

### Taller Final Electiva Bd 3

La decanatura de la facultad de ingeniería en aras de aumentar la calidad de sus programas; pretende identificar potenciales estudiantes con mayor probabilidad de deserción debido a causas socioeconómicas.

La decanatura ha decidido contratarlos como ingenieros científicos de datos; para elaborar un modelo sobre la información histórica de los estudiantes de la facultad de ingeniería.

1. Su objetivo será presentar un modelo de aprendizaje automático que ayude a identificar estudiantes con mayor probabilidad de deserción.

Se le otorgaran 3 conjuntos de datos.

Conjunto	Descripción
Entrenamiento	Conjunto de datos para realizar el entrenamiento de su modelo
Prueba	Conjunto de datos para que realice pruebas y valide la capacidad predictiva de su modelo
Evaluación	Conjunto de datos para entregar clasificado con el modelo que elabore.

La métrica de evaluación a tener en cuenta será el área bajo la curva.

2. El conjunto de datos de evaluación deberá ser clasificado y enviado a la decanatura para su evaluación.
3. Su modelo deberá ser implementado en un servicio web para su posterior uso.

La decanatura ha decidido contratar a una consultora externa para que también elabore un modelo.

Modelo	Max_depth	InstancesPerNode	Impurit	MinInfoGain	Auc validation	Auc prueba	Auc eval
Árbol	20	10	entropy	0	0.95	0.96	0.93

Lo ideal será que el modelo que usted elabore supere la performance de evaluación AUC del modelo realizado por la consultora la cual es de 0.93.

Todo el proceso deberá ser elaborado bajo la herramienta Spark.

