

# Wine Quality Prediction Documentation

## Introduction

This documentation provides an overview of the wine quality prediction project. The goal of this project is to predict the quality of wines based on various chemical attributes using a linear regression model.

## Dataset

### Wine Quality Dataset

- **Source:** [UCI Machine Learning Repository](#)
- **Format:** CSV

The dataset consists of two separate CSV files: **winequality-red.csv** and **winequality-white.csv**, containing data on red and white wines, respectively. Each dataset includes attributes such as fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, and wine quality (as the target variable).

## Code

### Overview

The code for this project is implemented in a Jupyter Notebook file named **wine\_quality\_prediction.ipynb**. It covers the following main aspects:

1. Data loading and preprocessing.
2. Model training using Linear Regression.
3. Model evaluation using mean squared error (MSE) and R-squared (R<sup>2</sup>) score.
4. Visualization of results, including scatter plots, histograms, and residual plots.

## **Operations**

### **Data Preprocessing**

- Data is loaded from the provided CSV files.
- Features are selected and cleaned as needed.
- The dataset is split into training and testing sets.

### **Model Training**

- A Linear Regression model is chosen for this task.
- The model is trained on the training data.

### **Model Evaluation**

- Model performance is assessed using Mean Squared Error (MSE) and R-squared (R<sup>2</sup>) score.

### **Visualizations**

#### **Scatter Plot**

- A scatter plot is used to visualize the predicted wine quality against the actual wine quality. This provides a visual assessment of the model's performance.

#### **Histogram of Predictions**

- Histograms are created to visualize the distribution of predicted and actual wine quality values. This helps in understanding the quality distribution.

#### **Residual Plot**

- A residual plot is generated to visualize the differences between predicted and actual wine quality values. A good model will exhibit residuals evenly distributed around zero.