

Trabalho Prático II - Individual

1. Objetivo Geral

Aplicar diferentes o algoritmo de agrupamento K-means em uma base de dados escolhida pelo professor, buscando encontrar a melhor configuração de parâmetros no cenário explorado, bem como gerar visualizações que expliquem as descobertas ou facilitem o entendimento destas.

2. Objetivos Específicos

2.1 - Explorar o *dataset* "Turkiye Student Evaluation Data Set":

- Dataset disponível em: <http://archive.ics.uci.edu/ml/datasets/turkiye+student+evaluation>
- Analisar os dados deste *dataset*, utilizando ferramentas apresentadas ou não em aula;
- Realizar pré-processamento caso necessário;
- Gerar visualização/métricas que facilitem o entendimento ou expliquem diferentes insights presente nesses dados;

2.2 – Implementar o K-Means, testando diferentes parâmetros;

2.3 – Gerar um breve relatório dos experimentos.

3. Entregáveis

3.1 – Notebook com os códigos dos experimentos contendo:

- Análises/visualizações do dataset (caso tenham utilizado jupyter notebook);
- Implementação dos algoritmos presentes no item 2.2;
- "Tunagem" de parâmetros (quando aplicável) do algoritmo presente no item 2.2;

3.2 – Relatório (em pdf) sobre as análises, testes e resultados;

4. Avaliação

4.1 – O presente trabalho possui um peso de **30%** no grau final da disciplina (os outros 70% referem-se ao Trabalho Prático I);

4.2 – A data de entrega do presente é, **impreterivelmente**, até às 23:55h do dia 30 de Setembro. Envios tardios serão desconsiderados;

4.3 – O peso de cada item avaliado é apresentado na tabela abaixo:

| Requisito | Peso |
|------------------------------|-------------|
| Exploração do <i>dataset</i> | 2,00 |
| Implementação K-Means | 1,50 |
| Tunagem de Parâmetros | 1,50 |
| Visualizações | 1,50 |
| Relatório | 2,00 |
| Criatividade | 1,50 |
| Total | 10 |

UniRitter – Pós-Graduação em Big Data e Data Science
Machine Learning I – Professor Willian Becker
09/09/2019 - Porto Alegre - RS

5. Considerações Finais

5.1 – O desenvolvimento dos algoritmos deverá ser feito em Python, mas para a visualização/exploração do *dataset* pode-se utilizar outras ferramentas;

5.2 – Para a entrega do trabalho será aberta uma sala de postagem no Blackboard. Todos os entregáveis devem ser entregues em uma pasta no formato “zip”, contendo o primeiro nome e o último sobrenome do aluno (ex: “WillianBecker.zip”);