



CLASIFICACIÓN DE FAKE NEWS

David G Nexans

Horacio Suarez

Yeferson Licet

Alejandro Jait

Jesús Cardozo

01 ¿QUÉ SON LAS
FAKE NEWS?

02 ¿CUÁLES LA
PROBLEMÁTICA?

03 ANÁLISIS DEL
DATASET

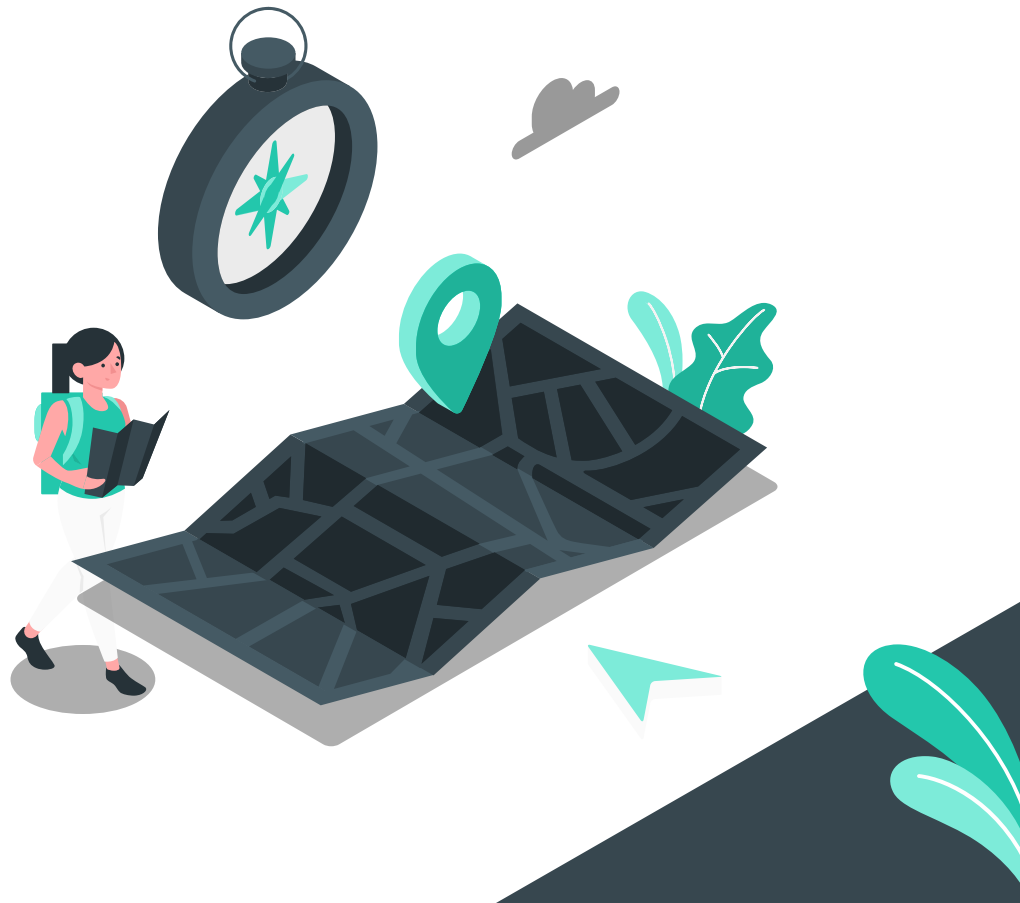
04 RESULTADOS CON
ML CLÁSICO

05 RESULTADOS
CON DNN

06 CONCLUSIÓN
Y CIERRE

¿QUÉ SON LAS FAKE NEWS?

Fake news es un anglicismo que hace referencia a un tipo de noticia que se difunde a través de internet, prensa, radio o televisión y **con el objetivo de *desinformar***



¿QUÉ PROPÓSITO TIENEN?

Las fake news presentan información falsa o sacada de contexto, que puede ser viral en los medios de comunicación y generalmente tienen dos propósitos:



PARA MANIPULAR LA OPINIÓN

Enaltecendo o desprestigiando una institución o persona.



PARA HACER DINERO

Con la ganancia de la publicidad y la viralización que pueden llegar a crear.



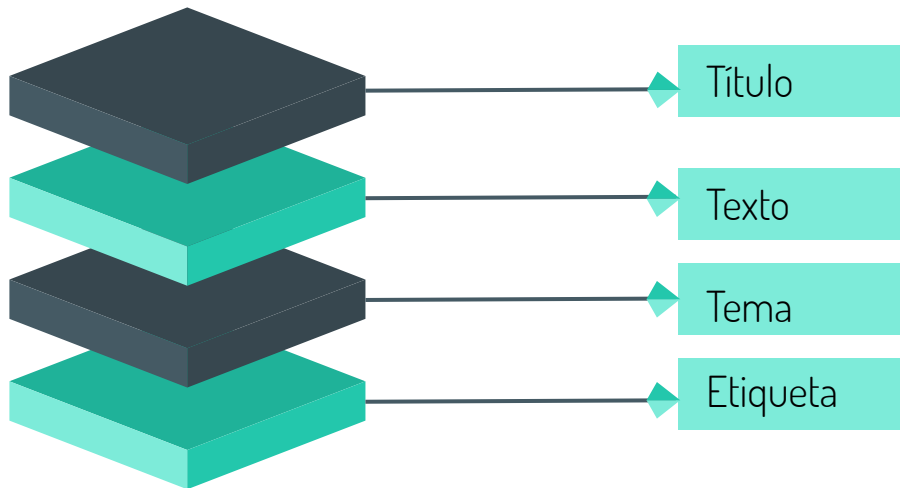
¿CUÁL ES LA PROBLEMÁTICA?

Pese a que las fake news han existido a lo largo de la historia, estas han sido recargadas con turbo con la tecnología digital y erosionan la confianza en los medios de comunicación y el ejercicio de la ciudadanía.

ANÁLISIS DEL DATASET

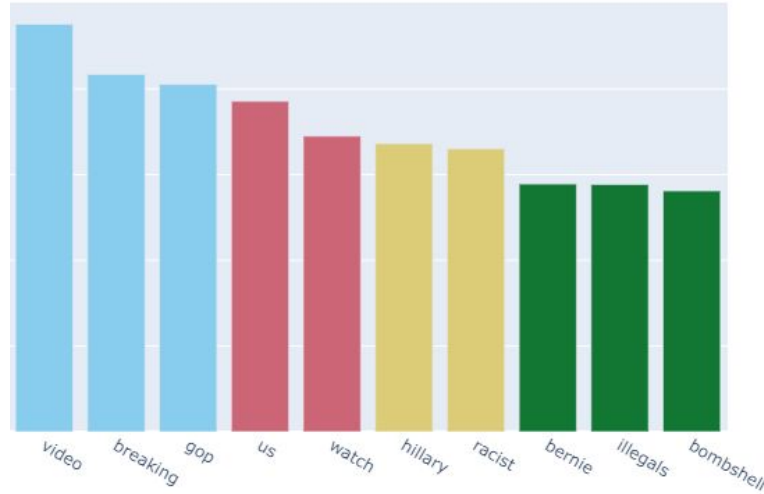


ESTRUCTURA DE LOS DATOS



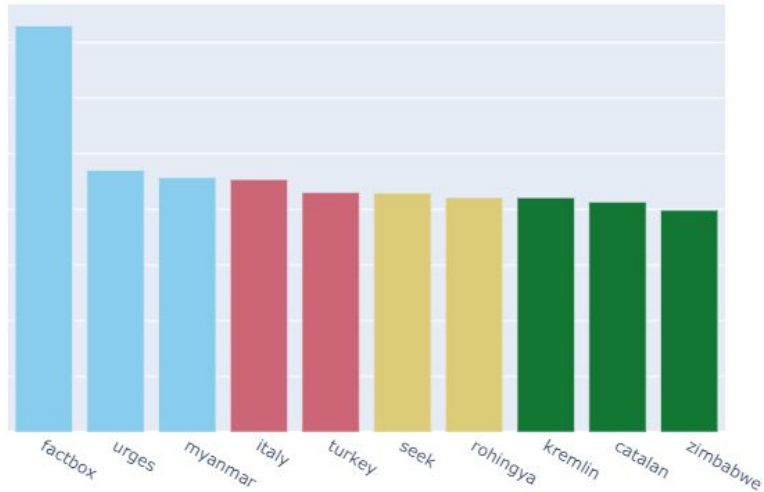
El dataset contiene una colección de **40.000** noticias etiquetadas en el idioma inglés, con fuerte presencia de contenido estadounidense.

PALABRAS TENDENCIA DE FAKE NEWS



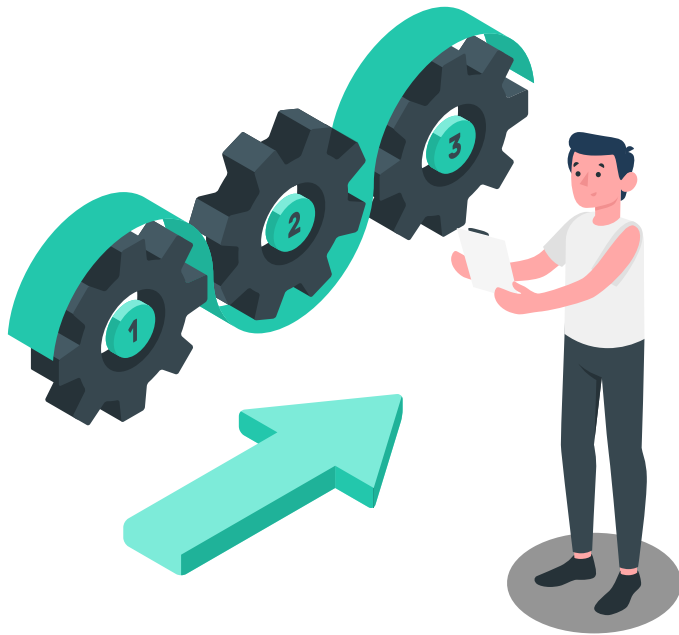
Mediante análisis por frecuencia y algoritmos de machine learning tradicionales encontramos **palabras tendenciosas que ayudan a distinguir fake news.**

PALABRAS CON TENDENCIA DE REAL NEWS



Similarmente, encontramos palabras que marcan una tendencia contraria y nos permiten distinguir noticias verdaderas de las que son fake news.

IMPLEMENTACIÓN DEL CLASIFICADOR



ALGORITMOS TRADICIONALES DE ML

Métodos de ensamble basados en árboles

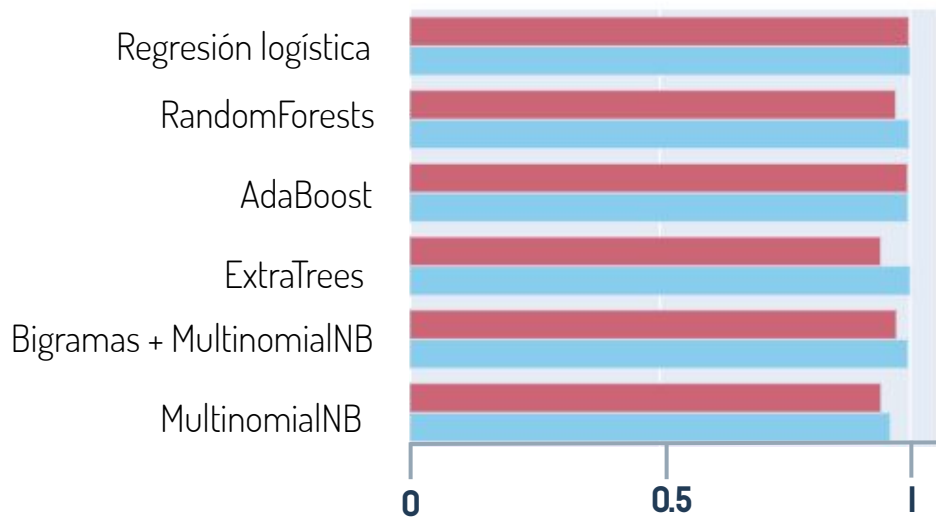
RED NEURONAL

De moderada complejidad

RED NEURONAL OPTIMIZADA

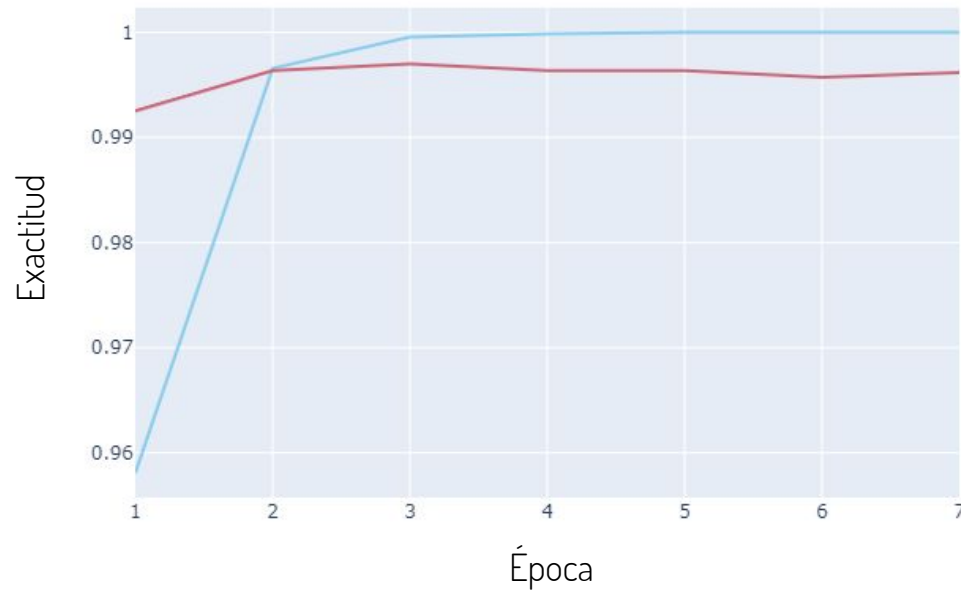
Para promover la generalización y combatir el sesgo dentro en el dataset.

ALGORITMOS TRADICIONALES DE ML



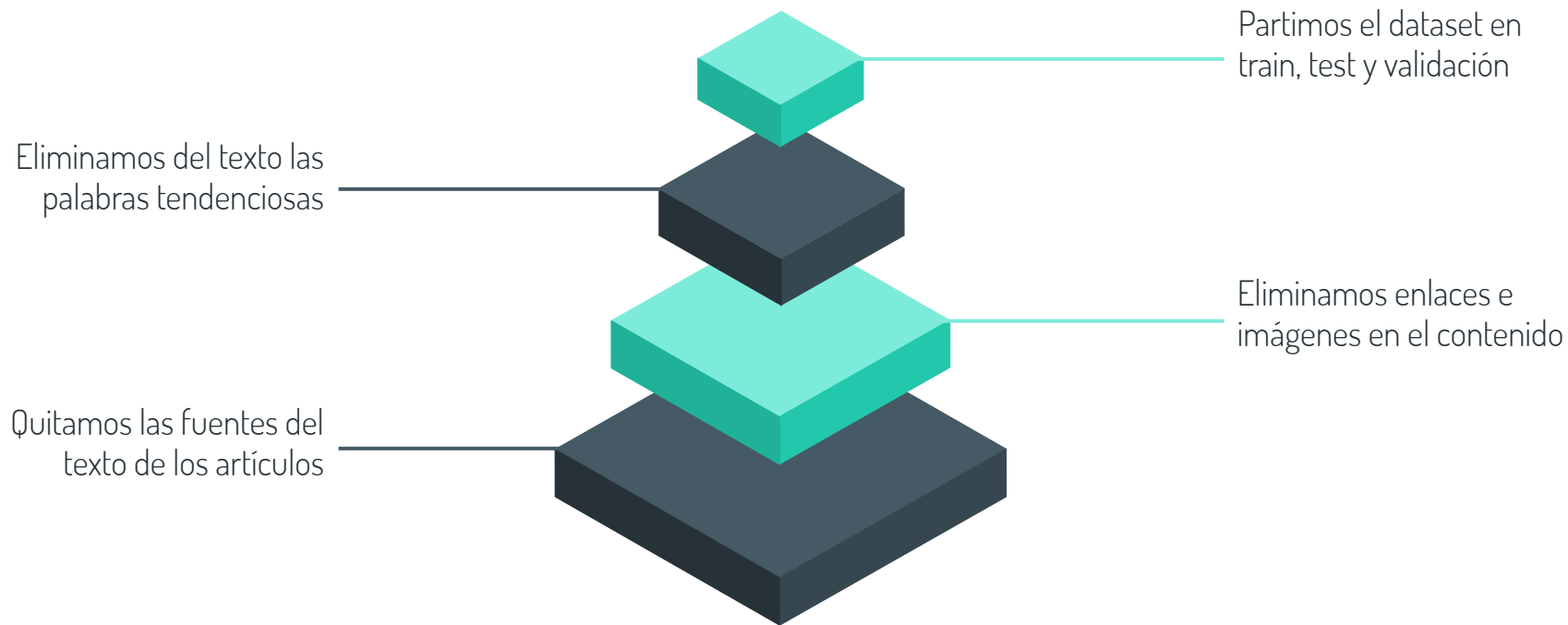
Los algoritmos de machine learning tradicionales dieron resultados excepcionales, especialmente la regresión logística y el AdaBoost.

RED NEURONAL BASELINE

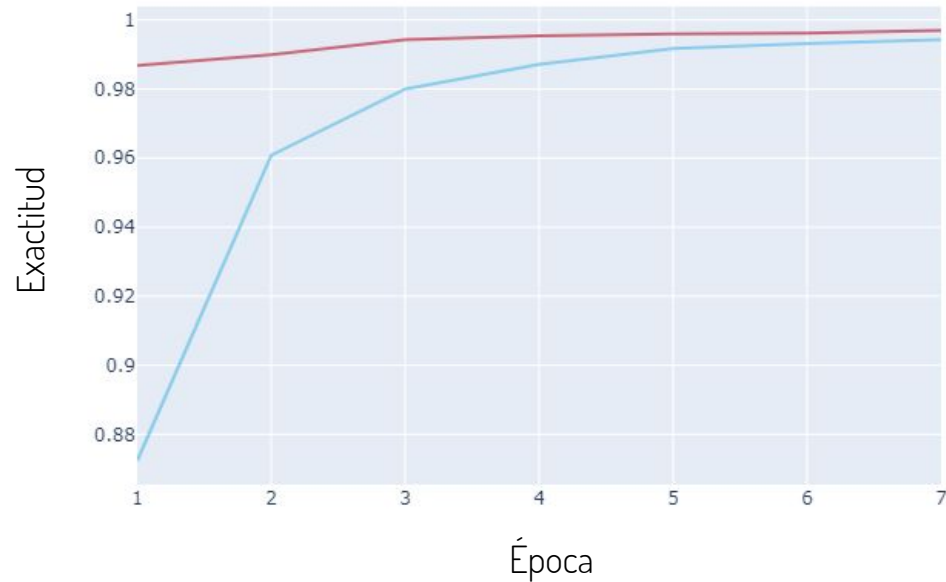


Continuamos aplicando una red neuronal simple, con la que obtuvimos resultados aún mejores, aunque con ligero overfitting.

REPROCESAMIENTO

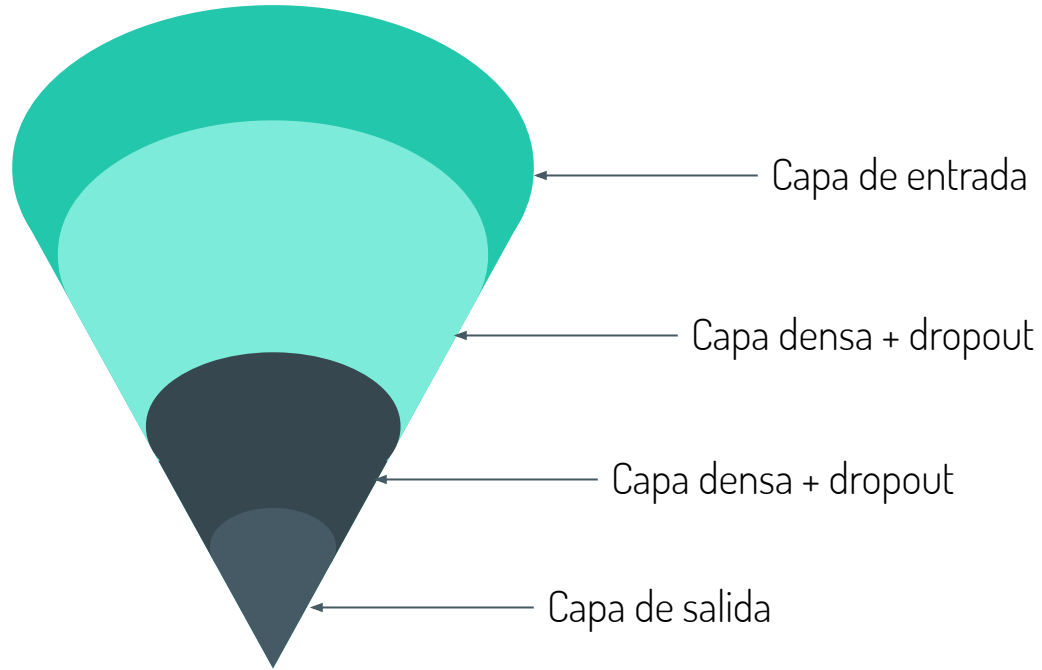


RED NEURONAL OPTIMIZADA



Para mejorar la generalización de la red neuronal aplicamos la técnica de dropout en dos capas escondidas.

ARQUITECTURA DE LA RED





CONCLUSIÓN

Implementamos un clasificador con buen rendimiento en este dataset. Sin embargo, creemos que la fuente de las noticias debe ser revisada y el etiquetado debe ser realizado por un humanos para reducir el sesgo.



GRACIAS

¿Preguntas?

Notebooks en cutt.ly/fakeNewsML

Slides en cutt.ly/fakeNewsPPT