Aula6 – Clustering.

1. Com o dataset ***iris*** demonstre o desempenho do algoritmo k-means nestes dados. Assuma três clusters e corra o algoritmo com o atributo objetivo “escondido”.

Asar a aba dos clusters, e não dos classifiers.

Usar a opção ***ignore attributes*** para ignorar as informações relativas ás classes, senão o algoritmo sabe rapidamente o número de clusters ideal.

1. Considere o dataset ***glass*** e compare os resultados de um processo de clustering obtidos dos algoritmos k-means e EM.
2. Apresente um dendrograma para este dataset e refira possíveis configurações de agrupamentos, identificação de outliers, etc.

Utilizar o clusterer hierarchial clusterer para visualizar o dendrograma.

1. Apresente um método para calcular o valor de k óptimo no algoritmo k-means. Ensaie esse método no dataset ***glass*** (sem atributo objetivo) e indique o valor de k obtido.

5) Considere o dataset ***Kaggle-employees*** (a descarregar do blackboard). Desenvolva um modelo de previsão para identificar os empregados que deixam esta empresa.

Deve estudar e selecionar os atributos adequados ao processo.

Apresente resultados para diferentes algoritmos com validação cruzada.

Apresente um estudo com métodos simples contra métodos de ensembles sobre esses métodos simples e.g. NaiveBayes versus AdaBoost sobre naiveBayes.