

CEIA - MODELO GRANDE DE LENGUAJE (LLM) E IA GENERATIVA

TF - DETECCIÓN DE FAKE NEWS - ING. JUAN I. MUNAR

El objetivo del presente trabajo es utilizar modelos de lenguaje grande en un problema de Procesamiento de Lenguaje Natural (NLP, por sus siglas en inglés). Se eligió un problema de clasificación binaria, se trata de diferenciar si una noticia es falsa o no utilizando un LLP.

El dataset se obtuvo de Kaggle, presenta algunos problemas de datos duplicados y desbalance (<https://www.kaggle.com/datasets/bhavikjikadara/fake-news-detection>).

El problema fue planteado inicialmente utilizando el modelo T5, pero los requisitos de tiempo/hardware hicieron imposible el entrenamiento. A su vez, se debe mencionar que buena parte del código disponible en repositorios públicos no funciona correctamente. Esto se debe a la constante actualización de las librerías lo que genera problemas de compatibilidad, recordar instalar las versiones correctas.

Brevemente se utilizó un modelo Llama pero, por alguna razón, fue retirada la licencia de uso (ya nos reencontraremos Mark).

Finalmente, para poder ajustar el modelo en tiempo y forma con una GPU T4 disponible en Google Colab, se utilizó un modelo Bert con las siguientes características:

- Se utilizó el modelo 'bert-base-uncased'.
- Dos clases (1-true, 0-fake).
- Longitud máxima de secuencia 256 tokens.
- Batch size de 32.
- Entrenamiento de 4 epochs.
- Learning rate de $2e-5$.
- Entrenamiento con 27428 registros.
- Validación con 6858 registros.

El resultado es demasiado bueno, el accuracy en validación es de 0.9978. Se quitaron caracteres especiales que pudieran identificar las noticias, también la firma de Reuters, y aún así el accuracy es prácticamente 1 lo que hace sospechar que hay algún tipo de sesgo obvio en el dataset.