

Trabalho prático

Objetivo: Desenvolver análise e modelagem de dados utilizando técnicas de aprendizado de máquina.

Os grupos devem apresentar:

1. **Problema prático** e objetivos da modelagem: classificação, regressão, previsão de séries temporais, processamento de linguagem natural, agrupamentos, ranqueamento de dados, entre outros.
2. **Etapas de preparação dos dados.** Exemplo: Pré-processamento dos dados, transformação de variáveis, identificação de pontos aberrantes, entre outros.
3. **Modelagem dos dados** utilizando pelo menos três métodos e produzindo comparação de desempenho entre eles, utilizando por exemplo algoritmos como Naive Bayes, KNN e extensões, árvore de classificação ou regressão, *bagging*, *boosting* e florestas aleatórias, análise discriminante, modelos de regressão, modelos de suporte vetorial, redes neurais artificiais, análise de cluster, análise de componentes principais, análise de correspondência, etc. Você deve justificar o uso das técnicas escolhidas e compará-las.
4. **Modelo escolhido e discussão de resultados.**

Valor atribuído: 30% da média final, conforme definido no início do curso.

Entrega: via e-disciplinas em formato zip com arquivos como relatório de análises, bancos de dados, comandos, etc. Pode ser substituído por link de um repositório github com todos os documentos. Na documentação do trabalho (relatório ou notebook) deve constar o nome completo de todos(as) os(as) alunos(as) que o desenvolveram.

Escolha dos dados: Cada grupo deve escolher um banco de dados. Caso algum grupo não defina, deve pedir à docente a sugestão de uma base de dados antes da data final para definir o tema.

Detalhes da execução: O trabalho deve ser feito em grupos de até **4 alunos**. A apresentação dos trabalhos será feita no dia 28 ou 29 de junho de 2023, sendo que cada grupo terá de 5 a 7 minutos para fazer a sua apresentação e responder a perguntas e comentários dos(as) colegas e docente. Caso algum(a) aluno(a) não queira participar da apresentação, deve se manifestar formalmente com a docente por e-mail com antecedência de pelo menos duas semanas.

Os temas dos trabalhos deverão ser definidos até o dia **31/05/2023** em planilha disponibilizada no e-disciplinas. O sorteio da ordem dos grupos será dia **07/06/2023**. A entrega final de todos os grupos será até o dia **05/07/2023** às 23h59 pelo e-disciplinas.