

Trabalho Final – MLOps na Prática

Introdução

O objetivo deste trabalho é aplicar conceitos e práticas de **MLOps** para desenvolver um pipeline de **Machine Learning** funcional e automatizado. Os grupos (no máximo 3 pessoas) deverão explorar um conjunto de dados real, implementar modelos preditivos e integrar o processo com ferramentas de monitoramento, versionamento e deploy.

O foco do projeto está na **construção de um fluxo completo**, contemplando desde a preparação dos dados até a entrega do modelo em produção, garantindo rastreabilidade e reprodutibilidade.

Dataset e Problema

Os grupos deverão trabalhar com um conjunto de dados estruturado e escolher um problema de **classificação ou regressão**. Sugestões de datasets:

- **Churn Prediction**: prever a saída de clientes de uma empresa.
- **Fraude em Cartões de Crédito**: identificar transações fraudulentas.
- **Previsão de Preço de Imóveis**: estimar o valor de um imóvel com base em suas características.

Os grupos poderão utilizar datasets públicos ou outros acordados previamente.

Requisitos Técnicos Sugeridos

- **Linguagem**: Python 3.10
- **Ferramentas sugeridas**:
 - **MLflow** (para rastreamento e versionamento de modelos)
 - **Evidently AI** (para monitoramento de drift)
 - **FastAPI/Flask** (para disponibilização do modelo via API, (ex. usando MLflow))
 - **GitHub/GitLab** (para controle de versão)

Etapas do Projeto (MLOps Pipeline)

1. Exploração e Pré-processamento dos Dados

- Análise exploratória e tratamento de valores ausentes.
- Normalização/Padronização dos dados quando necessário.

2. Treinamento e Avaliação do Modelo

- Implementação de pelo menos **dois modelos** e comparação de métricas.
- Utilização do **MLflow** para rastrear experimentos.

3. Versionamento e Armazenamento do Modelo

- Registro do modelo no **MLflow Model Registry**.

4. Implantação do Modelo

- Construção de uma API com **FastAPI** ou **Flask** para servir previsões (mlflow).
- Deploy local ou em nuvem (ex: AWS, GCP, Azure).

5. Monitoramento e Re-treinamento

- Implementação de monitoramento de drift de dados com **Evidently AI**.
- Definição de uma estratégia para re-treinamento automático do modelo.

6. Containerização e Documentação

- Instruções de execução/documentação do pipeline no repositório.

Entregáveis

- **Código-fonte** em um repositório Git (GitHub/GitLab) contendo:
 - Pipeline de dados e treinamento.
 - Código da API para inferência.
 - Scripts de monitoramento e re-treinamento.
 - Arquivos de configuração.
- **Relatório (.DOCX ou .PDF) explicando:**
 - Escolha do dataset e problema abordado.
 - Metodologia e ferramentas utilizadas.
 - Resultados e métricas dos modelos.
 - Fluxo completo do pipeline e considerações finais.

Considerações Finais

Esse projeto visa consolidar os aprendizados de **MLOps**, permitindo que os grupos desenvolvam um fluxo de Machine Learning robusto e aplicável ao mercado. A criatividade na solução, boas práticas de código e documentação serão consideradas na avaliação.

Prazo final de entrega: 21/04

Qualquer dúvida, entrem em contato com o professor.