Exploring Machine Learning Techniques for Wine Quality Prediction

Anonymous ACL submission

Abstract

This paper investigates the application of various machine learning techniques to predict the quality of wine based on its physicochemical properties. We utilize the Wine Quality dataset from the UCI Machine Learning Repository, focusing on red wines. Our aim is to evaluate the performance of different models in predicting wine quality, providing insights into the most effective approaches for this task.

1 Introduction

Wine quality is influenced by numerous factors, including its chemical composition, which can be quantitatively measured. Understanding these relationships can aid winemakers in producing higher-quality wines. In this project, we employ several machine learning algorithms to analyze the physicochemical attributes of wines and predict their quality ratings.

The Wine Quality dataset comprises multiple features, such as acidity levels, sugar content, and pH, which serve as inputs to our models. By leveraging various machine learning techniques, we aim to identify the methods that yield the most accurate predictions.

2 Machine Learning Techniques

To explore the relationships between the physicochemical properties and the quality of wine, we implement a range of machine learning models, including linear regression, decision trees, and support vector machines. Each model is trained and validated using a consistent approach to ensure fair comparisons.

The evaluation metrics, such as accuracy, precision, and recall, provide insights into the effectiveness of each model. We also employ cross-validation techniques to minimize overfitting and ensure that our results are generalizable.

3 Results and Discussion

Preliminary results indicate varying degrees of performance among the models. For instance, decision trees demonstrate a higher accuracy compared to linear regression. However, we discuss the trade-offs associated with each model, including their interpretability and computational efficiency.

By analyzing the results, we provide recommendations for winemakers on the most effective techniques for predicting wine quality. Our findings contribute to the ongoing discourse on the application of machine learning in agriculture and food science.

4 Limitations

While this study offers valuable insights, several limitations exist. The dataset may not capture all relevant factors influencing wine quality, and the models may not generalize to wines produced in different regions or under varying conditions. Further research could explore additional features and datasets to enhance the predictive capabilities of the models.

5 Ethics Statement

This research adheres to ethical guidelines, ensuring that the data is used responsibly and transparently. We acknowledge the importance of ethical considerations in the application of machine learning techniques, particularly in fields that impact public health and safety.

References