

Clasificación de Clickbait con NLP y Redes Neuronales

ANTON, Valentina; BARRAGAN, Jeronimo; CIFRE, Aldana; GANDOLFO, Lorenzo; HERLING, Lara

¹ Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Departamento de Computación.

Procesamiento de Lenguaje Natural
2º cuatrimestre 2025

1 - Detectando Clickbait

El objetivo de este proyecto es desarrollar y evaluar un modelo de clasificación de texto basado en redes neuronales para identificar si el titular de una noticia es "clickbait" o no.

Entendemos por **clickbait** al uso de titulares sensacionalistas o con información manipulada para capturar la atención del lector y atraer clicks en un artículo. Este tipo de contenido afecta la confianza en los medios, promueve la desinformación y dificulta la identificación de noticias confiables.

3 - Corpus Bilingüe

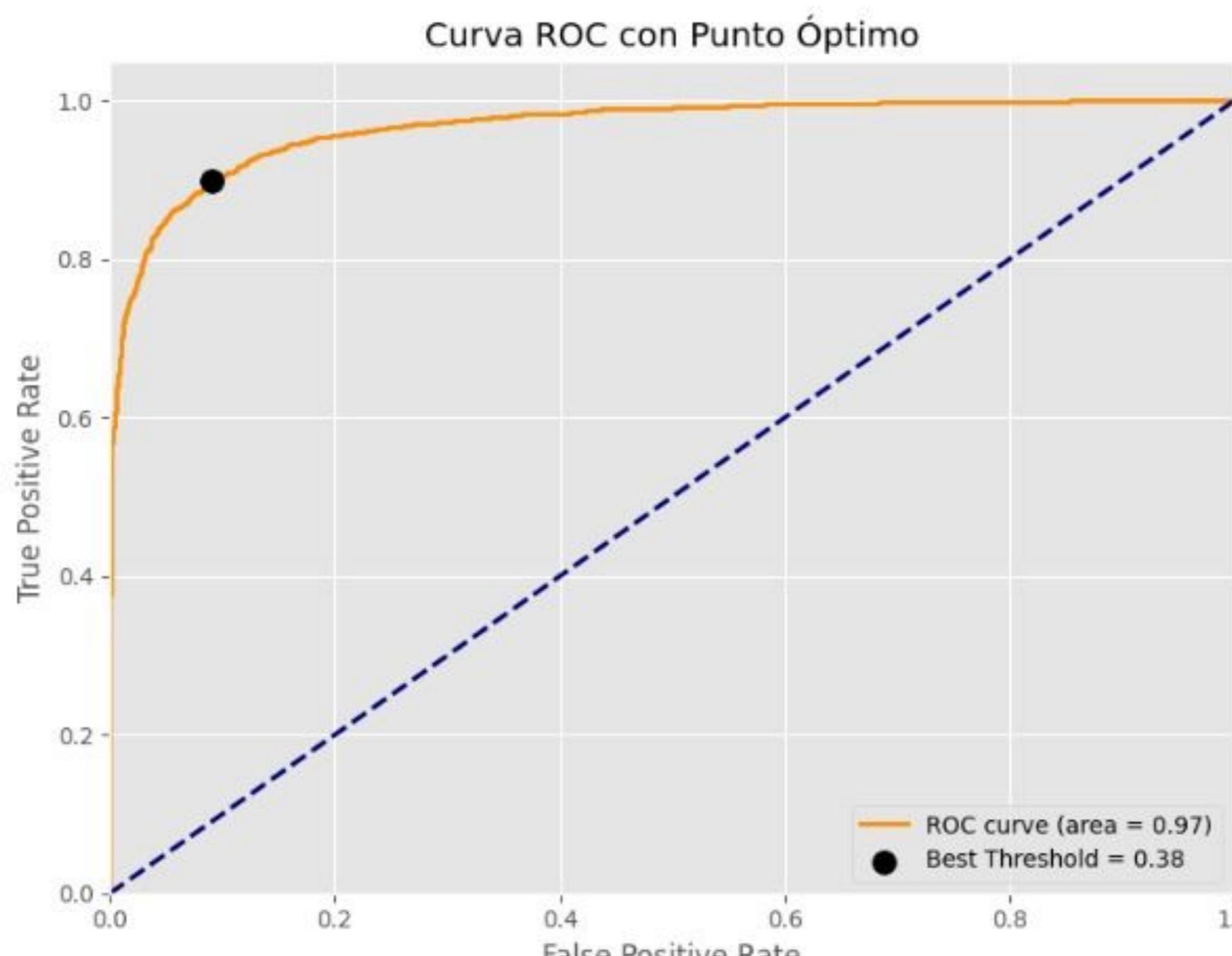
Utilizamos **53.661 instancias** que provenían de cuatro datasets en **inglés** ubicados en HuggingFace de distintos usuarios: '**clickbait detection dataset**'¹ y '**clickbait notclickbait dataset**'² de christinacdl, '**clickbait headline**'³ de evoreign, y '**clickbait title classification**'⁴ de marksverdhei, y **6.570 instancias** en **español** del dataset construido manualmente.

Nuestro dataset unificado:

- Columnas:
 - Text: el título de la noticia
 - Label: 1 si es clickbait y 0 si no es clickbait
 - Source: de qué dataset original es
 - Language: idioma en el que está el texto
- En total utilizamos **60.059 instancias**
- Distribución: Un **59% no eran clickbait** y el **41%** estaban etiquetados como **clickbait**.
- Usamos **48.047** para **entrenar** los modelos, **6.006** en **validación**, y **6.006** para **evaluación**.

4 - Evaluación

- Métricas obtenidas para la clasificación de los datos de evaluación:

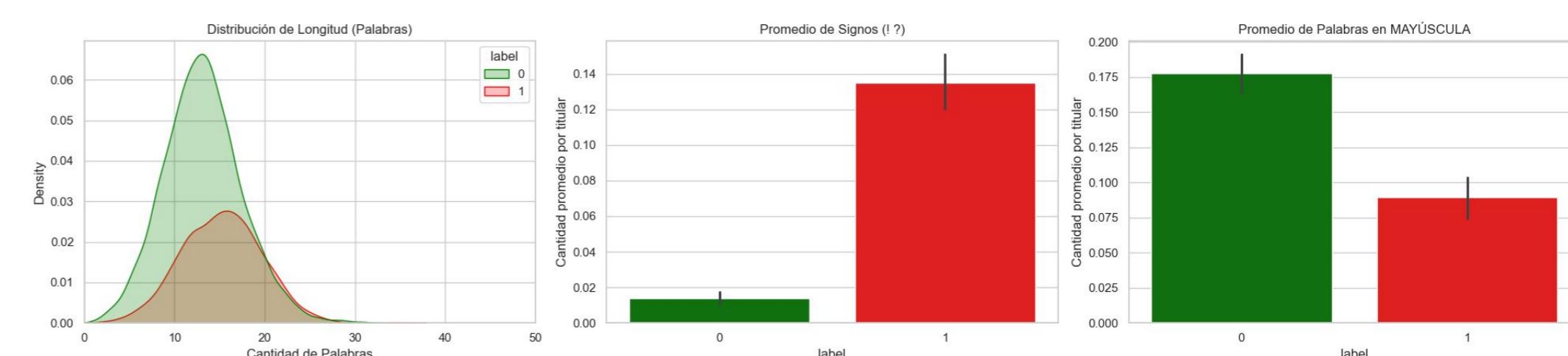


Matriz de confusión		
Etiqueta \ Valor real	Clickbait	No clickbait
Clickbait	2220	319
No clickbait	249	3218

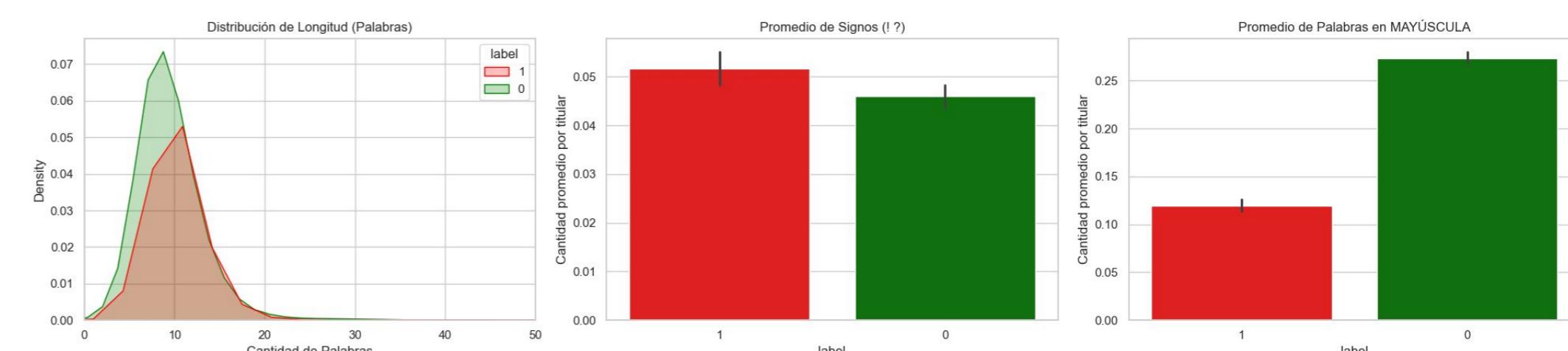
2 - Adaptando BERT

- **Modelo Base**
 - Se utiliza **BERT Multilingual** para manejar texto en múltiples idiomas.
- **Arquitectura**
 - Un clasificador de secuencia (*AutoModelForSequenceClassification*) que añade una capa lineal encima del embedding del token **[CLS]** de BERT.
- **Estrategia de Entrenamiento**
 - Full Fine-Tuning (ajustamos todos los pesos, incluyendo BERT).
 - Utilizamos una **función de pérdida ponderada** (Weighted Loss), de forma de penalizar más los errores en la clase minoritaria (clickbait) y solucionar el desbalance de clases.
- **Hiperparámetros**
 - # Epochs = 4
 - Batch Size = 32
 - Loss Function = **Cross-Entropy**.
 - MAX_LEN = **64 tokens**

➤ Análisis Estructural - Español



➤ Análisis Estructural - Inglés



	Precision	Recall	F1-Score	Support
No Clickbait	0.93	0.91	0.92	3537
Clickbait	0.87	0.90	0.89	2469
ROC AUC Score:				0.9677
Accuracy:				0.91

5 - Resultados

- ➔ "Qué hacer para obtener los dólares que regala el banco central." → How Healthy Is Ginger? **51,7% ES CLICKBAIT**
- ➔ "El banco central va a estar regalando dólares mañana." **78.9% ES CLICKBAIT** → The Markets Are Temperamental. Understand Your Risks. **75,8% NO ES CLICKBAIT**
- ➔ "Con Messi en la tribuna, River enfrentará al Barcelona" **91,4% NO ES CLICKBAIT** → Japan is facing a dementia crisis – can technology help? **83,0% NO ES CLICKBAIT**
- ➔ ¿Se acerca a River el '5' que pidió Gallardo? **72,7% ES CLICKBAIT** → Does Japan's megaquake warning mean the 'big one' is coming? **52,6% ES CLICKBAIT**



Referencias

1. Clickbait detection dataset: https://huggingface.co/datasets/christinacdl/clickbait_detection_dataset
2. Clickbait not clickbait dataset: [christinacdl/clickbait_notclickbait_dataset](https://huggingface.co/datasets/christinacdl/clickbait_notclickbait_dataset). Datasets at Hugging Face
3. Clickbait headline dataset: https://huggingface.co/datasets/evoreign/clickbait_headline
4. Clickbait title classification: https://huggingface.co/datasets/marksverdhei/clickbait_title_classification



Hugging Face