

Curso: Mestrado em Engenharia Biomédica
U.C.: Aprendizagem e Extração do Conhecimento

Ficha de Exercícios 07	
Docente:	Hugo Peixoto José Machado
Tema:	RapidMiner: k-Means Clustering
Ano Letivo:	2023-2024 – 1º Semestre
Duração da aula:	2 horas

1. Introdução

- 1] O que significa k-Means clustering?
- [2] Como se identificam os clusters e qual o processo que o rapidMiner usa para definir e colocar as observações num determinado cluster?
- [3] O que revela a Centroid Table ao utilizador? Como se interpretam os valores nessa tabela?
- [4] Depois do exercício introdutório pensar num problema que possa ser resolvido agrupando observações em clusters. Procurar na internet um dataset que possa ser utilizado e aplicado um modelo de k-Means.
 - [a] Garantir que os dados estão no formato CSV e importar os mesmos para o RapidMiner;
 - [b] Fase de preparação dos dados. Pode incluir componentes de inconsistência de dados, missing values, ou alteração do tipo de dados;
 - [c] Ligar um operador de k-means clustering ao dataset e alterar os parâmetros de acordo com a necessidade (sobretudo o k, para adequar ao problema em questão);
 - [d] Avaliar a Centroid Table, Folder View, e outras ferramentas de avaliação;
 - [e] Reportar todos os passos anteriores e as evidências encontradas. Discutir as iterações no modelo, e de que forma o que foi encontrado permite responder ao problema inicial.
- [5] Experimentar o mesmo dataset com diferentes operadores de k-Means como o Kernel ou Fast. Em que medida diferem do modelo original. Estes operadores mudam os clusters originais? Se sim, em que medida?