

FAKE NEWS DETECTOR

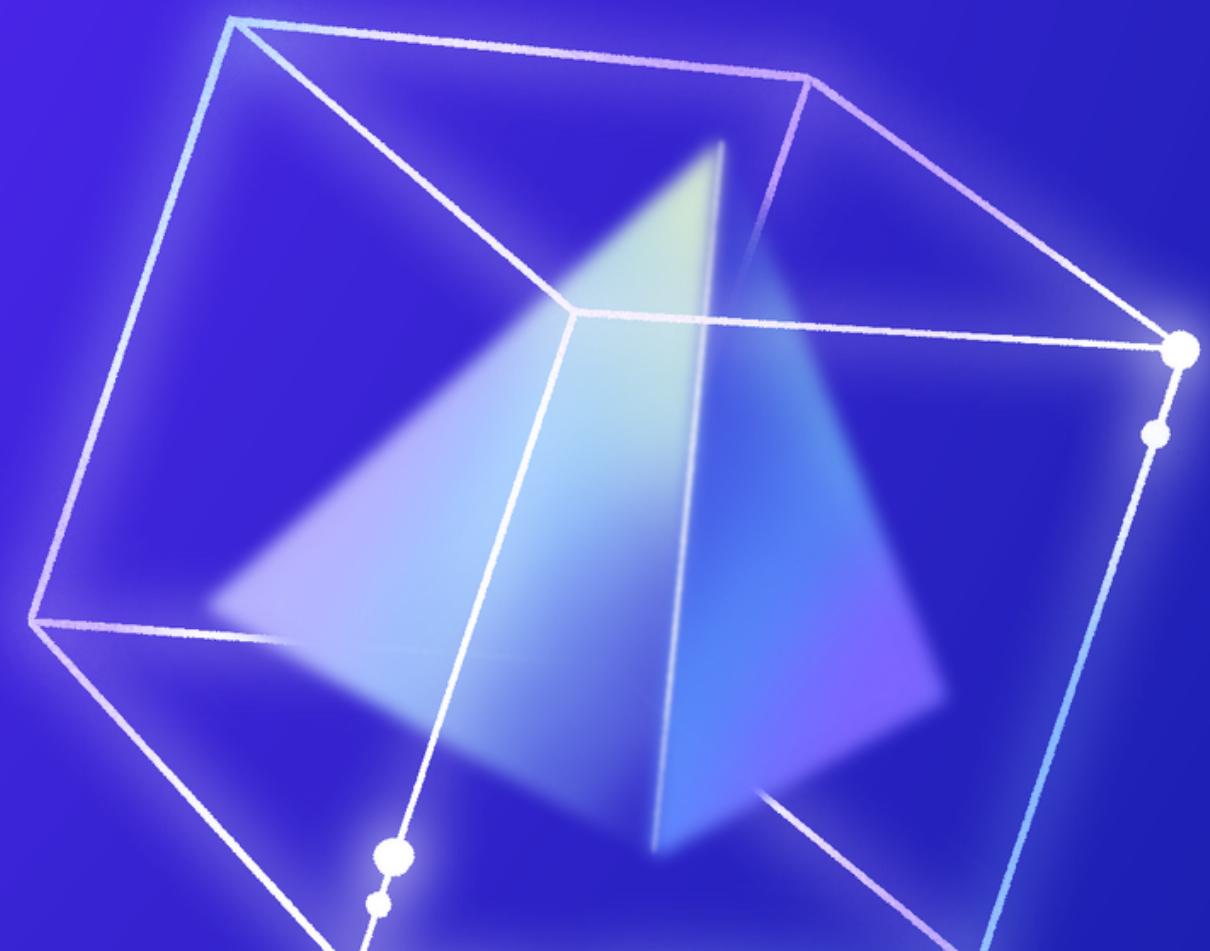
Presentacion realizada por Juan José
Rios Prieto





CONTENIDO

- Introducción
- Exploración de los Datos
- Nube de palabras
- Modelo
- Análisis del modelo
- Objetivos
- Resumen/ Conclusiones



INTRODUCCIÓN

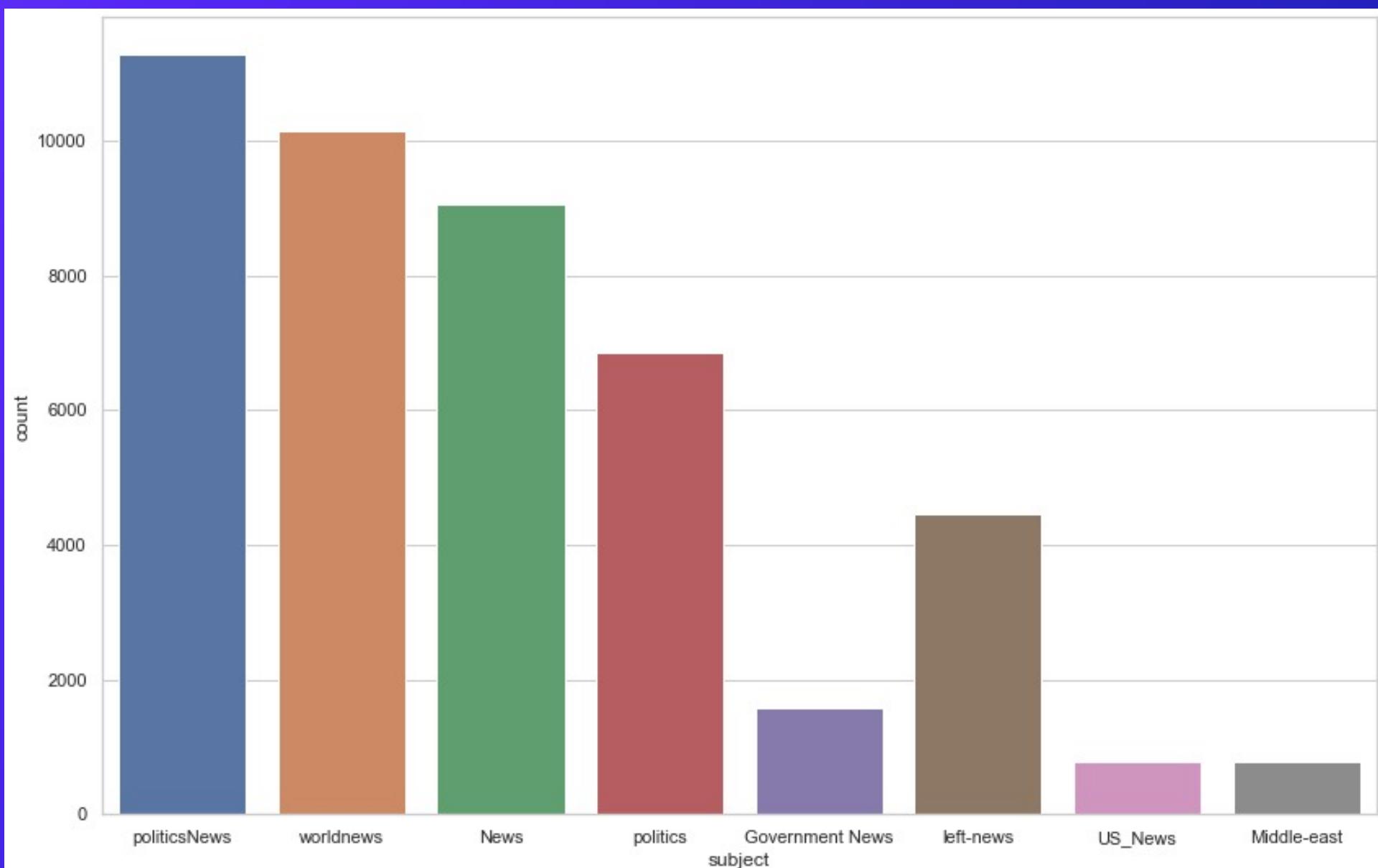
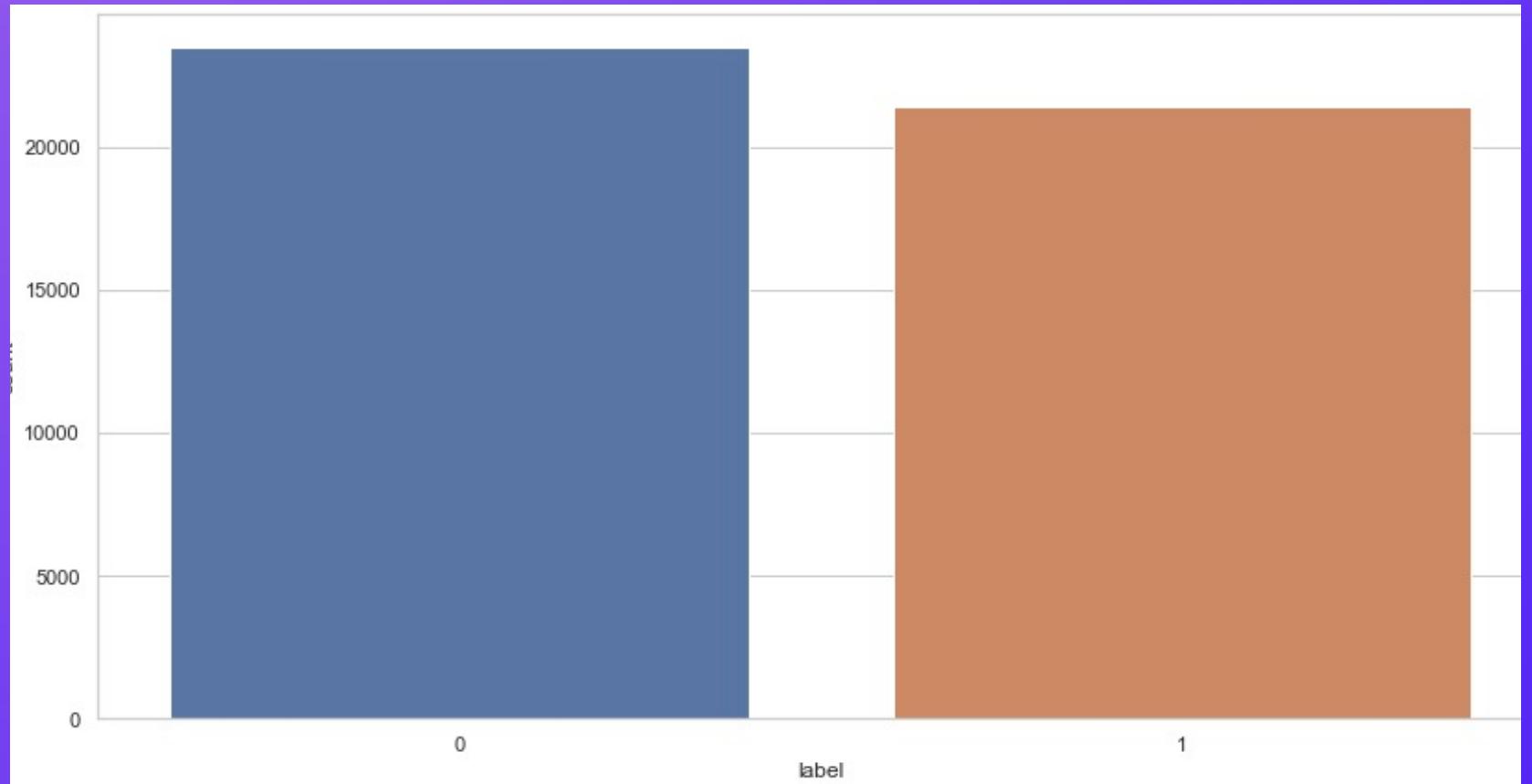
El propósito de este proyecto es hacer un modelo que me clasifica noticias falsas y noticias verdaderas , para esto utilice el modelo de Naive Bayes.

Descripción dataset

Para llevar a cabo esto, sacamos dos datasets de kaggle, uno con noticias verdaderas y otro con noticias falsas, estos datasets cuentan con 4 columnas las cuales son: Tittle, text, subject , date.



EXPLORACIÓN DE DATOS



NUBE DE PALABRAS

A dense word cloud centered around Donald Trump. The most prominent words are 'white house' (large green), 'united states' (large green), 'washington' (large green), 'reuters' (large green), 'trump' (large green), 'said' (large green), 'president' (large green), 'donald' (large green), 'chinese' (large green), and 'korean' (large green). Other visible words include 'china', 'korea', 'new york', 'issue', 'call', 'support', 'concern', 'remain', 'project', 'use', 'mean', 'around', 'program', 'deal', 'prime minister', 'new', 'york', 'issue', 'call', 'said', 'thursday', 'tuesday', 'reuters', 'white', 'house', 'united', 'states', 'washington', 'reuters', 'trump', 'said', 'president', 'donald', 'chinese', and 'korean'. The background is filled with many smaller, less distinct political terms.

FAKE NEWS

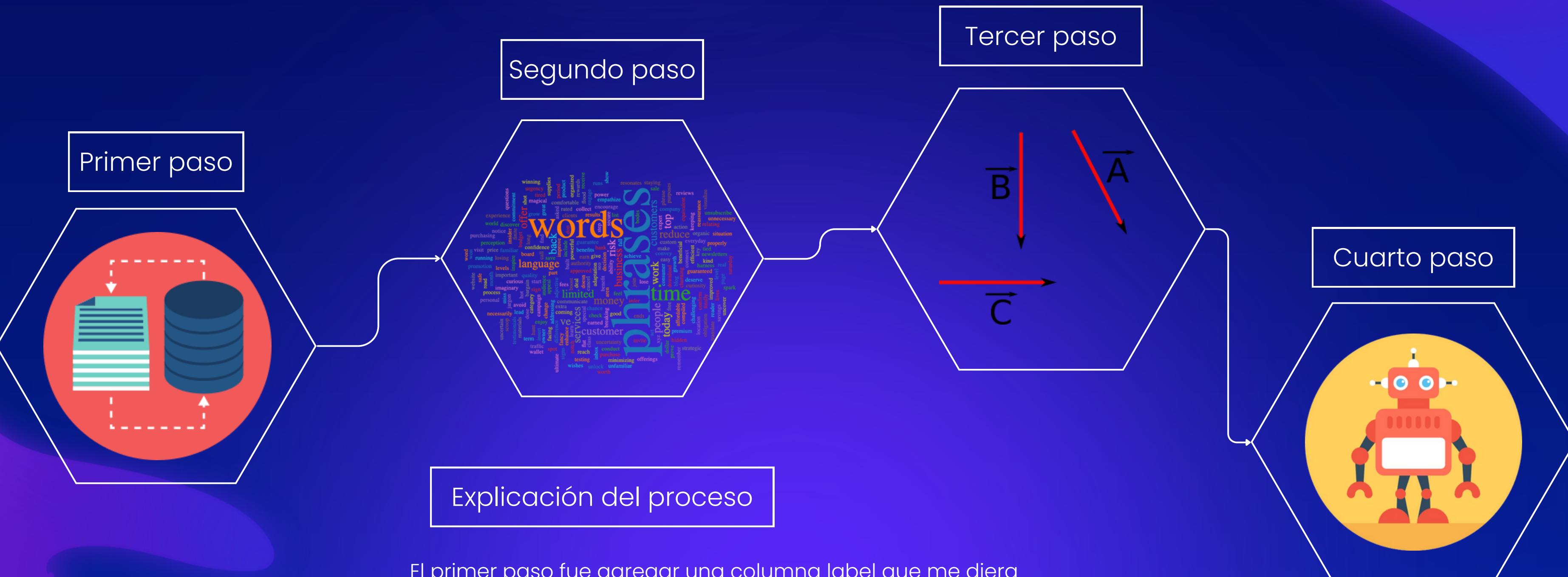
president trump said today that american people say they want one featured image think trump united states president donald trump even via twitter well right make case point white house

TRUE NEWS

MODELO



PROCESO



El primer paso fue agregar una columna label que me diera información sobre si la noticia era falsa o verdadera y luego juntarlo en un df, después quite las stopwords de las noticias para luego vectorizar las palabras y finalmente poder entrenar nuestro modelo.

ANÁLISIS DE MODELO



True Label

False

True

False

True

Predicted Label

	1771	80
True	64	1668
False		



ACURRACY SCORE

Cantidad de predicciones correctas:
0.9574 .

PRECISION_SCORE

De los True news que clasifico el modelo, el
0.9493 fueron bien clasificados.

RECALL_SCORE

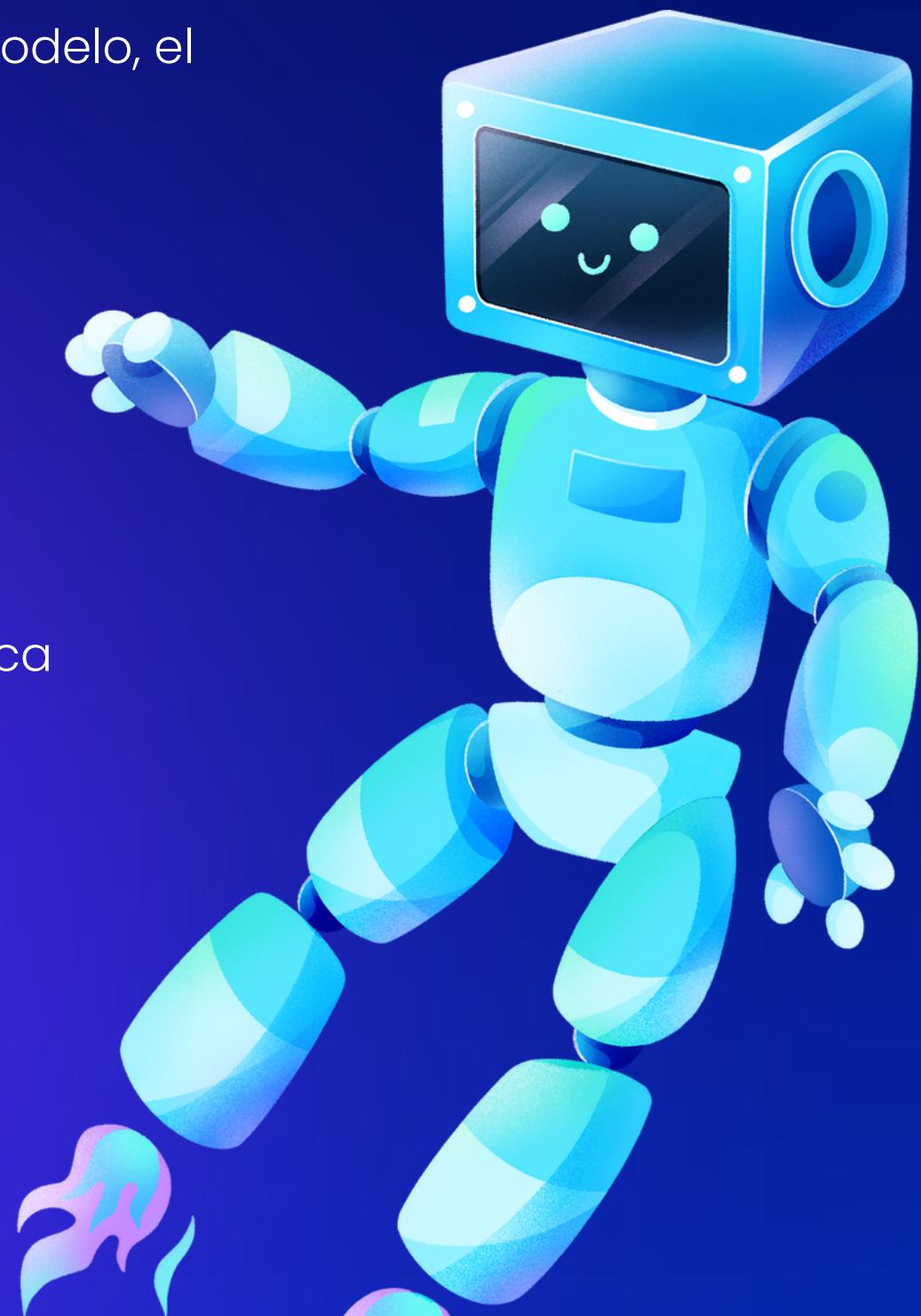
del total de True news el modelo
clasifico correctamente el : 0.9630

F1_SCORE

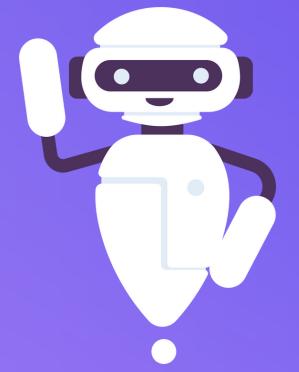
se entiende como la media armónica
de precisión y recall_score : 0.9561

ESPECIFICIDAD

El modelo clasifico 0.9651 de las Fake news
correctamente



OBJETIVOS



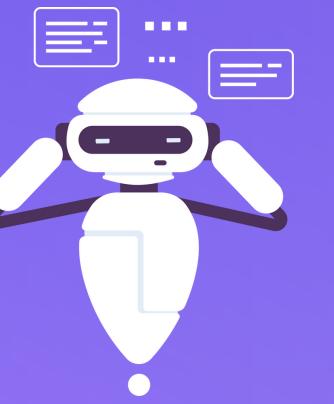
OBJETIVO 01

Mejorar el modelo para hacerlo escalable, ya que el gran volumen de fake news crea esta necesidad en la actualidad .



OBJETIVO 02

Hacer un modelo que pueda recibir la variable subject para tener una mejor clasificación y lograr encontrar patrones mas claros.



OBJETIVO 03

Llevar este modelo a la web para que las personas puedan verificar sus noticias.

CONCLUSIONES



MUCHAS GRACIAS

POR VER ESTA PRESENTACIÓN

GMAIL: JUANJRIOSP9@GMAIL.COM

LINKEDIN: JUAN JOSE RIOS PRIETO

