

1 Random forest

El random forest es una técnica de machine learning de uso común, este algoritmo combina el resultado de múltiples árboles de decisión para llegar a un resultado único.

Es una técnica fácil de interpretar, estable y que por lo general presenta buenas coincidencias y que se puede utilizar en tareas de regresión o clasificación.

Esta técnica de machine learning se utiliza ampliamente ya que reduce el riesgo de sobre ajuste, maneja datos grandes y con ruido, es escalable, versátil y menos sensible a datos atípicos.

2 Metodologia

El código del programa se ejecuto en un google collab, duplicando el notebook que el documento "AprendeMachineLearning" proporciona

Cuando los datos están listos para ejecutarse, usamos la siguiente función para entrenarlo con los datos dados

```
[12] from sklearn.ensemble import RandomForestClassifier

# Crear el modelo con 100 arboles
model = RandomForestClassifier(n_estimators=100,
                              bootstrap = True, verbose=2,
                              max_features = 'sqrt')

# entrenar!
model.fit(X_train, y_train)
```

building tree 1 of 100
building tree 2 of 100
building tree 3 of 100
building tree 4 of 100
building tree 5 of 100
building tree 6 of 100
building tree 7 of 100
building tree 8 of 100
building tree 9 of 100
building tree 10 of 100
building tree 11 of 100
building tree 12 of 100
building tree 13 of 100
building tree 14 of 100
building tree 15 of 100
building tree 16 of 100
building tree 17 of 100
building tree 18 of 100
building tree 19 of 100
building tree 20 of 100
building tree 21 of 100
building tree 22 of 100
building tree 23 of 100
building tree 24 of 100
building tree 25 of 100
building tree 26 of 100
building tree 27 of 100
building tree 28 of 100
building tree 29 of 100
building tree 30 of 100
building tree 31 of 100

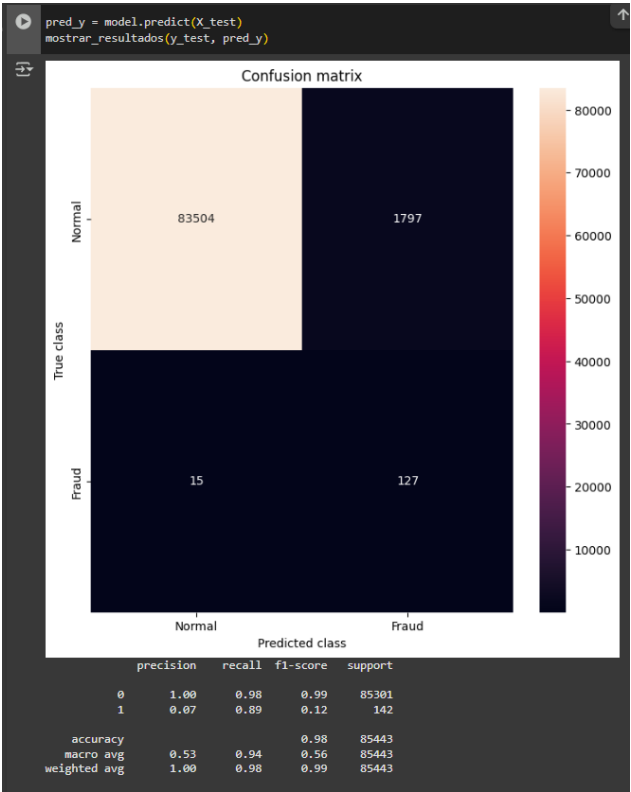
Después de esto se utiliza otro random forest mas veloz

```
[14] # otro modelo, variando hiperparametros
model = RandomForestClassifier(n_estimators=100, class_weight="balanced",
                             max_features = 'sqrt', verbose=2, max_depth=6,
                             oob_score=True, random_state=50)

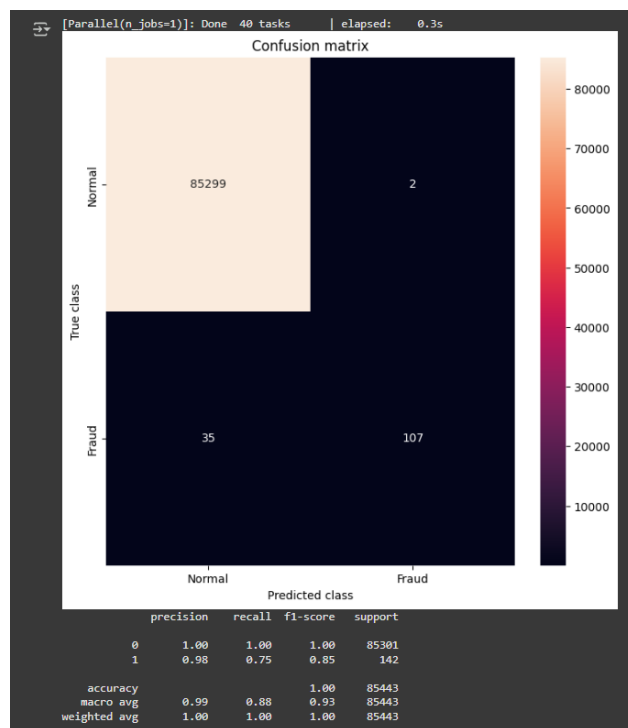
# a entrenar
model.fit(X_train, y_train)
```

3 Resultados

El test nos dio la siguiente matriz

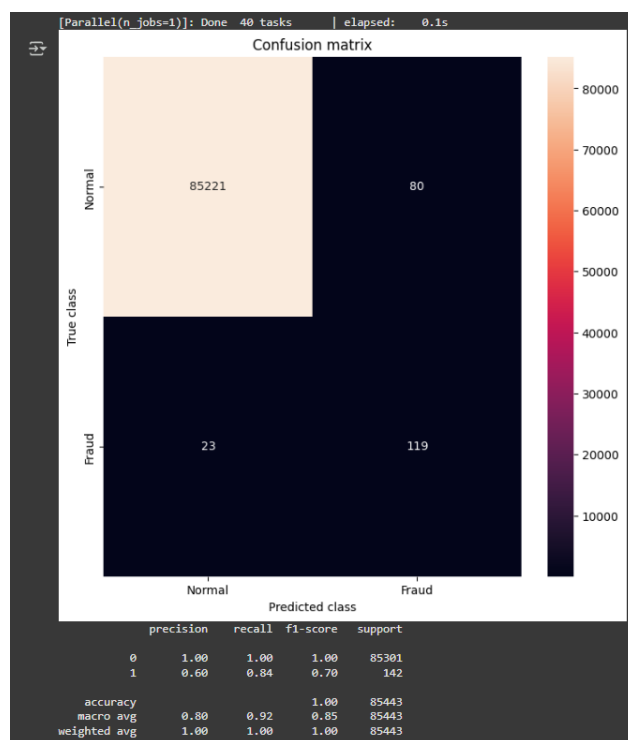


Se obtuvo un desempeño bueno considerando que es un test y que no se ha entrenado aun al modelo con el random forest El segundo modelo nos dio la siguiente matriz:



Este modelo tuvo un desempeño bastante mejor ya que clasifico menos normales como fraudes a comparacion del test.

Por ultimo con el ultimo modelado, usando un random forest mas rápido



Se obtuvieron resultados mejores que en el test, pero que no son tan buenos como los del segundo modelo.

4 Conclusión

Aprendí mas acerca de los random forest, como se utilizan y para que sirven, ya que aun que ya conocía cosas del machine learning, no sabia de los random forest, ademas con esta practica pude observar por mi mismo como cambia la eficacia del método cambiando la forma en la que el random forest trabaja con los datos.