

TRABALHO PRÁTICO PRIMEIRA UNIDADE

Entrega: 07/07/2021 – Horário: 23:59 – Local: via SIGAA

- 1- Escolha uma base de dados que melhor represente o seu interesse nesta disciplina (essa base de dados será utilizada nas próximas atividades desta disciplina). A base de dados deve ser uma escolha pessoal do aluno. Sendo assim, o aluno está livre para escolher uma base de dados de um assunto que o interesse. Nos repositórios <http://archive.ics.uci.edu/ml/>, <https://brasil.io/datasets/>, <https://dados.gov.br/>, <http://dados.ufrn.br/>, <https://datasetsearch.research.google.com/>, <https://basedosdados.org/>, <https://www.quandl.com/> e <https://www.kaggle.com/datasets>, o aluno pode encontrar vários exemplos de bases de dados (fica livre para escolher de outros repositórios). Diante da base de dados escolhida, responda às seguintes questões, **apresente-as no momento síncrono e anexe, no SIGAA, um relatório com a descrição do seu trabalho:**
 - a. Descreve o problema da base de dados escolhida (o objetivo).
 - b. Quantos atributos e quantas instâncias possui a sua base de dados?
 - c. Descreva a caracterização dos dados (tipo, escala)?
 - d. A base de dados escolhida contém atributos faltosos? Se sim, quantos? Qual o pré-processamento realizado diante deste problema?
 - e. É necessário normalizar a base de dados? Justifique sua resposta.
 - f. É necessário realizar a eliminação manual de atributos, integração de dados, amostragem de dados, balanceamento de dados, limpeza de dados, redução de dimensionalidade e transformação de dados? Descreva cada técnica utilizada.
- 2- É possível utilizar alguma técnica de visualização de dados antes da etapa de pré-processamento de dados? Se sim, qual/quais? Justifique sua resposta.

Bom trabalho!!!