



Tecnológico de Monterrey

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE
MONTERREY CAMPUS ESTADO DE MÉXICO

Inteligencia artificial avanzada para la ciencia de datos I.

Grp. 101.

Alumno:

Leonardo Cossio Dinorin.

Profesor:

Jorge Adolfo Ramírez Uresti.

Módulo:

Aprendizaje máquina

Uso de Framework o biblioteca de aprendizaje máquina para la implementación de una solución.

Fecha de entrega:

07 de septiembre del 2024.

Descripción:

Se implementó un modelo de machine learning utilizando un Framework.

Se utilizó el modelo de clasificación de Random Forest de la librería scikit-learn para poder clasificar noticias falsas y noticias verdaderas.

Dataset:

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

Métricas de evaluación:

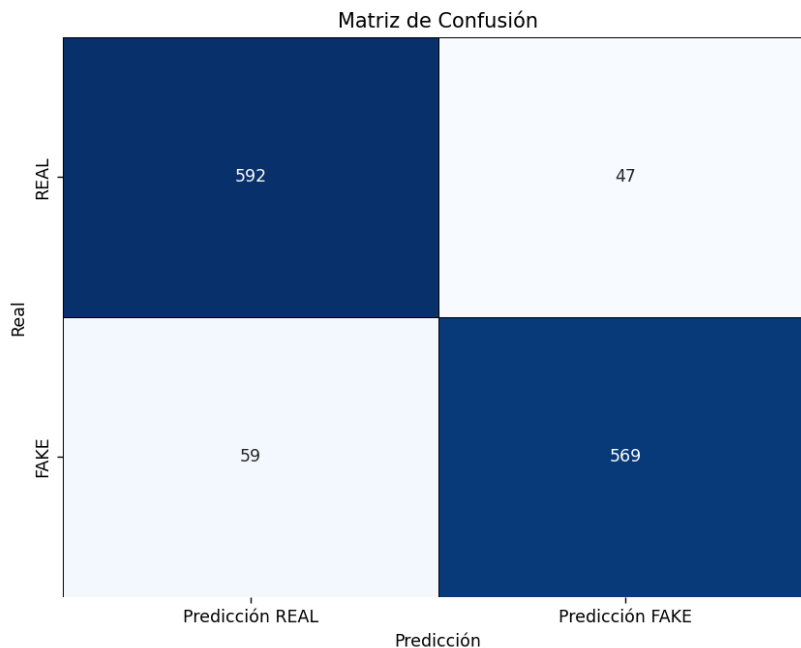
Se utilizaron las métricas presentes en la librería utilizada y se obtuvieron los siguientes resultados:

```
***** VALIDATION SCORE *****
Accuracy on validation set: 0.89
Recall on validation set: 0.88
Precision on validation set: 0.90
F1 score on validation set: 0.89
*****
```

```
***** TEST SCORE *****
Accuracy on test set: 0.92
Recall on test set: 0.91
Precision on test set: 0.92
F1 score on test set: 0.91
*****
```

Como se puede observar, el modelo utilizado consiguió una calificación promedio de 89% en el conjunto de validación y posteriormente mejoró en el conjunto de prueba, obteniendo un promedio de 91.5% en las métricas utilizadas.

Matriz de confusión:



En esta matriz se puede observar claramente cómo aproximadamente el 10% de los datos fueron falsos positivos y falsos negativos, lo que concuerda con las métricas mostradas anteriormente.

Predicciones individuales:

En el algoritmo se implementó una predicción individual, la cual es una noticia seleccionada al azar del conjunto de prueba, posteriormente se imprime el índice de la noticia, la etiqueta real y la predicción.

```
Índice de la muestra: 983  
Clase real: 0  
Clase predicha: 0
```

Conclusión:

Se logró implementar un algoritmo de machine learning utilizando la librería de sklearn en Python, se logró un accuracy del 92% en el conjunto de prueba utilizando random forest. Aunque los parámetros del modelo fueron elegidos de manera heurística, posiblemente se pueden obtener mejores valores utilizando un Grid search, el cual no pudo ser implementado debido al poder computacional que requiere (principalmente por tratarse de Random forest), sin embargo, en una computadora que cuente con un buen GPU y la posibilidad de estar encendida mucho tiempo ejecutando el programa, se podría obtener un mejor modelo.