

Assignment 2 Question:

Title: Wine Quality Classification Using Binary and Multiclass Classification

Description: In this assignment, you will explore the wine quality dataset and apply machine learning classification techniques to predict the quality of wines. The dataset encompasses various physicochemical properties of different wines alongside their quality ratings. You'll tackle two classification tasks: binary classification to distinguish high-quality wines from low-quality ones, and multiclass classification to categorize wines into quality classes. For the binary classification task, wines with quality ratings of 7 and above will be considered "high-quality" (positive class), while wines with ratings below 7 will be classified as "low-quality" (negative class). Your task is to construct classification models capable of accurately performing both binary and multiclass classification using One-vs-One (OvO) and One-vs-Rest (OvR) strategies, and analyze the results using confusion matrices.

Tasks:

Data Exploration:

1. Load the wine quality dataset and conduct exploratory data analysis (EDA) to grasp its structure and characteristics.
2. Analyze the distribution of wine quality ratings and discern any patterns or trends.

Data Preprocessing:

1. Preprocess the dataset by addressing any missing values and encoding categorical variables if necessary.
2. Partition the dataset into training and testing sets for model evaluation.

Binary Classification:

1. Simplify the classification problem into binary by categorizing wines as high-quality or low-quality based on quality ratings.
2. Train binary classification models using at least three different algorithms.
3. Evaluate each model using appropriate evaluation metrics (accuracy, precision, recall, F1-score), and visualize the results using confusion matrices.
4. Compare the performance of binary classification models and select the best-performing one.

Multiclass Classification:

1. Perform multiclass classification to categorize wines into quality classes using OvO strategy.
2. Train multiclass classification models using at least three different algorithms.
3. Evaluate each model using appropriate evaluation metrics (accuracy, precision, recall, F1-score), and visualize the results using confusion matrices.
4. Perform multiclass classification using OvR strategy.
5. Evaluate each model using appropriate evaluation metrics (accuracy, precision, recall, F1-score), and visualize the results using confusion matrices.
6. Compare the performance of multiclass classification models with OvO and OvR strategies.
7. Select the best-performing model for multiclass classification.

Interpretation and Analysis:

1. Analyze the importance of features in predicting wine quality using the selected binary classification model.

2. Interpret the confusion matrix and ROC curve analysis to understand the performance of the selected binary classification model.
3. Analyze the importance of features in predicting wine quality using the selected multiclass classification model.
4. Interpret the confusion matrix and ROC curve analysis to understand the performance of the selected multiclass classification model.

Submission:

- Submit your assignment report, including code snippets, visualizations, model evaluation results, and interpretations using Jupyter Notebook within 7 days after it is assigned.