

Práctica 4 Módulo 2

Es contratado por un blog de noticias pequeño, el cual recibe constantemente noticias de sus seguidores para publicar.

Sin embargo, un problema muy común al cual se enfrenta es que alrededor del 50% de las noticias que recibe son falsas.

Revisarlas una por una supone mucho esfuerzo por parte de los empleados.

Elabore un modelo clasificador con los datos "fake_news_train.csv" que sea capaz de identificar las noticias recién enviadas por los seguidores las cuales están en el archivo "fake_news_test.csv"

¿Cuántas noticias falsas determina su modelo que contiene el archivo "fake_news_test.csv"?

¿Qué propondría usted para que sea más eficiente la distinción de noticias falsas?

Instrucciones:

Realice un análisis exploratorio de los datos, así como la limpieza e ingeniería de variables. Finalmente construya un modelo de clasificación de los vistos en clase para ayudar a tomar la mejor decisión.

Entregables:

- Código en formato .ipynb con las siguientes características:
 - Limpieza y preprocesamiento de los datos
 - Análisis exploratorio de datos
 - Modelado
 - Todas las partes del proceso deben estar comentadas
- Archivo tipo docx en el que plantee los descubrimientos principales y su recomendación para este problema.