
Prova III – Programação para Dispositivos Móveis – DSM - Prof. Álvaro Gonçalves - 08/07/2025

Nome: _____

Valor: 10 pts.

Criação de README explicativo para o Pipeline de Treinamento de Modelo de PLN

Você implementou um pipeline completo de classificação de textos (fake news, sentimento ou similar) usando Processamento de Linguagem Natural (PLN) e aprendizado de máquina. Agora, seu desafio é elaborar um arquivo README.md para o repositório do seu projeto, explicando detalhadamente cada etapa do processo de treinamento do modelo.

Tarefa

Crie um arquivo README.md para o projeto, contendo as seguintes seções (com exemplos, se necessário):

1. Coleta e Limpeza dos Textos

- Explique como os dados/textos podem obtidos (origem do dataset, formato, volume).
- Descreva o processo de limpeza dos textos (remoção de pontuação, stopwords, conversão para minúsculas etc).
- Se possível, apresente um exemplo de texto antes e depois da limpeza.

2. Vetorização dos Textos

- Explique o que é vetorização de texto e por que ela é necessária para modelos de ML.
- Descreva como pode ser aplicada a vetorização (TF-IDF, embeddings etc).
- Mostre exemplos de como um texto é transformado em vetor/matriz.

3. Treinamento do Modelo

- Explique como o dataset foi dividido em treino/teste.
- Detalhe qual(is) algoritmo(s) foi(foram) utilizado(s) (Naive Bayes, Random Forest, BERT etc.), com breve descrição do funcionamento.
- Mostre como o modelo foi ajustado e treinado.

4. Avaliação dos Modelos

- Explique o significado das métricas de avaliação (acurácia, precisão, recall).
- Mostre como cada métrica foi calculada.
- Apresente os resultados obtidos e uma breve análise dos mesmos.

Entrega:

O arquivo README.md pronto para ser publicado no repositório GitHub do seu projeto, contemplando todos os itens acima.

Critérios de Avaliação:

- Clareza e organização das explicações.
- Detalhamento técnico e exemplos.
- Capacidade de tornar o projeto compreensível para novos desenvolvedores ou avaliadores.