

✦ Detección de noticias falsas con Machine Learning

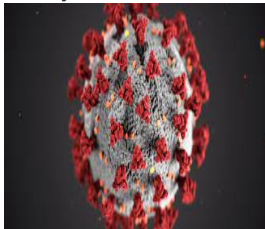
Diego Rojas, Javier Mariño, Nubia Suárez

<https://github.com/fake-news-unal/fake-news>

August 3, 2021

Introducción

¿Qué son las noticias falsas? Según el Ministerio de Tecnologías de la Información: "Se trata de historias o anuncios sensacionalistas, compartidos usualmente por redes sociales para involucrar emocionalmente a los lectores, generar atención y llevar a sus creadores a obtener rentabilidad por clics y viralización".



¿Por qué son peligrosas? Son utilizadas para desprestigiar a personajes públicos, robar y acceder a información personal, entorpecer el manejo de situaciones críticas (como el tratamiento de enfermedades) y obtener réditos políticos.

Problema

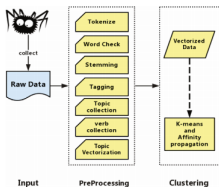
El problema a resolver por la estadística, y en específico el machine learning, es la detección de noticias falsas. Clasificar un texto como verdadero o falso de acuerdo a los procedimientos desarrollados dentro del Natural Language Processing.

El propósito de este proyecto fue comparar la efectividad, medida por la precisión, de distintos métodos del machine learning (neural networks, LSTM, Naive Bayes, Random Forest,...) para la detección de noticias falsas.

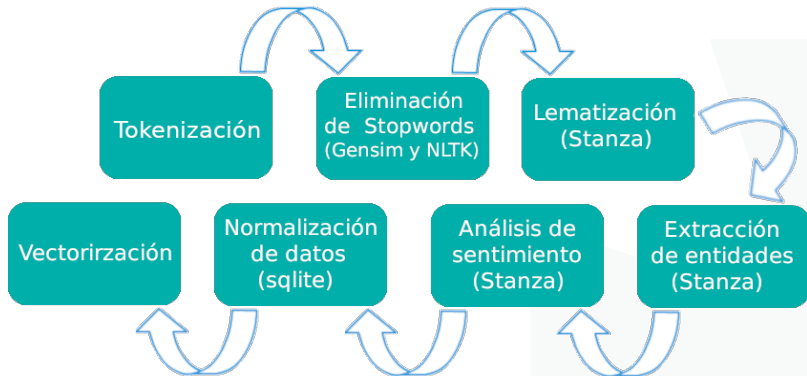


Metodología

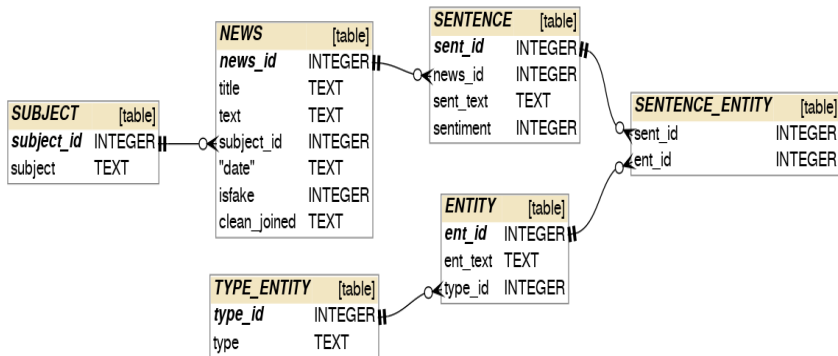
- ✦ La base de datos utilizadas para comparar los distintos métodos fue obtenida del portal Kaggle. Cuenta con un aproximado de 23.000 noticias falsas y 21.000 noticias verdaderas, con sus respectivos títulos, sobre política norteamericana.
- ✦ Los textos fueron tratados mediante las librerías Gensim y Stanza de Python. Con ellas se realizó el proceso de tokenización, lematización, eliminación de stop words y vectorización.
- ✦ Con los textos vectorizados se dividió la base tratada en un 80% de validación y 20% de testeo para la aplicación de los métodos.



Preprocesamiento



Preprocesamiento (Modelo de datos)



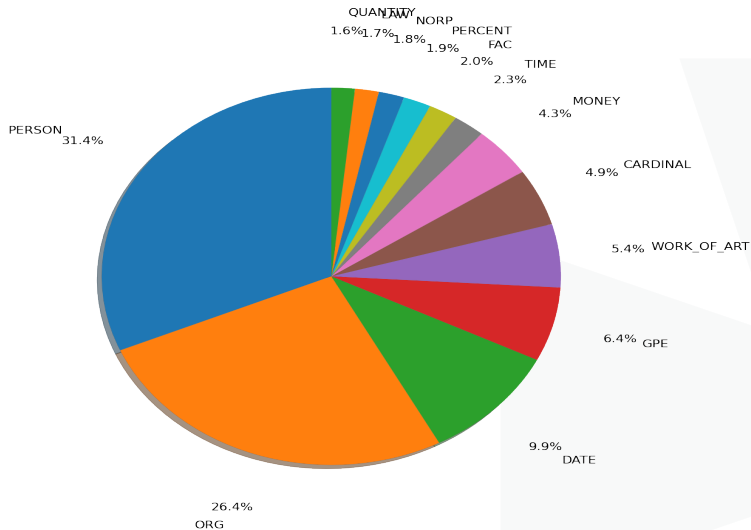
Visualización de los datos

Tabla de tipos de entidades

Tipo	Descripción
GPE	Países, ciudades o estados
NORP	Nacionalidades, grupos políticos o religiosos
ORG	Compañías, agencias o instituciones
DATE	Periodos y fechas
PERSON	Personas incluyendo ficción
PERCENT	Porcentajes
MONEY	Valores monetarios
CARDINAL	Números
ORDINAL	Números en letras "first", "second", etc
LAW	Documentos nombrados que representan leyes
WORK_OF_ART	Títulos de libros, canciones, etc
TIME	Tiempos que son inferiores a un día
FAC	Construcciones, edificios, avenidas, etc.
LOC	Ubicaciones diferentes a GPE como ríos, montañas, etc
EVENT	Huracanes, batallas, guerras, eventos deportivos, etc
QUANTITY	Medidas como pesos o distancias
PRODUCT	Objetos, carros, comida, etc (servicios no)
LANGUAGE	Cualquier idioma mencionado

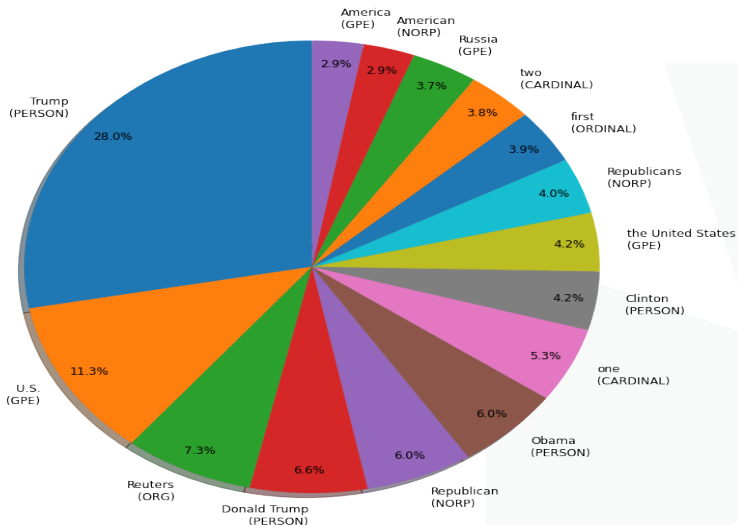
Visualización de los datos

Tipos de entidades



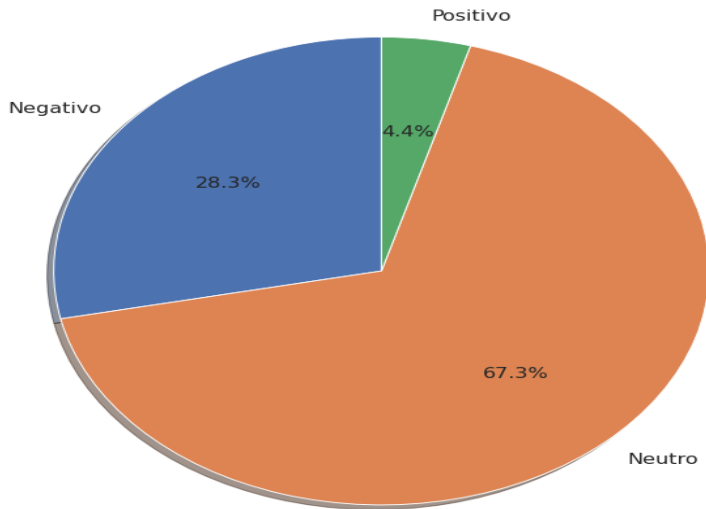
Visualización de los datos

Entidades



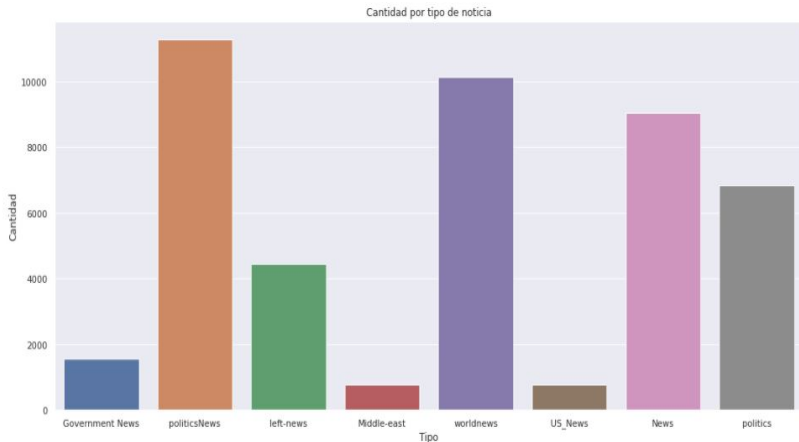
Visualización de los datos

Sentimiento



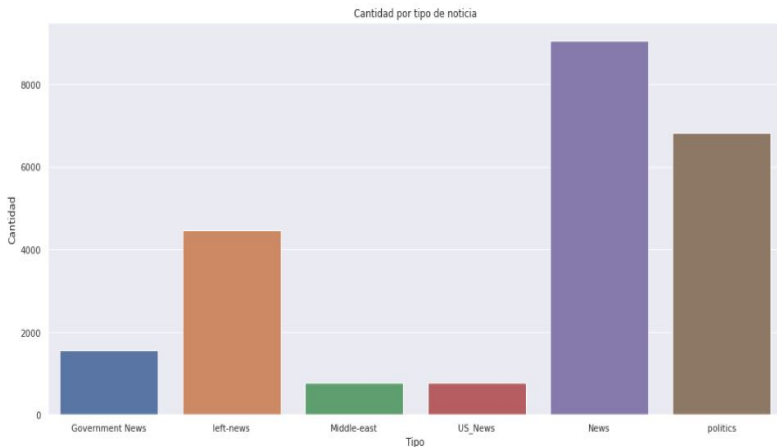
Visualización de los datos

TIPO: Noticias Reales



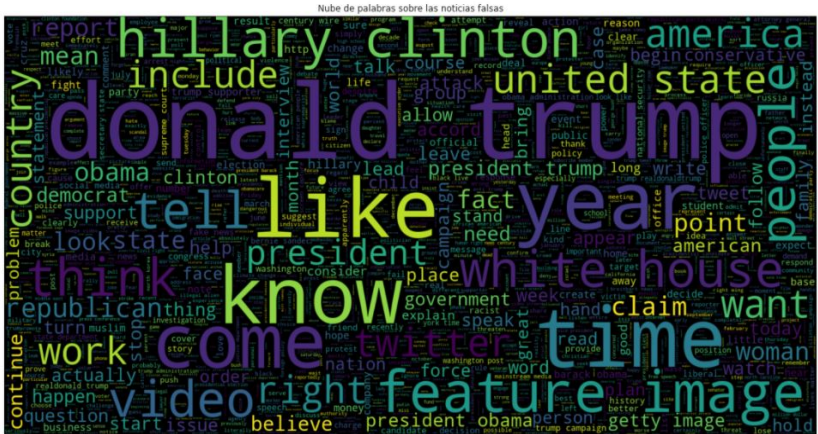
Visualización de los datos

TIPO: Noticias Falsas



Visualización de los datos

TIPO: Noticias Falsas

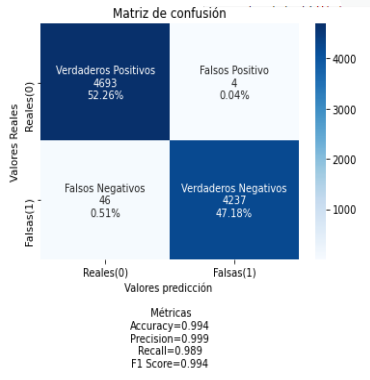


Red Neuronal

Para la construcción de la red neuronal se utilizaron como inputs los vectores generados por Gensim y Stanza. El optimizador utilizado fue "Adam" y la función de perdida la "binary-crossentropy". El modelo alcanzó, después de 2 épocas, una precisión de 0,9997 en entrenamiento y 0,9985 en validación.

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 40, 128)	1132157568
flatten (Flatten)	(None, 5120)	0
dense (Dense)	(None, 64)	327744
dropout (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 32)	2080
dropout_1 (Dropout)	(None, 32)	0
dense_2 (Dense)	(None, 16)	528
dropout_2 (Dropout)	(None, 16)	0
dense_3 (Dense)	(None, 1)	17
Total params: 1,132,487,937		
Trainable params: 1,132,487,937		
Non-trainable params: 0		



Naive Bayes

El clasificador bayesiano se construyó con base en los inputs generados por Gensim y Stanza. Se limitaron las secuencias a vectores de máximo 40 componentes con la función `pad_sequences` de TensorFlow. Además, se ajustó mediante la librería ScikitLearn de Python el algoritmo Gaussiano Naive Bayes de Clasificación.

Con un 80% de los datos dirigidos al entrenamiento y un 20% al testeo, se alcanzó la precisión más baja entre todos los modelos con tan sólo un 59% de noticias clasificadas correctamente.

LSTM (Long short-term memory)

Model: "sequential"

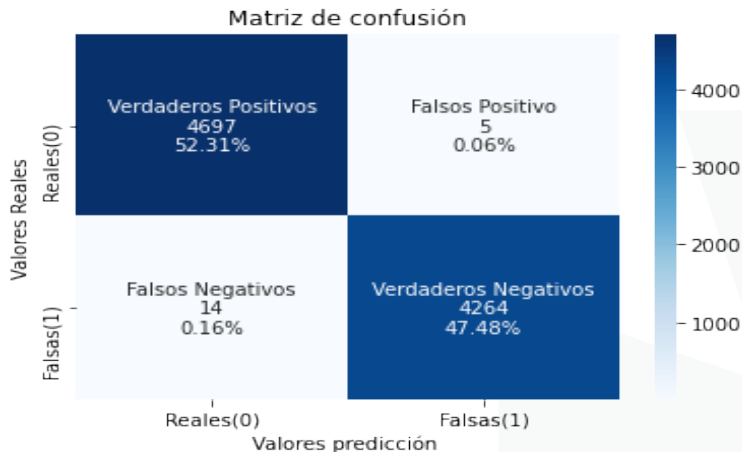
Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, None, 128)	1132157568
bidirectional (Bidirectional)	(None, 256)	263168
dense (Dense)	(None, 128)	32896
dense_1 (Dense)	(None, 1)	129

Total params: 1,132,453,761

Trainable params: 1,132,453,761

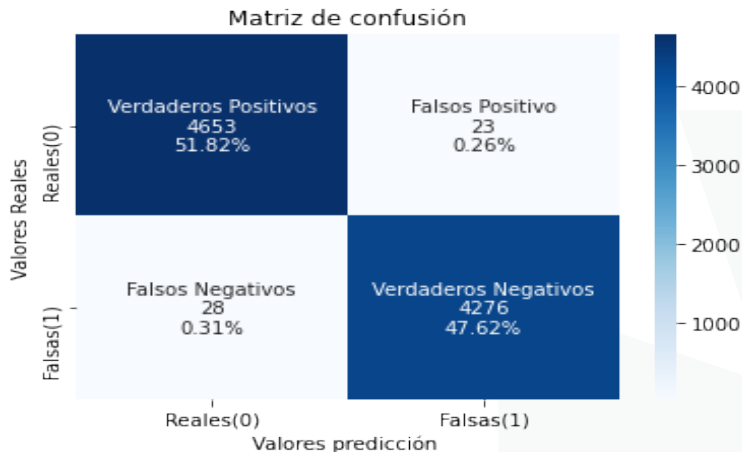
Non-trainable params: 0

LSTM (Long short-term memory)



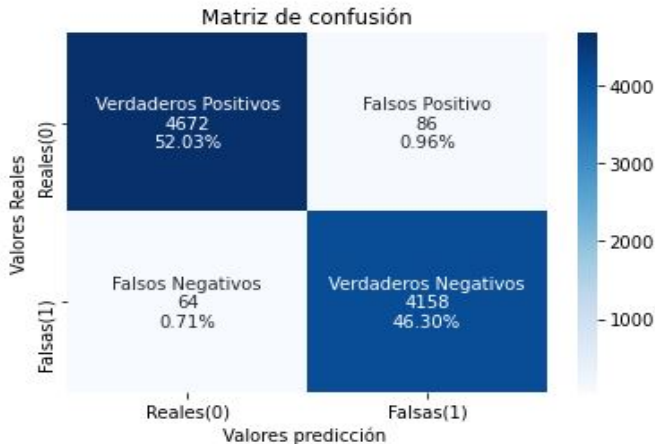
Métricas
Accuracy=0.998
Precision=0.999
Recall=0.997
F1 Score=0.998

SVM (Support vector machine)

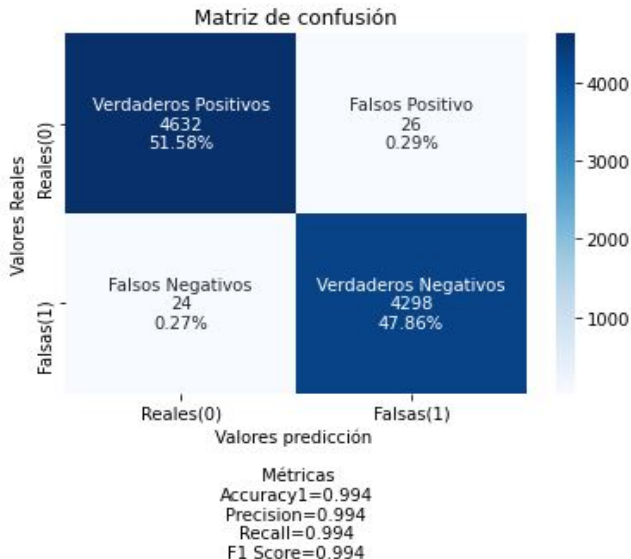


Métricas
Accuracy=0.994
Precision=0.995
Recall=0.993
F1 Score=0.994

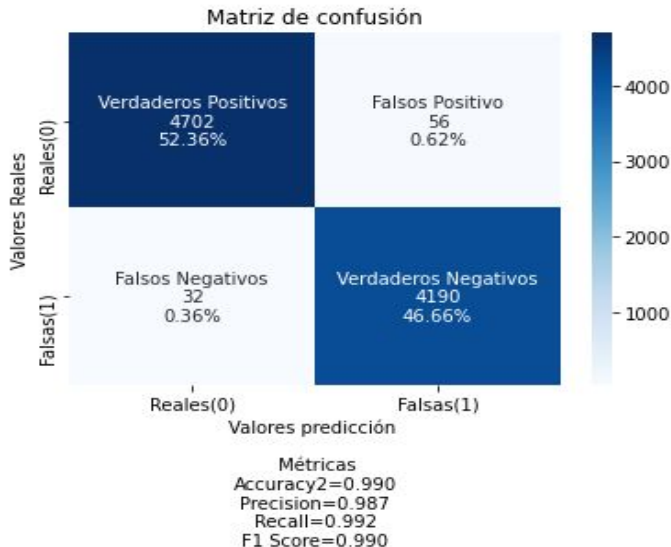
Regresión Logística



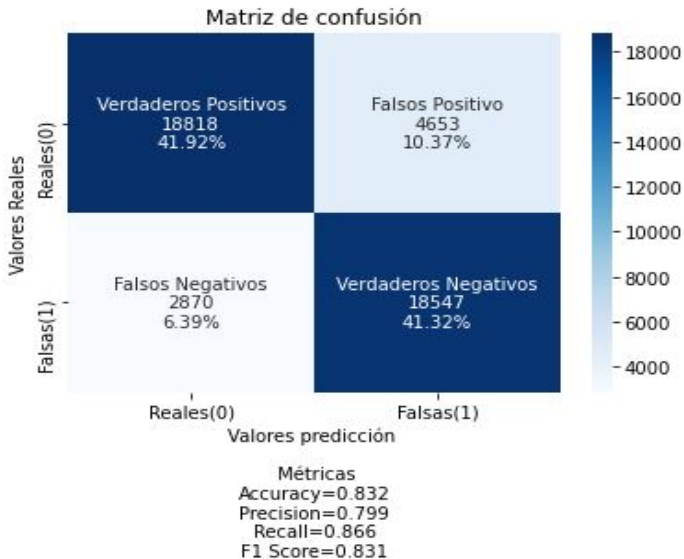
Árboles de decisión



Bosques Aleatorios



K-means (K-medias)






Comparación de resultados

	LSTM	SVM	Decision Tree	Neural Network	Logistic	Random Forest	K Means	Naive Bayes
Accuracy	0,998	0,994	0,994	0,994	0,983	0,990	0,832	0,589
Precision	0,999	0,995	0,994	0,999	0,980	0,987	0,799	0,542
Recall	0,997	0,993	0,994	0,989	0,985	0,992	0,866	0,852
F1 Score	0,998	0,994	0,994	0,994	0,982	0,990	0,831	0,663

Conclusiones

- ✚ El modelo que mejor resultados fue el LSTM con resultados muy cercanos a la red neuronal y el modelo con los resultados más bajos fue Naive Bayes

Referencias

-  Stanford NLP Group. (2014). Stanza v1.2.0 [software library]. Stanford, California. Disponible en <https://stanfordnlp.github.io>
-  Ahmed H, Traore I, Saad S. “Detecting opinion spams and fake news using text classification”, Journal of Security and Privacy, Volume 1, Issue 1, Wiley, January/February 2018.
-  Ahmed H, Traore I, Saad S. (2017) “Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. In: Traore I., Woungang I., Awad A. (eds) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. ISDDC 2017. Lecture Notes in Computer Science, vol 10618. Springer, Cham (pp. 127-138).