

Práctica 6.5

Entrega: durante la clase

Descargue el conjunto de datos [nba.txt](#) y lea la descripción del mismo en la cabecera del archivo. Este conjunto de datos será usado en los siguientes puntos, los cuales deben ser resueltos usando [scikit-learn](#).

1. Entrene un modelo de clasificación basado en árboles de decisión.
 - (a) Haga una partición del conjunto de datos, usando muestreo estratificado, en 70 % para entrenamiento y 30 % para test (operador: Split Data).
 - (b) Entrene el modelo
 - (c) Aplique el modelo al conjunto de test
 - (d) Mida el desempeño del modelo calculando exactitud, error de clasificación, precisión, recall, curva ROC y matriz de confusión.
2. Interprete el modelo obtenido:
 - (a) Grafique el árbol obtenido
 - (b) Cuál es el atributo más discriminante? Tiene sentido? De una explicación a partir del conocimiento del problema.
3. Comparación de modelos:
 - (a) Usando los datos de la NBA, entrene un modelo random forest.
 - (b) Compare el modelo con el árbol de decisión.
 - (c) Cuál de los dos modelos es mejor?

El taller debe enviarse como un [Jupyter notebook](#) a través del siguiente [Dropbox file request](#), antes de la medianoche de la fecha límite. El archivo debe nombrarse como `isi-practica6.5-unalusername.ipynb`, donde unalusername es el nombre de usuario asignado por la universidad. No incluya archivos adicionales.