



Instituto Federal de Educação, Ciência e Tecnologia

Câmpus Campinas

D1AED – Análise Estatística para Ciência de Dados

Professores: Ricardo Sovat (sovat@ifsp.edu.br)
Samuel Martins (samuel.martins@ifsp.edu.br)

Atividade em Dupla 2

1. Especificação

Nesta atividade, cada dupla deverá treinar um regressor linear para prever o consumo de cerveja da cidade de São Paulo, a partir do dataset *Beer Consumption - Sao Paulo*: <https://www.kaggle.com/dongearge/beer-consumption-sao-paulo>. O protocolo dos experimentos (p. ex., tamanho dos conjuntos de treinamento e teste, atributos escolhidos, etc) ficam a cargo da dupla.

Seguem os critérios a serem avaliados. Cada critério tem um conjunto de pontos que servirão como um guia para seu desenvolvimento. Outros pontos não mencionados aqui também podem ser considerados.

- **Análise exploratória simples [3 pontos]**
 - Analise os atributos do dataset por meio de gráficos, estatísticas descritivas, etc;
 - Tal análise visa apenas dar uma noção geral das principais características do dataset, logo não precisa ser algo muito complexo;
 - P. ex:
 - ◆ Como cada atributo é distribuído?
 - ◆ Existe alguma correlação entre atributos?
- **Protocolo de Experimentos: [1 ponto]**
 - Como dividiram a base em treinamento e teste?
 - Quais atributos foram considerados?
 - Houve a necessidade de pré-processar os dados?
 - Quais são as métricas para a avaliação dos resultados?
- **Códigos [3 pontos]**
 - Preparação da base de dados;
 - Divisão da base de dados;
 - Treinamento do modelo linear;
 - Predição dos resultados;
 - Cômputo das métricas, gráficos, etc;
- **Discussão dos resultados: [2 pontos]**
 - Forneça uma discussão sobre os resultados obtidos;
 - Como seu modelo se comportou? Ele forneceu previsões acuradas?
 - Que conclusões você toma a partir de tais resultados?
- **Relatório (Notebook): [1 ponto]**
 - Organização do relatório;
 - Clareza na apresentação dos textos e códigos;
 - Qualidade do código;

2. Submissão (prazo: 15/07/21)

A submissão desta atividade será feita em tarefa específica no Moodle da disciplina. A dupla poderá enviar (**apenas um membro**) um jupyter notebook (.ipynb) ou o link do repositório online com o código (ex., Google Colab, GitHub, Kaggle). No caso dos links para repositórios ou plataformas online, serão considerados apenas aqueles com atualização até o prazo de entrega desta atividade.

3. Material de Apoio

Dataset:

- Beer Consuption – Sao Paulo:
<https://www.kaggle.com/dongedorge/beer-consumption-sao-paulo>

Visualizações:

- <https://datavizcatalogue.com/index.html>
- <https://www.python-graph-gallery.com/>