

Análise da Qualidade de Maçãs

Introdução:

Este relatório documenta a análise realizada para prever a qualidade de maçãs com base em suas características físico-químicas. O objetivo é construir um modelo de Machine Learning capaz de classificar as maçãs como "boa" ou "ruim".

Dados:

Os dados utilizados neste projeto foram obtidos do arquivo "apple_quality.csv". O dataset contém informações sobre diversas características das maçãs, como acidez, doçura, suculência, etc.

Etapas:

1. Exploração e Preparação dos Dados:

- Importação das bibliotecas Pandas, NumPy, Scikit-learn, Matplotlib e Seaborn.
- Leitura dos dados do arquivo "apple_quality.csv" para um DataFrame Pandas.
- Análise exploratória dos dados:
 - Verificação das primeiras linhas do DataFrame.
 - Dimensões do DataFrame.
 - Valores únicos da coluna "Quality" (variável alvo).
 - Distribuição das classes (boa/ruim).
 - Verificação de dados duplicados.
 - Tipos de dados das colunas.
 - Valores mínimo e máximo da coluna "Quality".
 - Visualização da dispersão das features usando um gráfico de dispersão.
 - Cálculo e visualização da matriz de correlação entre as features.
- Preparação dos dados:
 - Remoção da coluna "A_id" (identificador).
 - Remoção de linhas com valores ausentes.
 - Conversão da coluna "Quality" para valores numéricos (0 para "ruim" e 1 para "boa").
 - Conversão da coluna "Acidity" para o tipo float.
 - Visualização da distribuição das classes usando um gráfico de barras.

2. Treinamento e Avaliação de Modelos:

- Divisão dos dados em conjuntos de treinamento e teste (70% para treinamento e 30% para teste).
- Treinamento e avaliação de diversos modelos de classificação:
 - XGBoost (com e sem ajuste de hiperparâmetros usando Grid Search).

- Regressão Linear (com ajuste de hiperparâmetros usando Grid Search).
- SVM (com ajuste de hiperparâmetros usando Grid Search).
- Naive Bayes.
- Árvore de Decisão (com ajuste de hiperparâmetros usando Grid Search).
- Random Forest (com ajuste de hiperparâmetros usando Grid Search).
- Regressão Logística (com ajuste de hiperparâmetros usando Grid Search).
- Para cada modelo:
 - Treinamento do modelo usando os dados de treinamento.
 - Previsão nos dados de teste.
 - Avaliação do desempenho do modelo usando métricas como acurácia, precisão, recall e F1-score.
 - Visualização da matriz de confusão e curva ROC.

Resultados:

Os resultados da avaliação dos modelos são apresentados no notebook. O modelo XGBoost com ajuste de hiperparâmetros apresentou o melhor desempenho, com uma acurácia de aproximadamente 90%.

Conclusão:

Este projeto demonstrou a aplicação de técnicas de Machine Learning para prever a qualidade de maçãs. O modelo XGBoost com ajuste de hiperparâmetros se mostrou promissor para essa tarefa.

Observações:

- Os códigos foram comentados para facilitar a compreensão.
- As bibliotecas utilizadas foram importadas no início do notebook.
- Os resultados da execução dos códigos foram omitidos neste relatório.

Integrantes

Caio Bonato - 822165248
Gabriel Castro - 822157975
Gabriel Dias - 822144973
Laysla Rodrigues - 821216057
Leonardo Freitas - 822135116
Lucas Quireza - 822229907
Vinicius Almeida - 821246691