



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
CAMPUS SENADOR HELVÍDIO NUNES DE BARROS – CSHNB
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO
DISCIPLINA DE TÓPICOS EM COMPUTAÇÃO

Atividade Prática I

As bases de dados Adult (Census Income)¹ e Titanic² são amplamente utilizadas em estudos de análise e reconhecimento de padrões, pois apresentam diferentes desafios no pré-processamento e análise. A base Adult contém informações socioeconômicas e ocupacionais, com problemas como valores ausentes, variáveis categóricas, desbalanceamento e necessidade de normalização. A base Titanic reúne dados históricos e pessoais dos passageiros do navio, com variáveis categóricas e numéricas, presença de valores faltantes e possibilidade de extração de novas variáveis (feature engineering). Diante disso, esta atividade tem como objetivo realizar uma análise exploratória e de pré-processamento dessas bases, desenvolvendo um pipeline de preparação de dados e análise de padrões.

1. Preparação de Dados:

- Faça o download do arquivo da base de dados e carregue os dados em Python, por exemplo, usando a biblioteca pandas.

2. Pré-processamento:

- Inspecione e trate dados valores ausentes ou inconsistências.
- Detecte e trate possíveis outliers.
- Realize a conversão de variáveis categóricas para numéricas.
- Normalize os dados para adequar à escala das variáveis.

3. Visualizações:

- Crie uma tabela de sumarização da base, apresentando estatísticas descritivas (média, mediana, variância e desvio padrão) para cada atributo.
- Produza visualizações em 2D, utilizando gráficos de dispersão, histogramas, boxplots e gráficos de barras agrupadas, de forma a destacar relações entre variáveis correlacionadas.

4. Análise de Padrões e Tendências:

- Adult (Census Income): verifique como variáveis como idade, escolaridade, horas de trabalho e ocupação influenciam na renda.
- Titanic: identifique padrões relacionados à sobrevivência dos passageiros, considerando variáveis como idade, gênero e classe.



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
CAMPUS SENADOR HELVÍDIO NUNES DE BARROS – CSHNB
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO
DISCIPLINA DE TÓPICOS EM COMPUTAÇÃO

- Em ambas as bases, explore correlações entre as variáveis numéricas e a variável alvo (classes).

5. Relatório Final:

- Elabore um relatório (utilizando o template da SBC) contendo:
 - a) Introdução, desenvolvimento e conclusão.
 - b) Descrição detalhada das etapas e métodos aplicados.
 - c) Gráficos, tabelas e resultados obtidos durante a análise.
- Submeta, via SIGAA, o relatório e o código-fonte com todo o processo descrito.
- Prazo de envio: até 04/09.

¹ <https://www.kaggle.com/datasets/wenruihu/adult-income-dataset>

² <https://www.kaggle.com/datasets/yasserh/titanic-dataset>