

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

| Ficha de Exercícios 09 | |
|-------------------------|--------------------------------|
| Docente: | Hugo Peixoto José Machado |
| Tema: | RapidMiner: k-Means Clustering |
| Ano Letivo: | 2024-2025 – 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. Encontrar um dataset online, e executar os seguintes passos:

[a] Preparação dos dados: Selecionar apenas os atributos numéricos e lidar com os valores nulos.

[b] Normalização: Normalizar os dados para garantir que todos atributos contribuem de igual forma para o clustering.

[c] Aplicar K-Means: Usar o K-Means da biblioteca sklearn para executar o clustering.

[d] Visualização: Fazer o plot e interpretar os dados.