Wine Quality Prediction — Project Overview

# Objective

To predict wine quality using physicochemical features (e.g., acidity, sulphates, alcohol) with the help of machine learning models. The project aims to assist in identifying key attributes that influence wine quality for better production control.

# What's in the Code

* Data Loading:  
  - Loaded red wine dataset from CSV.  
  - Basic overview of shape, data types, and value distribution.
* Exploratory Data Analysis (EDA):  
  - Visualizations of key attributes (alcohol, sulphates, acidity).  
  - Box plots, histograms, and heatmaps to understand correlations.
* Target Variable Creation:  
  - Converted wine quality scores into binary classification (goodquality) based on threshold (>=7).  
  - Checked class imbalance and label counts.
* Data Preprocessing:  
  - Applied StandardScaler for normalization.  
  - Split dataset into training and testing sets (80/20 split).
* Model Building:  
  - Trained multiple classification models:  
   • Random Forest  
   • Support Vector Machine (SVM)  
   • XGBoost  
  - Compared performance using:  
   • Accuracy Score  
   • Confusion Matrix  
   • Classification Report.
* Feature Importance:  
  - Visualized most influential features affecting wine quality.  
  - Alcohol and sulphates emerged as top predictors.

# Results

Achieved up to 92% accuracy with Random Forest.  
Alcohol content and sulphates were strong positive indicators of wine quality.

# Tools Used

Python, Pandas, Scikit-learn, Matplotlib, Seaborn, XGBoost

.