

Análise de Dados de Vinhos Portugueses

Gonçalo Candeias Amaro

15 Jan 2024

Introdução

- ▶ Pre-processamento e análise exploratória de dados.
- ▶ Previsão de nível de álcool, tipo e qualidade do vinho.
- ▶ Dúvidas e questões.

Análise Exploratória de Dados

- ▶ Dados com 12 colunas numéricas.
- ▶ Tratamento de *outliers* e valores nulos.
- ▶ Normalização dos dados com `RobustScaler`.
- ▶ Fusão e exportação dos dados tratados.

Tarefa de Regressão - Previsão de nível de álcool

Modelo	MSE	R2 Score
Regressão Linear	0.11	0.71
Regressão de Floresta Aleatória	0.05	0.87

Table: Resultados da tarefa de regressão.

- ▶ Melhor modelo: Regressão de Floresta Aleatória.
- ▶ Importância dos atributos analisada.

Tarefa de Classificação - Previsão do tipo de vinho

Modelo	Accuracy	F1 Score	Precisão	Recall
Regressão Logística	0.77	0.33	0.73	0.21
Floresta Aleatória	0.97	0.94	1.0	0.90

Table: Resultados da tarefa de classificação.

- ▶ Melhor modelo: Floresta Aleatória.
- ▶ Alta precisão e recall na classificação do tipo de vinho.

Tarefa de Previsão de Qualidade

Modelo	MSE	R2 Score	MAE
Gradient Boosting Regressor	0.33	0.37	0.47
K-Nearest Neighbors Regressor	0.35	0.32	0.46
Linear Regression	0.37	0.27	0.50
Random Forest Regressor	0.26	0.51	0.39
Support Vector Regressor	0.32	0.37	0.44

Table: Resultados da tarefa de previsão de qualidade.

- ▶ Melhor modelo: Random Forest Regressor.
- ▶ Mais eficaz na previsão da qualidade dos vinhos.

Conclusão

- ▶ Análise abrangente de vinhos portugueses "Vinho Verde".
- ▶ Modelos de machine learning aplicados com sucesso em tarefas de regressão e classificação.
- ▶ Destaque para os modelos de Floresta Aleatória em várias tarefas.
- ▶ Contribuição significativa para a compreensão e previsão de características importantes dos vinhos.

Referências

- ▶ Towards Data Science, Wine Quality Prediction Using Machine Learning.
- ▶ DataCamp, Random Forests Classifier Python.
- ▶ DataCamp, SVM Classification Scikit-learn Python.
- ▶ DataCamp, K-Nearest Neighbor Classification Scikit-learn.
- ▶ DataCamp, Understanding Logistic Regression Python.
- ▶ ChatGPT Queries, Debug python outputs.