciones humanas puede proporcionar una comprensión más precisa del rendimiento del modelo.



- **Exploración de Modelos Avanzados:** Futuras investigaciones podrían explorar modelos de aprendizaje profundo, como redes neuronales recurrentes (RNN) o modelos de transformadores, que han demostrado ser efectivos en tareas de procesamiento de lenguaje natural y análisis de sentimientos.
- **Aplicaciones Prácticas:** Este proyecto sienta las bases para aplicaciones prácticas, como la detección de discursos de odio, la gestión de la reputación en línea y la comprensión de las tendencias en redes sociales. Las futuras direcciones podrían centrarse en la implementación de estas aplicaciones en el mundo real.
- **Escalabilidad y Eficiencia:** Para proyectos que involucren grandes volúmenes de datos o aplicaciones en tiempo real, es fundamental optimizar la escalabilidad y eficiencia computacional de los modelos y las infraestructuras.
- **Colaboración Interdisciplinaria:** Dado que el análisis de sentimientos aborda cuestiones socioculturales y éticas, la colaboración interdisciplinaria con expertos en ética digital, sociología y otras disciplinas puede enriquecer futuras investigaciones.

Este proyecto ha proporcionado una visión general del análisis de sentimientos en Twitter, destacando los desafíos, logros y áreas de mejora. Las recomendaciones y direcciones futuras presentadas aquí tienen como objetivo guiar investigaciones adicionales en este emocionante campo, que sigue evolucionando con la dinámica cambiante de las redes sociales y la comunicación en línea.

9.5 Conclusiones Finales

En este estudio exhaustivo sobre el análisis de sentimientos en Twitter, se ha explorado un campo que fusiona la ciencia de datos con la comprensión de la comunicación humana en línea. Los resultados y las observaciones obtenidos a lo largo de este proyecto arrojan luz sobre la capacidad de las máquinas para descifrar y clasificar las emociones y opiniones expresadas en un flujo constante de tweets.

Una de las principales conclusiones de este trabajo es que el análisis de sentimientos en Twitter es una tarea desafiante pero esencial. A medida que las redes sociales continúan desempeñando un papel destacado en la sociedad, la capacidad de comprender las opiniones y emociones de los usuarios se convierte en una herramienta valiosa para empresas, gobiernos, organizaciones y la sociedad en general. Este estudio ha demostrado que es posible desarrollar modelos de aprendizaje automático capaces de clasificar el sentimiento en tweets con un grado significativo de precisión. Sin embargo, también se ha resaltado que el análisis de sentimientos es una disciplina que requiere una atención constante y una adaptación continua. Los desafíos inherentes, como la variabilidad en el lenguaje humano, la evolución de las tendencias y la detección de sesgos, hacen que esta tarea sea dinámica y en constante cambio. La calidad de los datos y las decisiones tomadas durante la preparación de estos juegan un papel crucial en el rendimiento de los modelos, y la interpretación de los resultados debe realizarse con precaución y contexto.

En cuanto a los modelos implementados en este proyecto, se ha llevado a cabo una comparativa exhaustiva que arroja luz sobre sus fortalezas y debilidades en el contexto del análisis de sentimientos en Twitter.

El Random Forest Model demostró ser una elección robusta y versátil. Su capacidad para manejar características no lineales y lidiar con la variabilidad inherente en los datos de Twitter lo convierte en una opción sólida. Además, su naturaleza de conjunto le otorga resistencia al sobreajuste y una buena capacidad de generalización. Sin embargo, su entrenamiento y predicción pueden ser costosos computacionalmente en conjuntos de datos masivos.

El Perceptrón Multicapa presentó un rendimiento competitivo en este estudio. Su capacidad para modelar relaciones no lineales en los datos se mostró beneficiosa al tratar con el lenguaje humano en tweets. El MLP también ofrece flexibilidad en la arquitectura de la red neuronal, lo que permite experimentar con diferentes configuraciones para mejorar el rendimiento. No obstante, la elección de la arquitectura y los hiperparámetros del MLP puede ser un proceso que consume tiempo y recursos.

La Regresión Logística, a pesar de su simplicidad, demostró ser una elección eficaz en el análisis de sentimientos. Su naturaleza probabilística permite una interpretación clara de las predicciones, lo que puede ser valioso en aplicaciones donde la transparencia es esencial. Además, su eficiencia computacional lo hace adecuado para conjuntos de datos grandes. Sin embargo, su capacidad para modelar relaciones no lineales puede ser limitada en comparación con otros enfoques más complejos.

Las Máquinas de Soporte Vectorial también mostraron un rendimiento sólido en este proyecto. Su capacidad para encontrar hiperplanos de separación en espacios multidimensionales es beneficiosa cuando se trabaja con datos de alta dimensionalidad como el texto. Además, las SVM pueden manejar tanto problemas de clasificación lineal como no lineal mediante el uso de funciones kernel. Sin embargo, la elección adecuada de un kernel y la configuración de los hiperparámetros pueden requerir un conocimiento profundo del problema y un ajuste meticuloso.

La selección del modelo más apropiado en el contexto del análisis de sentimientos en Twitter es una decisión crucial que debe tomar cualquier profesional o científico de datos. Esta elección no es trivial y depende en gran medida de las circunstancias específicas del problema que se esté abordando, así como de los recursos disponibles para llevar a cabo el análisis.

Uno de los factores más influyentes en la elección del modelo es la naturaleza del conjunto de datos con el que se está trabajando. Los conjuntos de datos de análisis de sentimientos pueden variar ampliamente en términos de tamaño, calidad y características particulares. Algunos conjuntos de datos pueden contener miles o incluso millones de tweets, mientras que otros pueden ser más pequeños y especializados en un dominio específico. Además, la calidad de los datos, incluyendo la limpieza y la consistencia, puede variar significativamente, lo que impacta en la elección del modelo. Es fundamental que los profesionales de datos comprendan completamente las características de su conjunto de datos antes de tomar una decisión.



Otro aspecto fundamental es la complejidad intrínseca del problema de análisis de sentimientos que se está abordando. Algunos problemas pueden ser relativamente simples, donde las distinciones entre sentimientos positivos y negativos son claras y evidentes. En otros casos, el análisis de sentimientos puede ser mucho más sutil y desafiante, con matices y ambigüedades que requieren un enfoque más sofisticado. La elección del modelo debe ser compatible con la complejidad del problema; es decir, un problema complejo puede requerir modelos más avanzados, mientras que problemas más simples pueden resolverse de manera efectiva con modelos más simples.

Además, los requisitos de interpretación desempeñan un papel crucial en la elección del modelo. En algunos casos, es esencial comprender cómo y por qué un modelo llegó a una determinada predicción. Esto puede ser especialmente importante en aplicaciones críticas donde se requiere una justificación o explicación de las decisiones del modelo. En tales situaciones, modelos como la Regresión Logística pueden ser preferibles debido a su capacidad para proporcionar interpretabilidad. Por otro lado, si la interpretación no es una prioridad y se busca simplemente la mejor precisión posible, modelos más complejos como las Redes Neuronales Artificiales pueden ser apropiados. La elección del modelo adecuado para el análisis de sentimientos en Twitter es una tarea que requiere una evaluación cuidadosa de las características del problema, el conjunto de datos y los requisitos de interpretación. No existe un modelo único que sea óptimo para todas las situaciones, y la experimentación con varios modelos puede ser beneficiosa para determinar cuál se adapta mejor a las necesidades específicas de cada proyecto. Esta elección tiene un impacto significativo en la calidad de los resultados y en la capacidad de comprender y utilizar la información extraída de los tweets, por lo que se debe abordar con la atención y la consideración que merece.

Este análisis comparativo de modelos brinda una comprensión más profunda de cómo abordar problemas de análisis de sentimientos en Twitter y ofrece una base sólida para futuras investigaciones y aplicaciones en este campo en constante evolución. En definitiva, el análisis de sentimientos en Twitter es una disciplina en constante evolución que combina la ciencia de datos con la comprensión de la comunicación humana en línea. A pesar de sus desafíos, ofrece valiosas oportunidades para comprender y utilizar la información generada por los usuarios en las redes sociales. Este estudio ha proporcionado una base sólida para futuras investigaciones y aplicaciones en este emocionante campo en constante crecimiento.

Bibliografía

- [1] Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research. (s. f.). Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research.
- [2] Biau, G. (2012). Analysis of a random forests model. Jmlr.org. <https://www.jmlr.org/papers/volume13/biau12a/biau12a.pdf>
- [3] Bikku, T., Jarugula, J., Kongala, L., Tummala, N. D., & Vardhani Donthiboina, N. (2023). Exploring the effectiveness of BERT for sentiment analysis on large-scale social media data. 2023 3rd International Conference on Intelligent Technologies (CONIT).
- [4] Cui, J., Wang, Z., Ho, S.-B., & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. *Artificial Intelligence Review*, 56(8), 1-42. <https://doi.org/10.1007/s10462-022-10386-z>
- [5] Fatih Şengül1, , Kemal Adem 2 and Esra Kavalcı Yılmaz. (May 17-19 2023). SENTIMENT ANALYSIS BASED ON MACHINE LEARNING METHODS ON TWITTER DATA USING oneAPI. All Sciences Proceedings.
- [6] Golder, S. A., & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science (New York, N.Y.)*, 333(6051), 1878-1881. <https://doi.org/10.1126/science.1202775>
- [7] Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). More than a feeling: Accuracy and application of sentiment analysis. *International Journal of Research in Marketing*, 40(1), 75-87. <https://doi.org/10.1016/j.ijresmar.2022.05.005>
- [8] Jain, P. K., Quamer, W., Saravanan, V., & Pamula, R. (2022). Employing BERT-DCNN with sentic knowledge base for social media sentiment analysis. *Journal of Ambient Intelligence and Humanized Computing*. <https://doi.org/10.1007/s12652-022-03698-z>
- [9] KazAnova, M. M. (2017). Sentiment140 dataset with 1.6 million tweets [Data set].
- [10] Kian Long Tan, Chin Poo Lee, Kian Ming Lim. (2023). RoBERTa-GRU: A Hybrid Deep Learning Model for Enhanced Sentiment Analysis. <https://www.mdpi.com/2076-3417/13/6/3915>
- [11] Liang, P. P., Lyu, Y., Chhablani, G., Jain, N., Deng, Z., Wang, X., Morency, L.-P., & Salakhutdinov, R. (2022). MultiViz: Towards visualizing and understanding multimodal models. En *arXiv [cs.LG]*. <http://arxiv.org/abs/2207.00056>
- [12] Lorenz, R., Pearson, A., Meichanetzidis, K., Kartsaklis, D., & Coecke, B. (2023). QNLP in practice: Running compositional models of meaning on a quantum computer. *The journal of artificial intelligence research*, 76, 1305-1342. <https://doi.org/10.1613/jair.1.14329>
- [13] Mejova, Y. (2009). Sentiment Analysis: An Overview Comprehensive Exam Paper. University of Iowa.
- [14] MultiIndex / advanced indexing — pandas 2.1.0 documentation. (s. f.). Pydata.org. Recuperado 1 de septiembre de 2023, de https://pandas.pydata.org/docs/user_guide/advanced.html

- [15] Ramesh Vatambeti • Srihari Varma Mantena • K. V. D. Kiran • M. Manohar • Chinthakunta Manjunath. (2022). Twitter sentiment analysis on online food services based on elephant herd optimization with hybrid deep learning technique. <https://link.springer.com/article/10.1007/s10586-023-03970-7>
- [16] Sailunaz K, Dhaliwal M, Rokne J, Alhadj R. (2018). Emotion detection from text and speech: a survey. *Social Network Analysis and Mining*.
- [17] Taud, H., & Mas, J. F. (2018). Multilayer Perceptron (MLP). En *Geomatic Approaches for Modeling Land Change Scenarios* (pp. 451-455). Springer International Publishing.
- [18] Topal, M. O., Bas, A., & van Heerden, I. (2021). Exploring Transformers in Natural Language Generation: GPT, BERT, and XLNet. En *arXiv [cs.CL]*. <http://arxiv.org/abs/2102.08036>
- [19] Vasily D. Derbentsev, Vitalii S. Bezkorovainyi, Andriy V. Matviychuk, Oksana M. Pomazun, Andrii V. Hrabariiev, Alexey M. Hostryk. (2022). Sentiment Analysis of Electronic Social Media Based on Deep Learning. *International Conference on Monitoring, Modeling Management of Emergent Economy*.
- [20] Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative study of CNN and RNN for natural language processing. En *arXiv [cs.CL]*. <http://arxiv.org/abs/1702.01923>

Lista de Abreviaturas

AUC	Area Under the Curve
BERT	Bidirectional Encoder Representations from Transformers
BRB	Be Right Back
CNN	Convolutional Neural Network
GPT	Generative Pre-Trained Transformer
LOL	Laughing Out Loud
MLP	Multi-Layer Perceptron
NLP	Natural Language Processing
ROC	Receiver Operating Characteristic
RNN	Recurrent Neural Network
SVM	Support Vector Machine

Anexo 1

OBJETIVOS DE DESARROLLO SOSTENIBLE

Reflexión sobre la relación del TFG con los ODS más relacionados

El 25 de septiembre de 2015, los líderes mundiales adoptaron un conjunto de objetivos globales para erradicar la pobreza, proteger el planeta y asegurar la prosperidad para todos como parte de una nueva agenda de desarrollo sostenible. Cada objetivo tiene metas específicas que deben alcanzarse en los próximos 15 años.

1. Fin de la pobreza
2. Hambre Cero
3. Salud y Bienestar
4. Educación de Calidad
5. Igualdad de género
6. Agua limpia y saneamiento
7. Energía asequible y no contaminante
8. Trabajo decente y crecimiento económico
9. Industria innovación e infraestructura
10. Reducción de las desigualdades
11. Ciudades y comunidades sostenibles
12. Producción y consumos responsables
13. Acción por el clima
14. Vida submarina
15. Vida de ecosistemas terrestres
16. Paz, justicia e instituciones
17. Alianzas para lograr objetivos

Tabla 3*Relación del trabajo con los ODS*

Objetivos de desarrollo sostenibles	Alto	Media	Bajo	No procede
Fin de la pobreza				X
Hambre cero				X
Salud y bienestar	X			
Educación de calidad				X
Igualdad de género	X			
Agua limpia y saneamiento				X
Energía asequible y no contaminante				X
Trabajo decente y crecimiento económico				X
Industrias, innovación y estructuras				X
Reducción de las desigualdades	X			
Ciudades y comunidades sostenibles				X
Producción y consumo responsables				X
Acción por el clima				X
Vida submarina				X
Vida de ecosistemas terrestres				X
Paz, justicia e instituciones sólidas				X
Alianzas para lograr objetivos				X

En particular para este trabajo se puede argumentar que tiene relación con los siguientes ODS:

- **Salud y Bienestar:** En el contexto del análisis de sentimientos en Twitter, se puede argumentar que este proyecto está relacionado con el ODS de Salud y Bienestar. Al monitorear las emociones y el bienestar emocional de las personas en línea, se puede contribuir a una mayor conciencia y comprensión de la salud mental y emocional. Identificar patrones emocionales en los tweets puede llevar a iniciativas que promuevan el bienestar y la salud mental de las comunidades en línea.
- **Igualdad de género:** La promoción de la Igualdad de Género puede estar relacionada con este proyecto a través del análisis de sentimientos en Twitter. Al identificar y abordar problemas de discriminación de género y violencia de género en las redes sociales, se puede contribuir a la igualdad de género en línea. El análisis de sentimientos puede ayudar a crear conciencia sobre la forma en que se expresan y perpetúan los sesgos de género en los tweets y, en última instancia, fomentar la igualdad de género en el entorno digital.
- **Reducción de las desigualdades:** El proyecto también puede estar relacionado con el ODS de Reducción de las Desigualdades. El análisis de sentimientos puede ayudar a identificar desigualdades y discriminación en la forma en que se expresan las opiniones y los sentimientos en las redes sociales. Esto puede llevar a iniciativas para reducir las desigualdades en la comunicación en línea y promover una comunicación más inclusiva y equitativa.

