



**Universidad
Europea**

UNIVERSIDAD EUROPEA DE MADRID

ESCUELA DE ARQUITECTURA, INGENIERÍA Y DISEÑO

MASTER UNIVERSITARIO EN ANALISIS DE DATOS MASIVOS (BIG DATA)

TRABAJO FIN DE MÁSTER

**Detección Automática de Sarcasmos en Textos
Mediante Modelos Basados en Transformers**

Ernesto Francisco Barrionuevo Aragonés

Dirigido por

Prof^a. María Cruz Gaya López

CURSO 2023-2024

Ernesto Francisco Barrionuevo Aragonés

TÍTULO: Detección automática de sarcasmos en textos utilizando modelos basados en Transformers

AUTOR: Ernesto Francisco Barrionuevo Aragonés

TITULACIÓN: Master en Big Data

DIRECTOR/ES DEL PROYECTO: María Cruz Gaya López

FECHA: Octubre de 2024

RESUMEN

Este proyecto aborda el problema de la detección automática de sarcasmos en español, un desafío complejo debido a la sutileza y el contexto necesario para identificar este tipo de lenguaje. La mayoría de los avances en procesamiento de lenguaje natural se han centrado en el inglés, dejando una brecha importante en otros idiomas como el español. La principal contribución de este trabajo ha sido el desarrollo de un modelo basado en Transformers, adaptado mediante fine-tuning a un conjunto de datos en español, que incluye titulares de noticias, diálogos de series y frases generadas por inteligencia artificial.

El proyecto también ha incluido un exhaustivo análisis exploratorio de los datos (EDA) para entender mejor las características del sarcasmo en diferentes contextos. Se implementaron y compararon enfoques tradicionales como las redes neuronales recurrentes (RNN) y las redes LSTM, con el modelo Transformer, demostrando que este último ofrece un rendimiento significativamente superior en la detección de sarcasmo.

El resultado principal es un sistema optimizado que puede detectar sarcasmo en textos escritos en español, con una capacidad de generalización mejorada. El proyecto concluye que, aunque el modelo Transformer es eficaz, aún existen limitaciones en la detección de sarcasmos más sutiles, abriendo la puerta a futuras investigaciones y mejoras en esta área.

Palabras clave: Detección de sarcasmo, Transformers, Procesamiento de lenguaje natural, Fine-tuning, Deep learning, Clasificación de textos

ABSTRACT

This project addresses the problem of automatic sarcasm detection in Spanish, a complex challenge due to the subtlety and context required to identify this type of language. Most advances in natural language processing have focused on English, leaving a significant gap in other languages like Spanish. The main contribution of this work has been the development of a Transformer-based model, adapted through fine-tuning to a Spanish dataset, which includes news headlines, dialogues from TV shows, and AI-generated phrases.

The project also included an exhaustive exploratory data analysis (EDA) to better understand the characteristics of sarcasm in different contexts. Traditional approaches, such as recurrent neural networks (RNN) and LSTM networks, were implemented and compared with the Transformer model, showing that the latter offers significantly superior performance in sarcasm detection.

The main result is an optimized system capable of detecting sarcasm in written texts in Spanish, with improved generalization capacity. The project concludes that, although the Transformer model is effective, there are still limitations in detecting more subtle forms of sarcasm, paving the way for future research and improvements in this area.

Keywords: Sarcasm detection, Transformers, Natural language processing, Fine-tuning, Deep learning, Text classification

AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a mi familia, en especial a mis padres, por su apoyo incondicional, no solo en este proyecto, sino a lo largo de toda mi trayectoria académica. Gracias por darme la fuerza y la motivación para continuar mis estudios, por creer en mí y por estar siempre a mi lado en cada paso de este camino.

También me gustaría agradecer a la profesora María Criz Gaya por su guía y apoyo durante el desarrollo de este proyecto. Su orientación ha sido fundamental para lograr los objetivos planteados y completar este trabajo de manera exitosa. Gracias por compartir su conocimiento y por ofrecerme su tiempo y dedicación a lo largo de todo el proceso.

"Quien mira hacia afuera, sueña; quien mira hacia adentro, despierta."

- Carl Gustav Jung

TABLA RESUMEN

	DATOS
Nombre y apellidos:	Ernesto Francisco Barrionuevo Aragonés
Título del proyecto:	Detección automática de sarcasmos en textos utilizando modelos basados en Transformers
Directores del proyecto:	María Cruz Gaya López
El proyecto se ha realizado en colaboración de una empresa o a petición de una empresa:	NO
El proyecto ha implementado un producto: (esta entrada se puede marcar junto a la siguiente)	NO
El proyecto ha consistido en el desarrollo de una investigación o innovación: (esta entrada se puede marcar junto a la anterior)	SI
Objetivo general del proyecto:	Desarrollar un sistema de detección de sarcasmos en español utilizando un modelo Transformer ajustado mediante fine-tuning

Índice

RESUMEN	3
ABSTRACT	4
TABLA RESUMEN	7
Capítulo 1. RESUMEN DEL PROYECTO	13
1.1 Contexto y justificación.....	13
1.2 Planteamiento del problema	13
1.3 Objetivos del proyecto.....	13
1.4 Resultados obtenidos	13
1.5 Estructura de la memoria.....	13
Capítulo 2. ANTECEDENTES / ESTADO DEL ARTE	14
2.1 Estado del arte	14
2.2 Contexto y justificación.....	47
2.3 Planteamiento del problema	55
Capítulo 3. OBJETIVOS	58
3.1 Objetivos generales	58
3.2 Objetivos específicos	58
3.3 Beneficios del proyecto	59
Capítulo 4. DESARROLLO DEL PROYECTO	60
4.1 Planificación del proyecto	60
4.2 Descripción de la solución, metodologías y herramientas empleadas.....	62
4.3 Recursos requeridos	75
4.4 Presupuesto	76
4.5 Viabilidad	77
4.6 Resultados del proyecto	77
Capítulo 5. DISCUSIÓN	104
5.1 Discusión de Resultados del EDA.....	104
5.2 Discusión de los modelos implementados.....	105
5.3 Discusión de las pruebas obtenidas.....	106

Capítulo 6.	CONCLUSIONES.....	108
6.1	Conclusiones del trabajo.....	108
6.2	Conclusiones personales.....	109
Capítulo 7.	FUTURAS LÍNEAS DE TRABAJO.....	110
Capítulo 8.	Bibliografía.....	111
Capítulo 9.	ANEXOS	114

Índice de Figuras

Ilustración 1: Separación en tokens de una frase	18
Ilustración 2: Análisis sintáctico de una oración.....	19
Ilustración 3: Relaciones semánticas y sintácticas en vectores Word2Vec	20
Ilustración 4: Esquema de una red neuronal con pesos y función de activación.	25
Ilustración 5: Gráfico de la función de activación Sigmoide.....	26
Ilustración 6: Gráfico de la función de activación ReLU.....	27
Ilustración 7: Gráfico de la función de activación Tangente Hiperbólica.....	28
Ilustración 8: Gráfico de la función de activación SoftMax.....	29
Ilustración 9: Gráfico de la función de pérdida con una curva sencilla	31
Ilustración 10: Esquema de una red neuronal convolucional para la clasificación de textos	33
Ilustración 11: Esquema de una red neuronal recurrente	34
Ilustración 12: Arquitectura del modelo Transformer de "Attention is All You Need".	38
Ilustración 13: Atención de producto punto escalado y atención multi-cabeza.	39
Ilustración 14: Bloque Encoder en la arquitectura Transformer	40
Ilustración 15: Bloque Decoder en la arquitectura Transformer.....	41
Ilustración 16: Ejemplo de noticia sarcástica de El Mundo Today.	64
Ilustración 17: Ejemplo de noticia no sarcástica de El Sur	65
Ilustración 18: Ejemplo de sarcasmo creado con Inteligencia Artificial.....	66
Ilustración 19: Ejemplo de extracción de noticias sarcásticas.....	79
Ilustración 20: Ejemplo de extracción de noticias no sarcásticas.....	79
Ilustración 21: Ejemplo de extracción de diálogos de series de televisión	80
Ilustración 22: Ejemplo de extracción de frases creadas por inteligencia artificial.....	80
Ilustración 23: Dataframe limpio	82
Ilustración 24: CSV guardado en el equipo	82
Ilustración 25: Distribución de las etiquetas de sarcasmo en el conjunto de datos.....	83
Ilustración 26: Descripción del DataFrame utilizado en el análisis de sarcasmo.	83
Ilustración 27: Estadísticas descriptivas de la longitud de los textos en el conjunto de datos. ...	84
Ilustración 28: Distribución de la longitud de los textos en el conjunto de datos.	84

Ilustración 29: Comparación de la longitud del texto según si es sarcástico o no.	85
Ilustración 30: Frecuencia de palabras en textos sarcásticos y no sarcásticos con y sin valores atípicos.	86
Ilustración 31: Top 10 términos más importantes según el análisis TF-IDF (conjunto completo)	87
Ilustración 32: Top 10 términos TF-IDF en textos sarcásticos y no sarcásticos con valores atípicos.	88
Ilustración 33: Top 10 términos TF-IDF en textos sarcásticos y no sarcásticos sin valores atípicos.	89
Ilustración 34: Top 10 términos TF-IDF en textos sarcásticos y no sarcásticos (solo valores atípicos).	89
Ilustración 35: Frecuencia de bigramas con stop words en textos sarcásticos y no sarcásticos con y sin valores atípicos.	90
Ilustración 36: Frecuencia de trigramas con stop words en textos sarcásticos y no sarcásticos con y sin valores atípicos.	91
Ilustración 37: Frecuencia de bigramas sin stop words en textos sarcásticos y no sarcásticos con y sin valores atípicos.	92
Ilustración 38: Frecuencia de trigramas sin stop words en textos sarcásticos y no sarcásticos con y sin valores atípicos.	93
Ilustración 39: Distribución de sentimientos en textos con y sin valores atípicos.	94
Ilustración 40: Distribución de sentimientos en textos sarcásticos y no sarcásticos.	95
Ilustración 41: Matriz de correlación entre longitud, sentimiento y sarcasmo (con valores atípicos).	96
Ilustración 42: Matriz de correlación entre longitud, sentimiento y sarcasmo (sin valores atípicos).	97
Ilustración 43: Comparación de variabilidad léxica (TTR) entre textos sarcásticos y no sarcásticos.	98
Ilustración 44: Interfaz gráfica para la detección de sarcasmo.	101
Ilustración 45: Comparativa de métricas entre modelos para la detección de sarcasmo.	105

Índice de Tablas

Tabla 1: Presupuesto del proyecto	77
Tabla 2. Resultados modelo basado en reglas	99
Tabla 3. Entrenamiento modelo basado en RNN	99
Tabla 4. Resultados modelo basado en RNN.....	99
Tabla 5. Entrenamiento modelo basado en LSTM.....	100
Tabla 6. Resultados modelo basado en LSTM	100
Tabla 7. Entrenamiento modelo sin optimizar basado en Transformers.....	100
Tabla 8. Mejores resultados obtenidos en la búsqueda de hiperparámetros	101
Tabla 9. Mejores hiperparámetros obtenidos para optimizar modelo Transformer	101
Tabla 10. Resultados obtenidos en la fase de pruebas	103

Capítulo 1. RESUMEN DEL PROYECTO

1.1 Contexto y justificación

El proyecto se desarrolla en el ámbito del procesamiento del lenguaje natural (PLN), abordando el problema de la detección automática de sarcasmo en textos en español. Dado que la mayor parte de los avances en esta área se han centrado en el inglés, existe una necesidad clara de adaptar estas tecnologías a otros idiomas. El uso de modelos basados en Transformers ha demostrado ser efectivo en múltiples tareas de PLN, lo que justifica su aplicación para la detección de sarcasmo en español, un reto complejo debido a la naturaleza contextual y sutil del sarcasmo.

1.2 Planteamiento del problema

El problema que este proyecto busca resolver es la dificultad de detectar sarcasmo en textos escritos en español, una tarea que, a nivel computacional, ha sido poco explorada. La pregunta principal a responder es: ¿cómo podemos mejorar la capacidad de un modelo de inteligencia artificial para identificar el sarcasmo en textos en español? El proyecto combina investigación y desarrollo para aportar una solución efectiva a este reto, utilizando técnicas avanzadas como el fine-tuning de Transformers para adaptarlos al contexto del sarcasmo.

1.3 Objetivos del proyecto

El objetivo principal es desarrollar un modelo basado en Transformers para la detección automática de sarcasmo en textos en español. Además, se plantean objetivos específicos como la creación de un conjunto de datos en español, el análisis exploratorio de los datos, la implementación de modelos tradicionales y de Transformers, y la validación y optimización del modelo final.

1.4 Resultados obtenidos

El proyecto logró implementar un modelo Transformer optimizado para la tarea de detección de sarcasmo en español, obteniendo resultados superiores a los enfoques tradicionales (como RNN y LSTM). Los resultados indican que el modelo Transformer es eficaz en la identificación de sarcasmo en la mayoría de los casos, aunque aún presenta algunas limitaciones en la detección de sarcasmo sutil.

1.5 Estructura de la memoria

La memoria se estructura en varios capítulos. En el Capítulo 2, se analiza el estado del arte y los antecedentes relevantes al proyecto. El Capítulo 3 expone los objetivos generales y específicos planteados. En el Capítulo 4, se describe el desarrollo del proyecto, incluyendo la metodología utilizada y los resultados obtenidos. En el Capítulo 5, se discuten los resultados obtenidos y el desempeño de los modelos implementados. El Capítulo 6 presenta las conclusiones del trabajo, mientras que en el Capítulo 7 se proponen posibles líneas de trabajo futuras.

Capítulo 2. ANTECEDENTES / ESTADO DEL ARTE

En este apartado, se abordarán sobre diferentes temas relacionados con el estado del arte de nuestro proyecto. En primer lugar, definiremos lo que es un sarcasmo, cuales son sus tipos y las dificultades que podemos encontrar a la hora de detectarlos en un texto escrito. Asimismo, se discutirá cuales eran las técnicas tradicionales utilizadas, los métodos clásicos de procesamiento de lenguaje natural y las limitaciones que estos enfoques podían encontrarse. Se continuará el estado del arte exponiendo los métodos basados en redes neuronales y **Deep Learning** para comprobar la funcionalidad del uso de **RNN, LSTM Y CNN** en las tareas de clasificación de texto además de una comparativa entre estos enfoques.

Posteriormente se analizará los modelos basados en Transformers, explicando su arquitectura y como mejoran el rendimiento en las tareas de PLN. Se realizará una breve explicación de los modelos más utilizados como **BERT, GPT o RoBERTa**, además de exponer diferentes ejemplos de estudios realizados previamente en este ámbito y cuales son sus ventajas y limitaciones.

Para contextualizar nuestro proyecto se comenzará exponiendo la importancia de la detección del sarcasmo, cual su relevancia en interacciones humanas en línea, que problemas puede causar la falta de una detección adecuada en aplicaciones con moderación de contenido y cuales son sus beneficios potenciales para diversas aplicaciones. Al mismo tiempo, se contemplará el estado de la tecnología actual, comparando su nivel de investigación en comparación con otras tareas de procesamiento de lenguaje natural (PLN), la revisión de los conjuntos de datos disponibles y exponiendo los retos actuales a los cuales se enfrenta este tipo de proyectos. Se debatirá la justificación del uso de Transformers frente a otros modelos debido a su capacidad de capturar dependencias contextuales más profundas. Además, se expondrán los avances más recientes en PLN, enfatizando cómo los Transformers han revolucionado tanto el procesamiento del lenguaje natural como la detección del sarcasmo.

Como último punto, abordaremos el planteamiento del problema, el cual se centra en cómo mejorar la precisión en la detección automática de sarcasmo utilizando modelos basados en Transformers frente a los métodos tradicionales de procesamiento de lenguaje natural (PLN). Este problema es crucial en el contexto actual de la comunicación digital, donde la ambigüedad del sarcasmo, influenciada por factores culturales, lingüísticos y pragmáticos, presenta retos significativos. Además, existen dificultades en la captura del contexto por parte de los modelos tradicionales, y limitaciones en los conjuntos de datos disponibles. El objetivo es desarrollar un modelo basado en Transformers que pueda ser evaluado en diversos conjuntos de datos públicos, comparándolo con modelos como BERT o RoBERTa, y considerando posibles mejoras en su arquitectura para una mejor detección del sarcasmo.

2.1 Estado del arte

En esta sección se revisará la literatura relacionada con la detección automática de sarcasmo en texto, definiendo en profundidad lo que es un sarcasmo, destacando las limitaciones de los enfoques tradicionales de PLN y los avances recientes con modelos basados en Transformers,

como BERT y RoBERTa, que han mejorado el rendimiento en esta tarea. Esta revisión contextualiza el problema y apoya el enfoque de este trabajo.

2.1.1 Sarcasmo en el lenguaje

El sarcasmo según la RAE se define como “*Una burla sangrienta, ironía mordaz y cruel con que se ofende o maltrata a alguien o algo*” (Real Academia Española, 2023). Como bien nos define la RAE el sarcasmo es un tipo de ironía con un grado más de crueldad en su intención, ya que busca no solo expresar lo contrario de lo que se dice si también provocar burla o desdén hacia el receptor. Para poder entender de manera correcta lo que es un sarcasmo primero debemos entender que es una ironía y en que lo diferencia de un sarcasmo.

2.1.1.1 Diferencia entre sarcasmo e ironía

Una ironía, según la RAE se puede definir como “*una burla fina y disimulada*” (Real Academia Española, 2023). A priori, estas definiciones podrían sugerir que es fácil diferenciar entre ambas, pero en la práctica, la distinción es más compleja de lo que parece. Esto se debe a que la ironía abarca un espectro más amplio de expresiones en las que el significado se encuentra implícito o es contrario a lo que se dice literalmente, mientras que el sarcasmo es solo una de sus formas más específicas. Además, la ironía puede manifestarse tanto en situaciones como en el lenguaje verbal; una circunstancia puede ser irónica sin la intervención directa de una persona. Por otro lado, el sarcasmo solo puede ser expresado verbalmente o en forma escrita, es decir, a través de la comunicación. Una persona puede ser sarcástica, pero no se puede decir lo mismo de una situación, ya que el sarcasmo requiere la intencionalidad de quien emite el mensaje, mientras que la ironía puede surgir de manera espontánea en el entorno o en los eventos.

2.1.1.2 Tipos de sarcasmo

Una vez se ha entendido lo que es el sarcasmo y cuales es su mayor diferencia con la ironía, se explicara cuales son los diferentes tipos de sarcasmos. Durante la realización de este trabajo se ha investigado en profundidad sobre cuales son los diferentes tipos de sarcasmos que podemos encontrar en la actualidad y se ha llegado a la conclusión de que existen dos tipos principales de sarcasmos. La distinción entre sarcasmo intencionado y sarcasmo percibido, también referida como sarcasmo codificado y sarcasmo decodificado, respectivamente, ha sido señalada en investigaciones previas (Vlad & Magdy, 2019).

- **Intencionado:** el sarcasmo intencionado se produce cuando una persona intenta realizar algún comentario hiriente acerca de alguien o algo en particular de forma cómica. Este es el tipo de sarcasmo más común ya que las personas cuando hacen un comentario de forma sarcástica siempre van con intención de herir a la persona u objeto. Además, porque una persona puede o no entender este sarcasmo y un objeto claramente no tiene la posibilidad de entender este comentario.
- **Percibido:** este tipo de sarcasmo se produce cuando la persona oyente entiende que el locutor le está lanzando una frase sarcástica hacia su persona. Este es el tipo de sarcasmo menos frecuente ya que normalmente no se entiende un sarcasmo cuando el

locutor no tiene intención de utilizarlo, pero si existe la posibilidad de que el oyente no entienda las intenciones sarcásticas del locutor.

2.1.1.3 Sarcasmo por contexto y zona geográfica

El sarcasmo, como se ha mencionado en esta primera etapa del proyecto, es algo que resulta complicado de entender a primera vista. De hecho, en muchos casos, si no se comprende el contexto o no se expresa de manera adecuada, incluso los oyentes pueden no captar el tono sarcástico del mensaje. Según estudios recientes, la interpretación del sarcasmo está influenciada no solo por el tono de voz y las expresiones no verbales, sino también por el contexto cultural, que puede hacer que un mismo comentario sea entendido de forma diferente en diversas regiones o comunidades. La interpretación del sarcasmo depende en gran medida de las señales verbales y no verbales, además del conocimiento previo sobre el hablante y el contexto (Everyday Speech). Por eso, entender el sarcasmo requiere una mayor atención y una astucia por encima de lo habitual. Además, cuanto más inteligente o cargado de contexto sea el sarcasmo, más mordaz y cómico puede ser, lo que también exige un mayor conocimiento para poder interpretarlo correctamente.

Otro aspecto importante a considerar al subcategorizar los sarcasmos es que, según la zona geográfica, estos pueden ser completamente diferentes entre sí. Como podemos observar en el siguiente mapa, una de las principales razones es que el humor que los sustenta depende del contexto, ya sea político, geográfico o cultural. Además, incluso entre personas de la misma región, el sarcasmo puede variar, ya que cada individuo interpreta y utiliza el sarcasmo de manera distinta, influenciado por sus experiencias y su entorno personal.



Un estudio reciente muestra que la comprensión del sarcasmo varía entre culturas debido a diferencias en valores como la distancia al poder y el colectivismo. Por ejemplo, los participantes chinos tienden a comprender el sarcasmo de manera diferente que los estadounidenses, destacando cómo los factores culturales influyen en la interpretación del sarcasmo en diferentes regiones y contextos sociales (Du, He, & Chu, 2024).

2.1.1.4 Dificultad del entendimiento del sarcasmo

Por último, para detectar completamente el sarcasmo en una frase, es fundamental prestar atención a una serie de factores externos que van más allá del simple uso del lenguaje. Estos factores incluyen elementos no verbales que ayudan a interpretar el verdadero sentido del mensaje. Por ejemplo, una sonrisa pícaro o maliciosa puede sugerir que lo que se ha dicho no debe tomarse de manera literal. Del mismo modo, el tono de voz juega un papel crucial: un tono que denota pereza, ironía, negatividad o incluso cierta agresividad puede ser una clara señal de sarcasmo. Estos matices, muchas veces sutiles, enriquecen la comprensión del sarcasmo, ya que, sin ellos, el mensaje puede perder su verdadera intención y ser malinterpretado.

2.1.2 Procesamiento de lenguaje natural

El procesamiento de lenguaje natural (PLN) es una rama clave del aprendizaje automático que se centra en la interacción entre las máquinas y el lenguaje humano. Según Amazon Web Services, el PLN es *“una tecnología de machine learning que brinda a las computadoras la capacidad de interpretar, manipular y comprender el lenguaje humano”* (Amazon Web Service, s.f.). Esta disciplina es esencial para analizar y extraer información de grandes volúmenes de datos de texto y voz de manera eficiente y precisa.

El PLN ha experimentado un crecimiento significativo gracias a décadas de investigación, y hoy en día es la base de una amplia gama de aplicaciones que impactan tanto en el ámbito personal como en el profesional.

Asimismo, el PLN sigue evolucionando y desarrollando diferentes técnicas avanzadas que han revolucionado el campo al mejorar la capacidad de las máquinas para captar el contexto y los matices del lenguaje. Esto ha permitido avances notables en tareas como la detección del sarcasmo, el resumen automático de textos y la generación avanzada de lenguaje natural.

2.1.2.1 Niveles de análisis en el procesamiento de lenguaje natural

En el procesamiento de lenguaje natural, el análisis de lenguaje se realiza en diferentes niveles, cada uno de ellos aborda un aspecto específico del texto y es fundamental para que los modelos de aprendizaje automático puedan entender y procesar el lenguaje humano de manera efectiva. Estos niveles permiten descomponer el lenguaje en unidades más manejables, lo que a su vez permite que la información extraída sea más relevante. Las cinco fases principales del procesamiento del lenguaje natural se pueden ver reflejadas en el artículo realizado por Gradient Ascendt las cuales son: Análisis léxico, Análisis sintáctico, Análisis semántico Integración del discurso y Análisis pragmático (Gradient Ascendt, 2023).

2.1.2.1.1 Análisis léxico

La tarea principal del análisis léxico se basa en preparar los datos para que ser procesados en niveles superiores, ayudando a los modelos a trabajar con palabras de manera eficiente. Este nivel se centra principalmente en el reconocimiento de las palabras individuales dentro de un texto, conocidas como **tokens**. La descomposición del conjunto de datos en sus componentes básicos, como pueden ser palabras o símbolos, es una de las primeras fases del preprocesamiento de datos llamada **Tokenización**. (Gradient Ascendt, 2023)

El análisis léxico también incluye la **normalización** de las palabras, el cual se basa en convertir todo a minúsculas y eliminar los signos de puntuación además de la **lematización** para poder reducir todas las palabras a su forma base.

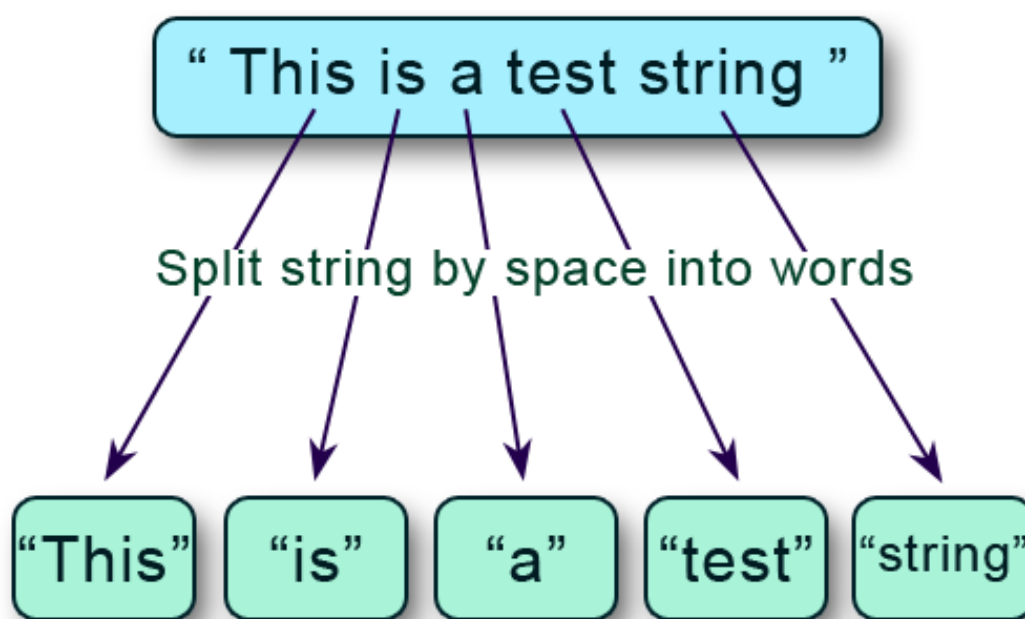


Ilustración 1: Separación en tokens de una frase

Fuente: Adaptado de w3resource (2023)

2.1.2.1.2 Análisis sintáctico

El análisis sintáctico, también conocido como Parsing analiza la estructura gramatical del texto, identificando como las palabras se relacionan entre si dentro de una oración. Para ello se construyen arboles sintácticos que representan la jerarquía gramatical de la frase (Gradient Ascendt, 2023). La función principal de los árboles sintácticos es mostrar como las palabras y las frases en una oración se agrupan y relacionan entre si según las reglas gramaticales del lenguaje. Este paso ayuda a comprender la organización del texto y establecer las relaciones entre los sujetos, predicados, objetos, etc. El parsing permite al modelo no solo trabajar con palabras individuales, sino también con las relaciones estructurales que definen el significado de las frases.

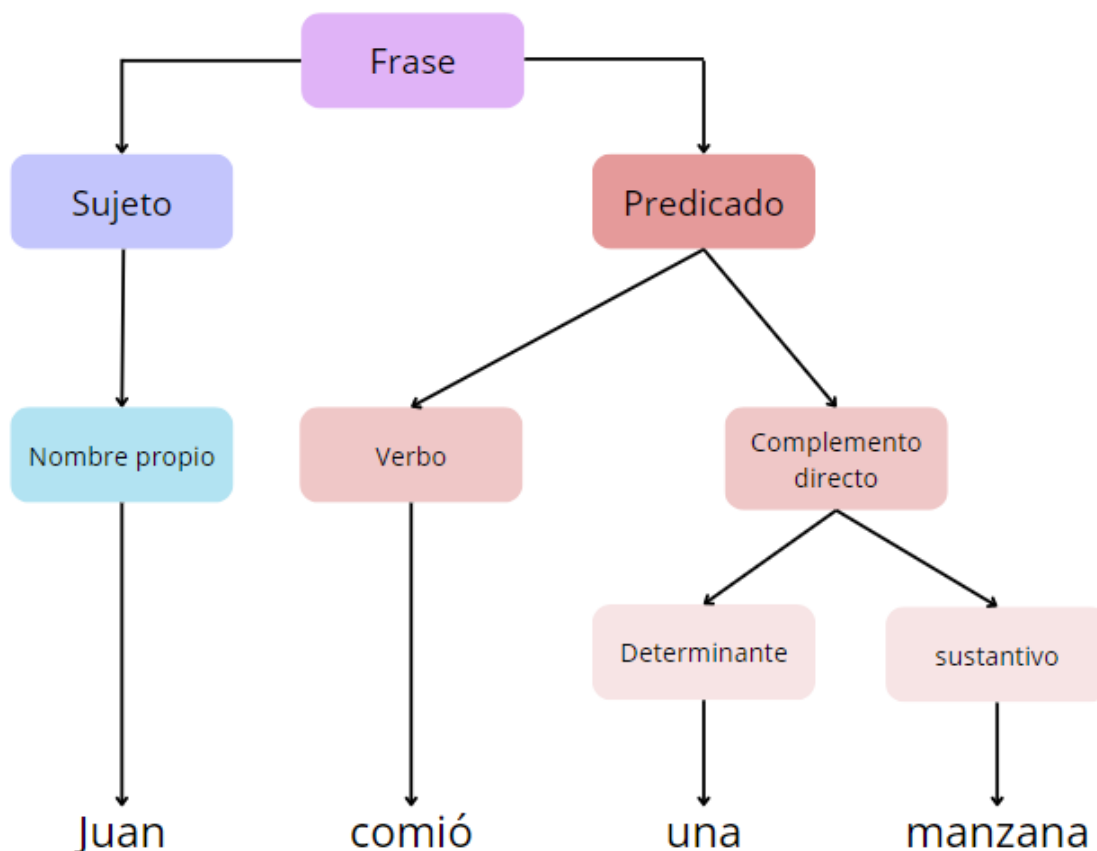


Ilustración 2: Análisis sintáctico de una oración

Fuente: Elaboración propia

2.1.2.1.3 Análisis semántico

El análisis semántico se enfoca en el significado de las palabras y frases más allá de su estructura gramatical (Gradient Ascendt, 2023). El objetivo es obtener una representación del significado que sea comprensible para la máquina. Adicionalmente se aborda la desambiguación semántica, que resuelve el significado de palabras que puedan tener múltiples sentidos dependiendo del contexto. El análisis semántico es crucial para que los modelos puedan entender el significado subyacente de los textos y no solo sus componentes superficiales.

Para que el sistema aprenda el significado de las palabras se pueden utilizar diferentes tipos de herramientas, como pueden ser los Words Embeddings o incluso algo mucho mas complejo como directamente aplicar un modelo transformer.

Un Word embedding es una representación matemática o vectorial de palabras en un espacio de características de dimensiones reducidas, que captura las relaciones semánticas y sintácticas entre las palabras además de ayudar a las computadoras a procesar el texto transformando las palabras en vectores numéricos. Los Word embedding resuelven estos problemas al asignar a cada palabra un vector de menor dimensionalidad, donde las relaciones semánticas y sintácticas

entre las palabras se reflejan en la geometría del espacio. Palabras que tienen un significado similar tienden a tener vectores que están más cerca en ese espacio, mientras que palabras con significados diferentes estarán más alejadas. Los Word embeddings más utilizados a día de hoy podrían decirse que son Word2Vec y GloVe.

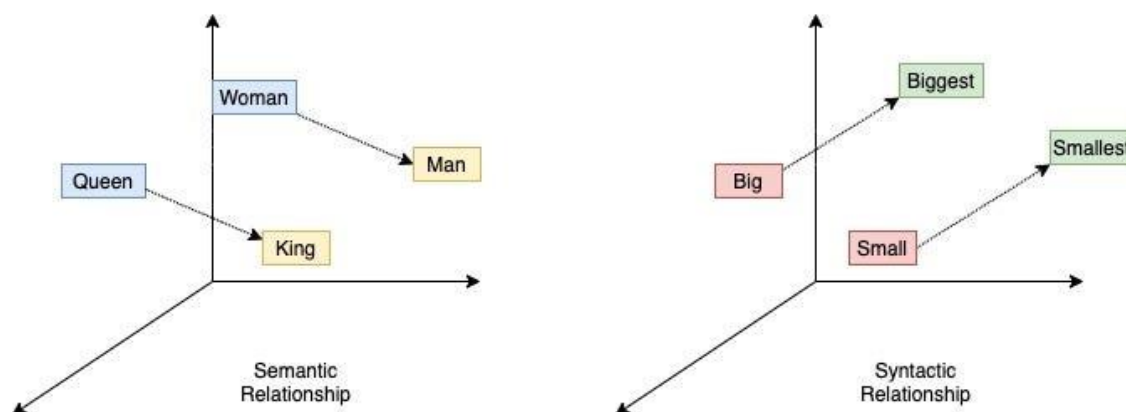


Ilustración 3: Relaciones semánticas y sintácticas en vectores Word2Vec

Fuente: Adaptado de Towards Data Science (2023)

2.1.2.1.4 Integración del discurso

Este nivel se centra en como las oraciones y las frases están conectadas entre si a lo largo de un texto o una conversación (Gradient Ascendt, 2023). El significado de una oración a menudo depende del contexto proporcionado. Por ejemplo, las palabras de referencia “él” o “ella” requieren información de oraciones anteriores para identificar a que persona se refieren. La integración del discurso ayuda a captar la cohesión y coherencia del texto, permitiendo un análisis más profundo que trascienda las frases individuales.

Las técnicas para la integración del discurso incluyen enfoques basados en reglas gramaticales y redes neuronales recurrentes como LSTM y GRU, que procesan secuencias de texto. Sin embargo, los modelos más avanzados son los transformers (e.g., BERT, GPT), que, gracias a su mecanismo de autoatención, capturan de manera efectiva las relaciones contextuales entre oraciones en textos largos.

2.1.2.1.5 Análisis pragmático

El análisis pragmático en PLN se ocupa de interpretar el significado en función del contexto situacional y de las intenciones del hablante o escritor, más allá del contenido explícito del texto (Gradient Ascendt, 2023). Este análisis es clave para detectar sarcasmos, ironía y otros matices del lenguaje donde el significado literal no coincide con la intención real.

El análisis pragmático implica el entendimiento de la relación entre el texto y su entorno, tomando en cuenta diferentes tipos de factores entre los que se encuentran: la intención comunicativa, el contexto extralingüístico, actos de habla e implicaturas.

En PLN, el análisis pragmático tradicionalmente ha sido difícil de implementar debido a la complejidad de la comprensión del contexto. Sin embargo, con la aparición de nuevos modelos basados en Transformers se ha avanzado significativamente en la detección de significados implícitos. Estos modelos pueden captar los matices pragmáticos utilizando mecanismos de atención, los cuales se explicarán más adelante en profundidad, para detectar incongruencias entre el significado literal y el contexto, lo que es crucial para identificar el sarcasmo y otras formas de lenguaje no literal.

2.1.2.2 Tareas principales del PLN

El procesamiento de lenguaje natural, como hemos comentado previamente, abarca una amplia variedad de tareas que permiten a los sistemas entender, interpretar y generar lenguaje humano. Entre una gran multitud de tareas que se pueden realizar referenciadas en un artículo realizado por IBM, podemos encontrar las siguientes (IBM, 2024):

- **Conversión de texto a voz:** esta tarea implica convertir datos de texto en habla, reproduciendo el texto como un discurso natural. Es comúnmente utilizado en aplicaciones de accesibilidad, como asistentes virtuales o dispositivos de lectura en voz alta.
- **Chatbots:** Los chatbots creados a utilizando técnicas de PLN para entender y responder de manera automática a las consultas de los usuarios. Esta tecnología permite interactuar de manera más efectiva y personalizada con los clientes, proporcionando respuestas basadas en el contexto e interacción del usuario.
- **Autocorrección:** La autocorrección se basa principalmente en la eliminación de errores en el texto, ofreciendo sugerencias alternativas basadas en el contexto.
- **Análisis de sentimiento:** Esta tarea sirve para analizar los sentimientos a través del tono y el contenido del texto para poder entender de una forma más efectiva el contexto de la frase. Este tipo de análisis tiene un alto valor ya que se pueden identificar emociones y actitudes, lo que es crucial para el análisis de opiniones y la reputación en línea.
- **Detección de urgencia:** Consiste en analizar el lenguaje utilizado por los usuarios para detectar la prioridad de las solicitudes. Es útil para priorizar tareas de acuerdo a la importancia o urgencia detectada en las consultas.
- **Reconocimiento de voz:** Permite a las aplicaciones entender las voces o usuarios y descifrar su significado. Este proceso es fundamental en asistentes de voz, sistemas de dictado y otras tecnologías donde se requiere la interacción a través del lenguaje hablado.
- **Detección de sarcasmos:** Esta tarea se basa en la detección automática de sarcasmos a través del contexto y el tono para descifrar un mensaje que normalmente suele tener un significado contrario al que el usuario envía. Este tipo de prácticas es muy útil para plataformas de creación de contenido moderado o de contenido infantil.

2.1.2.3 Contexto y objetivos del PLN en la detección de sarcasmos

El estudio del sarcasmo y el cómo detectarlo tanto oralmente como escrito se lleva realizando desde hace ya más de 10 años con el primer estudio realizado por Davidov, Tsur y Rappoport

llamado “Semi-supervised recognition of sarcasm in Twitter and Amazon” (Davidov, Tsur, & Rappoport, 2010) en donde se usa por primera vez una gran cantidad de datos extraídos de Twitter y reseñas de Amazon. A partir de este momento el estudio de la detección automática de sarcasmos en texto ha crecido de manera exponencial en los últimos años gracias en parte también al crecimiento que ha tenido el estudio de Procesamiento de lenguaje Natural (PLN).

Como bien se ha ido mencionando a lo largo de todo el estado de arte, el sarcasmo es un fenómeno lingüístico complejo, ya que a menudo el significado difiere de lo literal y puede depender en gran medida del contexto, tono o antecedentes culturales. Esto hace que la detección de un sarcasmo sea compleja tanto para los seres humanos como para las máquinas. En los textos, el sarcasmo puede aparecer como contradicciones entre lo que se dice y lo que se espera, el uso de exageraciones o incluso la combinación de afirmaciones positivas con contextos negativos. Todos estos matices son muy difíciles de identificar si un análisis profundo del contexto y pragmático anterior.

EL PLN se ha centrado en identificar estos retos mediante diferentes técnicas y modelos que evolucionan con el tiempo. Los primeros enfoques, como los basados en reglas y diccionarios de palabras clave eran efectivos en escenarios limitados, pero carecían de la capacidad de generalización necesaria para detectar sarcasmos en textos mas complejos y variados. Con la llegada de modelos mas avanzados como los basados en redes neuronales o Transformers, los investigadores han sido capaces de capturar patrones de lenguaje mas sutiles y contextuales, mejorando así la precisión de los sistemas de detección de sarcasmos.

En la detección de sarcasmos mediante PLN podemos encontrar múltiples objetivos:

- **Identificación de incongruencias:** detección de la presencia de discrepancias entre el significado literal de una frase y el contexto en el que se utiliza, lo que puede ser indicativo de sarcasmo (Gelbukh, 2022).
- **Captura de interacciones:** A través de los modelos de comprensión del lenguaje natural, no únicamente se busca por una comprensión literal de las palabras, si no también entender las interacciones detrás de estas.
- **Análisis contextual:** Gracias a la ayuda de los Transformers, es posible analizar el contexto global de una conversación o texto, lo que facilita la identificación de sarcasmos en secuencias mas largas y complejas del lenguaje.
- **Análisis de sentimiento:** La detección de los sentimientos en una frase sarcástica es fundamental, ya que en muchas situaciones las emociones expresadas no coinciden con el mensaje literal que se está comunicando.

2.1.3 Técnicas tradicionales de PLN para detectar sarcasmos

Como se mencionó en el párrafo anterior, el procesamiento de lenguaje natural ha experimentado un crecimiento vertiginoso en la última década, acompañado por la aparición de numerosas técnicas innovadoras que han mejorado significativamente su rendimiento. Sin embargo, este avance no siempre fue así. Para llegar al conjunto de herramientas sofisticadas que disponemos hoy en día, en el pasado se recurrió a técnicas más rudimentarias. Estas soluciones, aunque cumplían las mismas funciones que las actuales, ofrecían resultados menos

precisos y requerían un mayor esfuerzo tanto en términos de tiempo como de recursos. A continuación, se describirán los principales enfoques utilizados los cuales dominaron la detección de sarcasmos en sus primeros años.

2.1.3.1 Modelos basados en reglas

Un modelo de procesamiento de lenguaje natural (PLN) basado en reglas es un sistema que utiliza un conjunto predefinido de reglas para realizar tareas como el análisis, la extracción o la clasificación de información en textos o voz. Estas reglas, diseñadas por expertos humanos con conocimientos lingüísticos, se basan en elementos como la sintaxis, semántica, morfología o pragmática del idioma. Por ejemplo, en el análisis de sentimientos, un sistema basado en reglas podría utilizar listas de palabras y frases con connotaciones positivas o negativas para determinar la polaridad de un texto (Teki & Cicujano, 2024).

En el caso de la detección de sarcasmo, estos modelos emplean reglas que identifican patrones específicos que indican incongruencias entre el significado literal y el contexto. Por ejemplo, una frase positiva en un contexto negativo puede ser marcada como sarcasmo, o el uso de expresiones exageradas y adverbios como "obviamente" o "claramente" en situaciones irónicas puede ser otro indicador clave. Además, un modelo PLN basado en reglas puede ser propenso a errores y excepciones ya que en muchas ocasiones no puede manejar la ambigüedad, variabilidad o la creatividad del lenguaje natural (Teki & Cicujano, 2024).

2.1.3.2 Clasificadores supervisados tradicionales

Los clasificadores supervisados son modelos de aprendizaje automático que requieren un conjunto de datos etiquetados para entrenarse. Estos datos consisten en ejemplos en donde las características del texto están asociadas con etiquetas que indican si el texto es sarcástico o no. En base a estas etiquetas el modelo aprende a identificar patrones para después hacer predicciones sobre nuevos textos (Joshi, Bhattacharyya, & J Carman, 2016).

Existen una gran variedad de clasificadores supervisados, pero tradicionalmente los clasificadores que tenían una mayor aceptación para los modelos de procesamiento de lenguaje natural eran: Naive Bayes, Support Vector Machine y árboles de decisión. Asimismo, estos clasificadores normalmente eran entrenados con varias características de texto como la pragmática, léxica o semántica (Kumar, 2019).

Sin embargo, estos modelos presentan varias desventajas importantes en cuanto a detección de sarcasmos. La dependencia a grandes cantidades de datos etiquetados supone un gran costo en cuanto a tiempo y esfuerzo, además de la dificultad que le supone al modelo capturar relaciones complejas. Asimismo, están bastante limitados para encontrar nuevos contextos fuera de aquellos ya encontrados. Finalmente, su rendimiento es bastante inferior en comparación a técnicas más modernas, como las redes neuronales profundas o los Transformers (Kumar, 2019).

2.1.3.3 Modelos basados en diccionarios de palabras clave

Los modelos basados en diccionarios de palabras clave han sido uno de los primeros enfoques empleados para la detección de sarcasmo en textos. Estos modelos se basan en el uso de listas

predefinidas de palabras que son indicativos de sarcasmos. Se construyen diccionarios que contienen términos que suelen aparecer en comentarios sarcásticos, tales como superlativos, adjetivos extremos o expresiones exageradas. (Davidov, Tsur, & Rappoport, 2010)

EL enfoque de los diccionarios de palabras clave se centra en identificar términos que a menudo suelen estar asociados con el sarcasmo. Por ejemplo, palabras como “perfecto” o “genial” pueden usarse sarcásticamente dependiendo del contexto en donde el tono puede ser negativo o contradictorio. Asimismo, estos modelos pueden incorporar otros elementos léxicos, como pueden ser las mayúsculas, múltiples signos de interrogación o exclamación. Múltiples estudios, como el citado en el párrafo anterior, han demostrado que los estudios y palabras claves son efectivos cuando el sarcasmo es explícito y está muy bien definido (González-Ibáñez, Muresan, & Wacholder, 2011).

A pesar de la simplicidad y facilidad de uso de los modelos basados en diccionarios de palabras clave, estos presentan varias limitaciones. En primer lugar, no logran capturar de forma eficaz el contexto global en el que se expresan los comentarios sarcásticos, lo que puede acarrear una alta tasa de falsos positivos (Liebrecht, Kunneman, & van den Bosch, 2013). Al igual que los modelos de clasificación supervisados tradicionales, tiene dificultades para encontrar nuevos contextos ya que dependen de un conjunto limitado de palabras predefinidas.

2.1.3.4 Transición hacia enfoques más modernos

Los enfoques tradicionales, como se ha podido observar en los puntos anteriores, ofrecen soluciones iniciales simples y fáciles para poder abordar la compleja tarea de identificar los sarcasmos en textos. El problema radica en esta facilidad, ya que presentaron limitaciones significativas, particularmente en cuanto a su capacidad para comprender el contexto global, la generalización de nuevos dominios y las capturas más complejas en el lenguaje. Estos enfoques fueron fundamentales en los primeros esfuerzos para automatizar la detección de sarcasmos, pero su rendimiento limitado en situaciones más ambiguas y su alta dependencia de reglas o patrones fijos resaltaron la necesidad de técnicas más avanzadas.

La llegada de nuevas tecnologías, como las redes neuronales y el Deep learning, ha permitido un análisis mucho más profundo del lenguaje, permitiendo que los sistemas no solo comprendan mejor el significado literal de las palabras, sino también el contexto implícito que rodea el sarcasmo. Estos modelos más avanzados superan muchas de las restricciones que enfrentaban los enfoques tradicionales, abriendo un nuevo capítulo en la evolución de la detección de sarcasmo.

2.1.4 Modelos basados en Redes Neuronales y Deep Learning

2.1.4.1 Redes neuronales

Las redes neuronales son modelos de inteligencia artificial que simulan el funcionamiento del cerebro humano al procesar datos y reconocer patrones. Se basan en interconexiones de nodos que trabajan juntos para tomar decisiones y aprender de los datos. Además, las redes neuronales son la base del deep learning y de los modelos Transformers, los cuales son los modelos que utilizaremos más adelante. Amazon Web Service define una red neuronal como

“un método de inteligencia artificial que enseña a las computadoras a procesar datos de una manera similar a como lo hace un humano” (Amazon Web Service, s.f.). A veces las redes neuronales también son denominadas artificial neural networks (ANN) o simulated neural networks (SNN), y forman parte de aprendizaje automático y son la base del modelo del Deep learning (IBM, 2024).

2.1.4.1.1 ¿Cómo funciona una red neuronal?

Las redes neuronales consisten en capas de nodos (o neuronas artificiales): una capa de entrada, una o mas capas ocultas y una capa de salida. Cada nodo esta conectado a otros y tiene un **peso** y un umbral asociados. Si el valor de salida del nodo excede el umbral, ese nodo se activa y envía datos a la siguiente capa de la red. De lo contrario, no se transmite información a la siguiente capa.

Cada nodo es una red neuronal funciona de manera similar a una ecuación de regresión lineal, donde se reciben entradas (variables independientes), estas se multiplican por pesos y luego se suma un sesgo (bias) para ajustar el valor. La fórmula sería la siguiente:

$$\sum (Entrada \times peso) + sesgo$$

Después de calcular esta suma ponderada, el valor se pasa a través de una función de activación, que determina si el nodo se activa o no. Si el resultado del nodo excede un umbral específico definido por la función de activación, el nodo se activa y envía la información a la siguiente capa de la red. Si no, la señal no se transmite, lo que hace que el proceso sea eficiente y selectivo (IBM, 2024).

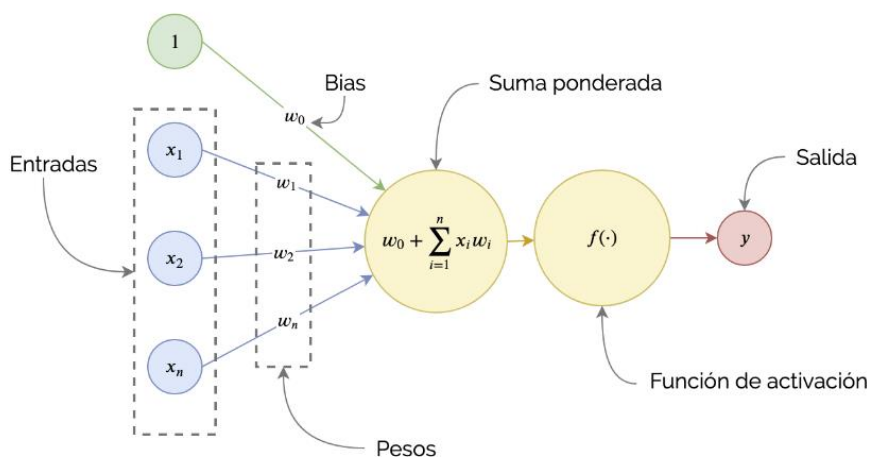


Ilustración 4: Esquema de una red neuronal con pesos y función de activación.

Fuente: Adaptado de Luis Castillo

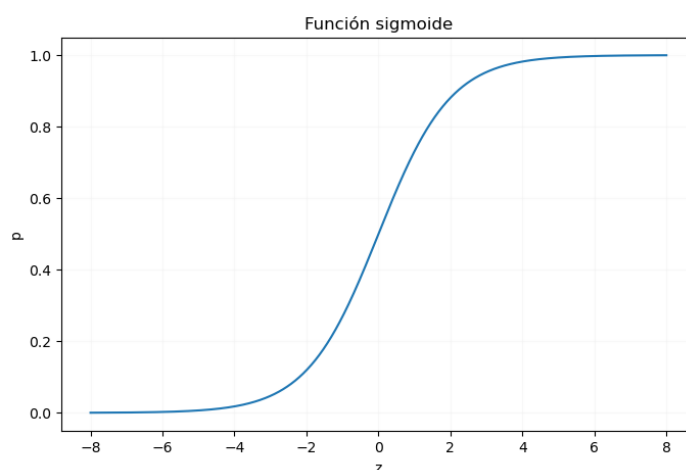
2.1.4.1.2 Función de activación

En el contexto de las redes neuronales, las funciones de activación determinan si una neurona se activa o no, permitiendo que la red capture la no linealidad en los datos. Estas funciones tienen el cometido de que las redes neuronales no limiten su capacidad para modelar datos complejos limitándose a comportarse de manera lineal. Hay una gran variedad de funciones de activación sin embargo presentaremos únicamente algunas de las más utilizadas. (Gabriel Gomila, 2024)

2.1.4.1.2.1 Función Sigmoidea

La función sigmoide transforma los valores de entrada en un rango entre 0 y 1, lo que es útil para problemas de clasificación binaria. La ecuación de la función sigmoide es:

$$f(x) = \frac{1}{1 + e^{-x}}$$



Fuente: Adaptado de InteractiveChaos

Ilustración 5: Gráfico de la función de activación Sigmoide

La función sigmoidea ofrece una salida suave y diferenciable, lo que facilita la retro propagación en las redes neuronales, pero tiene un problema con el desvanecimiento del gradiente, lo que puede hacer que el modelo aprenda de manera mas lenta. El método del gradiente descendente y la retro propagación se verán más adelante (Gabriel Gomila, 2024).

2.1.4.1.2.2 Función Rectified Linear Unit (ReLU)

La función ReLU es una de las funciones de activación más populares debido a su simplicidad y eficiencia en redes neuronales profundas. ReLU convierte los valores negativos en 0, mientras que deja pasar los valores positivos.

$$f(x) = \max(0, x)$$

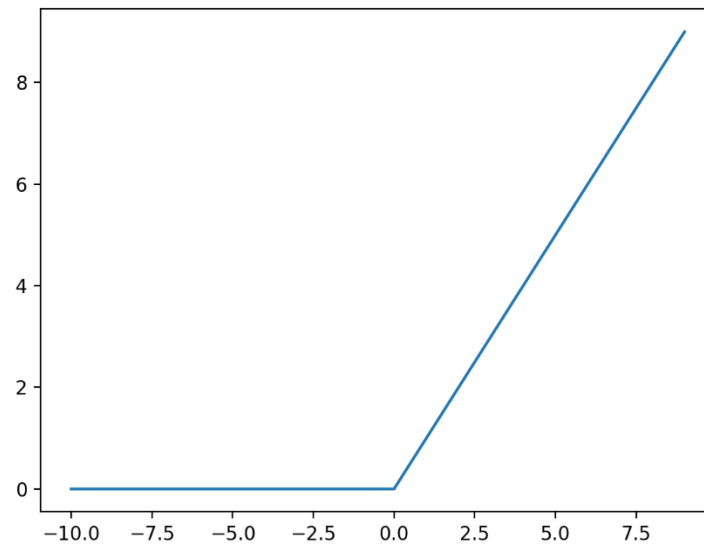


Ilustración 6: Gráfico de la función de activación ReLU.

Fuente: Adaptado de Alberto Rubiales en Medium

La función ReLU es computacionalmente eficiente y ayuda a resolver el problema de devastamiento del gradiente en redes profundas pero puede sufrir de problema de neuronas muertas, donde algunos nodos dejan de activarse si reciben valores negativos constantes (Gabriel Gomila, 2024).

2.1.4.1.2.3 Función Tanh (Tangente hiperbólica)

La función Tangente hiperbólica es similar a la sigmoidea, pero escala los valores de salida entre -1 y 1. Su ecuación sería la siguiente:

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

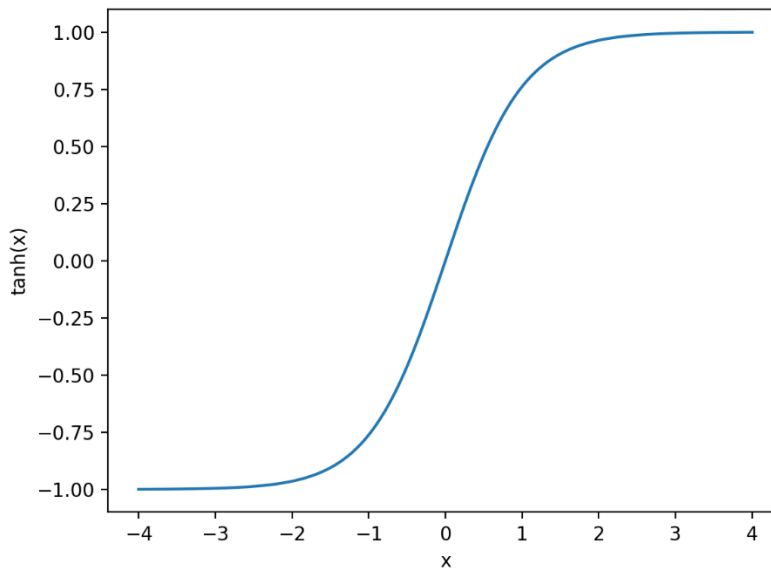


Ilustración 7: Gráfico de la función de activación Tangente Hiperbólica

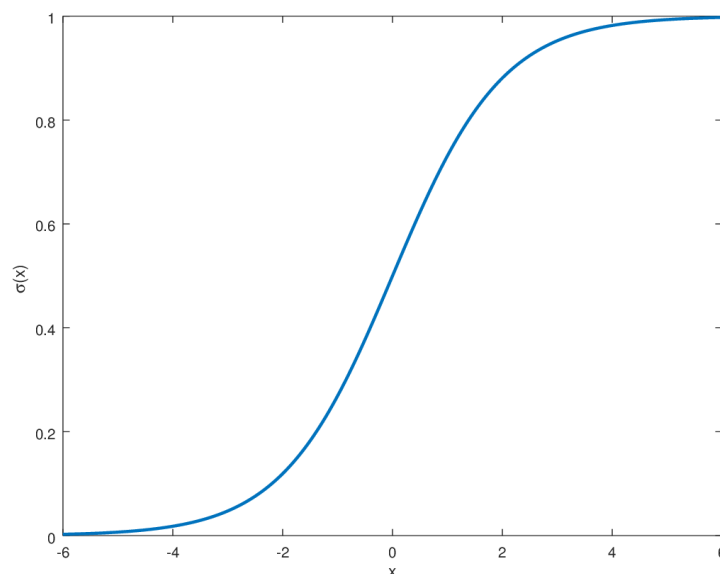
Fuente: Adaptado de Catedorial

La función de la tangente hiperbólica es más útil que la sigmoidea en algunas situaciones ya que la salida centrada en 0 puede ayudar a que las redes aprendan de manera más rápida, pero, al igual que la sigmoidea, también sufre del problema del desvanecimiento del gradiente (Gabriel Gomila, 2024).

2.1.4.1.2.4 Función Softmax

La función softmax es comúnmente utilizada en la última capa de una red neuronal multicategoría, ya que convierte las salidas en probabilidades para cada clase. Su ecuación es:

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}$$



Fuente: Adaptado de Alberto Rubiales en Medium

Ilustración 8: Gráfico de la función de activación SoftMax

La función softmax y la sigmoidea son funciones muy similares a la hora de representarlas gráficamente. La diferencia entre las dos, es debido a que la función sigmoidea se utiliza principalmente para clasificaciones binarias mientras que la función softmax se utiliza para clasificaciones multicategoría. El único problema de la función de activación softmax es que solo es adecuada para la capa de salida en problemas de clasificación de múltiples categorías (Gabriel Gomila, 2024).

2.1.4.1.3 ¿Cómo aprenden las redes neuronales?

El proceso de aprendizaje de las redes neuronales se basa en ajustar los pesos y sesgos de las conexiones entre las redes neuronales para mejorar su precisión. Este proceso comienza con la propagación hacia delante o feedforward, donde los datos pasan a través de varias capas. Cada neurona recibe los datos de las neuronas de la capa anterior, los pondera con los pesos asignados, les añade un sesgo y pasa el resultado por una función de activación. Este cálculo determina si la neurona se activa o no. Las salidas de una capa se convierten en las entradas de la siguiente, y así sucesivamente, hasta llegar a la capa de salida, que proporciona una predicción. Esta predicción se compara con el valor real a través de una función de costo, que mide la diferencia entre lo predicho y el resultado esperado.

El objetivo del aprendizaje es minimizar este error. A través de múltiples iteraciones, la red ajusta los pesos para reducir el error entre la predicción y el valor real. Este ajuste continuo, basado en la función de costo, permite que la red aprenda patrones cada vez más complejos en sus datos y mejore sus predicciones.

2.1.4.1.3.1 Backpropagation

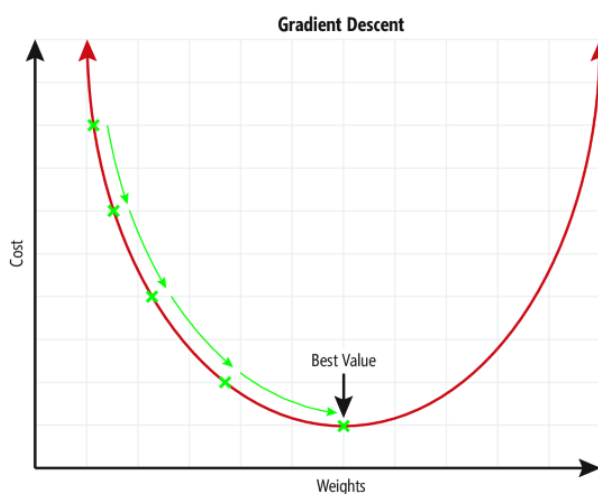
La propagación hacia atrás o backpropagation es un proceso clave en el entrenamiento de las redes neuronales que ajusta los pesos y sesgos de las conexiones entre las neuronas para mejorar la precisión del modelo. Este proceso compara las salidas actuales de la red con los resultados esperados y calcula la diferencia o error mediante la función de costes explicada en el punto anterior. Inicialmente, los pesos y sesgos suelen estar asignados de forma aleatoria, lo que genera un error elevado al inicio del entrenamiento.

Una vez calculado el error, la backpropagation ajusta los pesos hacia atrás a través de la red, minimizando la diferencia entre el resultado calculado y el valor correcto. Cada ajuste de los pesos se realiza con el objetivo de reducir el error y mejorar la precisión de la red. Este proceso se repite en ciclos llamados epochs, donde la red ajusta sus parámetros continuamente hasta alcanzar una pérdida mínima, lo que indica que la red ha aprendido a hacer predicciones más precisas (La, 2019).

2.1.4.1.4 Gradiente descendiente

El gradiente descendiente es el proceso mediante el cual se ajustan los pesos de las neuronas en una red neuronal durante cada ciclo de entrenamiento o epochs. La idea es minimizar la función de costes obtenida en los diferentes ciclos. Al visualizar la función de costes, esta tiene una curva con puntos altos y bajos que representan el rendimiento de la red. El objetivo del descenso por gradiente es encontrar el punto más bajo de esta curva, el cual representaría el error más bajo, ajustando los pesos en la dirección correcta. (La, 2019)

El algoritmo utiliza la pendiente de la curva para determinar si debe aumentar o disminuir los pesos. Si la pendiente es negativa, se suman valores a los pesos, y si es positiva, se restan. La velocidad de aprendizaje determina la magnitud de los ajustes: si es demasiado alta, la red puede pasarse del mínimo; si es demasiado baja, el proceso de aprendizaje será muy lento. (La, 2019)



Fuente: Adaptado de Frank La La en Microsoft Learn

Ilustración 9: Gráfico de la función de pérdida con una curva sencilla

Por otro lado, el gradiente descendente estocástico es una variante del método que introduce aleatoriedad al proceso. En lugar de calcular el gradiente exacto utilizando todo el conjunto de datos, este enfoque utiliza un subconjunto o muestra aleatoria para ajustar los pesos. Aunque puede ser menos preciso, el descenso estocástico es más rápido y puede evitar que la red se quede atascada en mínimos locales, ayudando a encontrar el mínimo global de la función de pérdida. (La, 2019)

2.1.4.1.4.1 Explosión del gradiente

La explosión del gradiente es un problema que ocurre cuando los gradientes crecen de forma descontrolada durante el entrenamiento de redes neuronales, especialmente en redes muy profundas o en redes neuronales recurrentes (RNN). Esto sucede porque, a medida que los gradientes se propagan hacia atrás durante la retropropagación, pueden multiplicarse por grandes valores, lo que hace que aumenten exponencialmente. Como resultado, los pesos del modelo se actualizan con pasos demasiado grandes, lo que provoca que el modelo sea inestable y el proceso de entrenamiento no pueda converger correctamente.

Para evitar este problema, se suelen aplicar técnicas como el clipping de gradientes, que limita los valores máximos de los gradientes durante el entrenamiento, evitando que crezcan demasiado. También se puede utilizar funciones de activación más adecuadas, como ReLU, que reducen la posibilidad de que los gradientes exploten, así como optimizadores avanzados como Adam o RMSprop, que ajustan las tasas de aprendizaje de manera adaptativa y ayudan a estabilizar el entrenamiento. Estas soluciones permiten entrenar redes profundas de manera más efectiva y evitar la explosión del gradiente.

2.1.4.1.4.2 Desvanecimiento del gradiente

El desvanecimiento del gradiente es otro problema común en el entrenamiento de redes neuronales profundas, y especialmente en redes neuronales recurrentes (RNN). Ocurre cuando los gradientes se vuelven extremadamente pequeños a medida que se propagan hacia atrás a través de las capas de la red durante la retropropagación. Esto causa que los cambios en los pesos de las primeras capas sean muy pequeños, lo que hace que la red aprenda muy lentamente o no aprenda en absoluto en esas capas.

El desvanecimiento del gradiente es especialmente problemático en redes profundas y en aquellas que trabajan con secuencias largas, como las RNN, ya que las primeras capas dejan de recibir suficiente información para actualizarse correctamente. Este problema se agrava cuando se utilizan funciones de activación como la sigmoide o la tangente hiperbólica (tanh), que comprimen los valores de salida en un rango pequeño, haciendo que los gradientes se reduzcan aún más.

En resumen, las redes neuronales han revolucionado la manera en que los sistemas de inteligencia artificial procesan y aprenden a partir de los datos, permitiendo modelar relaciones complejas y realizar predicciones precisas. Sin embargo, dentro de las redes neuronales, existen

arquitecturas especializadas que han sido diseñadas para abordar problemas específicos de manera más eficiente.

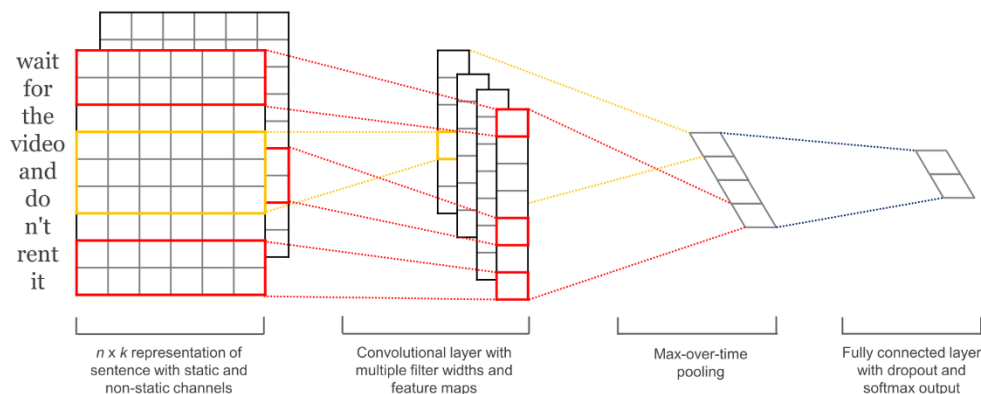
2.1.4.2 Redes neuronales convolucionales (CNN)

Las redes neuronales convolucionales (CNN) son un tipo específico de red neuronal diseñada principalmente para trabajar con datos tridimensionales, como imágenes, y son ampliamente utilizadas en tareas de clasificación de imágenes y reconocimiento de objetos. Las CNNs han sustituido los métodos manuales de extracción de características por un enfoque automatizado, que se basa en multiplicación de matrices para identificar patrones de una imagen. (IBM, 2024)

Aunque las CNNs están comúnmente asociadas a tareas de visión artificial, como el reconocimiento de imágenes, su capacidad para procesar datos secuenciales también las hace útiles en el campo del procesamiento de lenguaje natural. En PLN, las CNNs funcionan de manera similar a como lo hacen en la visión por computadora, pero en lugar de procesar píxeles de imágenes, trabajan con palabras o caracteres como datos de entrada. Antes de ser procesado por una CNN, el texto se convierte en vectores numéricos densos mediante técnicas de Word Embedding, lo que transforma las palabras en un formato adecuado para ser interpretado por una red neuronal (Korvage Information Technology, 2024).

Las CNNs suelen estar compuestas por múltiples capas, pero las más comunes son:

- **Capa convolucional:** Es el núcleo de la CNN, donde se realiza la mayor parte de los cálculos. Esta capa utiliza filtros o núcleos que recorren el texto para detectar características lingüísticas específicas como frases o combinaciones de palabras.
- **Capas de pooling:** Posteriormente, estas capas, reducen la dimensionalidad de los datos procesados por las capas convolucionales. Estas capas seleccionan los valores más importantes, lo que ayuda a destacar las características clave sin perder información crucial. Existen dos tipos principales: Max pooling, que selecciona el valor máximo en una región; y el average pooling, que calcula el promedio de los valores.
- **Capa completamente conectada:** En esta capa, cada neurona está conectada a las de la capa anterior, lo que permite clasificar los datos procesados por las capas convolucionales y de pooling. Generalmente, se utiliza una función **softmax** para generar probabilidades de clasificación (IBM, 2024) (Korvage Information Technology, 2024).



Fuente: Ilustrado de Danny Britz en Dannys Blogs

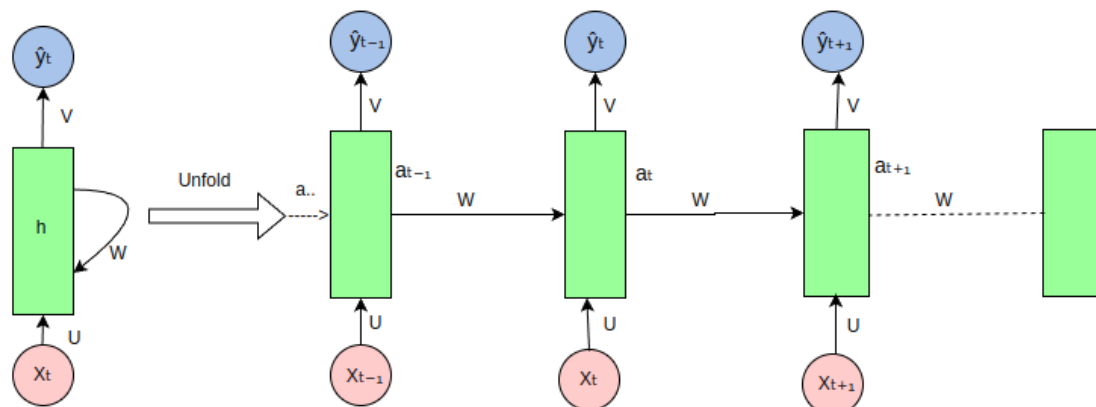
Ilustración 10: Esquema de una red neuronal convolucional para la clasificación de textos

2.1.4.3 Redes neuronales recurrentes (RNN)

Las redes neuronales recurrentes (RNN) son un tipo de red neuronal artificial que procesa datos secuenciales o series temporales. Son especialmente útiles para problemas en los que la secuencia o el orden de los datos es importante, como en el procesamiento de lenguaje natural, el reconocimiento de voz y la generación de subtítulos para imágenes. A diferencia de las redes neuronales tradicionales, las RNN son capaces de recordar información de entradas previas en la secuencia y utilizar esa información para influir en la salida actual, lo que les otorga una especie de “memoria”. Este enfoque es clave cuando el contexto es importante, como ocurre en frases o secuencias de sarcasmos.

Una de las principales características de las RNN es que comparten parámetros en cada capa de la red, lo que las distingue de las redes de propagación hacia delante. Utilizan un algoritmo llamado retropropagación a través del tiempo (BTPP) para ajustar sus pesos, similar a la retropropagación tradicional, pero optimizando secuencias de datos. Sin embargo, enfrentan desafíos como el desvanecimiento y explosión de gradientes, problemas que pueden limitar su capacidad de aprendizaje en secuencias largas (IBM, 2024).

Una red neuronal recurrente funciona recibiendo una entrada actual y un estado oculto que transporta información de los pasos anteriores. Esto permite que la RNN mantenga “memoria” de lo que ha sucedido previamente. La misma estructura se repite en cada paso de la secuencia utilizando los mismos pesos, lo que permite que la red capture dependencias temporales. Por último, cada paso, la RNN genera una salida basada tanto en la entrada actual como en la información acumulada.



Fuente: Adaptado de Sushmita Poudel en Medium

Ilustración 11: Esquema de una red neuronal recurrente

2.1.4.4 LSTM

Las Redes Neuronales Recurrentes básicas tienen limitaciones debido a su memoria de corto plazo, lo que hace que funcionen bien en secuencias cortas, pero fallen en secuencias largas, ya que la información inicial se va debilitando con cada paso.

Para superar estas limitaciones, se desarrollaron las redes LSTM (Long Short-Term Memory), que mejoran significativamente la capacidad de las redes para mejorar secuencias largas. Las LSTM introducen un componente clave llamado celda de estado, que actúa como una memoria a largo plazo. Esta celda permite que la red mantenga información relevante a lo largo del tiempo y deseche lo que no es necesario.

Además de la celda de estado otro componente fundamental son las compuertas. Las compuertas son mecanismos que controlan el flujo de información dentro de la red. Están compuestas por una función sigmoidea que genera valores entre 0 y 1, lo que le permite decidir cuanto de la información debe pasar a través de la red. Se diferencia entre los siguientes tipos de puerta:

- **Forget gate:** Decide qué información descartar de la memoria
- **Update gate:** Añade nueva información relevante a la celda de estado.
- **Output gate:** Filtra la información que se va a utilizar para generar el estado oculto actualizado y la salida final.

Estas características permiten que las LSTM manejen tanto memoria de corto plazo como a largo plazo, haciéndolas mucho mas efectivas en tareas donde las RNN fallan, como el procesamiento de lenguaje natural en secuencias largas. Además de resolver el problema del desvanecimiento del gradiente que comúnmente ocurría en las RNN tradicionales (Satoquirá, 2019).

2.1.4.5 Redes neuronales aplicadas a la detección de sarcasmos

Las redes neuronales profundas han supuesto un avance significativo en la detección de sarcasmos, mejorando considerablemente los enfoques tradicionales, como los basados en reglas y los clasificadores supervisados simples.

Uno de los grandes avances es la captura de patrones complejos. Los enfoques tradicionales, en muchas ocasiones dependen de un conjunto predefinido de normas para identificar palabras o frases que puedan ser sarcásticas. Sin embargo, el sarcasmo no siempre sigue unas reglas claras, ya que dependen en gran medida de contexto. Las redes neuronales profundas, mas en concreto las RNN y LSTM, son capaces de capturar dependencias complejas en los datos de secuencias largas, algo que los enfoques tradicionales no pueden hacer de manera eficiente. Esto permite que los modelos profundos detecten sarcasmo incluso cuando las señales de sarcasmo están distribuidas a lo largo del texto.

Otro paso hacia delante fue la generalización y el aprendizaje automático. A diferencia de los enfoques tradicionales, que requieren intervención manual para diseñar reglas y características específicas, las redes neuronales profundas pueden aprender patrones automáticamente a partir de grandes volúmenes de datos etiquetados. Modelos como CNN pueden identificar características locales del texto, como combinaciones de palabras que sugieren sarcasmo, mientras que las RNN y LSTM son capaces de entender relaciones mas profundas, lo que les permite generalizar mejor en datos nuevos y desconocidos.

Por último, los enfoques tradicionales, como los diccionarios de palabras clave, dependían de identificar palabras o frases que ya se sabe que son sarcásticas. Esto limita su efectividad a casos explícitos. Las redes neuronales profundas, al capturar contextos más amplios, son capaces de detectar sarcasmos implícitos, donde no hay palabras clave evidentes, pero el contexto sugiere sarcasmo (Kumar, 2019).

2.1.4.6 Limitaciones del Deep Learning en la detección de sarcasmos

Aunque el Deep learning ha permitido avances significativos en la detección de sarcasmos, aun presenta varias limitaciones que afectan a su rendimiento en esta tarea tan compleja:

En primer lugar, el sarcasmo depende mucho del contexto completo de una conversación u interacción, lo que implica que las redes neuronales deben captar relaciones entre palabras o frases que están distantes en una secuencia. Las RNN y LSTM pueden tener dificultades para mantener estas relaciones a lo largo de secuencias muy largas, lo que limita su capacidad para detectar sarcasmos de manera precisa cuando la clave esta distribuida en diferentes partes de la conversación (Pawar, 2024).

Uno de los mayores inconvenientes de las RNN y LSTM sería el procesamiento de forma secuencial, ya que impide la paralelización del entrenamiento. Esto resulta en una alta demanda computacional y ralentiza el entrenamiento, especialmente en secuencias largas. Si bien el LSTM mejora el manejo de las dependencias a largo plazo respecto a las RNN, todavía no pueden capturar de manera eficiente todas las relaciones en secuencias muy largas. Esto es crucial para

la detección de sarcasmos, donde el significado puede estar distribuido a lo largo de toda la oración o conversación.

Otra limitación se encontraría en la generación de embeddings contextuales. Aunque si bien es cierto que los LSTM y los RNN pueden generar embeddings que capturan el contexto, estos no son los suficientemente expresivos en comparación con enfoques mas recientes. Estos modelos a menudo generan representaciones de palabras que a menudo no logran capturar todas las sutilezas y matices que permiten identificar sarcasmo en situaciones complejas. (Pawar, 2024)

En conclusión, estos enfoques basados en redes neuronales y deep learning no solo abordan las carencias de los modelos secuenciales tradicionales, sino que también ofrecen una forma mas potente y eficiente de procesar secuencias de texto, haciendo posible la detección de sarcasmo de manera mas precisa y efectiva. Sin embargo, entre estos avances, ha surgido una arquitectura que ha revolucionado el campo del procesamiento del lenguaje natural. Este modelo se conoce como Transformers y en el siguiente punto se expondrá sus características y por que ha revolucionado el campo del procesamiento de lenguaje natural.

2.1.5 Modelos basados en Transformers

Dentro del ámbito del Deep learning, existen una variedad de modelos llamados Transformers, los cuales han revolucionado la forma en que procesamos el lenguaje natural. Un transformer es una red neuronal diseñada para aprender el contexto de datos secuenciales y generar nuevos datos a partir de ello. Explicado de una forma más sencilla, se trata de un modelo de inteligencia artificial que entiende y genera texto similar al humano mediante el análisis de patrones en grandes cantidades de datos textuales (Ferrer, 2024).

Lo que distingue a los Transformers de los modelos anteriores como las redes neruoanles recurrentes y las redes de memoria a largo plazo, es su capacidad para aprender contextos complejos sin necesidad de usar recurrencias. Mientras que el modelo clásico encoder-decoder depende principalmente de las RNNs para extraer información secuencial, los Transformers se basan en una técnica matemática conocida como attention (Ferrer, 2024). Esta técnica fue introducida por los investigadores de Google en el revolucionario e influyente articulo “Attention isa ll you need” en 2017, permitiendo al modelo comprender las relaciones entre diferentes elementos en una secuencia de datos sin necesidad de procesar palabra por palabra de manera secuencial (Vaswani et al., 2017). Este articulo marco un antes y un después en el procesamiento de lenguaje natural, proponiendo una arquitectura innovadora que elimino la dependencia de la recurrencia.

Los Transformers destacan principalmente por su capacidad de paralelización, permitiendo el procesamiento simultaneo de secuencias de entrada. A diferencia de las RNNs o LSTM, en donde el entrenamiento debe realizarse de forma secuencia, los Transformers logran una mayor eficiencia al entrenar y procesar los datos en paralelo resultando en un cómputo más rápido y escalable. Asimismo, utilizan el mecanismo de self-attention para capturar dependencias de largo alcance en las secuencias de entrada. Esto les permite considerar todas las posiciones de

una secuencia simultáneamente, lo cual genera representaciones contextuales más ricas y precisas que las utilizadas en los modelos tradicionales (Pawar, 2024).

Otra característica fundamental de los Transformers es su capacidad para procesar secuencias de manera bidireccional, lo que significa que considera tanto el contexto anterior como el posterior de una palabra en la secuencia. Este enfoque permite a los Transformers mejorar las sutilezas semánticas del lenguaje (Pawar, 2024).

Para entender de forma detallada el significado y funcionamiento de una Transformer es crucial entender la arquitectura de este. A continuación, se explicará su estructura desde el momento en que recibe el input hasta la generación del output, desglosando cada uno de los componentes clave que hacen posible el procesamiento eficiente de secuencias y la capacidad de generar resultados precisos.

2.1.5.1 Arquitectura de los Transformers

Para comprender a fondo cómo funciona un modelo basado en Transformers, es fundamental realizar un análisis profundo de su arquitectura. Esto implica examinar sus dos componentes principales: el encoder y el decoder. Además, es necesario analizar detalladamente cada capa por la que pasarán nuestros datos, comenzando por el embedding, seguido del positional encoding y la atención multi-cabeza, hasta llegar a la red completamente conectada feed-forward. Por lo tanto, desglosaremos esta arquitectura paso a paso para entender cada uno de sus elementos en detalle.

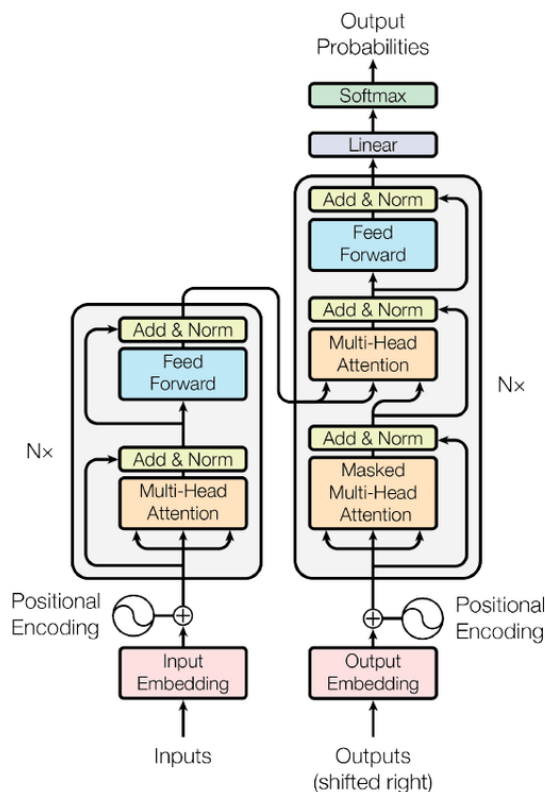


Ilustración 12: Arquitectura del modelo Transformer de "Attention is All You Need".

Fuente: Vaswani et al. (2017), Attention is All You Need.

2.1.5.1.1 Positional encoding

A diferencia de las redes neuronales recurrentes, los Transformers no introducen palabra a palabra de manera secuencial al modelo, sino que pasan todas las palabras de manera simultánea. Al realizar la ingesta de datos de forma simultánea, se necesita una forma de inyectarle información relativa o absoluta de la posición de los tokens en la frase. Es por ello que se añade un Positional encoding al principio tanto de encoder como del decoder. Es necesario que los Positional encodings tengan la misma dimensión que los embeddings, para que se puedan sumar (Vaswani et al., 2017). Según el paper original donde se presentó por primera vez la arquitectura de los Transformers completa se sugirieron estas dos fórmulas para calcular estos "positional embeddings", una para las posiciones pares y otra para los impares.

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{1000^{\frac{2i}{d_{model}}}}\right)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{1000^{\frac{2i}{d_{model}}}}\right)$$

En la formula, pos sería la posición de la palabra en la frase e i la dimensión. En esta aproximación, cada dimensión es representada por frecuencias y desplazamientos únicos de la

onda, con valores que van de -1 a 1, representando efectivamente una posición (Vaswani et al., 2017 & Ferrer, 2024).

2.1.5.1.2 Multihead attention

La Multi-Head Attention es un componente clave dentro de los modelos Transformer ya que permite al modelo enfocarse en diferentes partes de la secuencia de entrada de manera simultánea. Esto se logra aplicando varias capas de Scaled Dot-Product Attention de forma paralela, lo que aumenta la capacidad del modelo para capturar patrones complejos en los datos de secuencia.

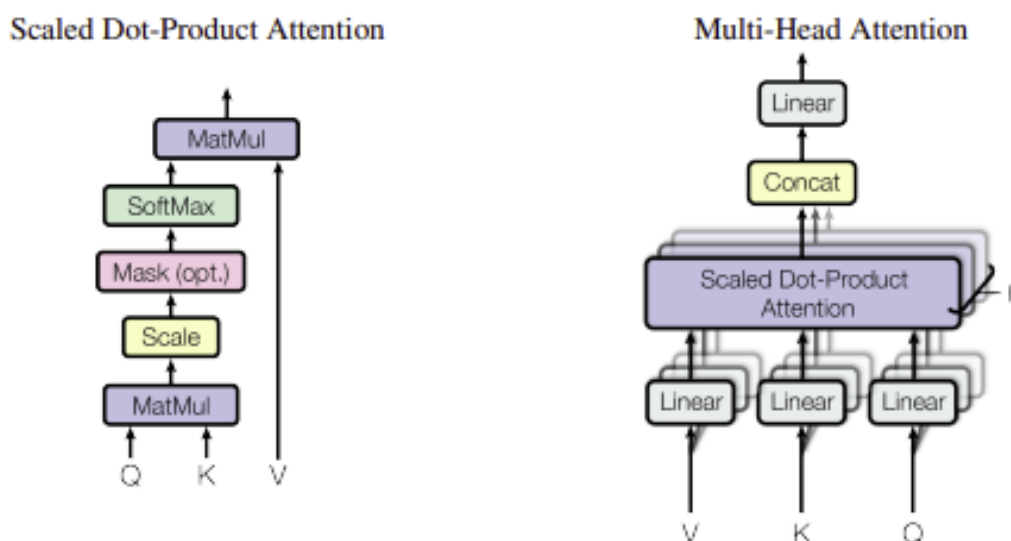


Ilustración 13: Atención de producto punto escalado y atención multi-cabeza.

Fuente: Vaswani et al. (2017), Attention is All You Need.

El Scaled Dot-Product, el cual se muestra en nuestra primera imagen, se construye usando tres matrices clave:

- **Queries:** Son vectores que representan lo que cada palabra busca en las demás palabras de la secuencia.
- **Keys:** Son vectores que describen cada palabra de la secuencia y actúan como “etiquetas”
- **Values:** Son los vectores que contienen la información que será transmitida a la siguiente capa.

El proceso comienza multiplicando las queries por las keys, lo que resulta en una matriz de atención que mide la similitud entre cada par de elementos de la secuencia. Este producto luego pasa por un escalado el cual consiste en dividirlo por la raíz cuadrada de la dimensión de las Keys para estabilizar los gradientes, y luego se aplica la función softmax para obtener las puntuaciones de atención. Finalmente, estas puntuaciones se usan para ponderar los valores, produciendo el resultado ponderado de la atención. Esta sería su función representada matemáticamente:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Este proceso de atención escalada se realiza varias veces de manera paralela, con diferentes proyecciones lineales par la queries, keys y values. Es decir, antes de aplicar la atención escalada, cada cabeza aplica una transformación lineal a los vectores Q,K y V. Esto permite que diferentes “cabezas” de atención se centren en distintas partes de la secuencia. Por último, los resultados de cada cabeza se concatenan y pasan por otra capa lineal para producir la salida final. (Vaswani et. al, 2017)

2.1.5.1.3 Feed forward

La capa Feed Forward en un Trasformer es una red simple que toma cada posición de la secuencia por separado. No hay interacción entre las posiciones, lo que significa que cada palabra o token se procesa de forma individual, y esta capa consiste en dos transformaciones lineales separadas por una función no líneas, generalmente una función ReLU. La función de la feed-forward es modificar las representaciones generadas por las capas de atención, permitiendo que el modelo capture patrones más complejos (Vaswani et. al, 2017).

Simplificando, si el transformador primero presta atención a las relaciones entre palabras, las capa feed-forward se asegura de que cada palabra pase por un ajuste adicional para mejorar la representación del contexto global. (Geva, Schuster, Berant, & Levy, 2021)

2.1.5.1.4 Enconder

En el contexto de los Transformers, el encoder actúa procesando la secuencia de entrada en capas repetidas que optimizan las representaciones de cada palabra. Este flujo se inicia cuando los inputs embeddings se combinan con el positional encoding, el cual introduce la información sobre el orden de las palabras. Este paso inicial garantiza que el modelo, el cual no es secuencial, tenga en cuenta las posiciones relativas de las palabras.

Una vez se ha añadido esta información posicional, la secuencia para a la capa Multi-Head Attention, donde cada palabra se compara con todas las demás. El modelo hace que se preste atención a diferentes aspectos de las relaciones entre palabras en paralelo, mejorando así la capacidad de entender contextos complejos, sin importar la distancia entre palabras.

A continuación, a la capa de atención se le aplica una conexión residual que suma la entrada original con esta salida, y luego se normaliza. Este proceso, llamado Add & Norm, estabiliza el flujo de información y evita que las representaciones sean alteradas drásticamente, manteniendo una coherencia en la transformación de los datos.

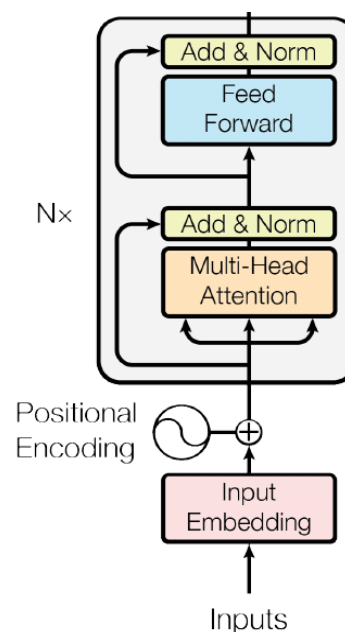


Ilustración 14: Bloque Encoder en la arquitectura Transformer

Fuente: Vaswani et al. (2017), Attention is All You Need.

Después de la normalización de la capa de atención, cada palabra se procesa individualmente en la capa feed-forward, la cual refina las representaciones obtenidas. Este paso ajusta los vectores de manera mas precisa, haciendo que las representaciones sean más útiles para tareas posteriores. Al igual que en el bloque anterior, otra conexión residual y normalización garantizan que las palabras no pierdan las características aprendidas anteriormente.

Este proceso se repite N veces (generalmente 6 como viene en el paper original o más), lo que permite que el modelo vaya afinando cada vez mas las representaciones. La salida final del encoder no solo tiene la información de cada palabra por separado, sino también como se relaciona con todas las demás en la secuencia, gracias al mecanismo de atención (Vaswani et. al, 2017).

2.1.5.1.5 Decoder

En el contexto de los Transformers, el decoder funciona generando secuencias de salida basándose en la información que ya ha procesado el encoder. El flujo de informacion comienza cuando los inputs embeddings, que representan los tokens de salida generados hasta ese momento, se combinan con el positional encoding, permitiendo que el decoder tenga en cuenta el orden de los tokens ya generados, pero en decoder es distinto ya que estas palabras tienen un desplazamiento o “shift” que evita que vea los futuros tokens generados.

A continuación, estos datos ya embebidos pasan a la Masked Multi-Head Attention, donde el decoder presta atención a los tokens ya generados, pero no a los futuros, gracias al enmascarado que impido al modelo tener acceso a los tokens que aún no han sido predichos. Esta atención enmascarada permite que el modelo genere secuencias de manera coherente, procesando cada palabra.

Una vez aplicada la atención enmascarada, se aplica una conexión residual, al igual que en el encoder, que suma la entrada original a la salida de esta atención enmascarada, estabilizando y refinando la representación. Asimismo, la información pasa a una segunda capa de Multi-Head Attention, que permite que cada token generado preste atención a las representaciones de la salida del encoder. Esto asegura que el decoder se base tanto en lo que ya ha generado como en la secuencia de entrada procesada por el encoder.

Después de la atención, se aplica una segunda capa de conexión residual para que la información aprendida en las capas anteriores no quede opacada por la nueva información aprendida.

Una vez la información ya está procesada, se pasa por la capa feed-forward, que ajusta las representaciones a través de una serie de transformaciones no lineales. Al igual que en el encoder, se utiliza una

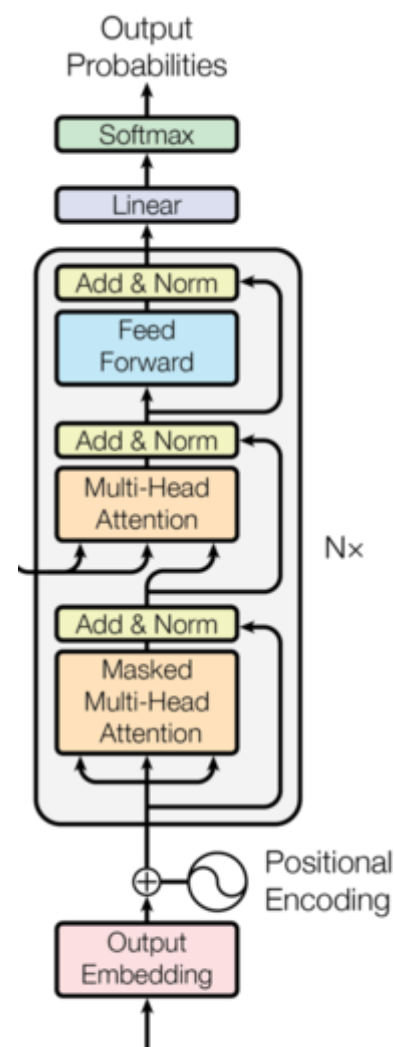


Ilustración 15: Bloque Decoder en la arquitectura Transformer.

Fuente: Vaswani et al. (2017), Attention is All You Need .

conexión residual y normalización, asegurando que las representaciones no se alteren drásticamente y preserven la coherencia de la información.

Por último, se vuelve a aplicar una tercera capa de conexión residual para evitar que la información adquirida en las capas anteriores se pierda en el olvido.

En la salida del decoder, después de pasar por las capas de atención y feed-forward, los resultados pasan por una capa lineal y luego por una función softmax. La capa lineal transforma las representaciones generadas en vectores que coinciden con el tamaño del vocabulario, asignando una puntuación a cada palabra posible en el vocabulario. Luego, la función softmax convierte estas puntuaciones en probabilidades, lo que permite que el modelo seleccione la palabra más probable como la salida siguiente en la secuencia. Este proceso se repite hasta que se genera toda la secuencia final (Vaswani et. al, 2017).

2.1.5.2 Fine-Tuning

Como se ha podido observar analizando la arquitectura de los Transformers, la cual involucra múltiples operaciones matemáticas y estructuras avanzadas de atención, surge una necesidad clave, el poder aprovechar estos modelos ya pre-entrenados para tareas específicas sin tener que empezar desde cero. Aquí es donde entra en juego el fine-tuning. El fine-tuning permite adaptar modelos ya entrenados, ajustando finamente sus parámetros optimizados en nuevas tareas. Esto es especialmente útil cuando se dispone de modelos preentrenados en grandes cantidades de datos generales, ya que facilita su ajuste para realizar tareas especiales con menores recursos y tiempo de entrenamiento. (Wikipedia, 2024).

Todo este proceso se basa en el transfer learning, donde un modelo previamente entrenado en una tarea más general se reutiliza y ajusta para realizar una tarea más específica y relacionada, lo que evita tener que entrenar un modelo desde 0 cero.

Existen diferentes enfoques de fine-tuning. El más común es el full fine-tuning, donde se actualizan todos los parámetros del modelo utilizando un conjunto de datos específico para la nueva tarea. Este método es efectivo, pero puede ser computacionalmente costoso, especialmente en Transformers. Para evitar estos altos costos, se pueden usar métodos más eficientes, como el partial fine-tuning, donde solo se ajustan algunas capas del modelo, mientras que las capas restantes no se actualizan (Bergmann, 2024)

Otra opción podría ser el parameter-efficient fine-tuning (PEFT), que reduce la cantidad de parámetros que deben actualizarse, lo que disminuye la demanda de recursos y hace que el proceso sea más estable. Este método es particularmente útil en modelos de lenguaje natural como BERT o GPT, donde se actualizan pequeñas partes del modelo (Bergmann, 2024).

2.1.5.3 Tipos de Transformers

En el mundo de los Transformers, existen muchas variantes de tipos de modelos, pero existen tres tipos que sobresalen sobre el resto. Estos son los modelos basados en encoders, los basados en decoders y los basados en encoder-decoder. Cada uno tiene sus propias características que lo hacen ideales para diferentes tareas en el procesamiento de lenguaje natural.

2.1.5.3.1 Modelos basados en encoders

En primer lugar, podemos encontrar los modelos basados en encoders, los cuales están diseñados para transformar las secuencias de entrada en representaciones de alta calidad, capturando tanto el contexto local como el global. La ventaja de estos modelos con respecto a sus antecesores, como LSTM, es la capacidad que tienen de procesar los tokens de manera simultánea. Al tener en cuenta tanto las palabras anteriores como las posteriores de una oración, los modelos basados en encoders son excelentes para tareas en donde es crucial comprender el significado global de un texto. Estos modelos suelen encontrar su utilidad en tareas como clasificación, análisis de sentimiento o hallar el contexto de un párrafo.

Uno de los modelos más representativos en esta categoría sería BERT. Este modelo es entrenado mediante una tarea llamada máscara de lenguaje, la cual le hace aprender a predecir palabras que faltan en una secuencia, lo que le permite capturar relaciones profundas entre palabras. Su capacidad bidireccional le permite mejorar significativamente los resultados en comparación con otros modelos que solo procesan la secuencia de manera unidireccional (Soren, 2023). Otro modelo basado en encoders es RoBERTa, una versión mejorada de BERT que ajusta la técnica de preentrenamiento para obtener mejores resultados en varias tareas. (Von platen, 2020)

2.1.5.3.2 Modelos basados en decoder

Los modelos basados en decoder están orientados principalmente a la generación de texto. En lugar de generar una representación compacta de una secuencia, como lo hacen los encoders, los decoders generan secuencias de salida de forma auto generativa, es decir, cada palabra generada depende de las palabras anteriores. Esto lo hace ideal para tareas como: traducción automática, asistencia virtual o diálogos automáticos. Una de las características clave de los modelos basados en decoders es que utilizan una atención enmascarada para evitar que el modelo pueda ver las palabras futuras en la secuencia durante el proceso de generación, así se garantiza que el modelo prediga a partir de las palabras anteriores (Soren, 2023)

El modelo más destacado en esta categoría es el conocido como GPT (Generative Pre-trained Transformer), cuya versión más reciente GPT-4, ha demostrado altas capacidades a la hora de generar texto coherente y relevante a partir de instrucciones simples.

2.1.5.3.3 Modelos basados en encoder-decoder

Por último, podemos encontrar los modelos que mantienen la arquitectura original presentado en el primer paper donde se habla de los transformers “Attention is all you need”. En esta arquitectura se combina tanto el encoder como el decoder en una arquitectura completa. Este diseño es particularmente útil en las tareas de secuencia a secuencia, como la traducción automática, donde una secuencia de entrada en un idioma debe transformarse en una secuencia de salida en otro idioma. El encoder transforma la secuencia de entrada en una representación intermedia, que luego el decoder utiliza para generar la secuencia de salida en el idioma al que se quiere traducir el texto original. (Hurtado, 2024)

Este tipo de arquitectura es especialmente eficiente porque permite que el encoder capture la mayoría de la información contextual de la secuencia de entrada, mientras que el decoder

genera la salida palabra por palabra basándose en la representación del encoder y las palabras generadas hasta el momento. BART Y T5 son dos de los modelos más utilizados que tienen este tipo de arquitectura. T5 convierte todas las tareas de procesamiento de lenguaje natural en una tarea de transformación de texto a text y BART es un modelo diseñado para realizar tareas como la generación de resúmenes, en donde el encoder recibe un texto como secuencia de entrada y el decoder lo transforma en una secuencia de salida más breve (Von platen, 2020).

2.1.5.4 Transformers aplicados a la detección de sarcasmos

Los Transformers han revolucionado tanto el campo del procesamiento del lenguaje natural como el de la detección de sarcasmos. Esto lo ha conseguido superando las limitaciones que sufrían los modelos utilizados previamente como las RNNs o LSTMs. Mientras que los sarcasmos dependen en gran medida del contexto y de relaciones implícitas entre palabras, las RNN y LSTM tienen bastantes problemas para poder capturar relaciones a largo plazo debido a su naturaleza secuencial. Estos modelos enfrentan varias dificultades como el desvanecimiento del gradiente o la explosión del gradiente, lo que reduce su capacidad para interpretar correctamente el sarcasmo en frases donde el contexto cambia abruptamente o se opone al significado literal. (Spain AI, 2021).

Por otro lado, los Transformers eliminan la dependencia secuencial mediante su mecanismo de atención, lo que permite capturar el contexto global de la oración y asignar importancia a diferentes partes de la secuencia sin importar la posición de los tokens. Este enfoque es imprescindible para la detección de sarcasmos, ya que una palabra puede adquirir un tono sarcástico debido a otra palabra en otra parte de la oración. (Alexandros, Siolas, & Stafylopatis, 2019)

Un ejemplo destacado de estas mejoras en la detección de sarcasmos se puede encontrar en el estudio realizado por Santosh Kumar Bharti llamado “Sarcasm detection in textual data: A supervised aproach” en donde utilizaron una variante de BERT para la detección de sarcasmos en Twitter. Este modelo fue ajustado específicamente con datos de redes sociales y logro superar el rendimiento a los enfoques tradicionales basados en LSTM, mostrando una mejora significativa en la clasificación de sarcasmos.

Por otro lado, RoBERTA, una variante del modelo de BERT, también ha sido utilizada con éxito en la detección de sarcasmos en conversaciones largas en plataformas como Reddit. En un estudio reciente, RoBERTa se entrenó en una combinación de conjuntos de datos de Twitter y Reddit, logrando una mayor precisión que los modelos de LSTM en la clasificación de sarcasmos dentro de los hilos de la conversación. Esta arquitectura permitió nop solo procesar el contenido, sino tambien el contexto mas amplio en el que se producen los comentarios (Dong, 2024)

Ademas de los modelos basados en encoders, también se han explorado modelos basados en decoders para la generación y detección de sarcasmos. GPT-3, por ejemplo, ha sido entrenado para la detección de textos sarcastico (Chiu, Collins, & Rohan, 2022). En este caso se han basado en un conjunto de comentarios realizados en youtube y reddit permitiendo que el modelo detectara el odio y sarcasmo en el conjunto de datos.

Los modelos encoder-decoder, como el T5, también han sido utilizados para la detección y generación de textos sarcásticos, especialmente en tareas de secuencia a secuencia. Un enfoque interesante es el uso de T5 para traducir una oración literal en una "traducción" sarcástica, permitiendo identificar patrones de sarcasmo que no serían obvios de otra manera. Esta capacidad de T5 para transformar una secuencia de entrada en otra secuencia ha sido eficaz en tareas de generación de resúmenes sarcásticos y clasificación de sarcasmos. (Tummala, 2024)

La investigación de Ghosh y Veale demostró que el uso de modelos híbridos como CNN-LSTM en combinación con embeddings preentrenados, aunque efectivo, no puede igualar la capacidad de los Transformers para procesar grandes cantidades de datos y capturar tanto las relaciones contextuales locales como globales dentro de una secuencia de texto. Este estudio subrayó las limitaciones de los enfoques más tradicionales frente a los avances recientes con Transformers. (Ghosh & Veale, 2017)

En resumen, los Transformers han transformado significativamente la detección de sarcasmos al superar las limitaciones que anteriormente presentaban modelos como las RNN y LSTM. Estos modelos tradicionales dependían de la secuencialidad y enfrentaban dificultades para capturar dependencias a largo plazo, lo que afectaba su capacidad para comprender el contexto global, crucial para identificar sarcasmos. En cambio, los Transformers, al utilizar mecanismos de atención, han demostrado una capacidad superior para procesar secuencias de texto completas, captando las relaciones entre palabras independientemente de su distancia en la oración. Esto es especialmente relevante en el sarcasmo, donde el tono implícito y el contexto son esenciales para su interpretación.

Modelos como BERT y RoBERTa han permitido una mejor comprensión del contexto bidireccional, mientras que GPT-2 y T5 han introducido mejoras tanto en la generación como en la detección de sarcasmos, lo que les ha permitido interpretar el lenguaje figurado y las contradicciones presentes en el sarcasmo. Estos modelos no solo pueden clasificar con precisión el sarcasmo en textos, sino que también pueden generar respuestas sarcásticas de manera coherente y fluida, mostrando una mayor adaptabilidad y eficiencia en comparación con los enfoques anteriores basados en redes recurrentes.

Además, la flexibilidad de los Transformers para entrenarse en grandes cantidades de datos y capturar tanto el contexto local como global ha demostrado ser crucial en aplicaciones de redes sociales y diálogos extensos, donde el sarcasmo es común. Los estudios recientes, como los que utilizan T5 en la traducción y detección de sarcasmos o el uso de RoBERTa en hilos de conversación, han mostrado cómo los Transformers han superado los enfoques previos al captar las sutilezas del lenguaje sarcástico. En conjunto, estas mejoras han permitido que la detección automática de sarcasmos alcance un nivel de precisión sin precedentes, consolidando a los Transformers como la arquitectura de referencia en este campo.

2.1.5.5 Limitaciones del uso de Transformers a la detección de sarcasmos

A pesar de todos los avances significativos que han traído los Transformers en la detección de sarcasmos, siguen enfrentándose a diversas limitaciones tanto generales como específicas del sarcasmo.

En primero lugar, uno de los mayores obstáculos en los Transformers es su alta demanda de recursos computacionales. Los Transformers, requieren enormes cantidades de memoria y capacidad de procesamiento debido a su complejidad en el mecanismo de atención, el cual aumenta conforme aumenta las secuencias de entrada. Este problema puede limitar mucho a su uso en aplicaciones en tiempo real o a usuarios que quieran hacer un modelo de este calibre en una computadora convencional.

Los Transformers también demandan una gran cantidad de datos para su entrenamiento, lo que es una desventaja en áreas donde los datos etiquetados son escasos, como en la detección de sarcasmos en español. En muchas ocasiones, los datos de sarcasmos en español disponibles son escasos o incluso nulos, lo que reduce la capacidad de los modelos para generalizar. Es por ello que es casi obligatorio para un usuario tener que realizar un fine-tuning para poder obtener resultados decentes.

El sarcasmo, en gran medida, depende principalmente de factores pragmáticos que van más allá del texto literal. Aunque los Transformers capturan bien el contexto, no son tan eficaces cuando se trata de entender las intenciones del hablante o el contexto social más amplio en el que se produce una conversación. Por ejemplo, una frase que parece literal dentro de un texto puede ser interpretada como sarcástica solo cuando se conoce el tono o el contexto extralingüístico, algo que los Transformers no pueden procesar.

Otra limitación sería la dependencia del texto literal. El sarcasmo muchas veces implica que el significado literal de una oración se contradice con su tono subyacente. Los Transformers, aunque capturan bien las relaciones entre palabra en un texto, tienden a depender demasiado de un texto literal, lo que puede hacer que no identifiquen correctamente el sarcasmo cuando el significado real está implícito en el tono o la intención del hablante.

Asimismo, en muchas ocasiones no se es capaz de diferenciar entre sarcasmo e ironía. Los transformes tienen dificultades para diferente ciar entre estos dos, ya que, aunque están relacionadas, no son exactamente lo mismo. El sarcasmo suele tener una intención crítica o negativa con tono de burla, mientras que la ironía puede ser más sutil. Debido a las similitudes estructurales entre ambas formas del lenguaje figurado, los Transformers a menudo confunden estos dos fenómenos lingüísticos.

Por último, en gran parte de las ocasiones, los sarcasmos no solo dependen de las palabras, sino también de señales no textuales como el tono de voz, expresiones faciales o lenguaje corporal. Los Transformers, los cuales trabajan principalmente con texto, no pueden interpretar estas señales, lo que limita su capacidad para identificar sarcasmos.

En conjunto, aunque los Transformers han mejorado significativamente la detección de sarcasmos al superar las limitaciones de modelos anteriores como las RNN y LSTM, todavía presentan desafíos importantes. Desde los altos requerimientos computacionales hasta la incapacidad de capturar el contexto pragmático y las señales multimodales, estos modelos tienen limitaciones que requieren ser abordadas para mejorar su aplicabilidad y eficiencia en la detección de sarcasmos en diferentes entornos.

2.2 Contexto y justificación

La detección de sarcasmos en el procesamiento de lenguaje natural es una tarea compleja y desafiante. El sarcasmo como hemos explicado en el apartado anterior, es una forma de comunicación en la que lo que se dice literalmente no coincide con la intención real del hablante, lo que lo convierte en un fenómeno lingüístico muy difícil de identificar. La dificultad se encuentra en que el sarcasmo a menudo requiere un profundo entendimiento del contexto y relaciones invisibles entre palabras, así como el conocimiento del contexto social.

El crecimiento de las interacciones en línea ha hecho que la detección del sarcasmo sea una necesidad en plataformas como redes sociales, foros y sistemas de chat, donde la ausencia de señales no verbales, como el tono de voz o las expresiones faciales, dificultan aún más la interpretación del significado. Además, el uso del sarcasmo suele ser bastante común en conversaciones sobre política, críticas sociales o conversaciones humorísticas, lo que aumenta la importancia de desarrollar herramientas automáticas que puedan comprender y clasificar este fenómeno.

Este proyecto tiene como objetivo abordar las limitaciones actuales en la detección automática de sarcasmos mediante el uso de modelos basados en Transformers. Estos modelos han demostrado ser superiores a enfoques más rudimentarios. Sin embargo, la detección de sarcasmos sigue siendo un área en la que hay mucho margen de mejora, especialmente cuando se trata de entender el contexto cultural en el que se desarrolla el lenguaje.

Además, la falta de conjuntos de datos etiquetados de sarcasmo en español limita la capacidad de los modelos para generalizar a diferentes contextos. Este proyecto no solo aborda estos desafíos mediante el uso de técnicas avanzadas, sino que también busca contribuir al desarrollo de nuevas bases de datos que optimicen la clasificación de sarcasmos. La aplicación de estos avances mejorará significativamente los sistemas de moderación de contenido, análisis de sentimiento y otras interacciones con asistentes virtuales, donde la comprensión del sarcasmo es fundamental para ofrecer respuestas más contextualmente adecuadas.

2.2.1 Importancia de la detección automática del sarcasmo

La detección automática del sarcasmo ha ganado importancia en el procesamiento de lenguaje natural debido a su impacto en las interacciones en línea y la moderación del contenido. El sarcasmo es un recurso comunicativo que introduce complejidad adicional al lenguaje, ya que la interacción real del hablante a menudo contradice el significado literal de las palabras. Esto lo convierte en un desafío único para los sistemas automatizados, que generalmente se basan en el análisis literal del texto. La falta de detección precisa del sarcasmo puede dar lugar a malentendidos, generar conflictos y reducir la eficacia de sistemas automatizados como los chatbots. (Fu et. al, 2024)

2.2.1.1 Relevancia del sarcasmo en relaciones humanas en línea

El sarcasmo juega un papel fundamental en las interacciones humanas, tanto en la vida cotidiana como en los entornos en línea. Se ha demostrado que el sarcasmo, al ser una forma indirecta de comunicación, promueve la creatividad y mejora la resolución de problemas en aquellos que lo

utilizan e interpretan. Las personas que utilizan el sarcasmo suelen establecer relaciones mas cercanas, ya que compartir este tipo de humor permite una mayor conexión y complicidad entre los interlocutores. El sarcasmo también puede servir como un mecanismo para suavizar críticas o comentarios negativos, haciéndolos más aceptables socialmente (Gino, 2015).

Otra característica que se observa en personas que utilizan mas el sarcasmo es el de un procesamiento cognitivo mas avanzado, ya que se debe entender tanto el significado literal de las palabras como el mensaje implícito que se transmite. Esto crea una oportunidad para que los individuos que lo usan o reciben desarrollen mejores habilidades cognitivas.

En el ámbito digital, el sarcasmo ha ganado relevancia debido al aumento de las interacciones a través de plataformas como Twitter, Facebook y los distintos foros de discusión. Sin embargo, a diferencia de las interacciones en persona, las señales no verbales están ausentes, lo que hace que el sarcasmo sea más difícil de detectar. Esta falta de señales visuales y auditivas incrementa la probabilidad de que los comentarios sarcásticos se malinterpreten, lo que puede generar conflictos y malentendidos en las relaciones interpersonales en línea.

En el contexto digital, el sarcasmo no solo es más difícil de detectar por la ausencia de señales no verbales, sino que también puede aumentar la complejidad en las interacciones textuales, donde el contexto y la interpretación correcta del tono juegan un papel clave. En las plataformas digitales, donde los mensajes suelen ser breves y carecen de un contexto amplio, el sarcasmo puede malinterpretarse con mayor frecuencia, lo que aumenta el riesgo de malentendidos y conflictos entre los usuarios. Este fenómeno se vuelve más problemático en discusiones sobre temas sensibles, donde un comentario sarcástico puede ser tomado literalmente, generando aún más tensiones.

Por otro lado, el uso del sarcasmo en línea puede tener implicaciones mas amplias en la dinámica de grupos sociales, particularmente en redes donde las relaciones se forman y mantienen en gran parte a través de interacciones textuales. La incapacidad de detectar el sarcasmo puede afectar negativamente la percepción de la personalidad o intenciones del emisor, lo que puede llevar a respuestas inapropiadas o a la exclusión social en entornos en línea. Por lo tanto, la detección automática del sarcasmo se convierte en una herramienta crucial para mejorar la convivencia de la comunidad digital y reducir el impacto negativo de los malentendidos. (Sobti et. al, 2024)

2.2.1.2 Problemas causados por la falta de detección de sarcasmos en aplicaciones con moderación de contenido

La falta de una detección precisa del sarcasmo en los sistemas de moderación de contenido puede causar graves consecuencias en la regulación de las plataformas digitales. Uno de los principales problemas podría ser la confusión entre sarcasmo y comentarios ofensivos, lo que podría llevar a sanciones equivocadas o por el contrario, la omisión de estos comentarios dañinos. Los sistemas automáticos de moderación de contenido, al basarse principalmente en análisis léxicos literales, no son capaces de distinguir cuando un comentario esta siendo sarcástico, lo que provoca la eliminación del contenido inofensivo o la perpetuación de mensajes que violan la reglas bajo la apariencia de un mensaje sarcástico.

Asimismo, esta falta de detección puede afectar de forma negativa la calidad de la experiencia del usuario. Esto no únicamente afecta a la experiencia del usuario, sino que también deteriora la confianza en las plataformas, especialmente en aquellas en donde el lenguaje esta limitado ya sea porque sea una plataforma con contenido para menores de edad o porque depende de un equilibrio entre libertad de expresión y el mantenimiento de un entorno seguro y respetuosos.

2.2.1.3 Beneficios de un sistema robusto de detección de sarcasmos automático

La implementación de un sistema robusto de detección automática de sarcasmos tiene importantes beneficios para diversas áreas del procesamiento de lenguaje natural, especialmente en plataformas digitales y aplicaciones con alto volumen de aplicaciones de usuarios. Uno de los principales beneficios seria la mejora del servicio de moderación de contenido, ya que la detección precisa del sarcasmo permite a los sistemas distinguir con mayor exactitud entre comentarios sarcásticos y aquellos que podrían estar violando las normas de la plataforma. Esto, como se ha mencionado antes, podría evitar la eliminación injustificada de comentarios legales y asegurar que los comentarios realmente ofensivos sean eliminados correctamente.

Asimismo, un sistema automatizado eficiente que detecte el sarcasmo podría mejorar el análisis de sentimiento en redes sociales y otras plataformas digitales. Al comprender mejor el tono de los mensajes, las empresas pueden extraer información más precisa sobre las opiniones y emociones de los usuarios, lo que podría ser de gran ayuda para campañas de marketing, análisis de marcas y la evaluación del estado de ánimo general del público.

En aplicaciones automáticas como chatbots o asistentes virtuales, la capacidad de detectar sarcasmos les permite ofrecer respuestas con más contexto y adaptadas al tono del usuario. Esto no solo mejora la experiencia del usuario, sino que también optimiza la interacción entre humanos y sistemas, haciendo que las respuestas de los asistentes virtuales sean más naturales y coherentes con la interacción real del hablante.

Además de todas las mejoras mencionadas, un sistema de avanzado de detección de sarcasmo puede optimizar el desarrollo de algoritmos de aprendizaje automático y modelos de inteligencia artificial, haciendo que los sistemas sean más útiles y eficientes al identificar patrones de lenguaje completos. Esto es particularmente útil en entornos donde el sarcasmo se utiliza como herramienta para desviar intenciones, como en foros políticos, donde los usuarios pueden disfrazar comentarios hirientes con un tono sarcástico. La detección automática de sarcasmos podría ayudar a identificar este tipo de comentarios para su posterior eliminación y así promover un espacio de discusión más saludable y respetuosos.

Otro beneficio de suma importancia se encuentra en el campo de la salud mental y el bienestar digital. Las interacciones sarcásticas en línea, si no son detectadas correctamente, pueden contribuir al ciberacoso o tener un impacto negativo en la autoestima de los usuarios. Los sistemas que pueden identificar sarcasmos y clasificarlos correctamente permiten una mejor moderación en plataforma donde se discuten temas sensibles o redes sociales, protegiendo así al usuario vulnerable de este tipo de comentarios satíricos.

En resumen, los beneficios de un sistema robusto de detección de sarcasmos automático de sarcasmos abarcan desde una mejor moderación de contenido hasta la mejora de la interacción humana en entornos digitales, protegiendo tanto la convivencia como el bienestar de los usuarios.

2.2.2 Estado de la tecnología actual

La detección de sarcasmos ha surgido como un desafío importante dentro del campo del procesamiento de lenguaje natural. El sarcasmo, al ser una forma de lenguaje figurado en donde se el significado literal de las palabras es opuesto a la intención del hablante, presenta dificultades únicas para los sistemas autónomos. A medida que la comunicación en plataformas digitales aumenta, la capacidad de identificar sarcasmos de manera eficiente se vuelve esencial para mejorar la calidad de las interacciones, la moderación de contenido y el análisis de sentimientos.

En los últimos años, los avances en la tecnología han permitido la introducción de modelos basados en Transformers, como BERT o RoBERTa, que han revolucionado el enfoque hacia la clasificación de sarcasmos al mejorar la comprensión lectora, las relaciones entre palabras y obteniendo un contexto más certero. Sin embargo, a pesar de los progresos, aún siguen existiendo retos importantes en este campo.

A lo largo de este apartado, se expondrá el estado actual de la investigación, los conjuntos de datos disponibles y los desafíos que todavía podemos encontrar al estudiar este campo.

2.2.2.1 Estado de la investigación

A medida que va creciendo el procesamiento de lenguaje natural, también lo hacen sus herramientas y sus estudios. A lo largo de los años el análisis de datos ha obtenido una popularidad desmesurada lo que ha hecho que aumente de manera exponencial en número de estudios realizados sobre todos los temas que abarca, incluyendo el estudio sobre la detección automática de sarcasmos en textos.

A medida que avanzan las tecnologías en el procesamiento del lenguaje natural, la detección automática de sarcasmos ha logrado mejoras notables gracias a los modelos basados en Transformers, como BERT y RoBERTa. Estos modelos permiten captar el contexto global de una oración, un aspecto clave para identificar el sarcasmo, ya que este tipo de lenguaje figurado depende del contraste entre el significado literal y la intención real del hablante. El uso de mecanismos de autoatención ha resultado ser esencial para mejorar la precisión, ya que los modelos pueden asignar diferentes niveles de importancia a cada palabra, capturando así las sutilezas que son características del sarcasmo.

En los estudios más recientes, se ha empezado a aplicar la detección multimodal de sarcasmo, lo que implica analizar no solo el texto, sino también otras señales, como imágenes, videos y audio. El estudio “A Survey of Multimodal Sarcasm Detection” prueba que la integración de modalidades adicionales permite que los modelos capten de manera más precisa el sarcasmo en entornos con más estímulos, como las redes sociales, ya que estos comentarios suelen estar acompañados de gifs o stickers. Este enfoque ha demostrado tener un gran impacto para

mejorar la comprensión del sarcasmo en contextos donde el lenguaje textual por si solo no es suficiente para detectar el sarcasmo (Zampieri et. al, 2024)

Por otra parte, el estudio "An Evaluation of State-of-the-Art Large Language Models for Sarcasm Detection" ha aportado una evaluación exhaustiva de modelos avanzados como GPT-3 y BERT, mostrando que estos modelos de lenguaje tienen un desempeño superior en comparación con enfoques más tradicionales. Uno de los hallazgos clave de este estudio es que las arquitecturas más recientes logran comprender mejor las relaciones contextuales a largo plazo, lo que es fundamental para identificar sarcasmos que dependen de un contexto extenso. Esto ha permitido que los sistemas de moderación de contenido y análisis de sentimiento en redes sociales sean más precisos (Zhou, 2023).

Otro avance significativo es el uso de redes generativas adversariales (GANs), que ha sido aplicado a la detección de sarcasmos mediante el aumento de datos, tal como se comenta en el estudio "Feature-Based Augmentation in Sarcasm Detection Using Reverse Generative Adversarial Network". Un **GAN** (Generative Adversarial Network) es un tipo de red neuronal que consta de dos partes: un **generador** y un **discriminador**, que trabajan en conjunto para crear nuevos datos que son muy similares a los datos reales. El generador crea ejemplos falsos, mientras que el discriminador intenta distinguir entre los ejemplos falsos y los reales. A medida que el generador mejora, el discriminador también mejora en su capacidad para detectar las falsificaciones, lo que lleva a una mejora continua en la calidad de los datos generados. En el contexto de la detección de sarcasmos, los **GANs** pueden utilizarse para generar ejemplos de sarcasmo que no están presentes en el conjunto de datos original, permitiendo a los modelos entrenarse con más información y mejorar su precisión. Esta técnica ha demostrado ser eficaz para resolver problemas de escasez de datos, uno de los mayores desafíos en la detección de sarcasmos. Utilizando GANs, los investigadores han sido capaces de generar ejemplos sintéticos que mejoran la calidad del entrenamiento de los modelos, lo que resulta en una mayor precisión, especialmente en entornos con conjuntos de datos limitados o desequilibrados (Suhartono, Handoyo, & Junior, 2023).

Estos estudios han mostrado cómo los enfoques basados en Transformers y las técnicas innovadoras, como las GANs, han avanzado significativamente en la detección automática de sarcasmos, logrando mejoras tanto en la precisión como en la capacidad de generalización de los modelos. La combinación de enfoques multimodales y el aumento de datos ha abierto nuevas vías para abordar los desafíos que el sarcasmo presenta, permitiendo a los investigadores construir sistemas más robustos que no solo comprendan el texto, sino también las señales no verbales que a menudo acompañan al sarcasmo en la comunicación digital.

2.2.2.2 Revisión de datasets disponibles

La detección automática de sarcasmos en textos ha impulsado la creación de diversos conjuntos de datos que se utilizan para entrenar y evaluar modelos de procesamiento de lenguaje natural. Uno de los conjuntos más utilizados es el SARC (Sarcasm corpus). Utilizado unas 22 veces en los últimos cuatro años, SARC es un dataset grande que contiene comentarios sarcásticos y no sarcásticos extraídos de las plataformas como Reddit. Este dataset ha sido fundamental para

permitir que los modelos de PLN puedan diferenciar entre comentarios literales y sarcásticos en un entorno con mucha diversidad como las redes sociales (Pelser & Murrell, 2019).

Otro conjunto de datos muy utilizado es el conocido como MUSTARD (Multimodal Sarcasm Detection). A diferencia de otros dataset, MUSTARD integra información visual y auditiva junto con el texto, lo que permite que los modelos capten mejor el tono y la ironía en la comunicación digital. Este enfoque multimodal ha demostrado mejorar significativamente la precisión en la detección del sarcasmo, especialmente en plataformas como Twitter y Facebook, donde los usuarios utilizan GIFs, fotos y otras señales no verbales para complementar su texto (Ray, Mishra, Nunna, & Bhattacharyya, 2022)

Otro recurso bastante relevante es el dataset de noticias sarcásticas en inglés que podemos encontrar disponible en Kaggle, el cual contiene titulares de noticias etiquetados como sarcásticos o no sarcásticos. Este conjunto de datos ha sido ampliamente utilizado en las investigaciones sobre la detección de sarcasmo gracias a su naturaleza textual, permitiendo a los modelos trabajar con noticias con un grado de humor bastante enrevesado, lo que hace que el modelo aprenda un sarcasmo más complejo.

En el ámbito español, no se ha realizado mucho hincapié en la construcción de dataset ricos en contenido por parte de la comunidad, sin embargo, podemos encontrar alguno como el dataset de diálogos sarcásticos de la serie de televisión South Park. Este conjunto de datos, aunque limitado, ha permitido avanzar en la detección de sarcasmos en español, un área que aun está en sus primeras etapas.

La falta de recursos en español sigue siendo un desafío para la investigación. A día de hoy, no existen datasets ampliamente disponibles en español que aborden de manera integral la detección de sarcasmos. Esto brinda una oportunidad para la creación de nuevos conjuntos de datos, ya que la mayor parte del trabajo en sarcasmo se ha centrado en el inglés. Para superar esta barrera, es necesario desarrollar datasets propios en español. Esto implicaría la recolección de ejemplos de sarcasmos de diferentes fuentes, como redes sociales o programas de televisión, y su etiquetado manual. Sin estos recursos, los modelos de PLN tendrán dificultades para generalizar a otros idiomas, lo que limita la capacidad para abordar la variabilidad cultural en la interpretación del sarcasmo.

Otro recurso que se está empezando a implementar sería el uso de técnicas como la generación de datos sintéticos mediante redes generativas adversariales (GANs). Esta técnica ha comenzado a jugar un papel importante en la ampliación de los conjuntos de datos disponibles. Estos modelos permiten crear ejemplos sin necesidad de recopilar y etiquetar datos manualmente.

2.2.2.3 Retos actuales de la investigación

A pesar de los grandes avances en la detección automática de sarcasmos, todavía podemos encontrar varios desafíos que limitan la precisión y la generalización de los modelos actuales. Uno de los mayores retos es la dependencia del contexto. El sarcasmo a menudo se basa en información implícita, como el tono, el conocimiento compartido o las referencias culturales, que no siempre están presentes en el texto. Aunque los modelos multimodales han intentado abordar este problema, su efectividad sigue siendo limitada en escenarios donde solo se dispone

de texto. Los enfoques actuales deben mejorar la capacidad de capturar estas señales adicionales para lograr una interpretación más precisa del sarcasmo en diferentes contextos (Zampieri, Kong , Ranasinghe, Farabi, & Kanojia, 2024).

Otro desafío importante es la falta de datasets multilingües. La mayoría de los estudios hasta la fecha se han centrado exclusivamente en el inglés, lo que deja un vacío en la investigación de otros idiomas, como el español o el francés, donde el sarcasmo puede expresarse de manera distinta. Esta limitación en la diversidad de los datasets dificulta que los modelos aprendan a identificar sarcasmos en múltiples lenguas y culturas, reduciendo su capacidad para generalizar más allá del inglés. En consecuencia, se requiere una mayor inversión en la creación de datasets etiquetados en diferentes idiomas para mejorar el rendimiento de los modelos en un contexto global.

Otro desafío significativo es el desequilibrio de datos. El sarcasmo, aunque frecuente, no es tan común como el lenguaje literal en muchos corpus de texto. Esto provoca que los modelos tengan una tendencia a sobreajustarse al lenguaje literal, lo que resulta en una alta tasa de falsos negativos en la clasificación de sarcasmos. Las redes generativas adversariales (GANs) han comenzado a ofrecer soluciones prometedoras a este problema, permitiendo generar datos sintéticos que equilibran mejor las clases y mejoran la calidad del entrenamiento del modelo. Sin embargo, la generación de datos sintéticos todavía presenta retos, especialmente en términos de la autenticidad y variabilidad de los ejemplos creados (Suhartono, Handoyo, & Junior, 2023).

La anotación inconsistente de los datos también representa un obstáculo importante. El sarcasmo es subjetivo, y lo que una persona puede etiquetar como sarcástico, otra podría interpretarlo de manera diferente. Esta variabilidad en la anotación introduce ruido en los datasets, lo que afecta directamente la precisión de los modelos. Una solución podría ser el desarrollo de estándares más claros para la anotación de sarcasmos, así como herramientas más sofisticadas que ayuden a identificar y clasificar estos casos de manera más consistente (Zampieri, Kong , Ranasinghe, Farabi, & Kanojia, 2024).

Finalmente, la eficiencia computacional es un reto continuo en la implementación de modelos de sarcasmo. Los modelos basados en Transformers requieren grandes cantidades de memoria y recursos computacionales, lo que los hace difíciles de desplegar en computadoras convencionales o en tiempo real. La optimización de estos modelos sin sacrificar la precisión es un área clave de investigación en el futuro (Suhartono, Handoyo, & Junior, 2023).

2.2.3 Aportaciones del proyecto al campo de estudio

El presente proyecto se sitúa en un punto muy importante de la investigación en detección de sarcasmos, y su relevancia se encuentra en el nuevo enfoque y en la contribución a la expansión de este campo en español. Dado que la mayoría de los estudios se han centrado exclusivamente en inglés, la creación de un nuevo dataset en español, combinado con un análisis exhaustivo y la comparación con modelos avanzados de aprendizaje profundo, establecerá una base sólida para futuras investigaciones. Estas aportaciones están divididas en dos grupos importantes, la novedad del enfoque y la contribución al estado del arte.

2.2.3.1 Novedad del enfoque

Una de las principales innovaciones del proyecto será la creación de un dataset único en español, lo que será crucial para poder avanzar en la investigación de un clasificador de sarcasmos automático. Este conjunto de datos está compuesto por tres fuentes principales: titulares de noticias, diálogos de series con toques humorísticos y frases generadas por inteligencia artificial. Al combinar estos datos reales con datos generados artificialmente, este proyecto ofrece una perspectiva más amplia y diversa del uso del sarcasmo en diferentes contextos. El análisis de este tipo de datos en español es único, dado que no existen recursos similares disponibles en ninguna plataforma pública para este idioma.

Además de un dataset innovador, el proyecto incluye un análisis exploratorio de los datos muy profundo, lo que podría identificar patrones y características lingüísticas del sarcasmo en español. Este análisis no solo acelerará la comprensión del sarcasmo, sino que ayudará a solidificar las bases de futuras investigaciones en el mismo campo.

Asimismo, se utilizará uno de los modelos de aprendizaje profundo más complejos y con mejores resultados, los modelos basados en Transformers usando transfer learning. Esta técnica permite aprovechar el conocimiento pre entrenado en modelos grandes y ajustarlo específicamente a los datos de sarcasmo en español, logrando una mayor precisión y adaptabilidad a este idioma. A diferencia de las aproximaciones más tradicionales basadas en RNN o LSTM, los Transformers con fine-tuning tiene la capacidad de capturar de manera más eficiente el contexto global, lo que es esencial para la identificación del sarcasmo, especialmente en frases donde el contraste entre lo literal y lo intencionado es muy sutil.

2.2.3.2 Contribución al estado del arte

Este proyecto contribuye de manera representativa al estado del arte en la detección automática de sarcasmos, abordando las principales áreas de investigación, desarrollo de datasets, mejora de modelos y comparativas con enfoques previos.

Una de las principales contribuciones al estado del arte es la profunda investigación que se ha realizado previamente para este proyecto, en donde se ha explorado de forma exhaustiva el fenómeno lingüístico del sarcasmo, el procesamiento de lenguaje natural y sus técnicas, desde las más rudimentarias hasta las actuales, y como se han aplicado a la detección de sarcasmos. Además, se ha hecho principal hincapié en los modelos basados en Transformers particularmente en la técnica de fine tuning y como estos pueden superar las limitaciones previas en la detección de sarcasmos. Al hacer esto, se genera un marco de referencia para futuras investigaciones en este campo, lo que contribuirá a mejorar la precisión y aplicabilidad de los modelos de PLN a nivel global.

Otra contribución que tiene suma importancia es la creación de un dataset único en español, que incluye noticias sarcásticas, diálogos de televisión con toques humorísticos y frases generadas por inteligencia artificial. Este dataset cubre una variedad de dominios que no solo proporciona diversidad textual, sino que también ayuda a comprender el sarcasmo desde múltiples puntos de vista. Dado que no existen muchos conjuntos de datos en español en español dedicados a esta tarea, este proyecto llena un vacío importante en la investigación,

permitiendo que futuros modelos aprendan sobre cómo actúa el sarcasmo en un idioma distinto al inglés.

Asimismo, este proyecto contribuye con la creación de un modelo basado en Transformers con fine tuning, más en concreto un modelo RoBERTa para la clasificación de textos, utilizando el innovador conjunto de datos en español. Al ajustar el modelo pre entrenado a los datos específicos de sarcasmo en español, el proyecto asegura que el modelo no solo capture de manera global, sino que también identifique las particularidades culturales y lingüísticas que definen el sarcasmo en este idioma. A diferencia de otros enfoques como RNNs o LSTMs, el Transformers permite manejar contextos largos y complejos, lo cual mejora la capacidad del modelo para identificar estos sarcasmos, incluyendo las frases más sutiles.

Por último, el proyecto ofrece una comparativa exhaustiva entre el modelo Transformers ajustado y otros enfoques más rudimentarios, como podrían ser CNNs, RNNs o LSTMs. Esta comparación permite medir con precisión las ventajas de los modelos basados en Transformers en la detección de sarcasmos en español. Los resultados esperados de esta comparativa mostraran como el uso de modelos avanzados con técnicas de fine tuning supera a los enfoques anteriores, tanto en capacidad de generalización como en eficiencia computacional.

2.3 Planteamiento del problema

El planteamiento del problema en este proyecto es el nexo de unión entre el estado del arte y los objetivos de la investigación. A partir de la revisión del estado del arte, se identificarán varias carencias en el campo de la detección automática de sarcasmos en español, un área que hasta ahora no ha sido especialmente abordada. Los avances significativos en la detección de sarcasmos, como el uso de Transformers o modelos multimodales, se han centrado principalmente en el inglés como idioma, lo que ha dejado un gran vacío en la investigación para otros idiomas, incluidos los contextos lingüísticos y culturales hispanohablantes. Esto incluye la falta de conjunto de datos en español que aborden esta tarea limita el desarrollo de soluciones para este campo.

2.3.1 Formulación del problema

El problema principal que se aborda en este proyecto es la falta de modelos y datasets adecuados para la detección automática de sarcasmo en español, una necesidad cada vez más solicitada, sobre todo en las redes sociales y foros de discusión con moderación de contenido. El sarcasmo, al ser un fenómeno lingüístico complejo es especialmente difícil de detectar, ya que suele tener un significado subyacente detrás de su significado literal además de la necesidad de tener conocimientos previos sobre el contexto al que se refiere, algo que los modelos de procesamiento de lenguaje natural tradicionales escasean. La mayoría de los modelos más contemporáneos se entrenan en inglés, donde el contexto y cultura difieren del español. En un contexto globalizado, donde el español es una de las lenguas más habladas, esta brecha en los recursos lingüísticos y tecnológicos crea una limitación importante en la aplicación de herramienta de PLN para hispanohablantes.

La correcta identificación del sarcasmo no solo requiere que se capte el significado literal de las palabras, sino que también comprendan el significado subyacente del emisor. El gran problema del sarcasmo en español, es que el significado de este puede variar enormemente en función del contexto cultural, el tono de la conversación e incluso el conocimiento individual de cada emisor y/o receptor. Es por ello que hay una imperante necesidad de la creación de un conjunto de datos públicos que abarquen todas estas brechas culturales que tiene la comunidad hispanohablante para que los modelos puedan entender estas sutilezas lingüísticas. Además, las diferencias sintácticas, semánticas y culturales entre el inglés y el español hacen que los modelos entrenados exclusivamente en inglés no puedan ser extrapolables al español.

2.3.2 Relevancia del problema en el contexto actual

La detección automática de sarcasmos en español ha cobrado una importancia creciente en la sociedad digital actual debido al auge de las plataformas de redes sociales, como foros de discusión, y el aumento en la cantidad de información generada por los usuarios. En este tipo de plataformas, el sarcasmo es un tipo de comunicación bastante recurrente, utilizado tanto para el humor como para críticas constructivas o destructivas. Sin embargo, la capacidad de los sistemas automáticos para detectar correctamente el sarcasmo puede conducir a interpretaciones erróneas de los comentarios de los usuarios, afectando la precisión del análisis de sentimientos y la eficacia de los sistemas de moderación de contenido. En un mundo donde las plataformas digitales son esenciales para la comunicación y donde se utilizan sistemas automatizados para gestionar grandes volúmenes de datos, la detección precisa del sarcasmo es crítica para evitar malentendidos y sanciones equivocadas.

En el contexto hispanohablante, este problema es aún más pronunciado debido a la falta de herramientas de PLN específicas para el español. Las redes sociales y las aplicaciones digitales que trabajan en español no cuentan con modelos robustos que sean capaces de captar las sutilezas del sarcasmo en este idioma, lo que genera un vacío tecnológico que afecta a millones de usuarios. Además, la capacidad de los sistemas actuales para identificar sarcasmo está limitada principalmente al inglés, lo que excluye una parte significativa de la población mundial de los beneficios de estas tecnologías. Esta falta de recursos afecta no solo a los usuarios individuales, sino también a las plataformas que dependen del análisis automatizado de opiniones para la toma de decisiones, la gestión de la reputación en línea y la moderación de contenido. En este sentido, la creación de modelos especializados para la detección de sarcasmos en español es un problema relevante en el contexto actual.

2.3.3 Hipótesis inicial

La hipótesis inicial del presente proyecto sostiene que la implementación de un Transformer ajustado mediante fine-tuning sobre un conjunto de datos en español, que incluye noticias, diálogos de televisión y frases generadas por inteligencia artificial, mejorará significativamente la precisión en la detección automática de sarcasmos en comparación con los modelos tradicionales basados en RNN o LSTM. Debido a la capacidad de los modelos basados en Transformers para captar el contexto global y asignar atención diferencial a las palabras clave de una oración, el sarcasmo pueda ser identificado de manera más efectiva en español.

Además, se postula que el análisis exploratorio y la construcción de un conjunto de datos en español, que contemple diversos estilos de lenguaje, lo que aportará nuevos conocimientos sobre las características lingüísticas y culturales que definen el sarcasmo en este idioma. Esto permitirá a los modelos ajustarse no solo a la estructura del lenguaje español, sino también a sus particularidades culturales, mejorando así la detección del sarcasmo en escenarios reales. La hipótesis también sugiere que este enfoque especializado en español puede ser extendido a otros idiomas, lo que ayudaría a solucionar la falta de herramientas multilingües robustas en este ámbito.

Capítulo 3. OBJETIVOS







Este capítulo del proyecto tiene como objetivo presentar de manera clara tanto el propósito general del trabajo como los objetivos específicos que complementan el objetivo general. Se establecerá una base para lo que se espera alcanzar con la investigación, conectando los elementos clave identificados en el planteamiento del problema con las soluciones propuestas.

Este apartado también busca detectar los beneficios tangibles que el proyecto aportara al campo de la detección automática de sarcasmos en español y su impacto en el procesamiento de lenguaje natural.

3.1 Objetivos generales

El objetivo principal de este proyecto es desarrollar un sistema de detección de sarcasmos en español utilizando un modelo Transformer ajustado mediante fine-tuning. El modelo será entrenado sobre un conjunto de datos propio, que incluye titulares de noticias, diálogos de series de televisión y frases generadas por inteligencia artificial en español. Este enfoque con fine-tuning permite adaptar un modelo preentrenado a las particularidades del sarcasmo en español, optimizando su rendimiento para captar mejor las sutilezas del lenguaje y el contexto cultural propio de este idioma.

3.2 Objetivos específicos

-  **Investigación:** Realizar un análisis exhaustivo del estado del arte en la detección de sarcasmos, enfocándose en los modelos basados en Transformers y técnicas de fine-tuning, identificando los principales retos y oportunidades en el contexto del idioma español.
-  **Recopilación de datos:** Obtener y organizar un conjunto de datos único en español, compuesto por titulares de noticias, diálogos de series de televisión y frases generadas por inteligencia artificial, con un enfoque especial en la detección de sarcasmos en diferentes contextos lingüísticos y culturales.
-  **ETL (Extracción, Transformación y Carga):** Procesar y limpiar los datos recopilados, asegurando que estén en el formato adecuado para ser utilizados en el modelo, eliminando posibles errores, redundancias y preparando las características necesarias.
-  **Análisis Exploratorio de Datos (EDA):** Realizar un análisis estadístico y visual de los datos para identificar patrones, outliers, y características clave que definan el sarcasmo en el dataset, aportando información útil para el desarrollo del modelo.
-  **Creación del modelo:** Desarrollar varios modelos de clasificación para poder enfrentar diferentes enfoques. Además de realizar un modelo basado en Transformers ajustado mediante fine-tuning, adaptándolo a las características y peculiaridades del sarcasmo en español.
-  **Entrenamiento de los modelos:** Entrenar los modelos con el dataset propio, optimizando los hiperparámetros del mejor modelo y asegurando que el modelo logre capturar las sutilezas del sarcasmo en las diferentes fuentes de datos.

- ✚ **Análisis comparativo de resultados:** Comparar el rendimiento del modelo basado en Transformers con otros enfoques tradicionales, como RNN o SVM, para evaluar las mejoras logradas en la detección de sarcasmos en español.
- ✚ **Optimización del modelo:** Ajustar el modelo en función de los resultados obtenidos durante la validación, buscando maximizar su rendimiento y eficiencia.
- ✚ **Validación y pruebas:** Realizar varias pruebas para validar el modelo y su rendimiento.
- ✚ **Documentación:** Elaborar una documentación exhaustiva del proceso, describiendo cada paso realizado, desde la recopilación de datos hasta la evaluación del modelo, incluyendo el análisis de los resultados obtenidos y las mejoras propuestas.

3.3 Beneficios del proyecto

El proyecto aportará beneficios tanto al campo del procesamiento del lenguaje natural (PLN) en español como a diferentes tipos de aplicaciones relacionadas con la detección automática de sarcasmos. En primer lugar, la creación de un dataset único en español, compuesto por titulares de noticias, diálogos de televisión y frases generadas por IA, llenará un vacío importante en los recursos actuales disponibles para el entrenamiento de modelos en este idioma. Este recurso no solo beneficiará a futuras investigaciones, sino que también proporcionará un estándar para el análisis de sarcasmos en español, un área poco explorada hasta el momento.

Además, el desarrollo y entrenamiento de un modelo Transformer con fine-tuning para la detección de sarcasmos permitirá una mayor precisión y eficiencia en la identificación de sarcasmos en español, mejorando aplicaciones de análisis de sentimientos y moderación de contenido en plataformas digitales. Esto contribuirá a una interacción más precisa entre las máquinas y los usuarios hispanohablantes, evitando malentendidos y mejorando la calidad de los servicios automatizados, como los sistemas de atención al cliente y los motores de búsqueda.

Finalmente, el proyecto beneficiará la comunidad científica al proporcionar análisis comparativos entre modelos avanzados y enfoques más rudimentarios, lo que servirá como punto de referencia para futuras investigaciones en la detección de sarcasmos y otras tareas de PLN en español y otros idiomas menos explorados.

Capítulo 4. DESARROLLO DEL PROYECTO

En este capítulo se detalla el proceso completo de desarrollo del proyecto, desde la planificación inicial hasta la obtención de resultados finales. Se comienza describiendo la planificación del proyecto, donde se definen los objetivos, las fases y los plazos clave para su ejecución. A continuación, se presenta la metodología utilizada, junto con las herramientas empleadas, asegurando que cada fase del proyecto se desarrolla de manera eficiente y siguiendo las mejores prácticas.

Además, se incluye un análisis detallado del presupuesto, especificando los recursos necesarios y su distribución a lo largo del proyecto. Este apartado está acompañado por una evaluación de la viabilidad técnica y económica, justificando la factibilidad del proyecto en términos de tiempo, costos y recursos disponibles.

Finalmente, se exponen los resultados obtenidos, presentando tanto el rendimiento de las herramientas implementadas como el cumplimiento de los objetivos planteados al inicio. Este análisis de resultados ofrece una visión clara del impacto y éxito del proyecto en su conjunto.

4.1 Planificación del proyecto

La planificación del proyecto es crucial para establecer una organización de las actividades realizadas, los hitos alcanzados y las reuniones que tuvieron lugar durante el desarrollo del trabajo. El plan ha sido diseñado para cumplir con los objetivos generales y específicos del proyecto, desde la fase inicial de investigación hasta la documentación final. A continuación, se presentará un diagrama de Gantt junto a una descripción de las actividades, el esfuerzo asociado, los hitos alcanzados y las reuniones programadas.

Fecha:	06-may	13-may	20-may	27-may	03-jun	10-jun	17-jun	24-jun
Actividad 1	Planificación proyecto							
Actividad 2			Entendimiento del problema					
Actividad 3								
Actividad 4								
Actividad 5								
Actividad 6								
Actividad 7								
Actividad 8								
Actividad 9								
Reuniones:	1ª reunión		2ª reunión			3ª reunión		
Hitos:		Anteproyecto				Introducción y estructura del proyecto		

Fecha:	01-jul	08-jul	15-jul	22-jul	29-jul	05-ago	12-ago	19-ago	26-ago	02-sep	09-sep	16-sep	23-sep	30-sep	07-oct
Actividad 1															
Actividad 2															
Actividad 3	Recopilación de datos														
Actividad 4		Análisis exploratorio													
Actividad 5			ETL												
Actividad 6							Desarrollo modelo								
Actividad 7											Evaluación y validación				
Actividad 8														Prueba final	
Actividad 9	Realizar informe														
Reuniones:	4ª reunión		5ª reunión					6ª reunión			7ª reunión		8ª reunión	9ª reunión	
Hitos:			Dataset			Datos preparados					Modelo	Modelo Validado	Modelo testeado	Documento	

4.1.1 Actividades del proyecto

1. **Planificación del proyecto (06-mayo - 13-mayo):**
 - En esta fase inicial se definieron los objetivos, se estableció el cronograma de trabajo y se coordinaron las reuniones clave con los colaboradores. La planificación involucró la creación de un esquema detallado del proyecto, incluyendo la identificación de recursos necesarios y la distribución de responsabilidades.
 - Esfuerzo: 10 horas.
 - Entregable: Anteproyecto
2. **Entendimiento del problema (13-mayo - 10-junio):**
 - Durante este período, se llevó a cabo un análisis profundo del problema de la detección automática de sarcasmos en español. Esto incluyó la revisión del estado del arte, la identificación de las limitaciones actuales en el PLN para este idioma, el contexto y justificación además de la formulación del problema del proyecto.
 - Esfuerzo: 20 horas.
 - Entregable: Estado del arte
3. **Recopilación de datos (01-julio - 15-julio)**
 - Se realizó la búsqueda, selección y recopilación de datos para el dataset en español. Esto incluyó noticias sarcásticas, diálogos de series de televisión y frases generadas por IA.
 - Esfuerzo: 40 horas.
 - Entregable: Conjunto de datos sin procesar y archivo .ipynb con el código de la extracción de los datos
4. **Análisis exploratorio de datos (EDA) (08-julio - 29-julio)**
 - Se realizó un análisis exploratorio del dataset para identificar patrones, anomalías y características clave relacionadas con el sarcasmo en español. Esto incluyó la creación de visualizaciones de datos y el cálculo de métricas descriptivas que ayudaron a comprender mejor la naturaleza de los datos recopilados.
 - Esfuerzo: 20 horas.
 - Entregable: Archivo. ipynb y documentación del EDA
5. **ETL (Extracción, Transformación y Carga) (15-julio - 12-agosto)**
 - En esta etapa, se procesaron y limpiaron los datos, asegurando que estuvieran listos para su uso en el modelo. Se transformaron las características necesarias y se eliminaron errores, valores en blanco y duplicados en el dataset. El proceso garantizó que los datos fueran adecuados para el entrenamiento del modelo.
 - Esfuerzo: 5 horas.
 - Entregable: Conjunto de datos procesado y .ipynb con el código usado para el preprocesado de datos
6. **Desarrollo del modelo (12-agosto - 09-septiembre)**

- En esta fase, se construyó un modelo basado en Transformers, más en concreto el modelo llamado RoBERTa, ajustado mediante fine-tuning. El modelo fue diseñado específicamente para detectar sarcasmos en el idioma español, utilizando el dataset propio generado durante el proyecto.
 - Esfuerzo: 30 horas.
 - Entregable: Modelo Roberta y .ipynb con el código usado para el modelo.
- 7. Evaluación y validación del modelo (09-septiembre - 23-septiembre)**
- El modelo fue evaluado y validado a través de pruebas exhaustivas, utilizando métricas de rendimiento como precisión, recall y F1-score. También se comparó su rendimiento con modelos más básicos como RNN y LSTM para determinar su eficacia en la detección de sarcasmos en español.
 - Esfuerzo: 15 horas.
 - Entregable: Modelo validado y .ipynb con el código usado para la validación.
- 8. Prueba final y optimización (23-septiembre - 30-septiembre)**
- Se realizaron pruebas finales para optimizar el modelo en función de los resultados obtenidos durante la evaluación. Esta fase incluyó ajustes de hiperparámetros y mejoras en la eficiencia del modelo para garantizar su rendimiento óptimo en aplicaciones prácticas.
 - Esfuerzo: 5 horas.
 - Entregable: .ipynb con el código usado para la prueba y optimización.
- 9. Documentación (30-septiembre - 07-octubre)**
- Finalmente, se elaboró un informe detallado del proyecto, documentando cada fase del desarrollo, desde la recopilación de datos hasta la validación del modelo. Se incluyeron los resultados obtenidos, las conclusiones alcanzadas y las recomendaciones para futuras investigaciones.
 - Esfuerzo: 20 horas.
 - Entregable: Documentación sobre el proyecto

4.2 Descripción de la solución, metodologías y herramientas empleadas

En este apartado se describirán las metodologías y herramientas tecnológicas empleadas a lo largo del desarrollo de este proyecto, cuyo objetivo principal es la creación de un sistema de detección de sarcasmos en español utilizando un modelo Transformer ajustado mediante fine-tuning. A lo largo del proyecto, se siguieron varias fases clave, que abarcan desde la obtención y preprocesamiento de datos, hasta la implementación y evaluación de los modelos. La metodología se basa en técnicas avanzadas de procesamiento de lenguaje natural (NLP) y aprendizaje profundo, adaptando un modelo preentrenado al conjunto de datos específico de sarcasmos.

4.2.1 Investigación

Para desarrollar un sistema de detección de sarcasmos efectivo, en primer lugar, fue necesario realizar una investigación exhaustiva en diversas fuentes, abarcando desde artículos periodísticos hasta cursos formativos avanzados. Este proceso no solo proporcionó una base

teórica sólida, sino que también facilitó la comprensión de los avances más recientes en el procesamiento de lenguaje natural y Transformers.

4.2.1.1 Fuentes periodísticas y académicas

En primer lugar, se consultaron artículos periodísticos y fuentes verificadas como la Real Academia Española (RAE) para obtener una visión clara de lo que es el fenómeno lingüístico conocido como sarcasmo en español. Estas fuentes proporcionaron una información valiosa sobre la definición de sarcasmo, la diferencia entre sarcasmo e ironía y los diferentes tipos de sarcasmo que pueden existir. Por ejemplo, se analizaron estudios que describen cómo el sarcasmo puede ser detectado por el uso de ciertos términos específicos, la entonación del lenguaje oral o la estructura de las frases en el lenguaje escrito.

También se exploraron plataformas con contenido de alta calidad, como Medium, donde expertos en procesamiento de lenguaje natural (PLN) publican regularmente sobre nuevas tendencias en la tecnología de redes neuronales y modelos Transformers. Para la búsqueda de artículos académicos especializados, se utilizó Google Scholar, que permitió acceder a estudios sobre el sarcasmo y su tratamiento en el análisis del lenguaje escrito y oral.

4.2.1.2 Cursos formativos

Además, se llevaron a cabo varias formaciones especializadas a través de diferentes plataformas en línea. El más relevante fue el curso de Transformers de Hugging Face, que proporciona una comprensión profunda en el uso y ajuste de modelos preentrenados como BERT y GPT. Hugging Face es una referencia clave en el ámbito del PLN, y sus cursos permiten adquirir habilidades avanzadas para la modificación de modelos a tareas específicas como la detección de sarcasmos. Las herramientas prácticas de Hugging Face fueron esenciales para implementar y ajustar modelos en Python.

Otro curso fundamental fue proporcionado por Spain AI, titulado “NLP de 0 a 100 con Hugging Face”, que ofrece una visión integral del PLN desde los conceptos más básicos hasta la implementación avanzada con Transformers. Este curso, junto con los recursos de Hugging Face, fue clave para desarrollar el modelo Transformer ajustado que se utiliza en este proyecto.

Asimismo, el curso de Udemy llamado “Deep Learning de A a Z: Redes neuronales en Python desde cero” impartido por Juan Gabriel Gomilas Salas fue de gran ayuda para entender el concepto detrás de las redes neuronales y el Deep Learning, algo imprescindible para entender el funcionamiento de los Transformers. Este curso proporcionó una introducción a herramientas como PyTorch, utilizadas posteriormente para la implementación de redes neuronales y modelos Transformer.

4.2.1.3 Investigación en la práctica

Esta combinación de fuentes teóricas y prácticas permitió contextualizar el problema del sarcasmo en español, tanto desde el punto de vista lingüístico como técnico. Al combinar el conocimiento obtenido de fuentes académicas, como las consultadas en Google Scholar, con las habilidades adquiridas en los cursos formativos de Udemy, Spain AI, y Hugging Face, fue posible

definir un enfoque metodológico claro y aplicar las mejores prácticas en el desarrollo de un modelo Transformer para detectar de manera óptima las frases sarcásticas en español.

Las herramientas tecnológicas empleadas durante esta fase inicial de investigación, como Google Scholar, Medium, Udemy, YouTube, y la documentación oficial de Hugging Face, resultaron esenciales para la correcta comprensión de las técnicas más recientes en el campo del procesamiento de lenguaje natural y redes neuronales.

4.2.2 Obtención de datos

Para el desarrollo de este proyecto, fue crucial la obtención de datos tanto sarcásticos como no sarcásticos en español provenientes de diferentes fuentes. El enfoque se centro en recopilar datos principalmente de tres tipos diferentes de fuentes: noticias sarcásticas y no sarcásticas, diálogos de series de televisión y frases generadas por inteligencia artificial.

4.2.2.1 Noticias sarcásticas

En primer lugar, se obtuvo una gran cantidad de datos del periódico satírico El Mundo Today, un periódico muy conocido en España que se dedica a crear noticias humorísticas. Utilizando técnicas de Web Scraping implementadas con las bibliotecas de Python, BeautifulSoup y Requests. Estas herramientas permitieron automatizar la extracción de titulares directamente del sitio web del periódico, lo que facilitó la recopilación de un gran volumen de datos representativos del sarcasmo en el contexto de noticias y titulares en español.



Ilustración 16: Ejemplo de noticia sarcástica de El Mundo Today.

Fuente: El Mundo Today

4.2.2.2 Noticias no sarcásticas

Para equilibrar los datos sarcásticos, también se necesitaban titulares no sarcásticos. Se utilizó un proceso similar de Web Scraping automatizado en periódicos populares como El País y El Diario Sur. Se creó una lista de palabras clave relevantes, y mediante un script de Python

desarrollado con BeautifulSoup y Requests, se realizaron búsquedas automatizadas en los sitios web de estos periódicos.



Ilustración 17: Ejemplo de noticia no sarcástica de El Sur

Fuente: El Diario Sur

4.2.2.3 Diálogos de series de televisión

Observando que los sarcasmos en las noticias eran sutiles y que el modelo podría sobre ajustarse si solo se entrenaba con estos datos, se decidió ampliar la diversidad del conjunto de datos con diálogos sarcásticos extraídos de series de televisión humorísticas. Para ello, se desarrolló un proceso de descarga automática de clips de series de televisión desde listas de YouTube, utilizando un script de Python con la ayuda de bibliotecas como Pytube para la descarga de videos. Series como Dos Hombres y Medio o Rick y Morty, que contienen altos niveles de sarcasmo en sus diálogos, fueron seleccionadas como fuentes.

Posteriormente, se extrajo el audio de estos clips utilizando ffmpeg para la conversión de video a audio, y con la ayuda de herramientas de transcripción automática, como AssemblyAI, se obtuvo la transcripción de los diálogos. Estas herramientas fueron especialmente útiles ya que incluyen la capacidad de detectar interlocutores en los diálogos, facilitando la identificación de quién habla en cada momento. Cada diálogo fue revisado manualmente y etiquetado para determinar si contenía sarcasmo o no, asegurando que se incorporaran ejemplos relevantes y diversos al conjunto de datos.

4.2.2.4 Sarcasmo creado con inteligencia artificial

Al comprobar que los diálogos de las series de televisión no eran suficientes y que existía un desequilibrio debido a que no todos los diálogos eran sarcásticos, se recurrió a la inteligencia artificial para generar más frases sarcásticas en español. Se utilizó ChatGPT, un sistema basado en modelos Transformer, para generar frases sarcásticas en diversos contextos, como el ámbito laboral, las relaciones amorosas y la vida cotidiana. Este enfoque permitió crear un gran volumen de datos adicionales que representaban el sarcasmo en situaciones más comunes y cotidianas.

Las frases generadas fueron revisadas manualmente y seleccionadas para asegurar que cumplieran con el objetivo de representar el sarcasmo de manera efectiva.

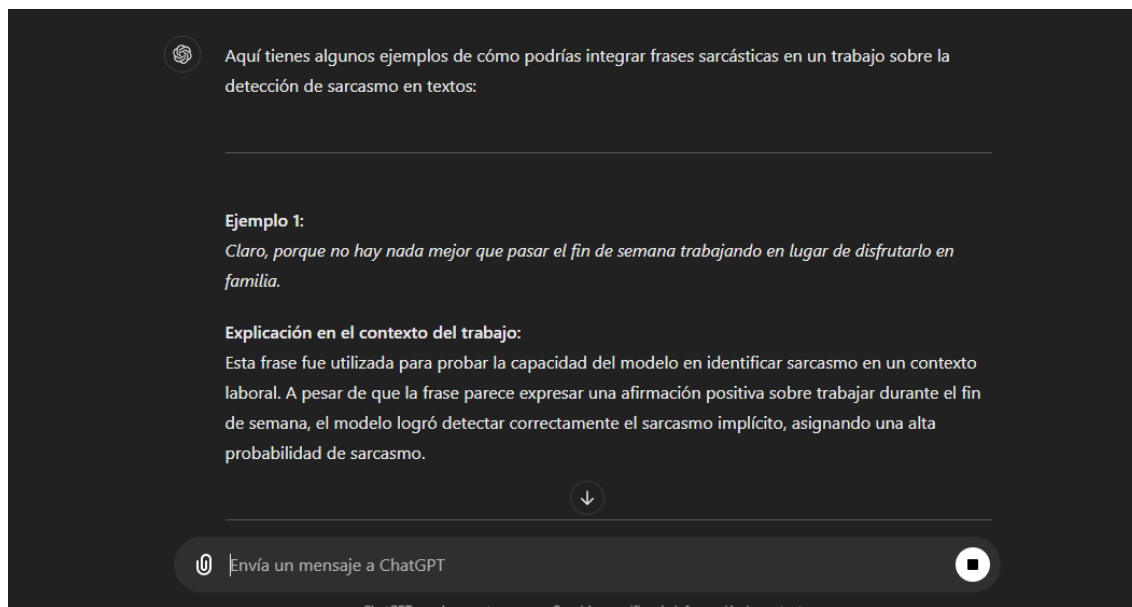


Ilustración 18: Ejemplo de sarcasmo creado con Inteligencia Artificial

Fuente: ChatGPT

4.2.3 ETL (Extracción, transformación y carga)

En esta etapa del proyecto se realizó una extracción, transformación y carga del conjunto de datos obtenido en el apartado anterior. Para llevarlo a cabo, se siguió una serie de pasos para asegurar que el conjunto de datos estuviese limpio y listo para el análisis posterior. Este proceso implicó la recopilación de datos de diversas fuentes, la transformación de los mismos y su posterior almacenamiento en un formato adecuado para ser utilizados en el análisis exploratorio de los datos y en la creación del modelo de machine learning.

4.2.3.1 Extracción de Datos

La extracción de datos fue el primer paso. Los datos estaban almacenados localmente en el equipo, provenientes de las diferentes fuentes descritas anteriormente. Para facilitar la extracción de estos datos, se utilizó un script de Python. Con las bibliotecas estándar como Pandas, se accedió a los archivos en formatos CSV y TXT, extrayéndolos del equipo para proceder con su procesamiento.

4.2.3.2 Transformación de Datos

Una vez extraídos, se procedió con la fase de transformación de los datos. A pesar de que la mayoría de los datos ya estaban etiquetados correctamente, fue necesario realizar ajustes para asegurar la coherencia y calidad del conjunto de datos. Para ello, se empleó Pandas, una herramienta fundamental que facilitó la combinación de todos los conjuntos de datos en un único dataframe. Los titulares de noticias, tanto sarcásticos como no sarcásticos, se unieron en

un solo dataset, dado que ya estaban etiquetados previamente. De la misma manera, los diálogos extraídos de las series de televisión, que habían sido etiquetados manualmente, también se integraron en el dataset final.

Las frases generadas por inteligencia artificial, que venían en archivos separados (sarcasmo por un lado y no sarcasmo por otro), fueron etiquetadas automáticamente en el script de Python y se añadieron al conjunto de datos global. Este paso aseguró que todos los datos estuvieran correctamente clasificados antes de su uso en las fases posteriores del análisis y modelado.

Posteriormente, se realizó una limpieza exhaustiva del dataset. Se eliminaron duplicados y registros con valores nulos, utilizando nuevamente Pandas para simplificar el proceso. Asimismo, se normalizaron los valores de la columna "Sarcasmo", asegurando que contenían únicamente valores enteros (0 y 1). Este paso garantizó la consistencia del conjunto de datos y facilitó su manejo durante el análisis exploratorio y el entrenamiento del modelo.

Como parte del proceso de limpieza, también se decidió eliminar los registros que contenían la palabra "Horóscopo", ya que esta categoría pertenecía a una sección del periódico "El Mundo Today" y se consideró irrelevante para el análisis del sarcasmo. Se realizó un filtrado de estos textos, eliminándolos del conjunto de datos, dado que no aportaban valor al objetivo del proyecto.

4.2.3.3 Carga de Datos

Finalmente, tras la transformación, los datos fueron almacenados nuevamente en el equipo en un formato adecuado para su posterior utilización. Se optó por guardar el dataframe final en formato CSV, lo que permitió trabajar de manera eficiente con los datos durante las siguientes fases del proyecto, como el análisis exploratorio de datos y el modelado.

Este proceso de ETL aseguró que los datos estuvieran limpios, consistentes y en un formato adecuado para su uso en el proyecto, permitiendo un análisis y modelado eficiente en las etapas posteriores.

4.2.4 Análisis exploratorio de los datos

El análisis exploratorio de los datos (EDA) se realizó para comprender mejor la estructura del conjunto de datos y detectar posibles patrones relevantes antes del entrenamiento del modelo. El análisis se centró en explorar las características clave del texto, como la longitud, las palabras más comunes, los bigramas y trigramas, el análisis de sentimientos y las correlaciones entre diferentes variables.

4.2.4.1 Análisis básico de los datos

El análisis exploratorio comenzó con una inspección básica del conjunto de datos, utilizando Pandas para revisar su estructura y asegurar que los datos estuvieran completos y organizados de manera adecuada. Durante esta etapa, se examinó la cantidad de registros y columnas, verificando la ausencia de valores nulos o inconsistencias en los datos. Además, se revisó la distribución de las etiquetas de sarcasmo, lo que permitió evaluar la proporción de textos sarcásticos frente a los no sarcásticos. Este análisis inicial fue clave para detectar posibles

desbalances en el conjunto de datos, aspecto fundamental para ajustar los modelos de manera eficiente en las siguientes fases del proyecto. Se generaron gráficos con Matplotlib y Seaborn para visualizar la estructura básica de los datos y verificar la limpieza del conjunto.

4.2.4.2 Análisis de longitud del texto y detección de valores atípicos

Una vez realizado el análisis básico de los datos, se procedió a estudiar la longitud de los textos para identificar posibles patrones relacionados con el sarcasmo y valores atípicos que pudieran influenciar en el análisis posterior. Este análisis permitió examinar la variabilidad de las frases en función de su longitud, observando si existía una correlación entre longitud y el sarcasmo.

Se calcularon las estadísticas descriptivas de la longitud de los textos, como la media, mediana, desviación estándar y percentiles. Esto permitió obtener una versión clara de como varían las longitudes y si existen textos que por su longitud podrían considerarse un valor típico. Además, se generaron diagramas de cajas y bigotes y grafico de barras para visualizar estos valores y comparar entre los textos sarcásticos y no sarcásticos.

A raíz de este estudio el análisis explicativo de los datos definirá tres caminos a seguir, un estudio con todos los valores, con únicamente los valores atípicos y sin los valores atípicos.

4.2.4.3 Análisis de palabras

Para la siguiente fase se optó por realizar un análisis de palabras en donde se trata de entender que palabras aparecen con mayor frecuencia en el conjunto de datos. Este análisis no solo permite identificar patrones en el uso del lenguaje, sino que también a detectar diferencias entre los textos sarcásticos y los no sarcásticos. Para llevar a cabo este análisis de manera efectiva hay que implementar las siguientes etapas:

- Limpieza del texto: se eliminaron las stopwords mediante la función de procesamiento de texto de NLTK para centrarse únicamente en palabras que aportan información significativa al texto.
- Cálculo de la frecuencia de las palabras: se identificó cuales fueron las palabras más repetidas en los textos tanto sarcásticos como no sarcásticos.

Para realizar un estudio más completo y ver la relevancia de los valores atípicos, como se explicó en el apartado anterior, se han realizado tres estudios diferentes: uno con el conjunto de datos al completo, otro sin los valores atípicos y otro con únicamente los valores atípicos.

Para poder visualizar la frecuencia las palabras y realizar una comparativa entre los diferentes estudios se generaron graficas de nubes de palabras y de barras generados con WordCloud y Matplotlib, tanto para los textos sarcásticos como para los no sarcásticos.

Por último, se realizó un análisis TF-IDF (Term Frequency - Inverse Document Frequency). Este método permite identificar tanto las palabras que aparecen con mayor frecuencia y que tan importantes son esas palabras para el conjunto de datos. Este método considera la frecuencia de una palabra en un documento en relación con su aparición en todo el conjunto de datos, lo que ayuda a reducir la relevancia de palabras comunes y destacar aquellas que son más distintivas en los textos.

4.2.4.4 *Análisis de bigramas y trigramas*

En la siguiente fase del análisis exploratorio de datos, se realizó un análisis de bigramas (combinaciones de dos palabras consecutivas) y trigramas (combinaciones de tres palabras consecutivas). Este tipo de análisis permite identificar secuencias frecuentes de palabras que pueden ser particularmente útiles para distinguir entre textos sarcásticos y no sarcásticos. Se emplearon las funciones de CountVectorizer de Scikit-learn para extraer estas secuencias.

Para profundizar en el análisis, se optó por realizar dos enfoques diferentes. En el primer análisis, no se eliminaron las palabras irrelevantes o stopwords, permitiendo así observar secuencias comunes completas que incluyen artículos, preposiciones, y otras palabras de uso frecuente. En el segundo análisis, se eliminaron estas palabras, centrándose exclusivamente en las palabras más relevantes. Ambos enfoques se compararon siguiendo la misma metodología que en los apartados anteriores, generando seis gráficos de barras para cada uno utilizando Seaborn, donde se comparaban los textos sarcásticos y no sarcásticos, teniendo en cuenta la presencia de valores atípicos.

4.2.4.5 *Análisis de sentimiento*

En este apartado del análisis exploratorio se llevó a cabo un análisis de sentimiento utilizando un modelo BERT multilingüe de Hugging Face, especializado en el análisis de sentimiento en español. Este modelo permite evaluar los sentimientos de los textos, asignando una puntuación en una escala del 1 al 5, donde 1 representa un sentimiento negativo y 5 un sentimiento positivo. EL objetivo del análisis es observar las emociones predominantes en los textos y como estas variaban entre los textos sarcásticos y no sarcásticos.

Para realizar el análisis, primero se cargó el modelo a través de la librería Transformers y se aplicó a cada uno de los textos del conjunto de datos. Esto permitió generar una nueva columna en el dataframe que contenía la puntuación de sentimiento asignada a cada texto. También se realizó una comparativa con un gráfico de barras entre el conjunto de datos completo y el conjunto de datos sin valores atípicos para asegurar que los resultados fueran consistentes y no estuvieran influenciados por la longitud de los textos inusuales.

Una vez generado el análisis de sentimiento, se organizó la información en distintos grupos para realizar comparaciones entre los textos sarcásticos y no sarcásticos. Este proceso permitió observar como varia el sentimiento en función dl sarcasmo y la longitud presente en los textos.

4.2.4.6 *Análisis de correlación*

Para la siguiente esta se ha llevado a cabo un análisis de correlación, en el cual se seleccionaron las tras variables clave: la longitud, el sentimiento y si el texto es sarcástico o no. Este análisis busca identificar relaciones lineales entre estas variables. En primer lugar, se decidió realizar el cálculo de la correlación utilizando el coeficiente de Pearsons, porque mide la relación lineal entre dos variables numéricas.

A continuación, se calculó la matriz de correlación entre las tres variables y se volvió a hacer dos análisis distintos: uno con el conjunto de datos completo y el otro sin los valores atípicos. La matriz resultante de nos dio valores entre -1 y 1, donde 1 indica una correlación positiva perfeta,

-1 una correlación negativa perfecta y 0 indica ausencia de correlación. Para facilitar la interpretación de los resultados, se visualizó esta matriz mediante un heatmap, una herramienta grafica que permite ver de forma intuitiva la fuerza de las correlaciones entre diferentes variables.

4.2.4.7 *Análisis de variabilidad léxica*

Finalmente se realizó un análisis de variabilidad léxica el cual consiste en determinar la diversidad del vocabulario dentro del conjunto de datos. Para ello se calcula el índice TTR ((Type-Token Ratio), el cual consiste en dividir el número de palabras únicas entre el número total de palabras en cada texto.

Se implementaron graficas de cajas y bigotes para poder visualizar la distribución del TTR en cada grupo de datos, textos sarcásticos y no sarcásticos, permitiendo observar posibles diferencias entre ellos.

4.2.5 Implementación del modelo

En este apartado vamos a hablar de cómo se llevó a cabo la implementación del modelo de detección de sarcasmo. Se explicará cómo se empezaron probando modelos más básicos, como los basados en reglas, RNN y LSTM, para luego pasar al modelo Transformer, que resultó ser el más efectivo para este proyecto. También se comentará cómo se ajustaron los hiperparámetros del modelo para mejorar su rendimiento y qué decisiones técnicas se tomaron durante el proceso de desarrollo. En resumen, se dará una visión general del proceso de implementación, resaltando lo más importante de cada fase.

4.2.5.1 *Elección de los primeros modelos*

En primer lugar, antes de implementar nuestro modelo basado en Transformers, se optó por probar algunos enfoques más tradicionales con el objetivo de tener una base de comparación y demostrar de manera práctica lo que se había observado en el estado del arte: que los modelos basados en Transformers tenían una superioridad ante los enfoques más tradicionales en el procesamiento de lenguaje natural y en la detección de sarcasmos.

Para empezar, se implementó el modelo basado en reglas, que fue configurado para identificar palabras o frases comunes que suelen estar asociadas al sarcasmo, como en nuestro caso, "claro", "seguro", o "obviamente". Este modelo no requirió un proceso de entrenamiento tradicional, ya que funciona a partir de reglas predefinidas. Sin embargo, para su correcta implementación, se utilizó Pandas para realizar un análisis del conjunto de datos, identificando términos que pudieran asociarse con el sarcasmo, y dichas reglas fueron codificadas manualmente en Python para realizar las predicciones. Este modelo se utilizó como referencia básica, sin necesidad de ajustar parámetros adicionales.

El segundo modelo implementado fue una Red Neuronal Recurrente (RNN). En este caso, el conjunto de datos fue preprocesado mediante la tokenización de los textos, utilizando Tokenizer de Keras para convertir las frases en secuencias numéricas que pudieran ser procesadas por el modelo. Además, se utilizó un embedding preentrenado en español de Wikipedia creado por

FastText (utilizando Gensim para cargar los embeddings), lo que permitió representar cada palabra en un espacio vectorial de alta dimensionalidad. Una vez configurada la capa de embedding con Keras, se añadió una capa recurrente que procesaba las secuencias de palabras y capturaba la información relevante en cada paso. El modelo fue entrenado con TensorFlow durante varias épocas con hiperparámetros básicos que no fueron optimizados exhaustivamente, ya que el objetivo era tener una referencia práctica del comportamiento de una RNN en esta tarea.

Por último, se procedió al entrenamiento de una Red de Memoria a Largo Plazo (LSTM). Al igual que en el caso de la RNN, se utilizó el embedding preentrenado de Wikipedia en español cargado con Gensim para representar las palabras. El modelo LSTM, que incluye mecanismos de memoria a largo plazo, fue configurado en Keras para capturar dependencias a largo plazo en las secuencias de texto. La arquitectura del modelo incluyó una o más capas LSTM, y el entrenamiento se realizó en TensorFlow utilizando hiperparámetros estándar. Aunque se utilizaron los embeddings preentrenados, no se realizaron ajustes finos para optimizar los hiperparámetros, ya que el objetivo de este modelo, al igual que el de la RNN, era únicamente servir como referencia para comparar con el Transformer en las etapas posteriores.

4.2.5.2 Implementación del modelo basado en Transformer

Una vez implementados y entrenados los modelos tradicionales, se procedió con la implementación de un modelo Transformer. El modelo elegido fue RoBERTa, una variante de BERT, conocida por su eficacia en tareas de procesamiento de lenguaje natural. Este modelo había sido preentrenado en el conjunto de datos más grande conocido hasta la fecha en español, que incluía 570GB de texto limpio, recopilado de los sitios web crawleados por la Biblioteca Nacional de España (BNE) entre los años 2009 y 2019.

El entrenamiento se realizó utilizando Google Colab, que permitió el uso de GPUs para acelerar el proceso de entrenamiento. Se comenzó cargando el modelo RoBERTa-base-bne preentrenado a través de la librería Hugging Face Transformers, utilizando la función `AutoModelForSequenceClassification` junto con el `AutoTokenizer`, que facilita el acceso tanto al modelo como a su tokenizador preentrenado. Posteriormente, se realizó el preprocesamiento del conjunto de datos de sarcasmo, que incluía titulares y diálogos previamente etiquetados como sarcásticos o no sarcásticos. Para el preprocesamiento, se utilizó Pandas para organizar y manipular los datos, y la tokenización de los textos se llevó a cabo utilizando el tokenizador de RoBERTa. Este tokenizador convierte cada frase en una secuencia numérica que el modelo puede entender, añadiendo automáticamente los tokens especiales de inicio ([CLS]) y fin ([SEP]) de secuencia, necesarios para que el modelo Transformer procese correctamente los límites de cada texto.

El fine-tuning del modelo RoBERTa implicó ajustar sus pesos a los ejemplos específicos de sarcasmo y no sarcasmo en el conjunto de datos. Se configuraron los hiperparámetros iniciales como la tasa de aprendizaje, el número de épocas y el tamaño del lote. Para ello, se utilizó la función `Trainer` de Hugging Face, que facilita la personalización del proceso de entrenamiento y

permite el seguimiento del rendimiento del modelo. Este ajuste fino se realizó durante varias épocas en Google Colab utilizando GPUs, optimizando así el tiempo de entrenamiento.

Este primer entrenamiento permitió tener una versión inicial del modelo adaptado al problema de detección de sarcasmo en español. Las bases para la siguiente fase de optimización de hiperparámetros y ajuste fino quedaron sentadas, mejorando el rendimiento del modelo en la detección de sarcasmos en textos españoles.

4.2.5.3 Búsqueda de hiperparámetros

Después de realizar el primer entrenamiento del modelo RoBERTa-base-bne mediante un proceso de fine-tuning básico, se llevó a cabo una búsqueda exhaustiva de hiperparámetros para optimizar el rendimiento del modelo en la tarea de detección de sarcasmo.

El objetivo de esta fase fue encontrar los valores óptimos para los siguientes parámetros clave:

- **Tasa de aprendizaje (Learning Rate):** Este hiperparámetro define la velocidad con la que el modelo ajusta sus pesos durante el proceso de entrenamiento.
- **Tamaño del Lote (Batch Size):** Se evaluaron diferentes tamaños de lote, lo que afecta tanto a la precisión del modelo como a la velocidad de entrenamiento.
- **Número de Épocas (Epochs):** El número de veces que el modelo pasa por completo sobre el conjunto de entrenamiento fue otro parámetro clave a ajustar. El objetivo era identificar un número de épocas que permitiera al modelo aprender adecuadamente sin tener problemas de sobreajuste.
- **Tasa de Decaimiento de la Tasa de Aprendizaje (Weight Decay):** Se probó la tasa de decaimiento para reducir gradualmente la tasa de aprendizaje durante el entrenamiento, para evitar el sobreajuste del modelo.

Para llevar a cabo la búsqueda de hiperparámetros, se utilizó una combinación de técnicas. Principalmente, se emplearon Grid Search y Random Search a través de la librería Scikit-learn y la integración con Hugging Face's Trainer API para probar las distintas configuraciones. Grid Search consistió en probar todas las combinaciones posibles de un conjunto predefinido de valores para cada hiperparámetro. Por otro lado, Random Search seleccionó aleatoriamente valores dentro de un rango específico, lo que permitió explorar más combinaciones en menos tiempo y de manera eficiente, dado el uso de GPU en Google Colab para acelerar los procesos.

El proceso de búsqueda de hiperparámetros fue dividido en varias etapas:

- En la primera fase, se ajustaron los valores de tasa de aprendizaje y tamaño del lote, ya que son dos de los hiperparámetros más relevantes en el rendimiento del modelo. Se probaron diferentes combinaciones de tasas de aprendizaje (1e-5, 2e-5, 3e-5) y tamaños de lote (8, 16, 32). Para ello, se utilizó la función Trainer de Hugging Face junto con los recursos de TensorBoard para visualizar el comportamiento del modelo durante el entrenamiento.
- En la segunda fase, se optimizó el número de épocas. A través del monitoreo en TensorBoard, se observaron los gráficos de accuracy y loss del modelo por cada época,

lo que ayudó a ajustar este valor para equilibrar un aprendizaje adecuado sin sobreentrenar el modelo.

- Finalmente, se evaluó el decaimiento de la tasa de aprendizaje, probando valores como 0.01, 0.1 y 0.5, con el fin de reducir el impacto de una tasa de aprendizaje alta en fases avanzadas del entrenamiento.

Para la selección del mejor modelo, se diseñó una métrica especial que evaluaba tanto la precisión (accuracy) del modelo como su capacidad de generalización. Esta métrica consistía en restar al accuracy el valor de la pérdida en el conjunto de validación multiplicado por 0,5, penalizando los modelos que, a pesar de tener un accuracy elevado, mostraran una alta pérdida en validación, lo que indicaba sobreajuste. La métrica se definió de la siguiente manera:

$$Validacion = Accuracy - (0,5 \times Validation Loss)$$

Durante todo este proceso, se utilizó un conjunto de validación extraído del dataset, evaluando el rendimiento del modelo en cada combinación de hiperparámetros. Los resultados de esta validación, registrados en TensorBoard, fueron determinantes para seleccionar los mejores parámetros, que luego se aplicaron en el entrenamiento final del modelo.

4.2.6 Evaluación del Desempeño del Modelo

Una vez finalizado el proceso de entrenamiento de todos los modelos, tanto los tradicionales como el modelo Transformer, más en concreto el modelo RoBERTa, se llevó a cabo una evaluación exhaustiva de cada uno. El objetivo de esta evaluación fue analizar cómo cada modelo se desempeñaba en la tarea de detección de sarcasmo, utilizando métricas que permitieran obtener una visión clara de su precisión y capacidad para generalizar. La evaluación se realizó en base al conjunto de datos de test, que no había sido utilizado durante el entrenamiento, para garantizar resultados más objetivos.

4.2.6.1 Métricas de Evaluación

Para la evaluación del desempeño de cada modelo, se utilizaron varias métricas comúnmente aplicadas en problemas de clasificación binaria, calculadas principalmente a través de las funciones de la librería Scikit-learn. Estas métricas permitieron analizar el rendimiento general de los modelos y su capacidad para clasificar correctamente los ejemplos de sarcasmo y no sarcasmo. En el caso del modelo Transformer, adicionalmente se utilizó la Cross Entropy Loss de PyTorch para monitorear la pérdida durante el proceso de entrenamiento:

- **Precisión (Accuracy):** La métrica principal utilizada para medir el porcentaje de predicciones correctas realizadas por el modelo, es decir, el número de ejemplos correctamente clasificados entre el total de ejemplos.
- **Recall (Sensibilidad):** Esta métrica fue utilizada para evaluar qué tan bien el modelo identificaba correctamente los ejemplos de sarcasmo.
- **Precisión (Precision):** La precisión se utilizó para medir cuántas de las predicciones positivas realizadas por el modelo eran efectivamente correctas.

- **F1-Score:** Para obtener una medida entre la precisión y el recall, se calculó el F1-Score. Esta métrica proporciona una visión más clara del desempeño del modelo al tener en cuenta tanto los falsos positivos como los falsos negativos.
- **Pérdida (Loss):** Durante la evaluación también se consideró la función de pérdida, que permite medir la diferencia entre las predicciones del modelo y los resultados reales. Esta métrica se utilizó para monitorear el ajuste del modelo y su capacidad de generalización.

Además de estas métricas tradicionales, para el modelo Transformer durante la búsqueda de hiperparámetros, se utilizó la métrica especial mencionada anteriormente, que penaliza las predicciones erróneas restando la pérdida en el conjunto de validación multiplicada por 0.5 al valor del accuracy. Esta métrica permitió realizar una evaluación más rigurosa y precisa para seleccionar el mejor modelo en términos de balance entre precisión y generalización.

4.2.6.2 Comparación entre Modelos

Una vez obtenidas las métricas para cada modelo, se procedió a la comparación de los resultados. Esta comparativa se llevó a cabo en tres fases clave, utilizando herramientas como Pandas para recopilar y organizar los resultados, y Matplotlib y Seaborn para generar visualizaciones claras que facilitaran la interpretación de los datos. Esta comparativa se realizó en tres fases:

- **Recopilación de Resultados:** Se registraron todas las métricas para cada uno de los modelos implementados: modelo basado en reglas, RNN, LSTM y RoBERTa-base-bne.
- **Análisis Comparativo:** Se realizó una comparación detallada entre los modelos tradicionales y el modelo Transformer. La comparación se centró en dos aspectos clave:
 - **Rendimiento en la detección de sarcasmo:** Se compararon los valores de precisión y F1-score para evaluar qué modelo era capaz de detectar el sarcasmo de manera más efectiva.
 - **Generalización:** Se analizó la pérdida y el recall de cada modelo para ver qué tan bien podían generalizar a datos no vistos.
- **Visualización de Resultados:** Para facilitar la comparación entre los modelos, se generaron gráficos que mostraban el rendimiento de cada uno de ellos en las diferentes métricas.

4.2.7 Interfaz grafica

Para evaluar el modelo de detección de sarcasmo de manera interactiva, se implementó una página web que permite a los usuarios ingresar una frase y obtener una respuesta inmediata sobre si el contenido es sarcástico o no. La metodología detrás de esta implementación se puede dividir en tres fases clave: diseño de la interfaz, integración del modelo de lenguaje, y presentación de resultados.

4.2.7.1 Diseño de la interfaz

La interfaz web fue diseñada para ser intuitiva y accesible, facilitando al usuario la interacción con el modelo. Se creó un formulario central donde el usuario puede escribir una frase. Esta

entrada es enviada al servidor, donde se conecta con el modelo de sarcasmo. Para lograr esto, se usó el archivo `index.html`, que define la estructura de la página, incluyendo el formulario y la sección donde se presentan los resultados. Se hizo uso de Bootstrap para asegurar que la página sea responsive, es decir, que se adapte a diferentes tamaños de pantalla, y se añadió una barra de progreso que indica la probabilidad de sarcasmo en el texto ingresado.

4.2.7.2 Integración del modelo de detección de sarcasmo

La lógica de back-end se maneja a través del archivo `app.py`. Este archivo gestiona el flujo de información entre la interfaz y el modelo de lenguaje. Cuando un usuario envía una frase, la aplicación captura el texto, lo procesa a través del modelo transformer que fue entrenado para detectar sarcasmo, y devuelve un resultado que indica si el texto es sarcástico o no. Para optimizar la interacción, se asegura que el procesamiento sea rápido, permitiendo que el usuario vea los resultados en tiempo real. Este archivo también se encarga de calcular la probabilidad de sarcasmo y enviarla a la página para su visualización.

4.2.7.3 Presentación de resultados

Una vez que el modelo procesa la frase, los resultados se muestran de forma clara en la página. El estilo de presentación se define en el archivo `style.css`, que fue diseñado para proporcionar una experiencia visual agradable y moderna. Los resultados incluyen la probabilidad de sarcasmo y un mensaje basado en la frase del usuario, todo ello presentado en una tarjeta con animaciones suaves que mejoran la experiencia de usuario. Además, se proporciona un indicador visual, en forma de barra de progreso, que muestra la confianza del modelo en su predicción.

4.2.8 Pruebas del Modelo

Para realizar una evaluación adicional del modelo final, se introdujeron manualmente 10 frases, 5 sarcásticas y 5 no sarcásticas, con el fin de comprobar el rendimiento del modelo en la detección de sarcasmo. Estas frases no formaban parte de los datos de entrenamiento ni de validación, y fueron seleccionadas de manera que cubrieran diferentes contextos y estilos de comunicación.

El objetivo de esta prueba era observar la capacidad del modelo para generalizar y detectar correctamente el sarcasmo en textos que nunca había visto antes. Las frases se procesaron utilizando el modelo previamente ajustado, y se registraron los resultados para evaluar la precisión del modelo. Estos resultados fueron almacenados en una tabla para un análisis detallado del rendimiento, permitiendo una fácil comparación entre las predicciones de frases sarcásticas y no sarcásticas.

4.3 Recursos requeridos

A continuación, se enumeran los recursos utilizados para la ejecución del proyecto, tanto a nivel técnico como de soporte externo:

- Hardware:

- Ordenador personal con procesador Intel Core i5, 16 GB de RAM y SSD de 464 GB.
- Google Colab con acceso a GPU para el entrenamiento de los modelos.
- Software y Herramientas:
 - Python (versión 3.8)
 - Hugging Face Transformers para la implementación y ajuste del modelo RoBERTa.
 - PyTorch para el entrenamiento y evaluación de los modelos.
 - Pandas para la manipulación y análisis de datos.
 - Scikit-learn para el cálculo de métricas de evaluación.
 - Matplotlib y Seaborn para la visualización de los resultados.
 - TensorBoard para monitorear el proceso de entrenamiento.
 - NLTK para el preprocesamiento de texto y eliminación de stopwords.
 - Gensim para el uso de embeddings preentrenados en español.
- Recursos de datos:
 - Web Scraping de fuentes de noticias (El Mundo Today, El País, Diario Sur) para la obtención de titulares.
 - Diálogos de series de televisión obtenidos mediante descarga automática de videos de YouTube y AssemblyAI para la transcripción automática.
 - Frases generadas por ChatGPT para textos sarcásticos adicionales.
- Asistencia Externa:
 - Cursos especializados:
 - Transformers de Hugging Face.
 - NLP de 0 a 100 con Hugging Face por Spain AI.
 - Curso de Redes Neuronales en Python de Udemy.

4.4 Presupuesto

Tipo de coste	Valor	Comentarios
Horas de trabajo en el proyecto	2666€	13,33€/h que cobra un analista de datos Junior promedio en España por 200 horas de trabajo invertidas en el proyecto.
Equipo técnico utilizado	650€	HP Laptop 15-da1xxx con procesador Intel Core i5-8265U, 16 GB de RAM.
Software utilizado	11,99 € (Colab) + aprox. 0,03 € (Hugging Face)	Un mes de Google Colab Pro (11,99 €) + CPU en Hugging Face (0,03 €/h, aprox. 500 horas).
Estudios e informes	0 €	No se ha requerido adquirir informes o estudios adicionales.

Materiales empleados	0 €	No se han utilizado materiales adicionales.
----------------------	-----	---

Tabla 1: Presupuesto del proyecto

4.5 Viabilidad

El coste total del proyecto ha sido de aproximadamente 3.350,99 €, considerando las horas de trabajo, el equipo técnico, y los costos de servicios como Google Colab y Hugging Face. Las horas de trabajo han sido valoradas en 2.666 €, lo que representa la mayor parte del coste del proyecto. Esto es un costo razonable, teniendo en cuenta que el desarrollo de modelos avanzados como los Transformers ajustados a tareas específicas de PLN suelen requerir recursos y tiempo significativos.

El principal beneficio del proyecto es la creación de un modelo ajustado específicamente para el idioma español, con aplicaciones en áreas como:

- **Análisis de contenido:** Detectar sarcasmo en comentarios, redes sociales, y titulares de noticias puede ser útil para empresas de monitoreo de medios y agencias de marketing.
- **Asistentes virtuales y chatbots:** Incluir la capacidad de detectar sarcasmo en estos sistemas puede mejorar su comprensión del lenguaje y proporcionar respuestas más contextuales.
- **Investigación académica:** El modelo puede servir como base para futuros estudios en PLN centrados en el sarcasmo, aportando al desarrollo académico en este campo.

A nivel de costes, se espera que el proyecto pueda recuperar la inversión si se comercializa el modelo como un servicio especializado (API), o si se licencia para uso en herramientas de análisis de medios, marketing o investigación. La demanda de herramientas de análisis avanzadas del lenguaje en español está en constante crecimiento, lo que puede facilitar la generación de ingresos.

En conclusión, el proyecto es viable tanto económica como técnicamente. Presenta un coste controlado que puede recuperarse fácilmente mediante la comercialización del modelo y cuenta con la capacidad de adaptarse y mantenerse vigente en el tiempo gracias a las herramientas utilizadas y la posibilidad de actualizarlo con nuevos datos.

4.6 Resultados del proyecto

En este apartado se presentarán los resultados obtenidos a lo largo del desarrollo del proyecto. El objetivo de este apartado es exponer cómo los diferentes objetivos específicos del trabajo han sido cumplidos a través de la implementación de las técnicas y herramientas mencionadas en la metodología.

Los resultados se presentarán siguiendo la estructura de los objetivos específicos definidos al inicio del proyecto, comenzando con la investigación del estado del arte, pasando por la

recopilación y procesamiento de datos, y terminando con la implementación, optimización, y evaluación del modelo final de detección de sarcasmo en español.

4.6.1 Investigación inicial

Durante la investigación, se revisaron diversas fuentes académicas que estudiaban el fenómeno lingüístico del sarcasmo, sus características y su detección mediante técnicas de procesamiento de lenguaje natural. Se inició con la definición del sarcasmo, diferenciándolo de la ironía, y se profundizó en los retos que presenta su detección automática, particularmente en el contexto del español.

En primer lugar, el análisis comenzó con la revisión de las técnicas tradicionales de PLN, como los modelos basados en reglas y clasificadores supervisados. Estos primeros enfoques, aunque útiles para la época y en ciertos contextos, presentaban limitaciones importantes, como la incapacidad de capturar las sutilezas del lenguaje y el contexto en el que el sarcasmo se produce. Asimismo, se exploró el avance en las redes neuronales, como las RNN, CNN y LSTM, las cuales mejoraron la capacidad de detectar patrones de lenguaje más complejos. Sin embargo, estas arquitecturas también enfrentaban dificultades, principalmente por su naturaleza secuencial, lo que afectaba su capacidad para interpretar secuencias largas de texto.

La investigación avanzó hacia el uso de modelos basados en Transformers, los cuales revolucionaron el PLN al superar las limitaciones de los modelos anteriores. Se realizó un análisis exhaustivo de las arquitecturas de modelos además de los modelos específicos como BERT, RoBERTa y GPT, destacando cómo su mecanismo de atención les permite capturar dependencias contextuales más complejas, lo cual es fundamental para la detección de sarcasmos. A través de estudios previos se validó la superioridad de estos modelos sobre las redes neuronales recurrentes (RNN y LSTM) en tareas de clasificación de texto y detección de sarcasmo. En particular, se destacó el uso de RoBERTa, que ha mostrado mejoras en la detección de sarcasmos en plataformas como Twitter y Reddit.

Uno de los aprendizajes más valiosos obtenidos durante esta fase fue el uso de la plataforma Hugging Face para realizar el fine-tuning del modelo preentrenado en español. Hugging Face proporciona un ecosistema completo para la manipulación de modelos Transformer, facilitando la implementación de tareas personalizadas a partir de modelos preentrenados. A través de este proceso, se aprendió cómo cargar, adaptar y entrenar estos modelos en un conjunto de datos específico de sarcasmo.

En conclusión, la revisión de modelos y enfoques permitió identificar que los Transformers, especialmente RoBERTa, son superiores a las redes neuronales tradicionales para la detección de sarcasmos. Los modelos basados en Transformers no solo captan el contexto de manera más eficiente, sino que también son más adaptables mediante el proceso de fine-tuning, lo que les permite especializarse en tareas como la detección de sarcasmo en español.

4.6.2 Recopilación de datos

Para el desarrollo de este proyecto, fue de suma importancia la obtención de un conjunto de datos tanto sarcásticos como no sarcásticos en español, provenientes de diferentes fuentes. El

enfoque se centró en recopilar datos de tres tipos de fuentes: titulares de noticias sarcásticas y no sarcásticas, diálogos de series de televisión y frases generadas por inteligencia artificial. Este conjunto de datos representaba diferentes contextos lingüísticos y culturales, lo que permitió crear un modelo más robusto y capaz de detectar sarcasmos en español.

4.6.2.1 Noticias sarcásticas

Se obtuvo una gran cantidad de datos del periódico satírico "El Mundo Today", un medio muy conocido en España que produce noticias humorísticas. A través de técnicas de web scraping, se extrajeron alrededor de 9.000 titulares. Esta fuente fue clave para obtener ejemplos claros de sarcasmos escritos en un formato de titulares de noticias, que es uno de los objetivos principales de este proyecto.

1. Se compra los nuevos calzoncillos cagados de Balenciaga y denuncia que se le han estropeado con el primer lavado
2. Feijóo se ofrece a pactar con la ultraderecha francesa para evitar que Pierre Sanchoise llegue al Palacio del Elíseo
3. Los consumidores sospechan que millones de reseñas negativas de mosquiteras, insecticidas y repelentes podrían haber sido escritas por los propios mosquitos
4. ¿Está tu casa preparada para una visita de Zazza el italiano?
5. Ningún ciudadano podrá morir sin rellenar antes una encuesta de satisfacción

Ilustración 19: Ejemplo de extracción de noticias sarcásticas

Fuente: Elaboración propia

4.6.2.2 Noticias no sarcásticas

Para equilibrar los datos sarcásticos, se decidió obtener titulares no sarcásticos de periódicos populares como El País y El Diario Sur. Se utilizó un proceso similar de web scraping automatizado, en el que se desarrolló un script en Python para extraer titulares de noticias. A través de búsquedas automáticas basadas en una lista de palabras clave, se recolectaron alrededor de 14.500 titulares no sarcásticos.

1. Hunter Biden, primer hijo de un presidente de Estados Unidos que se sienta en el banquillo
2. El jurado empieza a deliberar si Hunter Biden es culpable por la compra y posesión de un arma
3. Biden hace campaña con la inmigración
4. A favor o en contra: ¿qué futuro le espera al bidé?
5. Me gustaría no pensar que Biden va a perder
6. Un atisbo de esperanza para Gaza
7. La libertad como cortina de humo fiscal
8. EE UU, España y otros 15 países urgen a Israel y Hamás a aceptar la nueva propuesta de alto el fuego
9. Qué consecuencias tiene la condena del hijo de Biden para la campaña presidencial
10. Netanyahu enfría las esperanzas de un alto el fuego permanente en Gaza

Ilustración 20: Ejemplo de extracción de noticias no sarcásticas

Fuente: Elaboración propia

4.6.2.3 Diálogos de Series de Televisión

Debido a que los sarcasmos presentes en los titulares de noticias podrían resultar demasiado complejos para generalizar en el uso del sarcasmo cotidiano, también se extrajeron diálogos de series de televisión. Se seleccionaron programas de humor como Dos Hombres y Medio y House, conocidos por su uso frecuente de sarcasmo. Para ello, se descargaron automáticamente clips de estas series desde YouTube utilizando un script en Python. Posteriormente, se extrajo el audio de estos clips y, con la ayuda de herramientas de transcripción automática como AssemblyAI para un posterior etiquetado manual. Con este método se obtuvieron las

transcripciones etiquetadas de aproximadamente unas mil frases de diálogos, identificando los interlocutores.

```
Frase Anterior;Frase Actual;Sarcasmo
;Gracias por verme con tan poca antelación.;0
Gracias por verme con tan poca antelación.;No hay problema, Charlie. ¿Entonces, cuál es la emergencia?;0
No hay problema, Charlie. ¿Entonces, cuál es la emergencia?;¿Por qué ha de ser una emergencia? Tal vez solo quería venir.;0
¿Por qué ha de ser una emergencia? Tal vez solo quería venir.;Bueno, dejaste seis mensajes.;0
Bueno, dejaste seis mensajes.;Ah, ese fue un borracho que presionó el marcado rápido.;1
Ah, ese fue un borracho que presionó el marcado rápido.;Tengo 5 min. Antes de mi próximo paciente. Dime la noticia.;0
Tengo 5 min. Antes de mi próximo paciente. Dime la noticia.;Correcto. Salgo con una mujer.;0
Correcto. Salgo con una mujer.;Esa no es noticia. Es el nombre del diario.;1
Esa no es noticia. Es el nombre del diario.;Lo sé, lo sé. Pero ella es diferente del tipo de mujeres con las que salgo.;0
Lo sé, lo sé. Pero ella es diferente del tipo de mujeres con las que salgo.;¿Sí? ¿Diferente en qué?;0
¿Sí? ¿Diferente en qué?;¿Yousand? Bueno, es un poco mayor.;0
¿Yousand? Bueno, es un poco mayor.;No podría salir con una más joven sin tener que registrarte con las autoridades.;1
No podría salir con una más joven sin tener que registrarte con las autoridades.;Me refiero a que es mayor que yo.;0
Me refiero a que es mayor que yo.;Oh, ya está la noticia.;1
Oh, ya está la noticia.;Y estoy confundido porque me estoy enamorando mucho con ella.;0
Y estoy confundido porque me estoy enamorando mucho con ella.;¿La encuentras atractiva?;0
¿La encuentras atractiva?;Sí, es preciosa.;0
Sí, es preciosa.;¿Qué tal el sexo?;0
¿Qué tal el sexo?;Aún no lo hemos tenido.;0
Aún no lo hemos tenido.;Oh, dejen de imprimir esa.;1
```

Ilustración 21: Ejemplo de extracción de diálogos de series de televisión

Fuente: Elaboración propia

4.6.2.4 Sarcasmo generado por Inteligencia Artificial

Finalmente, para complementar los diálogos y titulares recopilados, se utilizaron sistemas de generación de texto como ChatGPT para crear más frases sarcásticas en español. Se obtuvieron alrededor de unas 700 frases, tanto sarcásticas como no sarcásticas.

```
1 Frases Sarcásticas
2 Nada como empezar el día con 10 correos urgentes que podrían esperar.
3 Increíble cómo los problemas siempre aparecen justo antes de salir.
4 "Qué bien, otra reunión que no cambiará absolutamente nada."
5 "La impresora atascada otra vez, ¡debe ser un día especial!"
6 "Por supuesto que necesito más trabajo de última hora, gracias."
7 ¿Un informe de 30 páginas para mañana? ¡Pan comido!
8 "Ah, sí, el cliente quiere más cambios... otra vez."
9 Nada dice productividad como reuniones eternas y sin conclusiones.
10 "Qué alivio, otro correo que no necesitaba leer."
11 Trabajar horas extra no remuneradas realmente mejora el ánimo.
12 La mejor parte del día es cuando la tecnología decide fallar.
13 ¿Plazos imposibles? ¡Todo un reto diario!
14 Qué sorpresa que nadie sabe lo que quiere hasta que ya es tarde.
15 "Ah, claro, todos los problemas son siempre ""urgentes""."
16 "Qué suerte la mía, otra vez haciendo el trabajo de los demás."
17 No hay nada como recibir tareas justo cuando estás por terminar.
```

Ilustración 22: Ejemplo de extracción de frases creadas por inteligencia artificial

Fuente: Elaboración propia

En resumen, se logró obtener un conjunto de datos completo compuesto por aproximadamente 19.000 frases, obtenidas de diversas fuentes que contienen una amplia variedad de contextos tanto lingüísticos como culturales. Los datos recopilados incluyen tanto sarcasmos claros y complejos como ejemplos no sarcásticos, lo que permitió crear un conjunto de datos balanceado y rico en ejemplos.

Las noticias sarcásticas de "El Mundo Today" proporcionaron una base sólida para entender sarcasmos con una alta complejidad, mientras que las noticias no sarcásticas de periódicos como "El País" y "El Diario Sur" permitieron generar un contraste para que el modelo pudiera aprender las diferencias entre ambos tipos de textos. Asimismo, los diálogos de series de televisión, ofrecieron ejemplos de sarcasmo en situaciones cotidianas. Finalmente, las frases generadas por inteligencia artificial complementaron el conjunto de datos, añadiendo ejemplos de sarcasmo en contextos variados que no estaban presentes en las fuentes originales.

Este proceso de recopilación no solo permitió obtener un conjunto de datos amplio, sino que también garantizó que el modelo entrenado pudiera captar las diferencias contextuales del sarcasmo en español. El conjunto final de datos es único en su composición, al incluir tanto titulares de noticias como diálogos televisivos y frases generadas por IA, lo que enriquece su diversidad y aplicabilidad para la tarea de detección automática de sarcasmos.

4.6.3 Proceso de ETL (Extracción, Transformación y Carga)

El proceso de Extracción, Transformación y Carga (ETL) fue fundamental para asegurar que el conjunto de datos recopilado estuviera limpio, organizado y listo para ser utilizado en el análisis exploratorio de datos (EDA) y en el posterior entrenamiento del modelo de detección de sarcasmo. Este proceso permitió transformar los datos en bruto obtenidos de diferentes fuentes en un formato unificado y adecuado para posteriores operaciones.

4.6.3.1 Extracción de Datos

El proceso de extracción se centró en reunir todos los datos que previamente se habían obtenido y almacenado en el equipo. En primer lugar, se extrajeron 23.500 titulares: 9.000 noticias sarcásticas de El Mundo Today y 14.500 titulares no sarcásticos de otros periódicos como El País y El Diario Sur. Estos datos ya estaban etiquetados, lo que facilitó su integración en el dataset.

También se extrajeron 1.000 diálogos de series de televisión como Dos Hombres y Medio y House, que habían sido transcritos y etiquetados previamente, proporcionando una fuente adicional de ejemplos para el análisis.

Por último, se extrajeron 14 datasets de frases generadas por ChatGPT, sumando 700 frases tanto sarcásticas como no sarcásticas. Estas frases abarcan diferentes contextos, complementando los datos de noticias y diálogos.

4.6.3.2 Transformación de Datos

Tras la extracción de los datos, se procedió a su transformación para asegurar que estuvieran listos para el análisis. En primer lugar, se logró fusionar los datasets de noticias sarcásticas y no sarcásticas en un único conjunto, sumando un total de 24.500 titulares debidamente etiquetados. Además, se integraron los 1.000 diálogos de series de televisión y las 500 frases generadas por inteligencia artificial.

El proceso de limpieza permitió eliminar registros duplicados y valores nulos, lo que garantizó que no hubiera entradas defectuosas en el conjunto de datos final. Se normalizaron las etiquetas de sarcasmo, asegurando que todas las entradas tuvieran valores consistentes de 0 (no

sarcástico) y 1 (sarcástico). Como parte de este proceso, se eliminaron alrededor de 200 textos relacionados con el horóscopo, que no se consideraron relevantes para el análisis de sarcasmo.

Finalmente, tras esta transformación, el conjunto de datos final quedó limpio, organizado y listo para ser utilizado, con un total de aproximadamente 19.000 registros.

	Texto	Sarcasmo
0	El fin de semana pasado salí con mis amigos y ...	0
1	Me encanta reunirme con mis amigos para ponern...	0
2	Este sábado tengo una quedada con amigos y est...	0
3	Siempre que nos reunimos con mis amigos, las c...	0
4	Anoche salimos a cenar con mis amigos y disfru...	0
...
24294	Un niño de siete años empieza a sospechar que ...	1
24295	Una señora cree realmente que su hija podría i...	1
24296	Miles de palomas españolas emigran a Alemania ...	1
24297	El Real Madrid ficha por error al actor Christ...	1
24298	Jesucristo se aparece en un capítulo de Los Si...	1

19096 rows × 2 columns

Ilustración 23: Dataframe limpio

Fuente: Elaboración propia

4.6.3.3 Carga de Datos

Después de la transformación, los datos procesados y limpios fueron almacenados nuevamente en el equipo en formato CSV, listos para su posterior análisis. El dataset consolidado, que incluye 19096 registros entre titulares de noticias, diálogos de series de televisión y frases generadas por inteligencia artificial, quedó guardado en un único archivo para su uso en el análisis exploratorio de datos (EDA) y el entrenamiento del modelo.


 Texto_procesado	05/10/2024 18:44	Archivo de valores...	1.529 KB
---	------------------	-----------------------	----------

Ilustración 24: CSV guardado en el equipo

Fuente: Elaboración propia

4.6.4 Análisis Exploratorio de los Datos

En este apartado, se presentarán los principales resultados obtenidos en el EDA, incluyendo el análisis básico de los datos, la distribución de la longitud de los textos, el análisis de frecuencia de palabras, los bigramas y trigramas más comunes, el análisis de sentimiento, las correlaciones entre las variables clave y el análisis de variabilidad léxica. Estos análisis proporcionan una base sólida para las siguientes fases del proyecto, facilitando el entrenamiento y evaluación de los modelos de machine learning.

4.6.4.1 Análisis básico de los datos

Para los resultados del análisis básico, el dataset está compuesto por 19,096 registros con dos columnas: una para el texto y otra que contiene las etiquetas binarias que indican si el texto es sarcástico (1) o no (0). A partir de la inspección de la estructura del conjunto de datos, se puede observar que no hay valores nulos, lo que asegura que el dataset está limpio y listo para su análisis. En la gráfica que muestra la distribución de etiquetas de sarcasmo, se observa que unos 11,500 registros no son sarcásticos y alrededor de 8,000 sí lo son.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19096 entries, 0 to 19095
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Texto       19096 non-null  object
1   Sarcasmo    19096 non-null  int64
dtypes: int64(1), object(1)
memory usage: 298.5+ KB
```

Ilustración 26: Descripción del DataFrame utilizado en el análisis de sarcasmo.

Fuente: Elaboración propia.

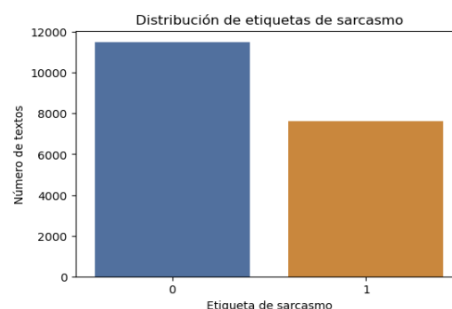


Ilustración 25: Distribución de las etiquetas de sarcasmo en el conjunto de datos.

Fuente: Elaboración propia

4.6.4.2 Análisis de la longitud del texto y detección de valores atípicos

Para el análisis de la longitud de los textos en primer lugar se muestran las estadísticas descriptivas en donde:

- Media: La longitud promedio de los textos es de 72 caracteres.
- Mediana: La mediana también es de 72 caracteres, lo que indica una distribución bastante centrada alrededor de este valor.
- Desviación estándar: La desviación estándar es de aproximadamente 33 caracteres, lo que sugiere que hay cierta variabilidad en la longitud de los textos, aunque no es extremadamente alta.
- Percentiles:
 - El 25% de los textos tienen una longitud menor o igual a 52 caracteres.
 - El 75% de los textos tienen una longitud menor o igual a 91 caracteres, lo que significa que la gran mayoría de los textos tienen entre 52 y 91 caracteres.

- Mínimo y máximo:
 - El texto más corto tiene una longitud de 2 caracteres.
 - El texto más largo alcanza los 1,224 caracteres, lo que sugiere la presencia de valores atípicos (outliers) que podrían influir en los resultados de análisis posteriores.

```
count    19096.000000
mean      72.072738
std       33.166319
min        2.000000
25%       51.000000
50%       72.000000
75%       91.000000
max      1224.000000
Name: longitud_texto, dtype: float64
```

Ilustración 27: Estadísticas descriptivas de la longitud de los textos en el conjunto de datos.

Fuente: Elaboración propia.

Asimismo, se realizó un histograma para ver la distribución de la longitud de los textos.

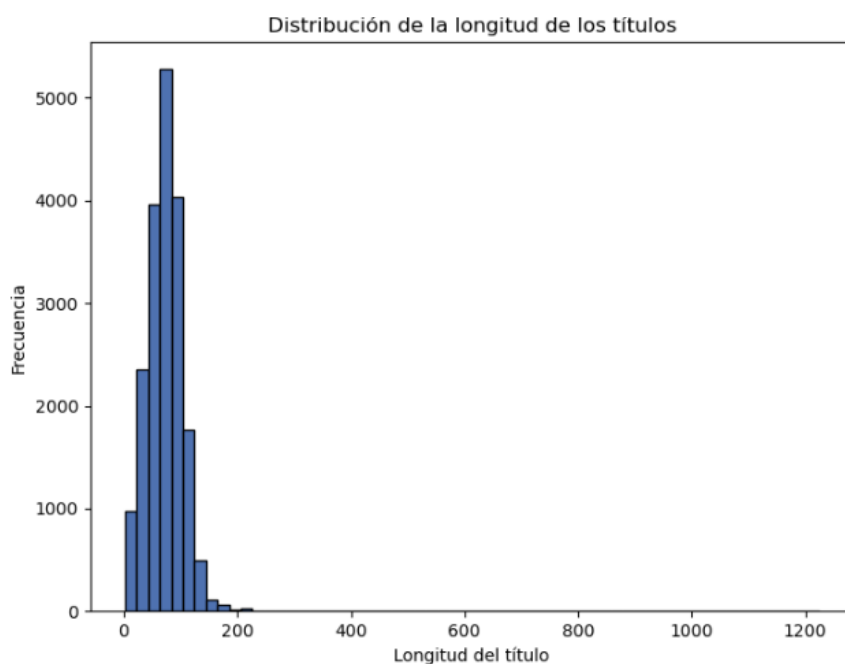


Ilustración 28: Distribución de la longitud de los textos en el conjunto de datos.

Fuente: Elaboración propia.

Como se puede ver en la gráfica la mayoría de los textos tienen una longitud que oscila entre 50 y 200 caracteres. La distribución tiene una cola larga hacia la derecha, lo que indica la presencia

de textos con una longitud considerablemente mayor que el promedio. Estos valores extremos, que corresponden a las observaciones fuera de lo común, serán considerados como valores atípicos en los análisis posteriores.

Además, se incluye un diagrama de cajas y bigotes que compara la longitud de los textos en función de si son sarcásticos o no.

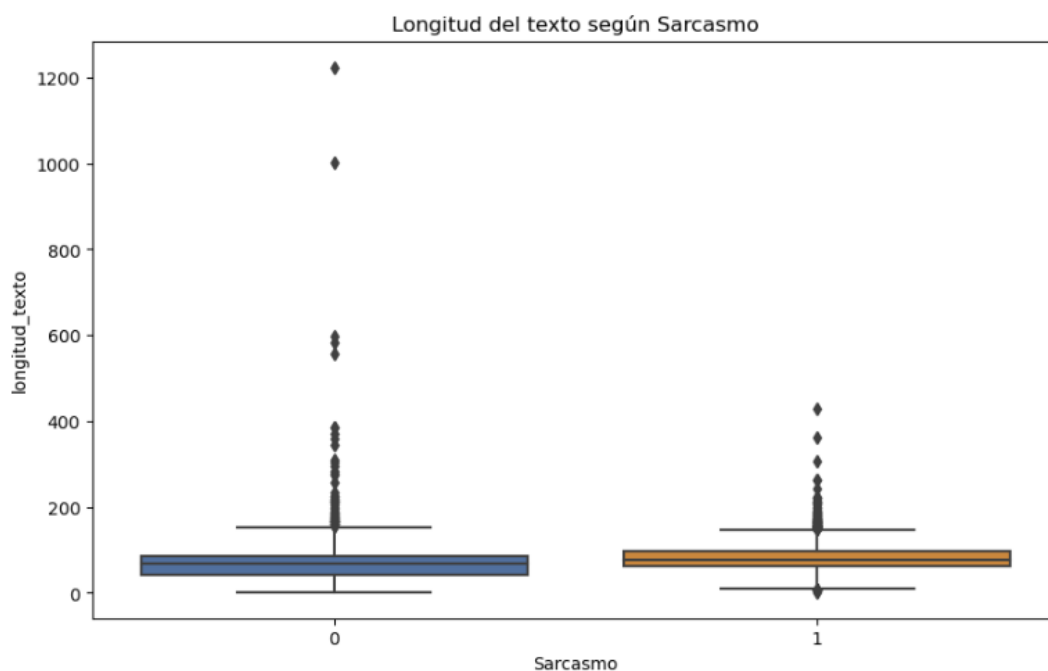


Ilustración 29: Comparación de la longitud del texto según si es sarcástico o no.

Fuente: Elaboración propia.

Como se ha analizado previamente en el histograma, los textos con una longitud superior a 200 caracteres son considerados valores atípicos por el diagrama de caja. Asimismo, la mediana de la longitud de los textos sarcásticos y no sarcásticos es muy similar. Esto podría indicar que la longitud del texto no parece ser un factor determinante para distinguir entre sarcasmo y no sarcasmo.

4.6.4.3 Análisis de frecuencia de palabras

En este apartado se presenta un análisis detallado de la frecuencia de palabras dentro de los textos sarcásticos y no sarcásticos del conjunto de datos. El objetivo principal de este análisis es identificar patrones lingüísticos recurrentes y observar cómo estos varían dependiendo de la presencia o ausencia de valores atípicos. Se han generado visualizaciones en forma de gráficos de barras y nubes de palabras que permiten comparar de manera clara el uso del lenguaje en ambos tipos de textos. Estos análisis proporcionan información valiosa sobre los términos más comunes en textos sarcásticos y no sarcásticos, y permiten evaluar si el sarcasmo influye en el uso de ciertas palabras en relación con los textos no sarcásticos.

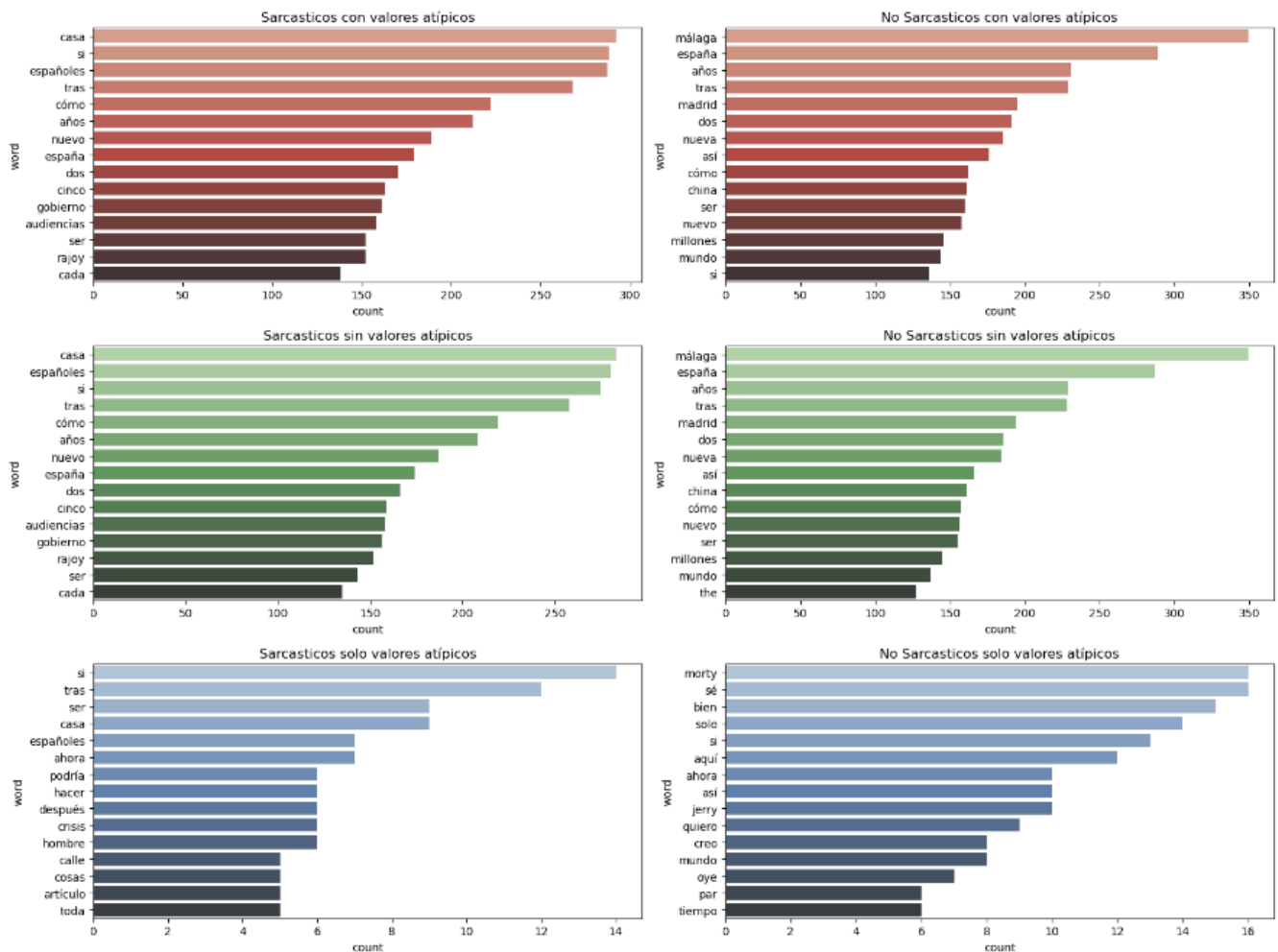


Ilustración 30: Frecuencia de palabras en textos sarcásticos y no sarcásticos con y sin valores atípicos.

Fuente: Elaboración propia.

En estos gráficos de barras se muestran seis comparativas sobre la frecuencia de palabras en textos sarcásticos y no sarcásticos, diferenciados por la presencia o ausencia de valores atípicos. Estos gráficos permiten observar variaciones en el uso del lenguaje entre los distintos grupos, proporcionando una visión clara de los temas más recurrentes.

En los gráficos rojos (sarcásticos y no sarcásticos con valores atípicos), se puede observar que en los textos sarcásticos aparecen palabras como "casa", "españoles" y "cómo", lo que sugiere una tendencia a tratar temas generales o políticos. En contraste, los textos no sarcásticos contienen palabras como "Málaga", "España" y "años", que tienden a centrarse más en descripciones geográficas y temporales.

En los gráficos verdes (sarcásticos y no sarcásticos sin valores atípicos), los temas siguen siendo similares, ya que aparecen palabras como "casa", "españoles", y "Málaga" de manera constante.

Esto indica que la eliminación de valores atípicos no altera significativamente los patrones lingüísticos.

Finalmente, en los gráficos azules (sarcásticos y no sarcásticos solo con valores atípicos), se pueden notar cambios más específicos. En los textos sarcásticos destacan palabras como "ser", "podría" y "tras", lo que sugiere un lenguaje más especulativo. En los textos no sarcásticos, la aparición de nombres propios como "Morty" y "Jerry" revela que los diálogos provienen de series de televisión, con un enfoque más narrativo y coloquial.

En resumen, el análisis muestra que el sarcasmo se expresa de manera similar en todos los grupos, mientras que los textos no sarcásticos varían más en función de sus fuentes, especialmente cuando provienen de diálogos televisivos.

4.6.4.4 Análisis TF-IDF

A continuación, realizaremos un análisis TF-IDF para identificar las palabras más relevantes en los textos sarcásticos y no sarcásticos, tanto con como sin valores atípicos. Este enfoque nos permitirá evaluar la importancia de cada palabra en relación con la frecuencia de su aparición en todo el conjunto de datos.

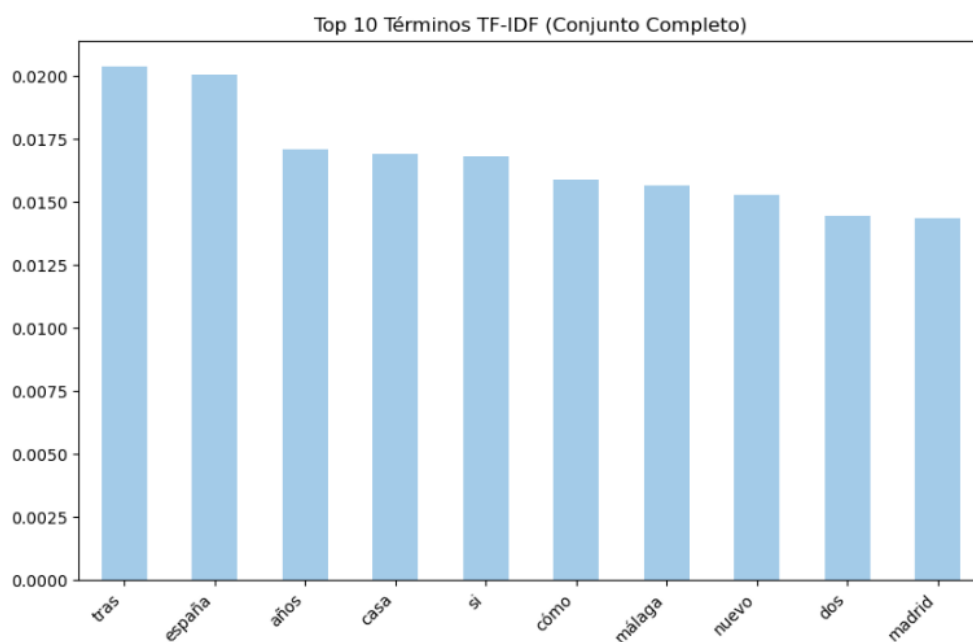


Ilustración 31: Top 10 términos más importantes según el análisis TF-IDF (conjunto completo)

Fuente: Elaboración propia.

En este gráfico se muestran los 10 términos más importantes según el análisis TF-IDF aplicado al conjunto completo de datos. El TF-IDF (frecuencia inversa de documentos) asigna un valor mayor a aquellas palabras que son importantes en un texto pero que no son tan comunes en el resto del corpus, resaltando así su relevancia en contextos específicos.

Las palabras con el mayor valor TF-IDF son "tras", "España", "años", "casa", y "sí". Estas palabras tienen un mayor peso dentro del análisis porque, aunque se repiten en diferentes textos, no son tan frecuentes en el conjunto completo, lo que las hace relevantes en contextos particulares.

Vemos una fuerte coincidencia con los términos más destacados en los análisis de palabras más frecuentes de textos sarcásticos y no sarcásticos, tanto con cómo sin valores atípicos. Palabras como "tras", "España", "años", y "casa" ya aparecían como términos predominantes en esos análisis visuales. Esto refuerza la idea de que estos términos son recurrentes y relevantes, no solo por su frecuencia, sino también por su importancia en diferentes contextos del corpus.

El TF-IDF nos muestra que, aunque estas palabras son frecuentes, su aparición en contextos específicos es significativa. Por ejemplo, en los textos no sarcásticos, "Málaga" y "Madrid" tienen un alto valor TF-IDF, lo que refleja su relevancia geográfica en los textos no sarcásticos, algo que también se vio reflejado en las nubes de palabras de textos no sarcásticos.

Aunque el análisis TF-IDF no distingue entre textos sarcásticos y no sarcásticos en este gráfico, las palabras identificadas como importantes coinciden en su mayoría con las observadas en los textos sarcásticos. Esto sugiere que términos como "casa", "tras" y "cómo" podrían tener un papel clave en el sarcasmo, aunque también son comunes en textos no sarcásticos.

Este análisis TF-IDF refuerza los patrones observados en el análisis de palabras más frecuentes: las palabras más frecuentes no solo aparecen con regularidad, sino que tienen un peso importante en contextos específicos, tanto en textos sarcásticos como no sarcásticos. El análisis TF-IDF nos permite afinar esta interpretación al destacar qué términos, aunque comunes, son especialmente relevantes en el corpus analizado.

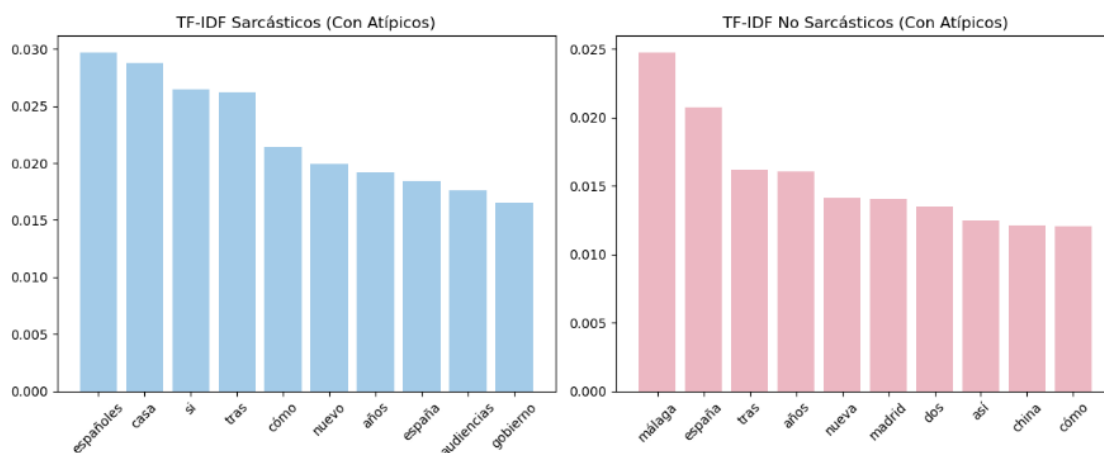


Ilustración 32: Top 10 términos TF-IDF en textos sarcásticos y no sarcásticos con valores atípicos.

Fuente: Elaboración propia.

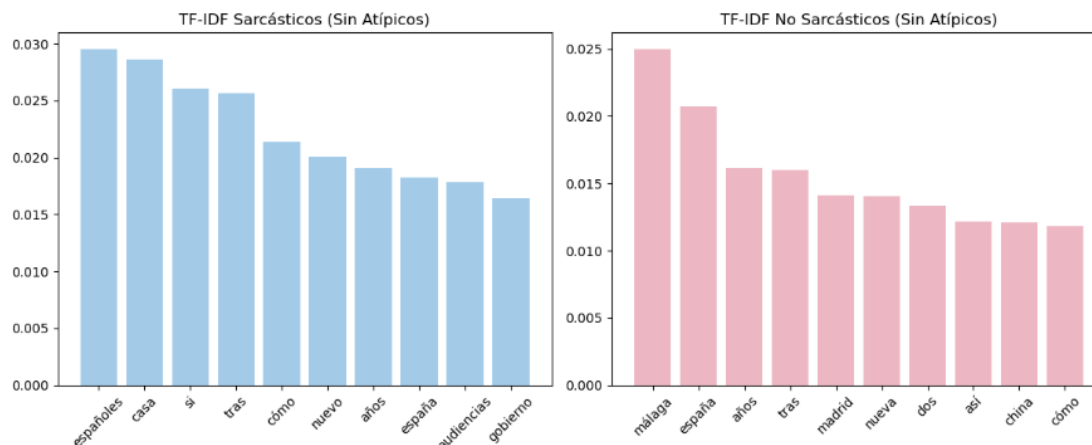


Ilustración 33: Top 10 términos TF-IDF en textos sarcásticos y no sarcásticos sin valores atípicos.

Fuente: Elaboración propia.

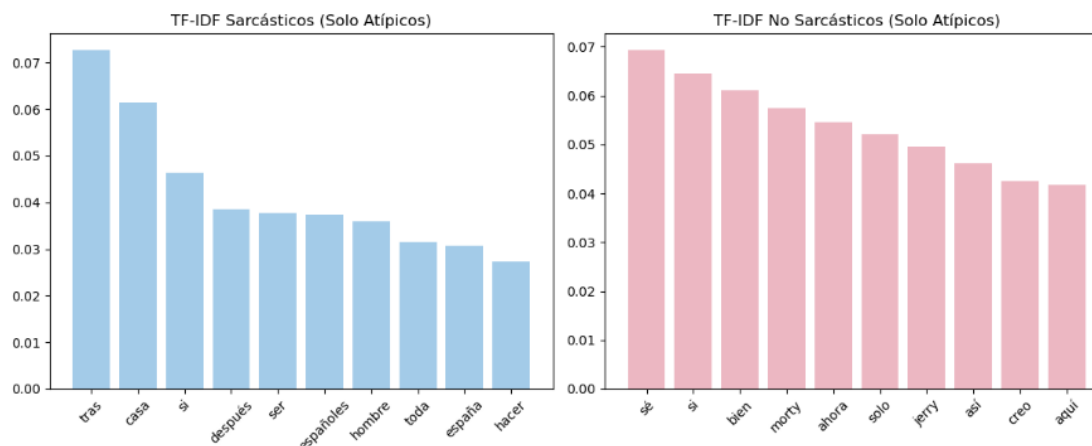


Ilustración 34: Top 10 términos TF-IDF en textos sarcásticos y no sarcásticos (solo valores atípicos).

Fuente: Elaboración propia.

En la comparativa de los gráficos de TF-IDF para textos sarcásticos y no sarcásticos con valores atípicos, se observa que en los textos sarcásticos predominan términos como "españoles", "casa", y "sí", lo que refuerza el enfoque en temas cotidianos y políticos, ya mencionado en análisis anteriores. Por otro lado, en los textos no sarcásticos, destacan palabras como "Málaga", "España", y "años", lo que sugiere que estos textos tienden a centrarse en descripciones más geográficas y temporales, como ya se vio en los análisis de palabras más frecuentes.

Al realizar el mismo análisis sin valores atípicos, los resultados obtenidos son prácticamente los mismos. Las palabras clave que aparecen en los textos sarcásticos y no sarcásticos sin valores atípicos coinciden con las que se observaron en los textos con valores atípicos, lo que sugiere que la presencia o ausencia de valores atípicos no afecta significativamente la relevancia de los términos en los textos.

4.6.4.5 Análisis de bigramas y trigramas

A continuación, pasaremos al análisis de bigramas y trigramas para identificar combinaciones de dos y tres palabras que aparecen con mayor frecuencia en los textos, lo que nos permitirá profundizar en las relaciones y patrones lingüísticos presentes en el texto. Para ello realizaremos dos enfoques, el análisis con stop words y sin stop words.

4.6.4.5.1 Análisis con stop words

Comenzaremos con el Análisis de bigramas y trigramas, enfocándonos en las combinaciones de dos palabras más frecuentes dentro del texto. Iniciaremos este análisis incluyendo stopwords (palabras vacías) para observar cómo estas palabras comunes, como preposiciones y artículos, forman parte de las combinaciones más utilizadas. Analizar tanto con cómo sin stopwords nos permitirá evaluar la estructura natural del lenguaje y las combinaciones que pueden ser relevantes en el contexto.

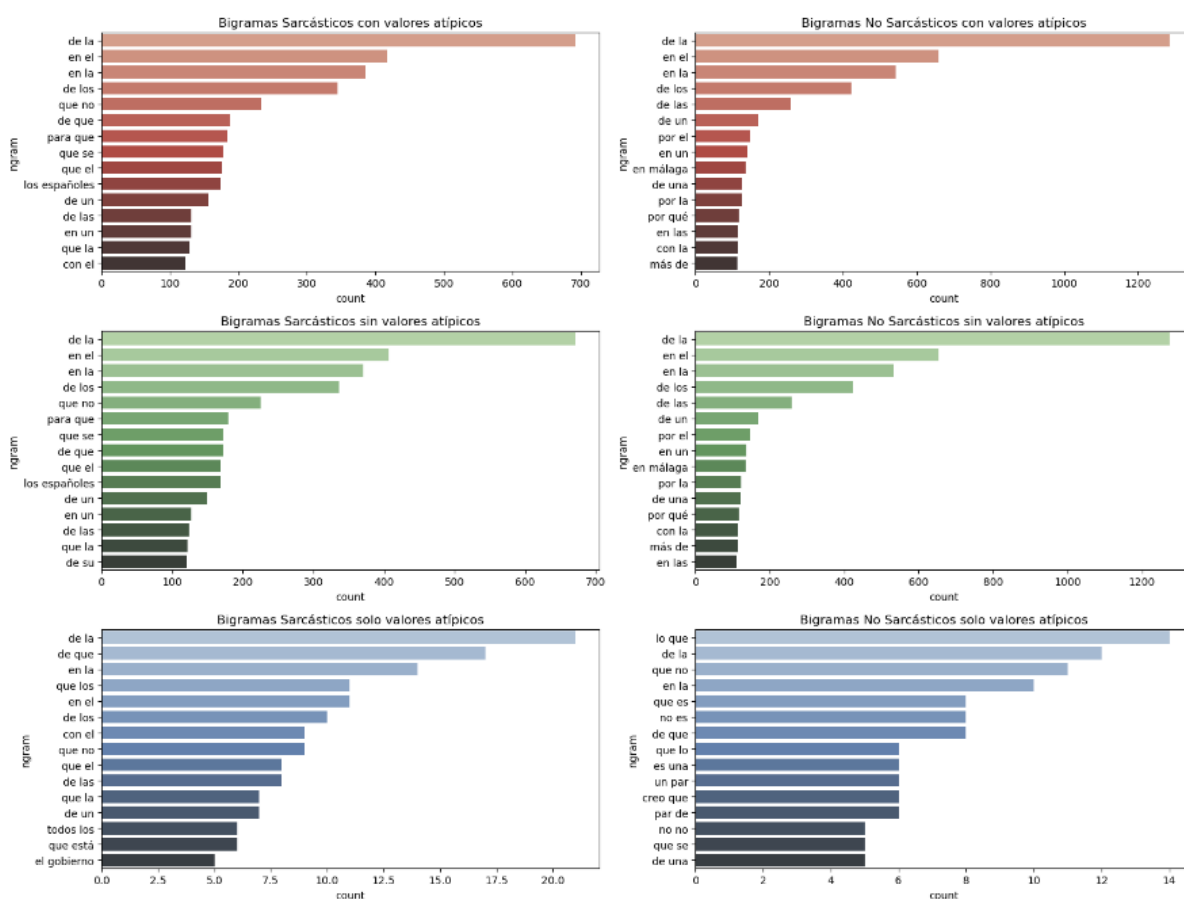


Ilustración 35: Frecuencia de bigramas con stop words en textos sarcásticos y no sarcásticos con y sin valores atípicos.

Fuente: Elaboración propia.

Fuente: Elaboración propia.

descriptivas como "costa del sol", "fin de semana" y "de junio de", lo que refuerza un enfoque más informativo y geográfico en estos textos.

Al comparar los textos con y sin valores atípicos, no se observan diferencias significativas en la mayoría de los trigramas dominantes. Las combinaciones son bastante similares, tanto en los textos sarcásticos como no sarcásticos, lo que indica que la presencia de valores atípicos no altera significativamente los patrones de uso del lenguaje. En los textos con solo valores atípicos, se mantienen patrones similares, aunque surgen algunas combinaciones más inusuales, como "después de que" o "lucha entre los", que podrían reflejar construcciones más complejas o fragmentos de diálogos.

4.6.4.5.2 Análisis sin stop words

En este apartado se probará realizar el análisis para ver cuales son los bigramas y trigramas de mayor importancia en nuestro corpus.

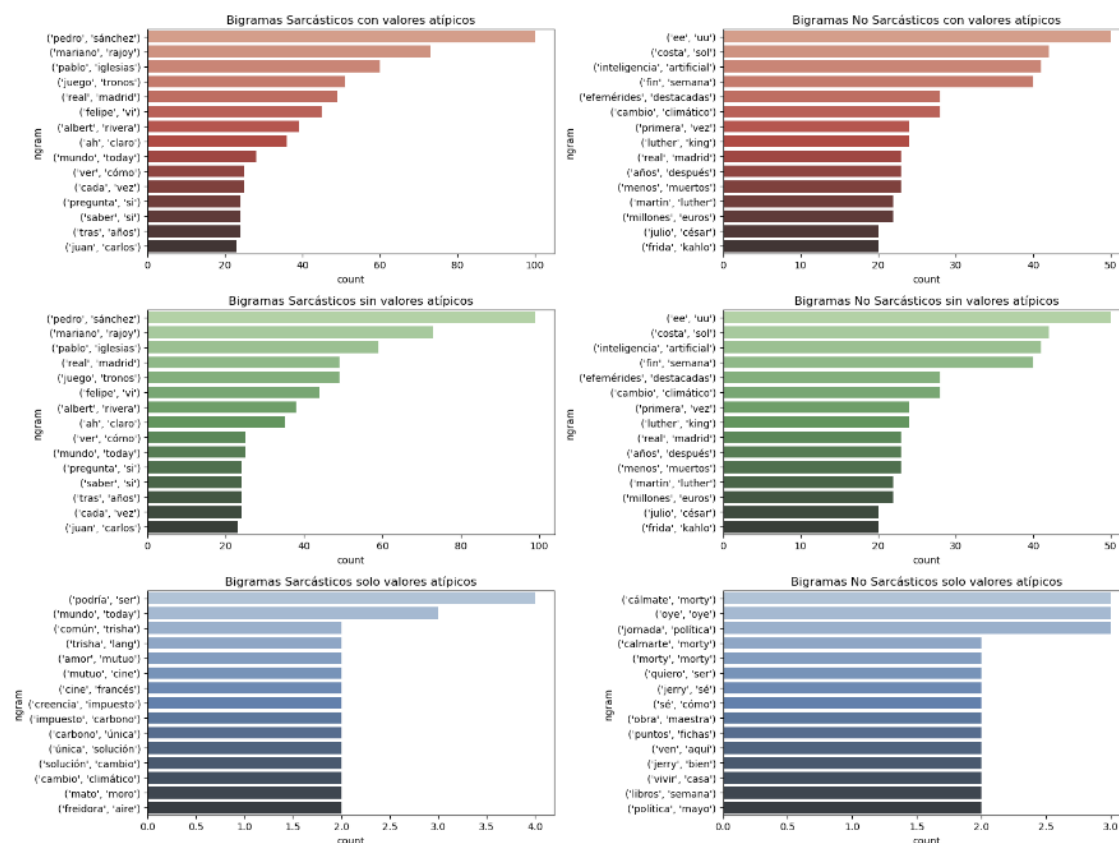


Ilustración 37: Frecuencia de bigramas sin stop words en textos sarcásticos y no sarcásticos con y sin valores atípicos.

Fuente: Elaboración propia.

En los gráficos de bigramas sin stopwords, observamos una mayor relevancia de nombres propios y combinaciones más específicas. En los textos sarcásticos, predominan términos como "Pedro Sánchez", "Mariano Rajoy", y referencias a figuras políticas y del entretenimiento, lo que sugiere un enfoque sarcástico hacia temas políticos y culturales. En los textos no sarcásticos,

destacan combinaciones más técnicas y descriptivas como "inteligencia artificial", "efemérides destacadas", y "costa del sol", lo que refuerza el carácter informativo y descriptivo de estos textos, que ya habíamos visto en análisis previos.

En comparación con los bigramas con stopwords, los bigramas sin stopwords resultan más útiles, ya que eliminan las combinaciones comunes de preposiciones y artículos, y en su lugar destacan nombres y términos específicos que aportan mayor valor semántico. En el caso de los textos sarcásticos, por ejemplo, se observan más referencias a personas públicas, mientras que en los no sarcásticos predominan términos más relacionados con lugares y conceptos.

Al igual que en los análisis anteriores, no se observa una diferencia significativa entre los textos con y sin valores atípicos. Los bigramas dominantes son prácticamente los mismos en ambos casos, lo que sugiere que la presencia de valores atípicos no afecta de manera considerable los patrones generales del uso del lenguaje en estos textos.

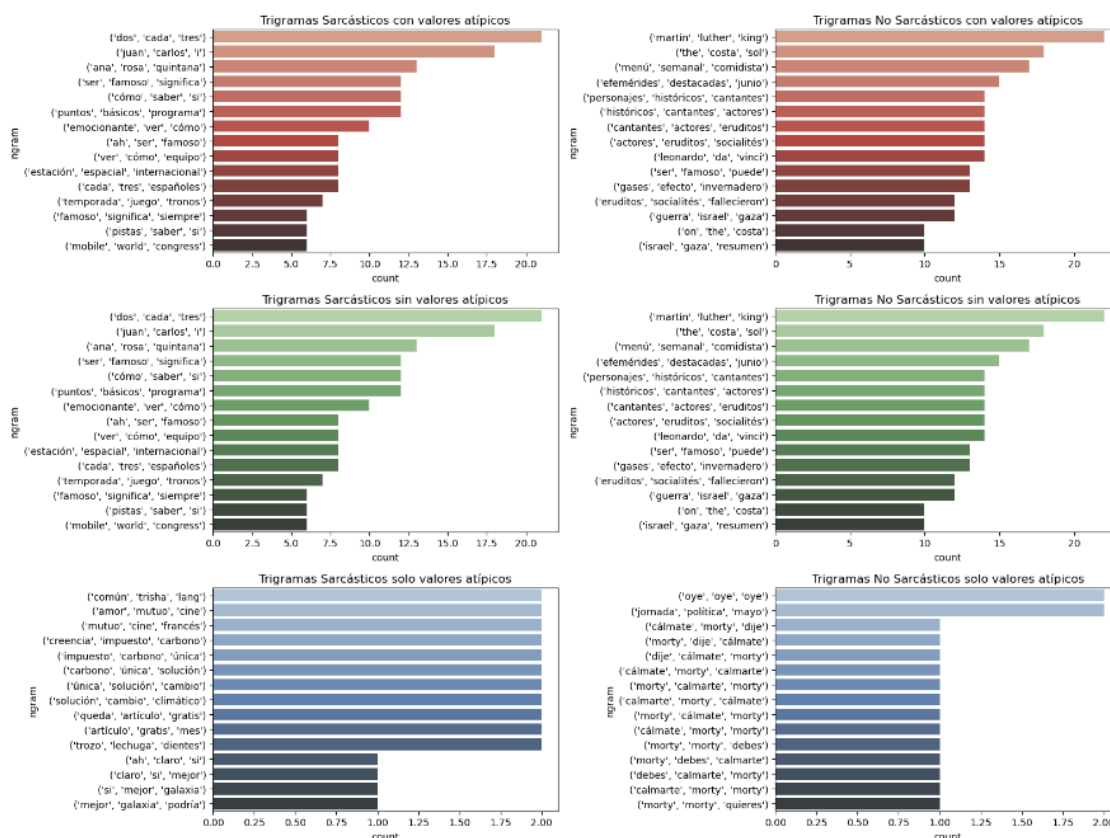


Ilustración 38: Frecuencia de trigramas sin stop words en textos sarcásticos y no sarcásticos con y sin valores atípicos.

Fuente: Elaboración propia.

En los gráficos de trigramas sin stopwords, observamos una mayor especificidad en los textos sarcásticos y no sarcásticos. En los textos sarcásticos, destacan combinaciones como "dos cada tres", "Juan Carlos I", y "ser famoso significa", lo que indica referencias irónicas o críticas hacia figuras públicas y situaciones comunes. Por otro lado, en los textos no sarcásticos, vemos trigramas más relacionados con descripciones históricas y culturales, como "Martin Luther

King", "menú semanal comidista", y "guerra Israel Gaza", mostrando un carácter más informativo y objetivo.

En comparación con los trigramas con stopwords, este análisis resulta más relevante, ya que elimina las combinaciones triviales y nos ofrece combinaciones más representativas y útiles para interpretar el contenido. Por ejemplo, en los textos sarcásticos, se observa un uso más enfático de personajes y situaciones, mientras que, en los no sarcásticos, predominan temas históricos y geográficos.

Al igual que en los análisis anteriores, no se observan grandes diferencias entre los textos con y sin valores atípicos. Los trigramas principales permanecen consistentes, lo que indica que los valores atípicos no alteran significativamente los patrones de lenguaje en los textos sarcásticos y no sarcásticos.

4.6.4.6 *Análisis de sentimiento*

En este apartado, llevaremos a cabo un análisis de sentimiento para los textos, utilizando una escala del 1 al 5, donde 1 representa un sentimiento negativo y 5 un sentimiento positivo. Para este análisis, hemos empleado el modelo BERT multilingüe sin restricciones de Hugging Face, que permite evaluar los sentimientos presentes en los textos en español. A continuación, aplicaremos este modelo para identificar y clasificar el sentimiento en cada texto y observar las tendencias en los datos.

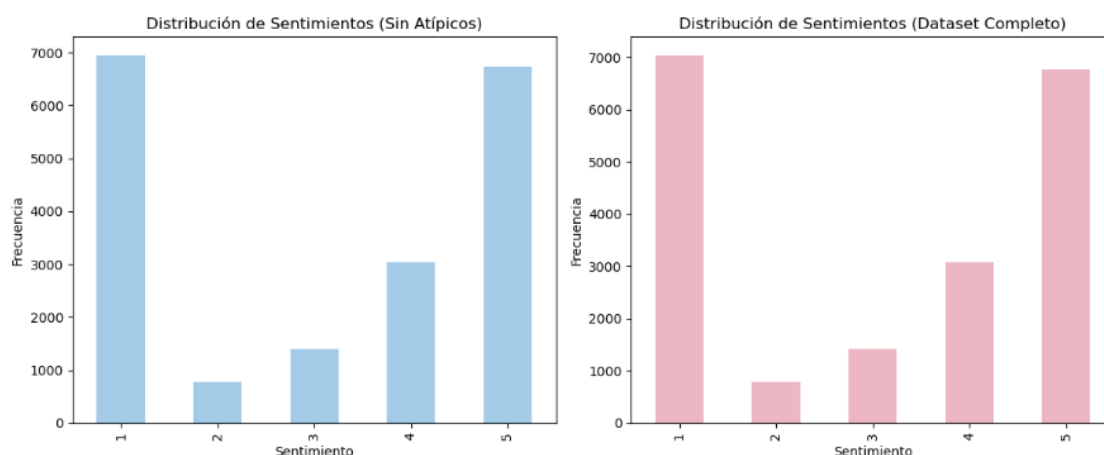


Ilustración 39: Distribución de sentimientos en textos con y sin valores atípicos.

Fuente: Elaboración propia.

En los gráficos se puede observar la distribución de sentimientos tanto en el conjunto de datos completo como en el conjunto de datos sin valores atípicos. En ambos casos, la mayoría de los textos están clasificados con un sentimiento extremadamente positivo (5) o extremadamente negativo (1), con picos evidentes en estas categorías. Las categorías intermedias (2, 3 y 4) presentan una menor frecuencia, lo que sugiere que los textos tienden a ser polarizados en cuanto a su carga emocional. No se observa una diferencia significativa entre los datos con y sin valores atípicos, ya que las distribuciones siguen patrones similares en ambos gráficos.

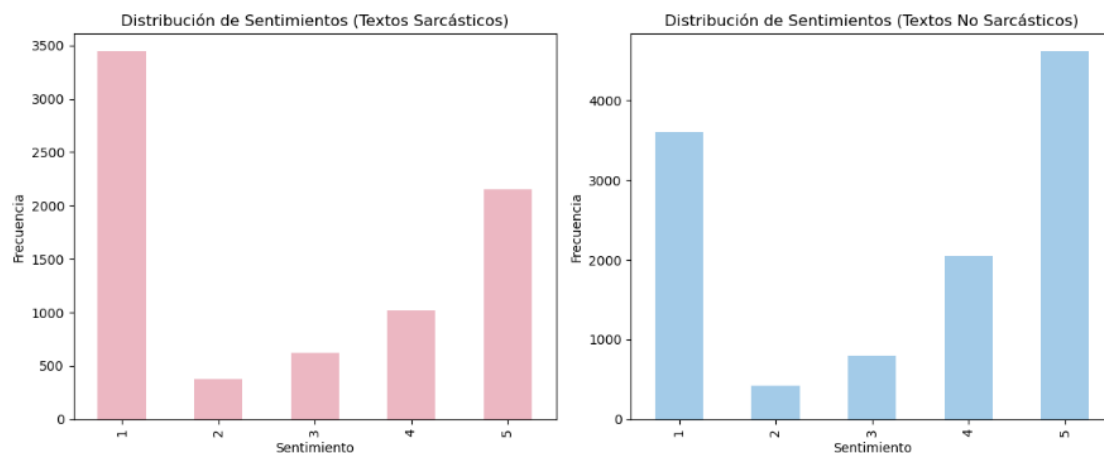


Ilustración 40: Distribución de sentimientos en textos sarcásticos y no sarcásticos.

Fuente: Elaboración propia.

En los gráficos se compara la distribución de sentimientos entre los textos sarcásticos y no sarcásticos. En los textos sarcásticos, se aprecia una fuerte tendencia hacia el sentimiento negativo (1), lo que refleja el tono característico del sarcasmo. Sin embargo, también hay una notable presencia de textos clasificados con sentimiento positivo (5). Por otro lado, en los textos no sarcásticos, la distribución es más equilibrada entre los sentimientos negativos (1) y positivos (5), aunque predomina el sentimiento positivo. Esto sugiere que, aunque el sarcasmo tiende a ser percibido como negativo, también puede transmitir ironía de manera positiva.

4.6.4.7 Análisis de correlación

A continuación, se presenta el Análisis de correlación. En este apartado, evaluaremos las relaciones entre las variables clave del estudio, como la longitud del texto, el sarcasmo y el sentimiento, para determinar si existe alguna correlación significativa entre ellas. Este análisis nos permitirá comprender mejor cómo se vinculan estas características en el conjunto de datos.

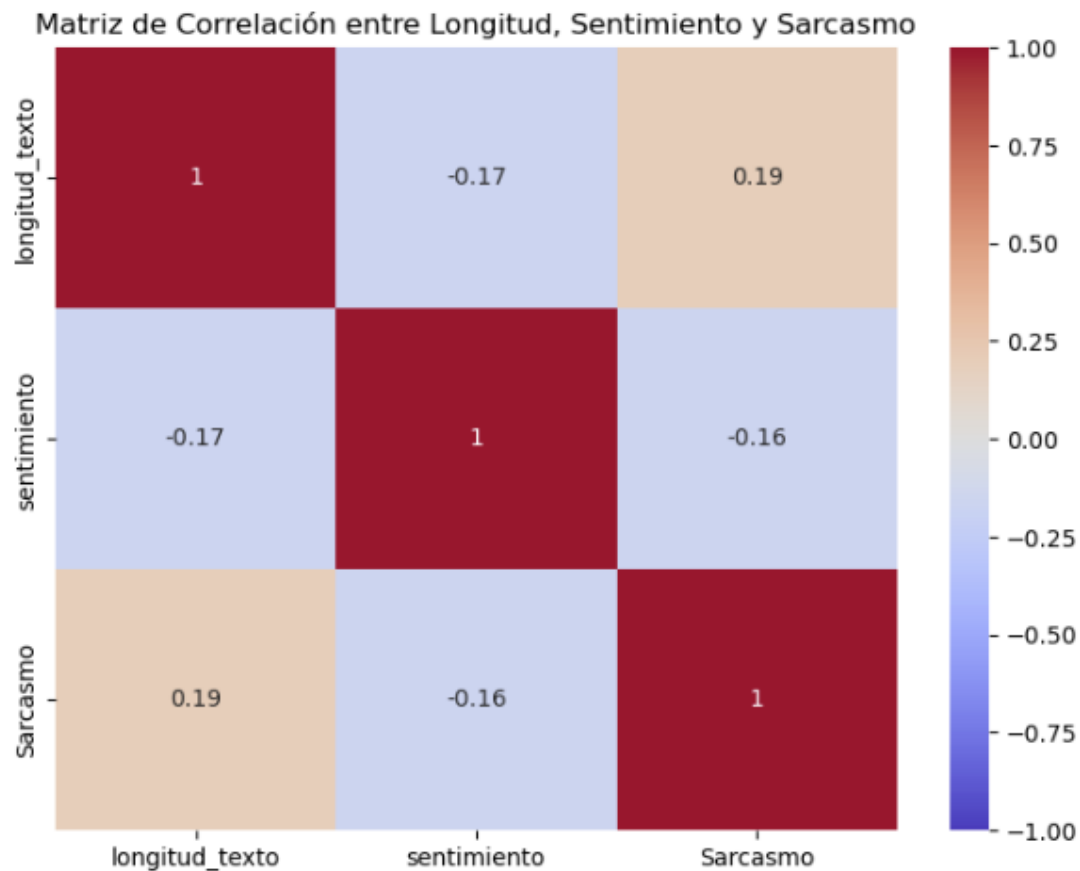


Ilustración 41: Matriz de correlación entre longitud, sentimiento y sarcasmo (con valores atípicos).

Fuente: Elaboración propia.

En la matriz de correlación presentada, se observa la relación entre tres variables clave: la longitud del texto, el sentimiento y el sarcasmo. La correlación más notable es la positiva entre la longitud del texto y el sarcasmo (0.19), lo que sugiere que los textos sarcásticos tienden a ser más largos. Por otro lado, se observa una correlación negativa débil entre el sarcasmo y el sentimiento (-0.16), lo que indica que los textos sarcásticos tienden a estar más asociados con sentimientos negativos. Además, la longitud del texto también muestra una ligera correlación negativa con el sentimiento (-0.17), aunque ninguna de estas correlaciones es muy fuerte.

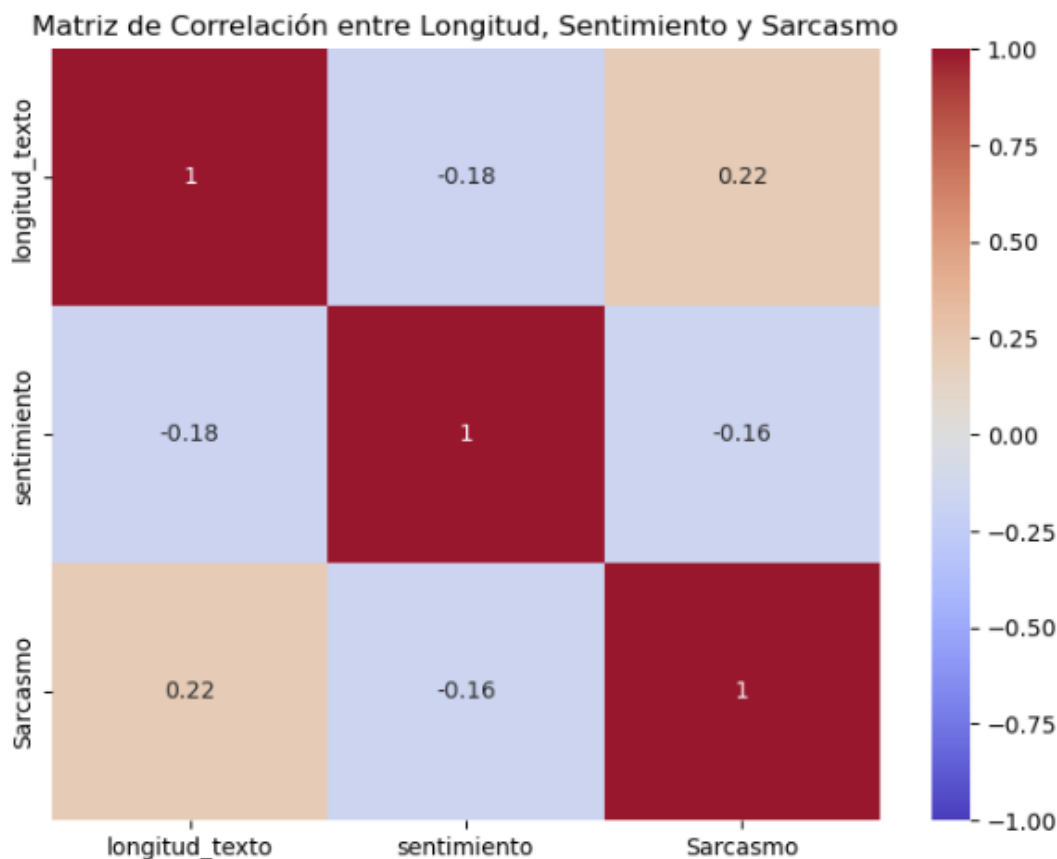


Ilustración 42: Matriz de correlación entre longitud, sentimiento y sarcasmo (sin valores atípicos).

Fuente: Elaboración propia.

En esta matriz de correlación sin valores atípicos, se observa un comportamiento muy similar al de la versión con atípicos. La correlación entre la longitud del texto y el sarcasmo es ligeramente mayor (0.22), lo que sigue indicando que los textos sarcásticos tienden a ser más largos. La correlación negativa entre el sarcasmo y el sentimiento se mantiene (-0.16), mostrando que los textos sarcásticos suelen estar asociados a sentimientos más negativos. La relación entre la longitud del texto y el sentimiento también permanece negativa (-0.18), aunque la magnitud es pequeña, lo que indica que, en general, la longitud del texto tiene un impacto limitado en el sentimiento expresado.

4.6.4.8 Análisis de variabilidad Lexica

En este apartado se abordará el análisis de la variabilidad léxica con el objetivo de medir la diversidad de vocabulario en los textos analizados. Este análisis es útil para determinar si los textos sarcásticos tienden a usar un vocabulario más variado o si, por el contrario, los textos no sarcásticos presentan una mayor riqueza léxica. Además, se evaluará si los valores atípicos afectan a esta variabilidad y cómo se distribuyen las palabras a lo largo de los diferentes textos en función de su longitud y sarcasmo.

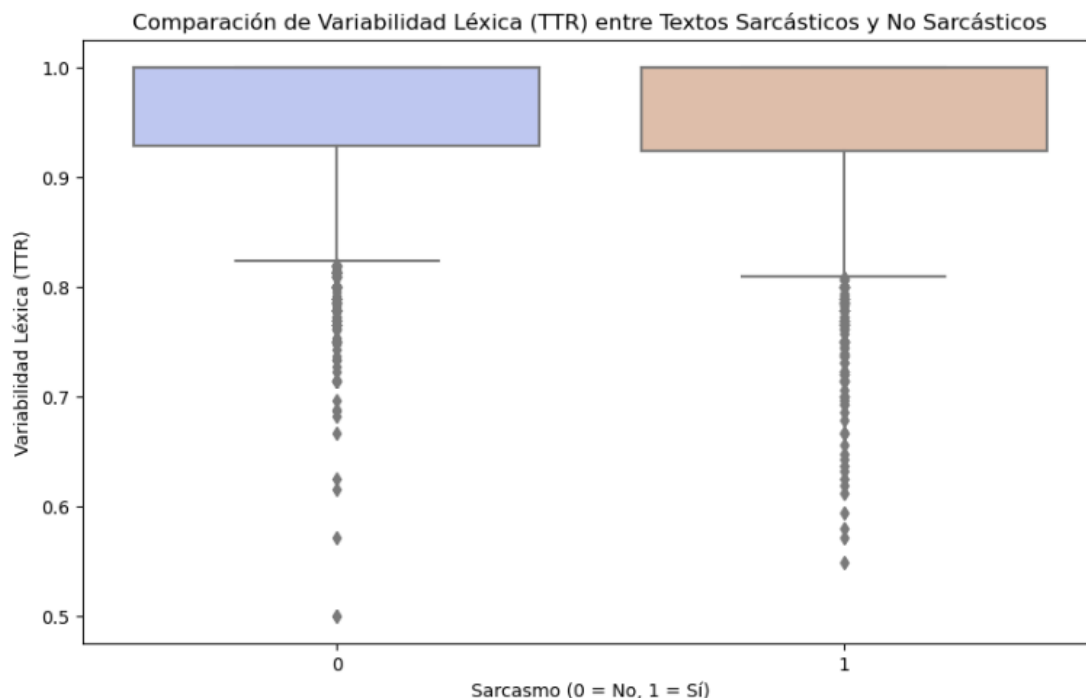


Ilustración 43: Comparación de variabilidad léxica (TTR) entre textos sarcásticos y no sarcásticos.

Fuente: Elaboración propia.

En la imagen se puede observar la comparación de la variabilidad léxica (medida como la Tasa de Tipo-Tokens, TTR) entre textos sarcásticos y no sarcásticos. Los valores más altos de TTR indican una mayor diversidad léxica en los textos, mientras que los valores más bajos sugieren un vocabulario menos variado.

El gráfico muestra que ambos grupos (sarcasmo y no sarcasmo) tienen una variabilidad léxica relativamente alta, cercana a 1, lo que indica que en general utilizan una amplia gama de vocabulario. Sin embargo, se observa que los textos sarcásticos tienden a tener una menor variabilidad léxica en algunos casos, como se evidencia por los puntos de valores atípicos que caen en niveles más bajos de TTR en comparación con los textos no sarcásticos.

4.6.5 Integración del modelo

En esta sección se presentan los resultados obtenidos durante la implementación, optimización y validación del modelo de detección de sarcasmo. El proceso se llevó a cabo en varias fases, comenzando con la implementación de modelos convencionales de aprendizaje automático como referencia comparativa. Posteriormente, se introdujo el modelo Transformer, basado en RoBERTa, que fue ajustado mediante fine-tuning para adaptarse a los datos de sarcasmo en español. Además, se optimizaron los hiperparámetros con el objetivo de maximizar el rendimiento del modelo. Finalmente, se realizaron pruebas exhaustivas con datos no vistos previamente para evaluar su capacidad de generalización.

4.6.5.1 Modelos basados en enfoques tradicionales

Para evaluar el rendimiento de modelos tradicionales en la tarea de detección de sarcasmos, se probaron enfoques basados en reglas, una red neuronal recurrente (RNN), y una red de memoria a largo plazo (LSTM). A continuación, se detallan los resultados de cada uno de estos enfoques:

4.6.5.1.1 Modelo basado en reglas

El primer enfoque implementado fue un modelo basado en reglas. Sus resultados fueron:

Accuracy	Precision	Recall	F1-Score
0.72185	0.72185	0.01432	0.02807

Tabla 2. Resultados modelo basado en reglas

4.6.5.1.2 Modelo basado en RNN

El siguiente modelo a implementar fue el modelo basado en redes neuronales recurrentes. Sus resultados del entrenamiento fueron:

Epoch	Training Loss	Accuracy	Validation Loss	Validation Accuracy
1	0.5530	0.7176	0.4779	0.7749
2	0.4308	0.8019	0.4098	0.7738
3	0.4628	0.7860	0.5291	0.7453
4	0.4618	0.7875	0.4454	0.7990
5	0.4483	0.7844	0.7353	0.7353

Tabla 3. Entrenamiento modelo basado en RNN

Los resultados de la evaluación fueron:

Training Loss	Accuracy	Precision	Recall	F1-Score
0.54488	0.73534	0.61202	0.88925	0.72505

Tabla 4. Resultados modelo basado en RNN

4.6.5.1.3 Modelo basado en LSTM

Para finalizar con los modelos con enfoques más tradicionales se optó por realizar un modelo LSTM. Sus resultados fueron:

Epoch	Training Loss	Accuracy	Validation Loss	Validation Accuracy
-------	---------------	----------	-----------------	---------------------

1	0.5357	0.7244	0.5174	0.7476
2	0.3981	0.8226	0.4324	0.8092
3	0.3330	0.8535	0.3889	0.8230
4	0.2735	0.8858	0.3897	0.8364
5	0.2424	0.9000	0.3712	0.8497

Tabla 5. Entrenamiento modelo basado en LSTM

Training Loss	Accuracy	Precision	Recall	F1-Score
0.23579	0.84973	0.794024	0.83322	0.81315

Tabla 6. Resultados modelo basado en LSTM

4.6.5.2 Modelo basado en Transformers

En este apartado, se presentarán los resultados obtenidos para el modelo basado en Transformers, utilizando el modelo preentrenado RoBERTa-base-bne adaptado para la detección de sarcasmo en español. Tras realizar varias épocas de entrenamiento, ajustando parámetros clave como la tasa de aprendizaje y el tamaño de lote, se ha logrado obtener un modelo optimizado para la tarea de clasificación de sarcasmo. A continuación, se presentan los resultados obtenidos en términos de precisión, recall, F1-score y la pérdida en el conjunto de validación.

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1-Score
1	0.2024	0.3138	0.8861	0.7996	0.9472	0.8671
2	0.0843	0.3318	0.9212	0.9341	0.8599	0.8954
3	0.0780	0.3915	0.9272	0.9040	0.9113	0.9076
4	0.0105	0.4064	0.9274	0.9208	0.8919	0.9061

Tabla 7. Entrenamiento modelo sin optimizar basado en Transformers

4.6.5.3 Búsqueda de hiperparámetros

A continuación, se presentan los resultados obtenidos tras la búsqueda de hiperparámetros óptimos para el modelo RoBERTa-base-bne. En primer lugar, se expondrán los resultados con los mejores hiperparámetros durante el proceso de búsqueda.

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1-Score
1	0.21260	0.19467	0.92921	0.90429	0.91700	0.91060

2	0.05260	0.24407	0.93578	0.94138	0.89224	0.91615
---	---------	---------	---------	---------	---------	---------

Tabla 8. Mejores resultados obtenidos en la búsqueda de hiperparámetros

En la siguiente tabla se expondrán los parámetros óptimos obtenidos durante la búsqueda de hiperparametros:

Learning Rate	Batch Size	Epochs	Weight Decay
2.949858019838e-05	32	2	0.21863411538200875

Tabla 9. Mejores hiperparámetros obtenidos para optimizar modelo Transformer

4.6.6 Interfaz grafica

En las pruebas realizadas a la interfaz gráfica con frases tanto sarcásticas como no sarcásticas, la interfaz demostró ser eficiente y precisa, proporcionando resultados en tiempo real. Esta herramienta interactiva permite realizar múltiples pruebas sin necesidad de conocimientos técnicos, y ha sido clave en la evaluación final del modelo de detección de sarcasmo.

En la siguiente imagen se muestra un ejemplo de la interfaz donde se introdujo la frase "Qué maravilloso, más trabajo en mi día libre", a la que el modelo asignó una probabilidad del 99,8% de sarcasmo. La barra de progreso visualiza esta confianza, lo que facilita la interpretación de los resultados.

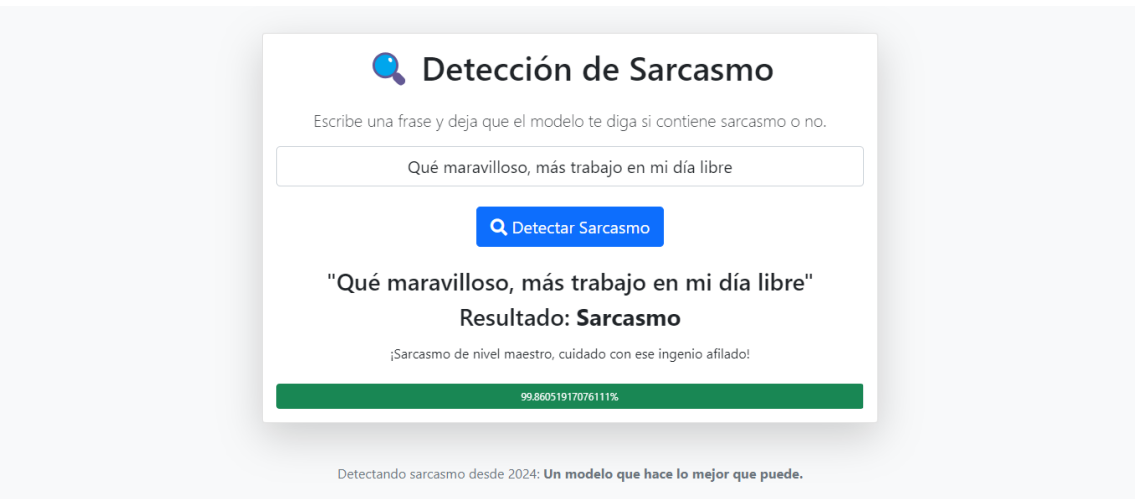


Ilustración 44: Interfaz gráfica para la detección de sarcasmo.

Fuente: Elaboración propia.

4.6.7 Pruebas del modelo optimizado

En la siguiente tabla se presentan los resultados obtenidos al realizar una serie de pruebas sobre el modelo Transformer entrenado para la detección de sarcasmo en español:

Frase	Sarcasmo detectado	Sarcasmo real
Cinco trucos para «okupar» la vivienda de Manu Tenorio	Sarcasmo	Sarcasmo
El principal miedo de los españoles ya es morir sin borrar el historial de navegación y que sus seres queridos vean que han visitado OK Diario con regularidad	Sarcasmo	Sarcasmo
¿Qué es la “solidaridad”, el arma secreta del Gobierno para solventar la crisis de los alquileres?	No sarcasmo	Sarcasmo
Después de 30 años de carrera, la Mala Rodríguez consigue ascender a Regular Rodríguez	Sarcasmo	Sarcasmo
El PP se abre a negociar la jornada laboral de 4 días siempre que el trabajador acuda a trabajar uno de sus días libres	Sarcasmo	Sarcasmo
Salud inicia este miércoles la vacunación de gripe y covid con los mayores de 80 años	No sarcasmo	No sarcasmo
La Moncloa encarga a Cuerpo abrir hueco para Bildu o PNV en el Banco de España	Sarcasmo	No sarcasmo
Dos diputados del PP y uno de Vox ofrecen su dimisión por	No sarcasmo	No sarcasmo

las rebajas de penas a etarras, pero sus partidos las rechazan		
La Asamblea Nacional de Venezuela pide romper relaciones con España y abolir la monarquía	No sarcasmo	No sarcasmo
La Asamblea Nacional de Venezuela pide romper relaciones con España y abolir la monarquía	No sarcasmo	No sarcasmo

Tabla 10. Resultados obtenidos en la fase de pruebas

Capítulo 5. DISCUSIÓN

La discusión en este capítulo se centrará en analizar los resultados obtenidos a lo largo del proyecto, abordando tanto los hallazgos derivados del análisis exploratorio de los datos (EDA) como el rendimiento de los modelos implementados, desde los enfoques tradicionales hasta el modelo basado en Transformers. Se expondrán las diferencias observadas entre los distintos modelos y se evaluará su capacidad para detectar sarcasmo en español, destacando las ventajas del uso de Transformers en comparación con enfoques más tradicionales como las redes neuronales recurrentes (RNN) o los modelos basados en reglas.

Además, se discutirán las limitaciones encontradas durante el desarrollo del proyecto, tanto a nivel de datos como en el aspecto tecnológico. Finalmente, se hará un análisis sobre la metodología utilizada y sobre los cambios que fue necesario realizar durante el transcurso del proyecto.

5.1 Discusión de Resultados del EDA

En esta sección, se discutirán los resultados obtenidos del análisis exploratorio de datos, evaluando cómo estos han permitido extraer conclusiones iniciales sobre la estructura y características del conjunto de datos. El EDA fue clave para comprender mejor los patrones en los textos sarcásticos y no sarcásticos, para identificar posibles valores atípicos y sesgos que podrían afectar el rendimiento de los modelos.

El análisis de la longitud de los textos reveló que la mayoría de los titulares y frases tenían entre 50 y 200 caracteres, con algunos valores atípicos, que podían alcanzar más de 1.200 caracteres. Estos valores atípicos, observados en el gráfico de distribución de la longitud, indicaban que algunos textos podían influir de manera significativa. La eliminación de los valores atípicos no mostró un cambio drástico en los patrones de sarcasmo, lo que sugiere que el sarcasmo no depende directamente de la longitud del texto, sino más bien del contenido y contexto lingüístico.

El análisis de palabras frecuentes, de bigramas y trigramas no mostró patrones sorprendentes en cuanto a secuencias de palabras que puedan indicar sarcasmo de forma clara. Esto reafirma que el sarcasmo es una característica más contextual y semántica, difícil de capturar solo mediante análisis de frecuencia de palabras o secuencias, lo que valida la necesidad de utilizar modelos más sofisticados como los basados en Transformers para capturar este fenómeno de manera efectiva.

Por otro lado, el análisis de sentimiento realizado expuso que los textos sarcásticos tendían a tener puntuaciones más extremas de sentimiento, tanto en positivo como en negativo. Sin embargo, los resultados no revelaron una relación de los sentimientos y la presencia de sarcasmo. Esto indica que el sarcasmo puede expresarse tanto en textos con sentimientos positivos como negativos y que el análisis de sentimiento por sí solo no es suficiente para distinguir de manera efectiva el sarcasmo.

Asimismo, el análisis de correlación entre la longitud del texto, el sentimiento y la presencia de sarcasmo tampoco mostró correlaciones significativas. Esto expone que no existe una relación lineal clara entre estas variables.

En conclusión, el EDA permitió comprender las características principales del dataset, como la distribución de la longitud de los textos y las palabras más frecuentes en los textos sarcásticos y no sarcásticos. Sin embargo, no se lograron identificar diferencias claras entre ambos grupos que permitieran definir patrones específicos para el sarcasmo, lo que sugiere que el sarcasmo es un fenómeno más complejo que no puede ser capturado únicamente a través del análisis superficial de los datos. Este resultado resalta la necesidad de utilizar modelos avanzados, como los basados en Transformers, para abordar la tarea de detección de sarcasmo de manera más efectiva.

5.2 Discusión de los modelos implementados

En este apartado, se discutirá sobre los diferentes modelos implementados para la detección de sarcasmo. Los enfoques tradicionales utilizados se comparan con el modelo Transformer, que ha demostrado ser más eficaz en la detección de los sarcasmos. A través de varias métricas, como precisión, recall, F1-score y accuracy, se analizará el rendimiento de cada uno de estos modelos para observar cuál ha sido más eficiente y eficaz.

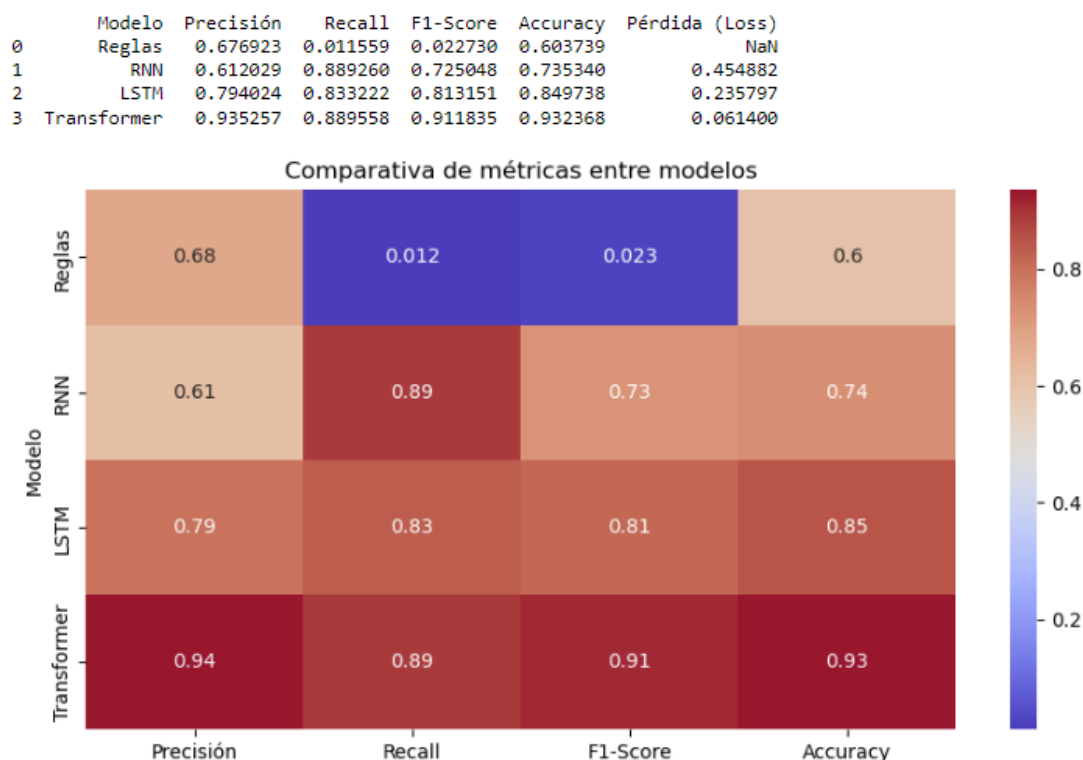


Ilustración 45: Comparativa de métricas entre modelos para la detección de sarcasmo.

Fuente: Elaboración propia.

5.2.1 Modelo basado en reglas

El modelo basado en reglas presentó los peores resultados entre todos los enfoques implementados. A pesar de que obtuvo una precisión del 67.69%, su capacidad para identificar correctamente el sarcasmo fue extremadamente baja, con un recall del 1.16% y un F1-score de solo 2.23%. Este resultado es bastante esperable, ya que el sarcasmo es una forma de comunicación compleja que difícilmente puede ser capturada a través de reglas fijas. Este modelo simplemente no es capaz de capturar el contexto ni las sutilezas necesarias para identificar el sarcasmo en español.

5.2.2 Modelo basado en RNN

El modelo basado en RNN mejoró significativamente con respecto al enfoque basado en reglas, logrando una precisión del 61.02% y un recall del 88.92%. Sin embargo, su F1-score de 72.50% y una accuracy de 73.53% indican que aún tiene problemas para generalizar en ciertos casos, lo que puede estar relacionado con la dificultad de capturar dependencias a largo plazo en los textos, que es crucial para entender el sarcasmo en contextos más complejos.

5.2.3 Modelo basado en LSTM

El modelo basado en LSTM se mostró una mejoría, con una precisión del 79.40%, un recall del 83.32% y un F1-score de 81.31%. El modelo LSTM, como era de esperar, obtiene una mejora clara respecto a los enfoques anteriores. La inclusión de mecanismos de memoria permite que este modelo sea más efectivo a la hora de identificar patrones complejos y contextos largos en los textos sarcásticos.

5.2.4 Modelo basado en Transformers

Finalmente, el modelo Transformer, basado en RoBERTa y ajustado mediante fine-tuning, ha demostrado ser el más efectivo en la detección de sarcasmo. Con una precisión del 93.52%, un recall de 88.95% y un F1-score de 91.18%, este modelo supera claramente a los enfoques tradicionales.

5.3 Discusión de las pruebas obtenidas

Después de implementar y optimizar el modelo Transformer, se hicieron varias pruebas con datos nuevos que no habían sido vistos antes. El objetivo de estas pruebas era ver qué tan bien el modelo podía detectar sarcasmo en frases y contextos diferentes a los utilizados en su entrenamiento.

En general, el modelo funcionó muy bien, identificando correctamente el sarcasmo en la mayoría de las frases. Por ejemplo, en frases como "Cinco trucos para okupar la vivienda de Manu Tenorio" o "El PP se abre a negociar la jornada laboral de 4 días," donde el sarcasmo es bastante obvio, el modelo lo detectó con precisión.

A pesar de los buenos resultados, también hubo algunos fallos. En la frase "¿Qué es la solidaridad, el arma secreta del Gobierno para solventar la crisis de los alquileres?", el modelo no detectó el sarcasmo, a pesar del tono irónico. Esto sugiere que todavía tiene dificultades para

captar sarcasmos más sutiles o contextuales, lo que podría requerir más ajustes o información adicional. En resumen, las pruebas del modelo Transformer han demostrado que puede identificar sarcasmo en la mayoría de las situaciones, especialmente en frases con un tono humorístico o exagerado. Sin embargo, todavía tiene algunas dificultades para captar sarcasmos más sutiles o en contextos más serios, lo que deja margen para futuras mejoras. Estos resultados coinciden con lo que se ha visto en otros estudios, donde los Transformers son muy eficaces, pero no perfectos, a la hora de entender el lenguaje complejo como el sarcasmo.

5.3.1 Limitaciones

Entre las principales limitaciones encontradas en este proyecto, una de las más destacadas fue el dataset utilizado. Aunque se recopiló una cantidad considerable de datos, incluyendo frases de distintos orígenes como noticias, diálogos y frases generadas por IA, la variedad del sarcasmo no siempre fue suficiente para capturar todas las sutilezas del lenguaje. El sarcasmo, al ser una forma de expresión muy dependiente del contexto, sigue siendo un desafío en muchos casos, especialmente en situaciones donde el tono es más sutil o está implícito.

Otro aspecto a considerar es que, a pesar del buen rendimiento de los modelos basados en Transformers, estos requieren una gran cantidad de recursos computacionales. Esto no solo alarga los tiempos de entrenamiento, sino que también puede ser un obstáculo en entornos con recursos limitados. Además, el proceso de búsqueda de hiperparámetros, aunque esencial para la optimización del modelo, también implica un costo computacional significativo, en nuestro caso de mínimo 35 minutos por ejecución, además de los recursos de pago en Google Colab.

Por último, una limitación importante fue la capacidad del modelo para generalizar a contextos muy diferentes a los utilizados en el entrenamiento. Aunque se realizaron pruebas con nuevos datos, ciertos tipos de sarcasmo más sutiles o con contextos más complejos no fueron detectados de manera óptima. Esto sugiere que, para mejorar el rendimiento, sería necesario continuar trabajando en la diversificación del dataset y en el refinamiento de los modelos.

Capítulo 6. CONCLUSIONES

En este capítulo sobre las conclusiones, se resumirá los principales hallazgos y resultados obtenidos durante el desarrollo del proyecto. Asimismo, se expondrá sobre cómo cada uno de los objetivos planteados al inicio fue abordado y alcanzado, destacando los logros conseguidos, las lecciones aprendidas y las posibles áreas de mejora. Además, se incluirá unas conclusiones personales para comentar la experiencia vivida a lo largo del proyecto y cómo este proceso ha contribuido tanto en el ámbito académico como en el personal.

6.1 Conclusiones del trabajo

El objetivo general de este proyecto se basó en desarrollar un sistema capaz de detectar sarcasmos en español utilizando un modelo Transformers. Después de un largo y complejo trabajo se consiguió realizar esta tarea con éxito. El modelo final, basado en RoBERTa y ajustado con fine-tuning, ha mostrado un rendimiento adecuado para la detección de sarcasmos, la cual es una tarea compleja incluso para un ser humano promedio.

En cuanto a los objetivos específicos, vamos a exponerlos uno a uno a continuación:

Uno de los primeros objetivos específicos era realizar una investigación exhaustiva sobre el estado del arte en la detección de sarcasmos. Este objetivo se cumplió ampliamente, ya que se analizaron y compararon distintos modelos y enfoques aplicados al procesamiento de lenguaje natural (PLN), con especial énfasis en los Transformers. Esto no solo permitió tener un entendimiento más profundo de las fortalezas y debilidades de los modelos, sino que también ayudó a definir una metodología clara para el desarrollo del proyecto.

En cuanto a la recopilación de datos, se logró obtener un dataset decente compuesto por noticias, diálogos y frases generadas por inteligencia artificial, todos enfocados en el sarcasmo en español. Este conjunto de datos fue clave para entrenar el modelo y obtener resultados fiables.

El proceso de ETL (Extracción, Transformación y Carga) también fue exitoso. Los datos recopilados fueron limpiados, organizados y preparados adecuadamente para su análisis y posterior uso en el modelo. Esto permitió asegurar la calidad de los datos utilizados, eliminando valores atípicos o irrelevantes que podrían haber afectado el rendimiento del modelo.

El análisis exploratorio de datos (EDA) proporcionó una visión clara de las características principales del dataset, como la longitud de los textos y la distribución de palabras frecuentes. Aunque el EDA no mostró diferencias evidentes entre los textos sarcásticos y no sarcásticos, fue crucial para entender mejor el comportamiento general del dataset y detectar posibles patrones.

Se implementaron varios modelos de machine learning, incluyendo enfoques tradicionales como las redes LSTM y RNN, que sirvieron como puntos de comparación. Si bien estos modelos ofrecieron resultados decentes, fue el modelo basado en Transformers el que destacó por su

capacidad de capturar el sarcasmo con mayor precisión, obteniendo métricas muy favorables tras su entrenamiento y optimización.

Finalmente, se realizaron varias pruebas y validaciones del modelo Transformer, las cuales confirmaron que era capaz de identificar sarcasmo con alta precisión en una amplia variedad de contextos. Aunque aún existen desafíos en la detección de sarcasmos más sutiles, el resultado general del proyecto ha sido muy positivo, cumpliendo con las expectativas y demostrando la eficacia de los Transformers en este tipo de tareas.

6.2 Conclusiones personales

Este proyecto ha sido una experiencia clave tanto a nivel personal como profesional. Trabajar con Transformers, que hoy en día son la tecnología más avanzada en todo lo relacionado con el procesamiento de lenguaje natural, ha sido un reto emocionante y a la vez una oportunidad de aprendizaje invaluable. Desde que empecé a investigar sobre el estado del arte, tenía claro que los Transformers no solo están revolucionando el campo del PLN, sino que son la base de muchísimas aplicaciones innovadoras que vemos cada día. Poder entender y aplicar esta tecnología en un proyecto propio ha sido una experiencia profundamente enriquecedora.

A nivel personal, uno de los mayores logros ha sido ver cómo un proyecto que he desarrollado desde el principio, pasando por la investigación, la recopilación de datos, la implementación del modelo y la validación final, ha conseguido funcionar de manera efectiva en algo tan complejo como la detección de sarcasmos en español. Este viaje, desde la teoría hasta la implementación práctica, me ha permitido no solo aprender sobre los aspectos técnicos, sino también valorar la dedicación y el esfuerzo que implica completar un proyecto de esta magnitud.

En lo profesional, este proyecto me ha ayudado a afianzar mis conocimientos en el uso de Transformers y técnicas de fine-tuning, algo que resulta crucial en un campo en constante evolución como el procesamiento de lenguaje natural. El hecho de haber logrado un sistema funcional, que puede detectar sarcasmo en textos, refuerza mi confianza en las capacidades que he desarrollado a lo largo del proyecto y me impulsa a seguir explorando nuevos horizontes en esta área tan fascinante.

Ver los resultados de este trabajo, y cómo un modelo entrenado por mí puede interpretar algo tan sutil como el sarcasmo, ha sido un gran estímulo para continuar trabajando con estas tecnologías y llevar mis habilidades al siguiente nivel.

Capítulo 7. FUTURAS LÍNEAS DE TRABAJO

A pesar de los avances logrados con este proyecto, existen varias áreas en las que se puede seguir trabajando para mejorar y expandir los resultados obtenidos. A continuación, se describen algunas de las principales líneas de trabajo que podrían ser exploradas en el futuro.

Aunque el dataset utilizado en este proyecto ha sido útil para desarrollar y entrenar el modelo, una futura línea de trabajo sería crear un conjunto de datos más amplio y representativo. Esto incluiría la recopilación de más ejemplos de sarcasmo en la vida cotidiana, en lugar de enfocarse exclusivamente en titulares de noticias o diálogos de series de televisión. Recopilar datos de interacciones cotidianas, redes sociales y conversaciones informales permitiría al modelo ser más flexible y captar sarcasmos que se utilizan en contextos más variados, mejorando así su capacidad de generalización.

Hasta ahora, el análisis exploratorio se ha centrado en técnicas tradicionales, como la correlación lineal entre variables. Una posible extensión sería realizar un análisis de la correlación no lineal entre variables clave, lo que podría aportar nuevos hallazgos sobre cómo las diferentes características del texto interactúan para producir el sarcasmo. Esto podría incluir el uso de técnicas avanzadas de análisis de datos, como redes neuronales para la detección de relaciones complejas entre las características de los textos.

Una línea ambiciosa de trabajo futuro sería construir un modelo Transformer desde cero, en lugar de utilizar modelos preentrenados. El preentrenamiento se realizaría exclusivamente con datos relacionados con el sarcasmo, permitiendo que el modelo aprenda características específicas de esta forma de comunicación desde el inicio. Este enfoque no solo permitiría una personalización completa del modelo para esta tarea, sino que también podría revelar nuevas formas en las que el sarcasmo se manifiesta en el lenguaje escrito.

Uno de los pasos más emocionantes sería llevar este modelo a una aplicación real. Por ejemplo, el sarcasmo es una forma común de expresión en las interacciones humanas cotidianas, y una futura aplicación podría ser un chatbot que sea capaz de detectar y responder adecuadamente a comentarios sarcásticos. Esto mejoraría la interacción usuario-máquina, permitiendo que el chatbot entienda mejor el tono de las conversaciones, y respondiera de manera más natural y efectiva en situaciones donde el sarcasmo juega un papel clave.

Otra línea de trabajo futura sería la creación de un modelo generativo basado en Transformers para producir textos sarcásticos de manera automática. Este tipo de modelo podría entrenarse con ejemplos de sarcasmo en español, lo que permitiría generar respuestas automáticas con tono sarcástico en distintas situaciones. Esta herramienta podría ser utilizada en chatbots, asistentes virtuales o incluso en aplicaciones humorísticas. Desarrollar un modelo de generación de textos permitiría explorar no solo la detección, sino también la creación de contenidos que utilicen el sarcasmo de manera eficaz y controlada.

Capítulo 8. Bibliografía

- Alexandros, R., Siolas, G., & Stafylopatis, A.-G. (12 de Noviembre de 2019). *A Transformer-based approach to Irony and Sarcasm detection*. Obtenido de <https://ar5iv.labs.arxiv.org/html/1911.10401>
- Amazon Web Service. (s.f.). *Machine Learning: Amazon Web Service*. Obtenido de Amazon Web Service.
- Bergmann, D. (15 de Marzo de 2024). *What is fine-tuning?* Obtenido de IBM: <https://www.ibm.com/topics/fine-tuning>
- Chiu, K.-L., Collins, A., & Rohan, A. (24 de Marzo de 2022). *Detecting Hate Speech with GPT-3 **. Obtenido de <https://arxiv.org/pdf/2103.12407>
- Davidov, D., Tsur, O., & Rappoport, A. (2010). Semi-Supervised Recognition of Sarcasm in Twitter and Amazon., (págs. 107-116). Uppsala.
- Dong, X. (2024). *Transformer-based Context-aware Sarcasm Detection in Conversation Threads from Social Media*. Obtenido de <https://ar5iv.labs.arxiv.org/html/2005.11424>
- Du, Y., He, H., & Chu, Z. (20 de Febrero de 2024). Cross-cultural nuances in sarcasm comprehension: a comparative study of Chinese and American perspectives.
- Everyday Speech. (s.f.). *Understanding Sarcasm: A Beginner's Guide to Recognizing and Interpreting Sarcastic Remarks*. Obtenido de Everyday Speech: <https://everydayspeech.com/>
- Ferrer, J. (9 de Enero de 2024). *How Transformers Work: A Detailed Exploration of Transformer Architecture*. Obtenido de DataCamp: <https://www.datacamp.com/tutorial/how-transformers-work>
- Fu, H., Liu, H., Wang, H., Xu, L., Jianli, L., & Jiang, D. (2 de Enero de 2024). *Multi-Modal Sarcasm Detection with Sentiment Word Embedding*. Obtenido de <https://www.mdpi.com/2079-9292/13/5/855>
- Gabriel Gomila, J. (2024). Deep Learning de la A a la Z. Udey.
- Gelbukh, A. (2022). *Sarcasm Detection Framework Using Context, Emotion and Sentiment Features*. Obtenido de Arxiv: <https://ar5iv.labs.arxiv.org/html/2211.13014>
- Geva, M., Schuster, R., Berant, J., & Levy, O. (2021). Transformer Feed-Forward Layers Are Key-Value Memories.
- Ghosh, A., & Veale, T. (2017). *Magnets for Sarcasm: Making Sarcasm Detection Timely, Contextual and Very Personal*. Obtenido de <https://aclanthology.org/D17-1050.pdf>
- Gino, F. (17 de Noviembre de 2015). *The Surprising Benefits of Sarcasm*. Obtenido de SCI AM: <https://www.scientificamerican.com/article/the-surprising-benefits-of-sarcasm/>

- González-Ibáñez, R., Muresan, S., & Wacholder, N. (2011). Identifying Sarcasm in Twitter: A Closer Look. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*.
- Gradient Ascendt. (2023 de Diciembre de 2023). *The Five Stages of Natural Language Processing*. Obtenido de Gradient Ascendt: <https://gradient-ascent.com/natural-language-processing-stages/>
- Hurtado, D. (2024). *Implementación de un sistema de traducción automática voz a voz mediante el uso de transformers*. Obtenido de <https://repositorio.uniandes.edu.co/server/api/core/bitstreams/7fc0cc0a-a3a5-40bc-90e5-e37549c0f12b/content>
- IBM. (2024). *¿Qué son las redes neuronales recurrentes?* Obtenido de IBM: <https://www.ibm.com/es-es/topics/recurrent-neural-networks>
- IBM. (6 de Junio de 2024). *¿What is NLP?* Obtenido de IBM: <https://www.ibm.com/>
- IBM. (2024). *What are convolutional neural networks?* Obtenido de IBM: <https://www.ibm.com/topics/convolutional-neural-networks>
- IBM. (2024). *What is a neural network?* Obtenido de IBM: <https://www.ibm.com/topics/neural-networks>
- Joshi, A., Bhattacharyya, P., & J Carman, M. (2016). Automatic Sarcasm Detection: A Survey.
- Korvage Information Technology. (2024). *How Is Convolutional Neural Network (CNN) Used In NLP ?* Obtenido de LinkedIn: <https://www.linkedin.com/pulse/how-convolutional-neural-network-cnn-used-nlp-korvage-wppcc>
- Kumar, S. (2019). Sarcasm Detection in Textual Data: A Supervised Approach.
- La, F. (2 de Agosto de 2019). *A Closer Look at Neural Networks*. Obtenido de Microsoft: <https://learn.microsoft.com/en-us/archive/msdn-magazine/2019/february/artificially-intelligent-a-closer-look-at-neural-networks>
- Liebrecht, C., Kunneman, F., & van den Bosch, A. (2013). The perfect solution for detecting sarcasm in tweets #not. *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Atlanta.
- OpenAI. (2024). *ChatGPT*. Obtenido de ChatGPT: <https://chatgpt.com/>
- Pawar, S. (7 de Marzo de 2024). *Why Transformers are better than LSTM and RNN*. Obtenido de Medium: <https://medium.com/@shridharpawar77/why-transformers-are-better-than-lstm-and-rnn-6a9dbc51b178>
- Pelser, D., & Murrell, H. (2019). Deep and Dense Sarcasm Detection.
- Ray, A., Mishra, S., Nunna, A., & Bhattacharyya, P. (2022). A Multimodal Corpus for Emotion Recognition in Sarcasm.
- Real Academia Española. (2023). Obtenido de <https://dle.rae.es/>

- Satoquirá, M. (20 de Julio de 2019). *¿Qué son las Redes LSTM?* Obtenido de Codificacionbits: <https://www.codificandobits.com/blog/redes-lstm/>
- Shaalán, K. (2010). Rule-based Approach in Arabic Natural Language Processing.
- Sobti, S., Agarwal, D. V., Tiwary, D. A., Naval, P., Jayabalan, B., & Sohal, J. (2024). Sarcasm in the Digital Age: An Opinion. *Nanotechnology Perceptions*.
- Soren, D. (13 de Febrero de 2023). *Understanding Transformer model architectures*. Obtenido de Practical Artificial Intelligence: <https://www.practicalai.io/understanding-transformer-model-architectures/>
- Spain AI. (2021 de Agosto de 2021). *Sesion 3: Transformers I*. Obtenido de NLP de 0 a 100 con Hugin Face: <https://www.youtube.com/watch?v=tVLwC4XKGeU>
- Suhartono, D., Handoyo, A. T., & Junior, F. A. (2023). Feature-Based Augmentation in Sarcasm Detection Using Reverse Generative Adversarial Network.
- Teki, S., & Cicujano, M. (2024). *¿Cuales son las ventajas y desventajas de un modelo de PNL basado en reglas?* Obtenido de LinkedIn: <https://www.linkedin.com/advice/0/what-advantages-drawbacks-rule-based-nlp-48qdc#:~:text=A%20rule%2Dbased%20NLP%20model%20is%20a%20system%20that%20relies,linguistic%20knowledge%20and%20domain%20expertise.>
- Tummala, P. (Enero de 2024). *Exploring T5 and RGAN for Enhanced Sarcasm Generation in NLP*. Obtenido de https://www.researchgate.net/publication/381571780_Exploring_T5_RGAN_for_Enhanced_Sarcasm_Generation_in_NLP
- 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.
- Vlad, S., & Magdy, W. (2019). *Arxiv*. Obtenido de Cornell University: <https://arxiv.org/>
- Von platen, P. (10 de Octubre de 2020). *Transformers-based Encoder-Decoder Models*. Obtenido de Hugging Face: <https://huggingface.co/blog/encoder-decoder>
- Wikipedia. (2024). *Fine-tuning (deep learning)*. Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Fine-tuning_\(deep_learning\)](https://en.wikipedia.org/wiki/Fine-tuning_(deep_learning))
- Zampieri, M., Kong, Y., Ranasinghe, T., Farabi, S., & Kanojia, D. (2024). A Survey of Multimodal Sarcasm Detection. *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*.
- Zhou, J. (2023). An Evaluation of State-of-the-Art Large Language.

Capítulo 9. ANEXOS

Anexo I. Extracción datos El mundo Today

Anexo II. Extracción Datos El país

Anexo III. Extracción Datos El SUR

Anexo IV. Extracción de videos de youtube a audio

Anexo V. Transformación de audio a diálogo

Anexo VI. Extracción, transformación y carga

Anexo VII. Texto_procesado.csv

Anexo VIII. Análisis exploratorio de los datos

Anexo IX. Modelos de clasificación y su comparación

Anexo X. Detección de sarcasmos utilizando modelo de Transformers Roberta

Anexo XI. Interfaz Gráfica Clasificador de Sarcasmos

[PÁGINA INTENCIONADAMENTE EN BLANCO]