




TruthFinder

“In a world of lies, the truth is your only weapon.”



Alex Nakamura Díaz Francés
David Hernández Uriostegui
Diego Javier Padilla Lara
Erick Daniel Arroyo Martínez
Sebastián Alamina Ramírez

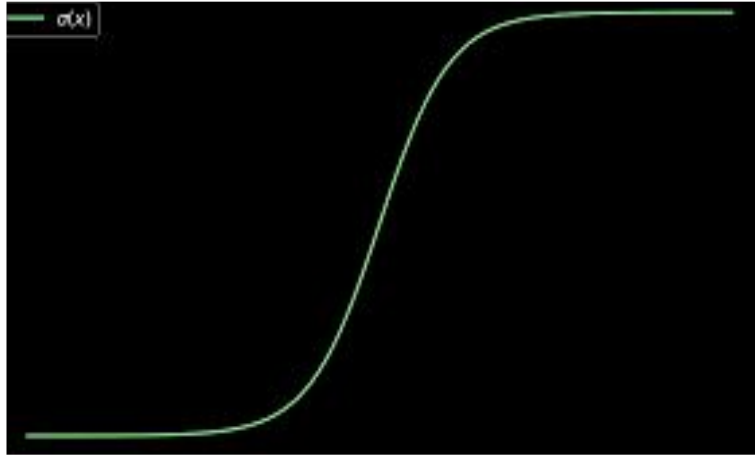
INTRODUCCIÓN

- Las **noticias falsas** son un tipo de periodismo amarillista que consta de piezas de noticias que suelen ser engaños transmitidos a través de las redes sociales y otros medios en línea.
- Su objetivo es engrandecer o imponer ideas, generalmente con propósitos políticos.
- Contienen afirmaciones falsas y/o exageradas, y son fácilmente viralizadas por algoritmos, por lo que es importante detectarlas prematuramente.
- Compararemos el desempeño de diferentes algoritmos de ML usados para la identificación de estas noticias.

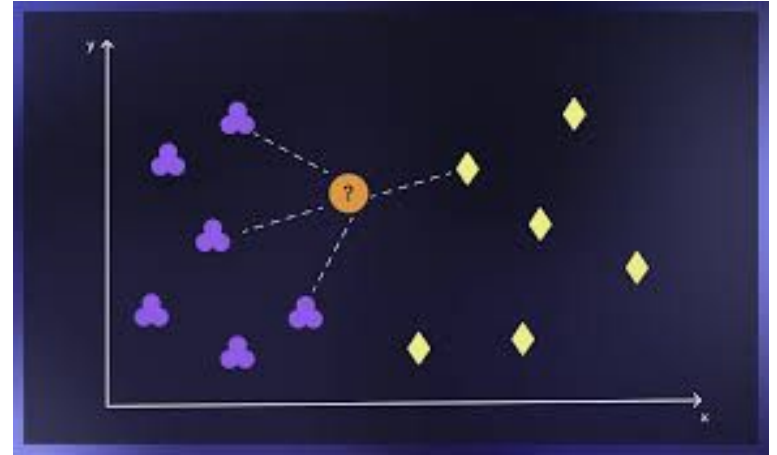


PREGUNTA DE INVESTIGACIÓN

¿Cuál es el rendimiento comparativo de la regresión logística y el algoritmo de K nearest neighbors para la detección y clasificación de noticias falsas en un conjunto de datos etiquetado para la detección y clasificación de Fake News en medios sociales, en términos de precisión y eficiencia en la clasificación?



VS



OBJETIVOS (cómo se responderá la pregunta de investigación)

- Utilizaremos el *dataset*
<https://www.kaggle.com/datasets/jainpooja/fake-news-detection>
- Contiene 17903 noticias falsas y 20826 noticias verdaderas.



METODOLOGÍA

- Preprocesamiento de datos

Con el propósito de mejorar la calidad y la eficacia del modelo, lo que implica el aumento de su capacidad para clasificar de manera precisa y efectiva FN. Las tareas a incluyen: la eliminación de datos faltantes o inconsistentes, la normalización de los valores, la eliminación de valores atípicos, la selección de características relevantes y la transformación de los datos a un formato adecuado para el modelo.



- Algoritmos de clasificación

Para escalas del sistema, haremos uso de los algoritmos planteados en la pregunta de investigación. Como sabemos dichos algoritmos nos permiten la clasificación dicotómica de nuestra variable dependiente. Aún sin esclarecer las herramientas que se emplearán para lograr nuestro cometido, podemos inferir el uso de bibliotecas como: Scikit-learn, Keras, PyTorch, etc.

Una vez definidas las necesidades específicas del proyecto se determinará el conjunto de herramientas que se utilizará para este proyecto.

