

Los archivos adjuntos que se proporcionan “train.csv” y “test.csv” son del dataset llamado “Cat in the Hat”. Una característica de este conjunto de datos es que contiene atributos categóricos. Usando este dataset realice lo siguiente:

1. Describir el dataset: qué describe, cuál es su objetivo, qué significan los atributos, cuál es el atributo clase, etc. Esta información se puede encontrar en [Cat in the Hat](#)
2. Realice el preprocesamiento necesario para tratar con los atributos categóricos de los archivos “train.csv” y “test.csv”. Se puede utilizar otro que conozcan pero hay que documentarlo y explicarlo en el reporte.
3. Utilice árboles de decisión de scikit-learn para clasificar los datos preprocesados. En este paso hay que tener en cuenta que los atributos en el dataset original son categóricos. Obtenga la precisión. Utilizar y comparar las funciones de división de nodo que vienen en la clase DecisionTreeClassifier().
4. Utilice algoritmo KNN de scikit-learn para clasificar los datos preprocesados. Obtenga la precisión.
5. Compare los resultados obtenidos con los dos algoritmos.

Entregables:

1. Reporte en pdf.
2. Archivo python clasificacion.py

SIN COMPRIMIR