

PIA PROJETO

TECNOLOGIAS E SISTEMAS DE INFORMAÇÃO PARA A WEB

PROJETO

Subject

Vinho Inteligente: Modelagem Preditiva com Machine Learning

Objectives and brief description

Descrição

Este projeto cruza o mundo da enologia com o do Machine Learning. Utilizando um conjunto de dados sobre vinhos, o objetivo é demonstrar as competências adquiridas ao longo da Unidade Curricular. Para tal, deverão ser desenvolvidos dois modelos de machine learning: um modelo de classificação para prever o tipo de vinho (branco ou tinto) e um modelo de regressão para estimar a qualidade do vinho baseada em suas propriedades químicas.

O projeto deve incluir uma análise completa do conjunto de dados fornecido, envolvendo diversas técnicas de exploração, pré-processamento, aplicação de algoritmos de machine learning, avaliação/otimização de resultados e sua discussão.

Objetivos

Classificação do Tipo de Vinho

Explorar o dataset e realizar uma análise exploratória dos dados (EDA) para entender as características que diferenciam os vinhos brancos dos tintos.

Desenvolver e treinar um modelo de classificação que utilize algoritmos como Árvores de Decisão, Random Forest ou SVM para prever corretamente o tipo de vinho.

Avaliar o desempenho do modelo utilizando métricas apropriadas como precisão, recall e F1-score.

Regressão da Qualidade do Vinho

Investigar quais fatores químicos influenciam mais na qualidade percebida do vinho.

Construir e treinar um modelo de regressão usando métodos como Regressão Linear, Random Forests e XGBoost para prever a qualidade do vinho.

Medir a acurácia do modelo através do erro quadrático médio (MSE) e o coeficiente de determinação (R^2).

Dataset

O dataset fornecido contém informações sobre diferentes amostras de vinho, incluindo dados como acidez, teor alcoólico, quantidade de açúcar, pH, entre outros. Cada amostra está etiquetada com seu tipo (branco ou tinto) e uma nota de qualidade dada por especialistas.

Entrega

Os grupos deverão submeter um notebook jupyter devidamente comentado e o power-point usado para a apresentação final.

A avaliação do projeto incluirá a apresentação, as técnicas implementadas e resultados obtidos.

Parâmetros de avaliação*Técnico:*

- Cumprimento das diversas etapas chave de análise;
- Discussão das diversas etapas de análise;
- Inclusão de técnicas de exploração e tratamento de dados, e otimização e avaliação de resultados, tal como lecionadas nas aulas;
- Espírito crítico e fundamentação;
- Diversidade de técnicas e algoritmos testados.

Notebook

- Conteúdo;
- Organização;
- Originalidade;

Os alunos deverão comentar no caderno de acordo com os principais segmentos do pipeline que seguem, bem como as técnicas que implementam (incluindo análise intermédia de interesse) e discussão dos resultados.

Apresentação:

- Conteúdo e organização;
- Clareza da mensagem;
- Capacidade de discutir e justificar.

As apresentações terão duração de 20 minutos por grupo (10 para apresentação + 10 para perguntas). Todos os membros do grupo devem falar durante a apresentação. A ordem das apresentações será oportunamente disponibilizada.

Prazo para envio: 21 de junho às 18h

Apresentação: 27 de junho 09h00