

Análisis y Clasificación de Tweets de Desastres

- Estudiante: Florencia de la Rosa
- Profesor:
- Tutor: Ezequiel Juan Bassano
- Comisión: #60945
- Entrega Final: 21/08/2024



Tabla de Contenido

1. Introducción

2. Objetivo

3. Metodología

4. Resultados

5. Conclusiones



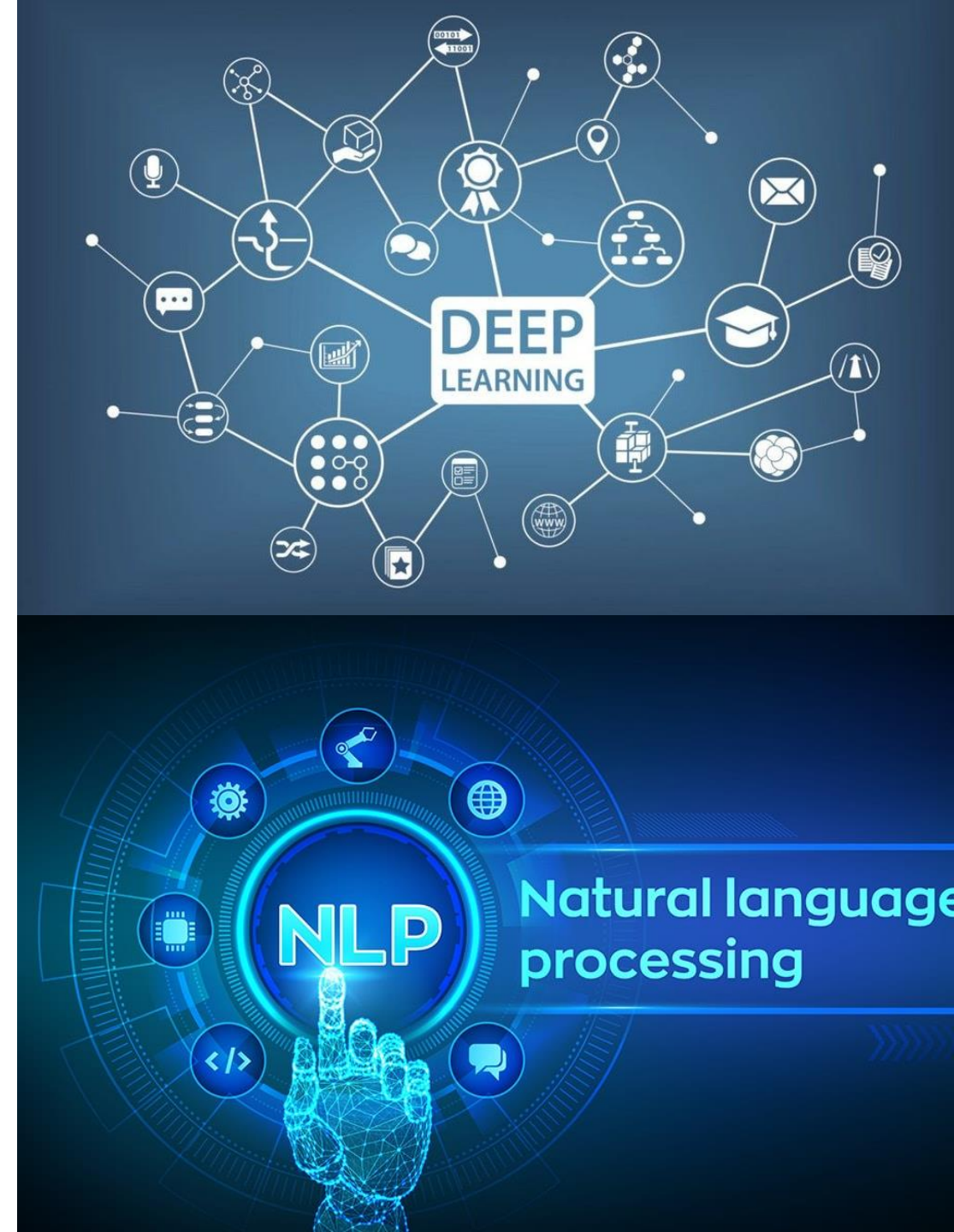
1. Introducción

Los datos utilizados contienen más de **11,000 tweets** con palabras clave relacionadas con **desastres**, recolectados el **14 de enero de 2020**.

Los temas incluyen la **erupción del volcán Taal**, incendios forestales en Australia y el **COVID-19**.

2. Objetivo

- Aplicar **técnicas de Procesamiento de Lenguaje Natural (NLP)** y **Aprendizaje Profundo (Deep Learning)**.
- **Clasificar tweets** en dos categorías: **"Desastre"** y **"No Desastre"** utilizando el modelo BERT.



3. Metodología

Dataset: Disaster tweets
<https://www.kaggle.com/datasets/vstepanenko/disaster-tweets>

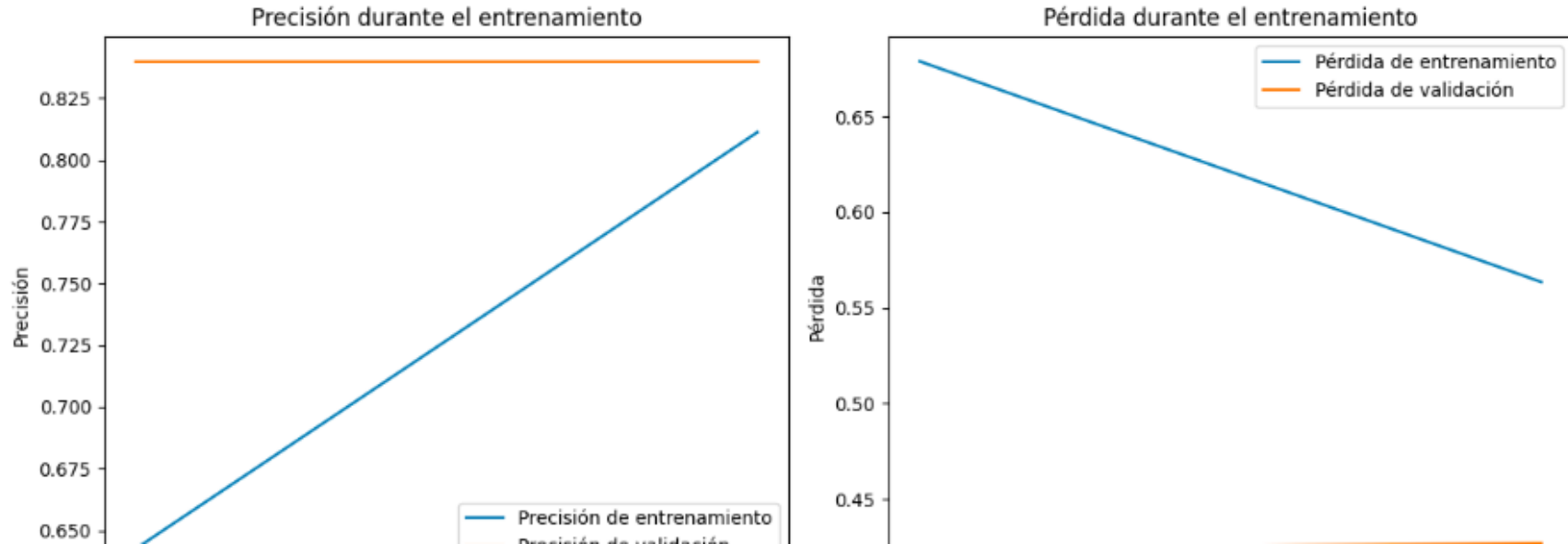
Modelo:

- **BERT** (Bidirectional Encoder Representations from Transformers)
- **Dataset:** Tweets con etiquetas de "Desastre" o "No Desastre".

Preprocesamiento:

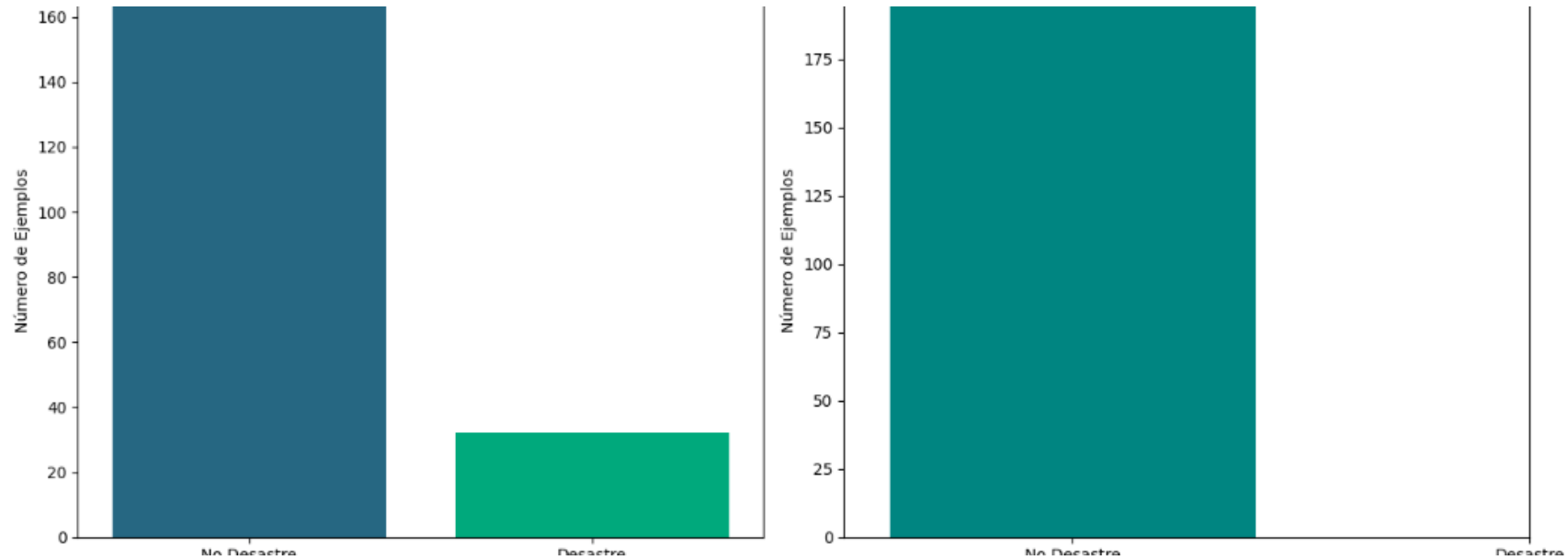
- **Tokenización** usando el BertTokenizer.
- **Preparación de datos** en entradas compatibles con BERT, incluyendo input_ids y attention_masks.





4. Resultados: Entrenamiento y Validación

- **Gráfica de Precisión:** La precisión aumenta en el entrenamiento pero se mantiene constante en la validación, lo que sugiere posible sobreajuste.
- **Gráfica de Pérdida:** La pérdida disminuye en el entrenamiento, mientras que en la validación permanece estable.



4. Resultados:

Distribución de Targets y Predicciones

- **Distribución Real:** La clase "No Desastre" domina, generando un desbalance.
- **Predicciones del Modelo:** El modelo predice en su mayoría la clase "No Desastre", mostrando dificultades para identificar correctamente "Desastre".

5. Conclusiones

- El modelo BERT presenta un sesgo hacia la clase mayoritaria ("No Desastre") debido al desbalance de clases.
- Se identificó sobreajuste, con un mejor rendimiento en el conjunto de entrenamiento que en el de validación.
- Se recomienda implementar técnicas como ajuste de clases (ponderación) y regularización para mejorar el rendimiento.

¡Muchas gracias!

Estudiante: Florencia de la Rosa

Profesor:

Tutor: Ezequiel Juan Bassano

Comisión: #60945

Entrega Final: 21/08/2024