



Aplicaciones de NLP y Modelos de Atención para la Identificación de Estrés en Textos de Redes Sociales

Francisco Javier Moya Ortiz

Tesis

Asesor

Santiago Hernández Torres

ESCUELA DE CIENCIAS
MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA

MEDELLÍN

2024

Aplicaciones de NLP y Modelos de Atención para la Identificación de Estrés en Textos de Redes Sociales

Francisco Javier Moya Ortiz

fjmoyao@eafit.edu.co

Director: Santiago Hernández Torres

MS. en Matemáticas

shernandez6@eafit.edu.co

Escuela de Ciencias Aplicadas e Ingeniería, Universidad EAFIT, Medellín, Colombia

Resumen

Las redes sociales se han consolidado como plataformas esenciales de comunicación, donde los individuos no solo comparten logros y actividades, sino también sus emociones y estados mentales. Este flujo constante de información brinda una oportunidad única para observar y analizar el bienestar emocional de los usuarios, incluyendo la detección del estrés, un trastorno mental prevalente y frecuentemente incapacitante. A diferencia de otros trastornos como la depresión, el estrés a menudo se manifiesta de manera sutil y sus expresiones lingüísticas son menos comprendidas. Este estudio emplea avanzadas técnicas de procesamiento de lenguaje natural (NLP), modelos de atención como RoBERTa, y grandes modelos de lenguaje (LLMs) para identificar patrones lingüísticos indicativos de estrés. La capacidad de estos modelos para discernir variaciones lingüísticas sutiles ofrece una herramienta valiosa tanto para la investigación académica como para aplicaciones prácticas en salud mental. Los hallazgos muestran que los modelos de atención, como RoBERTa, superan a las técnicas tradicionales de machine learning y a las basadas en conocimiento lingüístico específico, demostrando su capacidad para detectar incluso las señales más sutiles de estrés. Además, los LLMs se perfilan como una solución práctica, logrando un desempeño notable incluso sin ser entrenados específicamente para esta tarea. La automatización en la detección de señales de estrés facilita intervenciones tempranas, mejorando potencialmente los resultados para los individuos afectados. Además, contribuye a superar el estigma y las barreras que impiden buscar ayuda para trastornos mentales, promoviendo un entorno más receptivo y proactivo hacia la salud mental.

Palabras clave: Procesamiento de lenguaje natural (NLP), Modelos de atención, Detección de estrés, Análisis de sentimientos, Salud mental, Redes sociales

1. Introducción

1.1. Redes Sociales y Estrés: Una Perspectiva Analítica

En nuestro mundo cada vez más digitalizado, las redes sociales se han convertido en componentes cruciales de nuestras vidas cotidianas. Plataformas como Twitter, Reddit y Facebook permiten a los usuarios interactuar constantemente, publicando, tuiteando y enviando mensajes innumerables veces al día. Esta interconexión y la naturaleza saturada de información de nuestras existencias hacen que el estrés sea más prominente y observable que nunca antes. Además de funcionar como diarios personales donde las personas registran actividades, preferencias e intereses, las redes sociales se utilizan cada vez más para compartir estados mentales y solicitar apoyo social. Aunque trabajar con muestras no aleatorias y no representativas de usuarios de redes sociales presenta desafíos, estudios previos han identificado marcadores asociados con diversas condiciones, incluyendo depresión, esquizofrenia y estrés relacionado con eventos específicos como viajes o trabajo al analizar contenido de redes sociales (Lin et al., 2014; Ernala et al., 2017; S. C. Guntuku et al., 2017). Sin embargo, estos estudios no siempre logran captar la dimensión crónica del estrés, que puede persistir independientemente de eventos específicos. Por ejemplo, prepararse para un examen puede ser un evento estresante puntual, mientras que sentirse constantemente abrumado por responsabilidades refleja un estrés crónico relacionado con rasgos de personalidad.

Este documento aborda esta brecha de investigación y explora cómo las expresiones lingüísticas en redes sociales pueden indicar niveles de estrés. Se propone un marco para la detección de estrés puntual y específico asociado a textos. No obstante, el estudio del estrés en redes sociales abre la puerta a un seguimiento temporal y a un diagnóstico más personalizado, aunque este proyecto no implementa dicho seguimiento temporal. Se espera que las intervenciones basadas en redes sociales puedan promover estilos de vida de bajo estrés y ofrecer una comprensión más profunda de las variaciones regionales en el estrés.

1.2. Impacto de los Trastornos Mentales y Necesidad de Detección Temprana

Los trastornos mentales abarcan un amplio espectro de condiciones que impactan profundamente en los aspectos cognitivos, emocionales y conductuales de las personas, limitando significativamente su capacidad para interactuar eficazmente con su entorno, aprender y funcionar dentro de la sociedad (Hyman et al., 2006). Estas condiciones, que pueden ser iniciadas o agravadas por factores de riesgo tales como el acoso, el abuso, la violencia y la exposición a estresores crónicos, se encuentran entre las diez principales causas de deterioro de la salud a nivel mundial (Institute for Health Metrics and Evaluation, 2024). Además, afectan a personas de todas las edades, incrementando los costos en salud pública y reduciendo la productividad a nivel global (Institute for Health Metrics and Evaluation, 2024).

A pesar de su impacto devastador, el estigma asociado con los trastornos mentales persiste, obstaculizando la búsqueda de ayuda. Sin embargo, se ha demostrado que las intervenciones tempranas pueden mejorar de manera significativa los resultados para los afectados (Institute for Health Metrics and Evaluation, 2024). Este contexto subraya la urgencia de desarrollar herramientas que permitan una detección temprana y precisa, particularmente en plataformas como las redes sociales. En estas plataformas, las personas tienden a expresar sus emociones y experiencias cotidianas, revelando a menudo signos sutiles de estrés crónico

que podrían pasar desapercibidos en interacciones más formales o clínicas.

El desarrollo de métodos de detección temprana basados en el análisis de expresiones lingüísticas en redes sociales podría revolucionar el campo de la salud mental, proporcionando una base para intervenciones oportunas y personalizadas. Estas herramientas no solo ayudarían a identificar a individuos en riesgo, sino que también contribuirían a reducir el estigma al normalizar la discusión sobre salud mental en espacios digitales.

El estrés tiene una conexión profunda no solo con trastornos mentales como la depresión, el agotamiento profesional (burnout) y el trastorno de estrés postraumático (PTSD), sino que también influye significativamente en diversas condiciones físicas y cognitivas (Marin et al., 2011). Investigaciones muestran que el estrés contribuye al envejecimiento prematuro y es un factor determinante en el desarrollo de déficits cognitivos y afectivos desde etapas tempranas de la vida (Marin et al., 2011; Pechtel y Pizzagalli, 2011). Dada su prevalencia como un factor común en múltiples condiciones patológicas, el estrés emerge como un indicador crítico para otros trastornos mentales. La detección temprana de signos de estrés podría, por tanto, mejorar considerablemente los resultados de los tratamientos, facilitando intervenciones más efectivas y oportunas (Breton et al., 2021).

1.3. Métodos Tradicionales y Nuevos Enfoques para la Detección de Estrés

Tradicionalmente, el estrés se evalúa mediante pruebas psicológicas que intentan discernir el estado mental del paciente (Cox, 1985). Estas pruebas solo pueden ser realizadas por profesionales de la salud y, al basarse en cuestionarios psicológicos, están sujetas a los sesgos propios de los pacientes, quienes podrían no revelar completamente su situación médica. Además, se han realizado estudios que miden el estrés mediante biomarcadores como el cortisol (Marin et al., 2011) o utilizando señales fisiológicas como la respiración, el ritmo cardíaco, la respuesta galvánica de la piel, la electromiografía o el electrocardiograma. No obstante, estos métodos, además de ser invasivos debido a la necesidad de contacto directo con el paciente, presentan desafíos para el seguimiento continuo de la evolución del paciente, lo cual es crucial para desarrollar iniciativas e intervenciones eficaces.

Los trastornos mentales y la exposición al estrés pueden inducir transformaciones notables en la comunicación de las personas, impactando sus capacidades cognitivas, emocionales y expresivas, tanto en el habla como en la escritura (Buchanan et al., 2014; Rude et al., 2004). Estos cambios se manifiestan a través de varios indicadores lingüísticos que pueden ser cruciales para la detección temprana y no invasiva del estrés. Por ejemplo, se ha observado que las personas afectadas por depresión o estrés tienden a usar pronombres en singular con mayor frecuencia, reflejando un enfoque en uno mismo y una disminución en la sensación de pertenencia a un grupo (S. Guntuku et al., 2019).

Por lo anterior, se propone el uso de características lingüísticas que pueden verse afectadas por un estado mental sobrecargado o dificultades de concentración asociadas al estrés. Entre estas características se incluyen la longitud de las palabras y oraciones; en un estado mental alterado, se puede tender a usar un vocabulario reducido. Además, la diversidad léxica en los textos, medida por la Relación Tipo-Token (TTR), tiende a ser más baja, lo que sugiere un vocabulario limitado y repetitivo, potencialmente indicativo de pensamientos obsesivos o restrictivos. La sintaxis y estructura gramatical también se ven afectadas, con un uso reducido de adjetivos y adverbios, lo cual puede disminuir la expresividad emocional.

Notablemente, el incremento en el uso de la voz pasiva por parte de individuos deprimidos o estresados puede señalar una percepción de pasividad o de ser afectados por circunstancias externas, un rasgo común en narrativas de quienes experimentan impotencia o desesperanza. Se busca validar estadísticamente estas características lingüísticas ya que no solo reflejan el impacto del estrés, sino que también funcionan como predictores potenciales del mismo. Este enfoque permite utilizar el análisis lingüístico como un método confiable para predecir trastornos mentales, facilitando intervenciones efectivas antes de que los síntomas clínicos se manifiesten de manera severa. Esto subraya que las manifestaciones del estrés son tanto fisiológicas como emocionales y lingüísticas.

1.4. Aplicación de NLP y Aprendizaje Profundo en la Detección de Estrés

Con el entendimiento de cómo el estrés influye en el lenguaje, es natural avanzar hacia la aplicación de técnicas de Procesamiento de Lenguaje Natural (NLP) para detectar esta condición. El NLP, situado en la intersección entre la inteligencia artificial y la lingüística, permite extraer significados y semánticas de textos utilizando técnicas avanzadas de machine learning (Nadkarni et al., 2011). Esta disciplina ha logrado progresos notables en la extracción de información, análisis de sentimientos, detección de emociones y la traducción de textos. Además, el NLP ha demostrado ser eficaz en la supervisión de estados mentales, contribuyendo significativamente al diagnóstico y monitoreo temprano de trastornos mentales (Fine et al., 2020).

Otros estudios han explorado diversas metodologías para la detección de estrés, adoptando enfoques innovadores con tecnologías emergentes. Winata et al., 2018 implementaron redes Long Short-Term Memory (LSTMs) para identificar señales de estrés utilizando datos de X (anteriormente Twitter) y grabaciones de voz. Por su parte, S. Guntuku et al., 2019 analizaron publicaciones en Facebook y X para identificar usuarios con altos niveles de estrés mediante cuestionarios diagnósticos. Por otro lado, Lin et al., 2017 emplearon un modelo híbrido de redes neuronales convolucionales (CNNs) y grafos factoriales (Factor graph model) para detectar estrés en páginas de blogs y redes sociales. Estos estudios destacan el uso creciente de técnicas de aprendizaje profundo en la evaluación del bienestar emocional a través de plataformas digitales.

Nuestro estudio se inspira en trabajos previos, pero introduce un enfoque innovador al emplear modelos de atención avanzados, específicamente RoBERTa (A Robustly Optimized BERT Pretraining Approach) y grandes modelos de lenguaje o large language models (LLMs), como Llama3, para el análisis de textos en redes sociales, centrándose en publicaciones de Reddit. A diferencia de investigaciones anteriores que a menudo se concentran en datos más breves y etiquetados de manera superficial, como tweets con hashtags, nuestro enfoque se beneficia de la longitud y profundidad de las publicaciones en Reddit. Esto permite una detección más detallada y contextual de indicadores de estrés. RoBERTa fue seleccionado por su capacidad para manejar contextos largos y complejos, proporcionando una comprensión más rica y matizada del texto. Este modelo ofrece ventajas significativas sobre su predecesor, BERT (Bidirectional Encoder Representations from Transformers), que ha sido utilizado en estudios similares empleando el mismo conjunto de datos (Turcan y McKeown, 2019). El uso de RoBERTa no solo mejora la precisión y profundidad del análisis, sino que también permite una comparación directa con estudios anteriores que utilizaron BERT, pro-

porcionando un marco robusto para la comparación y validación de resultados. Por otro lado, se decidió emplear LLMs debido a su capacidad creciente para resolver problemas diversos y su destacado desempeño incluso sin un entrenamiento o ajuste específico para el problema determinado. Además, su reciente integración en tareas cotidianas sugiere que estos modelos tendrán una mayor influencia e integración en el futuro. Se eligió Llama3, un modelo de código abierto, lo que permitió una experimentación y reproducibilidad libre, fortaleciendo así la validez y aplicabilidad de nuestros hallazgos.

El estudio de Turcan y McKeown, 2019, que también utilizó datos de Reddit, es notablemente similar a nuestro enfoque. Sin embargo, nuestro estudio se diferencia en varios aspectos clave. Primero, utilizamos RoBERTa como modelo de atención para la clasificación del texto en estrés o no estrés, lo que ofrece una mayor precisión y profundidad analítica. Además, hemos desarrollado una herramienta práctica basada en RoBERTa para la clasificación automática de textos según niveles de estrés, facilitando intervenciones tempranas y personalizadas en el ámbito de la salud mental. Esta herramienta permite interactuar con un modelo preentrenado de RoBERTa, realizando tareas de clasificación de textos. Además, se proporciona un framework de chatbot, permitiendo una interacción natural que analiza los textos en segundo plano y notifica al usuario cuando detecta un texto con altos niveles de estrés, permitiéndole una autogestión con acciones recomendadas a tomar. Se puede interactuar con estas plataformas en: Inferencia RoBERTa, Chatbot RoBERTa.

La efectividad del modelo se evaluará mediante métricas de rendimiento estándar como precisión, recall y F1-score. Este enfoque no solo refuerza la robustez de nuestro modelo, sino que también subraya las posibles aplicaciones prácticas a largo plazo en la monitorización y mejora del bienestar emocional a través de plataformas digitales.

1.5. Contribuciones del Proyecto y Desarrollo de Herramientas Prácticas

Este proyecto promete aportar varias contribuciones significativas al campo de la salud mental y la tecnología del procesamiento de lenguaje natural. Estas incluyen la validación de hipótesis lingüísticas relacionadas con el estrés y el desarrollo de métodos avanzados para el análisis de sentimientos en contextos de estrés, lo cual podría revolucionar la detección temprana y el tratamiento de los trastornos mentales. Además, el desarrollo de una herramienta basada en Python para la clasificación de textos según niveles de estrés subraya la aplicabilidad práctica de nuestras investigaciones. Esta herramienta no solo valida nuestros enfoques teóricos, sino que también ofrece una solución tangible para profesionales de la salud, permitiendo la implementación de intervenciones más efectivas y personalizadas.

Los resultados esperados de este proyecto incluyen una mayor precisión en la detección de estrés en comparación con métodos anteriores. Esto se logrará mediante el uso de RoBERTa, que permite un análisis más detallado y contextual de los textos. Se espera que la herramienta desarrollada facilite intervenciones tempranas, reduciendo la progresión hacia trastornos más graves y promoviendo un mejor bienestar emocional. Las métricas de rendimiento, como precisión, recall y F1-score, se utilizarán para evaluar la efectividad del modelo, asegurando que los resultados sean robustos y aplicables en entornos clínicos.

2. Metodología y Diseño de Investigación

En este estudio, aplicamos técnicas avanzadas de Procesamiento de Lenguaje Natural (NLP) para clasificar textos según la presencia o ausencia de estrés, con el objetivo de desarrollar modelos que faciliten la identificación temprana del estrés y ayuden a los pacientes a gestionar su estado emocional y acceder a servicios de salud apropiados. La metodología adoptada consta de varios pasos críticos, como se ilustra en la **Figura 1**:

1. **Limpieza de Datos:** Iniciamos con una limpieza exhaustiva del dataset original para eliminar inconsistencias y errores. Este paso es crucial para asegurar la calidad del análisis y la eficacia del entrenamiento de modelos, preparando un dataset preprocesado. La limpieza incluye: eliminación de stopwords, corrección de errores ortográficos, manejo de datos faltantes y normalización del texto. Utilizamos herramientas y bibliotecas específicas como SpaCy y pandas para realizar estas tareas.
2. **Ingeniería de Características:** Continuamos con la ingeniería de características, dividiendo este proceso en dos enfoques : manual y automático.

El **enfoque manual** se basa en identificar y extraer aspectos lingüísticos que se sabe están asociados con el estrés. Estos aspectos incluyen:

- Frecuencia de uso de pronombres en singular, que puede indicar un enfoque en uno mismo y aislamiento.
- Diversidad léxica medida por la Relación Tipo-Token (TTR), reflejando la variedad y repetitividad del vocabulario.
- Longitud de oraciones y palabras, donde una tendencia hacia frases más cortas y palabras simples puede sugerir dificultad en la concentración.
- Frecuencia de construcciones pasivas, que puede indicar una percepción de falta de control.
- Análisis de la sintaxis para detectar patrones gramaticales que se asocian con estados de estrés.

El **enfoque automático** emplea técnicas avanzadas de NLP y aprendizaje automático (ML) para extraer características adicionales. Utilizamos:

- Embeddings que capturan el contexto semántico de las palabras.
- TF-IDF (Term Frequency-Inverse Document Frequency) para evaluar la relevancia de las palabras en el contexto de los documentos.

Estos dos enfoques combinados nos permiten desarrollar un conjunto de características robusto y exhaustivo, que mejora la capacidad de los modelos para detectar indicadores sutiles y complejos de estrés en los textos

3. **Entrenamiento de Modelos:** Con las características extraídas, procedemos a entrenar una variedad de modelos de aprendizaje automático y profundo para evaluar su efectividad en la detección de estrés. Entrenamos varios **modelos de aprendizaje automático**, incluyendo Regresión Logística, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), XGBoost, Random Forest, LightGBM y Naive Bayes. Estos modelos se

entrenan utilizando técnicas de validación cruzada para asegurar que los resultados sean robustos y generalizables. Cada modelo se evalúa en conjuntos de datos de entrenamiento y validación para medir su rendimiento y evitar el sobreajuste. Adicionalmente, utilizamos el texto crudo para el entrenamiento de los **modelos de aprendizaje profundo**, específicamente RoBERTa. RoBERTa es capaz de manejar contextos largos y complejos, proporcionando una comprensión más profunda y matizada del texto. Por su parte, los LLMs, aunque no fueron entrenados específicamente para esta tarea de clasificación binaria o análisis de sentimientos, se ajustaron a través de técnicas de prompt engineering para adaptarlos a esta necesidad. Para optimizar los parámetros de entrenamiento y la configuración de los modelos, utilizamos el Algoritmo Genético de Ordenación No Dominada II (NSGA-II). Este enfoque de optimización se seleccionó debido a su capacidad para realizar una búsqueda informada y eficiente, superando métodos más tradicionales como RandomSearch y GridSearch en términos de efectividad (Ali et al., 2023). NSGA-II permite explorar una amplia gama de combinaciones de hiperparámetros de manera más estratégica, basándose en principios de evolución genética.

4. **Evaluación y Comparación de Modelos:** Utilizamos el F1 score para medir el desempeño de cada modelo debido a su capacidad para equilibrar precisión y recall (Sokolova et al., 2006). En este caso específico, es crucial tener una métrica equilibrada porque tanto los falsos positivos como los falsos negativos pueden tener consecuencias significativas. Los falsos negativos podrían resultar en la falta de detección de individuos que realmente están experimentando estrés, lo cual puede impedirles recibir la ayuda necesaria. Por otro lado, los falsos positivos pueden generar alarmas innecesarias, desviando recursos y atención de otros casos críticos. Además del F1 score, consideramos otras métricas relevantes como precisión y recall.
5. **Despliegue del Modelo:** El modelo seleccionado con el mayor desempeño se integra en un sistema de inferencia implementado en una aplicación web para proporcionar herramientas de diagnóstico accesibles en tiempo real. Los modelos se cargan en HuggingFace, una plataforma y comunidad de machine learning y ciencia de datos que ayuda y soporta a los usuarios a desarrollar, desplegar y entrenar modelos de machine learning (HuggingFace, 2024a). En esta plataforma se establece un endpoint de inferencia que se integra en una interfaz de usuario desarrollada con Streamlit, una herramienta de código abierto diseñada para facilitar la creación de aplicaciones web interactivas y de visualización de datos con Python (Streamlit, 2024). Además, se empleó la API de Groq, que permite realizar inferencia de modelos de machine learning con una baja latencia (Groq, 2024), para la integración de un chatbot alimentado por un gran modelo de lenguaje o LLM.

Esto permite a los usuarios ingresar texto y recibir evaluaciones instantáneas de niveles de estrés. Este proceso asegura una implementación eficiente y accesible, facilitando la detección temprana de estrés a través de una plataforma amigable y optimizada.

La **Figura 1** resume visualmente estos pasos, destacando cómo cada etapa contribuye al desarrollo de soluciones prácticas para la detección de estrés basada en NLP. Este diseño metodológico asegura que cada paso se realice con precisión y contribuya directamente al objetivo del estudio de mejorar la detección y gestión del estrés.

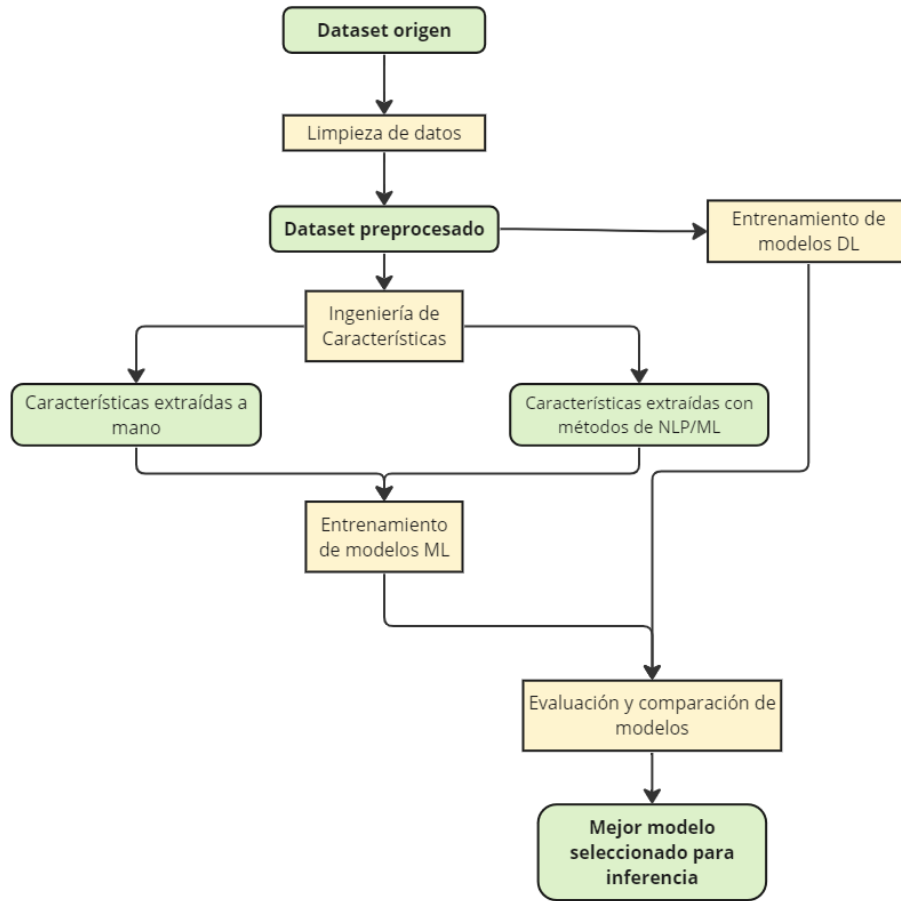


Figura 1: Diagrama de flujo de diseño de investigación

2.1. Recolección y Preparación de Datos

2.1.1. Dataset Original

El dataset utilizado en este estudio es Dreddit, creado por Turcan y McKeown, 2019, que consiste en aproximadamente 190,000 publicaciones de Reddit en inglés. Reddit es un sitio web que actúa como agregador de noticias y plataforma de discusión social, donde los contenidos son proporcionados y evaluados por sus usuarios (Anderson, 2015). Las publicaciones de Reddit, por su naturaleza más extensa y menos condicionada por el uso de hashtags o menciones explícitas de temas, ofrecen un medio ideal para el análisis detallado del estrés a través de indicadores lingüísticos sutiles.

Para este estudio, seleccionamos un subconjunto de 3,553 registros del dataset original, etiquetados específicamente para indicar la presencia o ausencia de estrés mediante el servicio Amazon Mechanical Turk. Cada registro contiene variables como el subreddit específico de la publicación, el contenido textual en inglés y la etiqueta binaria de estrés. Nos enfocamos principalmente en las variables relacionadas con el texto y las etiquetas para nuestro análisis, utilizando únicamente la variable que contiene el texto crudo y la etiqueta asociada. Esta selección permite un análisis concentrado y preciso de los indicadores lingüísticos del estrés en las publicaciones.

	text	id	label
0	he said he had not felt that way before, sugge...	33181	1
1	hey there r/assistance, not sure if this is th...	2606	0
2	my mom then hit me with the newspaper and it s...	38816	1
3	until i met my new boyfriend, he is amazing, h...	239	1
4	october is domestic violence awareness month a...	1421	1

Figura 2: Ejemplo de una entrada del dataset Dreddit mostrando el texto y su etiqueta correspondiente.

2.1.2. Limpieza de Datos

El proceso de limpieza y preparación de datos es crucial para optimizar el análisis textual posterior y el entrenamiento efectivo de modelos de aprendizaje automático. Estos procesos se dividen en dos: para los modelos de aprendizaje automático y para los modelos de aprendizaje profundo.

Modelos de Aprendizaje Automático

1. **Normalización del Texto:** Convertimos todas las letras a minúsculas para garantizar la consistencia en todo el conjunto de datos.
2. **Eliminación de Caracteres Especiales:** Retiramos caracteres que podrían interferir con el análisis textual, como símbolos y puntuación innecesaria. Sin embargo, se mantienen símbolos de puntuación específicos como comas y puntos, ya que son indicadores de parada o separación de oraciones.
3. **Tokenización:** Dividimos el texto en unidades básicas conocidas como tokens, lo que prepara los datos para un análisis más detallado y preciso.
4. **Remoción de Stopwords:** Eliminamos palabras comunes que generalmente no contribuyen significativamente al valor analítico, como preposiciones y artículos. En este paso, se mantienen los pronombres personales en singular, ya que son relevantes para la tarea de clasificación específica.
5. **Lemmatización:** Reducimos las palabras a su forma base o lema, lo que permite comparar efectivamente las diferentes formas de una misma palabra y mejorar la calidad del análisis.
6. **Separación de Conjuntos de Datos:** Dividimos los datos en conjuntos de entrenamiento y evaluación en una proporción de 9:1, asegurando que ambos conjuntos sean representativos y adecuados para su propósito.

Modelos de Aprendizaje Profundo

Para el aprendizaje profundo, utilizamos el tokenizador por defecto de RoBERTa, que se deriva del GPT-2 Tokenizer y utiliza byte-level Byte-Pair-Encoding (BPE) (HuggingFace,

2024b). Este método es altamente eficaz para manejar vocabularios grandes y diversas formas de palabras, permitiendo que el modelo procese texto de manera más eficiente y precisa. El tokenizador de RoBERTa convierte el texto en secuencias de sub-palabras que el modelo puede interpretar y procesar. Este enfoque es particularmente útil para el análisis de textos largos y complejos, ya que mantiene una representación rica y contextual del lenguaje. El tokenizador maneja automáticamente la normalización, la eliminación de caracteres especiales y la tokenización en un solo paso, lo que simplifica el preprocesamiento de datos y asegura que el texto se prepare de manera coherente para el modelo de aprendizaje profundo. Para este modelo se realiza la misma estrategia de separación de conjuntos de datos que en los modelos de aprendizaje automático.

2.2. Ingeniería de Características

2.2.1. Extracción Manual de Características

La extracción manual de características se basa en la premisa de que las características lingüísticas se alteran en personas bajo estrés, lo que resulta en diferencias sutiles en su forma de escribir en comparación con individuos que no están bajo estrés. Se establecieron las siguientes métricas para evaluar estas diferencias:

1. Uso de Pronombres en Singular (*singular_pronouns*)

- Métrica: $P_s = \frac{N_p}{N_w}$, donde P_s es la proporción de pronombres en singular, N_p es el número de pronombres en singular y N_w es el número total de palabras.

2. Longitud de Oraciones y Palabras (*avg_sentence_len / avg_word_len*)

- Métrica a: $L_w = \frac{\sum \text{largo de palabras}}{N_w}$, donde L_w es la longitud promedio de palabras.
- Métrica b: $L_s = \frac{\sum \text{largo de oraciones}}{N_s}$, donde L_s es la longitud promedio de oraciones y N_s es el número total de oraciones.

3. Diversidad Léxica (*lex_diversity*)

- Métrica: $D = \frac{N_u}{N_w}$, donde D es la diversidad léxica y N_u es el número de palabras únicas.

4. Sintaxis y Estructura Gramatical (*adj_adv*)

- Métrica: $S = \frac{N_a + N_v}{N_s}$, donde S es la suma de adjetivos y adverbios por oración, N_a es el número de adjetivos, N_v es el número de adverbios.

5. Frecuencia de Construcciones Pasivas (*passive_avg*)

- Métrica: $F_c = \frac{N_c}{N_s}$, donde F_c es la frecuencia de construcciones pasivas y N_c es el número de oraciones pasivas.

Para cada una de estas métricas, se llevaron a cabo pruebas estadísticas detalladas para verificar las hipótesis. Inicialmente, se utilizó la **prueba de Shapiro-Wilk** para evaluar la normalidad de las distribuciones de cada variable, como fue descrito por Shapiro y Wilk, 1965. La prueba de Shapiro-Wilk es adecuada para muestras pequeñas y se define como:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

donde $x_{(i)}$ son las observaciones ordenadas, a_i son los coeficientes generados a partir de los medios, varianzas y covarianzas de las muestras de la distribución normal, y \bar{x} es la media de las muestras.

Si la distribución de las variables no seguía una distribución normal, lo cual es común en datos lingüísticos, se procedió a utilizar la **prueba de Mann-Whitney** para comparar las distribuciones de dos grupos independientes (personas estresadas y no estresadas) según los métodos establecidos por Mann y Whitney, 1947. La prueba de Mann-Whitney es ideal para muestras no paramétricas y se calcula de la siguiente manera:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

donde n_1 y n_2 son los tamaños de las dos muestras independientes, y R_1 es la suma de los rangos de las observaciones de la primera muestra en la combinación de ambas muestras.

Solo se seleccionaron aquellas características para las cuales se demostró una diferencia estadísticamente significativa entre los individuos estresados y los no estresados, utilizando estos métodos estadísticos. Esto asegura que las características seleccionadas son verdaderamente indicativas de estrés y no productos de variaciones aleatorias en los datos.

2.2.2. Extracción Automática de Características

Term frequency – Inverse document frequency

Para la representación numérica del texto, se empleó la técnica de vectorización Frecuencia de Término - Frecuencia Inversa de Documento o Term frequency – Inverse document frequency (TF-IDF), que es fundamental en el procesamiento de texto para transformar texto en un formato que los modelos de aprendizaje automático pueden procesar eficientemente. La formula para TF-IDF se define como:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t)$$

donde $\text{TF}(t, d)$ es la frecuencia del término t en el documento d , y $\text{IDF}(t)$ es la frecuencia inversa del documento calculada como:

$$\text{IDF}(t) = \log \frac{N}{n_t}$$

con N siendo el número total de documentos en el corpus, y n_t el número de documentos donde aparece el término t . Esta medida ayuda a reducir la importancia de los términos que aparecen muy frecuentemente en el conjunto del corpus, que son típicamente menos informativos que los términos más raros Ramos, 2003.

La relevancia de TF-IDF en nuestro proyecto radica en su capacidad para resaltar palabras que son distintivas y potencialmente más informativas respecto al estado de estrés en comparación con palabras comunes en todos los documentos. Esta capacidad lo hace especialmente adecuado para tareas donde el contexto y la especificidad de los términos son cruciales para una clasificación precisa, como es el caso en la detección de señales de estrés a partir de texto.

La técnica TF-IDF ha sido ampliamente utilizada y validada en la literatura para una variedad de aplicaciones de procesamiento de texto, incluyendo clasificación de documentos, recuperación de información y filtrado de spam. Esta metodología es especialmente efectiva en nuestro contexto debido a que permite una identificación más precisa de términos relevantes en textos largos y complejos típicos de plataformas como Reddit, donde los indicadores de estrés pueden ser sutiles y dispersos a lo largo del texto.

Embeddings: MpNet y DistilRoBERTa

Además de TF-IDF, utilizamos métodos de embeddings, específicamente se emplearon los modelos MpNet y DistilRoBERTa. Los métodos de embeddings destacan porque generan una representación más fiel de los textos, incluyendo el contexto y la similitud de las palabras al codificar el lenguaje natural, permitiendo así una mejor representación de los textos.

MpNet, desarrollado por Song et al., 2020, representa palabras y oraciones en un espacio continuo de alta dimensión donde las palabras y oraciones con contextos similares se encuentran cerca unas de otras. Este enfoque mejora significativamente la capacidad de los modelos para captar relaciones semánticas y contextuales entre palabras y oraciones. El funcionamiento de MpNet se basa en una combinación de preentrenamiento en máscara (masked pre-training) y en el modelo Transformer.

- **Masked Pre-training:** Similar a BERT, MpNet utiliza enmascaramiento para predecir palabras en una secuencia. Este enfoque ayuda a capturar información contextual bidireccional, mejorando la representación de las palabras en distintos contextos.
- **Permuted Language Modeling (PLM):** MpNet también incluye permutaciones de secuencias, lo que permite al modelo aprender dependencias más complejas y de largo alcance en los textos. Esto mejora la capacidad del modelo para comprender relaciones semánticas en contextos variados.

DistilRoBERTa, desarrollado por Sanh et al., 2019 y basado en RoBERTa (Liu et al., 2019), es una versión más ligera y eficiente del modelo original. Se destaca por su capacidad para mantener una alta precisión mientras reduce el tamaño del modelo y el tiempo de inferencia, lo cual es crucial para aplicaciones en tiempo real y con recursos limitados.

- **Modelo Compacto:** DistilRoBERTa reduce aproximadamente el 40 % de los parámetros de RoBERTa sin una pérdida significativa de rendimiento, logrando un balance entre eficiencia y precisión.
- **Mantenimiento de Precisión:** A pesar de su menor tamaño, DistilRoBERTa conserva una gran parte de la precisión de RoBERTa al utilizar un proceso de destilación del conocimiento, donde se entrena el modelo más pequeño (DistilRoBERTa) para imitar el comportamiento del modelo más grande (RoBERTa).

La relevancia de MpNet y DistilRoBERTa en nuestro proyecto radica en su capacidad para capturar relaciones contextuales y semánticas tanto entre palabras como entre oraciones, lo cual es crucial para la detección de estrés. Por ejemplo, frases como 'me siento abrumado' y 'tengo mucha presión' pueden aparecer en contextos similares y tener una relación significativa con el estrés. Al utilizar embeddings de MpNet y DistilRoBERTa, nuestro modelo

puede identificar y aprovechar estas relaciones, mejorando la precisión en la clasificación del texto. Ambos modelos han demostrado ser innovaciones significativas en el procesamiento del lenguaje natural, proporcionando formas eficientes y eficaces de representar el texto que mejoran la comprensión y el análisis del lenguaje en diversas aplicaciones. En nuestro caso, permiten una mejor identificación de patrones lingüísticos sutiles asociados con el estrés, contribuyendo a un modelo más robusto y preciso para la detección de estrés en textos de redes sociales.

2.3. Modelos Aplicados

2.3.1. Entrenamiento de Modelos de Machine Learning

Para la clasificación de texto basada en las características extraídas, se emplearon diversos modelos de aprendizaje automático. Cada modelo tiene sus propias fortalezas y aplicaciones específicas en el campo del procesamiento de texto:

- **Regresión Logística:** Es un modelo estadístico que predice la probabilidad de una variable categórica en función de las variables predictoras. Es especialmente útil para casos binarios, donde la salida es una de dos posibles clases.
- **K-Nearest Neighbors (KNN):** Este algoritmo de aprendizaje supervisado clasifica un nuevo dato basándose en la similitud con los 'K' vecinos más cercanos en el espacio de características. Es simple y eficaz para conjuntos de datos pequeños.
- **Support Vector Machine (SVM):** Busca el hiperplano que mejor divide un conjunto de datos en clases, maximizando el margen entre las diferentes clases. Es potente para datos de alta dimensionalidad.
- **XGBoost:** Una implementación optimizada de gradient boosting que utiliza estructuras de árbol más complejas. Es altamente eficiente para conjuntos de datos grandes y ha demostrado ser muy efectivo en competiciones de ciencia de datos.
- **Random Forest:** Un ensamble de árboles de decisión entrenados con métodos de ensacado (bagging), lo que mejora la estabilidad y precisión del modelo al reducir el sobreajuste.
- **LightGBM:** Un framework de gradient boosting que usa algoritmos basados en histogramas, lo que permite un entrenamiento más rápido y un menor consumo de memoria. Es particularmente útil para grandes conjuntos de datos con alto rendimiento.
- **Naive Bayes:** es un modelo de clasificación basado en el teorema de Bayes, que asume que las características continuas siguen una distribución normal (gaussiana). Este clasificador es particularmente útil cuando se tiene un conjunto de datos donde las características se distribuyen normalmente

Cada uno de estos modelos se entrenó y evaluó utilizando técnicas de validación cruzada para asegurar que los resultados sean robustos y generalizables. Esta metodología permite medir el rendimiento de los modelos y evitar el sobreajuste, asegurando que el modelo final tenga un buen desempeño en datos no vistos.

Para una referencia detallada sobre estos modelos y su aplicación en tareas de clasificación, se recomienda consultar James et al., 2013, que proporciona una descripción exhaustiva de técnicas de aprendizaje automático.

Los modelos fueron entrenados utilizando dos conjuntos de características de entrada diferentes para evaluar el desempeño de cada combinación:

1. Solamente características extraídas manualmente.
2. Solamente características extraídas automáticamente (TF-IDF/Embeddings).

El desempeño de cada modelo fue evaluado utilizando técnicas de cross-validation con $k = 3$ folds. Esta metodología es crucial para validar la robustez del modelo en datos no vistos, minimizando el riesgo de sobreajuste y proporcionando una evaluación honesta del desempeño del modelo. James et al., 2013 también discute la importancia y metodología de cross-validation en el contexto de la evaluación de modelos predictivos, asegurando que los resultados sean generalizables a nuevos conjuntos de datos.

2.3.2. Entrenamiento de Modelos de Deep Learning

En la categoría de deep learning, seleccionamos RoBERTa (Robustly Optimized BERT Approach) y Llama3 para la clasificación de texto. RoBERTa es una variante optimizada del modelo BERT (Bidirectional Encoder Representations from Transformers), que ha sido mejorada para lograr un rendimiento superior mediante ajustes en el preentrenamiento del modelo original BERT. Llama3 es un modelo de lenguaje grande (LLM) que ha demostrado una capacidad destacada para el procesamiento de lenguaje natural a gran escala.

BERT: Bidirectional Encoder Representations from Transformers

BERT, desarrollado por Devlin et al., 2018, es un modelo de lenguaje bidireccional que utiliza Transformers para entender el contexto de una palabra basándose en todas las palabras de su entorno, tanto anteriores como posteriores. Esto lo diferencia de los modelos unidireccionales anteriores, que solo consideraban el contexto de una palabra desde una única dirección (izquierda a derecha o derecha a izquierda). BERT introduce dos tareas de preentrenamiento:

- **Máscara de Modelado de Lenguaje (Masked Language Modeling, MLM):** Algunas palabras en una oración son enmascaradas al azar y el modelo debe predecirlas.
- **Predicción de la Siguiente Oración (Next Sentence Prediction, NSP):** El modelo debe predecir si una oración sigue a otra en el texto original, lo que ayuda a comprender las relaciones entre oraciones.

Estas técnicas permiten que BERT capture una rica representación contextual del lenguaje natural, lo cual es crucial para tareas como la comprensión de lectura, la clasificación de texto y la respuesta a preguntas.

RoBERTa: Robustly Optimized BERT Approach

RoBERTa, introducido por Liu et al., 2019, es una mejora sobre BERT que optimiza su rendimiento ajustando varios aspectos del preentrenamiento. Las principales innovaciones de RoBERTa incluyen:

- **Eliminación de la Tarea de NSP:** Se elimina la tarea de predicción de la siguiente oración, basada en investigaciones que sugieren que no mejora el rendimiento.
- **Entrenamiento con Datos Más Amplios y Largos:** RoBERTa se entrena con lotes de datos más grandes y con secuencias de texto más largas, lo que permite una mejor captación del contexto.
- **Mayor Cantidad de Datos y Tiempo de Entrenamiento:** RoBERTa se entrena durante más tiempo y con más datos que BERT, utilizando fuentes adicionales para aumentar la cantidad y diversidad del entrenamiento.
- **Ajustes en los Hiperparámetros:** Se optimizan los hiperparámetros del modelo, como la tasa de aprendizaje y el tamaño del lote, para mejorar el rendimiento.

Estas mejoras permiten que RoBERTa capture contextos más ricos y complejos en los textos, haciendo que sea particularmente efectivo para tareas de comprensión y clasificación de texto complejo. Para adaptar RoBERTa a nuestra tarea específica, realizamos un proceso de fine-tuning utilizando los datos de nuestro estudio. Durante el fine-tuning, ajustamos los parámetros del modelo preentrenado a nuestro conjunto de datos específico durante un total de 5 epochs, permitiendo que el modelo ajuste sus pesos a las sutilezas lingüísticas y patrones específicos del estrés en las publicaciones textuales. En este proceso, se modificó la última capa de la red para adaptarla a la tarea de clasificación binaria, asegurando que el modelo pueda distinguir eficazmente entre textos con y sin estrés. Se utilizaron los pesos iniciales por defecto de RoBERTa, los cuales fueron posteriormente ajustados durante el entrenamiento. Este enfoque de fine-tuning aprovecha el poder del transfer learning, reutilizando un modelo desarrollado para una tarea general de lenguaje como punto de partida para una tarea específica y relacionada, maximizando así la eficacia del modelo en la detección de estrés.

Al igual que con los modelos de machine learning, la evaluación del desempeño de RoBERTa se llevó a cabo utilizando el conjunto de entrenamiento configurado para cross-validation con $k = 3$ folds. Esta técnica de validación cruzada es crucial para evaluar cómo el modelo generalizará a un conjunto de datos no visto anteriormente, proporcionando una medida robusta de rendimiento y ayudando a evitar el sobreajuste.

RoBERTa ha demostrado ser excepcionalmente efectivo para una amplia gama de tareas de procesamiento de lenguaje natural, según lo documentado en el estudio por Liu et al., 2019. Este estudio subraya cómo las modificaciones en la metodología de preentrenamiento pueden resultar en mejoras significativas sobre el modelo BERT original en tareas de benchmarking como GLUE, RACE, y SQuAD. Liu et al., 2019 también discute la metodología y eficacia de RoBERTa en varios contextos de NLP, estableciendo un precedente para su aplicación en la clasificación de texto y análisis semántico.

Modelos de Lenguaje Grande (LLMs)

Los modelos de lenguaje grande (LLMs) son una clase de modelos de aprendizaje profundo diseñados para entender y generar lenguaje natural. Estos modelos se entrenan con grandes volúmenes de datos textuales y utilizan arquitecturas avanzadas de transformers para aprender representaciones contextuales profundas del lenguaje. Los LLMs, como GPT-3, han revolucionado el campo del procesamiento de lenguaje natural (NLP) al proporcionar capacidades superiores en tareas como traducción, generación de texto, y respuesta a preguntas. Los LLMs son altamente versátiles y pueden adaptarse a una amplia gama de tareas con ajustes

mínimos, lo que los hace extremadamente útiles para aplicaciones en tiempo real y de gran escala.

Ingeniería de prompts

La ingeniería de prompts o prompt engineering es el proceso de diseñar y refinar los prompts utilizados para generar texto a partir de modelos de lenguaje, como GPT-3 o modelos similares. El objetivo de esta disciplina es mejorar la calidad y relevancia del texto generado mediante la elaboración cuidadosa de los prompts, de manera que se obtengan las respuestas deseadas del modelo. La ingeniería de prompts implica varios pasos esenciales. Primero, se debe seleccionar la arquitectura y los parámetros apropiados del modelo. Luego, se diseña el formato y la estructura del prompt, se elige la tarea y los datos de entrenamiento adecuados, y finalmente, se afina el modelo utilizando el prompt y los datos seleccionados.

Este proceso es crucial en el desarrollo de modelos de lenguaje, ya que puede influir significativamente en la calidad y efectividad de las respuestas del modelo. Al diseñar y refinar cuidadosamente los prompts, los investigadores y desarrolladores pueden mejorar la precisión y relevancia de las salidas del modelo, haciéndolo más útil para una amplia gama de aplicaciones, incluyendo chatbots, traducción de idiomas, creación de contenido, entre otros (OpenAI, 2024).

Existen diversas técnicas de ingeniería de prompts. Entre las más simples se encuentra el *Zero-shot*, en el cual se dan instrucciones simples al modelo. Por ejemplo: “clasifica el sentimiento del siguiente texto en positivo o negativo”. En contraste, en la técnica *Few-shot* se utilizan instrucciones en lenguaje natural y se adicionan ejemplos del comportamiento esperado, proporcionando al modelo una visión de la distribución de las etiquetas o la lógica de clasificación, lo cual puede producir mejores resultados.

Llama3

Llama3 es un modelo de lenguaje grande (LLM) de última generación desarrollado por Meta AI. Diseñado para ser altamente escalable y eficiente, Llama3 utiliza una arquitectura de transformers con mejoras significativas en la atención y la representación contextual. Este modelo se entrena con un conjunto de datos masivo y diverso, lo que le permite capturar una amplia gama de patrones lingüísticos y contextuales. A diferencia de modelos anteriores, Llama3 integra técnicas avanzadas de preentrenamiento que optimizan su capacidad para comprender y generar texto de manera coherente y precisa. Su arquitectura mejorada permite manejar tareas complejas de procesamiento de lenguaje natural sin necesidad de un entrenamiento específico para cada tarea, haciendo uso de técnicas de prompt engineering para adaptarse rápidamente a nuevas aplicaciones (Meta, 2023).

En este proyecto se empleó la versión más grande disponible de Llama3 (70B) y se implementó una estrategia de prompt engineering para ajustar este modelo a la tarea de clasificación binaria asociada a la detección de estrés en textos. Esto es crucial para maximizar la eficacia de Llama3 en la detección de estrés, dado que los LLMs no siempre requieren un entrenamiento específico para cada tarea, pero pueden beneficiarse enormemente de prompts bien diseñados que guíen su procesamiento y generación de texto. Se emplearon técnicas de *Zero-shot* y *Few-shot*. A continuación se ejemplifican los prompts empleados con Llama3:

- Prompt 1: *You are a classification model that labels text as 1 for stress and 0 for non-stress. You must respond with only a single digit, 1 or 0. Classify the following text as 1 for stress and 0 for non-stress: {input text}*

- Prompt 2: *You are a classification model that labels text as 1 for stress and 0 for non-stress.*

Examples of stress: example 1... example n

Examples of non-stress: example 1... example n

You must respond with only a single digit, 1 or 0. Classify the following text as 1 for stress and 0 for non-stress: {input text}

Llama3 fue utilizado de la siguiente manera: primero, se estableció una conexión al modelo mediante la API de Groq. A continuación, se especificó el prompt a utilizar. Para cada texto en el conjunto de prueba, se realizó una predicción pasando el prompt específico junto con el texto a clasificar, obteniendo así la clase correspondiente. Cabe destacar que, en el caso de los prompts con la estrategia few-shot, los ejemplos se extrajeron del conjunto de entrenamiento.

La relevancia de RoBERTa en nuestro proyecto radica en su capacidad para manejar contextos complejos y largos, proporcionando una herramienta poderosa para la detección de estrés en textos de redes sociales. RoBERTa mejora sobre su predecesor, BERT, con una mejor representación contextual y una mayor precisión. Esta capacidad es esencial para capturar las sutilezas lingüísticas y contextuales que indican estrés en textos largos y detallados, como las publicaciones en Reddit.

Por su parte, Llama3 permite realizar una clasificación out of the box, sin la necesidad de un proceso exhaustivo de entrenamiento y ajuste. Esto convierte a Llama3 en una alternativa altamente prometedora para la integración en sistemas ya existentes, como las plataformas de redes sociales. Gracias a su capacidad de adaptarse rápidamente a nuevas tareas mediante técnicas de prompt engineering, Llama3 facilita la identificación y el apoyo a un mayor número de individuos potencialmente afectados por el estrés. Esta eficiencia y versatilidad lo hacen una herramienta valiosa para la detección de estrés en tiempo real, ampliando significativamente el alcance y el impacto de las intervenciones de salud mental.

2.4. Evaluación de Modelos

La evaluación de los modelos en nuestro estudio se llevó a cabo utilizando el F1 Score, una métrica que proporciona un balance entre la precisión y el recall (sensibilidad). Esto es especialmente crucial en la detección de condiciones como el estrés en textos, donde es importante evitar tanto los falsos negativos como los falsos positivos (Sokolova et al., 2006). La etiqueta positiva (1) representa los registros con estrés, mientras que la etiqueta negativa (0) representa los registros sin estrés. La evaluación se realizó sobre el conjunto de evaluación.

El F1 Score es particularmente valioso en contextos donde las clases están desequilibradas o cuando se requiere igual importancia en la minimización de falsos negativos y falsos positivos. La fórmula para calcular el F1 Score es:

$$F1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Recall}}{\text{Precisión} + \text{Recall}}$$

Donde:

$$\text{Precisión} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

En estas fórmulas:

- *TP* (True Positives) son los casos correctamente identificados como positivos.
- *FP* (False Positives) son los casos incorrectamente identificados como positivos.
- *FN* (False Negatives) son los casos positivos que no fueron identificados.

Utilizamos el F1 Score para medir el desempeño de cada modelo debido a su capacidad para equilibrar precisión y recall. Este equilibrio es crucial en la detección de trastornos como el estrés porque garantiza que el modelo no solo identifique la mayoría de los casos reales de estrés (alta sensibilidad), sino que también mantenga un bajo número de falsas alarmas (alta precisión). Tanto los falsos negativos como los falsos positivos pueden tener consecuencias significativas. Los falsos negativos podrían resultar en la falta de detección de individuos que realmente están experimentando estrés, impidiéndoles recibir la ayuda necesaria. Por otro lado, los falsos positivos pueden generar alarmas innecesarias, desviando recursos y atención de otros casos críticos. Este balance es fundamental en aplicaciones prácticas donde las decisiones o intervenciones se basan en los resultados del modelo, asegurando así que las acciones tomadas sean efectivas y eficientes.

Además del F1 Score, reportamos las métricas de precisión y recall para proporcionar una visión más completa del rendimiento del modelo. La precisión nos ayuda a entender la proporción de verdaderos positivos entre todos los casos detectados como positivos, mientras que el recall mide la proporción de verdaderos positivos entre todos los casos que realmente son positivos. Estas métricas adicionales proporcionan una visión más completa del rendimiento de cada modelo, permitiéndonos identificar fortalezas y debilidades específicas.

En el contexto de los datos de texto, donde la expresión de estrés puede variar ampliamente y la contextualización de los términos es esencial, el F1 Score ayuda a asegurar que el modelo sea efectivo en capturar la sutileza y variabilidad del lenguaje relacionado con el estrés.

3. Resultados

3.1. Análisis exploratorio de los datos

Una verificación inicial del balance de etiquetas en el dataset confirmó que hay un equilibrio adecuado entre las categorías de 'Estrés = 1' y 'No Estrés = 0', como se visualiza en la **Figura 3**. Este balance es fundamental para evitar sesgos en el entrenamiento y la evaluación de modelos de machine learning, asegurando que las conclusiones sean válidas para ambas condiciones.

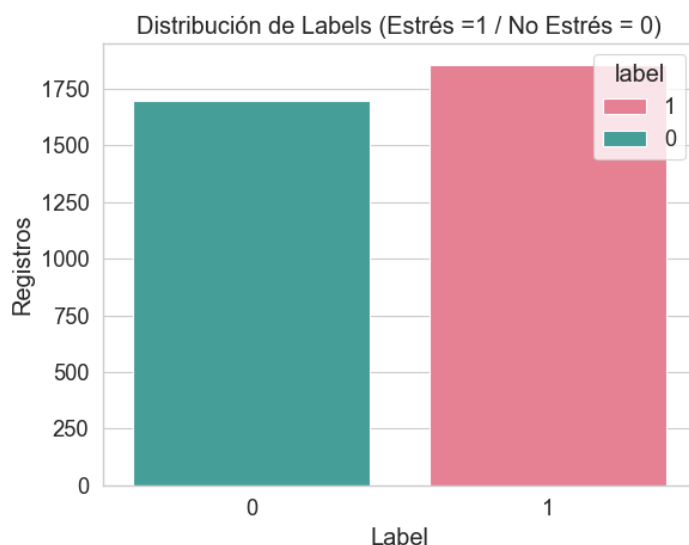


Figura 3: Distribución de las etiquetas en el conjunto de datos

Además, la longitud de las publicaciones, que varía de 6 a 310 palabras, muestra una diversidad significativa en el modo en que los usuarios expresan sus estados emocionales en textos de longitud variable (**Figura 4**). Esta variabilidad es crucial para desarrollar modelos que pueden generalizar eficazmente a través de diferentes estilos y longitudes de texto.

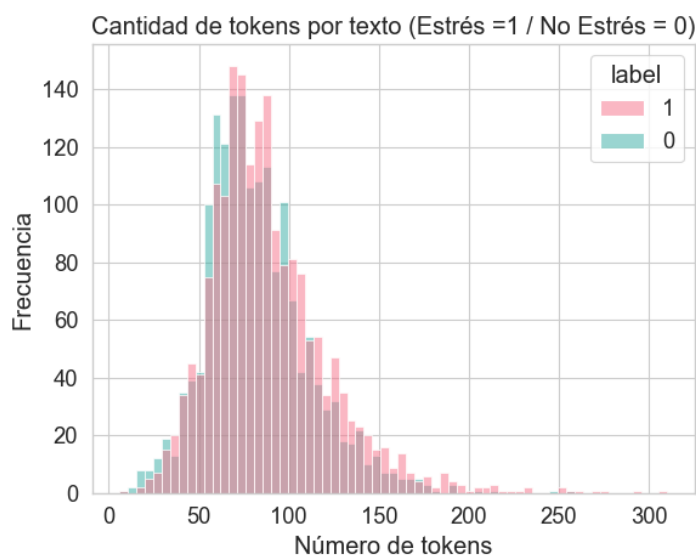


Figura 4: Distribución de la cantidad de palabras por texto estudiado

El análisis de las palabras más frecuentemente utilizadas reveló una predominancia de pronombres en singular, particularmente en textos de usuarios que reportan estrés (**Figura 5**). Este uso es coherente con la literatura que sugiere que el estrés puede llevar a una mayor auto-referencialidad lingüística (Turcan y McKeown, 2019 ; S. Guntuku et al., 2019).

Comparando el uso de palabras entre individuos estresados y no estresados (**Figura 5**), observamos que los estresados tienden a usar 'I' (yo) con mayor frecuencia, lo cual apunta a un enfoque más introspectivo o centrado en el yo, característico de discursos influenciados por el estrés.

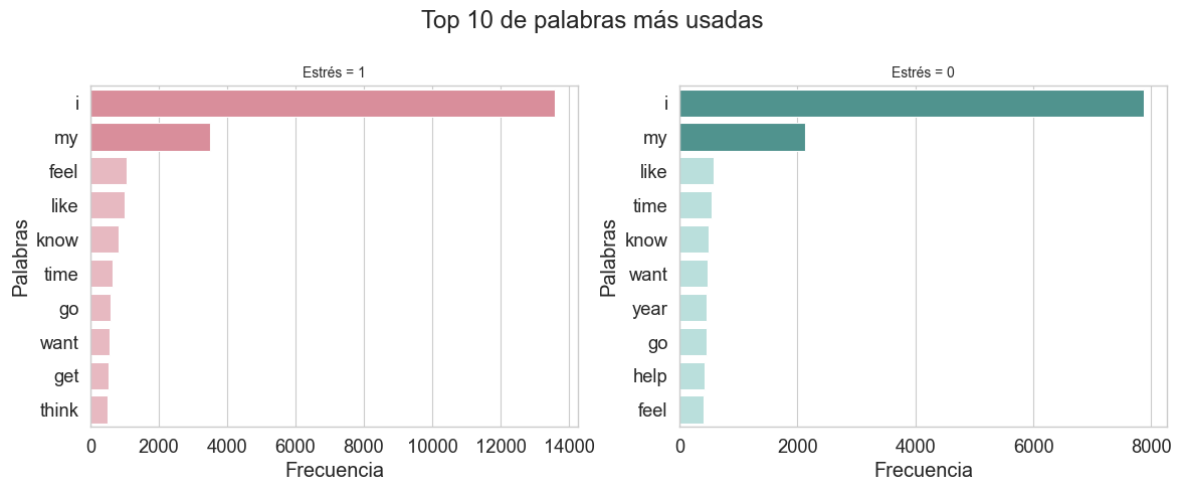


Figura 5: Frecuencia de las palabras utilizadas en individuos estresados y no estresados

Estos hallazgos subrayan la importancia de incluir análisis lingüísticos detallados en los modelos de detección de estrés, permitiendo no solo una mayor precisión en la identificación de señales de estrés, sino también realzando la relevancia clínica de las herramientas de diagnóstico automatizadas. La capacidad de los modelos para discernir variaciones sutiles en el uso del lenguaje ofrece una vía para intervenciones más personalizadas y oportunas.

3.2. Ingeniería de Características

3.2.1. Extracción Manual de Características

Las características manuales se derivan de características lingüísticas en los textos que han sido reportadas con una relación o se pueden ver afectadas por estados alterados como el estrés o los trastornos mentales. Estas características manuales son:

1. **Uso de Pronombres en Singular** (*singular_pronouns*)
2. **Longitud de Oraciones y Palabras** (*avg_sentence_len / avg_word_len*)
3. **Diversidad Léxica** (*lex_diversity*)
4. **Sintaxis y Estructura Gramatical** (*adj_adv*)
5. **Frecuencia de Construcciones Pasivas** (*passive_avg*)

Para validar la significancia de estas características en relación con la presencia de estrés en los textos, primero empleamos la prueba Shapiro-Wilk para confirmar la normalidad de estas características. Esta prueba tiene como hipótesis nula que las muestras provienen de una distribución normal. Utilizamos un nivel de significancia del 5 % ($\alpha = 0,05$). Los resultados de la prueba Shapiro-Wilk se presentan en la Tabla **Tabla 1** que muestra que el supuesto de normalidad solo se mantiene para la métrica de diversidad léxica en el grupo estresado. Para todas las demás variables evaluadas, se observa que siguen una distribución no normal dado que la hipótesis nula es rechazada.

Cuadro 1: Evaluación de Normalidad de Métricas Lingüísticas en Personas con y sin Estrés

Métrica	P-value (Estrés)	P-value (No Estrés)
Uso de pronombres en singular	$1,39 \times 10^{-6}$	$8,34 \times 10^{-17}$
Longitud promedio de palabras	$1,07 \times 10^{-10}$	$4,22 \times 10^{-21}$
Longitud promedio de oraciones	$3,94 \times 10^{-35}$	$1,36 \times 10^{-36}$
Diversidad Léxica	$2,43 \times 10^{-1}$	$1,60 \times 10^{-5}$
Sintaxis y Estructura Gramatical	$6,54 \times 10^{-35}$	$4,40 \times 10^{-33}$
Frecuencia de Construcciones Pasivas	$7,07 \times 10^{-70}$	$5,53 \times 10^{-68}$

Estos resultados indican que la mayoría de las características lingüísticas evaluadas no siguen una distribución normal, lo que tendrá implicaciones en la elección de las pruebas estadísticas que se utilizarán para analizar su relación con la presencia de estrés. Durante la evaluación de estas características, se aplicaron métodos estadísticos rigurosos para determinar diferencias significativas entre los textos escritos por individuos estresados y no estresados.

Estas relaciones de diferencia entre las muestras con y sin estrés pueden ser analizadas de manera visual en la **Figura 6**, donde las distribuciones de las características indican variaciones entre los grupos. Características como el uso de pronombres en singular, la longitud de palabras y la diversidad léxica parecen presentar diferencias entre los individuos estresados (1) y no estresados (0), lo que resalta cómo el estrés afecta aspectos específicos del lenguaje.

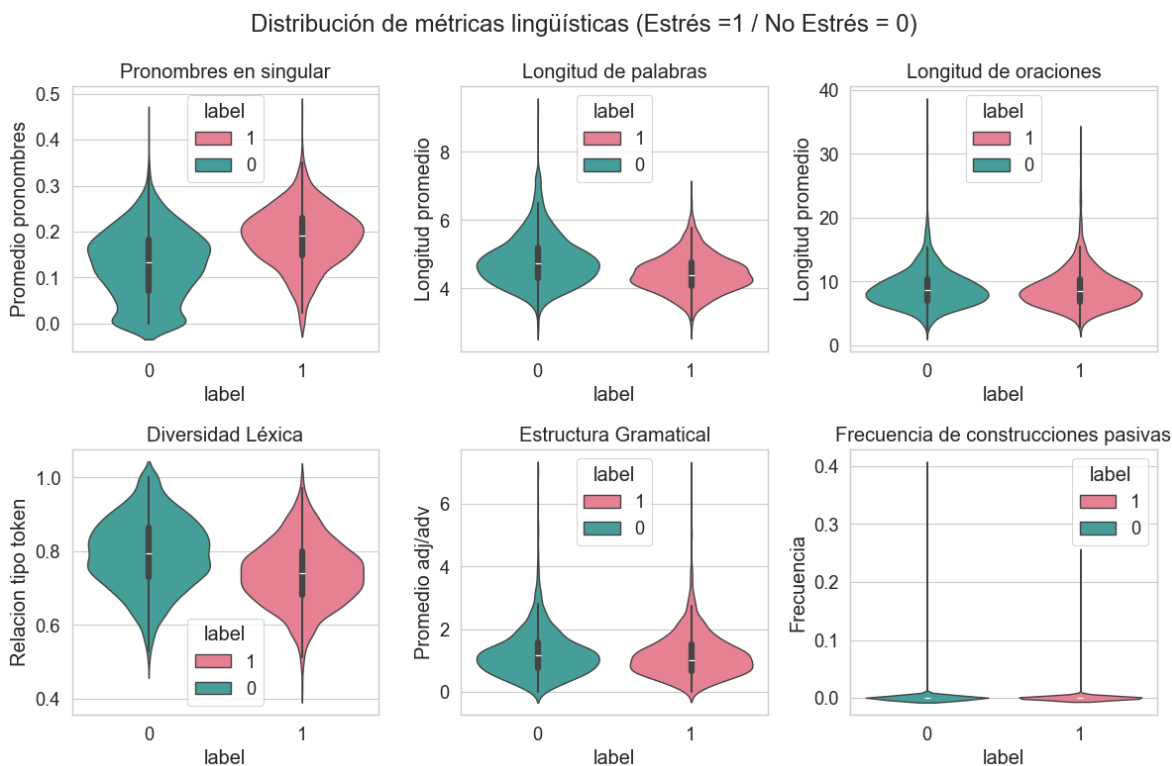


Figura 6: Distribución de métricas lingüísticas para individuos con y sin estrés

Con el fin de realizar un análisis robusto de las diferencias entre las muestras, se utiliza la prueba de Mann-Whitney con un nivel de significancia del 5% ($\alpha = 0,05$), gracias a lo

cual se confirmaron diferencias estadísticamente significativas en varias métricas lingüísticas fundamentales. Esta prueba no paramétrica se eligió porque no depende del supuesto de normalidad y es robusta, lo que la hace adecuada para analizar las características lingüísticas en este contexto.

A continuación se mencionan las hipótesis alternativas empleadas para la evaluación de cada variable, compartiendo todas la hipótesis nula de que no existe una diferencia significativa entre las muestras:

1. **Uso de Pronombres en Singular** (*singular_pronouns*): El uso de pronombres es singular es mayor en personas con estrés.
2. **Longitud de Oraciones y Palabras** (*avg_sentence_len / avg_word_len*): La longitud promedio de las palabras y oraciones utilizadas por personas estresadas es menor a sus contrapartes que no presentan indicios de estrés.
3. **Diversidad Léxica** (*lex_diversity*): La diversidad léxica de las personas estresadas es menor en comparación a las personas no estresadas.
4. **Sintaxis y Estructura Gramatical** (*adj_adv*): Las personas estresadas utilizan una sintaxis más simple que sus contrapartes.
5. **Frecuencia de Construcciones Pasivas** (*passive_avg*): la frecuencia del uso de construcciones pasivas en personas estresadas es mayor en comparación a personas no estresadas.

Al realizar las pruebas estadísticas, confirmamos que para el uso de pronombres en singular, la longitud promedio de palabras y la diversidad léxica existe una diferencia estadísticamente significativa entre las personas estresadas y personas no estresadas. En la **Tabla 2** se puede visualizar un resumen de las pruebas realizadas.

Cuadro 2: Comparación de Métricas Lingüísticas en Personas con y sin Estrés

Métrica	Hipótesis Alternativa	Pvalue
Uso de pronombres en singular	$F(x) > G(x)$	$3,56 \times 10^{-123}$
Longitud promedio de palabras	$F(x) < G(x)$	$2,26 \times 10^{-57}$
Longitud promedio de oraciones	$F(x) < G(x)$	$2,24 \times 10^{-1}$
Diversidad Léxica	$F(x) < G(x)$	$1,50 \times 10^{-61}$
Sintaxis y Estructura Gramatical	$F(x) < G(x)$	$1,15 \times 10^{-1}$
Frecuencia de Construcciones Pasivas	$F(x) > G(x)$	$3,56 \times 10^{-1}$

Donde $F(x)$ es la métrica evaluada en personas con estrés y $G(x)$ es la métrica evaluada en personas sin estrés.

A partir del análisis de correlación (**Figura 7**), se encontró una alta correlación entre algunas variables, lo que puede llevar a multicolinealidad y afectar la precisión y estabilidad de los modelos predictivos. En particular, la frecuencia del uso de pronombres singulares está altamente correlacionada a la longitud promedio de las palabras y a la diversidad léxica. Por este motivo, se decidió eliminar la variable del uso de pronombres singulares, conservando solo aquellas que proporcionan información única y valiosa (**Figura 8**). Esto mejora la generalización del modelo y reduce el riesgo de sobreajuste.

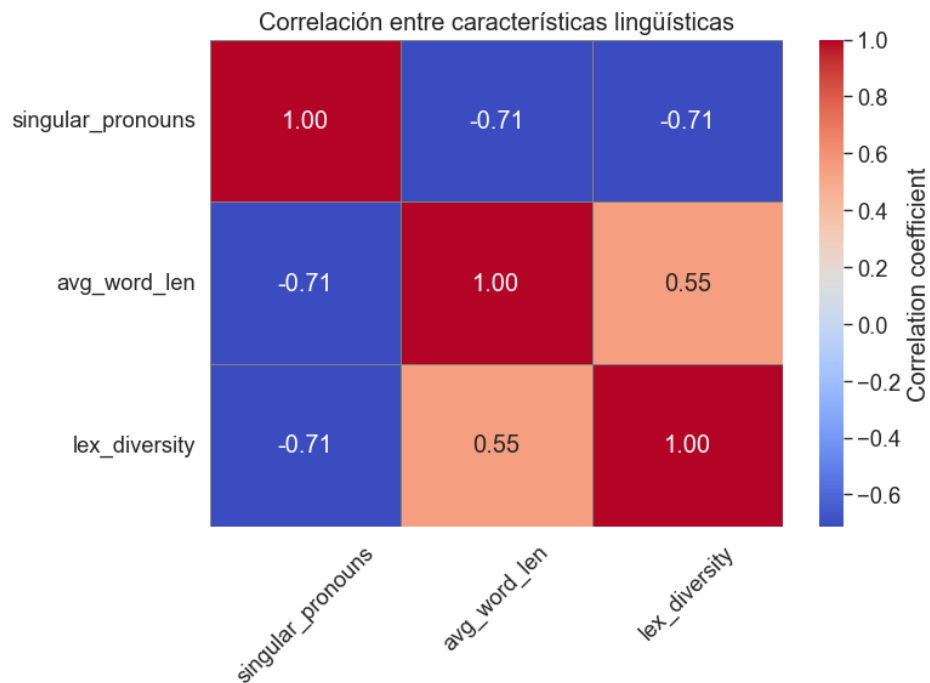


Figura 7: Heatmap de Correlación entre las características lingüísticas

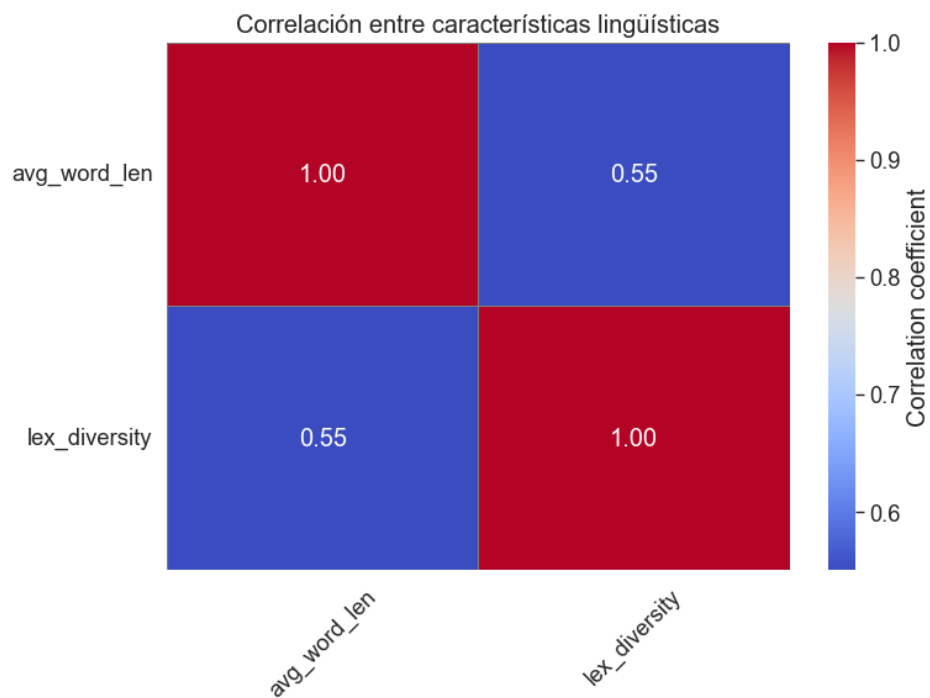


Figura 8: Heatmap de Correlación entre las características lingüísticas al eliminar variables altamente correlacionadas

La **Figura 9** muestra la importancia de las variables seleccionadas en el modelo predictivo.

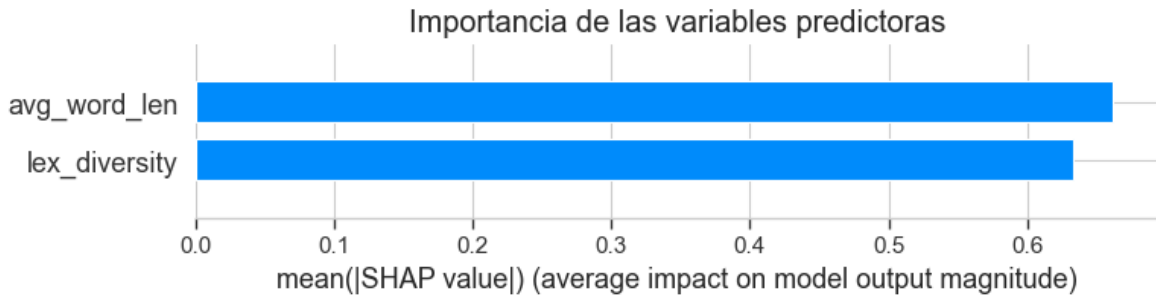


Figura 9: Importancia de las variables predictoras seleccionadas sobre el desempeño del modelo

Estos resultados ilustran la utilidad de métricas lingüísticas detalladas para identificar y clasificar niveles de estrés en textos, ofreciendo una base sólida para futuras aplicaciones en herramientas de diagnóstico e intervenciones tempranas. Además, subrayan la importancia de realizar una selección cuidadosa de características para optimizar el rendimiento y la aplicabilidad de los modelos de aprendizaje automático en contextos reales.

3.3. Entrenamiento de Modelos

3.3.1. Modelos de Machine Learning

Para los modelos de machine learning, se evaluaron diferentes configuraciones de preprocesamiento para determinar su efectividad:

1. **Características lingüísticas extraídas manualmente:** Utilización de características previamente identificadas, como el uso de pronombres en singular, la longitud de palabras y la diversidad léxica.
2. **TF-IDF:** Transformación del texto en una representación numérica mediante la técnica de TF-IDF (Frecuencia de Término - Frecuencia Inversa de Documento), que cuantifica la importancia de las palabras en el contexto del documento y del corpus completo.
3. **Embeddings:** Representación del texto utilizando embeddings de palabras, específicamente con métodos como MpNet y DistilRoBERTa, que capturan el contexto y la similitud semántica entre palabras.

Para cada una de estas configuraciones, se realizó un ajuste de hiperparámetros mediante el Algoritmo Genético de Ordenación No Dominada II (NSGA-II). Este método fue seleccionado por su capacidad para realizar una búsqueda informada y eficiente, maximizando el rendimiento de los modelos. Los hiperparámetros optimizados para cada modelo fueron:

- **Regresión Logística:** Inverso de la fuerza de regularización (C); debe ser un valor float positivo. Al igual que en las máquinas de vectores de soporte, valores más pequeños especifican una regularización más fuerte. Espacio de optimización: 1×10^{-10} - 1×10^{10} (scikit-learn, 2015)
- **K-Nearest Neighbors (KNN):** Número de vecinos a utilizar por defecto para las consultas de vecinos más cercanos (n_neighbors), espacio de optimización: [3,5,7].

Función de peso utilizada en la predicción (weights), espacio de optimización: [uniform, distance].

Tamaño de la hoja (leaf_size), espacio de optimización: 10 - 100 (scikit-learn, 2024a)

- **Support Vector Machine (SVM):** Parámetro de regularización (C). La fuerza de la regularización es inversamente proporcional a C. Debe ser estrictamente positivo, espacio de optimización: 1×10^{-10} - 1×10^{10} .

Penalty, Especifica la norma utilizada en la penalización, espacio de optimización: [l1, l2].

Peso de las clases (class_weight), espacio de optimización: [None, balanced] (scikit-learn, 2024b).

- **XGBoost:** Reducción del tamaño del paso utilizada en la actualización para prevenir el sobreajuste (lrate), espacio de optimización: 0,0001 - 0,4.

Profundidad máxima (max_depth), espacio de optimización: 10 - 1000.

Número de estimadores (n_estimators), espacio de optimización: 10 - 1000 (XGBoost, 2024)

- **Random Forest:** Número de árboles en el bosque (n_estimators), espacio de optimización: 10 - 1000.

La función para medir la calidad de una división (criterion). Los criterios soportados son 'gini' para la impureza de Gini y log_loss' y 'entropy' ambos para la ganancia de información de Shannon, espacio de optimización: [gini, entropy].

La profundidad máxima del árbol (max_depth), espacio de optimización: 10 - 1000 (scikit-learn, 2024d)

- **LightGBM:** Número de hojas (num_leaves), espacio de optimización: 10 - 1000.

Número de estimadores (n_estimators), espacio de optimización: 10 - 1000

Reducción del tamaño del paso utilizada en la actualización para prevenir el sobreajuste (lrate), espacio de optimización: 0,0001 - 0,4. (LightGBM, 2024)

- **Naive Bayes:** Porción de la mayor varianza de todas las características que se añade a las varianzas para la estabilidad del cálculo (var_smoothing). Espacio de optimización: 1×10^{-15} - 1×10^{-1} . (scikit-learn, 2024c)

En la Tabla 3 se muestra el desempeño de cada configuración de preprocesamiento usando los hiperparámetros más óptimos:

Modelo	Características Manuales	TF-IDF	MpNet	DistilRoBERTa
Regresión Logística	0.7076	0.7629	0.8025	0.8141
SVM	0.7053	0.7609	0.7981	0.8126
LightGBM	0.7021	0.7155	0.7952	0.8080
Random Forests	0.6903	0.7477	0.7922	0.8047
XGBoost	0.7019	0.7047	0.7959	0.8021
Naive Bayes	0.6971	0.7116	0.7685	0.7801
KNN	0.6155	0.6893	0.7651	0.7714

Cuadro 3: Desempeño (f1 score) de diferentes modelos y configuraciones de preprocesamiento.

Se observa que el modelo de machine learning con el mejor desempeño es la regresión logística, especialmente cuando se utiliza la representación de texto con embeddings de DistilRoBERTa, alcanzando un F1 score de 0.8141. Este rendimiento superior destaca la importancia crucial de la calidad de los datos de entrada y cómo técnicas más sofisticadas y complejas de representación de texto, como los embeddings, pueden mejorar significativamente la precisión del modelo.

Los resultados reflejados en la tabla muestran que en todos los modelos evaluados, el uso de embeddings, en particular DistilRoBERTa y MpNet, proporciona una ventaja notable sobre las características manuales y TF-IDF. Esto sugiere que los métodos de embeddings capturan mejor las relaciones semánticas y contextuales dentro del texto, lo cual es esencial para tareas como la detección de estrés, donde las señales pueden ser sutiles y contextualmente dependientes.

Además, el hecho de que la regresión logística, un modelo relativamente simple, alcance el mejor desempeño cuando se combina con una representación de texto avanzada como DistilRoBERTa, subraya que la complejidad del modelo puede ser menos crítica que la calidad de la representación de los datos. Este insight es crucial para futuras investigaciones y aplicaciones prácticas, indicando que invertir en técnicas de preprocesamiento y representación de datos puede ofrecer mayores beneficios que aumentar la complejidad del modelo.

El análisis también muestra una tendencia clara en el desempeño de los modelos: las características manuales obtuvieron el rendimiento más bajo, seguidas por TF-IDF, y finalmente, los métodos de embeddings lograron los mejores resultados. Esto no solo confirma la efectividad de los embeddings para la representación de texto, sino que también resalta la necesidad de utilizar técnicas avanzadas para aplicaciones sensibles y complejas, como la detección de estrés en textos de redes sociales.

3.3.2. Modelos de Deep Learning

Para los modelos de deep learning, se utilizaron los datos de entrada crudos para el entrenamiento y evaluación. En particular, se empleó RoBERTa, aprovechando su capacidad para manejar contextos largos y complejos, proporcionando una comprensión más rica y matizada del texto. RoBERTa fue ajustado mediante un proceso de fine-tuning utilizando nuestros datos específicos, lo que permitió al modelo captar las sutilezas lingüísticas y patrones asociados con el estrés en las publicaciones textuales. Durante el proceso de fine-tuning, se entrenó RoBERTa modificando los pesos iniciales del modelo a lo largo de varias épocas de entrenamiento. Este ajuste optimizó el rendimiento del modelo en la tarea de clasificación binaria

de estrés. Además, se adaptó la capa final de clasificación del modelo para alinearla con el problema específico de detección de estrés. Esto implicó ajustar la arquitectura de salida del modelo para que pudiera distinguir eficazmente entre textos que indican estrés y aquellos que no lo hacen.

El proceso de entrenamiento incluyó técnicas de validación cruzada para asegurar la robustez y generalización del modelo. En la Figura 10 se ilustra el proceso de entrenamiento del modelo, destacando la evolución de la función de pérdida a lo largo de las épocas de entrenamiento.

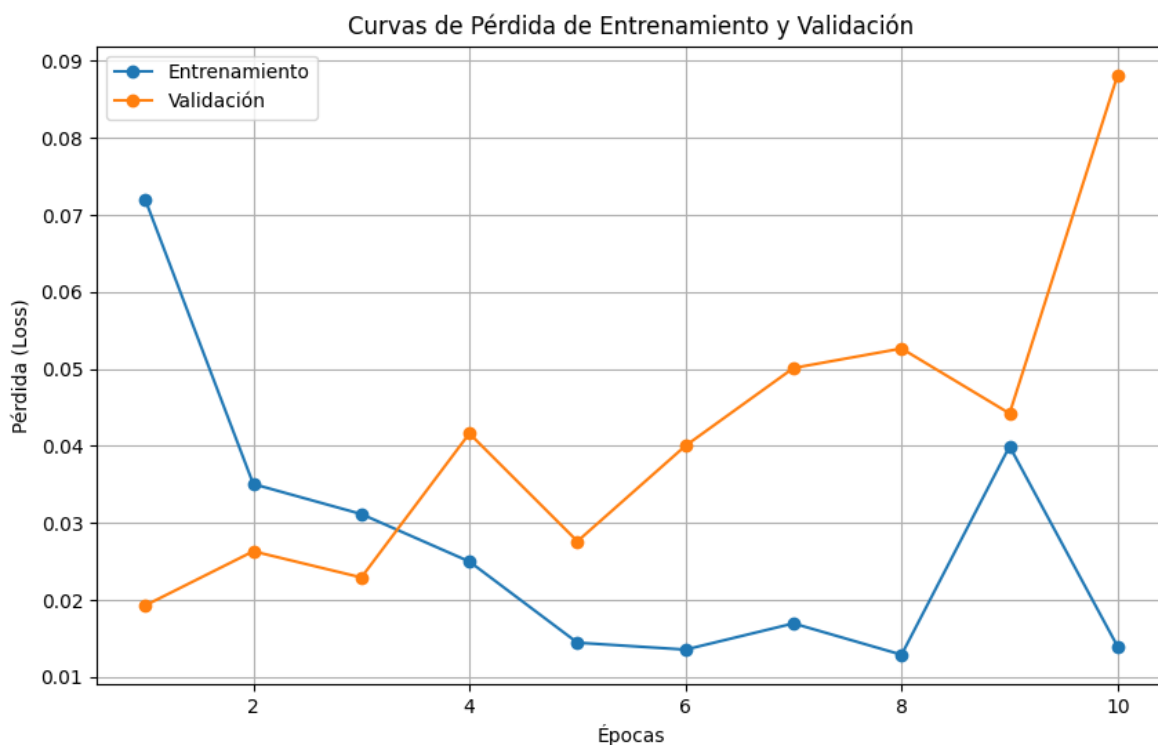


Figura 10: Curva de pérdida del modelo RoBERTa

La gráfica muestra las curvas de pérdida tanto para el conjunto de entrenamiento como para el de validación a lo largo de 10 epochs. Es evidente que la pérdida de entrenamiento disminuye de manera constante, lo que indica que el modelo está aprendiendo y ajustando sus pesos para minimizar el error. En nuestro caso, se decidió entrenar el modelo por 5 epochs, antes de que la pérdida de validación comenzara a aumentar notablemente, lo cual ayuda a prevenir el sobreajuste y asegura que el modelo generalice mejor en datos no vistos.

3.4. Evaluación de Modelos

Después del entrenamiento, se realizó una evaluación exhaustiva de los modelos. Todos los modelos fueron entrenados nuevamente utilizando la totalidad de los datos de entrenamiento y empleando la configuración de hiperparámetros más óptima encontrada durante la fase de entrenamiento. A continuación, se presenta el desempeño de todos los modelos en el conjunto de prueba.

Modelo	Características Manuales	TF-IDF	MpNet	DistilRoBERTa
Regresión Logística	0.7041	0.7646	0.7713	0.8162
SVM	0.7044	0.7593	0.7603	0.8226
LightGBM	0.6857	0.7120	0.7639	0.8204
Random Forests	0.6863	0.7161	0.7439	0.7989
XGBoost	0.6858	0.6928	0.7730	0.8011
Naive Bayes	0.7005	0.6755	0.7282	0.7639
KNN	0.6296	0.6277	0.7114	0.7568

Cuadro 4: Desempeño (f1 score) de diferentes modelos y configuraciones de preprocesamiento en la evaluación.

El análisis de la tabla de resultados de la evaluación revela una tendencia similar a la observada durante el entrenamiento: las representaciones de datos más complejas tienden a demostrar un mejor desempeño. Sin embargo, a diferencia de la fase de entrenamiento, donde la regresión logística mostró el mejor rendimiento en varias configuraciones, en la evaluación final, el modelo SVM utilizando embeddings de DistilRoBERTa logró el mejor desempeño general con un F1 score de 0.8226. Este resultado subraya la importancia de utilizar técnicas avanzadas de representación de texto, como los embeddings de DistilRoBERTa, que capturan mejor las relaciones semánticas y contextuales en los datos. Además, confirma que los modelos más sofisticados pueden beneficiarse significativamente de representaciones de datos de alta calidad.

La comparación entre los modelos de machine learning y los de deep learning también es crucial. Aunque los modelos de machine learning, como SVM y regresión logística, mostraron un rendimiento sólido, los modelos de deep learning, específicamente RoBERTa, demostraron ser altamente efectivos para la tarea de detección de estrés, aprovechando su capacidad para manejar contextos largos y complejos y adaptarse a los datos específicos mediante fine-tuning.

Modelo	F1 Score
RoBERTa	0.8235
SVM-distilroberta-embeddings	0.8226
Llama3-ejemplos	0.7490
Llama3	0.7427

Cuadro 5: Desempeño (f1 score) de los modelos de machine learning y deep learning en la evaluación.

El análisis de la tabla muestra que el mejor desempeño lo obtuvo el modelo RoBERTa con un F1 Score de 0.8235. Este resultado destaca la efectividad de RoBERTa en la tarea de detección de estrés en textos de redes sociales, gracias a su capacidad para manejar contextos largos y complejos y a su proceso de fine-tuning con datos específicos.

El modelo SVM con embeddings de DistilRoBERTa estuvo muy cerca del desempeño de RoBERTa, con un F1 Score de 0.8226. Esto refuerza la importancia de una buena representación de los datos. La similitud en el desempeño de ambos modelos no es sorprendente, dado que ambos utilizan RoBERTa como base para la representación de texto. Este resultado sugiere que, aunque RoBERTa puede ofrecer un rendimiento ligeramente superior debido a

su arquitectura y preentrenamiento más robusto, las técnicas de embeddings avanzadas como DistilRoBERTa pueden acercar significativamente el desempeño de modelos más simples como SVM.

Por otro lado, el modelo Llama3 obtuvo un F1 Score de 0.7427, mientras que Llama3 con ejemplos de prompt engineering alcanzó un F1 Score de 0.7490. Aunque estos resultados son menores en comparación con RoBERTa y SVM con DistilRoBERTa, son igualmente interesantes. Llama3 logra estos puntajes utilizando únicamente prompt engineering, una técnica mucho menos costosa en términos de recursos y tiempo. Esto indica que, aunque no alcanzan el máximo desempeño, los modelos basados en Llama3 pueden ser una opción viable y eficiente para aplicaciones donde los recursos computacionales y el tiempo son limitados.

Análisis de las Matrices de Confusión

Al analizar las matrices de confusión de los modelos RoBERTa, SVM con DistilRoBERTa embeddings, Llama3 con ejemplos y Llama3, se observa un patrón claro en la clasificación de textos de estrés y no estrés. El desafío principal para todos los modelos es la correcta clasificación de los textos sin estrés (0).

- **RoBERTa:** Este modelo muestra un buen desempeño general, pero aún presenta una cantidad significativa de falsos positivos (40) y falsos negativos (26). Sin embargo, RoBERTa logra clasificar correctamente un alto número de verdaderos positivos (154) y verdaderos negativos (135), lo que indica su capacidad para manejar contextos complejos y matices lingüísticos.
- **SVM con DistilRoBERTa embeddings:** Similar a RoBERTa, este modelo también muestra un buen desempeño con 39 falsos positivos y 27 falsos negativos. La cantidad de verdaderos positivos (153) y verdaderos negativos (136) es comparable a la de RoBERTa, lo que refuerza la importancia de utilizar embeddings de alta calidad para mejorar la clasificación.
- **Llama3 con ejemplos:** Este modelo enfrenta más dificultades, con 119 falsos positivos y solo 56 verdaderos negativos. Aunque logra un alto número de verdaderos positivos (179), la gran cantidad de falsos positivos sugiere que el modelo tiene problemas para distinguir entre textos de estrés y no estrés cuando se basa solo en prompt engineering.
- **Llama3:** Sin el uso de ejemplos específicos, Llama3 también muestra una alta tasa de falsos positivos (123) y una baja cantidad de verdaderos negativos (52). Similar a Llama3 con ejemplos, este modelo tiene 179 verdaderos positivos, lo que sugiere que aunque puede identificar textos de estrés, lucha significativamente con la clasificación de textos sin estrés.

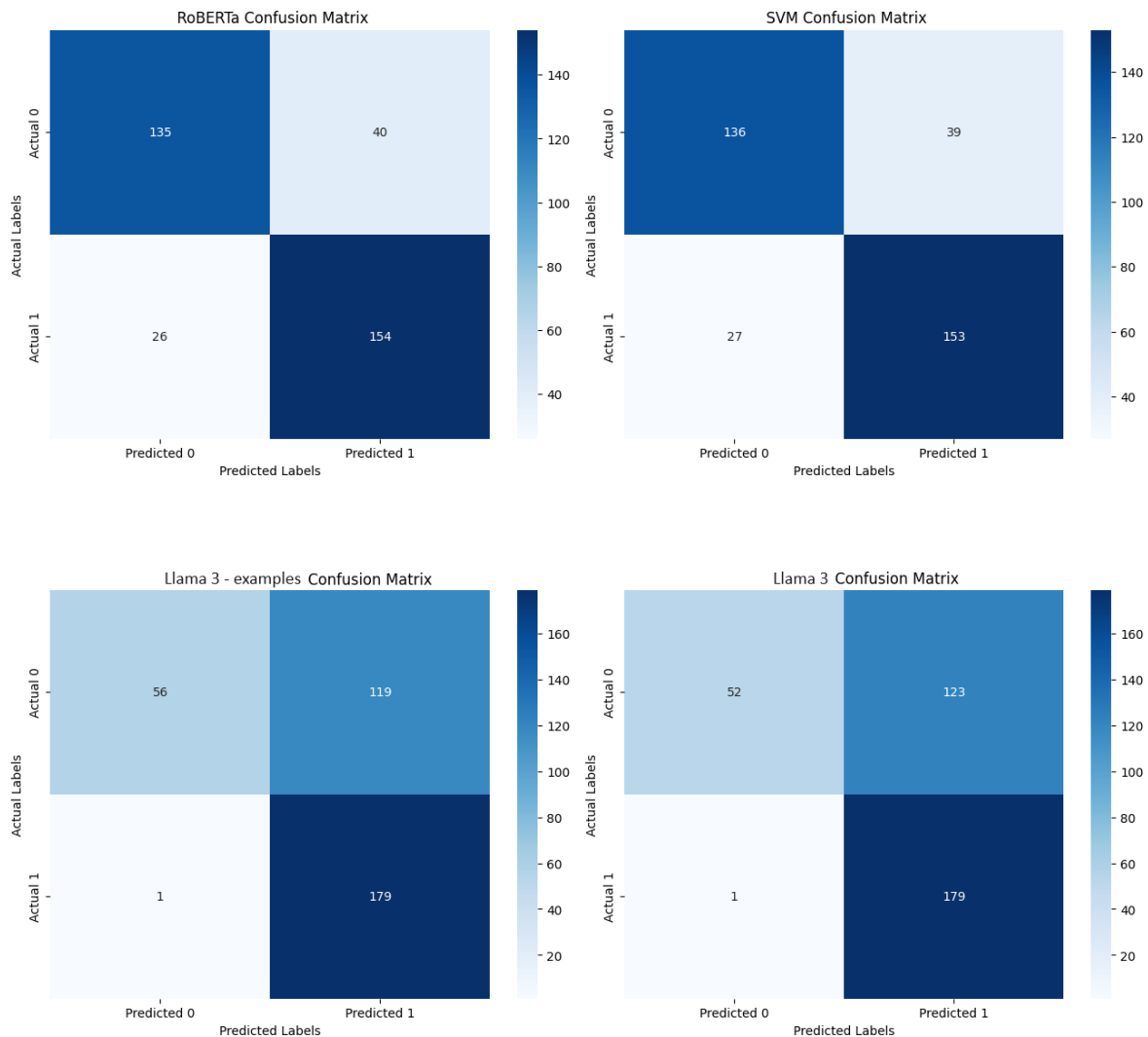


Figura 11: Matrices de confusión de los diferentes modelos

En resumen, los modelos basados en Llama3 tienden a generar muchos falsos positivos y fallan en clasificar correctamente los verdaderos negativos. Esto indica que, aunque pueden identificar textos con señales de estrés, tienen una tendencia a sobredetectar estrés en textos que no lo contienen. En contraste, RoBERTa y SVM con DistilRoBERTa embeddings muestran un equilibrio mejor en la clasificación de textos de estrés y no estrés, aunque todavía presentan algunas dificultades.

Este análisis subraya la importancia de continuar refinando los modelos y técnicas de preprocesamiento para mejorar la precisión y reducir los falsos positivos y negativos, especialmente en aplicaciones sensibles como la detección de estrés en textos de redes sociales.

3.5. Solución de Aplicación

Posterior al entrenamiento de modelos, se seleccionó el modelo de mejor desempeño para implementar una aplicación utilizando Streamlit. Se desarrollaron dos aplicaciones principales:

En primer lugar, se creó una aplicación que clasifica textos de entrada como 'estrés' o 'no

estrés'. En esta aplicación, el modelo RoBERTa se hospeda en HuggingFace, donde se genera un endpoint utilizado para la inferencia. La aplicación fue desarrollada en Python, específicamente utilizando la librería Streamlit, y está publicada en Streamlit Hub. Esta herramienta permite a los usuarios ingresar cualquier texto y recibir una evaluación instantánea sobre si el contenido expresa estrés.

Welcome to the Stress Detector with RoBERTa

This application uses an advanced natural language processing model to analyze texts and determine stress levels. Below, you can explore how the model works and test it with your own texts.

Functionalities

- **Text Analysis:** Enter any text and the model will determine if the content expresses stress.
- **Instant Results:** Get results in seconds.
- **Simple Interface:** An easy-to-use interface that does not require technical knowledge.

Enter the text for analysis:

I feel overwhelmed with all the tasks piling up at work. Every day feels like a race against time, and I'm constantly anxious about meeting deadlines. It's hard to sleep, and I can't seem to relax even when I try to take a break.

Analyze

Prediction Result:

Stress

Figura 12: Aplicación de detección de estrés utilizando RoBERTa

En segundo lugar, se desarrolló una aplicación adicional que integra un chatbot potenciado por Llama3. Este chatbot interactúa normalmente con los usuarios, pero también evalúa las interacciones en segundo plano utilizando el modelo RoBERTa. Cada vez que el modelo detecta un texto con altas probabilidades de estar relacionado con el estrés, envía un mensaje al usuario informándole sobre la detección de estrés y sugiriendo posibles acciones a tomar. El chatbot se construyó utilizando Python, específicamente con la librería Streamlit, y las capacidades de chatbot de Llama3 se implementaron usando la API de Groq, que facilita el acceso e integración de diversos LLMs en aplicaciones. La evaluación y clasificación del estrés se realiza de manera similar con RoBERTa, utilizando su endpoint integrado en HuggingFace.



Llama3 Chatbot

Hello!

Hello! It's nice to meet you. Is there something I can help you with, or would you like to chat?

Honestly, I feel a little overwhelmed as my midterms are close. I'm a scholarship student so I'm under a lot of pressure

I totally understand! It can be really tough to deal with academic pressure, especially when there's a lot at stake like your scholarship. It's normal to feel overwhelmed, and it's great that you're acknowledging your feelings.

Take a deep breath and let's break it down together. What are your biggest concerns right now? Are there specific subjects or exams that are worrying you, or is it more about the overall workload? Are there any particular challenges you're facing, like time management or understanding certain concepts?

Remember, I'm here to listen and offer support. We can work together to come up with a plan to help you feel more confident and prepared for your midterms.

Hello! I hope you're doing okay. I've noticed some signs of stress in our conversation — it's completely okay to feel this way. If you're feeling overwhelmed, please remember to take care of yourself. Help is available if you need it, and you're not alone. You can reach out to someone at these numbers: XXXXX. Also, if you're curious about how I can detect signs of stress, feel free to explore our methods here: <https://github.com/fjimoyao/depression-detection>. Your well-being matters to us!

Figura 13: Aplicación de chatbot con funcionalidad de detección de estrés utilizando Llama3

Es importante mencionar que ambas aplicaciones están en inglés, ya que los modelos fueron entrenados en este idioma y generan mejores resultados de esta manera. Estas aplicaciones no solo funcionan como pruebas de concepto, sino que también demuestran la importancia de la integración de la investigación con productos o funcionalidades que pueden tener un impacto real en la sociedad.

Las aplicaciones desarrolladas ofrecen una perspectiva práctica sobre cómo los avances en modelos de procesamiento de lenguaje natural pueden ser utilizados para mejorar la salud mental. La capacidad de detectar estrés en tiempo real permite intervenciones tempranas y personalizadas, promoviendo un enfoque más proactivo en la gestión del bienestar emocional. La combinación de modelos avanzados como RoBERTa y Llama3 proporciona una base sólida para el desarrollo de herramientas efectivas y accesibles, destacando el potencial de la inteligencia artificial en la mejora de la calidad de vida de las personas.

Se puede interactuar con estas plataformas en: Inferencia RoBERTa, Chatbot RoBERTa.

4. Discusión

Este estudio resalta el potencial significativo de las técnicas avanzadas de procesamiento de lenguaje natural (NLP) para detectar indicadores de estrés en textos de redes sociales utilizando modelos de atención como RoBERTa. Los resultados demuestran que RoBERTa supera notablemente a los modelos tradicionales de machine learning, evidenciando su capacidad para identificar patrones lingüísticos sutiles indicativos de estrés. Esta habilidad para discernir variaciones finas en el uso del lenguaje constituye una herramienta esencial para la detección temprana y precisa de trastornos mentales, facilitando intervenciones proactivas que pueden mejorar significativamente los resultados terapéuticos.

La evaluación también reveló que el modelo SVM con embeddings de DistilRoBERTa obtuvo un desempeño muy cercano al de RoBERTa, lo que subraya la importancia de una buena representación de los datos. La similitud en el desempeño de ambos modelos, basados en la misma técnica de embeddings, destaca cómo las técnicas de representación de texto más sofisticadas pueden igualar, e incluso superar, el rendimiento de modelos de deep learning más complejos.

Además, los resultados obtenidos con Llama3, aunque inferiores a los de RoBERTa y SVM con DistilRoBERTa, son igualmente interesantes. Llama3 alcanzó un F1 Score de 0.7427, y 0.7490 cuando se utilizó prompt engineering, lo que resalta su potencial como una alternativa menos costosa en términos de recursos y tiempo.

Sin embargo, es fundamental reconocer las limitaciones inherentes a esta aproximación. La dependencia de datos provenientes de redes sociales plantea desafíos relacionados con la representatividad y la privacidad. Los datos de redes sociales pueden no ser completamente representativos de la población general debido a sesgos demográficos y de comportamiento en el uso de estas plataformas. Además, la interpretación de los datos está sujeta a la precisión de los modelos y a la calidad del texto disponible, que puede variar ampliamente.

A pesar de estos desafíos, las aplicaciones prácticas derivadas de este estudio son prometedoras. La implementación de modelos de NLP en sistemas de salud mental ofrece nuevas posibilidades para el monitoreo continuo y la intervención personalizada. Herramientas basadas en NLP integradas en aplicaciones móviles de salud mental podrían proporcionar a los usuarios retroalimentación en tiempo real sobre su estado emocional, junto con recomendaciones personalizadas para gestionar su estrés de manera efectiva.

Mirando hacia el futuro, se recomienda explorar la integración de datos de múltiples modalidades, incluyendo entradas audiovisuales y biométricas, para mejorar la detección y análisis del estrés. También es esencial continuar perfeccionando los algoritmos de NLP para que manejen con mayor precisión la diversidad y complejidad del lenguaje humano, asegurando que las herramientas desarrolladas sean inclusivas y adaptables a distintos contextos lingüísticos y culturales.

Finalmente, es fundamental adoptar un enfoque ético y transparente en la recopilación y el uso de datos personales, asegurando que las intervenciones basadas en NLP respeten la privacidad de los usuarios y fomenten su bienestar. La consideración ética no solo es crucial para mantener la confianza del público, sino también para la implementación efectiva de tecnologías avanzadas en el campo de la salud mental.

En resumen, este estudio subraya la potencialidad de los modelos de atención, como RoBERTa, para aplicaciones prácticas en la detección temprana de estrés. El enfoque basado en modelos de deep learning puede ser especialmente útil en entornos donde los recursos para el análisis detallado son limitados, ofreciendo una herramienta poderosa y eficiente para intervenciones tempranas y precisas en el ámbito de la salud mental. La capacidad de generalización de estos modelos los hace adecuados para su uso en diversos contextos y poblaciones, aumentando su valor práctico y su aplicabilidad a nivel global. Estos resultados destacan la importancia de continuar explorando y desarrollando modelos de deep learning en el campo de la detección de estrés y otros trastornos relacionados, potenciando el uso de la inteligencia artificial en la mejora de la salud mental.

5. Conclusiones

Este estudio ha demostrado el potencial significativo de los modelos avanzados de procesamiento de lenguaje natural (NLP), especialmente RoBERTa, en la detección y clasificación de estados de estrés a partir de textos en redes sociales. Los hallazgos subrayan la eficacia de los modelos de atención en interpretar contextos complejos y capturar matices lingüísticos críticos para el análisis de sentimientos y emociones.

La implementación de una aplicación basada en RoBERTa facilita la detección temprana y precisa del estrés, permitiendo intervenciones en tiempo real y apoyo en la toma de decisiones en contextos clínicos y de salud mental. Este avance representa un paso importante hacia la integración de la inteligencia artificial en el campo de la salud mental, con el potencial de mejorar significativamente los resultados del tratamiento y la calidad de vida de los pacientes.

Además, este proyecto ha realizado contribuciones sustanciales al campo de la salud mental y la tecnología NLP. Hemos validado hipótesis lingüísticas que ilustran cómo el estrés altera el lenguaje, desarrollado técnicas refinadas de análisis de sentimientos para contextos de estrés, y creado una herramienta en Python que facilita la clasificación de texto según niveles de estrés. Estos avances no solo enriquecen nuestra comprensión académica y pública del estrés, sino que también promueven la colaboración interdisciplinaria y fomentan la innovación en herramientas de diagnóstico, haciendo estos métodos más accesibles y aplicables a nivel global.

En conclusión, este estudio no solo resalta el potencial de los modelos de atención como RoBERTa para aplicaciones prácticas en la detección temprana de estrés, sino que también abre nuevas vías para la investigación y el desarrollo de herramientas innovadoras en el campo de la salud mental. La colaboración interdisciplinaria y el enfoque ético en el uso de estas

tecnologías serán fundamentales para maximizar su impacto positivo a nivel global.

Referencias

- Ali, Y. A., Awwad, E. M., Al-Razgan, M. S., & Maarouf, A. (2023). Hyperparameter Search for Machine Learning Algorithms for Optimizing the Computational Complexity. *Processes*. <https://doi.org/10.3390/pr11020349>
- Anderson, K. (2015). Ask me anything: what is Reddit? *Library Hi Tech News*, 32, 8-11. <https://doi.org/10.1108/LHTN-03-2015-0018>
- Breton, J. M., Barraza, M., Hu, K. Y., Frias, S. J., Long, K. L. P., & Kaufer, D. (2021). Juvenile exposure to acute traumatic stress leads to long-lasting alterations in grey matter myelination in adult female but not male rats. *Neurobiology of Stress*, 14, 100319. <https://doi.org/10.1016/j.ynstr.2021.100319>
- Buchanan, T., Laures-Gore, J., & Duff, M. (2014). Acute stress reduces speech fluency. *Biological Psychology*, 97, 60-66. <https://doi.org/10.1016/j.biopsycho.2014.02.005>
- Cox, T. (1985). The nature and measurement of stress. *Ergonomics*, 28(8), 1155-63. <https://doi.org/10.1080/00140138508963238>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
- Ernala, S. K., Rizvi, A. F., Birnbaum, M. L., Kane, J. M., & Choudhury, M. D. (2017). Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia. *Proc ACM Hum-Comput Interact*, 1(1), 43.
- Fine, A., Crutchley, P., Blase, J., Carroll, J., & Coppersmith, G. (2020). Assessing population-level symptoms of anxiety, depression, and suicide risk in real time using NLP applied to social media data. *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, 50-54.
- Groq. (2024). Groq: Extreme Performance for AI and Machine Learning. <https://wow.groq.com/why-groq/>
- Guntuku, S., Buffone, A., Jaidka, K., Eichstaedt, J. C., & Ungar, L. H. (2019). Understanding and Measuring Psychological Stress Using Social Media. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01), 214-225. <https://doi.org/10.1609/icwsm.v13i01.3223>
- Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18, 43-49.
- HuggingFace. (2024a). Hugging Face. <https://huggingface.co/>
- HuggingFace. (2024b). Tokenizers: Summary [Accessed: 2024-06-19].
- Hyman, S., Chisholm, D., Kessler, R., Patel, V., & Whiteford, H. (2006). Mental Disorders. En WHO (Ed.), *Disease control priorities related to mental, neurological, developmental and substance abuse disorders* (pp. 1-2). <https://apps.who.int/iris/handle/10665/43565>
- Institute for Health Metrics and Evaluation. (2024). *Health Risks and Issues: Mental Health*. Institute for Health Metrics y Evaluation. Consultado el 29 de abril de 2024, desde <http://www.healthdata.org/research-analysis/health-risks-issues/mental-health>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer. <https://doi.org/10.1007/978-1-4614-7138-7>

- LightGBM. (2024). *LightGBM Quick Start Guide*. <https://lightgbm.readthedocs.io/en/latest/Quick-Start.html>
- Lin, H., Jia, J., Guo, Q., Xue, Y., Li, Q., Huang, J., Cai, L., & Feng, L. (2014). User-level psychological stress detection from social media using deep neural network. *Proceedings of the ACM Multimedia Conference (ACM MM)*.
- Lin, H., Jia, J., Qiu, J., Zhang, Y., Shen, G., Xie, L., Tang, J., Feng, L., & Chua, T.-S. (2017). Detecting stress based on social interactions in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 29(09), 1820-1833.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*.
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18(1), 50-60. <https://doi.org/10.1214/aoms/1177730491>
- Marin, M.-F., Lord, C., Andrews, J., Juster, R.-P., Sindi, S., Arseneault-Lapierre, G., Fiocco, A. J., & Lupien, S. J. (2011). Chronic stress, cognitive functioning and mental health. *Neurobiology of Learning and Memory*, 96(4), 583-595. <https://doi.org/10.1016/j.nlm.2011.02.016>
- Meta. (2023). Meta AI Introduces LLaMA 3: A Leap Forward in Language Models [Accessed: 2024-06-27].
- Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association : JAMIA*, 18(5), 544. <https://doi.org/10.1136/AMIAJNL-2011-000464>
- OpenAI. (2024). Prompt Engineering [Accessed: 2024-06-30]. <https://platform.openai.com/docs/guides/prompt-engineering>
- Pechtel, P., & Pizzagalli, D. (2011). Effects of early life stress on cognitive and affective function: an integrated review of human literature. *Psychopharmacology*, 214, 55-70. <https://doi.org/10.1007/s00213-010-2009-2>
- Ramos, J. (2003). Using TF-IDF to Determine Word Relevance in Document Queries. *Proceedings of the First Instructional Conference on Machine Learning*, 242, 133-142.
- Rude, S. S., Gortner, E. M., & Pennebaker, J. W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 18(8), 1121-1133. <https://doi.org/10.1080/02699930441000030>
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
- scikit-learn. (2015). *sklearn LogisticRegression*. https://scikit-learn.org/0.16/modules/generated/sklearn.linear_model.LogisticRegression.html
- scikit-learn. (2024a). *sklearn KNeighborsClassifier*. <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>
- scikit-learn. (2024b). *sklearn LinearSVC*. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC>
- scikit-learn. (2024c). *sklearn naive bayes GaussianNB*. https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html#sklearn.naive_bayes.GaussianNB
- scikit-learn. (2024d). *sklearn RandomForestClassifier*. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591-611. <https://doi.org/10.2307/2333709>
- Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. En A. Sattar & B.-h. Kang (Eds.), *AI 2006: Advances in Artificial Intelligence* (pp. 1015-1021). Springer Berlin Heidelberg.
- Song, K., Tan, X., Qin, T., Lu, J., & Liu, T.-Y. (2020). Mpnet: Masked and permuted pre-training for language understanding. *arXiv preprint arXiv:2004.09297*.
- Streamlit. (2024). Streamlit: A faster way to build and share data apps. <https://streamlit.io/>
- Turcan, E., & McKeown, K. (2019). Dreaddit: A Reddit Dataset for Stress Analysis in Social Media. *ArXiv, abs/1911.00133*. <https://doi.org/10.18653/v1/D19-6213>
- Winata, G. I., Kampman, O. P., & Fung, P. (2018). Attention-based LSTM for psychological stress detection from spoken language using distant supervision. *CoRR, abs/1805.12307*. <http://arxiv.org/abs/1805.12307>
- XGBoost. (2024). *XGBoost Parameters*. <https://xgboost.readthedocs.io/en/stable/parameter.html>