Red Wine Quality Prediction Using Machine Learning

Problem Statement:

Wine quality plays a crucial role in customer satisfaction and business growth in the beverage industry. Determining the quality of wine based on its physicochemical properties can be challenging for non-experts. This project aims to leverage machine learning to predict the quality of red wine using its chemical attributes, both as a **regression** and a **classification** problem.

Objective:

To build and compare two machine learning models that:

1. **Predict the wine quality score** on a numeric scale (regression approach)

2. **Classify the wine as "good" or "bad"** based on a quality threshold (classification approach)

This project helps in understanding core ML workflows and how different modeling techniques can be applied to the same dataset for different insights.

Domain:

Data Mining / Food & Beverage Analytics / Machine Learning

Motivation:

- To practice and brush up on foundational machine learning skills

- To gain hands-on experience in both regression and classification tasks

- To explore how predictive models can assist in improving business processes such as wine production, quality control, and targeted marketing

Dataset:

- **Wine Quality Dataset** (commonly available on UCI ML Repository)

- Contains physicochemical properties of red wine such as:
    - Fixed acidity
    - Volatile acidity
    - Citric acid
    - Residual sugar
    - Chlorides
    - Free sulfur dioxide
    - Total sulfur dioxide
    - Density
    - pH
    - Sulphates
    - Alcohol

- **Quality** (score between 0–10, which is the target variable)

Tools & Technologies Used:

- **Languages:** Python
- **Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
- **ML Models:** Linear Regression, Random Forest, Logistic Regression, Decision Tree, etc.
- **Environment:** Jupyter Notebook / Google Colab

Project Workflow:

1. Data Preprocessing & EDA

- Load data into pandas DataFrame
- Handle missing values, if any
- Explore data using visualizations and summary statistics
- Analyze feature correlations using heatmaps and scatterplots
- Identify key ingredients influencing wine quality

2. Regression Approach

- Treat the **'quality'** column as a continuous variable
- Train a regression model (e.g., Random Forest Regressor)
- Evaluate using MAE, RMSE, and R^2 Score
- Use feature importance to understand influential ingredients

3. Classification Approach

- Convert quality scores into binary labels (e.g., quality ≥ 7 → "Good", else → "Bad")
- Train classification models like Logistic Regression, Decision Tree, or Random Forest
- Evaluate using Accuracy, Precision, Recall, F1-Score, and Confusion Matrix
- Compare model performances

4. Insights & Business Impact

- Determine which features most impact wine quality
- Use results to suggest possible production tweaks or quality control strategies
- Highlight how classification thresholds can help tailor offerings for premium wine lines

Key Learnings:

- Understanding of both **supervised regression and classification**
- Hands-on with **data preprocessing and exploratory data analysis**
- Importance of **feature selection and correlation analysis**
- Comparative analysis of different modeling approaches on the same problem
- Application of machine learning in real-world F&B industry scenarios

Conclusion:

This project demonstrates the versatility of machine learning in analyzing and predicting product quality using chemical attributes. It offers dual insights through both regression and classification perspectives and serves as a solid foundational project for any data science learner.