



Análisis de Sentimientos

Luis Alberto Soto

Diego Ros Pagan

CAPÍTULO 1

Introducción

El análisis de sentimientos en redes sociales resulta fundamental para captar percepciones públicas y obtener insights en tiempo real sobre temas, productos y eventos. Twitter, con su formato breve y su diversidad de usuarios, es una fuente ideal para esta minería de opiniones. Sin embargo, su variabilidad en estilos de escritura, uso de emojis, abreviaciones y menciones requiere técnicas avanzadas de procesamiento de lenguaje natural (NLP) y modelos robustos de clasificación para interpretar con precisión el sentimiento de cada publicación.

Para abordar esta tarea, se evaluarán distintas versiones de BERT y se aplicarán ajustes de hiperparámetros específicos para optimizar cada modelo en función de los datos en español. La evaluación del rendimiento se centrará en la métrica F1, que permite medir tanto la precisión general como la efectividad en cada clase de sentimiento, observando de esta forma la capacidad de cada modelo para generalizar en el conjunto de prueba.

En conclusión, este proyecto busca establecer una base sólida para la clasificación de sentimientos en redes sociales en español y evaluar la influencia de distintas versiones de BERT en la precisión final del modelo. La metodología demuestra cómo los modelos de transformadores, y especialmente BERT, pueden adaptarse para enfrentar los desafíos propios de textos informales y breves, como los tweets.

CAPÍTULO 2

Análisis de Sentimiento

El análisis de sentimientos es una rama del procesamiento de lenguaje natural (NLP) enfocada en extraer y categorizar opiniones, emociones y actitudes presentes en textos. Es una técnica ampliamente utilizada en aplicaciones como el análisis de opiniones en redes sociales, la monitorización de marcas y la evaluación de la satisfacción del cliente en diversas plataformas. En el caso de Twitter, cada tweet aporta una perspectiva única que refleja la percepción del usuario sobre temas específicos, lo que convierte a esta red en una fuente rica y variada de datos de opinión.

Este proyecto de análisis de sentimientos se centra en la clasificación de tweets en español en cuatro categorías de sentimiento: positivo (P), negativo (N), neutral (NEU) y sin sentimiento aparente (NONE). La categoría positiva se asigna a tweets con una percepción claramente favorable, mientras que la categoría negativa se utiliza para aquellos con una percepción desfavorable. Los tweets clasificados como neutral incluyen tanto opiniones positivas como negativas, y la categoría sin sentimiento aparente representa aquellos en los que no se observa un sentimiento relevante.

2.1 Metodología

Para esta tarea de clasificación de sentimientos, se implementaron y evaluaron varios modelos basados en BERT, un modelo de transformadores preentrenado en tareas de comprensión de lenguaje natural. BERT cuenta con diversas versiones que se probarán en este proyecto, cada una con diferencias en sus parámetros y dimensiones de *embeddings*, lo cual impacta en la eficiencia computacional. Los *embeddings* son vectores que representan tanto cada palabra como su contexto, permitiendo una representación semántica precisa. En la tabla a continuación se muestran las distintas versiones de BERT con sus dimensiones y parámetros.

Modelo	Dimensiones Embedding	Parámetros
DistilBERT	30.522 x 768	66M
BERT	30.522 x 768	110M
RoBERT	50.265 x 768	125M

BERT ha demostrado un rendimiento sobresaliente en tareas de NLP, como la clasificación de texto y la detección de emociones, gracias a su capacidad para capturar dependencias contextuales de largo alcance dentro del texto. En este proyecto, se probarán las siguientes versiones de BERT:

- 1) **DistilBERT**: Es una versión comprimida de BERT, diseñada para ser rápida y menos demandante en términos de recursos. Utiliza una técnica llamada destilación, que le permite conservar cerca del 97% del rendimiento del modelo original, pero con la mitad de los parámetros, lo cual lo hace ideal en contextos donde la eficiencia computacional es prioritaria.
- 2) **DistilBERT Multilingüe**: Similar al modelo anterior, pero con la capacidad de procesar múltiples idiomas, incluido el español. Su enfoque multilingüe permite manejar datos en diferentes idiomas manteniendo la velocidad y eficiencia de DistilBERT. Además, es *cased*, lo que significa que diferencia entre mayúsculas y minúsculas, una característica relevante en idiomas donde la capitalización altera el significado.
- 3) **BERTweet**: Este modelo fue entrenado específicamente en datos de Twitter en inglés, lo que le permite captar con mayor precisión el lenguaje informal y abreviado propio de esta red social. Gracias a su entrenamiento en tweets, es más efectivo al interpretar abreviaciones, emojis, hashtags y menciones, optimizándose para el estilo particular de Twitter. Este modelo es exclusivo para el inglés.
- 4) **BERT-base en español**: Es una adaptación de BERT orientada exclusivamente al español, y emplea la técnica de Whole Word Masking (WWM). A diferencia del BERT estándar, que enmascara sub-palabras, WWM enmascara palabras completas, lo que mejora su capacidad para entender la estructura del idioma y las relaciones entre palabras en español. También es *cased*, por lo que distingue entre mayúsculas y minúsculas, relevante en tareas donde la capitalización influye en el significado.
- 5) **RoBERTa-BNE**: Variante optimizada de BERT para el español, que utiliza el marco de RoBERTa, una versión ajustada de BERT que maximiza el rendimiento mediante un entrenamiento más intensivo y una mayor cantidad de datos en español. Su arquitectura *large* cuenta con más parámetros, lo que permite captar matices complejos del idioma, aunque requiere mayor capacidad de procesamiento. Este modelo es ideal para aplicaciones que requieren alta precisión en el análisis de textos en español.

2.1.1 Preprocesamiento del Texto

Para nuestra tarea en particular contamos con 1008 tweets de entrenamiento y 506 tweets de validación, lo cual es una cantidad baja para entrenar modelos Transformers. Debido a esto, nuestras precisiones no serán las más altas ni se acercarán al estado del arte.

Para cada versión de BERT empleada en el entrenamiento y evaluación del modelo, los datos fueron preprocesados para mejorar la calidad de las entradas. Este proceso incluyó los siguientes pasos:

- **Normalización de emojis:** Cada emoji fue convertido a su descripción textual para que el modelo los interprete como expresiones de emoción, facilitando el análisis de sentimientos.
- **Eliminación de menciones y enlaces externos:** Se eliminaron las menciones a otros usuarios y los enlaces, ya que estos elementos no aportan información relevante al análisis de sentimientos.
- **Balanceo de clases:** Al no tener tantas etiquetas de algunas clases se producía un desbalanceo de clases. Para corregirlo hacemos que el modelo preste más atención a las clases con menos ejemplos.

El objetivo del preprocesamiento es estandarizar el texto de manera que el modelo pueda enfocarse en el contenido semántico de cada tweet, optimizando así su capacidad para identificar el sentimiento subyacente en los datos.

2.1.2 Tokenización y Ajuste del Modelo

Para preparar los datos, se utilizó un tokenizador preentrenado específico de cada versión de BERT (BERT, DistilBERT y RoBERTa) que convierte el texto en una secuencia de tokens, donde cada token es representado por un vector en el espacio de embeddings, permitiendo al modelo procesar el contexto de cada tweet. BERT y DistilBERT utilizan una tokenización basada en subpalabras para manejar palabras raras y derivadas, mientras que RoBERTa emplea una tokenización basada en Byte-Pair Encoding (BPE) y fue entrenado en un corpus más amplio, lo que le permite captar mejor ciertos contextos.

Posteriormente, se realiza el ajuste de los modelos (fine-tuning) sobre el conjunto de datos de entrenamiento. Este ajuste permite optimizar los modelos específicamente para la tarea de clasificación de sentimientos, maximizando su precisión en el análisis de los datos y adaptando sus parámetros para obtener mejores resultados en el contexto de redes sociales en español.

2.2 Evaluación de los Modelos

La evaluación del rendimiento de los modelos se basó en la métrica F1-Score, que proporciona una visión completa sobre la capacidad de cada modelo para clasificar correctamente las diferentes categorías de sentimiento en el conjunto de prueba. Los resultados obtenidos son los siguientes:

Modelo	F1-Score
DistilBERT	44.5%
DistilBERT Multilingüe	44.5%
BERTweet	45.6%
BERT-base en español	60.7%
RoBERTa-BNE	65.1%

Los resultados muestran claramente que los modelos entrenados específicamente en español presentan un rendimiento superior en la clasificación de sentimientos en tweets en español. DistilBERT y su versión multilingüe (ambos con F1-Score de 44.5%) presentan resultados limitados debido a su naturaleza comprimida y enfoque general, lo que reduce su capacidad para captar los matices específicos del idioma. BERTweet, aunque está adaptado al estilo de Twitter, solo mejora ligeramente (F1-Score: 45.6%) debido a que fue entrenado en inglés, limitando su comprensión de los patrones lingüísticos del español.

En contraste, BERT-base en español (F1-Score: 60.7%) demuestra una mejora significativa gracias a su entrenamiento exclusivo en español y al uso de la técnica de Whole Word Masking (WWM), que le permite capturar relaciones lingüísticas completas en este idioma. Finalmente, RoBERTa-BNE (F1-Score: 65.1%) ofrece el mejor rendimiento de todos los modelos evaluados, superando al resto gracias a su optimización intensiva y al entrenamiento en grandes volúmenes de texto en español. Esto lo convierte en el modelo más adecuado para la tarea de clasificación de sentimientos en este contexto.

2.2.1 Elección Modelo y Fine-Tuning

Los resultados previos destacan a dos modelos sobre el resto: BERT-base en español y RoBERTa-BNE. Debido a su rendimiento superior, se seleccionaron estos modelos para realizar un proceso de fine-tuning, evaluando cuál logra una mejor puntuación F1. Es importante señalar que BERT cuenta con aproximadamente el 90% de los parámetros de RoBERTa, lo que reduce su carga computacional y lo hace una opción eficiente en comparación. En el proceso de fine-tuning, se experimentó con diferentes valores de learning rate, tamaño de batch y niveles de patience para evitar el sobre entrenamiento y optimizar la F1-Score.

F1-Score/Modelo	BERT-base en español	RoBERTa-BNE
Base	60.7%	65.1%
Learning Rate	63.6%	65.3%
Batch Size	65.1%	65.5%
Patience	65.6%	65.8%
% Mejora	4.9%	0.7%

Los resultados del fine-tuning muestran que BERT-base en español logró una mejora significativa del 4.9% en F1-Score, alcanzando un 65.6%, mientras que RoBERTa-BNE solo mejoró en un 0.7%, alcanzando un 65.8%. Aunque RoBERTa-BNE ofrece el mejor rendimiento en términos absolutos, la eficiencia computacional y la mejora sustancial de BERT sugieren que es una alternativa viable, especialmente en entornos con limitaciones de recursos. Finalmente, podemos

notar que el mayor incremento de F1 se produce cuando ajustados el learning rate y batch size lo que nos confirma lo importante que son estos dos hiperparámetros a la hora de entrenar.

2.2.2 Salida del Modelo

Al seleccionar BERT-base en español como nuestro modelo final, estamos preparados para generar su salida. Cada tweet en el conjunto de prueba será clasificado y etiquetado con su categoría correspondiente. Posteriormente, se generará un archivo de salida en formato .txt que incluirá los identificadores de cada tweet junto con su etiqueta asignada, siguiendo el formato:

ID TWEET	ETIQUETA
1	N
2	P
3	NEU
4	NONE

Este archivo vendrá adjunto con la memoria, que servirá como referencia final de las predicciones del modelo, facilitando el análisis y evaluación del desempeño en la clasificación de sentimientos.

CAPÍTULO 3

Conclusiones

El modelo BERT-base en español se consolidó como la opción óptima para la clasificación de sentimientos en tweets en español. Aunque RoBERTa-BNE obtuvo una F1-Score ligeramente superior en las pruebas iniciales, el proceso de fine-tuning demostró que ambos modelos alcanzaron un rendimiento prácticamente equivalente, con BERT logrando una F1-Score de 65.6% frente al 65.8% de RoBERTa. La ventaja de BERT radica en su menor número de parámetros, lo que reduce su carga computacional y lo hace más eficiente, especialmente en entornos con limitaciones de recursos.

A lo largo de este trabajo, se exploraron y evaluaron diversos modelos, desde versiones comprimidas y multilingües hasta aquellos entrenados específicamente en español. Los resultados evidencian que los modelos adaptados al idioma español logran un mejor desempeño, subrayando la importancia de la especialización lingüística en el análisis de textos informales y complejos, como los tweets. El proceso de fine-tuning, que incluyó ajustes en el learning rate, tamaño de batch y niveles de patience, optimizó el rendimiento de los modelos sin llevarlos al sobre entrenamiento.

En resumen, BERT-base en español ofrece un equilibrio ideal entre precisión y eficiencia, cumpliendo con los objetivos del proyecto y proporcionando una solución robusta para la clasificación de sentimientos en redes sociales en español.