



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE
SISTEMAS

Análisis de los sentimientos y emociones en las redes sociales
de equipos de una liga de fútbol utilizando inteligencia artificial

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTORES:

Adrianzen Zapata, Marco Antonio (orcid.org/0000-0001-7798-0615)

Fernandez Olaya, Juan (orcid.org/0000-0001-8993-0279)

ASESOR:

Mg. More Valencia, Rubén Alexander (orcid.org/0000-0002-7496-3702)

LÍNEA DE INVESTIGACIÓN:

Sistemas de Información y Comunicaciones

LÍNEA DE ACCIÓN DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo económico, empleo y emprendimiento

PIURA - PERÚ

2023

DEDICATORIA

A mis padres y mi abuela por su amor incondicional y por creer en mí desde el primer día. Por sus sacrificios y su apoyo constante que han sido la clave de mi éxito.

A mi tutor académico, por su dedicación y pasión por la enseñanza y por guiarme en mi camino.

Marco Antonio A.

A mi docente, que de una u otra manera, brindó su apoyo, para la realización de este trabajo.

A mis Padres, esposa e hijos, quienes han sido mi mayor motivación a lo largo de mi carrera profesional.

Juan Fernández O.

AGRADECIMIENTO

En primer lugar, a mi Dios por darme la vida y fortaleza para poder seguir en camino. A todos aquellos que han sido una parte integral de mi camino académico y personal.

Marco Antonio A.

El principal agradecimiento a Dios y a la Virgen, por haberme acompañado, guiado y por su bendición a lo largo de mi carrera.

A Marco Antonio, por haber sido un excelente compañero de Tesis y buen amigo.

A todas las personas, que me brindaron su confianza, apoyo, motivación, para poder crecer profesionalmente.

Juan Fernández O.



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Declaratoria de Autenticidad del Asesor

Yo, MORE VALENCIA RUBEN ALEXANDER, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - PIURA, asesor de Tesis titulada: "Análisis de los sentimientos y emociones en las redes sociales de equipos de una Liga de futbol utilizando inteligencia artificial", cuyos autores son ADRIANZEN ZAPATA MARCO ANTONIO, FERNANDEZ OLAYA JUAN, constato que la investigación tiene un índice de similitud de 13.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

PIURA, 08 de Diciembre del 2023

Apellidos y Nombres del Asesor:	Firma
MORE VALENCIA RUBEN ALEXANDER DNI: 02897931 ORCID: 0000-0002-7496-3702	Firmado electrónicamente por: RMOREV el 08-12- 2023 23:55:02

Código documento Trilce: TRI - 0689239



Declaratoria de Originalidad de los Autores

Nosotros, ADRIANZEN ZAPATA MARCO ANTONIO, FERNANDEZ OLAYA JUAN estudiantes de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - PIURA, declaramos bajo juramento que todos los datos e información que acompañan la Tesis titulada: "Análisis de los sentimientos y emociones en las redes sociales de equipos de una Liga de futbol utilizando inteligencia artificial", es de nuestra autoría, por lo tanto, declaramos que la Tesis:

1. No ha sido plagiada ni total, ni parcialmente.
2. Hemos mencionado todas las fuentes empleadas, identificando correctamente toda cita textual o de paráfrasis proveniente de otras fuentes.
3. No ha sido publicada, ni presentada anteriormente para la obtención de otro grado académico o título profesional.
4. Los datos presentados en los resultados no han sido falseados, ni duplicados, ni copiados.

En tal sentido asumimos la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de la información aportada, por lo cual nos sometemos a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

Nombres y Apellidos	Firma
JUAN FERNANDEZ OLAYA DNI: 43709242 ORCID: 0000-0001-8993-0279	Firmado electrónicamente por: FFERNANDEZOLA el 08-12-2023 11:57:10
MARCO ANTONIO ADRIANZEN ZAPATA DNI: 78007389 ORCID: 0000-0001-7798-0615	Firmado electrónicamente por: MADRIANZENZ el 08-12-2023 18:15:07

Código documento Trilce: TRI - 0689241

ÍNDICE DE CONTENIDOS

CARÁTULA.....	I
DEDICATORIA	II
AGRADECIMIENTO	III
DECLARATORIA DE AUTENTICIDAD DEL ASESOR.....	IV
DECLARATORIA DE ORIGINALIDAD DE LOS AUTORES.....	V
ÍNDICE DE CONTENIDOS	VI
ÍNDICE DE TABLAS	VIII
ÍNDICE DE FIGURAS	IX
RESUMEN	X
ABSTRACT.....	XI
I. INTRODUCCIÓN.....	1
II. MARCO TEÓRICO	4
III. METODOLOGÍA.....	11
3.1 Tipo y diseño de investigación	11
3.2 Variables y operacionalización.....	11
3.3 Población, muestra y muestreo.....	15
3.4 Técnicas e instrumentos de recolección de datos	16
3.5 Procedimiento	16
3.6 Método de análisis de datos.....	16
3.7 Aspectos éticos.....	17
IV. RESULTADOS	18
4.1 Web Scraping	18
4.2 Extracción de la información de Facebook.....	20
4.3 Filtrado y limpieza de los datos	22
4.4 Entrenamiento del modelo de aprendizaje automático	24
4.5 Análisis de los sentimientos	32
4.6 Matriz de confusión y métricas asociadas	32
4.7 Limitaciones del modelo	33
V. DISCUSIÓN	35
VI. CONCLUSIONES.....	39
VII. RECOMENDACIONES	41
REFERENCIAS	43

ANEXOS	46
Anexo I – Código de Web Scraping	46
Anexo II – Código para la clasificación manual de los sentimientos	55
Anexo III – Código para el entrenamiento automático del modelo.....	65
Anexo IV – Matriz de Consistencia	70
Anexo V – Certificado	72

ÍNDICE DE TABLAS

Tabla 1 <i>Variables de la investigación</i>	12
Tabla 2 <i>Resultados de la extracción de la información mediante web scraping</i>	21
Tabla 3 <i>Resultado del filtrado y limpieza</i>	24
Tabla 4 <i>Diccionario de palabras</i>	24
Tabla 5 <i>Diccionario de frases</i>	25
Tabla 6 <i>Polaridad de los emoticones</i>	30
Tabla 7 <i>Reglas para los emoticones</i>	30
Tabla 8 <i>Resultados de la polaridad de los sentimientos</i>	32
Tabla 9 <i>Matriz de confusiones</i>	32

ÍNDICE DE FIGURAS

Figura 1 <i>Fases para el análisis de resultados</i>	18
Figura 2 <i>Proceso de web scraping</i>	19
Figura 3 <i>Número total de información extraída</i>	22
Figura 4 <i>Interfaz para el entrenamiento manual del modelo</i>	28
Figura 5 <i>Resultados del etiquetado manual</i>	29
Figura 6 <i>Modelo de Naïve Bayes</i>	31

RESUMEN

Esta investigación se centra en el análisis exhaustivo de los sentimientos y emociones expresados en las redes sociales por los seguidores de los equipos de fútbol más destacados de la Liga 1 peruana. Utilizando técnicas avanzadas de inteligencia artificial, el estudio se propuso examinar la polaridad emocional de la comunidad de seguidores en respuesta a eventos específicos, actuaciones del equipo y otras variables contextuales. Para lograr esto, se implementó un modelo de aprendizaje automático utilizando el clasificador Naïve Bayes, que ha demostrado ser eficiente en este tipo de análisis.

La investigación no solo se centra en la clasificación de los sentimientos, sino también en la identificación de patrones y tendencias a lo largo del tiempo, lo que demandó una evaluación continua del rendimiento del modelo utilizando diversas métricas de desempeño, permitiendo ajustes iterativos para mejorar la precisión y relevancia del análisis.

Adicionalmente, se expusieron las limitaciones del modelo, proporcionando una visión equilibrada de los resultados marcando las áreas para futuras mejoras y enriquecimientos metodológicos. Finalmente, esta investigación busca ofrecer una perspectiva integral sobre la dinámica emocional en el ámbito futbolístico peruano, proporcionando información valiosa tanto para los equipos como para los profesionales del marketing y la gestión de la reputación.

Palabras clave: Análisis de sentimientos, Inteligencia artificial, Clasificación de sentimientos.

ABSTRACT

This research focuses on the exhaustive analysis of the feelings and emotions expressed on social networks by the followers of the most prominent soccer teams in the Peruvian League 1. Using advanced artificial intelligence techniques, the study set out to examine the emotional polarity of the fan community in response to specific events, team performances, and other contextual variables. To achieve this, a machine learning model was implemented using the Naïve Bayes classifier, which has proven to be efficient in this type of analysis.

The research not only focuses on the classification of sentiments, but also on the identification of patterns and trends over time, which demanded continuous evaluation of the model's performance using various performance metrics, allowing iterative adjustments to improve the accuracy and relevance of the analysis.

Additionally, the limitations of the model were exposed, providing a balanced view of the results, marking areas for future improvements and methodological enrichments. Finally, this research seeks to offer a comprehensive perspective on the emotional dynamics in the Peruvian soccer field, providing valuable information for both teams and marketing and reputation management professionals.

Keywords: Sentiment analysis, Artificial intelligence, Sentiment classification.

I. INTRODUCCIÓN

Desde su aparición las redes sociales se han enfocado en permitir a las personas conectarse y comunicarse digitalmente ofreciendo diversas funciones que facilitan la interacción social, como la creación de perfiles personales, la publicación de contenido, la compartición de información, la interacción con otros usuarios mediante comentarios y mensajes, y la formación de comunidades basadas en intereses comunes. Las redes sociales pueden ser utilizadas para establecer y mantener relaciones personales, compartir experiencias, informarse, entretenerse y promover ideas. Además, las redes sociales también han adquirido un papel relevante en el ámbito empresarial, brindando oportunidades para la promoción de marcas, el marketing digital y la interacción con clientes y seguidores (Gozálvez y Cortijo, 2023).

Los sentimientos y emociones son aspectos intrínsecos de la experiencia humana que implican respuestas afectivas y subjetivas ante estímulos internos o externos. Los sentimientos se refieren a estados afectivos más duraderos, como la alegría, la tristeza, el miedo, la ira o el amor, que pueden estar influenciados por factores cognitivos y sociales. Por otro lado, las emociones son respuestas afectivas más intensas y efímeras que surgen en situaciones específicas, como la sorpresa, el disgusto, la admiración o el temor. Ambos juegan un papel crucial en la vida diaria, influyendo en la toma de decisiones, las interacciones sociales y el bienestar general de las personas (Leal, 2023).

Hoy en día, la inteligencia artificial ha obtenido amplia aceptación en la sociedad debido a su enfoque en desarrollar sistemas capaces de suplantar la inteligencia humana mediante el uso de algoritmos y modelos matemáticos. La inteligencia artificial permite que las máquinas capturen, procesen, aprendan, analicen, tomen decisiones y resuelvan problemas de manera autónoma (sin la intervención humana). Estos sistemas se caracterizan por su capacidad de procesar datos desde distintas fuentes de información, identificar patrones, reconocer imágenes, comprender el lenguaje natural y adaptarse a nuevas situaciones. La inteligencia artificial ha evolucionado significativamente en los últimos años llegando a incursionar en múltiples sectores como la medicina, la industria, la investigación científica y el comercio (Barrios et al., 2020).

La inteligencia artificial (IA) ha revolucionado en su capacidad de analizar los

sentimientos y emociones expresados en las principales redes sociales permitiendo el procesamiento automático y la interpretación de grandes volúmenes de datos. Los algoritmos de aprendizaje automático han demostrado ser herramientas eficaces para analizar el contenido de las publicaciones y determinar la polaridad emocional asociada a ellas, es decir, si expresan sentimientos positivos, negativos o neutros (Darad y Krishnan, 2023). Según (More et al., 2020) el Procesamiento del Lenguaje Natural (PNL) es una herramienta útil para analizar opiniones sobre temas sociales complejos, ya que permite identificar los temas más importantes que se abordan en las opiniones, así como las perspectivas y opiniones de los diferentes actores involucrados.

En Perú, como en el resto de los países del mundo, el fútbol es considerado como el deporte más popular ya que se vive con intensidad en todos los rincones, desde temprana edad, los peruanos muestran un fervor por el fútbol, ya sea jugándolo en las calles y canchas improvisadas o siguiendo a su equipo favorito en los estadios o a través de las transmisiones televisivas. Asimismo, los clubes de fútbol peruanos tienen una base de seguidores apasionados siendo los equipos más con mayor hinchada el Sporting Cristal, Universitario de Deportes y Alianza Lima que cuentan con una ferviente hinchada que los respalda en cada encuentro. Lo que hace que los equipos de fútbol utilicen las redes sociales para mantener contacto con sus hinchas y mantener e incrementar esa pasión que el deporte genera.

En este sentido, en la actual era digital, las redes sociales juegan un papel fundamental a través de las cuales los hinchas del fútbol generan una alta interacción social mediante la expresión de opiniones, ideas, emociones y sentimientos, lo que ha generado que los internautas no sean solo consumidores de información, sino que pasen a ser generadores de la misma. Por lo tanto, el alto volumen de datos generados en las redes sociales dificulta la capacidad de analizar la manera, el nivel y la intensidad los sentimientos y las emociones expresados en estos medios digitales.

En este sentido, se formula la siguiente pregunta de investigación ¿De qué manera los internautas expresan sus sentimientos y emociones en las redes sociales de sus equipos favoritos de la Liga 1 peruana y cómo la inteligencia artificial permite realizar el análisis de éstas publicaciones como insumo que guíe a los Community Managers en su planificación de publicaciones?

Por lo antes expuesto, la presente investigación pretende explorar y evaluar diferentes enfoques de inteligencia artificial relacionados con el análisis de los sentimientos expresados en medios digitales, utilizando técnicas de inteligencia artificial capaces de identificar y clasificar automáticamente el tono emocional de las interacciones generadas en redes sociales de los diferentes equipos de fútbol peruano. Adicionalmente, se pretende proporcionar una comprensión más profunda de las emociones expresadas en las redes sociales, que pueden ser replicadas en otros ámbitos. En este sentido, el objetivo general de la presente investigación es el de Implementar un modelo de inteligencia artificial que permita analizar los sentimientos y emociones en las redes sociales de los equipos de una liga de fútbol, y los objetivos específicos: Desarrollar un modelo de aprendizaje automático que permita analizar y clasificar los sentimientos y emociones expresados en los textos de las publicaciones y comentarios en redes sociales, recopilar y preparar un conjunto de datos etiquetados adecuado que represente una variedad de sentimientos y emociones en diferentes contextos de las redes sociales, identificar patrones y tendencias en la expresión de sentimientos y emociones en las redes sociales a través del análisis de los resultados obtenidos y evaluar y mejorar el rendimiento del modelo de inteligencia artificial aplicando las diferentes métricas de desempeño.

II. MARCO TEÓRICO

A lo largo del marco teórico, se examinará los conceptos clave relacionados con las redes sociales, las emociones humanas, el análisis de sentimientos y las técnicas avanzadas de procesamiento del lenguaje natural, que son fundamentales para abordar este emocionante campo de investigación.

Los sentimientos son experiencias subjetivas que surgen como respuesta a una situación o evento específico, son estados afectivos complejos y duraderos que implican la interpretación cognitiva de la experiencia. Los sentimientos pueden variar en intensidad y pueden estar influenciados por factores cognitivos, culturales y contextuales. Por otro lado, las emociones son respuestas afectivas rápidas y automáticas ante estímulos emocionales. Son respuestas biológicas y psicológicas que ocurren en un nivel más primario y básico. Las emociones son respuestas innatas y universales, como el miedo, la alegría, la ira, la tristeza, entre otras, y suelen tener una duración más corta que los sentimientos (Murillo y Jiménez, 2019). Las redes sociales han surgido como una poderosa fuerza transformadora en la era digital, redefiniendo la forma en que las personas se conectan, comparten información y construyen comunidades en línea, estas plataformas digitales permiten a individuos y grupos de todo el mundo crear perfiles, compartir contenido, interactuar con otros usuarios y participar en conversaciones sobre una amplia variedad de temas. En este sentido, brindan un espacio virtual donde las personas pueden expresar sus emociones de manera inmediata y a gran escala, permitiendo que sus sentimientos sean visibles para su red de contactos y potencialmente para una audiencia más amplia. La expresión de sentimientos en redes sociales es un fenómeno relevante en la actualidad, ya que proporciona una ventana a las actitudes, experiencias y preocupaciones emocionales de las personas en la era digital (Gozálvez y Cortijo, 2023).

La inteligencia artificial (IA) desempeña un papel vital para analizar altos volúmenes de datos, considerando que una de sus principales características se basa en la capacidad para extraer y procesar información de manera eficiente. Con el incremento exponencial de datos que se generan en el ciberespacio, se vuelve cada vez más difícil y laborioso para los seres humanos analizarlos y extraer conocimientos relevantes. En este escenario, la IA se convierte en un factor crucial al utilizar algoritmos que permiten analizar y comprender estos vastos conjuntos de

datos de manera efectiva (Zapata, 2020).

La IA tiene la capacidad de detectar patrones, tendencias y relaciones ocultas en los datos, lo que posibilita la revelación de información relevante que permite tomar decisiones en base a los datos obtenidos. Esto es especialmente relevante en áreas como el comercio electrónico, la medicina, la industria financiera, la investigación científica y muchas otras, donde el análisis de grandes volúmenes de datos es fundamental para el desarrollo de estrategias efectivas, la detección de anomalías y la personalización de servicios (López, 2021).

La IA posee características diferenciadas que le permite procesar y analizar los datos de forma rápida y precisa, impulsa la eficiencia y la precisión en la identificación de patrones y tendencias. Esta habilidad tiene un impacto significativo en la innovación, la mejora de servicios y la optimización de procesos en diversos sectores. Además, la inteligencia artificial tiene la capacidad de automatizar tareas de análisis de datos que antes requerían una inversión considerable de tiempo y esfuerzo, permitiendo a los profesionales enfocarse en tareas más complejas y estratégicas.

En la actualidad existen varios métodos de IA, mismo que están compuestos por un conjunto diverso de enfoques y técnicas que permiten que las máquinas puedan aprender a realizar tareas que normalmente requerirían de la inteligencia humana. Uno de los principales métodos utilizados para el análisis de sentimientos es el aprendizaje automático supervisado mismo que mediante el uso de técnicas de aprendizaje automático permite entrenar los modelos para que puedan clasificar automáticamente los textos en categorías de sentimiento, como positivo, negativo o neutro. Para lo cual, se proporcionan ejemplos etiquetados de textos con sus respectivos sentimientos para que posteriormente el modelo aprenda a reconocer patrones y características que están asociados con diferentes sentimientos (Brito y Brito, 2018). Es por esto que el aprendizaje automático supervisado es ampliamente utilizado en diversas aplicaciones relacionadas con la detección de fraudes, el análisis de sentimientos, la visión por computadora y muchas más, donde se requiere hacer predicciones precisas basadas en datos previamente etiquetados.

Otro de los métodos, también, utilizado es el de redes neuronales recurrentes (RNN) su característica principal radica en su capacidad para procesar secuencias

de texto y capturar las dependencias temporales en los datos, se utilizan en el análisis de sentimientos para capturar el contexto y la estructura de los textos en las redes sociales. Las RNN pueden aprender a modelar las relaciones entre las palabras y comprender cómo influyen en la expresión emocional global (Darad y Krishnan, 2023). Sin embargo, las RNN tradicionales tienen limitaciones, como el problema de desvanecimiento o explosión del gradiente, que dificulta el aprendizaje de dependencias a largo plazo en secuencias. Para superar estas limitaciones, se han desarrollado variantes más avanzadas de RNN, como las redes neuronales LSTM (Long Short-Term Memory) y GRU (Gated Recurrent Unit), que incorporan mecanismos de memoria más sofisticados y son capaces de manejar secuencias más largas y complejas. Estas variantes han demostrado ser muy efectivas en una amplia gama de aplicaciones de procesamiento de secuencias.

El último método de IA es el de procesamiento del lenguaje natural (NLP), mismo que está diseñado para comprender, interpretar y generar texto en lenguaje humano de manera efectiva, emplean técnicas avanzadas de aprendizaje automático, en particular el aprendizaje profundo, para analizar y procesar datos de texto de una manera que se asemeje a la comprensión humana del lenguaje. Su versatilidad abarca una amplia gama de aplicaciones, desde la traducción automática y la generación de texto hasta la extracción de información y el análisis de sentimientos en redes sociales. Un hito importante en el campo de los modelos de NLP fue la introducción de Transformers, que permiten el procesamiento contextualizado de palabras y han dado lugar a modelos pre-entrenados como BERT y GPT que han revolucionado la capacidad de las máquinas para entender y generar texto de manera más natural y coherente. (Peña et al., 2019).

Así también existen modelos clasificadores de IA que se utilizan para asignar objetos o instancias a diferentes categorías o clases, estos modelos requieren de un entrenamiento con datos previamente etiquetados para aprender patrones y características que permitan realizar predicciones precisas sobre nuevas instancias no etiquetadas. El modelo de Naïve Bayes se basa en la suposición "ingenua" de que las características utilizadas para la clasificación son independientes entre sí, es decir, que la presencia de una característica en particular no afecta la presencia de otras características. Aunque esta suposición puede no ser cierta en todos los casos, Naïve Bayes a menudo funciona sorprendentemente bien en una amplia

variedad de aplicaciones, como la clasificación de correos electrónicos como spam o no spam, la detección de sentimientos en texto, la clasificación de documentos y más. (Rico et al., 2019).

Así también, el modelo de árboles de decisión cuya clasificación la representa mediante una estructura de árbol; para lo cual, cada uno de los nodos internos representan una pregunta o característica, y cada rama la posible respuesta. Estos árboles se construyen mediante la división recursiva del conjunto de datos en función de las características más relevantes, buscando maximizar la pureza de las clases en cada subdivisión. A medida que se desciende por el árbol, las instancias se asignan a una clase final en las hojas. Los árboles de decisión son fáciles de interpretar y visualizar, y pueden manejar tanto características numéricas como categóricas. (Atamari et al., 2022).

Adicionalmente, existe el modelo de máquinas de vectores de soporte (SVM) cuyo uso es mayoritariamente para la clasificación binaria, pudiendo extenderse a problemas de clasificación multiclase, estas máquinas tienen el objetivo de maximizar el margen entre las instancias de las diferentes clases y el hiperplano de separación. Las SVM son conocidas por su eficacia en problemas de clasificación en los que los datos son linealmente separables, es decir, se pueden dividir en dos clases mediante una línea, un plano o un hiperplano. Sin embargo, una característica poderosa de las SVM es su capacidad para manejar datos no linealmente separables mediante el uso de "kernels" que transforman los datos en un espacio de características de mayor dimensión donde la separación lineal es posible. Los kernels más comunes incluyen el kernel lineal, el kernel polinómico y el kernel radial (RBF). (Aparicio et al., 2023).

Otro modelo altamente conocido es el de vecinos más cercanos (KNN), cuyo algoritmo es utilizado para problemas de clasificación y reconocimiento de patrones, donde las instancias se clasifican en función de las clases de sus vecinos más cercanos acorde a sus características; por lo que, el valor de "K" simboliza el número de vecinos más cercanos considerados para la clasificación. El KNN busca instancias similares en términos de distancia y asigna la clase más común entre los vecinos seleccionados. Este enfoque es especialmente útil en conjuntos de datos donde la separación entre clases es clara. El KNN es fácil de entender e implementar, y es utilizado en una gran variedad de aplicaciones, desde

reconocimiento de imágenes hasta sistemas de recomendación. Sin embargo, es importante elegir adecuadamente el valor de "K" y tener en cuenta el impacto de las características y la métrica de distancia utilizada en el rendimiento del modelo. (Contreas et al., 2023).

El último modelo en ser descrito es el de regresión logística, éste permite predecir la probabilidad de pertenencia a una clase en lugar de predecir un valor numérico, este modelo se utiliza una función logística para transformar una combinación lineal de las características en una probabilidad, que representa la probabilidad de pertenecer a la clase positiva. Para clasificar nuevas instancias, se utiliza un umbral de probabilidad para decidir a qué clase pertenecen. Este tipo de regresión es ampliamente utilizado por su simplicidad, interpretabilidad y eficiencia computacional. Así también, puede ser utilizado para resolver problemas de clasificación multiclase utilizando técnicas como la regresión logística multinomial o la regresión logística ordinal. (Mazuera et al., 2021).

Para comprobar la fiabilidad de los modelos, existen diversas métricas de desempeño que son utilizadas para evaluar la calidad y el rendimiento de los clasificadores con inteligencia artificial, su importancia radica en su capacidad de evaluar y comparar el desempeño relacionado con la capacidad de clasificación, la exactitud y la capacidad de detectar correctamente instancias positivas y negativas. Para lo cual, la precisión (Accuracy) es una métrica que permite evaluar la calidad de un clasificador, su precisión proporciona una medida general del rendimiento del clasificador y es especialmente útil cuando las clases están equilibradas en el conjunto de datos. Sin embargo, puede ser engañosa cuando las clases están desequilibradas pudiendo tener un desempeño deficiente en la clasificación de la clase minoritaria. Por lo tanto, es importante considerar otras métricas y tener en cuenta el contexto del problema para obtener una evaluación más completa del rendimiento del clasificador (Aparicio et al., 2023).

Adicionalmente, la matriz de confusión (Confusión Matrix) permite mostrar la relación entre las clasificaciones reales y las clasificaciones predichas, expresados en una matriz de confusión, en donde cada fila representa la clase real de las instancias y cada columna representa la clase predicha por el modelo. Siendo la diagonal principal de la matriz la que represente las instancias correctamente clasificadas, mientras que los valores que se encuentran por fuera de dicha

diagonal indican las instancias clasificadas incorrectamente (Brito y Brito, 2018). Así también, la sensibilidad (Recall) es una métrica fundamental utilizada para medir la capacidad de un modelo para identificar correctamente las instancias positivas. Su cálculo es el resultado de dividir el total de verdaderos positivos para la suma de los verdaderos positivos y los falsos negativos. Por otro lado, el cálculo de la tasa de verdaderos negativos (especificidad), es considerada una métrica que mide la capacidad del modelo para identificar adecuadamente las instancias negativas. Su cálculo es el resultado de dividir el total de verdaderos negativos para la suma de los verdaderos negativos y los falsos positivos. En este sentido, las dos medidas son métricas complementarias y brindan información sobre la capacidad de respuesta del modelo (Rosso et al., 2022).

Otra métrica importante para considerar es la de valor F1 score, misma que se consta de un índice que tiene la capacidad de combinar la precisión y la sensibilidad en una única medida, se utiliza comúnmente en problemas de clasificación donde las clases están desequilibradas. Su cálculo provee una medida equilibrada del rendimiento del clasificador al incluir en su cálculo tanto la capacidad de clasificar correctamente instancias positivas como negativas. El valor F1 oscila entre 0 y 1, donde 1 representa una clasificación perfecta. Un valor F1 alto indica un buen equilibrio entre la precisión y la sensibilidad, lo que significa que el clasificador logra tanto una baja tasa de falsos positivos como una baja tasa de falsos negativos (Atamari et al., 2022).

Una última métrica que debe ser considerada es la conocida como Curva ROC (Receiver Operating Characteristic), que contempla una representación gráfica en forma de curva que es utilizada mayoritariamente en problemas de clasificación para evaluar el desempeño de un modelo en la detección de verdaderos positivos y falsos positivos sobre diferentes umbrales de clasificación. Dentro de la curva ROC cada punto representa un umbral de clasificación diferente y, en general, cuanto más cercana esté la curva al rincón superior izquierdo del gráfico, significa que el modelo tiene un buen desempeño. La curva ROC también se utiliza para calcular el área bajo la curva (AUC), que proporciona una medida relacionada con la calidad del modelo. Un AUC cercano a 1 indica un rendimiento excelente, mientras que un valor cercano a 0.5 indica un rendimiento aleatorio. Por lo tanto, estas dos curvas son herramientas útiles para comparar modelos y seleccionar el

umbral de clasificación adecuado según los requisitos del problema. (Contreas et al., 2023).

Finalmente, la ética en el análisis de sentimientos reviste una importancia creciente en un mundo digital interconectado. Cuando se utilizan técnicas de inteligencia artificial y procesamiento del lenguaje natural para comprender las emociones y opiniones expresadas en redes sociales, comentarios en línea y otros datos textuales, es esencial considerar cuestiones éticas clave. Esto incluye la protección de la privacidad de los usuarios, la transparencia en la toma de decisiones automatizada, la identificación y mitigación de sesgos, así como la responsabilidad en la interpretación de resultados. También se debe prestar atención a la equidad y la no discriminación, ya que los modelos de análisis de sentimientos pueden influir en la percepción pública y en las decisiones comerciales. La ética en el análisis de sentimientos implica un equilibrio delicado entre la innovación tecnológica y la responsabilidad hacia las personas cuyos datos y emociones están siendo analizados, con el objetivo de garantizar que estos avances beneficien a la sociedad en su conjunto.

III. METODOLOGÍA

3.1 Tipo y diseño de investigación

3.1.1 Tipo de investigación

Se utilizó la investigación aplicada, considerando que se caracteriza por la aplicación de conocimientos teóricos y prácticos existentes para resolver problemas o desafíos concretos (Aguirre et al., 2013). En el análisis de sentimientos y emociones en redes sociales, se desarrolló y evaluó los algoritmos de aprendizaje automático específicos, adicionalmente se implementaron sistemas de procesamiento de lenguaje natural para clasificar publicaciones y generar reportes para comprender mejor las tendencias emocionales en redes sociales.

3.1.2 Diseño de investigación

Se desarrolló utilizando el diseño experimental, considerando que este tipo de diseño puede establecer grupos de control y grupos de tratamiento para evaluar el impacto de diferentes técnicas o algoritmos (Rodríguez et al., 2021) de inteligencia artificial para el análisis de sentimientos en redes sociales. Para lo cual, se pueden aplicar diferentes enfoques de procesamiento del lenguaje natural o algoritmos de aprendizaje automático a los datos recopilados de las redes sociales y comparar los resultados. Este diseño permitirá establecer relaciones causa-efecto y determinar qué técnicas son más eficaces en el análisis de sentimientos.

3.2 Variables y operacionalización

3.2.1 Definición conceptual

En el análisis de sentimientos en redes sociales utilizando la inteligencia artificial, las variables dependientes e independientes pueden variar al momento en que se cuente con la información extraída; sin embargo, a continuación, se presentan las variables dependientes e independientes identificadas al momento de redactar el presente documento.

Tabla 1*Variables de la investigación*

Tipo de variables	Descripción
Dependientes	<p><u>Temas emocionales:</u> Esta variable se refiere a los temas o categorías emocionales identificados en las publicaciones, como alegría, tristeza, enfado, miedo, etc.</p> <p><u>Polaridad emocional:</u> Esta variable representa la clasificación de la publicación en redes sociales como positiva, negativa o neutra en términos de sentimiento o emoción expresada.</p> <p><u>Intensidad emocional:</u> Esta variable refleja la medida de la fuerza o grado de la emoción expresada en la publicación, indicando si es fuerte o débil.</p>
Independientes	<p><u>Técnicas de procesamiento:</u> Esta variable representa las técnicas utilizadas para procesar y analizar el contenido textual.</p> <p><u>Algoritmos de aprendizaje automático:</u> Esta variable se refiere a los algoritmos utilizados para clasificar y analizar los datos, como algoritmos de clasificación, redes neuronales, modelos basados en reglas, etc.</p> <p><u>Características del texto:</u> Esta variable incluye aspectos del texto de las publicaciones, como la longitud del texto, el uso de palabras clave, el tono o estilo lingüístico, etc.</p>

3.2.2 Definición operacional

La definición operacional de la presente investigación implicó la especificación clara y concreta de los procedimientos utilizados. A continuación, se presentan los elementos clave de la definición operacional:

- **Recopilación de datos:** Se extrajeron los datos de las publicaciones realizadas por los equipos más populares de la liga 1 peruana durante el segundo semestre del año 2023 utilizando las API's oficiales de dichas redes.
- **Extracción de características:** Para la identificación de características

relevantes para el análisis de las publicaciones se aplicaron técnicas como el conteo de palabras, el análisis de frecuencia, la identificación de términos clave o la representación vectorial del texto mediante modelos de lenguaje pre entrenados.

- Clasificación de sentimientos: Utilizando la IA se clasificaron las publicaciones en categorías de sentimiento, como positivo, negativo o neutro. Esto puede involucrar la implementación de algoritmos de aprendizaje automático, como clasificadores basados en reglas, SVM (Máquinas de Vectores de Soporte) o redes neuronales, entrenados con conjuntos de datos etiquetados.
- Evaluación y validación: Se aplicaron métricas de rendimiento, como precisión, exhaustividad y puntuación F1, comparando los resultados obtenidos por el modelo de análisis de sentimientos con un grupo de datos de prueba o con etiquetas de sentimiento analizados manualmente (Darad y Krishnan, 2023).

3.2.3 Indicadores

Al realizar el análisis de sentimientos en redes sociales a través de la IA, se aplicaron distintos indicadores para evaluar y medir los resultados obtenidos. A continuación, se presentan algunos ejemplos de indicadores relacionados con esta área de investigación:

- Velocidad de procesamiento: Mide la rapidez con la que el modelo de análisis de sentimientos es capaz de procesar y clasificar las publicaciones en redes sociales. En este sentido, es importante considerar como indicador el porcentaje de precisión, mismo que permite medir la proporción de clasificaciones correctas realizadas por el modelo de análisis de sentimientos en comparación con las etiquetas de sentimiento de referencia o con las evaluaciones realizadas manualmente. Una alta precisión indica una mayor capacidad del modelo para clasificar correctamente las publicaciones en redes sociales en términos de su polaridad emocional.
- Tiempo de entrenamiento: Este tiempo puede variar ampliamente según la complejidad del modelo, el tamaño de los datos y la potencia computacional disponible. Para lo cual, es importante considerar como indicador al

porcentaje de sensibilidad que mide la proporción de clasificaciones positivas reales que un modelo ha identificado correctamente, este valor se expresa típicamente como un porcentaje y muestra qué tan efectivo es un modelo en detectar positivos verdaderos en relación con todos los casos positivos reales. Una sensibilidad alta indica que el modelo tiene una capacidad sólida para identificar ejemplos positivos; por otro lado, una baja sensibilidad indica que el modelo no es efectivo en la identificación de positivos verdaderos.

- Promedio de la longitud del texto: Se refiere a la medida que representa la extensión típica de las publicaciones, un promedio de longitud del texto más corto sugiere mensajes concisos y directos, mientras que un promedio más largo indica contenido más detallado. Este indicador es valioso para comprender la complejidad de las emociones expresadas, detectar contenido relevante y adaptar enfoques de análisis. Así también permite cuantificar la frecuencia de ciertos términos específicos que permite identificar rápidamente la emoción expresada.
- Frecuencia de emociones: Se refiere a la proporción de las diferentes emociones expresadas en las publicaciones, permite cuantificar y comprender qué emociones son más prevalentes en una plataforma específica, lo que puede ser fundamental para detectar tendencias emocionales en eventos, campañas o discusiones en línea. El seguimiento de la frecuencia de emociones es esencial para la toma de decisiones basada en datos.
- Polaridad emocional promedio: Muestra la distribución de las publicaciones en redes sociales en términos de su polaridad emocional, permitiendo identificar el porcentaje de publicaciones positivas, negativas y neutras en el conjunto de datos analizados. Este indicador proporciona una visión general de la distribución de sentimientos en las publicaciones.
- Distribución de intensidad: Se refiere a cómo se distribuyen las intensidades emocionales en el contenido de las publicaciones, esta métrica permite identificar si las emociones expresadas son mayoritariamente leves o intensas. Comprender la distribución de intensidad es fundamental para evaluar la profundidad de las emociones expresadas en línea y para ajustar

estrategias de acuerdo con la intensidad emocional deseada en una determinada campaña o contexto de redes sociales.

3.3 Población, muestra y muestreo

3.3.1 Población

La población estuvo dada por los 18 equipos que para el año 2023 se encuentran disputando la Liga 1.

3.3.2 Muestra

La muestra estuvo seleccionada considerando a los equipos que actualmente se encuentra jugando en la Liga 1, ya que su popularidad se da por la cantidad de hinchas que tienen factor que mantiene directa proporción con el número de seguidores en redes sociales y la interacción que cada publicación genera en el mundo digital.

Adicionalmente a seleccionar a las cuentas de redes sociales de los equipos que formaron parte del estudio es importante definir que las publicaciones a ser analizadas para identificar los sentimientos y emociones fueron las publicaciones realizadas en el segundo semestre del año 2023.

3.3.3 Muestreo

Para la presente investigación no se utilizó ninguna técnica de muestreo considerando primero que el número de equipos que juegan la liga 1 no permite la aplicación de una muestra y como segunda consideración se debe tener en cuenta que de manera manual es imposible determinar el número de publicaciones en las redes sociales de cada uno de los equipos y mucho menos la interacción de sus seguidores.

3.3.4 Unidad de análisis

La unidad de análisis fueron las publicaciones realizadas a través de las redes oficiales de los equipos de fútbol que disputan la Liga 1 peruana.

3.4 Técnicas e instrumentos de recolección de datos

Para la presente investigación se utilizó web scraping como técnicas e instrumentos de recolección de datos, considerando que implica la extracción automatizada de datos de las redes sociales a través del uso de herramientas y bibliotecas de programación para extraer publicaciones, comentarios o mensajes relevantes de las plataformas sociales (Rosso et al., 2022). Estos datos se pueden recopilar en función de palabras clave, hashtags, ubicaciones geográficas, perfiles de usuarios, entre otros criterios.

3.5 Procedimiento

El procedimiento para la recolección de datos incluyó los siguientes pasos:

- Seleccionar las redes sociales: Identificar la red social más utilizada por los equipos de fútbol para el análisis de sentimientos.
- Obtener acceso a los datos: Se debe obtener acceso a los datos a través de web scraping.
- Definir los criterios de búsqueda: Determinar los criterios de búsqueda para filtrar y obtener los datos relevantes para el análisis de sentimientos. Esto puede incluir palabras clave, hashtags, ubicaciones geográficas, fechas o perfiles de usuarios específicos.
- Almacenar y organizar los datos: Se requiere guardar los datos recopilados en un formato adecuado y asegurarse de que estén estructurados y organizados para facilitar el posterior análisis.
- Anonimizar y proteger la privacidad: Es necesario proteger la privacidad de los usuarios, mediante el uso de los datos únicamente para responder a los objetivos de esta investigación.

3.6 Método de análisis de datos

El análisis de datos se realizó aplicando los siguientes métodos:

- Análisis basado en reglas: Este método implica el uso de reglas y patrones predefinidos para clasificar las publicaciones en categorías de sentimiento. Las reglas pueden incluir la presencia de palabras clave o patrones gramaticales específicos que indiquen una polaridad emocional determinada.

- Aprendizaje automático supervisado: Este método implica entrenar un modelo de aprendizaje automático con un conjunto de datos etiquetados, donde cada publicación tiene una etiqueta de sentimiento conocida. El modelo aprende a clasificar nuevas publicaciones en función de las características extraídas del texto. Se pueden utilizar algoritmos como Naive Bayes, Máquinas de Vectores de Soporte (SVM) o Bosques Aleatorios para entrenar y evaluar el modelo.

3.7 Aspectos éticos

Durante la ejecución de la presente investigación se consideraron ciertos aspectos éticos propios de los proyectos relacionados con la inteligencia artificial:

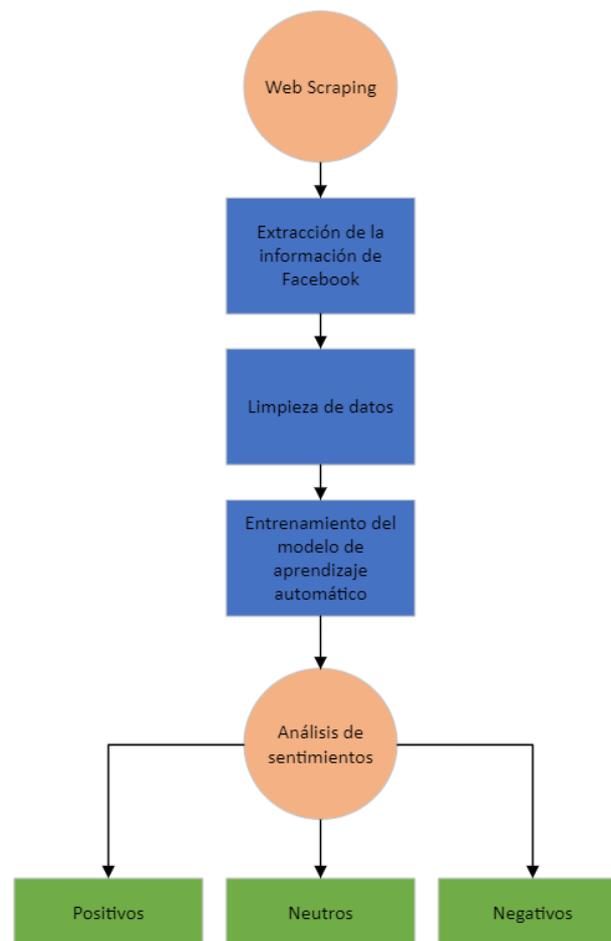
- Protección de datos y privacidad: Es imperante garantizar la privacidad de los usuarios y proteger sus datos personales durante la recopilación y análisis de sentimientos en redes sociales; para lo cual, se deben cumplir estrictamente las regulaciones y normativas relacionadas con la protección de datos de los usuarios asegurando que la información recopilada se utilice únicamente con fines legítimos y dentro de los límites establecidos.
- Sesgo y discriminación: Es crucial realizar una evaluación continua y mitigar cualquier sesgo en los modelos de inteligencia artificial utilizados, garantizando que los resultados sean justos, equitativos y no discriminatorios.
- Transparencia: Es importante que los resultados sean comprensibles y explicables, de manera que los usuarios y las partes interesadas puedan entender cómo se obtienen y utilizan los datos.
- Uso responsable de los resultados: Los resultados obtenidos como producto del análisis de sentimientos en redes sociales deben utilizarse de manera responsable y ética. Es importante evitar el uso indebido de la información para manipular o dañar a los usuarios, así como para generar perjuicios o discriminación.

IV. RESULTADOS

Para el proceso de análisis de los sentimientos expresados en las redes sociales de los principales equipos de fútbol de la Liga 1 peruana, acorde a lo descrito en la Figura 1 se realizaron los siguientes procesos que garantizaron la extracción, limpieza y clasificación de los sentimientos.

Figura 1

Fases para el análisis de resultados



Nota. Elaboración Propia (2023).

4.1 Web Scraping

La importancia de seleccionar la técnica apropiada de web scraping se refleja directamente en la precisión al extraer datos de diversas Fan Pages objeto de este estudio, permitiendo la obtención fiable de información que sustenta los análisis posteriores. Asimismo, una técnica adecuada asegura la eficacia del proceso y el apego a las regulaciones legales y éticas, elementos fundamentales para obtener

datos confiables y oportunamente actualizados. En este contexto, se detallan a continuación las definiciones establecidas para el desarrollo del web scraping.

Figura 2

Proceso de web scraping



Nota. Tomado de MIOTI (2022).

4.1.1 Definición de datos a extraer

Para el análisis de los sentimientos expresados en las publicaciones de Facebook de todos los equipos de la liga 1 del fútbol peruano se ha definido que la información a extraer corresponde a las publicaciones realizadas en las Fan Pagues oficiales y a todos los comentarios generados por sus seguidores.

4.1.2 Definición de las fuentes de información

Las fuentes de información de las que se extraerá la información corresponden a las Fan Pagues oficiales de cada uno de los equipos, las identificadas son:

- <https://www.facebook.com/atleticograu>
- <https://www.facebook.com/UnionComercioOficial>
- <https://www.facebook.com/AcademiaCantolao>
- <https://www.facebook.com/CuscoFCoficial>
- <https://www.facebook.com/ClubSportBoys>
- <https://www.facebook.com/camannucci>
- <https://www.facebook.com/DeportivoBinacionalFC>
- <https://www.facebook.com/ClubCiencianoOficial>
- <https://www.facebook.com/ucvclubdefutbol>

- <https://www.facebook.com/ClubDeportivoGarcilasoOficial>
- <https://www.facebook.com/sporthuancayopaginaoficial>
- <https://www.facebook.com/ClubDeportivoMunicipal>
- <https://www.facebook.com/FBCMelgar>
- <https://www.facebook.com/CSCristal>
- <https://www.facebook.com/Universitario>
- <https://www.facebook.com/ClubAlianzaLima>
- <https://www.facebook.com/CuscoFCoficial>
- <https://www.facebook.com/adtdtarma>
- <https://www.facebook.com/clubUTC.pe>

4.1.3 Definición de la frecuencia del scraping

Para el entrenamiento del modelo se ha extraído toda la información generada entre los meses de agosto, septiembre y octubre del presente año; sin embargo, el web scraping desarrollado está en la capacidad de extraer información en el rango de tiempo que los usuarios definan.

4.1.4 Definición de la herramienta de web scraping

Para garantizar el correcto desarrollo del web scraping se utilizaron las siguientes herramientas, mismas que son las más utilizadas en desarrollo similares alrededor del mundo:

- Entornos de Desarrollo Integrado (IDE): Jupyter Notebook
- Lenguaje de programación: Python
- Librería: facebook-scraper

4.2 Extracción de la información de Facebook

La extracción de la información corresponde el primer paso para el posterior análisis de los sentimientos, en el Anexo 1 se encuentra el código utilizado para el efecto. En la Tabla 2 se muestra el total de información extraída para la presente investigación, se puede evidenciar que de 7 equipos no se pudo extraer información: “Club Sport Boys”, “Club Ciencia no Oficial”, “Sport Huancayo página oficial”, “CS Cristal”, “Universitario”, “ADT dtarma”, “Club UTC.pe”, mismo que por

la falta de información se excluyen del análisis.

Tabla 2

Resultados de la extracción de la información mediante web scraping

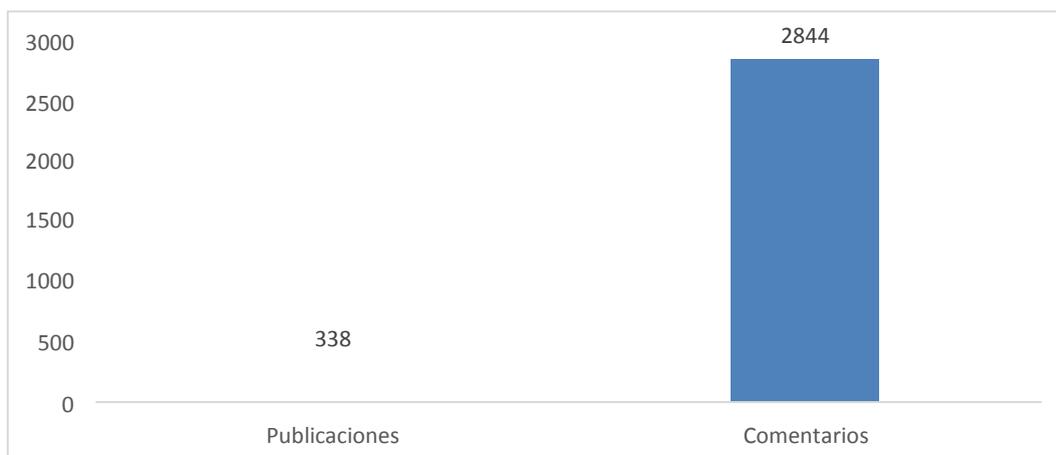
Equipo	Publicaciones	Comentarios
UCO	71	1094
CAG	100	358
CAN	100	175
CUS	19	5
SBA	0	0
MAN	18	246
BIN	14	128
CIE	0	0
UCV	1	8
GAR	3	53
HUA	0	0
MUN	7	183
MEL	2	40
SC	0	0
U	0	0
AL	3	554
ADT	0	0
UTC	0	0
TOTALES	338	2,844

Nota. Elaboración Propia (2023).

Resumiendo, se obtuvieron desde scrapers de Facebook un total de 338 publicaciones, mismas que generaron 2.844 comentarios de sus seguidores (ver Figura 3) lo que significa que en promedio por cada publicación se generan aproximadamente 8 comentarios. Este resultado representa los datos primarios de esta investigación, los cuales fueron obtenidos desde el código python “facebook_posts_comments_scraper.py”, mismo que exporta a carpetas independientes con el nombre de cada equipo archivos en formato JSON con toda la información extraída.

Figura 3

Número total de información extraída



Nota. Elaboración Propia (2023).

4.3 Filtrado y limpieza de los datos

El filtrado y la limpieza de datos son dos fases clave en el proceso de preparación de datos para el análisis. El filtrado implica la identificación y la eliminación de datos no deseados o irrelevantes, como valores atípicos, duplicados, o información que no contribuye significativamente al análisis. Por otro lado, la limpieza de datos se centra en corregir errores, completar datos faltantes y estandarizar el formato para mejorar la calidad general de los datos. En este sentido los criterios aplicados fueron:

- **Todo el texto es cambiado a minúsculas:** Ayuda a estandarizar el texto y a reducir la variabilidad de las palabras, ya que las palabras en minúsculas y mayúsculas se consideran equivalentes. Además, al convertir todo el texto a minúsculas, se evita que el modelo considere dos palabras idénticas en minúsculas y mayúsculas como diferentes, mejorando así la coherencia y precisión del análisis.
- **Cambio del texto con tildes:** La eliminación de tildes asegura que las palabras acentuadas se representen de la misma manera que las no acentuadas, simplificando el procesamiento y asegurando una mayor coherencia en el texto.
- **Eliminación de links:** Al eliminar los enlaces, se simplifica el texto, reduciendo posibles problemas relacionados con la interpretación errónea

de los algoritmos de análisis de sentimientos. Esta práctica se lleva a cabo para concentrarse en el contenido textual principal y evitar que los enlaces, que son cadenas de caracteres con información específica, influyan en la evaluación del sentimiento del texto.

- **Eliminación de usuarios mencionados:** La eliminación de estos nombres de usuario puede ser considerada durante la etapa de limpieza de datos para evitar que estos identificadores específicos influyan en el análisis de sentimientos. Al eliminar los nombres de usuario, se busca centrar el análisis en el contenido del texto y no en las menciones a usuarios individuales, permitiendo un análisis más general de las emociones y opiniones expresadas.
- **Eliminación de stopwords:** Las stopwords son palabras comunes y frecuentes en un idioma que generalmente no aportan significado contextual al análisis, como "el", "un", "de", entre otras. Al eliminar estas palabras del texto, se reduce el ruido y se simplifica el análisis, ya que permite enfocarse en las palabras más relevantes que pueden aportar más información sobre el sentimiento expresado. Esta práctica contribuye a mejorar la eficacia de los algoritmos de análisis de sentimientos al concentrarse en las palabras que transmiten la carga emocional o de opinión, lo que puede llevar a una mejor comprensión de los sentimientos expresados en el texto.
- **Eliminación de comentarios duplicados:** Este proceso implica identificar y eliminar comentarios idénticos o muy similares, ya que su presencia no aporta información adicional al análisis y puede sesgar los resultados. Al eliminar los comentarios duplicados, se mejora la precisión de los resultados y se evita el impacto desproporcionado que estos comentarios podrían tener en el análisis de sentimientos.

Finalmente, como se describe en la Tabla 3 una vez que se aplicaron todos los criterios de filtrado y limpieza se queda con un total de 2.844 datos, mismos que pasarán a la siguiente fase del proyecto que es el análisis de la polaridad de los sentimientos.

Tabla 3*Resultado del filtrado y limpieza*

Criterio	Total
Datos extraídos	2.844
Datos luego del filtrado y limpieza	2.844

Nota. Elaboración Propia (2023).

4.4 Entrenamiento del modelo de aprendizaje automático

4.4.1 Definición del diccionario de palabras y frases

Este diccionario de palabras y frases se concibe como una herramienta fundamental para descifrar la complejidad de las expresiones emocionales en este contexto específico. En un universo donde cada partido desencadena una sinfonía de comentarios y reacciones, este compendio se convierte en un mapa lingüístico que busca clasificar las expresiones en positivas, negativas o neutras. Desde exaltadas celebraciones hasta las críticas más fundamentadas, estas frases encapsulan la diversidad de emociones que inundan las plataformas digitales, permitiendo así una interpretación más precisa y contextualizada del pulso emocional de la comunidad futbolística en línea. A continuación, en la Tabla 4 se describe el diccionario de palabras utilizado para la clasificación de los sentimientos.

Tabla 4*Diccionario de palabras*

Clasificación	Palabras
Positivo	<ul style="list-style-type: none"> • Victoria • Éxito • Celebración • Admiración • Orgullo • Fantástico • Increíble • Maravilloso • Triunfo

Clasificación	Palabras	
Negativo	• Apoyo	
	• Derrota	
	• Fracaso	
	• Decepción	
	• Desastre	
	• Error	
	• Frustración	
	• Descontento	
	• Crítica	
	• Desilusión	
	• Dolor	
	Neutro	• Partido
		• Jugador
• Equipo		
• Estadio		
• Resultado		
• Temporada		
• Comentario		
• Información		
• Noticia		
• Entrenamiento		

Nota. Elaboración Propia (2023).

Así también, en la Tabla 5 se definen las frases que podrían encontrarse en las redes sociales relacionadas con los sentimientos expresados alrededor del fútbol.

Tabla 5

Diccionario de frases

Clasificación	Frase
Positivo	• Gran actuación del equipo
	• Increíble juego, chicos
	• Orgulloso de ser parte de esta victoria
	• Admirando el esfuerzo y la dedicación

Clasificación	Frase
	<ul style="list-style-type: none"> • Triunfo merecido, felicidades • Fantástico rendimiento en el campo • Maravillosa exhibición de habilidades • Estamos en una racha ganadora • Apoyo total al equipo, siempre • Emocionado por los éxitos por venir • Qué gran partido Estamos imparables • Hoy demostramos por qué somos los mejores • Felicidades al equipo por esa victoria increíble • Orgulloso de nuestros jugadores • Increíble desempeño, equipo • Este triunfo nos llena de alegría y esperanza • Celebrando cada gol. • Vamos por más • Maravilloso verlos jugar con tanta pasión • Éxito merecido • Se lo ganaron con esfuerzo • El apoyo de la afición hace la diferencia
Negativo	<ul style="list-style-type: none"> • Decepcionante derrota hoy • Frustrado con el resultado final • Descontento con el rendimiento • Errores costosos en el partido • Dolorosa derrota para todos • Desastre en el campo esta vez • Críticas justificadas a la actuación • Desilusionado por el resultado • Desafíos que necesitan superarse • Momento difícil para el equipo • Una derrota dolorosa. Vamos a levantarnos • Frustrante ver errores que podríamos evitar

Clasificación	Frase
Neutro	<ul style="list-style-type: none"> • Desilusión total. Esperábamos más del equipo • Error tras error. Necesitamos mejorar urgente • Descontento con el rendimiento de hoy • La derrota duele, pero aprenderemos de ello • Crítico con algunas decisiones del partido • Desastre en el campo. Necesitamos cambios • Dolorosa derrota. No es el resultado esperado • Desilusionados, pero seguiremos apoyando • Buena temporada hasta ahora • El equipo sigue entrenando fuerte • Información sobre el próximo partido • Noticias sobre cambios en la alineación • Comentario sobre la táctica utilizada • Datos sobre el rendimiento histórico • Publicación sobre el estadio y la afición • Actualización sobre el estado físico • Estamos enfocados en el siguiente juego • Entrenamiento regular esta semana • El equipo jugó un partido regular • El resultado fue dentro de lo esperado • Buena temporada en general • El rendimiento del jugador fue estándar • Un partido sin eventos destacados • El entrenamiento fue parte de la rutina • La noticia del día: empate en el partido • Comentario sobre el estado del equipo • Información sobre próximos encuentros • Noticias sobre la gestión del club

Nota. Elaboración Propia (2023).

4.4.2 Entrenamiento manual del modelo

El entrenamiento del modelo para el análisis de sentimientos es un proceso fundamental en el aprendizaje automático, donde se ajustan los parámetros del modelo utilizando un conjunto de datos etiquetado. Durante este proceso, el modelo "aprende" a reconocer patrones y relaciones entre las características del texto y las etiquetas de sentimiento asociadas, como positivo, negativo o neutro.

Figura 4

Interfaz para el entrenamiento manual del modelo

Clasificación de Sentimientos y Emociones, Equipos de Fútbol del Perú

Selecciona un Equipo de Fútbol: Atlético Grau

No.	Publicación	Fecha	Hora
1	Los albitos que van por la clasificación en la ciudad de Arequipa 🗣️ ¡Vamos Albitos! ❤️❤️ #PatrimonioDePiura #PiuraEsGrau	2023-10-30	15:41:03
2	¡Estamos con ustedes Albitos! 🗣️ ¡Con todo! ¡Vamos Grau! ❤️❤️ #PatrimonioDePiura #PiuraEsGrau	2023-10-30	14:23:11
3	Final del primer tiempo 🕒 ¡Vamos Grau! ❤️❤️ #PatrimonioDePiura #PiuraEsGrau	2023-10-29	16:58:16
4	Así salen nuestros albos al terreno de juego 🗣️ ¡Vamos Grau! ❤️❤️ #PatrimonioDePiura #PiuraEsGrau	2023-10-29	15:43:54
5	Ya estamos en Bernal 🏠 Listos 🗣️ ¡Vamos Grau! ❤️❤️ #PatrimonioDePiura #PiuraEsGrau	2023-10-29	14:56:58
6	🗣️ El Patrimonio de Piura y su último viaje a Bernal en la temporada 2 1 2 3	2023-10-29	13:03:00

No.	Comentario	Clasificación
1	Aseguraron su grati....	<input type="radio"/> Positivo <input checked="" type="radio"/> Neutro <input type="radio"/> Negativo
2	Que bueno que haya desendido un equipo de altura ya muchos	<input type="radio"/> Positivo <input checked="" type="radio"/> Neutro <input type="radio"/> Negativo
3	Oswald Ramirez pero subió el chankas	<input type="radio"/> Positivo <input checked="" type="radio"/> Neutro <input type="radio"/> Negativo
4	Anibal Seminario pero con binacional era uno más.	<input type="radio"/> Positivo <input checked="" type="radio"/> Neutro

Comenzar a Entrenar...

Nota. Elaboración Propia (2023).

A medida que el modelo se expone a más ejemplos de entrenamiento, se adapta y mejora su capacidad para predecir el sentimiento en textos no etiquetados, este proceso implica la optimización de hiperparámetros, como la tasa de aprendizaje, y la evaluación continua del rendimiento del modelo en un conjunto de datos de validación para garantizar que sea preciso y generalice bien a nuevos datos. Para lo cual, mediante el uso del Framework Flask se creó un micrositio web que permite el etiquetado manual de la polaridad (positiva, negativa o neutra) para cada comentario recopilados de las páginas de Facebook de los equipos de fútbol (ver Figura 4), cuyo código se encuentra en el Anexo 2.

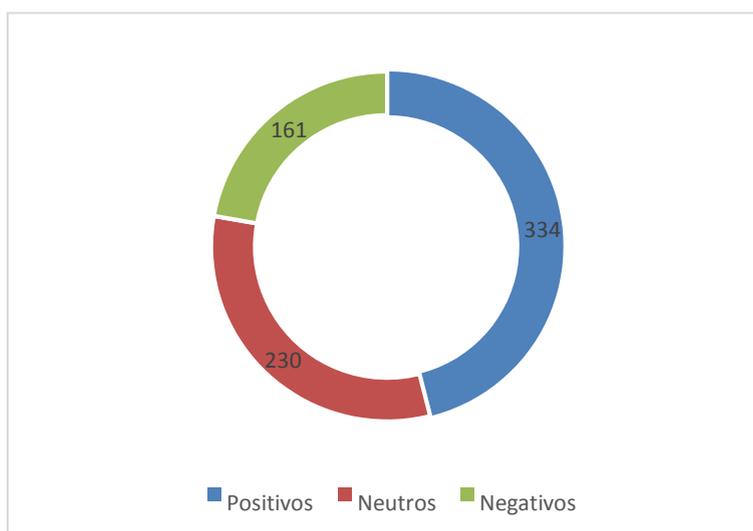
Para este entrenamiento se sumó al equipo de desarrollo a un voluntario, contando de esta manera con 3 personas encargadas de leer cada uno de los comentarios y

etiquetarlos manualmente utilizando la interfaz antes mencionada, misma que de una manera ágil y sencilla les permite a los voluntarios seleccionar el comentario a ser etiquetado.

Para contar con una muestra representativa de etiquetado manual de la polaridad de los sentimientos se decidió revisar el 25% de la información extraída; para lo cual, cada persona diariamente analizó 100 comentarios, lo significa que el proceso de etiquetado manual de 725 comentarios tomó aproximadamente 3 días, dado como resultados preliminares que el 46,07% de los comentarios generaron sentimientos positivos, un 22,20% generaron sentimientos negativos y el restante 31,72% sentimientos neutros,

Figura 5

Resultados del etiquetado manual



Nota. Elaboración Propia (2023).

4.4.3 Definición de reglas para los emojis

La expresión de texto mediante emojis es una práctica común para transmitir emociones, tono o contexto en todo tipo de plataformas digitales, en el análisis de sentimientos, los emojis a menudo son interpretados como indicadores adicionales de la emoción detrás del texto, en la Tabla 6 se muestra la polaridad de los principales emojis utilizados.

Tabla 6*Polaridad de los emoticones*

Emoji	Polaridad
:)	Positivo
(:	Positivo
;)	Positivo
:-)	Positivo
(-:	Positivo
:D	Positivo
: -D	Positivo
:P	Positivo
:-P	Positivo
:(Negativo
):	Negativo
;(Negativo
:-()-:	Negativo
D:	Negativo
D-:	Negativo
:´()-´:	Negativo

Nota. Elaboración Propia (2023).

Al considerar los emojis, se puede captar mejor la carga emocional de un comentario o publicación, lo que puede enriquecer el análisis de sentimientos al brindar una comprensión más completa de la intención emocional detrás del texto; para lo cual, en la Tabla 7 se definen las reglas a aplicar considerando que en un texto puede haber uno o varios emojis de distinta polaridad.

Tabla 7*Reglas para los emoticones*

Regla	Clasificación
Sólo positivos	Positivo
Más positivos que negativos	Positivo

Regla	Clasificación
Igual positivos que neutros	Neutro
Más negativos que positivos	Negativo
Sólo negativos	Negativo

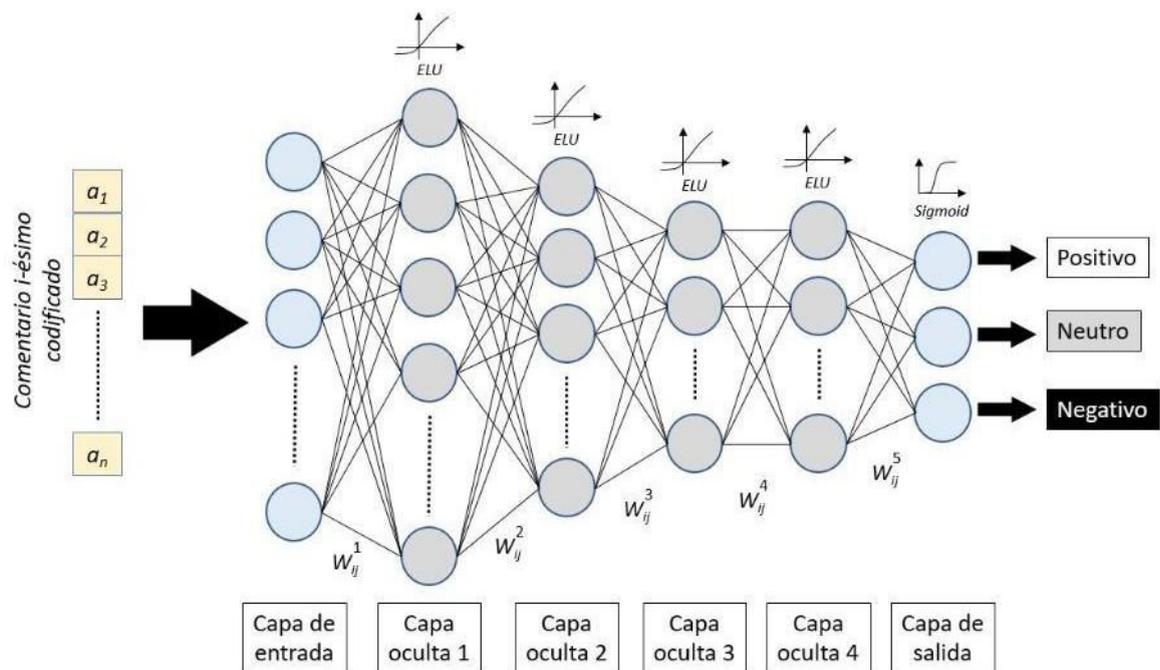
Nota. Elaboración Propia (2023).

4.4.4 Naïve Bayes

El clasificador Naïve Bayes es una herramienta efectiva para el análisis de sentimientos en redes sociales, este enfoque se basa en el teorema de Bayes, que calcula la probabilidad de que un mensaje pertenezca a una categoría específica dadas ciertas características observadas acorde a lo descrito en la Figura 6.

Figura 6

Modelo de Naïve Bayes



Nota. Elaboración Propia (2023).

En el contexto del análisis de sentimientos, estas características pueden incluir palabras clave, frases específicas o patrones lingüísticos asociados con emociones positivas, negativas o neutras. Aunque el modelo asume independencia entre las características (de ahí el término "naïve" o ingenuo), para el presente proyecto, Naïve Bayes ha demostrado ser eficaz al proporcionar resultados rápidos y

razonablemente precisos en la clasificación de sentimientos en grandes conjuntos de datos de redes sociales.

4.5 Análisis de los sentimientos

Finalmente, una vez que el modelo ha sido entrenado y que se evidencia un óptimo rendimiento, se procesa toda la información extraída, en la Tabla 8 se muestran los resultados obtenidos de la clasificación de la polaridad de sentimientos contenidos en 2.844 publicaciones en las redes sociales de los equipos de la Liga 1 peruana, dando como principal resultado que más de la mitad de los sentimientos expresados son positivos.

Tabla 8

Resultados de la polaridad de los sentimientos

Polaridad	Total	Porcentaje
Positivo	952	33,47%
Neutro	1.090	38,33%
Negativo	802	28,20%

Nota. Elaboración Propia (2023).

4.6 Matriz de confusión y métricas asociadas

La matriz de confusión se utiliza en el análisis de sentimientos en redes sociales para evaluar el rendimiento de un modelo de clasificación en la tarea de asignar sentimientos a publicaciones o textos. Esta matriz muestra los resultados de la clasificación comparando las predicciones del modelo con las categorías reales de sentimiento (positivo, negativo o neutro). Los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos se representan en esta matriz, lo que permite comprender cómo el modelo está acertando en sus predicciones y en qué áreas comete errores.

Tabla 9

Matriz de confusiones

Matriz de confusión		Negativos	Positivos	Precisión d/(b+d)
Real	Negativo	148 (a)	13 (b)	
	Positivo	16 (c)	318 (d)	96,07%

	Sensibilidad $d/(d+c)$	Especificidad $a/(a+b)$	Exactitud $(a+d)/(a+b+c+d)$
	95,21%	91,93%	94,14%

Nota. Elaboración Propia (2023).

- **Métrica de Exactitud:** Es una medida comúnmente utilizada para evaluar el rendimiento de un modelo de clasificación, el valor obtenido es de un 94,14%, es decir que el modelo está clasificando de manera acertada la mayoría de las publicaciones en sus respectivas categorías de sentimiento.
- **Métrica de Precisión:** La precisión indica la proporción de las predicciones positivas que son verdaderamente positivas en comparación con el total de predicciones clasificadas como positivas; el valor obtenido es del 96,07% por lo que se puede concluir que el modelo está clasificando de manera acertada las predicciones positivas.
- **Métrica de sensibilidad:** Esta métrica evalúa la proporción de predicciones positivas correctas sobre el total de observaciones que son verdaderamente positivas en los datos, el valor calculado es del 95,21%, es decir que el modelo identifica los sentimientos positivos expresados en los comentarios de Facebook con mucha precisión.
- **Métrica de Especificidad:** Representa la proporción de las predicciones negativas correctas (verdaderos negativos) sobre el número total de observaciones que en realidad son negativas en los datos, el valor calculado es del 91,93%, es decir que el algoritmo identifica las publicaciones con sentimientos negativos de una manera correcta.

4.7 Limitaciones del modelo

El modelo para el análisis de sentimientos enfrenta desafíos sustanciales asociados con el acceso restringido a información de Facebook mediante webscraping; para lo cual, hay que considerar las políticas de privacidad de la red social que prohíbe explícitamente el webscraping no autorizado, lo que puede resultar en restricciones de acceso o incluso en bloqueos de la cuenta que realiza el scraping, por lo que para acceder a la mayor cantidad de información fue necesario la creación de varias cuentas de correo que a medida que se extraía la información eran bloqueadas.

Así también, las restricciones de velocidad (rate limiting) puede ser considerada como un obstáculo, ya que Facebook controla la velocidad de las solicitudes para evitar un acceso excesivo o abusivo. Este límite impone desafíos en términos de recopilación eficiente de datos en tiempo real y puede afectar la capacidad del modelo para proporcionar análisis.

V. DISCUSIÓN

Los resultados del análisis de sentimientos en las redes sociales de los equipos de fútbol peruanos revelan una compleja interacción de emociones entre los seguidores. Si bien el modelo demostró una capacidad considerable para identificar y clasificar sentimientos, es esencial reconocer las limitaciones inherentes, como posibles sesgos lingüísticos y la interpretación variable de ciertos términos. Se observó una fuerte polarización de opiniones, con expresiones apasionadas tanto positivas como negativas hacia los equipos, los eventos deportivos clave y cambios en la composición de los equipos influyeron significativamente en las tendencias emocionales.

El proceso de extracción de la información mediante webscraping presentó limitaciones en cuanto al número de publicaciones y comentarios extraídos, debido a las políticas de seguridad de la información de Facebook lo que hace que el modelo funcione correctamente para la extracción de información generada no más allá de dos días atrás. Posterior a la recopilación y limpieza de la información, para su análisis fue necesario evaluar los diferentes modelos existentes hasta llegar a determinar que el clasificador de Naïve Bayes presenta las mejores características para garantizar los resultados deseados; para lo cual, a continuación, se describen las características principales en las cuales se basaron nuestras decisiones:

- Su simplicidad y eficiencia computacional permiten un procesamiento rápido de grandes volúmenes de datos, lo que es esencial en el dinámico entorno de las redes sociales.
- La capacidad para manejar características categóricas y, en el contexto del análisis de sentimientos, puede adaptarse fácilmente a la naturaleza discreta de las expresiones emocionales en los textos.
- La transparencia del modelo ofreciendo una interpretación clara de los resultados al proporcionar probabilidades asociadas a las clasificaciones, esta transparencia es valiosa en entornos donde comprender la lógica detrás de las decisiones del modelo es esencial.
- Escalabilidad, debido a su simplicidad y eficiencia, Naïve Bayes es escalable a conjuntos de datos grandes y puede adaptarse bien a situaciones en tiempo real, como el monitoreo continuo de las redes sociales.

- Su buen desempeño en texto, su capacidad para capturar patrones en datos de texto lo hace especialmente útil en aplicaciones relacionadas con el procesamiento del lenguaje natural.

En el análisis de sentimientos en redes sociales, donde la interpretación precisa de las emociones es crucial, estos atributos hacen que Naïve Bayes sea una opción pragmática y efectiva; sin embargo, existen otras investigaciones relacionadas con el análisis de sentimientos en las que el uso del modelo de máquinas de vectores de soporte (SVM) les brindaron los mejores resultados para la clasificación de sentimientos sobre todo en el ámbito de elecciones presidenciales en distintos países de Latinoamérica (Valladares, 2022)

Adicionalmente es importante mencionar ciertos sesgos y limitaciones inherentes al modelo, siendo el principal el sesgo lingüístico, donde ciertos términos pueden tener connotaciones distintas en diferentes regiones del país, podría influir en la precisión del análisis; además, la incapacidad del modelo para comprender plenamente el sarcasmo o expresiones coloquiales puede afectar la interpretación precisa de las emociones. La evolución constante de las tendencias lingüísticas y la introducción de nuevos modismos podrían desafiar la capacidad del modelo para mantenerse actualizado. Además, el modelo puede no captar adecuadamente las complejidades emocionales que se manifiestan de manera sutil en el contexto del fútbol. Estos sesgos también los identificó Maldonado (2022) en su análisis de sentimientos en la red social Twitter llegando a destacar la importancia de una interpretación crítica de los resultados y la consideración de factores culturales y lingüísticos específicos en el análisis de sentimientos en este contexto particular. Para esto, la interacción entre humanos y máquinas en el análisis de sentimientos en redes sociales es fundamental para aprovechar lo mejor de ambos mundos, por un lado, las máquinas, mediante algoritmos de inteligencia artificial, pueden procesar enormes cantidades de datos a una velocidad impresionante, identificando patrones sutiles y correlaciones; sin embargo, la interpretación de las emociones humanas, sarcasmo, y matices culturales a menudo escapa a la pura lógica algorítmica, es aquí es donde la intervención humana se vuelve crítica aportando con su comprensión contextual, empatía y discernimiento, refinando y ajustando los modelos de inteligencia artificial para que sean más precisos y culturalmente sensibles.

La simbiosis entre la capacidad analítica de las máquinas y la intuición humana no solo mejora la exactitud del análisis de sentimientos, sino que también asegura que la interpretación sea más completa y alineada con las complejidades del comportamiento humano en línea. Además, la interacción humano-máquina permite un ciclo de retroalimentación continuo, a medida que los humanos interpretan y aplican su comprensión contextual a los resultados proporcionados por la inteligencia artificial, estas interpretaciones alimentan el proceso de aprendizaje de la máquina, este ciclo iterativo de mejora continua permite que los modelos se adapten a cambios en el lenguaje y las expresiones emocionales en constante evolución en las redes sociales. Esta interacción también facilita la interpretación de matices emocionales que podrían no ser evidentes para un modelo puramente automatizado, garantizando que el análisis de sentimientos sea más preciso y reflexivo en la interpretación de la complejidad inherente a las interacciones humanas en línea.

La precisión del modelo para la identificación y clasificación de sentimientos es notable, reflejando su capacidad para discernir eficazmente entre expresiones positivas, negativas y neutras. Sin embargo, es esencial contextualizar que alcanzar una precisión del 100% puede ser impracticable y, en algunos casos, indeseable. esta precisión considerando la naturaleza subjetiva del lenguaje y la diversidad de expresiones emocionales.

Por lo que es crucial examinar también las métricas relacionadas, como la sensibilidad y la especificidad, para obtener una visión más completa de cómo el modelo se desempeña en situaciones específicas, como identificar correctamente tanto los comentarios positivos como los negativos. Un equilibrio adecuado entre estas métricas revelará la capacidad del modelo para capturar la diversidad de expresiones emocionales en las redes sociales deportivas peruanas. En última instancia, la discusión sobre la precisión del modelo debe considerar la complejidad intrínseca del análisis de sentimientos y buscar un equilibrio entre la capacidad predictiva y la adaptabilidad a la riqueza y variabilidad del lenguaje emocional en línea.

La estabilidad temporal del modelo para el análisis de sentimientos es un factor crucial que merece atención ya que demuestra la robustez para mantener un rendimiento consistente a lo largo del tiempo, proporcionando confianza en su

capacidad para adaptarse a las variaciones del discurso en línea, que puede ser influenciado por factores como eventos deportivos, cambios en el rendimiento del equipo o eventos externos. Esta estabilidad también asegura que el modelo sea resistente a posibles sesgos temporales y pueda proporcionar una evaluación coherente de la percepción emocional de los usuarios a lo largo de diferentes períodos. Por lo que es necesario constantemente analizar esta estabilidad temporal con el objetivo de garantizar la validez del modelo a lo largo del tiempo y realizar los ajustes de ser necesario para adaptarse a los constantes cambios de la comunicación virtual.

Así también, el análisis a profundidad de los falsos positivos y negativos revela aspectos cruciales para perfeccionar el modelo, la identificación de falsos positivos sugiere la posibilidad de que el modelo interprete incorrectamente expresiones neutras como positivas o negativas, mientras que los falsos negativos indican ocasiones en las que el modelo podría pasar por alto emociones expresadas. Para mejorar, es esencial examinar estas instancias en detalle, identificar patrones específicos de lenguaje o contextos que causen interpretaciones erróneas y ajustar el modelo en consecuencia. Para lo cual, podrían implementarse estrategias más avanzadas de procesamiento del lenguaje natural y algoritmos de aprendizaje automático que se adapten a las complejidades del discurso futbolístico y consideren la evolución del lenguaje en las redes sociales. La retroalimentación continua y la revisión manual de casos difíciles pueden contribuir a la mejora progresiva del modelo, asegurando una interpretación más precisa y contextual de las emociones expresadas por los seguidores en línea.

Finalmente, una de las aplicaciones prácticas del modelo para el análisis de sentimientos en las redes sociales puede ser instrumento de mucha importancia para la gestión de la reputación de los equipos y la toma de decisiones estratégicas. La capacidad de identificar patrones emocionales permite a los equipos adaptar sus estrategias de compromiso, responder proactivamente a situaciones delicadas y comprender las preferencias y expectativas de la base de seguidores. Además, esta herramienta puede ser esencial para evaluar la efectividad de las campañas de marketing y patrocinio, brindando información valiosa sobre la percepción pública y el impacto emocional de las iniciativas.

VI. CONCLUSIONES

La creación de un modelo de aprendizaje automático para analizar y clasificar los sentimientos y emociones expresados en los textos de publicaciones y comentarios en redes sociales representa un avance significativo en la comprensión de la dinámica emocional en entornos digitales, este modelo está diseñado para procesar grandes volúmenes de datos de manera eficiente, identificando patrones y tendencias emocionales en tiempo real. Así también, este trabajo representa un paso hacia la mejora de la comunicación en línea, la gestión de la reputación digital y la comprensión profunda de las experiencias y opiniones de la comunidad en las redes sociales.

La tarea de recopilar y preparar un conjunto de datos etiquetados que represente una amplia variedad de sentimientos y emociones en diversos contextos de las redes sociales es esencial para el desarrollo exitoso de modelos de análisis emocional. La calidad y diversidad de los datos desempeñan un papel crucial en la capacidad del modelo para generalizar y adaptarse a la complejidad del discurso en línea, este proceso no solo implica la recopilación de una cantidad significativa de datos, sino también la meticulosa anotación que refleje la riqueza y matices de las expresiones emocionales. La representación precisa de diferentes contextos y matices lingüísticos en el conjunto de datos preparado no solo mejora la robustez del modelo, sino que también garantiza una comprensión más completa y precisa de las emociones expresadas en el variado paisaje de las redes sociales.

Finalmente, la identificación de patrones y tendencias en la expresión de sentimientos y emociones en las redes sociales, a través del análisis de los resultados obtenidos mediante modelos de inteligencia artificial, revela un valioso panorama sobre la dinámica emocional de las comunidades en línea. La observación de estos patrones permite entender mejor cómo las emociones fluctúan en respuesta a eventos específicos, tendencias culturales o cambios en la narrativa pública. La capacidad de discernir estos patrones no solo proporciona clasificaciones inmediatas, sino que también

permite una preparación proactiva para futuros escenarios. La capacidad de identificar y comprender estos patrones establece las bases para una comprensión más profunda y contextualizada de la expresión emocional en las redes sociales.

La evaluación y mejora continua del rendimiento de un modelo de inteligencia artificial, mediante la aplicación de diversas métricas de desempeño, es un aspecto crítico en el desarrollo de sistemas efectivos de análisis de sentimientos. La aplicación rigurosa de métricas como precisión, sensibilidad, especificidad y exactitud proporciona una visión holística de la capacidad del modelo para clasificar emociones de manera precisa con resultados superiores al 90%. La interpretación cuidadosa de estas métricas no solo permite identificar áreas de mejora, como la gestión de falsos positivos o falsos negativos, sino que también facilita ajustes y optimizaciones específicas, este enfoque iterativo garantiza que el modelo se adapte a las complejidades cambiantes del lenguaje y las dinámicas emocionales en línea. La mejora constante del rendimiento es esencial para mantener la relevancia y eficacia del modelo en un entorno digital dinámico y en constante evolución.

VII. RECOMENDACIONES

Se recomienda implementar una segmentación temática más refinada al analizar los sentimientos al dividir las conversaciones en temas específicos, como partidos, jugadores, gestión técnica y eventos, proporcionará una comprensión más detallada de las emociones asociadas a aspectos particulares, facilitando acciones específicas para mejorar la relación con los fanáticos.

Para garantizar la calidad y representatividad del análisis de sentimientos, se sugiere implementar mecanismos de etiquetado cuidadosos y detallados que consideren la variabilidad de tonos emocionales, matices lingüísticos y contextos específicos asociados a las interacciones relacionadas con los equipos de fútbol peruanos. La calidad del conjunto de datos no solo impactará directamente en la precisión del modelo de análisis de sentimientos, sino que también permitirá una comprensión más profunda de las dinámicas emocionales en la comunidad de seguidores, contribuyendo así a estrategias más efectivas de gestión de la reputación en línea.

Se aconseja implementar un ciclo continuo de evaluación y mejora del rendimiento del modelo de inteligencia artificial, haciendo hincapié en la aplicación sistemática de diversas métricas de desempeño. Este enfoque implica monitorear regularmente el rendimiento del modelo mediante métricas como la precisión, sensibilidad, especificidad y el valor F1. Además, es crucial considerar métricas específicas relacionadas con el contexto del análisis de sentimientos en redes sociales, como la capacidad para identificar matices emocionales y la adaptación a cambios en el lenguaje y las tendencias en línea. A través de este proceso iterativo de evaluación y ajuste, se logrará una mejora continua en la precisión y relevancia del modelo, asegurando su capacidad para capturar de manera efectiva la complejidad de las expresiones emocionales en el contexto específico de los equipos de fútbol peruanos en redes sociales.

Se sugiere dedicar esfuerzos a una investigación más profunda para identificar patrones y tendencias específicas en la expresión de sentimientos y emociones, este enfoque implica el uso de herramientas de visualización

de datos avanzadas y técnicas de análisis estadístico para explorar las dinámicas emocionales en profundidad. Identificar patrones emergentes y tendencias a lo largo del tiempo o en respuesta a eventos específicos puede proporcionar información valiosa sobre la evolución de las actitudes y emociones de la comunidad de seguidores de los equipos de fútbol peruanos.

REFERENCIAS

- Aguirre, P., Anaya, M., Laurencio, R., & Casco, J. (2013). INVESTIGACIÓN APLICADA E INTERDISCIPLINARIEDAD EN LAS CIENCIAS DE LA COMUNICACIÓN. *Prisma Social*, 11, 294-320. <https://www.redalyc.org/articulo.oa?id=353744535009>
- Aparicio, L. E., Amaya, O. J., & Devia, P. A. (2023). Aplicación de Deep Learning para la identificación de defectos superficiales utilizados en control de calidad de manufactura y producción industrial: una revisión de la literatura. *Ingeniería*, 28(1). <https://doi.org/https://doi.org/10.14483/23448393.18934>
- Arango, C. A., & Osorio, C. F. (2021). Aislamiento social obligatorio: un análisis de sentimientos mediante machine learning. *Suma de Negocios*, 12(26), 1-13. <https://doi.org/10.14349/sumneg/2021.V12.N26.A1>
- Atamari, J., Flores, C., Mamani, J., & Rondon, S. (2022). Uso de árboles de decisión para detectar si una habitación está ocupada usando Python. *Innovación y Software*, 3(1), 58-66. <https://www.redalyc.org/journal/6738/673870840005/673870840005.pdf>
- Attachmedia. (2023). *INTELIGENCIA ARTIFICIAL: QUÉ ES, PARA QUÉ SIRVE Y SUS APLICACIONES*. Sitio web: <https://attachmedia.com/blog/inteligencia-artificial-aplicaciones/>
- Barrios, H., Díaz, V., & Guerra, Y. (2020). Subjetividades e inteligencia artificial: desafíos para 'lo humano'. *Veritas*(47), 81-107. <https://www.redalyc.org/jatsRepo/2911/291166073004/291166073004.pdf>
- Brito, C. D., & Brito, C. F. (2018). Un modelo nuevo variacional para la clasificación binaria en el contexto de aprendizaje automático supervisado. *Ingeniería*, 22(2), 1-8. <https://www.redalyc.org/journal/467/46758579001/46758579001.pdf>
- Contreas, L. E., Nieves, N., & González, K. (2023). Prediction of University-Level Academic Performance through Machine Learning Mechanisms and Supervised Methods. *Ingeniería*, 28(1). <https://doi.org/https://doi.org/10.14483/23448393.19514>
- Darad, S., & Krishnan, S. (2023). Análisis de sentimiento de los datos de twitter de COVID-19 utilizando modelos de aprendizaje profundo y aprendizaje de

- máquina. *Ingenius. Revista de Ciencia y Tecnología*(29).
<https://doi.org/https://doi.org/10.17163/ings.n29.2023.10>
- Gozálvez, V., & Cortijo, G. (2023). Desarrollo humano y redes sociales en sociedades digitales. *Sophia, Colección de Filosofía de la Educación*(34), 41-64. <https://doi.org/https://doi.org/10.17163/soph.n34.2023.01>
- Gozálvez, V., & Cortijo, G. (2023). Desarrollo humano y redes sociales en sociedades digitales. *Sophia, Colección de Filosofía de la Educación*(34), 41-64. <https://doi.org/https://doi.org/10.17163/soph.n34.2023.01>
- Leal, M. S. (2023). INTELIGENCIA EMOCIONAL Y COMPROMISO LABORAL. *Ciencias Administrativas*(21).
<https://doi.org/https://doi.org/10.24215/23143738e116>
- López, J. C. (2021). Desenmascarando datos: Igualdad e Inteligencia Artificial. *Revista IUS*, 15(48), 137-156.
<https://doi.org/https://doi.org/10.35487/rius.v15i48.2021.740>
- Maldonado, E. (2022). ANÁLISIS DE SENTIMIENTOS EN LA RED SOCIAL TWITTER MEDIANTE EL PROCESAMIENTO DE LENGUAJE NATURAL. Sitio web:
<http://dspace.unach.edu.ec/bitstream/51000/10103/1/Maldonado%20R.%2c%20Erik%20S.%20%282022%29%20Análisis%20de%20sentimientos%20en%20la%20red%20social%20Twitter%20mediante%20el%20procesamiento%20del%20lenguaje%20natural.pdf>
- Mazuera, R., Albornoz, N., & Briceño, R. (2021). Seguridad alimentaria, Covid-19 y crisis: una aplicación del modelo de regresión logística binomial para su estudio. *Revista mexicana de ciencias políticas y sociales, LXVI*(242), 443-473.
<https://doi.org/https://doi.org/10.22201/fcpys.2448492xe.2021.242.78085>
- More, R., Quito, C., & Torres, L. (2020). Decima Novena Conferencia Iberoamericana en Sistemas, Cibernética e Informática. *Análisis De Texto Por Procesamiento Del Lenguaje Natural (PNL) Con Software R Desde Opiniones Sobre 'Feminicidio' En Diarios Del Perú*. Callaos N.C., Horne J., Sanchez B., Tremante A.
<https://www.scopus.com/record/display.uri?eid=2-s2.0-85096641630&origin=inward&txGid=7ab00925eeb5e403d0e06e56ec3ce21>

- Murillo, J. F., & Jiménez, M. R. (2019). LA AUTOGESTIÓN DE EMOCIONES Y SENTIMIENTOS PARA “SER BUENAS PERSONAS”. *International Journal of Developmental and Educational Psychology*, 3(1), 281-294. <https://www.redalyc.org/articulo.oa?id=349853365029>
- Peña, J. A., Giraldo, S., Arango, C. A., & Bucheli, V. A. (2019). Un chatbot para asistir a las necesidades de información en tiempos de COVID-19. *Ingeniería y competitividad*, 24(1). <https://www.redalyc.org/journal/2913/291371829011/291371829011.pdf>
- Rico, A., Gaytán, N. D., & Sánchez, D. (2019). Construcción e implementación de un modelo para predecir el rendimiento académico de estudiantes universitarios mediante el algoritmo Naïve Bayes. *Diálogos sobre educación*, 10(19). <https://doi.org/10.32870/dse.v0i19.509>
- Rodríguez, M., Guzmán, R., & Huamán, M. (2021). APLICACIÓN DE UN DISEÑO EXPERIMENTAL EN LA SÍNTESIS. *Revista Boliviana*, 38(4), 133-147. <https://www.redalyc.org/journal/4263/426369860001/426369860001.pdf>
- Rosso, A. E., Montilla, Y. M., & Garzón, S. C. (2022). Metodología para obtención y análisis de datos. *Ingeniería*, 27(3). <https://doi.org/https://doi.org/10.14483/23448393.17952>
- Valladares, J. (2022). ANÁLISIS DE SENTIMIENTOS PARA TEXTOS CORTOS EN ESPAÑOL, UNA REVISIÓN DEL ESTADO DEL ARTE. Sitio web: <https://dspace.ups.edu.ec/bitstream/123456789/22180/1/UPS%20-%20TTS666.pdf>
- Wolde, B. G., & Boltana, A. S. (2021). REST API composition for effectively testing the Cloud. *Journal of applied research and technology*, 19(6), 676-693. <https://doi.org/https://doi.org/10.14482/INDES.30.1.303.661>
- X3 Media. (2023). ¿Cuáles son las redes sociales más populares en Colombia? Sitio web: <https://inbound-marketing.xtresmedia.com/blog/cuales-son-las-redes-sociales-mas-populares-en-colombia>
- Zapata, J. A. (2020). Editorial. Inteligencia artificial para la toma de decisiones. *Revista Perspectiva Empresarial*, 7(2), 3-5. <https://www.redalyc.org/articulo.oa?id=672271538001>

ANEXOS

Anexo I – Código de Web Scraping

'''Para ejecutar este código solo debe instalar las dependencias.

Abra la línea de comandos de windows (CMD) y ejecute una a una los siguientes comandos.

```
pip install facebook-scraper'''
```

```
#import facebook_scraper as fs
```

```
from facebook_page_scraper import Facebook_scraper
```

```
import json
```

```
import sys
```

```
import time
```

```
import requests
```

```
import os
```

```
# Aquí debe especificarse el correo electrónico y contraseña de usuario de  
facebook, para así tener acceso a
```

```
# los mensajes públicos de las páginas de los equipos de fútbol de Perú
```

```
USER = 'alexander.pascau@gmail.com'
```

```
PASSWORD = '!!Juan11'
```

```
page_list = ['adtdtarma']
```

```
'''
```

```
, 'camannucci', 'ClubAlianzaLima', 'ClubCiencianoOficial', 'ClubDeportivoGarcilasoOfi  
cial',
```

```
'ClubDeportivoMunicipal', 'ClubSportBoys', 'clubUTC.pe', 'CSCristal', 'DeportivoBinaci  
onalFC', 'FBCMelgar',
```

```
'sporthuancayopaginaoficial', 'ucvclubdefutbol', 'UnionComercioOficial', 'Universitario  
'
```

```
'''
```

```

proxy_port = 10001
posts_count = 100
browser = "firefox" #Chrome
timeout = 600
# Moving on to the headless browser variable. Type false as a boolean if you want
to see the scraper in action.
# Otherwise, write true and run the code in the background.
headless = False

if not sys.warnoptions:
    import warnings
    warnings.simplefilter("ignore")

# Estas son las p'aginas de los equipos de futbol
PAGES = [
    #"https://www.facebook.com/atleticograu",
    "https://www.facebook.com/UnionComercioOficial",
    "https://www.facebook.com/AcademiaCantolao",
    "https://www.facebook.com/CuscoFCoficial",
    "https://www.facebook.com/ClubSportBoys",
    "https://www.facebook.com/camannucci",
    "https://www.facebook.com/DeportivoBinacionalFC",
    "https://www.facebook.com/ClubCiencianoOficial",
    "https://www.facebook.com/ucvclubdefutbol",
    "https://www.facebook.com/ClubDeportivoGarcilasoOficial",
    "https://www.facebook.com/sporthuancayopaginaoficial",
    "https://www.facebook.com/ClubDeportivoMunicipal",
    "https://www.facebook.com/FBCMelgar",
    "https://www.facebook.com/CSCristal",
    "https://www.facebook.com/Universitario",
    "https://www.facebook.com/ClubAlianzaLima",
    "https://www.facebook.com/CuscoFCoficial",
    "https://www.facebook.com/adtdtarma",

```

```

    "https://www.facebook.com/clubUTC.pe",
]
class FacebookCommentsScraper():

    def __init__(self, facebookUser=None, facebookPassword=None):
        """
        facebookUser: Email del usuario de Facebook que se utiliza para hacer el
        scraping.
        facebookPassword: Password del usuario de Facebook que se utiliza para
        hacer el scraping.
        """
        self.user = facebookUser
        self.password = facebookPassword
        self.credentials = None if not self.user or not self.password else (self.user,
        self.password)

    def get_post_data(self, post, pageUniqueNameOrId):
        global full_path_older
        """
        Hace scraping a un post de Facebook y devuelve sus datos.

        post: URL o ID de la publicación de la que se deben obtener los datos.
        return: Devuelve un objeto con los datos del post, incluyendo los comentarios
        y estadísticas.
        """
        try:
            # Hace el scraping de la publicación en Facebook.
            gen = fs.get_posts(
                post_urls=[post], # list, URLs or post IDs to extract posts from. Alternative
                to fetching based on username
                extra_info=True, # if true the function will try to do an extra request to get
                the post reactions. Default is False
                options={"comments": True, # to extract comments

```

```

        "reactors": True, # to extract the people reacting to the post
        "progress": True, # to get a tqdm progress bar while extracting
comments and replies
        "allow_extra_requests" : False, # to disable making extra requests
when extracting post data
                                # (required for some things like full text and
image links)
        "posts_per_page": 200 # to request x posts per page. The default is
4
    },
    timeout=60,
    # pages=5, # how many pages of posts to request, the first 2 pages may
have no results, so try with a
    # number greater than 2. Default is 10
    credentials=self.credentials
)
postData = next(gen) # Los datos del post estan en el primer elemento del
generador.

# Extrae de los datos del post, los textos de los comentarios de los usuarios.
texts = []
#comments_full = fs.utils.safe_consume(postData['comments_full'])
# Preguntamos primero si se obtuvieron los comentarios en comments_full,
sino entonces se procede a obtener los comentarios
# guardados en el mismo archivo pero de la carpeta vieja (old_folder)
if 'comments_full' in postData.keys():
    for comment in postData['comments_full']:
        texts.append(comment['comment_text'])
        for reply in comment['replies']:
            texts.append(reply['comment_text'])
else:
    # leemos entonces el archivo .json respectivo de la carpeta old_folder, y
leemos sus estadisticos,

```

```

# incluido los comentarios...
archivo_json = pageUniqueNameOrId + '_' + post + '.json'
ruta_archivo_json = os.path.join(full_path_older, archivo_json)
if os.path.exists(ruta_archivo_json):
    with open(ruta_archivo_json) as archivo_datos:
        data = json.load(archivo_datos)
        # Ahora completamos la informacion que falta...
        postData.update({'likes' : data['user_reactions']['statistics']['likes']})
        postData.update({'comments' :
data['user_reactions']['statistics']['comments']})
        postData.update({'shares' : data['user_reactions']['statistics']['shares']})
        postData.update({'reaction_count' :
data['user_reactions']['statistics']['reaction']})
        postData.update({'reactions' : data['user_reactions']['reactions']})
        texts = data['user_reactions']['texts']
    else:
        postData.update({'likes' : 0})
        postData.update({'comments' : 0})
        postData.update({'shares' : 0})
        postData.update({'reaction_count' : 0})
        postData.update({'reactions' : {"like": 0, "love": 0, "haha": 0, "wow": 0,
"care": 0}})
        texts = ""
    # Devuelve un objeto que contine los textos de comentarios extraidos y
también
    # contiene estadísticas de la cantidad de interacciones y sus tipos.
    return {
        'statistics': {
            'likes': postData['likes'],
            'comments': postData['comments'],
            'shares': postData['shares'],
            'reaction': postData['reaction_count']
        },

```

```

        'reactions': postData['reactions'],
        'texts': texts
    }
except requests.exceptions.RequestException as req_error:
    print('Error de solicitud a Facebook:', req_error)
except Exception as e:
    print('Error no especificado:', str(e))

def to_file(self, fileName, data, operation):
    """Agrega una nueva línea (data) de datos al archivo (fileName)."""
    try:
        with open(fileName, operation, encoding='utf-8') as fileOut:
            fileOut.write(str(data) + '\n')
    except Exception as e:
        print('Error: Guardando en fichero!')
        print('Exception: ' + str(e))

def get_posts_on_page(self, pageUniqueId, verbose=False):
    global full_path_base
    """
    Obtiene publicaciones de una página de Facebook

    pageUniqueId: Nombre único que sale en la URL de la página o el
    identificador.
    verbose: Poner a True para ver mensajes en la línea de comandos.
    return: Devuelve una lista de objetos donde cada objeto tiene los datos de una
    publicación.

    Los datos de cada publicación incluyen estadísticas y evaluación del
    sentimiento
    presente en los comentarios de los usuarios y sus reacciones.
    """
    first = True

```

```

try:
    for post in fs.get_posts(pageUniqueNameOrId, pages=10,
credentials=self.credentials):
        try:
            postData = {
                'username': post['username'],
                'user_id': post['user_id'],
                'page_id': post['page_id'],
                'post_id': post['post_id'],
                'timestamp': post['timestamp'],
                'time': post['time'].isoformat(),
                'post_url': post['post_url'],
                'post_text' : post['post_text']
            }
            if verbose:

print('_____')
                for key in list(postData.keys()):
                    print(f'{key}:', postData[key])
                postData['user_reactions'] = self.get_post_data(post['post_id'],
pageUniqueNameOrId)
                postData['sentiment_manual'] = 0 # por defecto, todos son valorados
como NEUTRO
                #Crear la carpeta
                # Crea un fichero JSON con los datos.
                fileName = os.path.join(full_path_base,
"{}_{}.json".format(pageUniqueNameOrId, postData['post_id']))
                self.to_file(fileName, json.dumps(postData), 'w')

                # Crea un fichero CSV con los comentarios.
                for comment in postData['user_reactions']['texts']:
                    line = "{}","{}","{}","{}".format(0.5, pageUniqueNameOrId,
postData['post_id'], comment)

```

```

        fileName = os.path.join(full_path_base,
"{_all_comments.csv".format(pageUniqueId))
        if first:
            first = False
            if os.path.exists(fileName):
                os.remove(fileName)
            self.to_file(fileName, line, 'a')
        except Exception as e:
            if verbose:
                print('Error! Obteniendo una publicación de la página
"{_all_comments.csv".format(pageUniqueId))
                print('Exception:', str(e))
            except Exception as e:
                print('Error: Obteniendo los post de la pagina
"{_all_comments.csv".format(pageUniqueId))
                print('Exception:', str(e))

```

INICIO DE LA EJECUCION...

carpeta base, al mismo nivel de la app .py, donde se archivarian las publicaciones
y

comentarios de cada equipo de fútbol peruano

base_folder = 'posts'

Se tiene otra carpeta temporal, que contiene informacion que quizas no se
encuentre al buscar comentarios y a su vez el texto

de la publicacion, todo combinarse y colocarse en la carpeta base

old_folder = 'data'

Crea el objeto para probar el funcionamiento del algoritmo.

scraper = FacebookCommentsScrapper(USER, PASSWORD)

PAGE_SELECTOR = 0

full_path_base = "

full_path_older = "

#for page in page_list:

```

# proxy = f'alexande:password@us.smartproxy.com:{proxy_port}'
scraper = Facebook_scraper(page_list, posts_count, browser, timeout=timeout,
headless=headless)
json_data = scraper.scrap_to_json()

while PAGE_SELECTOR < len(PAGES):
    #Crear la carpeta ./posts/<nombre equipo>
    if not os.path.exists(base_folder):
        os.mkdir(os.path.join(os.getcwd(),base_folder)) # creo la carpeta base
    pageUniqueNameOrId = PAGES[PAGE_SELECTOR].split('.')[1]
    full_path_base = os.path.join(os.getcwd(),base_folder, pageUniqueNameOrId)
    full_path_older = os.path.join(os.getcwd(),old_folder, pageUniqueNameOrId)
    if not os.path.exists(full_path_base):
        # creo la carpeta <nombre equipo>
        os.mkdir(full_path_base)
    scraper.get_posts_on_page(pageUniqueNameOrId, True)
    PAGE_SELECTOR += 1

```

Anexo II – Código para la clasificación manual de los sentimientos

```
from flask import render_template, Flask, request, url_for
#Declaracion de librerias
import os
from transformers import BertTokenizer, BertForSequenceClassification
import torch
from torch.utils.data import DataLoader, TensorDataset
from transformers import AdamW
import json
import html
import datetime
from markupsafe import Markup

#Instanciamos con la clase Flask, usando directorios templates y static/...
app = Flask(__name__, template_folder='templates', static_folder='static')
#clave secreta
app.secret_key = 'MarcoAntonioJuan2023'

# Definir etiquetas para sentimientos
labels = ["positiva nivel máximo", "positiva nivel intermedio", "positiva nivel mínimo",
"negativa nivel máximo", "negativa nivel intermedio", "negativa nivel mínimo"]

# Valores de chequeo de Radio button para la Polaridad seleccionada
status_rb_Polarity = {
    1 : { # Polaridad Positiva
        'CheckPos' : 'checked',
        'CheckNeu' : "",
        'CheckNeg' : ""
    },
    0.5 : { # Polaridad Neutra
        'CheckPos' : "",
        'CheckNeu' : 'checked',
```

```

        'CheckNeg' : "
    },
    -1 : { # Polaridad Negativa
        'CheckPos' : "",
        'CheckNeu' : "",
        'CheckNeg' : 'checked'
    },
    # Sin asignacion de Polaridad (esto es importante a la hora que se desee
    entrenar el modelo y solo se escojan
    # los comentarios ya clasificados, y los que no puedan entonces automaticamene
    ser pronosticados por el modelo)
    0 : {
        'CheckPos' : "",
        'CheckNeu' : "",
        'CheckNeg' : "
    }
}

```

Directorio base donde se encuentran los archivos .json de cada publicacion y sus comentarios, para cada equipo de futbol peruano

data_folder = 'posts'

```

@app.route("/")
def init():
    listado_publicaciones = ""
    '<tr><td>&nbsp;</td><td>&nbsp;</td><td>&nbsp;</td><td>&nbsp;</td></tr>'
    listado_comentarios = '<tr>'
    listado_comentarios += '<td valign="top"><div align="center"></div></td>'
    listado_comentarios += '<td valign="top"><div align="left"></div></td>'
    listado_comentarios += '<td valign="top"><div align="center"></div></td>'
    listado_comentarios += '</tr>'
    equipo_sel = 'None'
    return render_template("clasificacion.html",

```

```

listado_publicaciones=Markup(listado_publicaciones),
    listado_comentarios=Markup(listado_comentarios), equipo_sel=equipo_sel)

# pub_comments es un diccionario que tiene como llaves iniciales al nombre
# generico de cada equipo, luego cada publicacion y luego
# cada comentario, dentro de éstos estaria la Pluralidad de referencia
#
# pub_comments = {
#     |
#     "atleticograu": [
#         |
#         - { "id" : 1
#             "post_id" : "1050761329712097",
#             "post_text" : "\u00a1Empieza la Liguilla para nuestros...",
#             "Fecha" : "2023-11-03", "Hora" : "13:51:51",
#             "comments" : [
#                 { "id" : 1,
#                   "text" : "A campeonar albitos para darle esos 02
# puntos al principal que lo necesita",
#                   "Polarity" : 0 (sin asignar),
#                   "Polarity_Predict" : 1 (polaridad dada por el modulo
# IA)
#                 },
#                 { "id" : 2,
#                   "text" : "Disculpen saben si lo van a transmitir",
#                   "Polarity" : 0.5, (neutra)
#                   "Polarity_Predict" : 0.5 (polaridad dada por el
# modulo IA, correcta)
#                 }, ...
#             ]
#         },
#     { "id" : 2
#       "post_id" : "1051225379665692",

```

```

#             "post_text" : "Final del partido \u23f1\u00a1A
S...",
#             "Fecha" : "2023-11-09", "Hora" : "11:15:32",
#             "comments" : [
#                 { "id" : 1,
#                   "text" : "A campeonar albitos para darle esos 02
puntos al principal que lo necesita",
#                   "Polarity" : 0 (sin asignar),
#                   "Polarity_Predict" : 0 (polaridad sin pronosticar por
el modulo IA)
#                 },
#                 { "id" : 2,
#                   "text" : "Jalen a muchos de la reserva para...",
#                   "Polarity" : 1, (positiva)
#                   "Polarity_Predict" : 1 (polaridad dada por el modulo
IA, correcta)
#                 }, ...
#             ]
#         },
#     "CuscoFCoficial" : [
#         :
#     ]
# }
#
#
pub_comments = dict()

```

```

@app.route("/listar_pub/<nom_equipo>/<int:id_p>")
def listar_publicaciones(nom_equipo = 'None', id_p = 0):
    global status_rb_Polarity

```

```

listado_publicaciones = "
listado_comentarios = "
equipo_sel = 'None'
id_pub = 'None'
if nom_equipo!='None':
    equipo_sel = nom_equipo
    id_pub = "pub"+str(id_p)
    # Aqui se listarían todas las publicaciones del equipo elegido, desde el
directorio base
    full_path_base = os.path.join(os.getcwd(),data_folder,nom_equipo)
    if os.path.exists(full_path_base):
        list_files_json = os.listdir(full_path_base)
        # Crear una lista de tuplas (nombre_archivo, fecha_modificacion)
        # ordenadas por fecha de modif de mas reciente al menos reciente
        files_with_mdate = []
        for archivo in list_files_json:
            ruta_archivo = os.path.join(full_path_base, archivo)
            info_archivo = os.stat(ruta_archivo)
            fecha_modificacion
            =
datetime.datetime.fromtimestamp(info_archivo.st_mtime)
            files_with_mdate.append((archivo, fecha_modificacion))
        # Ordenar la lista por fecha de modificación
        files_with_mdate.sort(key=lambda x: x[1])
        # Comenzamos a listar...
        n = 1
        list_pub = list()
        for archivo, mdate in files_with_mdate:
            complete_path_name = os.path.join(full_path_base, archivo)
            if os.path.isfile(complete_path_name) and archivo.endswith(".json"):
                #leemos el archivo json...
                with open(complete_path_name) as archivo_datos:
                    data = json.load(archivo_datos)
                    a_pub = dict() # nuevo dicc para la actual pub

```

```

# Ahora preguntamos si la publicacion actual tiene comentarios...
if 'user_reactions' in data.keys() and data['user_reactions'] is not None:
    # De ser asi, se incluye en el listado...
    listado_publicaciones += '<tr id="pub'+str(n)+'">'
    listado_publicaciones      +=      '<td      align="center"
onmouseover="sombrear_pub('+str(n)+')">'+str(n)+'</td>' # Id.
    a_pub.update({"id" : n})
    a_pub.update({"post_id" : data["post_id"]})
    a_pub.update({"post_text" : data["post_text"]})
    listado_publicaciones      +=      '<td      onclick="sel_pub('+str(n)+')"
onmouseover="sombrear_pub('+str(n)+')">'+html.escape(data['post_text'])+'</td>'
    if data['time']==None:
        fecha_formateada = '-'
        hora_formateada = '-'
    else:
        fecha_formateada, hora_formateada = data['time'].split("T")
    a_pub.update({"Fecha" : fecha_formateada})
    a_pub.update({"Hora" : hora_formateada})
    listado_publicaciones      +=      '<td      onclick="sel_pub('+str(n)+')"
onmouseover="sombrear_pub('+str(n)+')">'+fecha_formateada+'</td>'
    listado_publicaciones      +=      '<td      onclick="sel_pub('+str(n)+')"
onmouseover="sombrear_pub('+str(n)+')">'+hora_formateada+'</td>'
    listado_publicaciones += '</tr>'
# Ahora preguntamos si la publicacion actual es la seleccionada...
if n == id_p:
    # para entonces listar los comentarios de dicha publicacion...
    list_comments = list()
    listado_comentarios = '<tr>'
    lista_comentarios = data['user_reactions']['texts']
    i_com = 1
    for comentario in lista_comentarios:
        a_com = dict()

```

```

a_com.update({"id" : i_com})
a_com.update({"text" : comentario})
listado_comentarios += '<td align="center">'+str(i_com)+'</td>'
listado_comentarios += '<td>'+html.escape(comentario)+'</td>'
# Columna tercera, donde se clasifica la polaridad ( Positivo,
Neutro y Negativo)
# Se valora la llave sentiment_manual para seleccionar la
opcion de Polaridad para cada comentario
Polarity = round(data['sentiment_manual'])
a_com.update({"Polarity" : Polarity})
a_com.update({"Polarity_Predict" : 0})
listado_comentarios += '<td><table width="200">'
listado_comentarios += '<tr>'
listado_comentarios += '<td><label>'
listado_comentarios += '<input'
type="radio" name="clasificacion'+str(i_com)+'" value="1"
'+status_rb_Polarity[Polarity][ 'CheckPos'+']>Positivo'
listado_comentarios += '</label></td>'
listado_comentarios += '</tr>'
listado_comentarios += '<tr>'
listado_comentarios += '<td><label>'
listado_comentarios += '<input'
type="radio" name="clasificacion'+str(i_com)+'" value="0"
'+status_rb_Polarity[Polarity][ 'CheckNeu'+']>Neutro'
listado_comentarios += '</label></td>'
listado_comentarios += '</tr>'
listado_comentarios += '<tr>'
listado_comentarios += '<td><label>'
listado_comentarios += '<input'
type="radio" name="clasificacion'+str(i_com)+'" value="-1"
'+status_rb_Polarity[Polarity][ 'CheckNeg'+']>Negativo'
listado_comentarios += '</label></td>'
listado_comentarios += '</tr>'

```

```

        listado_comentarios += '</table>'
        listado_comentarios += '</td>'
        listado_comentarios += '</tr>'
        list_comments.append(a_com)
        i_com += 1
        a_pub.update({"comments" : list_comments})
    else:
        a_pub.update({"comments" : list()})
    list_pub.append(a_pub)
    n += 1
    pub_comments.update({equipo_sel : list_pub})
# Verificando resultado de las dos listas...
if listado_publicaciones=="":
    listado_publicaciones = Markup('<tr><td>&nbsp;</td><td>&nbsp;</td><td>&nbsp;</td><td>&nbsp;</td></tr>')
if listado_comentarios=="":
    listado_comentarios = Markup('<tr>
    listado_comentarios += '<td valign="top"><div align="center"></div></td>'
    listado_comentarios += '<td valign="top"><div align="left"></div></td>'
    listado_comentarios += '<td valign="top"><div align="center"></div></td>'
    listado_comentarios += '</tr>')
return render_template("clasificacion.html",
listado_publicaciones=Markup(listado_publicaciones),
    listado_comentarios=Markup(listado_comentarios), equipo_sel=equipo_sel,
id_pub=id_pub)

def training():
    global data_folder
    # Inicializa una lista para almacenar los datos
    data = []
    # Itera a través de los archivos JSON en la carpeta
    for filename in os.listdir(data_folder):
        if filename.endswith(".json"):

```

```

file_path = os.path.join(data_folder, filename)
with open(file_path, 'r', encoding='utf-8') as file:
    json_data = json.load(file)
    data.append(json_data)
# Cargar el modelo BERT y el tokenizador
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
model = BertForSequenceClassification.from_pretrained('bert-base-uncased',
num_labels=len(labels))
# Tokenizar y preparar los datos
input_ids = []
attention_masks = []
labels_encoded = []
for example in data:
    inputs = tokenizer(example["text"], padding="max_length", max_length=128,
truncation=True, return_tensors="pt")
    input_ids.append(inputs["input_ids"])
    attention_masks.append(inputs["attention_mask"])
    labels_encoded.append(labels.index(example["label"]))
input_ids = torch.cat(input_ids, dim=0)
attention_masks = torch.cat(attention_masks, dim=0)
labels_encoded = torch.tensor(labels_encoded)
dataset = TensorDataset(input_ids, attention_masks, labels_encoded)
# Dividir el conjunto de datos en entrenamiento y prueba
train_size = int(0.8 * len(dataset))
test_size = len(dataset) - train_size
train_dataset, test_dataset = torch.utils.data.random_split(dataset, [train_size,
test_size])
# Cargar los datos en lotes (batch)
train_dataloader = DataLoader(train_dataset, batch_size=32, shuffle=True)
test_dataloader = DataLoader(test_dataset, batch_size=32, shuffle=False)
# Definir el optimizador
optimizer = AdamW(model.parameters(), lr=2e-5, eps=1e-8)
# Entrenar el modelo

```

```

model.train()
for epoch in range(3): # Cambia el número de épocas según tus necesidades
    for batch in train_dataloader:
        input_ids, attention_mask, labels = batch

        optimizer.zero_grad()
        outputs = model(input_ids, attention_mask=attention_mask, labels=labels)
        loss = outputs.loss
        loss.backward()
        optimizer.step()
    # Evaluar el modelo
    model.eval()
    correct = 0
    total = 0
    with torch.no_grad():
        for batch in test_dataloader:
            input_ids, attention_mask, labels = batch
            outputs = model(input_ids, attention_mask=attention_mask)
            logits = outputs.logits
            predicted = torch.argmax(logits, dim=1)
            total += labels.size(0)
            correct += (predicted == labels).sum().item()
    accuracy = correct / total
    print(f"Accuracy: {accuracy * 100:.2f}%")

if __name__ == '__main__':
    app.run(debug = True)

```

Anexo III – Código para el entrenamiento automático del modelo

```
// JavaScript Document para ENTRENAR.HTML

function entrenar_evaluar()
{
    //Aqui invocamos tres veces al servidor para obtener los resultados del
    entrenamiento de los modelos
    //clasificadores: SVM, NB y NN

    // Crear instancia de objeto XMLHttpRequest para SVM (Support Vector
    Machine) ...
    var xhrSVM = new XMLHttpRequest();
    // Configurar la solicitud
    xhrSVM.open('GET', '/entrenar_svm', true);
    // Configurar el evento de carga completa
    xhrSVM.onload = function() {
        if (xhrSVM.status === 200) {
            // Obtener el contenido de la tabla
            var tabla = xhrSVM.responseText;
            document.querySelector('#tabla_SVM').innerHTML = tabla;
        }
    };
    // Enviar la solicitud
    xhrSVM.send()

    // Crear instancia de objeto XMLHttpRequest para NB (Navies Bayes) ...
    var xhrNB = new XMLHttpRequest();
    // Configurar la solicitud
    xhrNB.open('GET', '/entrenar_nb', true);
    // Configurar el evento de carga completa
    xhrNB.onload = function() {
        if (xhrNB.status === 200) {
```

```

        // Obtener el contenido de la tabla
        var tabla = xhrNB.responseText;
        document.querySelector('#tabla_NB').innerHTML = tabla;
    }
};
// Enviar la solicitud
xhrNB.send()

// Crear instancia de objeto XMLHttpRequest para NN (Neural Network) ...
var xhrNN = new XMLHttpRequest();
// Configurar la solicitud
xhrNN.open('GET', '/entrenar_nn', true);
// Configurar el evento de carga completa
xhrNN.onload = function() {
    if (xhrNN.status === 200) {
        // Obtener el contenido de la tabla
        var tabla = xhrNN.responseText;
        document.querySelector('#tabla_NN').innerHTML = tabla;
    }
};
// Enviar la solicitud
xhrNN.send()
}

```

```

funcion polaridad_global()
{
    //Aqui invocamos tres veces al servidor, de forma sincronica, para obtener
    las graficas resultantes de tipo barra
    //para cada clase y la frecuencia y/o porciento por cada clase...
    //clasificadores: SVM, NB y NN
    return new Promise((resolve, reject) => {
        var xhrSVM = new XMLHttpRequest();
        xhrSVM.open('GET', '/polaridad_global_svm', true);
    });
}

```

```

xhrSVM.responseType = 'blob';
xhrSVM.onreadystatechange = function() {
    if (xhrSVM.readyState === 4 && xhrSVM.status === 200) {
        // Crear una URL temporal para el gráfico
        var url = URL.createObjectURL(xhrSVM.response);
        // Crear una imagen y establecer la URL como fuente
        var img = document.createElement('img');
        img.src = url;
        // Modificar la anchura del gráfico
        img.style.width = '100%';
        // Añadir la imagen a la celda de la tabla
        var
                celda
                =
document.getElementById('Graph_PolaridadGlobal_SVM');
        celda.appendChild(img);

        resolve(xhrSVM.response);
    } else if (xhrSVM.readyState === 4 && xhr1.status !== 200) {
        reject(xhrSVM.statusText);
    }
};
xhrSVM.send();
}).then(grafico1 => {
    var xhrNB = new XMLHttpRequest();
    xhrNB.open('GET', '/polaridad_global_nb', true);
    xhrNB.responseType = 'blob';
    xhrNB.onreadystatechange = function() {
        if (xhrNB.readyState === 4 && xhrNB.status === 200) {
            // Crear una URL temporal para el gráfico
            var
                    url
                    =
URL.createObjectURL(xhrNB.response);
            // Crear una imagen y establecer la URL como
fuente
            var img = document.createElement('img');

```

```

        img.src = url;
        // Modificar la anchura del gráfico
        img.style.width = '100%';
        // Añadir la imagen a la celda de la tabla
        var celda =
document.getElementById('Graph_PolaridadGlobal_NB');
        celda.appendChild(img);

        resolve(xhrNB.response);
    } else if (xhrNB.readyState === 4 && xhrNB.status !==
200) {
        reject(xhrNB.statusText);
    }
};
xhrNB.send();
}).then(grafico2 => {
    var xhrNN = new XMLHttpRequest();
    xhrNN.open('GET', '/polaridad_global_nn', true);
    xhrNN.responseType = 'blob';
    xhrNN.onreadystatechange = function() {
        if (xhrNN.readyState === 4 && xhrNN.status === 200) {
            // Crear una URL temporal para el gráfico
            var url =
URL.createObjectURL(xhrNN.response);
            // Crear una imagen y establecer la URL como
fuente

            var img = document.createElement('img');
            img.src = url;
            // Modificar la anchura del gráfico
            img.style.width = '100%';
            // Añadir la imagen a la celda de la tabla
            var celda =
document.getElementById('Graph_PolaridadGlobal_NN');

```

```
        celda.appendChild(img);

        resolve(xhrNN.response);
    } else if (xhrNN.readyState === 4 && xhrNN.status !==
200) {
        reject(xhrNN.statusText);
    }
};
xhrNN.send();
}).catch(error => {
// Manejar el error en caso de que alguna de las peticiones falle
console.error(error);
});
}
```

Anexo IV – Matriz de Consistencia

TÍTULO	PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES	DIMENSIÓN	INDICADORES	METODOLOGÍA
Análisis de los sentimientos y emociones en las redes sociales de equipos de una Liga de fútbol utilizando inteligencia artificial	GENERAL: De qué manera los internautas expresan sus sentimientos y emociones en las redes sociales de sus equipos favoritos de la Liga 1 peruana y cómo la inteligencia artificial permite realizar el análisis de estas publicaciones como insumo que guíe a los Community Managers en su planificación de publicaciones	GENERAL: Implementar un modelo de inteligencia artificial que permita analizar los sentimientos y emociones en las redes sociales de los equipos de una Liga de fútbol. ESPECÍFICOS: Desarrollar un modelo de aprendizaje automático que permita analizar y clasificar los sentimientos y emociones expresados en los textos de las publicaciones y comentarios en redes sociales. Recopilar y preparar un conjunto de datos etiquetados adecuado que represente una variedad de sentimientos y emociones en diferentes contextos de las redes		VI: Técnicas de procesamiento	Tipo de técnica. Volumen de datos.	Velocidad de procesamiento. Porcentaje de precisión.	DISEÑO: Experimental TIPO: Aplicada MUESTRA: 18 Equipos más populares de la Liga 1 peruana MUESTREO: No probabilístico TÉCNICAS E INSTRUMENTO DE RECOLECCIÓN DE DATOS: Web scraping ESCALA: Razón
				Algoritmos de aprendizaje automático	Tipo de algoritmo. Enfoque del aprendizaje automático.	Tiempo de entrenamiento. Porcentaje de Sensibilidad	
				Características del texto	Longitud del texto. Diversidad de vocabulario.	Promedio de la longitud del texto. Frecuencia de Términos Específicos	
				VD: Temas emocionales	Emociones básicas Emociones sociales	Frecuencia de emociones. Análisis de sentimientos.	
				Polaridad emocional	Emociones Positivas. Emociones negativas.	Polaridad Promedio. Intensidad Emocional	
				Intensidad emocional	Efecto de los Eventos. Efecto de la Audiencia	Distribución de intensidad. Puntuación de intensidad.	

TÍTULO	PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES	DIMENSIÓN	INDICADORES	METODOLOGÍA
		<p>sociales.</p> <p>Evaluar y mejorar el rendimiento del modelo de inteligencia artificial aplicando las diferentes métricas de desempeño.</p> <p>Identificar patrones y tendencias en la expresión de sentimientos y emociones en las redes sociales a través del análisis de los resultados obtenidos.</p>					

Anexo V – Certificado

Piura, 31 de octubre del 2023

Mg. Elmer Alfredo Chunga Zapata

De nuestra especial consideración, tenemos a bien saludarle.

Somos los estudiantes Adrianzén Zapata Marco Antonio, con código de matrícula N° 7001134875 y Fernández Olaya Juan, con código de matrícula N° 7002524336. Nos dirigimos a Ud. muy respetuosamente, para manifestarle lo siguiente:

Ambos somos compañeros de tesis, llevamos el curso de "Desarrollo del Proyecto de la Investigación" del décimo ciclo, carrera profesional de Ingeniería de Sistemas. Nuestro Proyecto de Investigación, lleva como título "Análisis de los sentimientos y emociones en las redes sociales de equipos de una Liga de fútbol utilizando inteligencia artificial". El motivo del presente, es para comunicarle que nuestro proyecto no está vinculado a ninguna empresa, por lo que es de investigación libre, por lo cual no contamos con el permiso de alguna organización como lo solicita trilce, así mismo con permiso de escuela se hace llegar este documento con nuestras firmas, para los fines que se crea conveniente.

Nos despedimos de usted, reiterándole una vez más nuestro distinguido aprecio y estima.

Atentamente,



Adrianzén Zapata, Marco Antonio
DNI 78007389



Fernández Olaya, Juan
DNI 43709242