Capstone Final Report

Executive Summary

This project explores predicting wine quality using physicochemical features from red and white wine datasets. A Random Forest model was trained to predict wine quality ratings. The final model achieved 83% accuracy on the test set.

Introduction

The goal is to help winemakers identify key factors that influence wine quality, potentially improving production strategies. The dataset was sourced from the UCI Machine Learning Repository and includes red and white wines.

Modeling Approach

Both datasets were combined with an added 'wine_type' feature. Features were scaled, and a Random Forest model was trained with tuned hyperparameters. Train/test split was 80/20. The model was evaluated using accuracy, precision, recall, and F1 score.

Recommendations

1. Separate modeling pipelines for red and white wines could enhance accuracy.

2. Focus on alcohol and sulphate content, which showed strong correlation with quality.

3. Use predicted quality scores to guide pricing strategies and inventory decisions.

Further Research

Future analysis could explore regression instead of classification, or incorporate external data such as climate, grape type, or expert reviews. Deep learning models could also be tested to capture non-linear patterns.