

Random Forest

Valeria Rodríguez

30 de Marzo del 2025

1 Introducción

La metodología de Random Forest es un ensamble en Machine Learning donde se combinan diversos árboles y la salida de cada uno de ellos será un *voto*, y la opción más votada será la respuesta del Bosque Aleatorio. Este modelo es de aprendizaje supervisado para clasificación, aunque también suele ser usado para problemas de regresión.

El random forest surge debido al overfitting producido en ocasiones con un árbol de decisión, cuando a este se le da una profundidad suficiente para que este "memorice" las soluciones en vez de generalizar el aprendizaje. Es por ello que la solución surge en crear muchos árboles y que estos trabajen en conjunto.

2 Metodología

Para la realización de esta actividad se llevaron a cabo una serie de pasos encaminados a la recreación del problema de Random Forest presentado en el libro *Aprende Machine Learning* del autor Juan Ignacio Bagnato, páginas 92-96.

2.1 Creación de carpeta de trabajo

Se creó una carpeta llamada Random_forest, en la cual se copió el archivo csv y py de la actividad anterior de Regresión Logística.

2.2 Codificación

Con el archivo de la práctica anterior se recicló la parte de importación, lectura de archivo y selección de información en las distintas variables. Posteriormente, se creó el modelo de random forest.

```
model = RandomForestClassifier(  
    n_estimators=100,  
    bootstrap = True, verbose=2,  
    max_features = 'sqrt')  
# a entrenar!
```

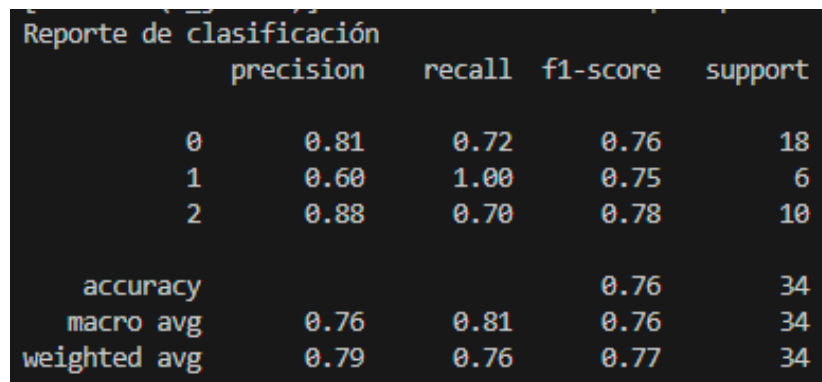
```
model.fit(X_train, Y_train)
```

Posteriormente, se evaluaron los resultados del Random Forest con un reporte de clasificación.

```
predictions=model.predict(X_validation)
print("Reporte de clasificación")
print(classification_report(Y_validation, predictions))
```

3 Resultados

El reporte de clasificación arroja un recall de 0.72, un valor bastante bueno para el modelo. Así bien, el F1-Score macro avg es de 0.76.



Reporte de clasificación				
	precision	recall	f1-score	support
0	0.81	0.72	0.76	18
1	0.60	1.00	0.75	6
2	0.88	0.70	0.78	10
accuracy			0.76	34
macro avg	0.76	0.81	0.76	34
weighted avg	0.79	0.76	0.77	34

Figure 1: Reporte de clasificación.

4 Conclusiones

Con la práctica de esta metodología se puso a prueba los múltiples árboles para evitar el overfitting en un árbol de decisión. Considero que fue una práctica relativamente sencilla. A comparación de otras actividades, en esta práctica no hubo complicaciones mayores a la hora de la codificación, por lo que no fue necesario hacer ajustes en el código proporcionado por el libro de trabajo.

5 Referencias

Bagnato, J. (2019). Aprende Machine Learning. Leanpub.
Materiales de clase (2025). UANL.