

Case Study: Wine Quality Classification using Supervised Machine Learning

1. Introduction

The wine quality dataset consists of physicochemical properties of red wine samples along with a sensory quality rating provided by wine tasters. The goal is to predict the wine quality (good or bad) based on measurable chemical features.

Dataset Details:

- Rows: 1599
- Columns: 12 + 1 label column
- Class distribution: Good wines: 855, Bad wines: 744

Features include: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol.

Target Variable: A new target label 'quality_label' is created: 'good' if quality ≥ 6 , else 'bad'.

2. Steps in Building a Classification Model

Data Collection and Understanding

- Gather the relevant dataset for your problem.
- Load the data using pandas.
- Perform Exploratory Data Analysis (EDA) to identify outliers, trends, and relationships.

Data Preprocessing

- Handle missing values (if any).
- Clean data (remove duplicates, correct types).
- Feature engineering (if necessary).
- Convert quality scores into binary labels: bad (0), good (1).
- Split data into features (x) and target (y).

Encoding and Scaling

- No categorical columns in this dataset, so encoding is skipped.
- Apply StandardScaler for feature scaling.

Splitting Data

- Use train_test_split from sklearn.model_selection.
- Typical split is 70% training, 30% testing.

Model Selection and Training

- - Apply LogisticRegression for this binaryclassifier problem
- - Apply DecisionTreeClassifier for this binaryclassifier problem
- - Apply Pruning Technique
- - Apply RandomForestClassifier for this binary classification problem.
- -Train the model

Model Evaluation

- - Evaluate with accuracy, precision, recall, and F1-score.
- - Use confusion matrix for visual inspection.

Model Optimization

- - Compare with other classifiers like SVM, Logistic Regression, XGBoost, AdaBoost, Gradient Boost.

3. Conclusion

The wine quality classification model allows us to predict wine quality using supervised learning techniques. This enables winemakers or quality controllers to automate and improve wine assessment processes based on chemical analysis.