

Project Report: Red Wine Quality Analysis

Introduction

This project explores the Red Wine Quality dataset using various data mining techniques to uncover insights about wine characteristics and predict quality ratings.

Key Findings

1.Exploratory Data Analysis (EDA): The dataset contains 1599 entries with no missing values. Most features are numerical, with the target variable `quality` as an integer. Feature distributions, such as alcohol content and acidity levels, were visualized, revealing typical ranges and distributions. A correlation matrix indicated that `alcohol`, `sulphates`, and `citric acid` positively correlate with wine quality, while `volatile acidity` shows a negative correlation.

2. Classification (Predicting Wine Quality): A Random Forest Classifier was trained to predict wine quality categories (Low, Medium, High). The model achieved an accuracy of approximately 86.56%, identifying key patterns among features like reviews and acidity levels.

3. Clustering (Grouping Similar Wines): K-Means clustering was applied to group wines based on features such as alcohol

content, acidity, and ratings. Three distinct clusters were identified, including high-quality wines with lower acidity, medium-quality wines with average characteristics, and low-quality wines that may require improvement.

4. Association Rule Mining: Relationships were discovered, such as high-quality wines often having higher alcohol content and lower volatile acidity. Visualizations helped illustrate these associations, providing insights into wine characteristics.

5. Regression (Predicting Quality Score): A Linear Regression model was developed to predict the continuous quality score based on wine features. The model showed moderate predictive capability, with an R-squared value of approximately 0.40, indicating room for improvement.

Conclusion

This analysis demonstrates the effectiveness of data mining techniques in understanding wine quality and characteristics. Future work could enhance predictive models with more detailed data, leading to applications such as personalized wine recommendations and quality improvement strategies for producers.