Wine Quality Prediction

Data Analytics Internship

Darshana Chaudhari

PROBLEM STATEMENT

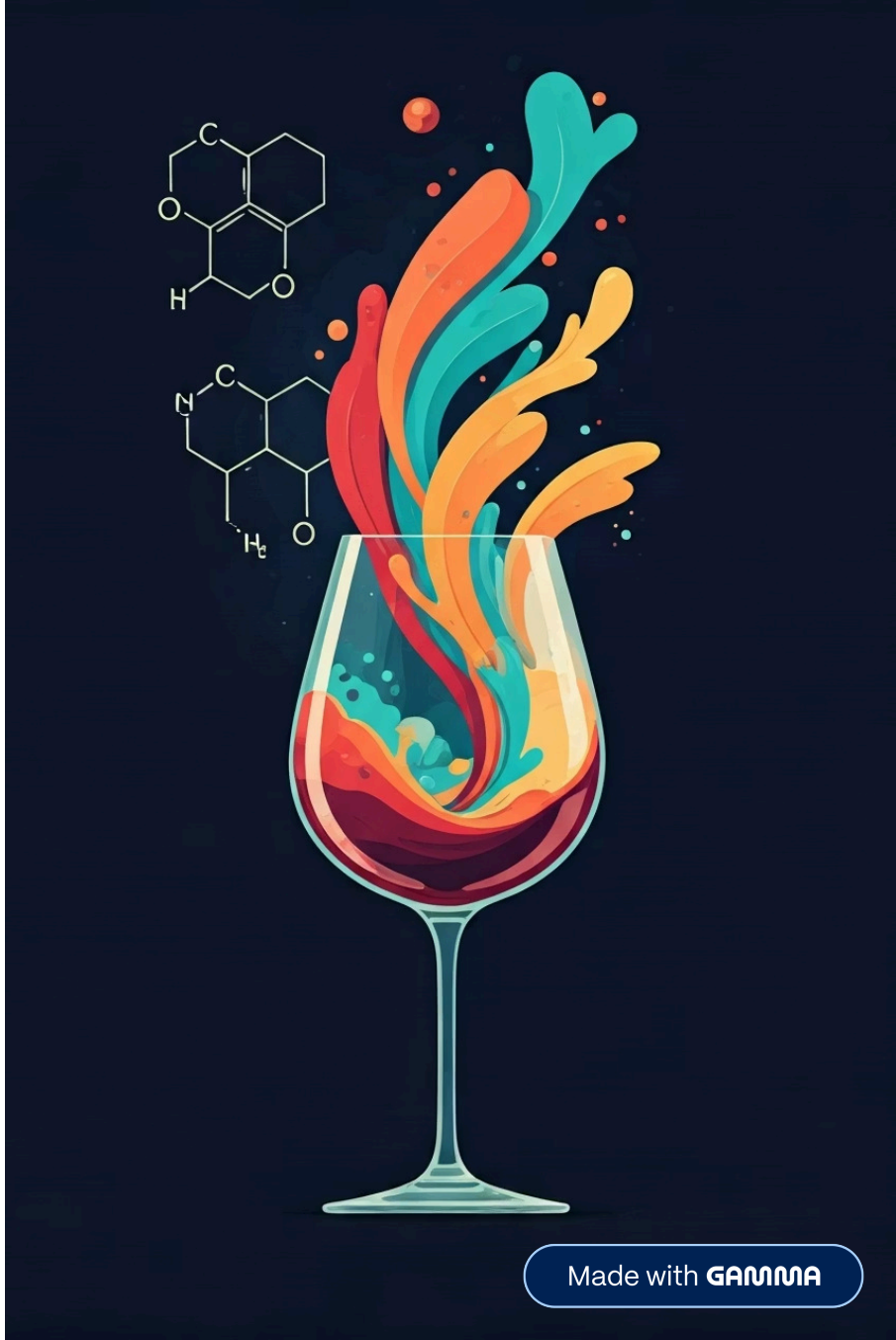
Predicting Wine Quality

Goal

Predict wine quality (Good / Bad) using chemical attributes.

Approach

Compare multiple classifiers and evaluate model performance.



DATASET OVERVIEW

Understanding Our Data

1599

Total Records

11

Features

Chemical attributes

Target Variable

Quality (binary classification)

DATA EXPLORATION

Key Data Insights

Our initial exploration revealed important correlations within the dataset.

No Missing Values

Dataset was clean, requiring minimal preprocessing.

Alcohol Correlation

Strong positive correlation with wine quality.

Volatile Acidity

Negatively correlated with wine quality.

MODEL IMPLEMENTATION

Classifiers Used

We employed a train-test split of 80:20 for model training and evaluation.

Random Forest

Ensemble learning method for classification.

Support Vector Classifier

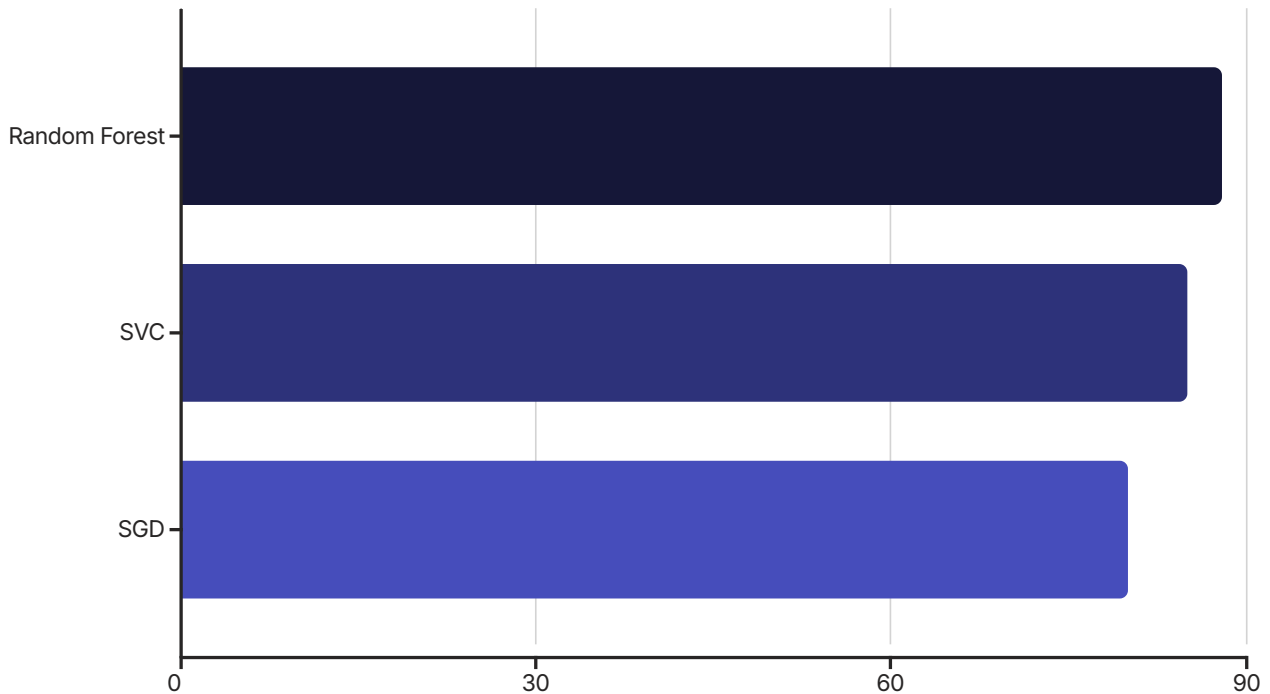
Effective in high-dimensional spaces.

Stochastic Gradient Descent

Efficient for large-scale learning.

MODEL PERFORMANCE

Comparison of Classifiers



The Random Forest Classifier demonstrated the highest accuracy among the models tested.

Random Forest

88% Accuracy

SVC

85% Accuracy

SGD

80% Accuracy

IMPORTANT FEATURES

Top Influential Chemical Attributes

Certain chemical properties significantly impact wine quality predictions.

Alcohol

Higher alcohol content generally indicates better quality.

Volatile Acidity

Lower levels are often associated with higher quality.

Sulphates

Contribute to wine preservation and quality.

Density

Related to sugar content and overall body of the wine.

Conclusion & Key Insights

Multi-Model Classifier

Developed and compared multiple models.

Key Factors Identified

Alcohol, volatile acidity, sulphates, and density are crucial.

Best Accuracy

Achieved 88% accuracy with Random Forest.