Wine Quality Prediction System Report (Interface included)

Group members:
Izza Farhat 22K-4120

Abstract

This report presents a wine quality prediction system developed using machine learning techniques. The system utilizes various classification algorithms to predict the quality of wine based on its physicochemical properties. A user-friendly interface is created using Streamlit, allowing users to input wine characteristics and receive predictions in real-time. The model is trained and tested on a dataset, showcasing the effectiveness of different machine learning algorithms and providing insights into the performance metrics of each approach.

Introduction

Wine quality is a significant factor in the wine industry, influencing consumer preferences and market dynamics. Accurate prediction of wine quality can assist producers in enhancing their products and guiding consumers in their selections. This project aims to develop a machine learning model that predicts wine quality based on several input features, including acidity, sugar content, and alcohol levels. The system employs multiple algorithms to ensure robust predictions and offers a graphical user interface for ease of use.

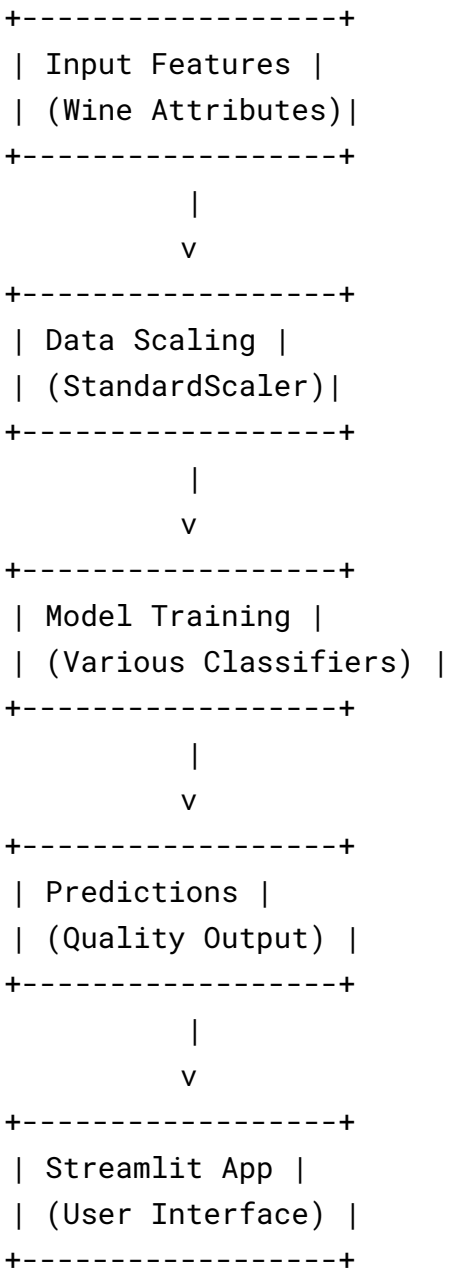
Proposed System

The proposed wine quality prediction system consists of the following components:

- **Data Preprocessing:** Scaling and transforming the dataset for uniformity.
- **Model Training:** Implementing various classification algorithms such as K-Nearest Neighbors, Decision Trees, Support Vector Machines, Logistic Regression, Naive Bayes,

Random Forest, Perceptron, and Multi-Layer Perceptron.
- **User Interface:** A Streamlit application to allow users to input wine attributes and receive predictions.

System Diagram



Experimental Setup

The dataset used for this project is a wine quality dataset containing various physicochemical properties of wine samples. Each sample is characterized by the following features:

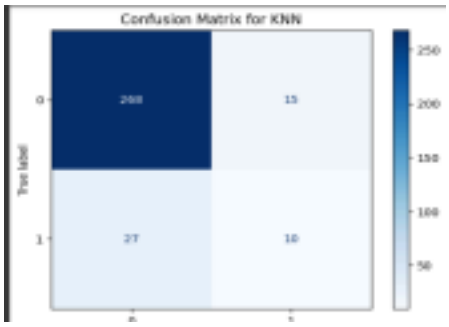
- Fixed Acidity
- Volatile Acidity
- Citric Acid
- Residual Sugar
- Chlorides
- Free Sulfur Dioxide
- Total Sulfur Dioxide
- Density
- pH
- Sulphates
- Alcohol
- Quality (target variable)

The dataset is split into training and testing sets, and a random state is used to ensure reproducibility. The random state value used in this project is **42**. The classification models are trained on the training dataset, and their performance is evaluated on the testing dataset.

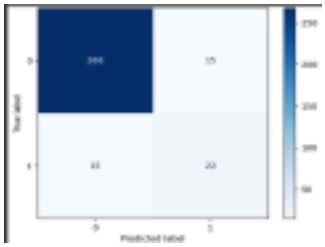
Results and Discussion

The performance of various classifiers is evaluated using accuracy scores and classification reports, which include precision, recall, and F1 scores. Below are the results obtained for each algorithm:

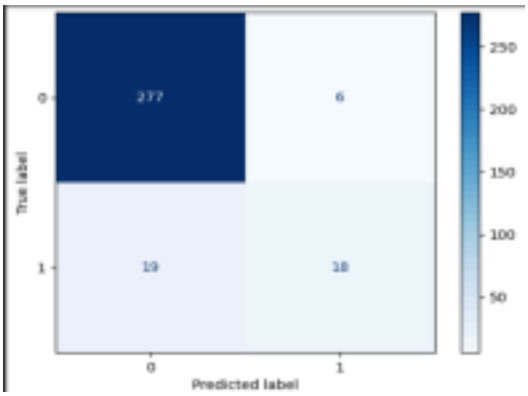
1. **K-Nearest Neighbors (KNN)**
    - Accuracy: **0.86875**
    - Classification report:



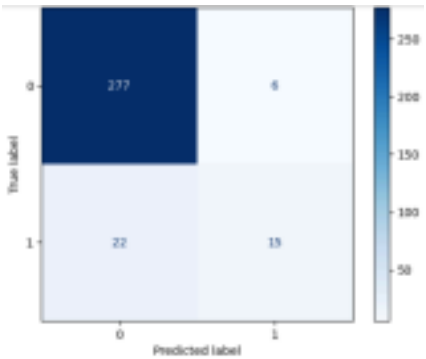
2. **Decision Tree**
    - Accuracy: **0.90625**
    - Classification report**:**



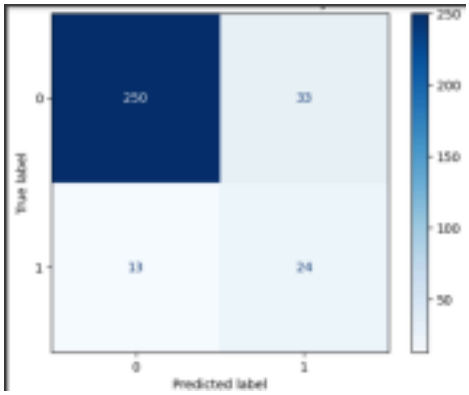
3. **Support Vector Machine (SVM)**
    - Accuracy: **0.921875**
    - Classification report:



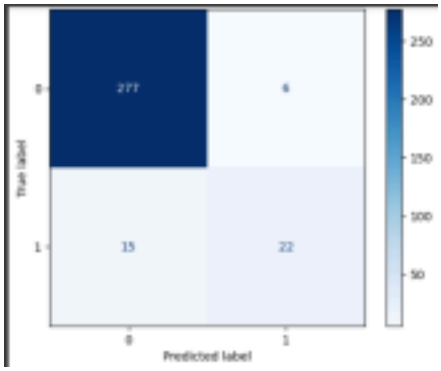
4. **Logistic Regression**
    - Accuracy: **0.9125**
    - Classification report:



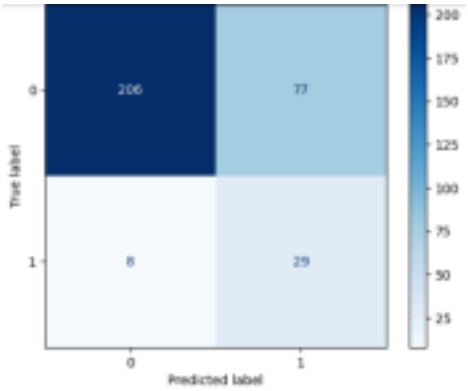
5. **Naive Bayes**
    - Accuracy: **0.85625**
    - Classification report:



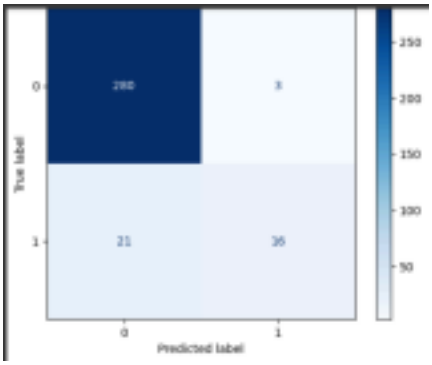
6. **Random Forest**
    - ○ Accuracy: **0.934375**
    - ○ Classification report:



7. **Perceptron**
    - ○ Accuracy on Test Data: **0.7344**
    - ○ Classification report:



8. **Multi-Layer Perceptron (MLP)**
    - Accuracy on Test Data: **0.9250**
    - Classification report:



The results indicate that Random Forest achieved the highest accuracy at 0.934375, followed closely by SVM and MLP. The classification reports provide a detailed breakdown of each model's performance, enabling a comprehensive analysis of their strengths and weaknesses.

Interface driven result:

We can adjust the levels of all the column names accordingly to figure out if the desired adjustment of the column names will predict a good quality or a bad quality wine

Conclusion

In conclusion, the wine quality prediction system successfully demonstrates the application of machine learning algorithms in predicting wine quality based on physicochemical properties. The integration of various models allows for a robust comparison, highlighting the potential for improvement in wine production and consumer guidance. Future work may involve further fine-tuning of the models, exploring additional features, or incorporating more advanced

algorithms to enhance prediction accuracy.