

Uso de framework o biblioteca de aprendizaje máquina para la implementación de una solución

Elisa María Bonilla Martín
A01028576

Instituto Tecnológico de Estudios Superiores de Monterrey
Campus Estado de México

1 Introducción

Random Forest es un algoritmo de aprendizaje supervisado. Para esta entrega decidí usarlo debido a que probé diferentes datasets y es un algoritmo flexible y fácil de utilizar. Básicamente el algoritmo crea árboles de decisión con datos que se seleccionan aleatoriamente.

2 Modelo

La primera base de datos que escogí es de la plataforma Kaggle y habla acerca de si un paciente muere debido a un problema cardíaco. La base de datos se encuentra en mi repositorio de github pero la liga original es la siguiente: <https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data>. La base de datos no contaba con valores nulos ni con variables categóricas, debido a esto fue bastante sencillo utilizarla. Es importante mencionar que el dataset lo dividí en training y test para probar la precisión del modelo creado.

Para la primera prueba utilicé los siguientes parámetros:

- `n_estimators = 100`
- `max_leaf_nodes = 10`
- `n_jobs = 1`
- `random_state = 50`
- `warm_start = True`

Y la precisión del modelo fue la siguiente:

```
Precisión: 0.8111111111111111
```

En la segunda prueba decidí eliminar la mayoría de los hiperparametros y solamente deje `n_estimators = 100`, la precisión del modelo fue la siguiente:

```
Precisión: 0.7888888888888889
```

Para la tercera y última prueba cambie la función que mide la calidad de un split en el algoritmo, en default es 'gini' pero la cambie a 'entropy'. La precisión resultante fue la más alta, con eso podemos llegar a la conclusión que modificar los parámetros si cambia la precisión del modelo, es importante ver todos los parámetros que se pueden modificar dentro del modelo ya programado.

```
Precisión: 0.8222222222222222
```