Wine Quality Prediction

Data Analytics Internship

Anisha More

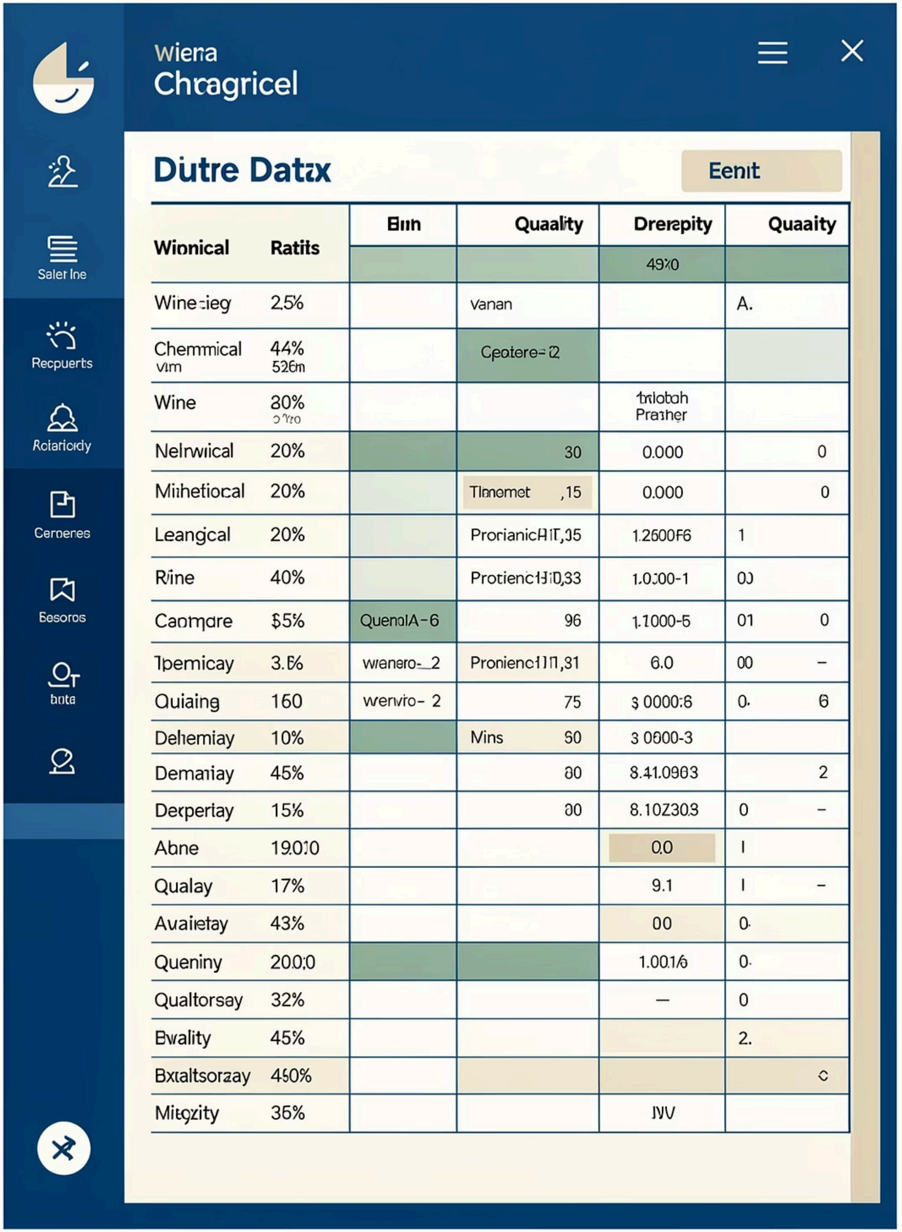
Problem Statement

Predict wine quality (Good / Bad)

Use chemical attributes

Compare multiple classifiers

Evaluate model performance



Dataset Overview

1599

Total Records

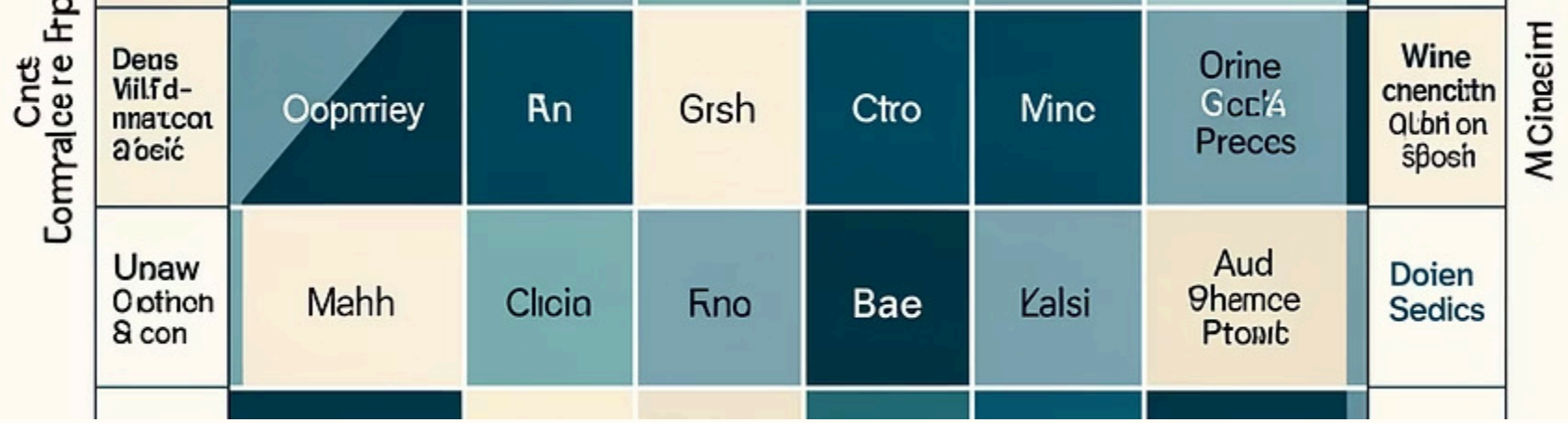
11

Features

Chemical attributes

Target Variable: Quality

Converted to binary classification



Data Exploration

- No major missing values
- Correlation analysis performed
- Alcohol strongly correlated with quality
- Volatile acidity negatively correlated

Models Implemented

Random Forest Classifier

Support Vector Classifier (SVC)

Stochastic Gradient Descent (SGD)

Train-Test Split (80:20)

Model Performance Comparison

Random Forest Accuracy

SVC Accuracy

SGD Accuracy

Random Forest performed best

Important Features

Top Influential Features:

Alcohol

Volatile Acidity

Sulphates

Density

Visualization Insights

- Heatmap for correlation
- Distribution plots
- Count plot for quality
- Feature importance chart

Key Insights

Higher alcohol → Better quality

High acidity → Lower quality

Random Forest most reliable model

Dataset moderately balanced

Conclusion

01

Built multi-model classifier

02

Compared performance

03

Identified key chemical factors

04

Achieved 88% best accuracy