Wine Quality Prediction Using Machine Learning

Objective:

The main objective of this project is to build a machine learning model that can predict the quality of wine (red or white) based on its chemical properties. The goal is to assist wine producers and quality analysts in identifying high-quality wine using measurable factors.

Dataset Used:

Kaggle – Wine Quality Dataset

https://www.kaggle.com/datasets/ruthgn/wine-quality-data-set-red-white-wine/data

Description:

This dataset combines red and white wine data, containing 13 columns:

Features:

Fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, type.

Target:

Wine quality (score from 0–10).

Models Chosen:

Multiple classification algorithms were applied to predict wine quality:

- Random Forest
- Decision Tree
- K-Nearest Neighbors (KNN)
- Support Vector Classifier (SVC)
- Gradient Boosting

These models were trained on scaled and preprocessed data and evaluated to choose the best-performing one.

Performance Metrics:

To evaluate model performance, the following metrics were used:

	Models	Accuracy (%)	Precision	Recall	F1 Score
0	LR	74.377395	0.748047	0.743774	0.743646
1	SVC	80.076628	0.811684	0.800766	0.800050
2	KNN	82.950192	0.841568	0.829502	0.828846
3	DT	83.237548	0.833159	0.832375	0.832462
4	RF	88.170498	0.883140	0.881705	0.881765
5	GBC	79.454023	0.799616	0.794540	0.794398

Among all models, **Random Forest** performed the best, with balanced precision and recall for both red and white wines.

For Random Forest:

Accuracy: 88.17%

Precision: 0.8831

Recall: 0.8817

F1-Score: 0.8817

Challenges & Learnings

Handling Imbalanced Data: Certain wine quality scores were underrepresented, which required careful handling during training.

Combining Red and White Wines: Managing different distributions between red and white wine required additional feature engineering.

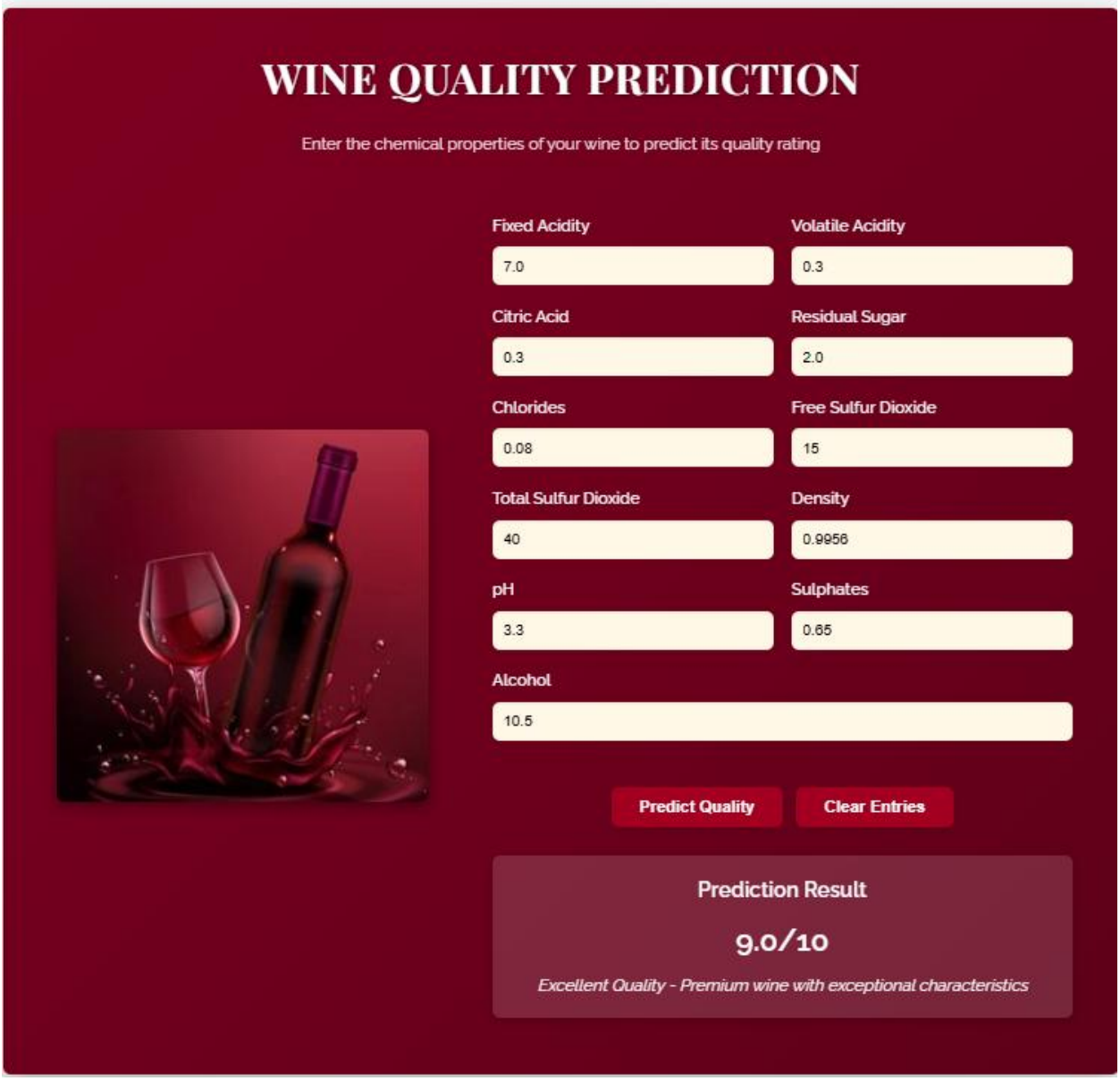
Data Cleaning: Although the dataset was mostly clean, ensuring proper scaling and encoding was crucial for model performance.

Power BI Integration: Creating an interactive dashboard and embedding live metrics was a key learning, especially for non-technical stakeholders.

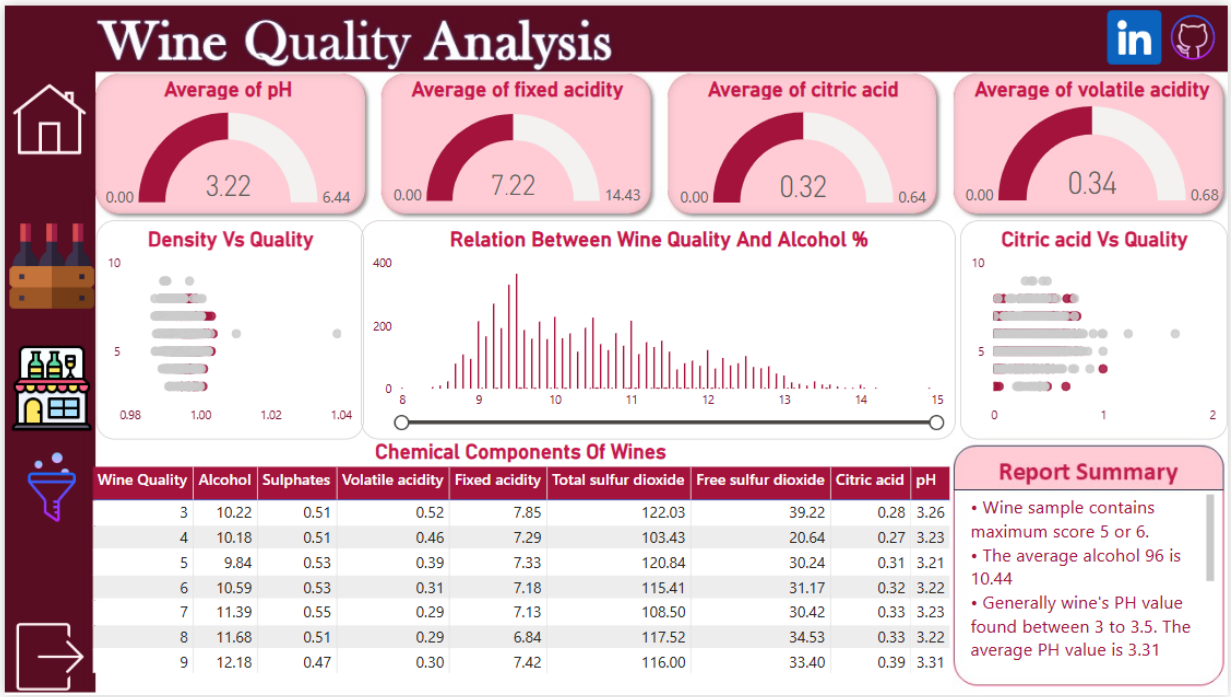
Real-time Prediction UI: Integrated a web interface using FastAPI and deployed it on Render for public access.

Live Previews

Web UI: Wine Quality Prediction UI [View]



Power BI Dashboard: [Power BI Live View]



GitHub:

https://github.com/sonikirtan110/Wine-Quality-Analysis/tree/main