

Disciplina: Inteligência Artificial – BCC IBILCE/UNESP 2025

Prof. Dr. Wallace Casaca <wallace.casaca@unesp.br>

Monitor: Thiago Amorim de Faccio <thiago.faccio@unesp.br>

<http://tiny.cc/iaunesp2025>

Trabalho-Prova (P1)

O Trabalho-Prova (P1) contém questões teóricas cujas respostas devem ser entregues, via Google Classroom, juntamente com os códigos da resolução prática de cada questão (as respostas podem ser apresentadas em um único arquivo Notebook).

1. (1.0 ponto) Encontre no Kaggle (<https://www.kaggle.com>) ou repositório equivalente na internet uma base de dados que possa ser explorada utilizando: (a) Métodos de Aprendizado de Máquina para Classificação, ou ainda (b) Métodos de Aprendizado de Máquina para Regressão. Insira o link da base escolhida no arquivo público da turma postado no Classroom. Explicar os atributos de entrada, o atributo-meta (saída), e sua motivação para trabalhar com o problema de classificação (ou regressão) escolhido.
2. (6.0 pontos) Empregar três algoritmos de Aprendizado de Máquina (AM) nos conjuntos de dados selecionados na Questão 1 (algoritmos vistos ou não vistos em aula, à sua escolha). Siga as instruções:
 - a) Conduza as etapas necessárias de pré-processamento do *dataset* escolhido visando “preparar o dado” para a aplicação dos algoritmos de AM. Forneça detalhes dos processos de pré-processamento que utilizou.
 - b) Aplique os três algoritmos nos dados pós-processados. Varie os hiperparâmetros dos algoritmos, nos casos pertinentes, de forma a melhor ajustar os modelos, melhorando assim os resultados.
 - c) Para quem optou por trabalhar com a aplicação de Classificação: (i) apresente os resultados obtidos utilizando ao menos duas métricas de avaliação; exiba também a Matriz de Confusão. (ii) Aplique alguma(s) ferramenta(s) de visualização (escolha livre) para analisar os dados. (iii) Compare os resultados obtidos pelos três algoritmos.
 - d) Para quem optou por trabalhar com a aplicação de Regressão: (i) apresente os resultados obtidos utilizando ao menos duas métricas de avaliação, como o Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), entre outras de sua escolha. (ii) Aplique alguma(s) ferramenta(s) de visualização (escolha livre) para analisar os dados. (iii) Compare os resultados obtidos pelos três algoritmos.
3. (3.0 pontos) A partir da base de dados Wine Quality, disponível no link (<https://archive.ics.uci.edu/ml/datasets/wine+quality>), responda:
 - a) Análise de Outliers: Utilize boxplots para analisar a presença de outliers em pelo menos duas variáveis contínuas da base de dados. Esses outliers podem influenciar as previsões de modelos de aprendizado de máquina?
 - b) Gere uma matriz de correlação entre as variáveis da base de dados. Quais atributos têm maior correlação com a qualidade do vinho?
 - c) Algumas variáveis possuem escalas muito diferentes entre si. Escolha uma técnica de normalização (Min-Max Scaling ou Padronização) e aplique a pelo menos três variáveis. Explique como essa transformação pode influenciar o desempenho de um algoritmo de aprendizado de máquina.