

## Ciência da Computação


### Exercícios de Inteligência Artificial

Cristiane Neri Nobre

Data de entrega: 28/03

1. Faça um resumo do artigo **“Balanceamento de dados”** que está no SGA
2. Faça um resumo do artigo **“Estudo Comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de Decisão”** que está no SGA
3. Faça um resumo do artigo **“A Systematic Review on K-Means Clustering Techniques”** que está no SGA
4. Faça um resumo do artigo **“A Survey on Clustering Techniques for Big Data Mining”** que está no SGA
5. Cite e explique os principais parâmetros que podem ser ajustados para melhorar o desempenho do classificador Random Forest. Quais os valores default destes parâmetros na ferramenta WEKA? Explique cada um deles.
6. Considere o seguinte conjunto de 8 exemplos de pares de pontos no plano real, Dados os exemplos indicados na tabela abaixo, caracterizados por dois atributos (pontos em  $R^2$ ), aplique o algoritmo *Kmeans* para determinar uma partição adequada desses dados em 2 grupos. Realize a primeira execução do algoritmo, partindo dos exemplos 3 (Ex3 -> X=0 e Y=2) e 4 (Ex4 -> X=0 e Y = -2). Utilize a distância euclidiana para o cálculo de distância.

Exemplo	X	Y
Ex1	-5	2
Ex2	5	-2
Ex3	0	2
Ex4	0	-2
Ex5	-5	1
Ex6	-5	-2
Ex7	5	2
Ex8	5	-1

7. Cite e explique o funcionamento de **duas métricas** para avaliação de agrupamento.
8. Abra a base de dados Iris.arff no WEKA (ou em qualquer outra ferramenta de aprendizado de máquina) e rode o Kmeans. A partir disso, pede-se:
  - a) Veja visualmente os grupos obtidos. O que aconteceu com as instâncias? Houve erro de agrupamento? Por quê?
  - b) Agora marque a opção  **Classes to clusters evaluation** e rode novamente o Kmeans. Avalie novamente os agrupamentos encontrados. Comente sobre os resultados encontrados.
9. Assista o vídeo:

[https://www.youtube.com/watch?v=E2M\\_yTulcmU](https://www.youtube.com/watch?v=E2M_yTulcmU)

E analise as limitações deste algoritmo