

Aprendizado de Máquina e Mineração de Dados 2025.2
Prof. Leonardo Rocha

Trabalho Final de Aprendizado de Máquina e Mineração de Dados

Objetivo:

Desenvolver um projeto completo de Aprendizado de Máquina ou Mineração de Dados aplicando todas as etapas do pipeline, inclusive a geração de artefatos reutilizáveis e o model serving, em um dataset real de médio porte..

Orientações:

- A entrega deve ser composta por **dois elementos obrigatórios**:

1. Repositório GitHub organizado

- README.md claro e bem estruturado, incluindo:
 - descrição resumida do problema e da solução desenvolvida;
 - organização das pastas do repositório;
 - instruções para executar o código e carregar/rodar os modelos localmente;
 - dependências necessárias (requirements.txt).
- Código-fonte completo, incluindo:
 - código de treinamento e exportação dos modelos;
 - implementação da API para servir os modelos (model serving);
 - arquivos auxiliares necessários para execução (ex.: pré-processamento, configuração, utilitários).
 - Notebooks utilizados no desenvolvimento (se houver).
- Modelos exportados, preferencialmente em um dos formatos padrão:
 - .onnx, .pt, .h5 ou equivalente.

2. Relatório técnico em PDF

- Descrição da modelagem e etapas da implementação;
 - Resultados obtidos (valores, gráficos, análises);
 - Avaliação crítica da solução (limitações e possíveis melhorias);
 - Instruções para reprodução dos experimentos, se necessário.
-
- Podem ser formadas equipes de até 2 pessoas.
 - Uso correto de IA generativa: permitido apenas como apoio (ex: explicações, depuração pontual).

- Geração automática de código ou relatório sem domínio do conteúdo será considerada plágio.

Atividades:

1. Escolha um dataset real (i.e. não sintético) e uma tarefa de aprendizado de máquina, e preencha o formulário para validação pelo professor (até 26/11):
 - <https://forms.gle/3bDg4KWi5pLL2Mw37>
2. Executar o pipeline completo já visto em sala:
 - Tratamento dos dados
 - Análise exploratória
 - Engenharia de features
 - Treinamento de Modelos
 - Validação e comparação entre diferentes modelos e hiperparâmetros
3. Exportação dos melhores modelos (.onnx, .pt, .h5 ou similares)
4. Implementação de uma API para servir os modelos (sugestão: usar FastAPI e onnx Runtime).
5. Organizar resultados no repositório conforme orientações, garantindo que o trabalho possa ser reproduzido e testado facilmente.