

Reporte Módulo 2; Uso de framework de aprendizaje de máquina para la implementación de una solución

Para esta entrega se implementó un algoritmo de clasificación de random forests, utilizando el módulo de scikit-learn correspondiente.

Se entrenó este modelo utilizando el siguiente dataset:

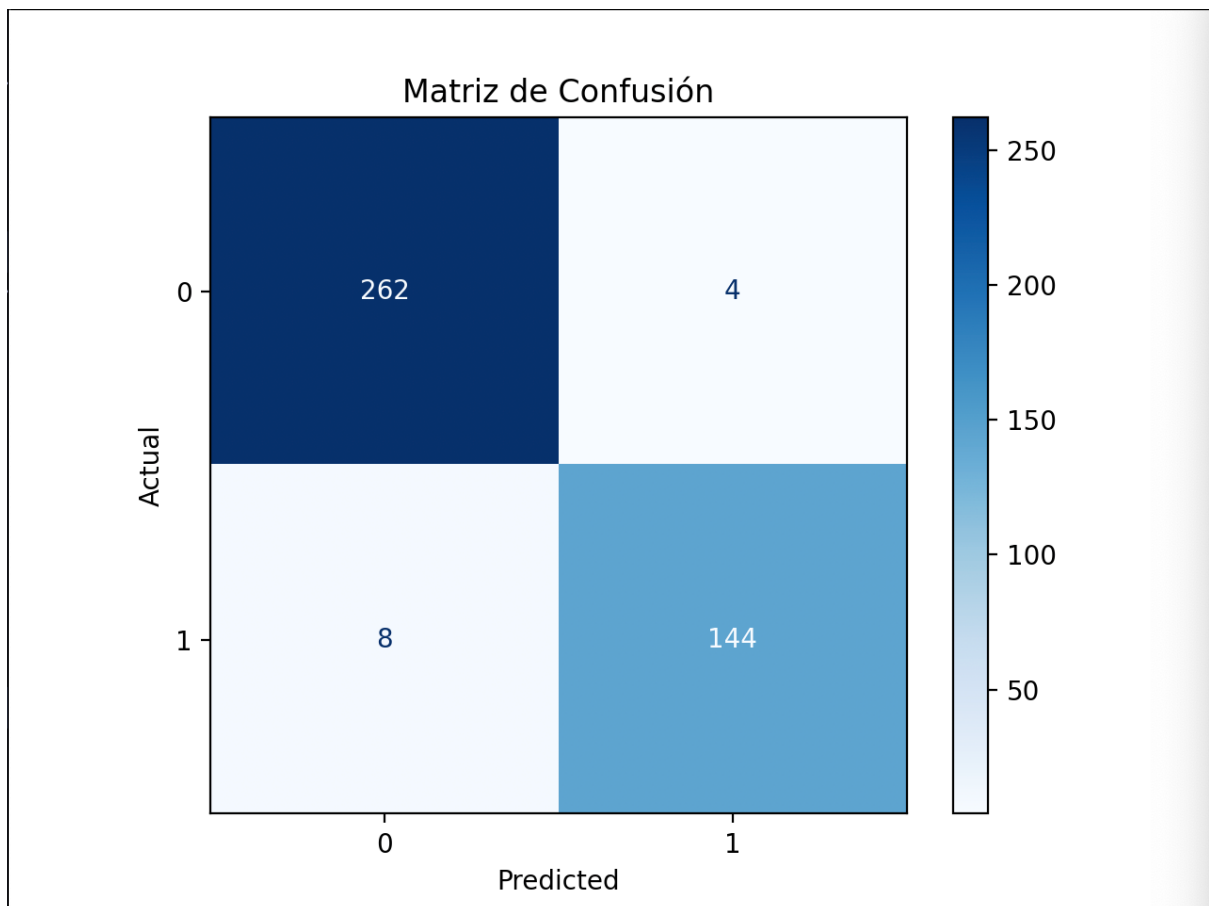
<https://www.kaggle.com/competitions/titanic/data>

Este dataset ya provee tanto datos de entrenamiento como datos de prueba, así como un archivo con las etiquetas de los datos de prueba para comprobaciones, utilizando estos datos se entrenó el modelo.

El modelo fue configurado de la siguiente manera para el primer intento:

```
model = RandomForestClassifier(n_estimators=100, max_depth=5, random_state=1)
```

Y se corrieron los tests, de estos tests salió esta matriz de confusión



Como podemos ver en esta matriz, el modelo es bastante preciso, teniendo pocos falsos negativos y poco falsos positivos comparativamente con la cantidad de datos que se están utilizando.

Además, el valor de accuracy global proporcionado por el mismo scikit-learn es del 97%, al igual que la precisión para la clase positiva.

El recall es del 95% y el F1-score es del 96% por lo que podemos decir que el modelo tiene una buena precisión y no tiene muchos falsos positivos realmente.