Análisis de datos con Python
Ejercicios Parte 2 - Clase 4 - UNSAM 2018
Árboles de desición - Random Forest - Medidas de Performance

1. Crear un DataFrame (df) con la base de datos Titanic.csv:
   (a) ¿Existen valores NANs en el df? ¿En qué variables aparecen? ¿De qué tipo son esas variables? (categóricas, continuas). Una opción sería eliminar los datos donde se presentan NANs. Otra opción es reemplazar los NANs por valores razonables y no perder ese dato. ¿Cómo completaría los NANs presentes en variables que indican categoría? ¿Y en los casos de las variables continuas? Realizar los cambios necesarios para obtener un df sin NANs y sin eliminar ninguna fila.
   (b) Crear una variable X con las columnas Pclass, Sex, Age, Fare, Embarked y una variable y con las columna Survived.
   (c) Crear $X_{train}, X_{test}, y_{train}, y_{test}$.
   (d) Realizar una clasificación con el algoritmo **DecisionTreeClassifier**.
   (e) Evaluar la exactitud, la precisión, recall y F_1-score de la clasifiación realizada.
   (f) Realizar la matriz de confusión. Utilizar la función **heatmap** dentro de la librería **seaborn** para plotear la matriz de confusión.
   (g) Dentro de **DecisionTreeClassifier** está la opción de crear clasificadores con una cantidad de niveles establecida o considerando solo cierta cantidad de atributos.
   (h) Realizar una nueva clasificación experimentando con distintas opciones de las características: **max depth** y **max features**. Comparar la nueva clasificación con la anterior.
2. Crear un DataFrame (df) con la base de datos **iris**:
   (a) Crear $X_{train}, X_{test}, y_{train}, y_{test}$.
   (b) Realizar una clasificación con el algoritmo **RandomForestClassifier**.
   (c) Evaluar la exactitud, la precisión, recall y F_1-score de la clasifiación realizada.
   (d) Realizar la matriz de confusión. Utilizar la función **heatmap** dentro de la librería **seaborn** para plotear la matriz de confusión.

(e) Dentro de **RandomForestClassifier** está la opción de crear cierta cantidad de arboles para el bosque y/o considerando solo cierta cantidad de niveles de cada árbol. Realizar una nueva clasificación experimentando con distintas opciones de las características: **n estimators** y **max depth**. Comparar la nueva clasificación con la anterior.