



## **Trabajo Final de Graduación**

Análisis de sentimientos en redes sociales con modelos Transformer:  
Aplicación al fútbol argentino.

Carrera de grado: Licenciatura en Analítica de Negocios

Autor: Gianluca Mazzocchi

DNI: 44.967.918

Tutora del trabajo: Paula Mariela Iaccarino

Materia: Analítica de Negocios Avanzada

Año: 2024

## Resumen

El presente trabajo se centra en la implementación de técnicas avanzadas de procesamiento de lenguaje natural (NLP) y aprendizaje automático para analizar los sentimientos expresados por los hinchas en redes sociales, específicamente en el ámbito del fútbol argentino. Twitter, como una de las plataformas más utilizadas por la hinchada, permite recopilar opiniones y reacciones en tiempo real sobre equipos, jugadores y eventos deportivos. Sin embargo, la naturaleza dispersa y no estructurada de estos datos dificulta su análisis eficiente.

Para abordar este problema, se empleó un modelo basado en la arquitectura Transformer, con un enfoque particular en BERT (Bidirectional Encoder Representations from Transformers), utilizado para clasificar automáticamente tweets en categorías de sentimientos positivos, negativos y neutros. El modelo fue aplicado a un conjunto de datos compuesto por más de mil tweets, previamente procesados para garantizar la calidad del análisis.

El trabajo demuestra que el uso de herramientas modernas como BERT permite captar matices emocionales complejos y mejorar significativamente la precisión del análisis de sentimientos. Entre otras cosas, se discuten aplicaciones prácticas del modelo en la optimización de estrategias de comunicación y marketing para clubes y asociaciones deportivas así como también una mayor fidelización y comunicación con la hinchada a la cual responde cada club. Este estudio no solo aporta una herramienta innovadora para comprender las dinámicas emocionales de la hinchada argentina, sino que también sienta las bases para futuras investigaciones en el cruce entre el deporte y la tecnología de datos.

# Índice

<b>Resumen</b>	<b>2</b>
<b>Índice</b>	<b>3</b>
<b>Introducción</b>	<b>4</b>
1.1 El problema del análisis de sentimientos en redes sociales	4
1.2 Data Product Canvas	5
1.3 Herramientas y tecnologías aplicadas	10
<b>Marco Teórico</b>	<b>13</b>
2.1 Origen de los Transformers y el Procesamiento de Lenguaje Natural (NLP)	13
2.2 Modelos de lenguaje actuales y su aplicación en redes sociales	14
2.3 Enfoque BERT para el análisis de sentimientos	16
2.4 Modelo VADER	17
2.5 TextBlob: Una biblioteca python para el procesamiento del lenguaje natural	18
2.6 Métricas de evaluación	19
<b>Desarrollo</b>	<b>21</b>
3.1 Metodología	21
3.2 Recolección de los datos	22
3.3 Limpieza y preprocesamiento de tweets	23
3.4 Análisis exploratorio de los datos (EDA)	24
3.5 Análisis de sentimientos con Transformer	29
3.6 Resultados del análisis	31
3.7 Comparación con otros modelos	33
<b>Dashboard</b>	<b>37</b>
<b>Proyecciones</b>	<b>39</b>
<b>Conclusiones</b>	<b>40</b>
<b>Referencias</b>	<b>43</b>
<b>Anexo</b>	<b>45</b>

# Introducción

## 1.1 El problema del análisis de sentimientos en redes sociales

El fútbol argentino no es solo un deporte, sino una parte esencial de la cultura nacional que moviliza pasiones y une a millones de personas. Este fenómeno encuentra en las redes sociales un espacio privilegiado para la expresión de emociones y opiniones, especialmente durante los partidos y eventos clave. Twitter, como una de las plataformas más dinámicas, permite a los hinchas compartir en tiempo real sus reacciones ante jugadas, decisiones arbitrales o noticias deportivas, generando un flujo masivo y constante de datos que refleja la intensidad emocional del fútbol argentino.

El auge del procesamiento de lenguaje natural (NLP) ha abierto nuevas posibilidades para analizar este tipo de interacciones. Herramientas avanzadas como BERT (Bidirectional Encoder Representations from Transformers), basadas en la arquitectura Transformer, han revolucionado la interpretación del texto, superando las limitaciones de los métodos tradicionales al captar no solo el significado literal de las palabras, sino también sus matices contextuales y emocionales. Estas capacidades son especialmente valiosas en un contexto como el del fútbol argentino, donde las expresiones coloquiales, regionalismos y sarcasmos complican el análisis de sentimientos.

El objetivo principal de este trabajo es analizar y comparar el desempeño de diferentes técnicas de procesamiento de lenguaje natural (NLP) aplicadas al análisis de sentimientos en redes sociales en el contexto del fútbol argentino. Esto incluye la implementación y evaluación de modelos basados en Transformers, como BERT, junto con otras metodologías tradicionales, como VADER y TextBlob. Además, se busca identificar las fortalezas y limitaciones de cada enfoque para clasificar automáticamente los sentimientos expresados por hinchas en Twitter, abordando las emociones positivas, negativas y neutras. Este enfoque no solo busca superar los desafíos técnicos y lingüísticos inherentes al análisis de datos no estructurados, sino también generar insights accionables que puedan ser utilizados por clubes y asociaciones deportivas para mejorar en varios aspectos como la comunicación, estrategias y aspectos relacionados tanto a lo deportivo como a lo social.

Este estudio integra conceptos clave de la carrera de analítica de negocios, como la transformación de datos no estructurados en valor práctico, la aplicación de inteligencia

artificial en contextos complejos y la generación de herramientas para la toma de decisiones informadas. Además, se alinea con estudios previos en análisis de sentimientos y NLP, contribuyendo a la literatura mediante la adaptación de estas técnicas al contexto único del fútbol argentino.

Asimismo, el análisis de sentimientos en redes sociales enfrenta desafíos particulares, como la velocidad con la que se generan tweets durante los eventos deportivos y las restricciones impuestas por la API de Twitter, que limitan el acceso a datos en tiempo real. Para mitigar estas barreras, se recurre al uso de datos sintéticos que emulan escenarios reales, permitiendo evaluar la efectividad del modelo y capturar la volatilidad emocional de los hinchas en tiempo real.

Este trabajo no solo tiene relevancia práctica en el ámbito deportivo, sino que también aporta una perspectiva innovadora sobre cómo las herramientas de NLP pueden integrarse en el diseño de estrategias basadas en datos en sectores dinámicos y culturalmente significativos. Se puede decir que sienta las bases para futuras investigaciones que puedan extender este enfoque a otros deportes o entornos culturales, consolidando la intersección entre tecnología, análisis de datos y comportamiento humano.

## **1.2 Data Product Canvas**

El Data Product Canvas proporciona una estructura para guiar la implementación de este modelo, asegurando que cada decisión esté alineada con los objetivos estratégicos de los clubes del fútbol argentino y la AFA. Este enfoque facilita la transformación de datos desestructurados en información accionable para apoyar decisiones clave.

### Problema:

Los clubes del fútbol argentino y la AFA enfrentan dos desafíos principales relacionados con la gestión de las emociones de sus hinchas:

- *Comprender y responder a las emociones de los hinchas:* No cuentan con herramientas para interpretar de manera efectiva los sentimientos predominantes en redes sociales, lo que limita su capacidad de diseñar estrategias de comunicación que fortalezcan la relación con su audiencia.
- *Tomar decisiones basadas en insights emocionales:* La falta de análisis estructurado de las opiniones y reacciones de los hinchas dificulta la identificación de

oportunidades para mejorar su conexión emocional y aumentar la fidelidad de su público.

#### Solución:

La solución propuesta es un modelo de análisis de sentimientos basado en BERT, implementado para procesar el lenguaje futbolero argentino en redes sociales. Este modelo permite a los clubes y la AFA:

- *Clasificar automáticamente los sentimientos* expresados en Twitter en categorías de positivos, negativos o neutros.
- *Generar insights procesables* que faciliten el diseño de estrategias de comunicación, gestión de crisis y optimización de experiencias para sus hinchas.
- *Proveer dashboards interactivos* que permitan interpretar fácilmente los resultados y tomar decisiones informadas.

#### Datos:

El proyecto utiliza datos sintéticos para superar las limitaciones de acceso a datos reales, asegurando que el modelo sea representativo del lenguaje y las emociones de los hinchas.

- *Origen:* Dataset FIFA World Cup 2022 (Anexo 1), adaptado para reflejar interacciones típicas de hinchas argentinos.
- *Características:* Textos enriquecidos con jerga futbolera, emojis y expresiones comunes del fútbol argentino.
- *Volumen:* Se generaron más de 3,000 registros iniciales para validar el modelo, asegurando una representación adecuada del contexto emocional.

#### Hipótesis:

Se plantea que los insights derivados del análisis de sentimientos permitirán a los clubes y a la AFA comprender de manera más profunda las emociones y opiniones de los hinchas en relación con jugadores, equipos, arbitrajes y competencias. Esta comprensión facilitará la toma de decisiones estratégicas basadas en datos, como ajustes en las dinámicas deportivas, el fortalecimiento de las relaciones con los hinchas y la mejora de la gestión de eventos y competencias.

Se espera que, a través del modelo, los clubes y la AFA puedan identificar patrones emocionales recurrentes, como picos de descontento o entusiasmo, y actuar en consecuencia para mitigar conflictos o potenciar aspectos positivos. Por ejemplo, ante un sentimiento generalizado de frustración por una decisión arbitral o el desempeño de un equipo, los datos podrán guiar a las instituciones hacia acciones concretas que aborden estas preocupaciones, fortaleciendo la percepción de cercanía y receptividad hacia las opiniones de los hinchas.

### Actores:

Existe una dinámica única entre los distintos actores clave del ecosistema deportivo. Cada uno de ellos juega un rol particular en la generación, uso y aprovechamiento de los datos analizados por el modelo, con beneficios claros y específicos que fortalecen sus respectivas estrategias.

#### *Hinchas (Generadores de datos)*

- Rol: Los hinchas son los principales generadores de los datos no estructurados que alimentan el modelo. A través de sus publicaciones en redes sociales, como tweets, expresan sus emociones, opiniones y percepciones sobre equipos, jugadores, arbitrajes y eventos deportivos, que constituyen el insumo esencial del análisis.
- Beneficio: Aunque no son usuarios directos del modelo, los hinchas se benefician indirectamente al experimentar una mayor conexión emocional con los clubes. Las estrategias que los clubes diseñen, respaldadas por datos obtenidos de los sentimientos de los hinchas, permitirán una comunicación más empática y alineada con sus expectativas y emociones. Como consecuencia de esto, se puede fortalecer la relación hincha-club y generar un sentido de pertenencia más sólido.

#### *Clubes de fútbol (Usuarios principales)*

- Rol: Son los usuarios principales del producto. Utilizan los insights generados por el análisis para tomar decisiones estratégicas que impactan tanto en su desempeño deportivo como en su relación con los hinchas. Esto incluye decisiones como transferencias de jugadores, renovaciones de contratos, comunicación institucional, y campañas de marketing dirigidas a los hinchas.

- Beneficio: Los clubes pueden dejar de basar sus decisiones exclusivamente en opiniones internas o intuiciones y, en cambio, respaldarlas con datos concretos. Esto no solo mejora la efectividad de las decisiones, sino que también fortalece su imagen ante los hinchas, quienes perciben un mayor compromiso con sus sentimientos. Por ejemplo, si los hinchas expresan mayor apoyo hacia un jugador emergente, el club puede priorizar su renovación o dar más visibilidad a su desarrollo.

#### *AFA (Organizador del ecosistema)*

- Rol: La Asociación del Fútbol Argentino, como ente regulador, puede utilizar los datos analizados para entender las percepciones generales del público sobre sus decisiones. Esto incluye la organización de torneos, el formato de competencias y la gestión de eventos.
- Beneficio: Con acceso a los insights generados por el modelo, la AFA podría diseñar estrategias que busquen recuperar la confianza de los hinchas. Por ejemplo, si las opiniones reflejan un descontento con torneos de formato extenso, la AFA podría considerar una reformulación basada en estas percepciones, fortaleciendo su legitimidad como ente rector.

#### Acciones:

A partir de los insights generados por el modelo, se pueden realizar las siguientes acciones:

- *Club del fútbol argentino:* Diseñar estrategias de engagement basadas en emociones detectadas, las decisiones que se tomen no serán al azar o en base a opiniones de la comisión directiva sino que se va a poder tomar decisiones respaldadas por los datos, ¿Que opina el hincha? ¿Es bueno o malo? ¿Venimos bien? Estas decisiones incluyen compra de jugadores, elección de director técnico, etc. Este acercamiento al hincha sin duda permitiría tener ese respaldo desde la hinchada al club que hoy en día, en algunos clubes, no se tiene.
- *AFA:* Coordinar la confección de los torneos en base al análisis de las opiniones de los hinchas. Un ejemplo breve es, los hinchas piensan que se pierde prestigio al realizar un torneo de 30 equipos, bueno lo analizamos. Actualmente y como dije en la presentación de la idea de negocio, la gente ODIA al ente máximo del fútbol argentino debido a las malas decisiones, entonces con este modelo van a poder tener también opiniones de gente que conoce y que no tiene intereses propios.



### KPI's:

Para evaluar el éxito del modelo y su impacto en los objetivos del negocio, se definen:

- ***KPIs técnicos:***

- *Precisión del modelo:* Se evaluará el porcentaje de clasificaciones correctas de sentimientos utilizando métricas estandarizadas como Accuracy, Precision, Recall y F1-Score. Estas métricas permitirán determinar qué tan bien el modelo clasifica los tweets en las categorías de positivo, neutral y negativo, asegurando que no exista un sesgo significativo hacia ninguna clase. El objetivo será mantener una diferencia mínima entre las métricas de cada clase, con una variación de menos del 5% en el F1-Score.
- *Velocidad de procesamiento:* Se medirá el tiempo necesario para analizar grandes volúmenes de tweets, evaluando la capacidad del modelo para manejar datos en tiempo real. La meta será que el sistema procese 1,000 tweets en menos de 30 segundos, asegurando una latencia óptima para aplicaciones dinámicas como el monitoreo de eventos deportivos en vivo. Este indicador refleja la escalabilidad del modelo y su aplicabilidad en escenarios de alta demanda.

- ***KPIs de negocio:***

- *Aumento del engagement:* Incremento en la interacción de los hinchas con los clubes y la AFA tras implementar estrategias basadas en los insights del modelo.
- *Satisfacción de los hinchas:* Mejora en las métricas de percepción y lealtad de los hinchas hacia los clubes y las competencias organizadas.
- *Impacto financiero:* Medir cómo estas acciones se traducen en beneficios económicos. Por ejemplo, entender si las campañas de marketing dirigidas, basadas en los datos del modelo, generan más ingresos por venta de entradas, merchandising o patrocinios. Si el modelo detecta, por ejemplo, un interés creciente por un jugador en particular, se podrían diseñar estrategias específicas para capitalizar esa información y aumentar las ganancias.

### Valores:

El modelo proporciona valor al:

- *Mejorar* la capacidad de los clubes y la AFA para comprender y responder a las emociones de los hinchas.
- *Fortalecer* la conexión emocional entre las instituciones y sus públicos.
- *Optimizar* la toma de decisiones basadas en datos procesables, generando resultados medibles en términos de engagement y fidelización.

#### Riesgos:

- *Representatividad limitada:* Los datos sintéticos podrían no capturar completamente la riqueza del lenguaje futbolero.
- *Evolución del lenguaje:* El modelo debe ser actualizado periódicamente para mantenerse alineado con las expresiones y tendencias emergentes.

#### Impacto:

El modelo permite decisiones más informadas que benefician tanto a los clubes como a la AFA, mejorando la experiencia de los hinchas y fortaleciendo los lazos emocionales en el ecosistema del fútbol argentino. Esto, a su vez, contribuye a un mayor nivel de satisfacción y lealtad por parte de los aficionados.

### **1.3 Herramientas y tecnologías aplicadas**

El desarrollo del modelo de análisis de sentimientos se basó en un conjunto de herramientas y tecnologías seleccionadas para abordar los desafíos específicos del proyecto. Estas herramientas permitieron procesar grandes volúmenes de datos, aplicar técnicas avanzadas de procesamiento de lenguaje natural (NLP) y garantizar la efectividad del modelo en el contexto del fútbol argentino.

#### **Datos sintéticos**

La primera etapa del proyecto consistió en la generación de datos sintéticos a partir del dataset FIFA World Cup 2022, obtenido de Kaggle (Anexo 1). Este dataset sirvió como base para simular interacciones propias de los hinchas argentinos en redes sociales. Los registros fueron adaptados para reflejar dinámicas características del lenguaje futbolero argentino, como el uso de términos coloquiales, regionalismos y expresiones emocionales típicas.

#### **Preprocesamiento y limpieza**

Python fue el lenguaje principal utilizado, gracias a su versatilidad y la disponibilidad de bibliotecas especializadas en NLP. Herramientas como **SpaCy** y **NLTK** fueron clave en el preprocesamiento de los datos, realizando tareas como:

- *Tokenización*: Fragmentación de texto en palabras o frases clave.
- *Lematización*: Reducción de palabras a su forma base para mejorar la consistencia.
- *Stopwords*: Limpieza de palabras no relevantes para el análisis.

Adicionalmente, bibliotecas como **Pandas** y **NumPy** facilitaron la organización y manipulación de los datos, estructurándolos en formatos adecuados para su integración en el pipeline de procesamiento.

### **Implementación del modelo**

Primero se realizó un análisis simple de sentimientos con un etiquetado manual en cada tweet en la columna inicial “Sentiment” luego la idea fue implementar un modelo de análisis de sentimientos automatizado como lo es *nlptown/bert-base-multilingual-uncased-sentiment*, una variante de BERT diseñada para la clasificación de sentimientos. Este modelo fue elegido por su capacidad de interpretar el contexto bidireccional de las palabras, captando matices y tonalidades emocionales en los textos, lo que es esencial para el análisis de los tweets de los hinchas.

### **Desarrollo y evaluación**

El desarrollo del MVP se realizó en **Google Colab**, un entorno que ofrece acceso gratuito a recursos computacionales avanzados, como GPU, lo que permitió acelerar procesos intensivos como la evaluación y ajuste del modelo. Este entorno facilitó la implementación inicial y la demostración de los resultados de forma eficiente.

### **Posible despliegue en la nube**

Aunque el MVP fue desarrollado en Google Colab, para un despliegue en el mundo real se recomienda utilizar plataformas en la nube como **Google Cloud Platform (GCP)** o **AWS**. Estas plataformas permiten:

- **Escalabilidad**: Procesar grandes volúmenes de datos en tiempo real, adaptándose a las necesidades de los usuarios finales (AFA y clubes del fútbol argentino).

- **Disponibilidad:** Garantizar el acceso continuo a los modelos y dashboards desde cualquier ubicación.
- **Automatización:** Integrar pipelines de datos automáticos que actualicen los insights en tiempo real.
- **Costos optimizados:** Ofrecer soluciones eficientes en términos de costo-beneficio para tareas intensivas como el procesamiento de lenguaje natural.

El despliegue en la nube ampliará el alcance del modelo, permitiendo su uso continuo y accesible por parte de los clubes y la AFA, quienes podrían integrar los insights directamente en sus sistemas de toma de decisiones.

### **Visualización de resultados**

Los resultados del análisis se integraron en un dashboard interactivo desarrollado en **Looker Studio**, que permite visualizar tendencias emocionales de manera clara e intuitiva. Este dashboard facilita la interpretación de los datos, proporcionando insights accionables que permiten a los clubes y a la AFA tomar decisiones informadas en función de las emociones detectadas.

## **Marco Teórico**

El marco teórico establece los fundamentos conceptuales y técnicos que sustentan el desarrollo del modelo de análisis de sentimientos presentado en este trabajo. A lo largo de esta sección, se explorará la evolución del procesamiento de lenguaje natural (NLP), los modelos actuales basados en la arquitectura Transformer y específicamente el enfoque de BERT, que constituye el núcleo del análisis propuesto.

### **2.1 Origen de los Transformers y el Procesamiento de Lenguaje Natural (NLP)**

El procesamiento de lenguaje natural (NLP) es una rama de la inteligencia artificial que busca capacitar a las computadoras para comprender y procesar texto humano de manera que sea útil para tareas específicas. Desde sus inicios, este campo ha experimentado avances significativos, que van desde enfoques manuales basados en reglas lingüísticas hasta modelos de aprendizaje profundo que emplean redes neuronales complejas.

Los primeros enfoques en NLP dependían de reglas gramaticales explícitas definidas manualmente. Aunque efectivos en escenarios muy limitados, estos métodos no podían capturar las complejidades inherentes al lenguaje humano, como el sarcasmo, las ambigüedades y las referencias culturales. Posteriormente, el surgimiento de métodos estadísticos permitió modelar el lenguaje en función de probabilidades derivadas de grandes corpus textuales. Sin embargo, estos enfoques seguían siendo limitados al no poder manejar relaciones contextuales a largo plazo.

El punto de inflexión en el campo llegó con la introducción de la arquitectura Transformer por Vaswani et al. (2017). Este modelo reemplazó las redes neuronales recurrentes (RNN) y las redes de memoria a largo plazo (LSTM) que procesaban texto de manera secuencial. Los Transformers, en cambio, procesan las palabras en paralelo, lo que aumenta significativamente la eficiencia del entrenamiento y la capacidad para capturar relaciones contextuales a largo plazo. Su mecanismo de atención, conocido como "self-attention", permite al modelo asignar pesos específicos a diferentes palabras en función de su relevancia contextual dentro de la oración (Vaswani et al., 2017).

En el análisis de sentimientos aplicado a redes sociales, los Transformers ofrecen ventajas inigualables. Son capaces de interpretar el lenguaje informal y desestructurado típico de plataformas como Twitter, donde los textos suelen incluir abreviaturas, emojis, regionalismos y sarcasmo. Estas características los hacen ideales para abordar la complejidad del discurso futbolero argentino, caracterizado por un alto contenido emocional y expresiones únicas (DataScientest, n.d.).

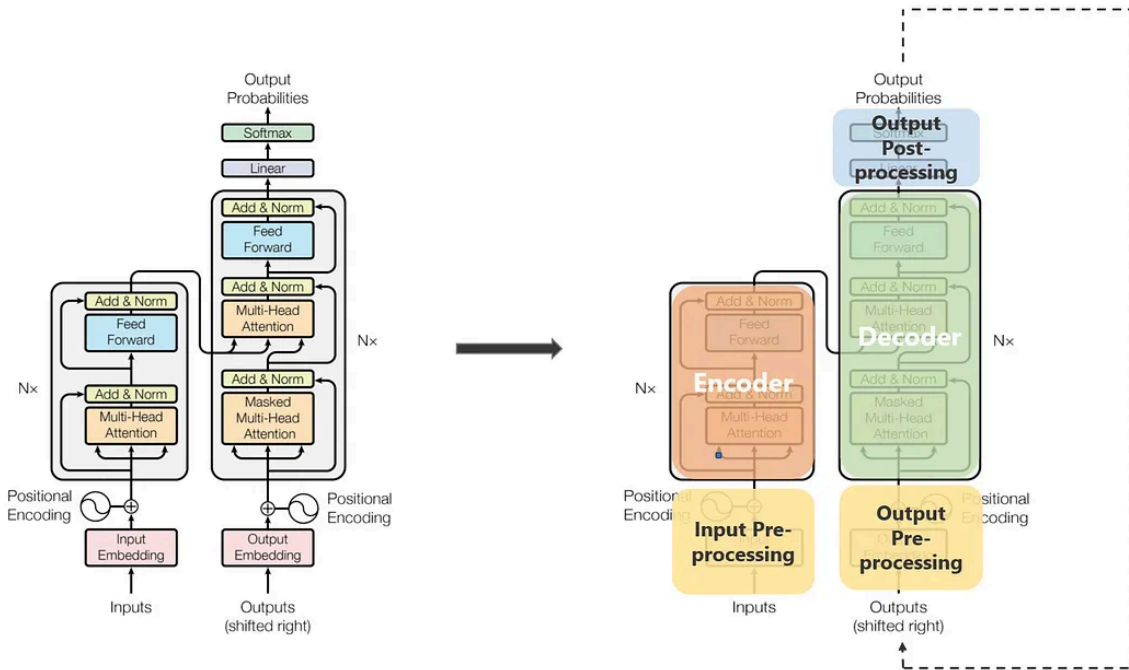


Figura 1: Arquitectura Transformer. Imagen extraída de [9].

## 2.2 Modelos de lenguaje actuales y su aplicación en redes sociales

Los avances en el procesamiento de lenguaje natural han permitido el desarrollo de modelos que transforman la forma en que se aborda el análisis de texto, especialmente en el contexto de las redes sociales. Plataformas como Twitter presentan textos breves, cargados de matices y elementos no convencionales, como emojis, abreviaturas y expresiones coloquiales. Estos mensajes, en muchos casos, están influenciados por emociones intensas y regionalismos, características que complican su interpretación para enfoques tradicionales.

Como dije anteriormente, en este trabajo se utilizó el modelo preentrenado *nlptown/bert-base-multilingual-uncased-sentiment*, diseñado para clasificar sentimientos en múltiples idiomas. Este modelo, basado en la arquitectura Transformer, se empleó

directamente sin realizar ajustes adicionales, aprovechando su capacidad para interpretar tanto el contexto anterior como posterior de las palabras dentro de un texto.

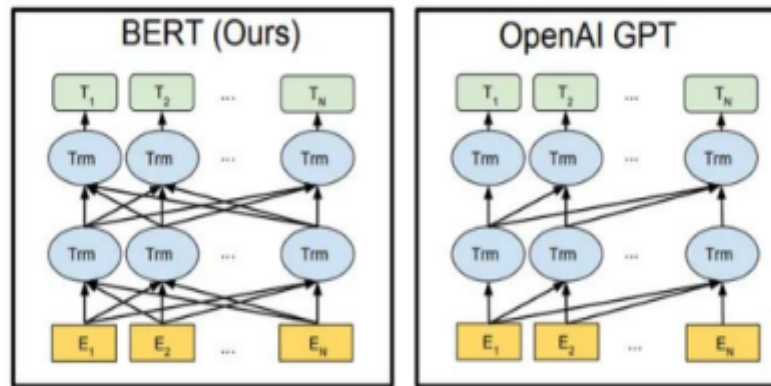


Figura 2: Comparativa entre la arquitectura bidireccional de BERT y la unidireccional de GPT. Imágen extraída de [2].

Esto resulta fundamental en el análisis de sentimientos en redes sociales, donde el significado de un mensaje puede variar significativamente dependiendo del contexto.

El lenguaje futbolero argentino, en particular, añade una capa adicional de complejidad al análisis de sentimientos. Los hinchas expresan sus emociones de manera única, utilizando frases cargadas de ironía, sarcasmo o regionalismos que, en muchos casos, requieren un entendimiento profundo del contexto. Por ejemplo, una frase como “Un clásico, jugamos como campeones y terminamos como siempre” refleja un elogio inicial seguido de una crítica implícita, que puede pasar desapercibida para métodos que no integren el contexto completo.

BERT, gracias a su enfoque bidireccional, es capaz de descomponer estos mensajes y asignar relevancia a las palabras clave según su relación con el resto del texto. Esto permite interpretar matices emocionales y lingüísticos con mayor precisión, ofreciendo insights relevantes sobre el sentir de los hinchas durante eventos deportivos. La aplicación del modelo preentrenado a los datos sintéticos utilizados en este proyecto asegura que los resultados reflejen las dinámicas emocionales propias del público objetivo.

El uso de modelos avanzados como BERT ha demostrado ser una solución efectiva para abordar los desafíos inherentes al análisis en redes sociales. En este trabajo, permitió procesar y analizar interacciones textuales de manera eficiente, obteniendo una visión más clara y

detallada de las emociones expresadas en un entorno culturalmente específico como el fútbol argentino.

## 2.3 Enfoque BERT para el análisis de sentimientos

BERT ha marcado un antes y un después en la forma en que se abordan tareas de procesamiento de lenguaje natural. Una de las características más destacadas de este modelo es su capacidad para comprender el contexto completo de un texto mediante un enfoque bidireccional, lo que lo hace especialmente efectivo en escenarios donde el lenguaje está cargado de matices, como ocurre en redes sociales. Este avance ha permitido superar las limitaciones de enfoques anteriores, como los modelos unidireccionales, que procesaban las palabras en una sola dirección y tenían dificultades para interpretar relaciones complejas entre palabras.

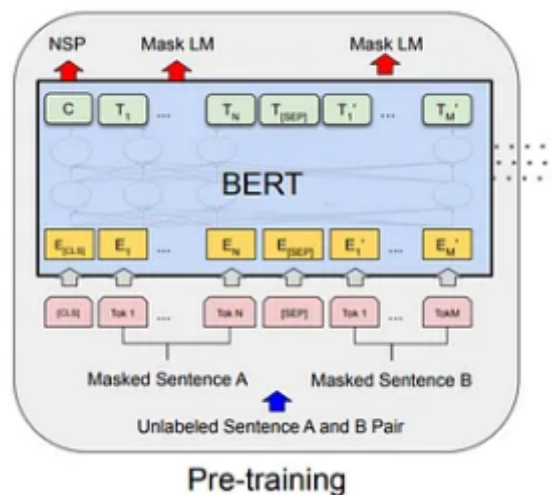


Figura 3: Estructura de preentrenamiento de BERT. El modelo utiliza dos técnicas principales: Masked Language Modeling (MLM), para predecir palabras ocultas en una oración, y Next Sentence Prediction (NSP), para evaluar relaciones entre pares de oraciones. Estas capacidades permiten a BERT captar relaciones semánticas bidireccionales y comprender el contexto completo del texto (Devlin et al., 2018). Imagen extraída de [2].

En el ámbito de las redes sociales, los textos presentan desafíos únicos: son breves, informales y, en muchos casos, contienen elementos como emojis, abreviaturas y sarcasmo. Además, el dinamismo de estas plataformas implica que las expresiones y el lenguaje evolucionan constantemente, lo que aumenta la complejidad del análisis. Ante este panorama, la arquitectura Transformer, base de BERT, ha demostrado ser una herramienta indispensable.



El mecanismo de atención que emplea le permite identificar qué palabras en un texto son más relevantes para determinar su significado general, lo que resulta crucial en textos tan desestructurados como los tweets.

En el contexto del fútbol argentino, donde las emociones de los hinchas suelen expresarse con un lenguaje altamente coloquial y regional, BERT aporta un nivel de análisis que modelos anteriores no podían alcanzar. Por ejemplo, el uso de expresiones irónicas o frases que combinan elogios y críticas dentro de un mismo mensaje son características comunes del discurso futbolero. La capacidad de BERT para procesar todo el contexto de una oración, en lugar de limitarse a interpretar cada palabra de manera aislada, le permite identificar correctamente estos matices emocionales y lingüísticos.

En este trabajo, el modelo ha sido implementado en un entorno que prioriza la eficiencia y la precisión, permitiendo analizar grandes volúmenes de datos de manera ágil. Su uso no solo facilita la clasificación de emociones, sino que también abre la puerta a aplicaciones más complejas en el futuro, como la identificación de patrones en el comportamiento de los hinchas o la generación de alertas en tiempo real para clubes y asociaciones deportivas. Además, su flexibilidad para adaptarse a diferentes conjuntos de datos garantiza que pueda seguir siendo relevante en escenarios dinámicos y en constante evolución.

Finalmente, el análisis de sentimientos realizado mediante BERT no se limita a una simple clasificación de textos, sino que proporciona una herramienta estratégica que puede ser utilizada para tomar decisiones informadas. Al ofrecer una interpretación más precisa de las emociones de los hinchas, este modelo permite a los clubes y asociaciones del fútbol argentino identificar tendencias emocionales, anticiparse a posibles crisis y diseñar estrategias de comunicación más efectivas. Esto no solo mejora la conexión con los aficionados, sino que también refuerza la toma de decisiones basadas en datos, un componente esencial en el fútbol moderno.

## **2.4 Modelo VADER: Un enfoque lexicográfico para el análisis de sentimientos en redes sociales**

VADER es un analizador de sentimientos basado en un léxico y un conjunto de reglas, diseñado específicamente para capturar la naturaleza de los sentimientos expresados en el lenguaje informal propio de las redes sociales. A diferencia de otros modelos que se basan en

técnicas de aprendizaje automático, VADER utiliza un enfoque lexicográfico, lo que significa que asigna una puntuación de polaridad a cada palabra o frase en función de su significado semántico y su contexto (Hutto & Gilbert, 2014).

El proceso de análisis de sentimientos con VADER implica la tokenización del texto de entrada, la asignación de una puntuación de polaridad a cada token basada en el diccionario propio, la consideración del contexto mediante la aplicación de reglas para ajustar la puntuación de polaridad y, finalmente, el cálculo de una puntuación compuesta que representa la polaridad general del texto.

VADER presenta varias ventajas, como su eficiencia, precisión en la identificación de sentimientos en textos de redes sociales, sensibilidad a la negación y la intensificación, y facilidad de uso. Sin embargo, también tiene limitaciones, como la dependencia de la calidad del diccionario, dificultades para capturar sarcasmo e ironía y sensibilidad a la jerga y los slang.

El modelo VADER encuentra numerosas aplicaciones en el análisis de sentimientos, incluyendo el monitoreo de la opinión pública, la evaluación de la satisfacción del cliente, la detección de crisis en las redes sociales y el análisis de la opinión política.

## **2.5 TextBlob: Una biblioteca python para el procesamiento del lenguaje natural**

TextBlob es una versátil biblioteca de Python diseñada para el procesamiento del lenguaje natural (PLN). Ofrece una amplia gama de herramientas que facilitan la manipulación y análisis de texto, incluyendo tareas como la tokenización, la lematización, el etiquetado de partes del discurso y, por supuesto, el análisis de sentimientos.

En cuanto al análisis de sentimientos, TextBlob emplea un clasificador de Naïve Bayes preentrenado para determinar la polaridad de un texto. Este clasificador ha sido expuesto a un corpus extenso de reseñas de películas, lo que le permite identificar con relativa precisión si una oración expresa un sentimiento positivo, negativo o neutro. Además de la polaridad, TextBlob también puede cuantificar la subjetividad de un texto, indicando el grado en que expresa una opinión personal en lugar de un hecho objetivo.

Una de las grandes ventajas de TextBlob reside en su facilidad de uso. Su API intuitiva la hace accesible tanto para usuarios principiantes como para aquellos con mayor experiencia en programación. Además, su capacidad para realizar diversas tareas de PLN más allá del análisis de sentimientos la convierte en una herramienta muy completa para proyectos de procesamiento de texto.

Sin embargo, es importante reconocer que TextBlob, al igual que cualquier herramienta de análisis de sentimientos, tiene sus limitaciones. El clasificador de Naïve Bayes, aunque eficaz en muchos casos, puede presentar dificultades cuando se enfrenta a textos que contienen jerga, lenguaje técnico o expresiones idiomáticas. Además, su rendimiento puede verse afectado si el texto de entrada difiere significativamente del corpus utilizado para entrenar el clasificador.

## **2.6 Métricas de evaluación**

En el análisis de sentimientos aplicado al estudio del fútbol argentino en plataformas digitales como Twitter, la elección de las métricas de evaluación es fundamental para asegurar la precisión y relevancia de los resultados obtenidos. La naturaleza apasionada del deporte y la variabilidad en la cantidad y tipo de expresiones en las redes sociales requieren métricas que no solo midan la exactitud de las clasificaciones, sino que también reflejen la complejidad del sentimiento del público. A continuación, se detallan las métricas consideradas en este estudio:

*Accuracy* es una medida que indica el porcentaje total de clasificaciones correctas. No obstante, en el contexto del fútbol argentino, donde las emociones pueden ser extremadamente polarizadas y los volúmenes de tweets pueden diferir significativamente entre equipos, esta métrica puede ser engañosa. Un ejemplo claro sería si la mayoría de los tweets sobre un equipo determinado son positivos; un modelo que siempre predijera "positivo" podría alcanzar una alta precisión, pero fallaría en detectar cualquier sentimiento negativo hacia ese equipo.

*Precisión* se refiere a la proporción de tweets clasificados como positivos que realmente son positivos. En el ámbito del fútbol, una alta precisión es indicativa de que cuando el modelo clasifica un tweet como positivo hacia un equipo, es muy probable que la clasificación sea correcta. Esto es esencial para identificar correctamente el entusiasmo y apoyo hacia los equipos.

*Recall*, o sensibilidad, mide la proporción de tweets positivos que el modelo detecta correctamente. Es decir, de todos los tweets que realmente son positivos, ¿cuántos el modelo clasifica como tales? Un alto recall indica que el modelo está capturando adecuadamente la mayoría de los sentimientos positivos expresados hacia los equipos.

*F1 Score* es la media armónica entre precisión y recall, proporcionando un balance entre ambas métricas. Esta es especialmente útil cuando se trabaja con datos desbalanceados, una situación común en el análisis de sentimientos en redes sociales donde no hay una igual distribución entre tweets positivos y negativos. Un F1 Score cercano a 1 sugiere que el modelo tiene un buen desempeño tanto en precisión como en recall.

## Desarrollo

### 3.1 Metodología

Este trabajo propone un enfoque sistemático para el desarrollo de un modelo de análisis de sentimientos aplicado a tweets relacionados con el fútbol argentino. Dada la imposibilidad de acceder a datos reales debido a restricciones en la API de Twitter, se diseñó una prueba de concepto que incluye la generación de un dataset sintético, su procesamiento y limpieza, y la implementación de un modelo basado en transformers para realizar el análisis de sentimientos.

La metodología se estructura en cuatro etapas principales. En la primera etapa, se genera un conjunto de datos sintético que permite simular tweets a partir de frases modelo enriquecidas dinámicamente con nombres de equipos, jugadores y eventos futbolísticos. Cada tweet se clasifica manualmente con etiquetas de sentimiento (positivo, negativo o neutral) y se complementa con metadatos adicionales como la fecha de creación, el número de "me gusta" y la fuente del tweet.

En la segunda etapa, se realiza un preprocesamiento exhaustivo del texto para garantizar su calidad y estructura. Este proceso incluye pasos como la eliminación de caracteres especiales y stopwords, la normalización de los textos a minúsculas, la tokenización y la lematización. Además, se calcula la longitud de cada tweet mediante una nueva columna que proporciona información útil sobre las características de los datos.

La tercera etapa se centra en la exploración y visualización del dataset. Se analizan las distribuciones de las etiquetas de sentimiento y las longitudes de los tweets, además de generar nubes de palabras que destacan los términos más frecuentes en el conjunto de datos. Estas visualizaciones permiten validar la calidad del dataset y descubrir patrones iniciales que pueden ser relevantes para el análisis de sentimientos.

Finalmente, en la cuarta etapa, se emplea la biblioteca Transformers de Hugging Face para implementar un modelo preentrenado que pueda analizar los sentimientos de los tweets. Este modelo es integrado en el pipeline para evaluar su desempeño en la clasificación de sentimientos y explorar su capacidad para generalizar patrones a partir de los datos sintéticos generados.

Esta metodología no solo valida la viabilidad del enfoque propuesto, sino que también proporciona un marco reproducible que puede adaptarse a escenarios con datos reales en futuras implementaciones.

Otra cuestión que debe tenerse en cuenta es que la idea fue realizar un análisis lo más detallado posible con la data sintética para que en un futuro, si puedo contar con el acceso, no deba cambiar mucho código porque lo que esto es una idea de negocio fructífera aún no utilizada en este fútbol argentino donde los sentimientos suelen bastantes pasionales ya sea positivos o negativos.



Figura 4: Diagrama de flujo del trabajo. Imágen de confección propia.

### 3.2 Recolección de los datos

Como mencioné anteriormente, la idea en un principio era realizar este análisis de sentimientos con datos reales pero debido a un cambio en las políticas de privacidad por parte de Elon Musk, se necesita un acceso developer premium el cual en este momento no estoy para afrontarlo. Lejos de rendirme, arme un dataset sintético en base a un dataset real el cual mencione antes y está adjunto (Anexo 1).

El dataset fue creado utilizando la librería **faker**, generando 3,000 registros que combinan frases modelo con nombres de equipos, jugadores y eventos relevantes. Estas frases fueron clasificadas manualmente con etiquetas de sentimiento (positivo, negativo o neutral) y

enriquecidas con metadatos como la fecha de publicación, el número de "likes" y la fuente desde donde supuestamente se emitió el tweet (Twitter for Iphone, Android o Web App). Por ejemplo, una frase como *"El arbitraje fue un desastre, favorecieron a {team}"* pudo convertirse en *"El arbitraje fue un desastre, favorecieron a River Plate"*, lo que añade dinamismo y diversidad al conjunto de datos.

Una vez generado, el dataset fue cargado en un entorno de análisis mediante **pandas**. A partir de allí, se inició un proceso de limpieza y transformación para garantizar que los textos estuvieran libres de ruido y listos para ser utilizados en las etapas posteriores.

### 3.3 Limpieza y preprocesamiento de los tweets

El proceso comenzó con la limpieza de ruido textual (si bien en el dataset sintético no lo tenemos, en el mundo real si y como dije antes, este análisis está hecho como si la data fuera real). También se eliminaron puntuaciones innecesarias y se normalizó el texto transformándolo a minúsculas, lo que garantizó que términos como "Gol" y "gol" se trataran como equivalentes. Además, se revisó que no existieran registros duplicados, evitando cualquier sesgo en el análisis.

En la etapa de preprocesamiento, se realizaron varias transformaciones clave utilizando herramientas avanzadas como **NLTK** y **SpaCy**:

1. **Eliminación de stopwords:** Se eliminaron palabras comunes como "de", "la", "el" y "y" que, aunque frecuentes, no aportan significado relevante para el análisis. Esto permitió reducir el peso innecesario en los datos.
2. **Tokenización:** Cada tweet se dividió en palabras individuales, conocidas como tokens. Este paso facilitó la descomposición de las frases en unidades mínimas de análisis. Por ejemplo, el texto *"El arbitraje fue un desastre"* se convirtió en ["el", "arbitraje", "fue", "un", "desastre"].
3. **Lematización:** Cada palabra se redujo a su forma base o lema. Esto permitió uniformar variantes de una misma palabra, como "jugando", "jugador" y "jugaron", tratándolas todas como "jugar". La lematización fue realizada con **SpaCy**, una herramienta eficiente para este tipo de operaciones.

Por último, se añadió una columna adicional llamada **Tweet Length**, que calculó automáticamente la longitud de cada tweet en caracteres. Esta métrica resultó útil para

identificar patrones relacionados con la extensión del texto y su posible relación con los sentimientos expresados, proporcionando un insight valioso como ¿Es mejor un tweet largo o corto? Esto lo vemos después en un gráfico que confeccione que relaciona esta columna con los likes.

El resultado de todas estas operaciones fue un dataset limpio y enriquecido, listo para ser utilizado en las siguientes etapas del proyecto.

### 3.4 Análisis exploratorio de los datos (EDA)

El análisis exploratorio de datos (EDA, por sus siglas en inglés) permite comprender en profundidad las características de un dataset y también permite validar su calidad antes de proceder a las etapas de modelado. El objetivo principal de esta fase fue identificar patrones, evaluar la distribución de las etiquetas y descubrir posibles irregularidades o sesgos en los datos.

Una de las primeras tareas realizadas en el EDA fue analizar la distribución de los likes y de la longitud de los tweets (columna agregada), algo básico como para tener un primer overview del dataset generado.

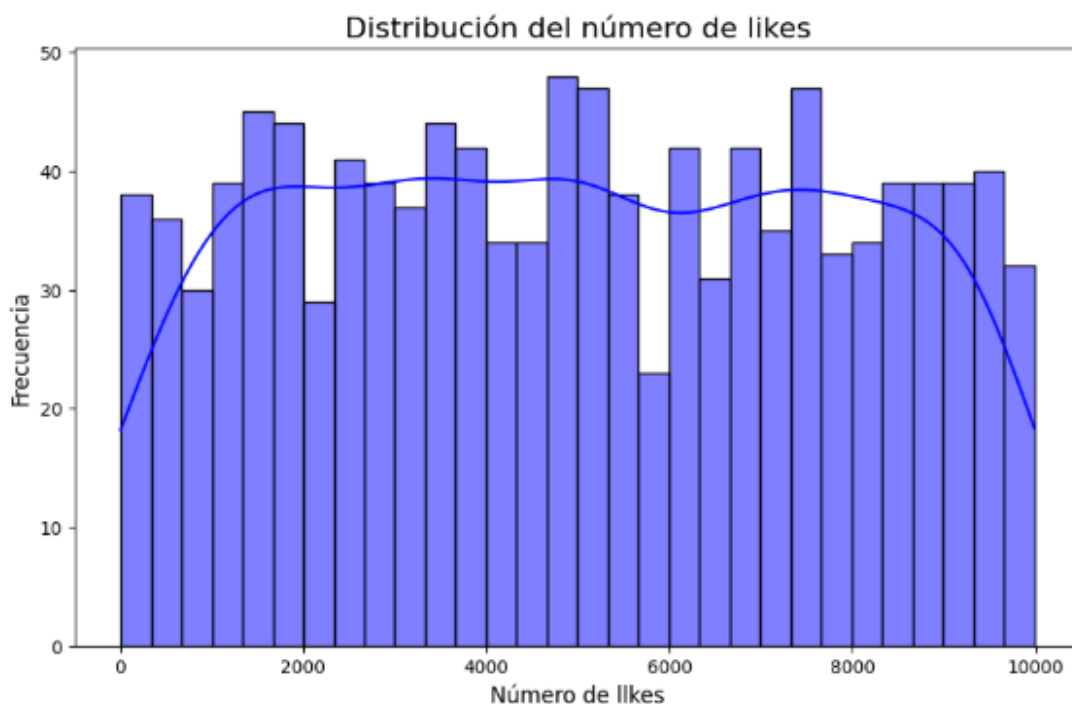


Figura 5: Distribución de los likes en el dataset. Imágen de confección propia.



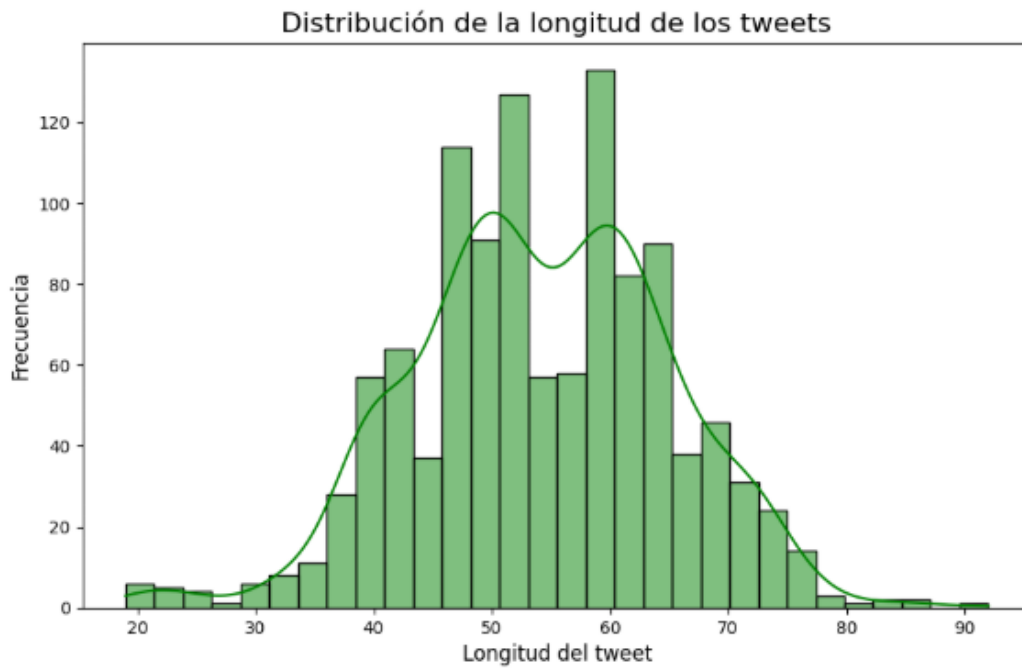


Figura 6: Distribución de los caracteres de los tweets del dataset. Imágen de confección propia.

Con estas dos visualizaciones básicas podemos ahora sí comenzar a realizarnos preguntas para tratar de mostrar o identificar ciertos patrones. Usando estas dos distribuciones podemos preguntarnos; **¿Qué prefieren los hinchas leer, tweets largos o cortos?**

Esto puede responderse fácilmente unificando las dos figuras mostradas anteriormente utilizando un **scatter plot** (gráfico de dispersión).

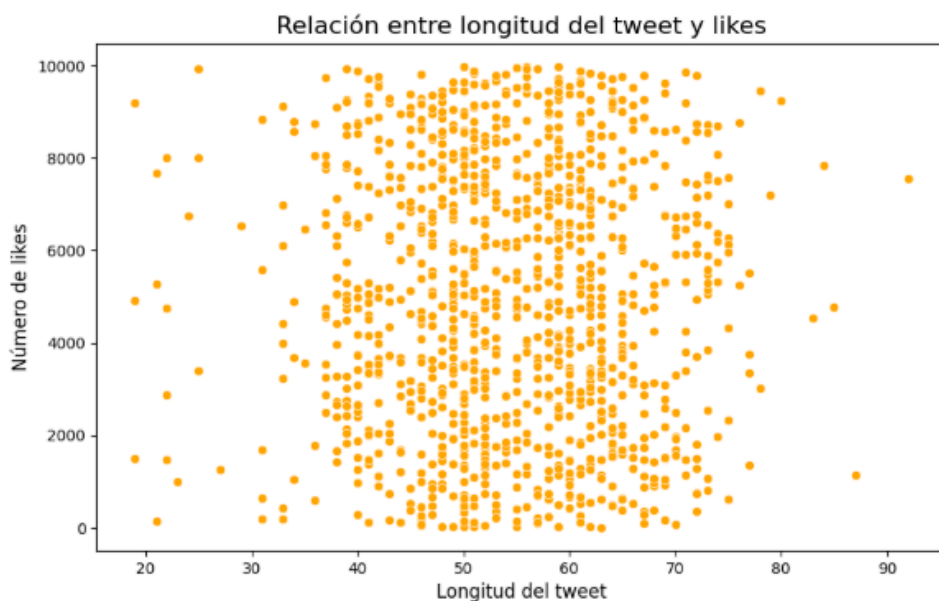
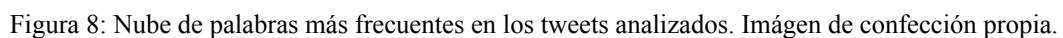


Figura 7: Relación entre las columnas "Tweet Length" y "Number of Likes". Imágen de confección propia.

El objetivo principal de este trabajo como se mencionó antes es analizar qué sienten o que reflejan los hinchas y usar esos sentimientos como un medio de análisis. Como un primer paso para esto, resulta importante utilizar una **WordCloud** (nube de palabras) a fin de ver las palabras más frecuentes en los tweets que estamos analizando. Acá es donde entra el juego el preprocesamiento y todo lo mencionado anteriormente debido a que si eso no se hacía correctamente, lo que veríamos ahora serían artículos y la verdad que un artículo como “la”, “el”, etc. no genera valor.



26

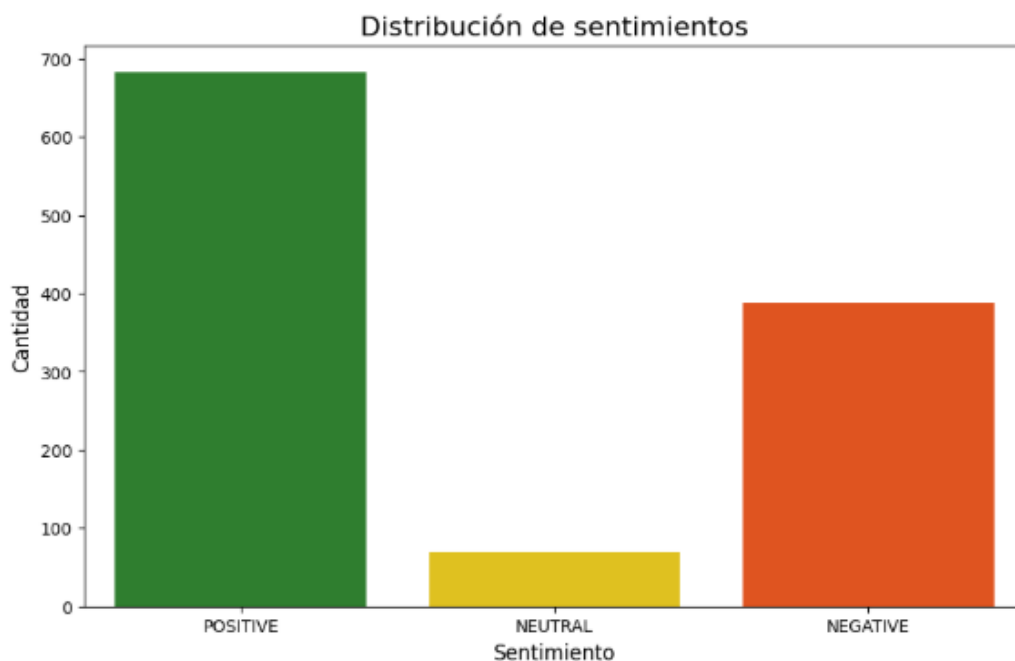


Figura 9: Distribución de los sentimientos. Imágen de confección propia.

Gracias a la columna “Sentiment” (etiquetada manualmente) podemos analizar la distribución de los tweets. Tiene mucho sentido el hecho de que no haya muchos tweets de carácter neutral debido a que cuando una persona entra a Twitter a transmitir un sentimiento, lo hace para expresar dos sentimientos, **alegría** y/o **descontento**. Twitter debido a las pocas limitaciones de contenido, permite ver muchísimos tweets de carácter muy negativo los cuales expresan frustración, lamentos, CRÍTICAS, etc. Y en consecuencia hay otros hinchas que se identifican con ese sentimiento o pensamiento y le dan like, lo que refleja que de alguna manera están de acuerdo con eso que otro hincha escribió. Si la crítica va hacia el director técnico de un equipo, el mal estado del campo de juego, los jugadores, el formato del campeonato, entre muchas otras cosas, **¿no estaría bueno que los clubes de fútbol y la AFA sepan eso para de alguna manera evitar posibles conflictos y mejorar lo que se está haciendo mal?** Así también se expresó esto, sería bueno también que sepan lo que se está haciendo bien para seguir replicando o teniéndolo en cuenta antes de tomar una decisión que empeore las cosas. Es por esto que este MVP es una manera innovadora y super útil en este país tan lindo pero tan pasional y esa pasión a veces lleva a ciertos episodios de violencia que podrían tranquilamente evitarse (Anexo 2).

Pero el gráfico de distribución si bien nos ayuda a tener un panorama general, no dice mucho como para hacer una análisis aún más profundo, esto se hace mediante WordClouds como dije antes pero no una con todas las palabras en general sino más bien una segmentada por

sentimientos y por color (se creó una paleta personalizada para que las visualizaciones sean más atractivas visualmente)



Figura 10: Palabras positivas. Imágen de confección propia.



Figura 11: Palabras neutrales. Imágen de confección propia.

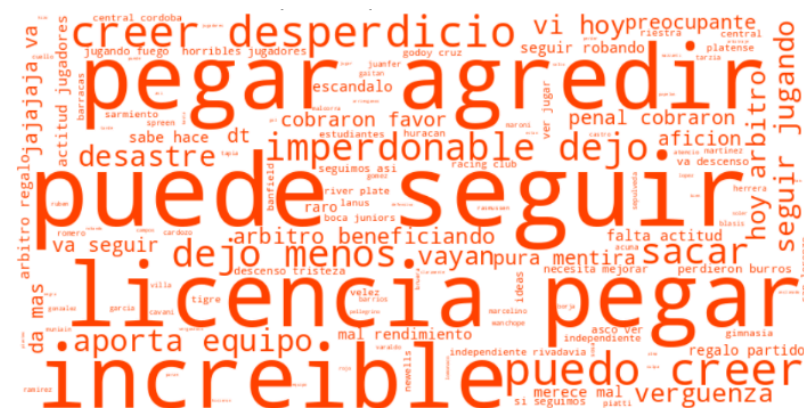


Figura 12: Palabras negativas. Imágen de confección propia.

Con un buena limpieza, un preprocesamiento que agregue valor y un análisis exploratorio podemos llegar a eso, dividir las palabras por sentimiento y en base a eso poder sacar insights. Las palabras positivas destacan jugadas del partido como “pase”, “magia”, las neutrales “camiseta”, “transpirar” y en las negativas jugadas extrafutbolísticas como pueden ser “agredir” y “licencia” que en la jerga futbolera, hace referencia a que un jugador puede hacer lo que quiera sin ser sancionado.

Inicialmente, el análisis de sentimientos se llevó a cabo mediante un etiquetado manual de los tweets. Si bien este enfoque permitió obtener resultados precisos y comprensibles, su naturaleza intensiva en recursos humanos lo convirtió en un proceso lento y tedioso. La creciente popularidad de plataformas como Twitter, caracterizada por un volumen exponencial de publicaciones, evidenció la necesidad de desarrollar métodos más eficientes para realizar este tipo de análisis. Surge así la siguiente interrogante: **¿Existen técnicas de procesamiento del lenguaje natural que permitan automatizar el etiquetado de sentimientos, garantizando al mismo tiempo un alto nivel de precisión?**

### 3.5 Análisis de sentimientos con Transformer

Este etiquetado manual se optimiza con modelos de machine learning, en este caso un transformer;

#### Configuración del análisis

El pipeline de análisis de sentimientos se configuró cargando el modelo *nlptown/bert-base-multilingual-uncased-sentiment* y utilizando una función personalizada para mapear las etiquetas predichas del modelo en tres categorías simplificadas: **NEGATIVE**, **NEUTRAL** y **POSITIVE**. Este mapeo se realizó mediante un diccionario que agrupaba las etiquetas de estrellas devueltas por el modelo en sentimientos generales, como se muestra a continuación:

- **1 y 2 estrellas:** Se asignaron a la etiqueta **NEGATIVE**.
- **3 estrellas:** Corresponden a **NEUTRAL**.
- **4 y 5 estrellas:** Se asignaron a la etiqueta **POSITIVE**.

Cada tweet procesado fue enviado al pipeline, y las predicciones resultantes fueron almacenadas en una nueva columna denominada **Transformer\_Sentiment** dentro del dataset.

Este proceso garantizó una integración fluida entre las etiquetas predichas por el modelo y el análisis posterior.

### **Comparativa entre análisis simple y Transformer**

Para evaluar el desempeño del modelo, se realizó una comparación con un análisis de sentimientos básico basado en palabras clave. Los datasets provenientes de ambos métodos fueron combinados utilizando la columna **Processed Tweet**, lo que permitió una comparación directa de las etiquetas asignadas por cada enfoque.

1. **Distribución de sentimientos:** Se generaron gráficos que mostraban la distribución de las etiquetas de sentimiento producidas por ambos métodos. Los resultados destacaron diferencias significativas en la asignación de etiquetas, con el modelo Transformer mostrando una mayor sensibilidad hacia tweets más complejos en comparación con el análisis simple.

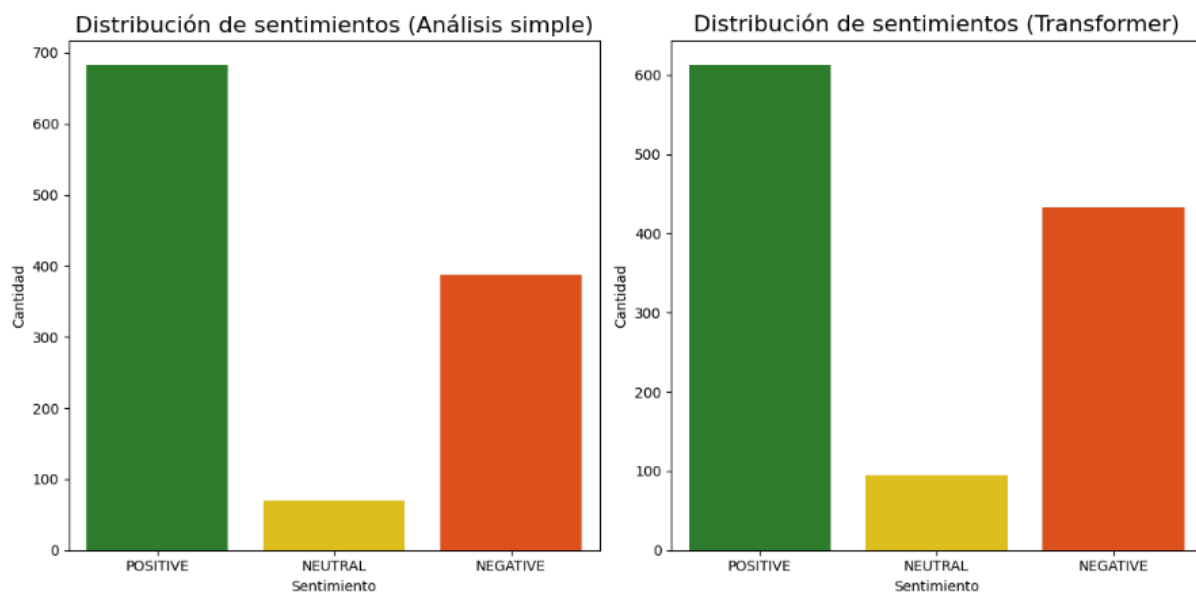


Figura 13: Comparación de análisis de sentimientos simples vs Transformer. Imágen de confección propia.

2. **Coincidencias y diferencias:** Se calculó el porcentaje de coincidencia entre las etiquetas asignadas por ambos métodos. Los resultados mostraron una tasa de coincidencia del **73,79%**, lo que indica que, aunque ambos métodos tienen enfoques diferentes, existe un alto grado de concordancia. Las diferencias se atribuyeron

principalmente a tweets con lenguaje más ambiguo o matices que solo el Transformer pudo captar.

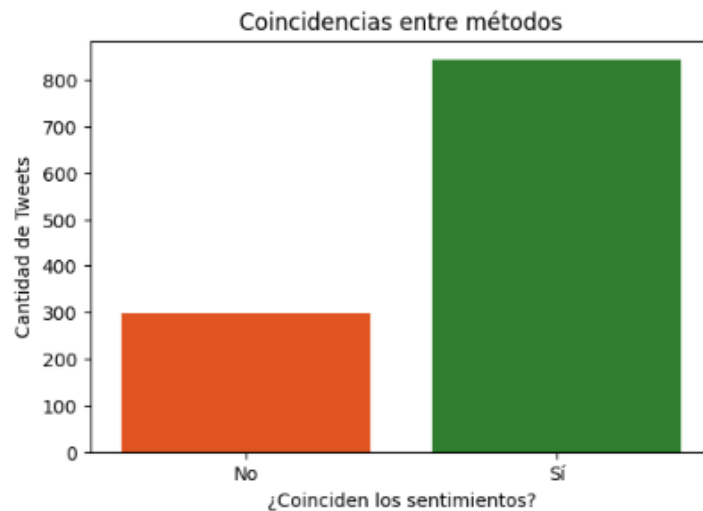


Figura 14: Diferencia entre métodos de análisis de sentimientos. Imágen de confección propia.

### 3.6 Resultados del análisis

Para explorar cómo la adición de nuevos datos impacta en el desempeño del modelo, se generaron 100 tweets adicionales utilizando frases modelo enriquecidas con nombres de equipos y jugadores distintos a los del dataset original. Estos nuevos tweets fueron sometidos a las mismas etapas de limpieza y preprocesamiento que los datos originales, incluyendo la eliminación de caracteres especiales, normalización, eliminación de stopwords, lematización y cálculo de la longitud de los tweets.

Una vez procesados, los nuevos tweets fueron clasificados utilizando el modelo Transformer, y sus resultados se combinaron con los del dataset original para una evaluación conjunta. Se analizaron métricas como **Accuracy**, **Precision**, **Recall** y **F1-Score**, comparando el rendimiento del modelo antes y después de la incorporación de los nuevos datos.

#### Métricas de desempeño

Los resultados mostraron que la combinación del dataset original con los nuevos tweets mejoró ligeramente el desempeño del modelo:

- **Accuracy:** Incrementó de 74% a 75%.
- **Precision:** Incrementó de 72% a 73%.

- **Recall:** Incrementó de 73% a 74%.
- **F1-Score:** Incrementó de 72% a 74%.

Aunque las mejoras no fueron drásticas, indican que el modelo es capaz de generalizar mejor con un conjunto de datos ampliado.

### Curva ROC

Para validar el desempeño del modelo en la clasificación multiclase, se generaron curvas ROC para las tres clases principales;

- **Clase 0 = NEGATIVE**
- **Clase 1 = NEUTRAL**
- **Clase 2 = POSITIVE**

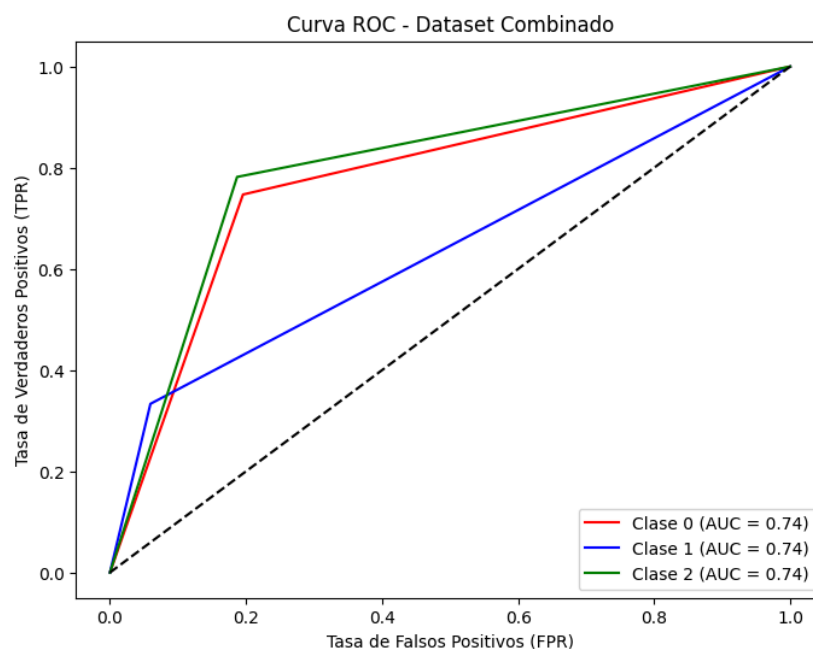


Figura 15: Curva ROC. Imágen de confección propia.

Todas las clases tienen un AUC de **0.74**, lo que refleja un desempeño consistente en la clasificación. Las tres curvas siguen una tendencia similar, básicamente el modelo no tiene un sesgo particular hacia ninguna clase, sino que su desempeño es equilibrado en todas ellas. Esto sugiere que el modelo generaliza bien en el conjunto combinado, pero que la introducción de datos más específicos o un ajuste fino en el dominio podría mejorar su capacidad para captar matices emocionales más sutiles. Si nos ponemos más analistas, tiene sentido que lo más difícil de clasificar sea lo neutral porque este modelo es automatizado



pero no está maximizado y siempre se juega para un lado ya sea positivo o negativo. Sumado a eso y como dije en secciones anteriores, generalmente la gente no entra a twitter y escribe algo neutral, siempre escribe algo bueno o algo malo.

### 3.7 Comparación con otros modelos

Con el fin de evaluar la solidez de la aproximación propuesta basada en el modelo *nlptown/bert-base-multilingual-uncased-sentiment*, se llevó a cabo una comparación exhaustiva con dos métodos alternativos de análisis de sentimiento: VADER Y TextBlob. El objetivo de esta comparativa fue determinar en qué medida la elección de BERT como núcleo del MVP era adecuada y justificar su selección como solución central.

En primer lugar, se analizaron los resultados obtenidos por VADER y TextBlob, herramientas comúnmente empleadas en tareas de análisis de sentimiento gracias a su simplicidad e independencia de un entrenamiento previo extenso. Estos métodos, sin embargo, se basan principalmente en diccionarios de polaridad y reglas heurísticas, lo que limita su capacidad para comprender matices semánticos, contexto cultural o ironías tan frecuentes en el lenguaje coloquial utilizado por los aficionados al fútbol. Al aplicar ambas herramientas sobre la colección de tuits del fútbol argentino, se observó un desempeño marcadamente inferior en términos de exactitud, precisión, exhaustividad y puntuación F1. Esta brecha se explica por la incapacidad de VADER y TextBlob para capturar el contexto completo y las múltiples variantes del lenguaje futbolero, incluyendo el uso del lunfardo, términos propios de la jerga futbolística y expresiones regionales.

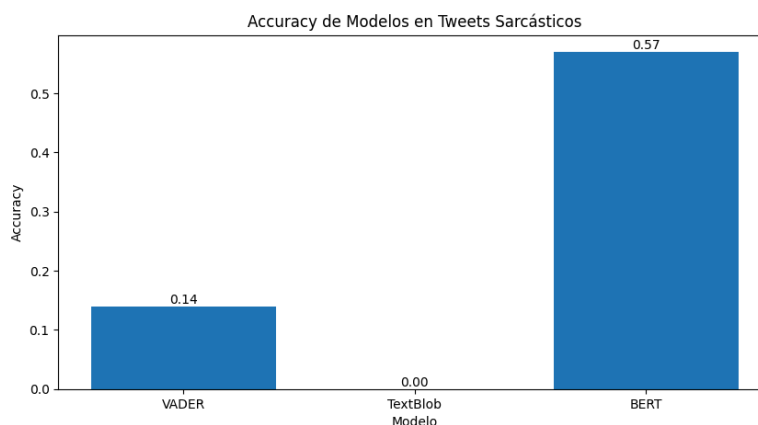


Figura 16: Análisis de sarcasmo. Imágen de confección propia.

En este gráfico se puede apreciar la diferencia para captar sarcasmo en el modelo BERT implementado. Textblob por el contrario, no logra identificar sarcasmo debido a que se basa en diccionarios de palabras con puntuaciones de polaridad. El sarcasmo, sin embargo, a menudo implica un uso irónico de palabras que, por sí solas, podrían tener una connotación positiva o negativa. TextBlob no está equipado para comprender este nivel de contexto y doble sentido. El sarcasmo suele depender del contexto de la conversación, las relaciones entre los hablantes y la situación comunicativa. TextBlob no tiene la capacidad de analizar el contexto más amplio de un tweet, lo que limita su habilidad para identificar el sarcasmo. TextBlob es una herramienta más simple en comparación con modelos más avanzados como BERT. Este último utiliza arquitecturas de transformadores que le permiten capturar relaciones semánticas más profundas y entender el lenguaje de manera más similar a los humanos.

Hablando de VADER, está diseñado principalmente para detectar emociones básicas como alegría, tristeza, ira, miedo, sorpresa y neutralidad. El sarcasmo, al ser más sutil y contextual, puede no encajar perfectamente en estas categorías. Puede tener dificultades para identificar la negación y el contexto en el que se expresa el sarcasmo, lo cual es crucial para su detección. Además VADER se basa en un léxico de palabras clave y sus puntuaciones asociadas. Esto puede ser limitante, ya que el sarcasmo a menudo se expresa de forma más sutil y no siempre involucra palabras claramente negativas o positivas. A modo de ejemplificación un tweet que incluye una carcajada “*Jaja no expulsaron a Herrera*”, Vader y BERT lo clasifican correctamente como NEGATIVE, pero TextBlob lo clasifica como neutral. Y un ejemplo más “*Malcorra es un crack, sobre todo para errar goles cantados.*” El único modelo en identificar el sarcasmo es BERT, el modelo implementado en el MVP, los otros dos modelos lo clasifican como neutral.

En entornos altamente dinámicos como el fútbol, donde las narrativas cambian con frecuencia (nuevos jugadores, hechos relevantes, cánticos, expresiones y contextos informativos muy variables), la capacidad de adaptación y la flexibilidad del modelo resultan fundamentales.

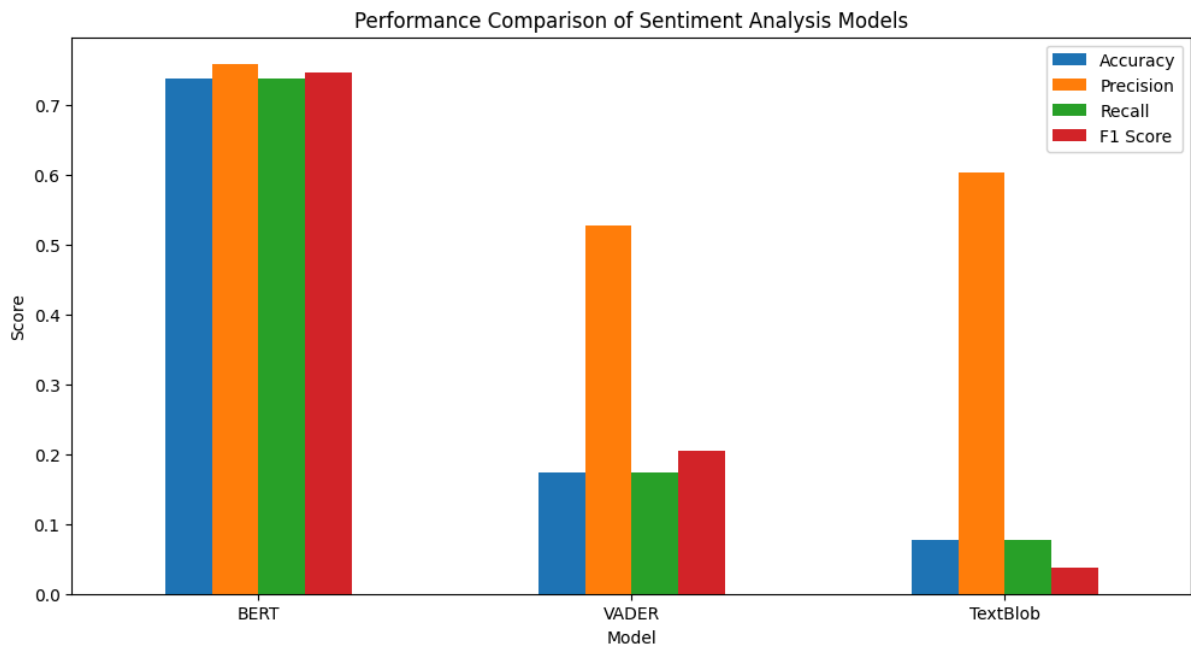


Figura 17: Comparativa de diferentes modelos. Imágen de confección propia.

El enfoque basado en `nlptown/bert-base-multilingual-uncased-sentiment` logró un equilibrio sólido entre la complejidad del problema y el desempeño alcanzado, superando a los métodos basados en léxicos como VADER y TextBlob, que no capturan dependencias contextuales. La interpretación de las métricas revela que BERT tiene una ventaja significativa: su alta precisión minimiza los falsos positivos, lo cual es fundamental para evitar estrategias erróneas basadas en clasificaciones incorrectas de tweets neutrales. Su recall, por otro lado, es vital para capturar todos los sentimientos relevantes, especialmente en situaciones críticas como la identificación de reacciones negativas durante una crisis. Mientras que la accuracy muestra cómo BERT maneja bien la clasificación general, aunque puede ser engañosa en datos desbalanceados, el F1 Score subraya su robustez al equilibrar precisión y recall, mostrando que BERT es capaz de entender y clasificar correctamente las emociones en el contexto específico del fútbol argentino, captando matices semánticos y culturales que otros modelos no pueden.

Considerando el contexto futbolístico argentino, caracterizado por un lenguaje dinámico, con abundantes expresiones coloquiales, jerga futbolera en general y referencias culturales complejas, la capacidad de BERT para comprender y adaptarse a estos matices destaca ampliamente. La robustez que presenta este modelo al enfrentar vocabulario cambiante y la posibilidad de mantener un buen rendimiento sin una ingeniería de características exhaustiva resultan en un enfoque más escalable y sostenible a largo plazo. Además, la superioridad

evidenciada en las métricas clave, junto con su flexibilidad para incorporar nuevos escenarios y datos, justifican plenamente la elección de BERT como eje del MVP de análisis de sentimiento en el ámbito del fútbol argentino, superando las limitaciones observadas en las alternativas evaluadas.

En el gráfico se observan las métricas Accuracy, Precision, Recall y F1 Score de los cuatro enfoques utilizados para clasificar los sentimientos en el contexto del fútbol argentino. La altura de las barras refleja la calidad de cada modelo en la tarea de análisis de sentimiento. Se puede apreciar claramente la superioridad del enfoque basado en BERT, que supera significativamente a los métodos léxicos como VADER y TextBlob. Sin embargo, es importante señalar que el modelo BERT también presenta sus propias limitaciones. Entre estas, destacan los altos requerimientos computacionales que pueden hacer que su implementación sea costosa en términos de recursos. Además, BERT requiere un proceso de ajuste fino (fine-tuning) con datos específicos para alcanzar su máximo potencial en dominios particulares como el del fútbol argentino. Este proceso no solo implica tiempo adicional para la configuración y entrenamiento del modelo sino que también demanda un volumen significativo de datos de alta calidad para lograr un ajuste óptimo. Por último, el procesamiento de grandes volúmenes de datos con BERT puede resultar en tiempos de respuesta más largos, lo cual es un factor crítico en aplicaciones que requieren análisis en tiempo real o cerca del mismo. Estos aspectos deben considerarse al evaluar la viabilidad y eficiencia de implementar BERT en contextos prácticos.

## Dashboard

Para ejemplificar lo que podría ser un análisis usando este modelo, adjunto un dashboard (Anexo 3) y un posterior análisis al mismo.

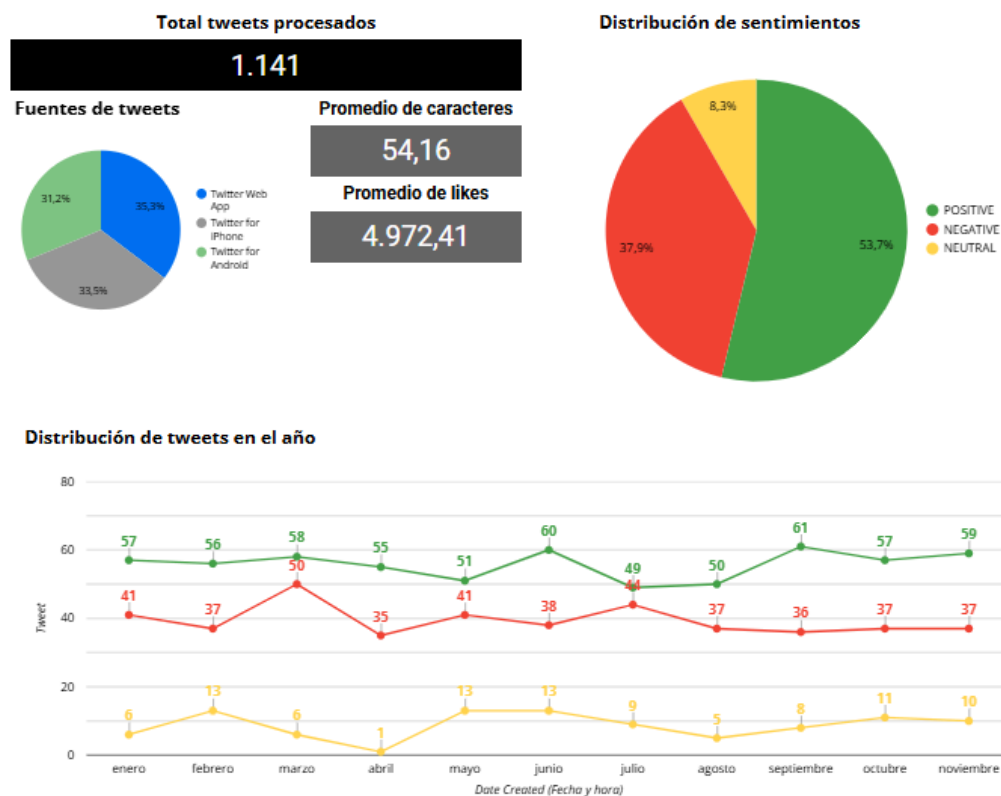


Figura 18: Dashboard de negocio. Imágen de confección propia en anexo 3.

Aunque los datos son sintéticos, podemos utilizarlos como ejemplo para derivar **KPIs de negocio** y posibles **insights** aplicables en un entorno real.

Cuando hablamos de un 54% de mensajes positivos, queda claro que el fútbol sigue siendo una fuente inagotable de alegría, pasión y orgullo, muchas veces motivado por triunfos, jugadas épicas, un golazo inesperado o esa conexión única entre los jugadores y su gente. Pero, ojo, porque ese 38% de sentimientos negativos también tiene su peso, y seguramente esté relacionado con derrotas dolorosas, arbitrajes polémicos, “robos” que se sienten en carne propia o decisiones institucionales que no terminan de cerrar.

Para los clubes, este tipo de información es oro puro. Saber qué hace feliz o qué enoja a sus hinchas puede ser la diferencia entre ganarse el corazón de la hinchada o perderla en medio de un mar de críticas. Después de una derrota dura o un papelón dirigenal, salir a dar la cara

con un mensaje auténtico, reconocer errores o demostrar compromiso puede ayudar a calmar las aguas. Al mismo tiempo, identificar lo que genera alegría, como un clásico ganado o el debut de un pibe de las inferiores, puede ser la oportunidad perfecta para reforzar ese sentido de pertenencia que hace que el hincha siga al equipo "en las buenas y en las malas".

Por el lado de la AFA, este tipo de análisis debería ser una herramienta básica para tomar decisiones que no solo apunten a la organización de los torneos, sino también al corazón de los hinchas. Los formatos confusos, los cambios de último momento o los arbitrajes que generan sospechas no solo generan memes en las redes, sino también un malestar profundo que afecta la confianza en el torneo. Si la AFA empieza a leer mejor estas señales, podría evitar muchas de las críticas que suelen lloverle semana tras semana. Además, los momentos de mayor actividad positiva en redes, como el inicio de los torneos, las finales o los mercados de pases, son una oportunidad inmejorable para reforzar la imagen del fútbol argentino como uno de los más pasionales del mundo. Es en esos momentos donde se puede conectar con la hinchada desde un lugar de orgullo y unidad, mostrándoles que, a pesar de todo, seguimos teniendo "el fútbol más lindo del planeta".

## Proyecciones

Este proyecto ha sentado las bases para desarrollar un sistema sólido y eficiente de análisis de sentimientos aplicable al dominio del fútbol argentino. Sin embargo, para alcanzar su máximo potencial, es necesario avanzar hacia el uso de datos reales y la implementación de técnicas más avanzadas que permitan superar las limitaciones actuales y mejorar significativamente el desempeño del modelo.

- **Lograr acceso a la API de Twitter**, lo que permitiría recolectar tweets auténticos en tiempo real. Este acceso no solo brindaría un flujo constante de datos representativos del lenguaje y las emociones de los hinchas, sino que también habilitaría la creación de un dataset mucho más diverso y enriquecido.
- **Mejora del modelo Transformer**: Una vez que se disponga de datos reales, realizar “fine-tuning” para que el modelo ya realizado pueda adaptarse mejor al contexto y mejorar su capacidad para captar matices emocionales complejos, como el sarcasmo y las expresiones irónicas. Por ejemplo, un tweet como *"El árbitro siempre nos favorece (irónico)"* podría ser interpretado correctamente como negativo.
- **Uso de Word2Vec**: Entrenado con datos del dominio futbolístico, podría mejorar la comprensión contextual del modelo. Permitiría capturar relaciones semánticas específicas entre palabras y expresiones comunes en este ámbito, enriqueciendo aún más la capacidad del modelo para interpretar correctamente los tweets.

## Conclusión

El análisis de sentimientos sobre el fútbol argentino en la plataforma Twitter ha sido un ejercicio revelador que no solo ha capturado la intensidad emocional de los hinchas sino que también ha proporcionado un marco sólido para entender cómo la tecnología puede transformar la comunicación en el deporte. Este proyecto, a través de la implementación de un MVP, ha explorado diversas facetas del análisis de datos, desde la recolección y preprocesamiento hasta la evaluación comparativa de modelos de aprendizaje automático para el análisis de sentimientos.

El fútbol argentino, con su pasión desbordante, genera un caudal inmenso de opiniones en las redes sociales, especialmente en Twitter. La problemática central que este MVP ha abordado es la dispersión y desorganización de estas opiniones, que dificultan su utilización por parte de clubes, medios de comunicación, y patrocinadores. La solución propuesta fue la creación de un sistema que pudiera analizar automáticamente estos sentimientos, proporcionando insights valiosos para la toma de decisiones estratégicas. Debido a limitaciones en el acceso a la API de Twitter para datos en tiempo real, se optó por la generación de un dataset sintético, basado en el contexto del fútbol argentino, que simulara de manera realista la variedad de opiniones y reacciones de los hinchas.

El análisis exploratorio del dataset permitió identificar patrones relevantes, como la preferencia de los usuarios por tweets cortos e impactantes, lo que se refleja en la relación entre la longitud de los tweets y el número de likes. Las nubes de palabras por sentimiento destacaron términos clave asociados con emociones positivas, negativas y neutras, evidenciando cómo el lenguaje futbolístico refleja la intensidad emocional del público en Twitter. Los picos de actividad emocional correlacionados con eventos deportivos significativos resaltan el potencial del modelo para proporcionar insights valiosos, que pueden ser utilizados para optimizar estrategias de comunicación y fortalecer la conexión emocional entre clubes, hinchas y marcas.

En cuanto al análisis de sentimientos, la comparación entre métodos simples y aquellos basados en machine learning, particularmente con el uso de Transformers como BERT, demostró diferencias significativas en precisión y capacidad de captar matices del lenguaje. La precisión de BERT superó a las metodologías tradicionales como VADER y TextBlob,



especialmente en la detección de sarcasmo, lo cual es crucial en el contexto de los comentarios deportivos donde el tono puede ser ambiguo.

Estos hallazgos no solo se alinean con el objetivo del proyecto de aplicar técnicas de NLP para entender y clasificar los sentimientos en el contexto del fútbol argentino, sino que también proporcionan herramientas para una mejor gestión de la comunicación y estrategia deportiva. La capacidad de procesar y analizar grandes volúmenes de datos de Twitter en tiempo real abre la puerta a una toma de decisiones más informada por parte de clubes, asociaciones y marcas en el ámbito deportivo.

El núcleo del estudio fue la comparación de diferentes modelos de análisis de sentimientos, con resultados que resaltan varios hallazgos:

- *BERT*: Se destacó ampliamente en todas las métricas evaluadas, todas por encima de 70%, lo cual indica una alta precisión y capacidad para captar la complejidad emocional y contextual en los tweets relacionados con el fútbol argentino. Este modelo no solo es superior en términos de exactitud y precisión, sino que también logra un balance sobresaliente entre Recall y la Precision, reflejado en un alto F1 Score. BERT tiene una capacidad significativamente mayor para detectar sarcasmo, con un Accuracy de 57%. Este resultado subraya la versatilidad de BERT no solo en el análisis de sentimientos generales sino también en la interpretación de matices lingüísticos más sutiles como el sarcasmo, lo que es crucial en el contexto del fútbol donde la ironía y el sarcasmo son comunes.
- *VADER*: Con un Accuracy de 17% y un F1 Score de 20%, VADER evidenció limitaciones al tratar con el lenguaje en español, especialmente en el contexto argentino. Su aproximación basada en reglas y léxicos predefinidos no captura bien las sutilezas culturales y lingüísticas que caracterizan las discusiones futbolísticas, siendo menos efectivo en la detección de sarcasmo.
- *TextBlob*: La simplicidad de TextBlob para el análisis de sentimientos resultó en un Accuracy de 7% y un F1 Score de 4%. lo cual indica una capacidad limitada para adaptarse a las particularidades del español argentino en contextos de análisis de sentimientos. Su enfoque simplificado basado en reglas y léxicos no permite una interpretación eficaz de las complejidades emocionales y contextuales presentes en las conversaciones sobre fútbol, lo que se evidencia en su baja efectividad en el manejo del sarcasmo y la interpretación de emociones complejas.

Los hallazgos confirman que BERT no solo es el modelo más efectivo para el análisis general de sentimientos en el contexto del fútbol argentino sino que también es superior en capturar el sarcasmo y las complejidades lingüísticas. Esto lo convierte en una herramienta invaluable para el análisis en tiempo real de la opinión pública en redes sociales, no solo en el deporte sino en cualquier campo donde el lenguaje y la emoción juegan un papel crucial. La capacidad de BERT para entender y clasificar con precisión el sarcasmo y otros matices del lenguaje abre nuevas vías para su aplicación en la comunicación, el marketing, la política, y más allá, donde el sarcasmo y la ironía pueden influir significativamente en la percepción y reacción del público.

Este MVP no solo ha validado la capacidad técnica de los modelos Transformers en el análisis de sentimientos sino que también ha mostrado aplicaciones prácticas significativas. Para los clubes, el análisis de sentimientos puede guiar estrategias de comunicación y gestión de la imagen pública. Para los medios de comunicación, ofrece una herramienta para reportar de manera más informada sobre la percepción pública de eventos deportivos. Las marcas y patrocinadores pueden utilizar estos datos para medir el impacto de sus campañas y alinearse mejor con la comunidad de hinchas.

A modo de cierre para este trabajo final de graduación, este proyecto ha sido un paso significativo hacia la utilización de tecnologías avanzadas de inteligencia artificial para entender y responder a las dinámicas emocionales en el deporte. Ha demostrado cómo un dataset cuidadosamente curado y un modelo de aprendizaje automático bien seleccionado pueden transformar el ruido de las opiniones en datos accionables. El análisis detallado de diferentes modelos ha proporcionado insights sobre las limitaciones y las fortalezas de cada enfoque, resaltando la necesidad de modelos que no solo midan la polaridad sino que capturen la esencia de las emociones humanas en un contexto cultural específico. Este MVP abre la puerta a futuras investigaciones y aplicaciones que podrían beneficiarse de la integración de modelos más avanzados con datos en tiempo real, prometiendo una nueva era en la gestión y comprensión de la opinión pública en el deporte.

## Referencias

- [1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need*. *Advances in Neural Information Processing Systems*, 30. <https://arxiv.org/abs/1706.03762>
- [2] Vig, J. (2020, June 4). *Keeping up with the BERTs*. Towards Data Science. <https://towardsdatascience.com/keeping-up-with-the-berts-5b7beb92766>
- [3] Verma, M. (2020, August 18). *The art of tokenization: Breaking down text for AI*. Towards Data Science. <https://towardsdatascience.com/the-art-of-tokenization-breaking-down-text-for-ai-43c7bccae-d25>
- [4] DataScientest. (n.d.). *Introducción al procesamiento de lenguaje natural (NLP)*. <https://datascientest.com/es/nlp-introduccion>
- [5] TyC Sports. (2020, July 26). *Big data en el fútbol: qué es, su utilidad y cómo se aplica en la Argentina*. <https://www.tycsports.com/al-angulo/big-data-en-el-futbol-que-es-su-utilidad-y-como-se-aplica-en-la-argentina-20200726.html>
- [6] UNIR. (n.d.). *Big Data en el fútbol: Cómo se está revolucionando el deporte*. <https://www.unir.net/revista/ingenieria/big-data-futbol/>
- [7] Nussbaumer Knafllic, C. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley. <https://www.asrepayesh.com/assets/asrepayesh.com/repo/file/storytelling-with-data-cole-nussbaumer-knafllic.pdf>
- [8] Hugging Face. (n.d.). *Fine-tuning pretrained models*. Hugging Face NLP Course. <https://huggingface.co/learn/nlp-course/chapter6/5?fw=pt>
- [9] Chakravarthy, S. (2020, June 3). *Transformers: The foundation of modern NLP*. Towards Data Science. <https://towardsdatascience.com/transformers-89034557de14>
- [10] KiwiQA Services. (n.d.). *ChatGPT for test automation: A comprehensive guide*. <https://www.kiwiqa.com/chatgpt-for-test-automation/>
- [11] Learn Prompting. (n.d.). *Introduction to prompting*. Learn Prompting. <https://learnprompting.org/docs/intro>
- [12] DAIR AI. (n.d.). *Prompt engineering guide*. GitHub. <https://github.com/dair-ai/Prompt-Engineering-Guide>

- [14] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186.  
<https://doi.org/10.18653/v1/N19-1423>
- [15] Loria, S. (2018). *TextBlob: Simplified Text Processing [Software]*. Retrieved de <https://textblob.readthedocs.io/en/latest/>
- [16] Hutto, C. M., & Gilbert, E. E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international<sup>1</sup> conference on computational linguistics* (pp. 743-752). Association for Computational Linguistics.
- [17] Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, <sup>1</sup> 2(1), 37–63.

## Anexo

*1 - Trabajo inspiración “FIFA World Cup 2022”*

<https://www.kaggle.com/code/aks777sp/fifa-world-cup-day-1-tweets>

*2 - Presentación de negocio - Producto de datos*

[https://www.canva.com/design/DAGXoxqiBnA/6OtdbdHTFw9m3H04Yf6QXA/edit?utm\\_content=DAGXoxqiBnA&utm\\_campaign=designshare&utm\\_medium=link2&utm\\_source=sharebutton](https://www.canva.com/design/DAGXoxqiBnA/6OtdbdHTFw9m3H04Yf6QXA/edit?utm_content=DAGXoxqiBnA&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton)

*3 - Dashboard Looker Studio - Producto de datos*

<https://lookerstudio.google.com/reporting/7802e475-44b9-40a7-89f5-2fff74f3efaa>

*4 - Repositorio GitHub - Notebook y demás*

<https://github.com/Gianmazzocchi/TFG---LIAN>