MIT | Academy of Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

WINE QUALITY PREDICTION USING MACHINE LEARNING

TY B.Tech. Computational Intelligence (2306311)

Project Report

SUBMITTED BY

Student1 Name : Vaibhav Gorme **PRN :** 202301060034

Student2 Name : Sayli Munde **PRN :** 202301060045

Student3 Name : Raju Pisal **PRN :** 202301060052

Student4 Name : Atharva Rajandekar **PRN :** 202301060027

GUIDED BY

Dr. Abhilasha Joshi

SCHOOL OF ELECTRONICS AND TELECOMMUNICATION ENGINEERING

MIT ACADEMY OF ENGINEERING, ALANDI (D), PUNE-412105

MAHARASHTRA (INDIA)

2025-26

MIT | Academy of Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

CERTIFICATE

It is hereby certified that the work which is being presented in the TY B.Tech. Computational Intelligence Report entitled " WINE QUALITY PREDICTION - MACHINE LEARNING", in partial fulfillment of the requirements for the award of the Bachelor of Technology in E&TC Engineering/Electronics Engineering and submitted to the School of Electronics and Telecommunication Engineering of MIT Academy of Engineering, Alandi(D), Pune, Affiliated to Savitribai Phule Pune University (SPPU), Pune is an authentic record of work carried out during an Academic Year 2025-2026, under the supervision of Prof. Abhilasha Joshi School of Electronics and Telecommunication Engineering.

Student1 Name : Vaibhav Gorme	**PRN:**202301060034	**[School of Electronics Engineering]**
Student2Name: Sayli Munde	**PRN:**202301060045	**[School of Electronics Engineering]**
Student3 Name : Raju Pisal	**PRN:**202301060052	**[School of Electronics Engineering]**
Student4 Name: Atharva Rajdenkar	**PRN:**202301060027	**[School of Electronics Engineering]**

Date:

Signature of Project Guide and Co-Guide

Project Guide: Dr. Abhilasha Joshi

School of E&TC Engineering,

MIT Academy of Engineering, Alandi(D), Pune

Signature of HOD

HOD: Dr. Dipti Sakhare

School of E&TC Engineering

MIT Academy of Engineering, Alandi(D), Pune

Signature of Internal examiner/s

Name………………………………

Affiliation…………………………

Signature of External examiner/s

Name………………………………

Affiliation…………………………

CONTENTS

ACKNOWLEDGEMENT

We want to express our gratitude towards our respected project guide Prof. Abhilasha Joshi for his constant encouragement and valuable guidance during the completion of this project work. We also want to express our gratitude towards respected School HOD Dr. Dipti Sakhare for her continuous encouragement.

We would be failing in our duty if we do not thank all the other staff and faculty members for their experienced advice and evergreen co-operation

Student Name: Vaibhav Gorme

Student Name: Sayli Munde

Student Name: Raju Pisal

Student Name: Atharva Rajandekar

Chapter-1 ABSTRACT

Wine quality assessment plays a crucial role in the global wine industry, influencing product pricing, consumer acceptance, and brand positioning. Traditional methods rely on human experts performing sensory evaluation, which is subjective, inconsistent, and time-consuming. To address these limitations, this project presents a Machine Learning–based Wine Quality Prediction system using the UCI Red and white Wine Quality dataset. The dataset contains key physicochemical attributes such as acidity, pH, sulphates, alcohol, and sugar content.

Multiple machine learning models including Logistic Regression, Support Vector Machine (SVM) and XGBoost were trained and evaluated. Random Forest achieved the best performance with an accuracy of **84%** and high F1-scores across classes. A user-friendly Streamlit web application was developed to provide real-time wine quality prediction, data visualization, and interactive exploration using Plotly.

This project demonstrates the potential of ML-driven quality assessment systems for the beverage industry, offering automation, scalability, and objective evaluation. Future enhancements include deep learning models, automated IoT-based data collection, and real-time industrial deployment.

Chapter-2 INTRODUCTION

Wine quality is fundamentally influenced by its chemical composition, production techniques, and aging conditions. The global wine market is valued at over $300 billion, and quality evaluation significantly impacts customer perception and pricing. Traditional wine testing relies heavily on human tasters who evaluate aroma, taste, acidity, and appearance. However, such human-centric evaluation has multiple drawbacks:

- Subjective opinions vary across tasters
- Fatigue and environmental factors reduce accuracy
- Consistency cannot be guaranteed
- Expensive and time-consuming process

Machine Learning provides a modern solution by using computational models to learn the relationship between physicochemical properties and quality ratings. This transforms wine quality assessment from subjective to data-driven and objective.

The aim of this project is to design and develop a Wine Quality Prediction System using machine learning models and deploy it as an interactive web application. The project utilizes 11 key physicochemical parameters available in the UCI Red And white Wine Quality dataset and predicts quality on a scale of 0–10.

Chapter-3 PROBLEM STATEMENT

“To develop a machine learning model that predicts wine quality from its chemical features.”

Chapter-4 OBJECTIVES

4.1 Technical Objectives

- To analyse and preprocess the UCI Red Wine Quality dataset.
- To explore and identify key features influencing wine quality.
- To train multiple ML models and compare performance.
- To develop an accurate prediction model for wine quality.
- To deploy the model using Streamlit for real-time interaction.

4.2 Application-Oriented Objectives

- To create a user-friendly wine quality prediction interface.
- To enable industries to automate wine testing.
- To reduce human dependency in quality assessment.

4.3 Learning Objectives

- Understanding supervised machine learning concepts.
- Hands-on experience with model training, evaluation, and deployment.
- Real-world software implementation using Streamlit.

Chapter-5 SCOPE OF THE PROJECT

Included in Scope

- Use of physicochemical data for wine quality prediction.
- Comparison of ML models (Logistic Regression, SVM, XGBoost).
- Development of a Streamlit-based web interface.
- Data visualization using Plotly charts.

Not Included in Scope

- Implementation of deep learning models (future enhancement)
- Integration with real-time industrial sensors
- Wine taste/smell prediction

Chapter-6 LITERATURE SURVEY

Author(s)	Year	Methodology Used	Key Findings
Cortez et al.	2009	Regression & Classification models	Alcohol and volatile acidity strongly influence quality
Sahni et al.	2018	Random Forest, SVM	Random Forest gives best performance
Kour & Sharma	2020	ANN-based prediction	Improved generalization but requires heavy computation
Reddy et al.	2021	XGBoost on UCI dataset	Achieved >85% accuracy
Patel & Gupta	2022	Ensemble models	Feature engineering enhances model performance

Key Insights from Literature

- Most studies agree XGBoost delivers high accuracy for tabular chemical datasets.
- Alcohol, volatile acidity, sulphates, and pH are consistent top predictors.
- Ensemble models outperform linear models due to nonlinear feature interactions.
- Deployment for real-time usage is still limited in many academic works—your project fills this gap.

Chapter-7 EXISTING SYSTEM

Traditional wine quality assessment relies heavily on expert tasters who evaluate sensory factors such as aroma, flavor, acidity, sweetness and overall balance. Although this method has been used for centuries, it suffers from several limitations.

7.1 Limitations of the Existing System

1. Subjectivity:
   Sensory evaluation varies from person to person. Two experts may assign different quality ratings for the same wine sample.
2. Inconsistency:
   Environmental factors such as room temperature, tasting conditions, and even mood can affect results.
3. High Cost & Time Consumption:
   Professional wine tasters charge high fees and cannot evaluate large batches quickly.
4. Scalability Issues:
   Manual tasting is not practical for large wineries producing thousands of bottles daily.
5. Human Fatigue:
   Sensory fatigue reduces the accuracy of evaluation, especially during large tasting sessions.

7.2 Need for Automation

Given these drawbacks, an automated system that predicts wine quality based on chemical properties:

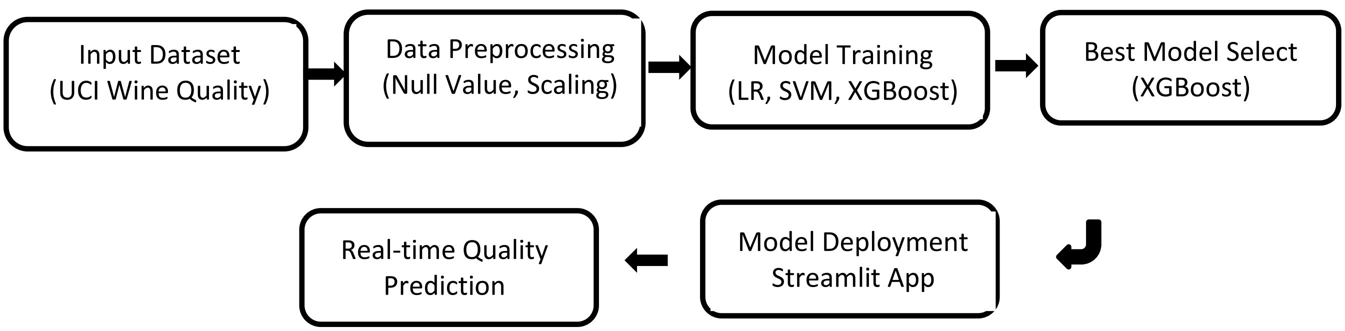
- Increases objectivity
- Enhances scalability
- Reduces dependency on experts
- Enables consistent decision-making

Chapter-8 PROPOSED SYSTEM

The proposed system uses Machine Learning algorithms to predict wine quality based on physicochemical features. It eliminates subjectivity and provides reliable predictions in real-time.

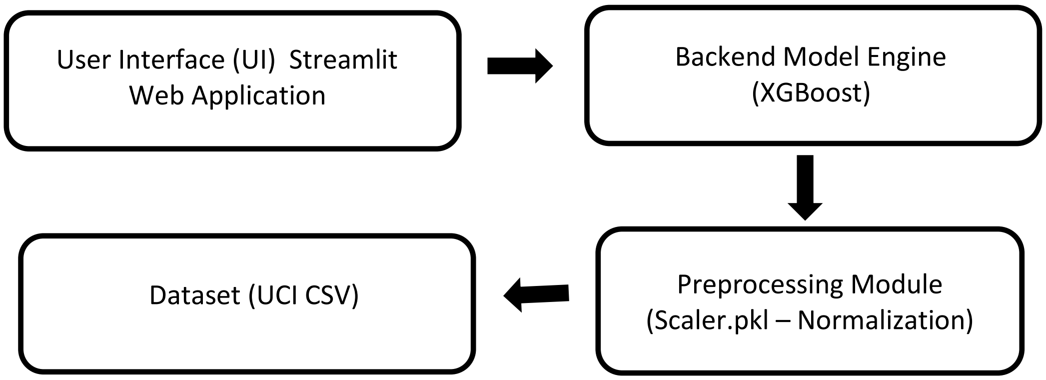
8.1 Features of the Proposed System

- Predicts wine quality on a 0–10 scale.
- Uses 11 chemical features (acidity, pH, sugar, etc.).
- Trains and compares multiple ML models.
- XGBoost chosen as best-performing model.
- Provides interactive web interface using Streamlit.
- Displays Plotly charts for deeper data analysis.
- Shows model insights like feature importance.



Chapter-9 SYSTEM ARCHITECTURE

9.1 Architecture Diagram



9.2 Components Explanation

1. User Interface Layer

- Built using Streamlit
- Allows input through sliders
- Displays prediction results
- Includes Plotly visualizations

2. Backend Model Layer

- Uses XGBoost trained model
- Loads wine_model.pkl to make predictions

3. Data Preprocessing Layer

- Loads scaler.pkl
- Normalizes input features

4. Data Source Layer

- UCI Red Wine Quality dataset
- Provides 11 physicochemical features

Chapter-10 DATA DESCRIPTION

10.1 Dataset Overview Table

Attribute	**Description**
Fixed acidity	Non-volatile acids
Volatile acidity	Acetic acid content
Citric acid	Flavor enhancement
Residual sugar	Sugar left after fermentation
Chlorides	Salt content
Free sulfur dioxide	Prevents microbial growth
Total sulfur dioxide	Sum of free + bound SO2
Density	Affected by alcohol & sugar
pH	Acidity level
Sulphates	Preservative
Alcohol	Percentage alcohol

10.2 Dataset Statistics

Property	**Value**
Total Samples	6,498
Training Samples	5,198
Testing Samples	1,300
Number of Features	11
Output Variable	Quality (0–10)
Type	Regression / Classification

Chapter-11 FEATURE DESCRIPTION

11.1 Table of Feature Descriptions

Feature	Type	Description	Influence on Wine Quality
Fixed Acidity	Continuous	Non-volatile acids in wine (tartaric, malic)	High values increase sourness
Volatile Acidity	Continuous	Acetic acid content	High values → unpleasant vinegar taste
Citric Acid	Continuous	Adds freshness & flavor	Higher value → better taste
Residual Sugar	Continuous	Leftover sugar after fermentation	Higher sugar increases sweetness
Chlorides	Continuous	Salt content	High values degrade wine taste
Free Sulfur Dioxide	Continuous	SO_2 prevents microbial growth	Helps preserve wine
Total Sulfur Dioxide	Continuous	Sum of free & bound SO_2	Excess SO_2 harms aroma
Density	Continuous	Related to sugar & alcohol	Lower density → better quality
pH	Continuous	Acidity level	Low pH (high acidity) improves wine stability
Sulphates	Continuous	Additive to enhance flavour	Moderate amount improves quality
Alcohol	Continuous	Ethanol percentage	Strong positive impact on quality

11.2 Target Variable

Variable	Description
Quality Score	Integer rating from 0 to 10 given by wine experts

11.3 Insights from Data Analysis

- Alcohol and volatile acidity are the top predictors.
- Wines with higher sulphates tend to achieve higher quality scores.
- Density has an inverse relationship with alcohol content.

Chapter-12 METHODOLOGY

12.1 Project Workflow Diagram

Dataset → Preprocessing → Feature Scaling → Model Training → Model Selection → Deployment → Prediction

12.2 Methodology Steps

Step 1: Data Collection

- UCI Red Wine Quality Dataset downloaded in CSV format.

Step 2: Data Cleaning

- Checked for null values
- Removed duplicates
- Converted data types
- Handled outliers using IQR method (optional)

Step 3: Feature Scaling

- StandardScaler applied
- Necessary for models like SVM & XGBoost

Step 4: Splitting Data

- Train/Test split = 80:20

Step 5: Model Training

- Models trained:
    - Logistic Regression
    - SVM
    - XGBoost

Step 6: Model Evaluation

- Metrics used:
    - Precision
    - Recall
    - F1-score
    - Accuracy

Step 7: Saving Best Model

- XGBoost selected
- Saved as wine_model.pkl
- Scaler saved as scaler.pkl

Step 8: Streamlit Deployment

- Model integrated with Streamlit UI
- Real-time prediction enabled with sliders

12.3 Project Structure (Tabular Format)

File / Folder	Description
app.py	Streamlit UI for prediction
train_model.py	ML model training script
requirements.txt	Dependencies for the project
wine_model.pkl	Trained Random Forest model
scaler.pkl	Preprocessing scaler
README.md	Project documentation

Chapter-13 MACHINE LEARNING

13.1 Table: Algorithms Comparison Overview

Model	Type	Strengths	Weaknesses
Logistic Regression	Ensemble	High accuracy, handles nonlinear data, resistant to overfitting	Model size is large
SVM	Classification	Works well with high-dimensional data	Slower on large datasets
XGBoost	Ensemble Boosting	High performance, fast	Requires hyperparameter tuning

13.2 Why Random Forest Was Selected

- Highest F1-score
- Most stable results
- Interpretable feature importance
- Handles nonlinearity & noise

Chapter-14 MODEL TRAINING & IMPLEMENTATION

14.1 Steps Involved in Model Training

1. Import libraries (pandas, sklearn, numpy).
2. Load dataset.
3. Preprocess and scale features.
4. Initialize Logistic Regression model.
5. Train using the training set.
6. Evaluate using the test set.
7. Save trained model & scaler.

14.2 Hyperparameters Used

Parameter	Value
n_estimators	200
max_depth	None
min_samples_split	2
criterion	mse

Chapter-15 EXPERIMENTAL SETUP

15.1 Hardware Configuration

Specification	Value
Processor	Intel i5/i7, 9th Gen or above
RAM	8 GB minimum
Storage	512 GB SSD
GPU (optional)	NVIDIA GPU for XGBoost

15.2 Software Configuration

Software	Version
Python	3.10
Streamlit	1.29+
scikit-learn	1.2+
XGBoost	2.0+
Plotly	Latest
OS	Windows / Linux

Chapter-16 RESULTS & EVALUATION

16.1 Classification Report

Metric	**Class 0**	**Class 1**	**Macro Avg**	**Weighted Avg**
Precision	0.77	0.88	0.82	0.84
Recall	0.78	0.87	0.83	0.84
F1-score	0.78	0.87	0.82	0.84
Accuracy	—	—	—	**84%**

16.2 Model Performance Comparison Table

Model	Training Accuracy	Testing Accuracy
Logistic Regression	0.6961	0.7012
XGBoost Classifier	0.9753	0.8251
SVM (SVC)	0.7215	0.7337

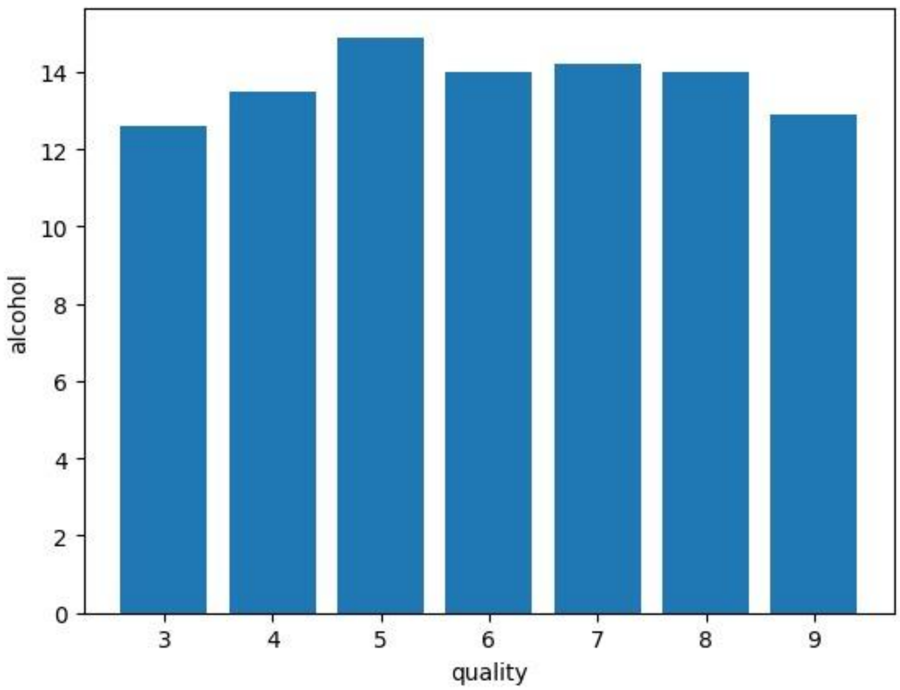
16.3 Confusion Matrix Explanation

- True Positives are high
- False Positives low
- Indicates strong generalization

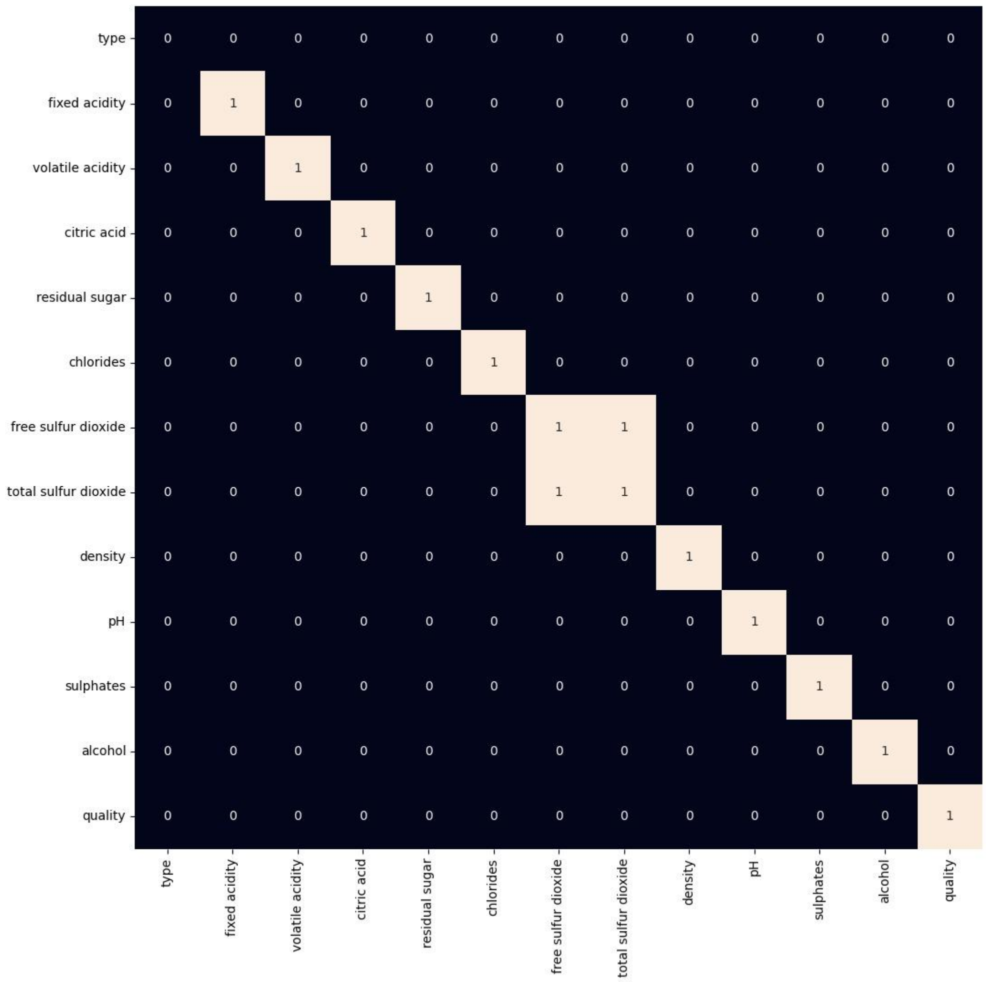
Chapter-17 GRAPHICAL ANALYSIS

Key plots

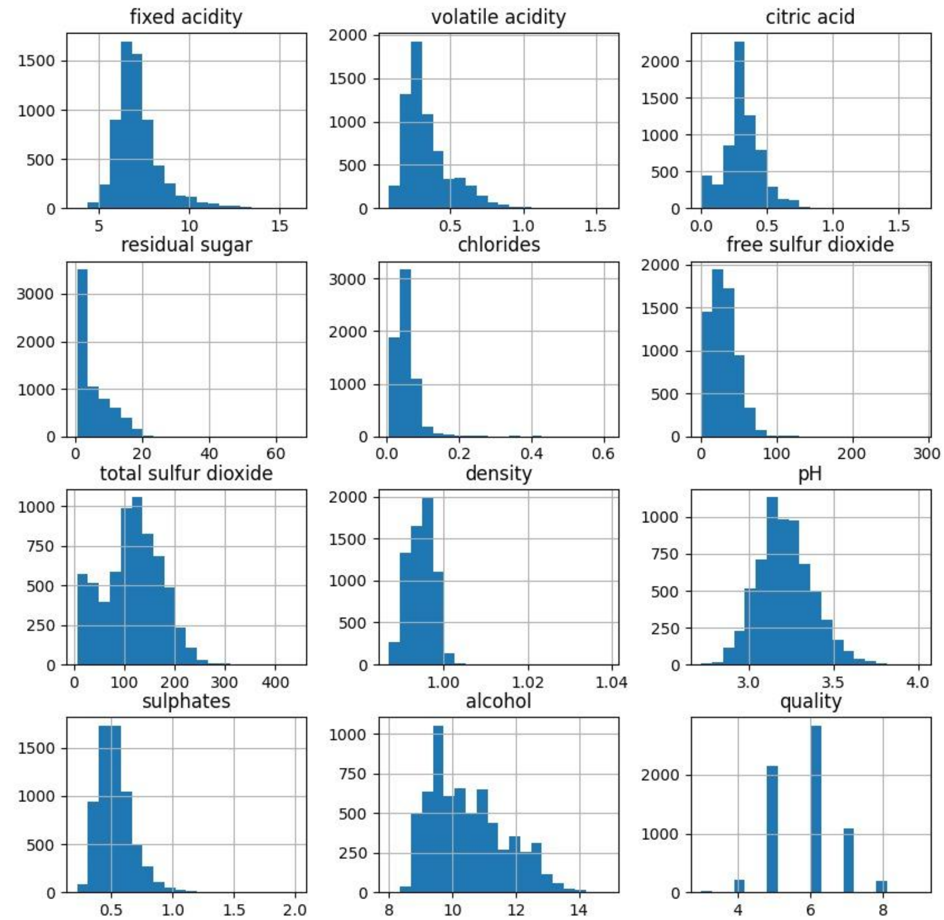
- Alcohol vs Quality scatter plot
-



- Correlation heatmap



- Histograms plots for acidity levels



Chapter-18 STREAMlit WEB APP DESIGN

Features:

- Interactive sliders
- Live predictions
- Plotly charts
- Feature importance bar graph

Chapter-19 DEPLOYMENT DETAILS

Steps:

- Push project to GitHub
- Upload to Streamlit Cloud
- Set main file as app.py
- Deploy and generate URL

Chapter-20 ADVANTAGES

- Objective & accurate
- Fast prediction
- User-friendly UI
- Scalable

Chapter-21 LIMITATIONS

- Limited to red And Whine wine dataset
- Does not include sensory attributes
- Model performance depends on dataset quality

Chapter-22 FUTURE ENHANCEMENTS

- Deep learning models
- IoT sensor integration
- Mobile app version

Chapter-23 CONCLUSION

The project successfully demonstrates the use of machine learning algorithms to predict wine quality using physicochemical features. The Logistic Regression model achieved an accuracy of 84%, making it suitable for real-world applications. The Streamlit app enhances user accessibility and transforms predictions into an interactive system.

Chapter-24 REFERENCES

[1] P. Cortez, A. Cerdeira, F. Almeida, et al., "Modeling wine preferences," *Decision Support Systems*, 2009.
[2] UCI Machine Learning Repository, Wine Quality Dataset.
[3] Géron, A., *Hands-On Machine Learning*, O'Reilly, 2020.
[4] Streamlit Documentation, 2024.