



Treinamento CIS - 3º Período (Clusterização)

1. Conteúdos:

- a. Introdução a Clusterização;
 - i. Conceitos básicos e aplicações;
 - ii. Modelos;
 - iii. Diferença entre classificação e clusterização;;
- b. Algoritmos de clusterização;;
 - i. K- means;
 - ii. Clusterização hierárquica;
 - iii. DBSCAN (Density-Based Spatial Clustering of Applications with Noise);
- c. Validação e ajuste de modelos;
 - i. Técnica da curva do cotovelo;
 - ii. Silhouette Score;
 - iii. Coesão e Separação em Clusters;

2. Materiais:

- a. [Introdução](#) - Introdução a Clusterização;
- b. [Vídeo explicando](#) - Como aplicar o algoritmo;
- c. [Vídeo explicando](#) - Usar o K-Means para Clusterizar (Agrupar os Dados);
- d. [Vídeo explicando](#) - Centroids;
- e. [Artigo explicando](#) - K- Means ++

3. Materiais para quem quer avançar:

- a) [Alura](#) - Validação e ajuste de modelo;

4. Tarefas:

Para esse período, NÃO TERÁ competição no Kaggle:

1. Visualize o scatterplot em matriz com hue = "Gender". Dá para notar alguma diferença? Se sim, qual e por quê?
2. Crie um DataFrame apenas com colunas Annual Income e Spending Score.
3. Import o K-means.
4. Crie o modelo em cluster padrão, sem especificar a quantidade K de clusters.
5. Crie uma curva de cotovelo. Qual seria o número ideal de clusters?
6. Crie novamente o cluster com kmeans com a quantidade de n_clusters ajustada. Compare os resultados e comente.
7. Quais foram os centroids? Imprima os centroids
8. Mostra o gráfico do resultado com os centroids
9. Analise o resultado. Para qual grupo você direcionaria com maior prioridade a campanha de marketing?



Universidade de Brasília Faculdade de
Tecnologia, Departamento de Engenharia
Elétrica