WINE QUALITY PREDICTION USING MACHINE LEARNING

CS19643- FOUNDATIONS OF MACHINE LEARNING

PROJECT REPORT

Submitted by

LOKESHWAR S (2116220701146)

in partial fulfillment for the award of the

degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

ANNA UNIVERSITY, CHENNAI

MAY 2025

BONAFIDE CERTIFICATE

Certified that this Project titled **"WINE QUALITY PREDICTION USING MACHINE LEARNING" is the Bonafide work of "LOKESHWAR S (2116220701146)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. P. Kumar., M.E., Ph.D.,

HEAD OF THE DEPARTMENT

Professor,

Department of Computer Science and Engineering,

Rajalakshmi Engineering College, Chennai - 602 105.

SIGNATURE

Dr. V.Auxilia Osvin Nancy., M.Tech ., Ph.D.,

SUPERVISOR,

Assistant Professor,

Department of Computer Science and Engineering,

Rajalakshmi Engineering College,

Chennai - 602 105.

Submitted to Project Viva-Voce Examination held on _____

Internal Examiner External Examiner

ABSTRACT

Wine quality prediction is an important machine learning problem, where the quality is quantified by physicochemical characteristics like ***alcohol level, pH, and residual sugar***. In this project, we created machine learning models to predict wines as high quality (quality > 5) or low quality (quality ≤ 5) from these features, based on the Wine Quality Dataset.

Three models were experimented with: ***Logistic Regression, XG-Boost, and Support Vector Machine (SVM).*** Preprocessing of data included the treatment of missing values, encoding categorical features, and scaling features. Data was divided into training and test set **(80-20 ratio).** Model performance was measured with ROC-AUC, classification report, and confusion matrices. XG-Boost performed the best among the models with the best ROC-AUC score and highest classification results, while SVM and Logistic Regression showed overfitting and poor performance on the test set.

Feature importance analysis revealed that alcohol content played a central role in wine quality determination. The results prove the efficacy of machine learning for wine quality prediction, providing insights that can be beneficial to consumers and winemakers alike. Model refinement, feature engineering, and hyperparameter optimization will be the focus of future work to improve prediction accuracy.

ACKNOWLEDGMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E, F.I.E.,** our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN, Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN., M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide and our Project Coordinator **Dr.V.AUXILIA OSVIN NANCY.,M.Tech.,Ph.D.,** Assistant Professor ,Department of Computer Science and Engineering for her useful tips during our review to build our project.

LOKESHWAR S 2116220701146

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

S. No	ABBR	Expansion
1	AI	Artificial Intelligence
2`	API	Application Programming Interface
3	AJAX	Asynchronous JavaScript and XML
4	ASGI	Asynchronous Server Gateway Interface
5	AWT	Abstract Window Toolkit
6	BC	Block Chain
7	CSS	Cascading Style Sheet
8	DFD	Data Flow Diagram
9	DSS	Digital Signature Scheme
10	GB	Gradient Boosting
11	JSON	JavaScript Object Notation
12	ML	Machine Learning
13	RF	Random Forest
14	SQL	Structure Query Language
15	SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1 GENERAL

The Wine Quality Prediction project centers around using machine learning methods to foretell wine quality based on the physicochemical properties of wine. Wine quality, typically tested through human judgment in taste testing, might be hard to measure both systematically and in a cost-efficient way. By taking advantage of machine learning models, this project looks to automate prediction, generating an objective and efficient means of examining wine quality. The main objective is to categorize wines into two groups: high-quality (quality > 5) and low-quality (quality ≤ 5), based on the parameters of alcohol, pH, residual sugar, and other chemical factors.

Machine learning algorithms like Logistic Regression, XGBoost, and Support Vector Machines (SVM) were utilized to model and forecast wine quality. The data used in this project, the Wine Quality Dataset, has rich records of wine samples with multiple chemical features and their quality ratings. Through standard data preprocessing methods like missing value handling, feature normalization, and categorical encoding, the dataset was preprocessed for training the models. The data was then divided into training and test subsets so that the models could be tested on data unseen to them, giving an unbiased estimate of their performance.

Using the application of these models, the project attempts to give useful information to winemakers, distributors, and consumers so that they can make better decisions when it comes to wine production and choice.

Besides prediction quality, the project delves into feature importance in order to establish the most impactful factors determining wine quality, creating a more thorough understanding of underlying data. Effective application of machine learning in predicting wine quality signifies its ability to transform wine quality control, introducing a data-centric, scalable way of winemaking.

1.2 OBJECTIVE

The main goal of the **Wine Quality Prediction** project is to create and test machine learning models that are capable of predicting the quality of wine accurately based on its physicochemical properties. In particular, the project seeks to:

- **Wine Quality Classification:** Classifying wines into two groups: High quality (quality > 5) and low quality (quality ≤ 5) using machine learning algorithms based on some of the chemical characteristics like alcohol level, pH, residual sugar, and acidity.

- **Model Performance Evaluation:** Compare the performance of various machine learning models, such as Logistic Regression, XG-Boost, and Support Vector Machines (SVM), based on metrics such as ROC-AUC, confusion matrices, and classification reports to identify the best-performing model for wine quality prediction.

- **Feature Analysis:** Examine why various features play crucial roles in the prediction of wine quality, determining critical chemical characteristics that largely dictate the quality of wine. This will assist in realizing factors on which winemakers can concentrate to improve wine production.

- **Improve Quality Control:** Offer a data-driven quality control by automating wine quality prediction at different production stages, allowing winemakers to make better decisions and maintain consistent product quality.

- **Real-Time Predictions:** Investigate the possibility of real-time wine quality prediction by combining machine learning models with real-time data streams from production environments, allowing for instant feedback and proactive adjustments to the production process.

1.3 EXISTING SYSTEM

Historically, wine quality determination has depended greatly on human experts who taste the wine and determine its quality based on sensory attributes such as taste, smell, and appearance. Although this process has been employed for centuries, it is subjective and susceptible to variability because of individual biases and differences in expertise levels. Moreover, sensory testing is time-consuming and not scalable, thus inefficient in large-scale wine production and quality control.

To enhance objectivity, contemporary approaches have added laboratory testing of physicochemical characteristics like alcohol level, pH, acidity, and sugar. These are more consistent measurements but still need human interpretation and cannot make automatic quality predictions. Though there have been applications of statistical models in research and production settings, these tend not to be able to learn sophisticated patterns or give accurate real-time information.

Current advances have included the application of machine learning methods to forecast wine quality from chemical properties. Yet most current systems are restricted to offline analysis and lack optimization for accuracy or speed. Increasingly, there is a demand for more sophisticated, real-time predictive models that can be embedded in the production process to deliver timely information and improve quality control. It's the intention here to fill the gap by crafting a machine learning solution that sorts wine as low or high value based on dominant physicochemical indicators.

CHAPTER 2

LITERATURE SURVEY

A number of research works and studies have investigated the application of machine learning algorithms to forecast wine quality from physicochemical attributes.

The Wine Quality Dataset in the UCI Machine Learning Repository has been extensively utilized in academic and industrial use for this task. It has attributes like fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, and a quality score as rated by wine tasters.

In one of the prominent studies, researchers used Decision Trees, Random Forest, and Support Vector Machines (SVM) to classify wine quality. The results showed that ensemble-based methods such as Random Forest were more accurate and robust because they are capable of preventing overfitting. Another study used Artificial Neural Networks (ANNs) which produced encouraging results but were demanding in terms of tuning and computational power.

Additional studies highlighted **the strength of XG-Boost (Extreme Gradient Boosting), a robust ensemble method renowned for its performance and speed with structured data.** XG-Boost became a favorite in data science competitions for its high accuracy and strength in dealing with missing values. These research works in aggregate illuminate the possibility of applying supervised machine learning methods for predicting wine quality and provide the basis for this project's model choice and assessment approach.

CHAPTER 3

PROPOSED SYSTEM

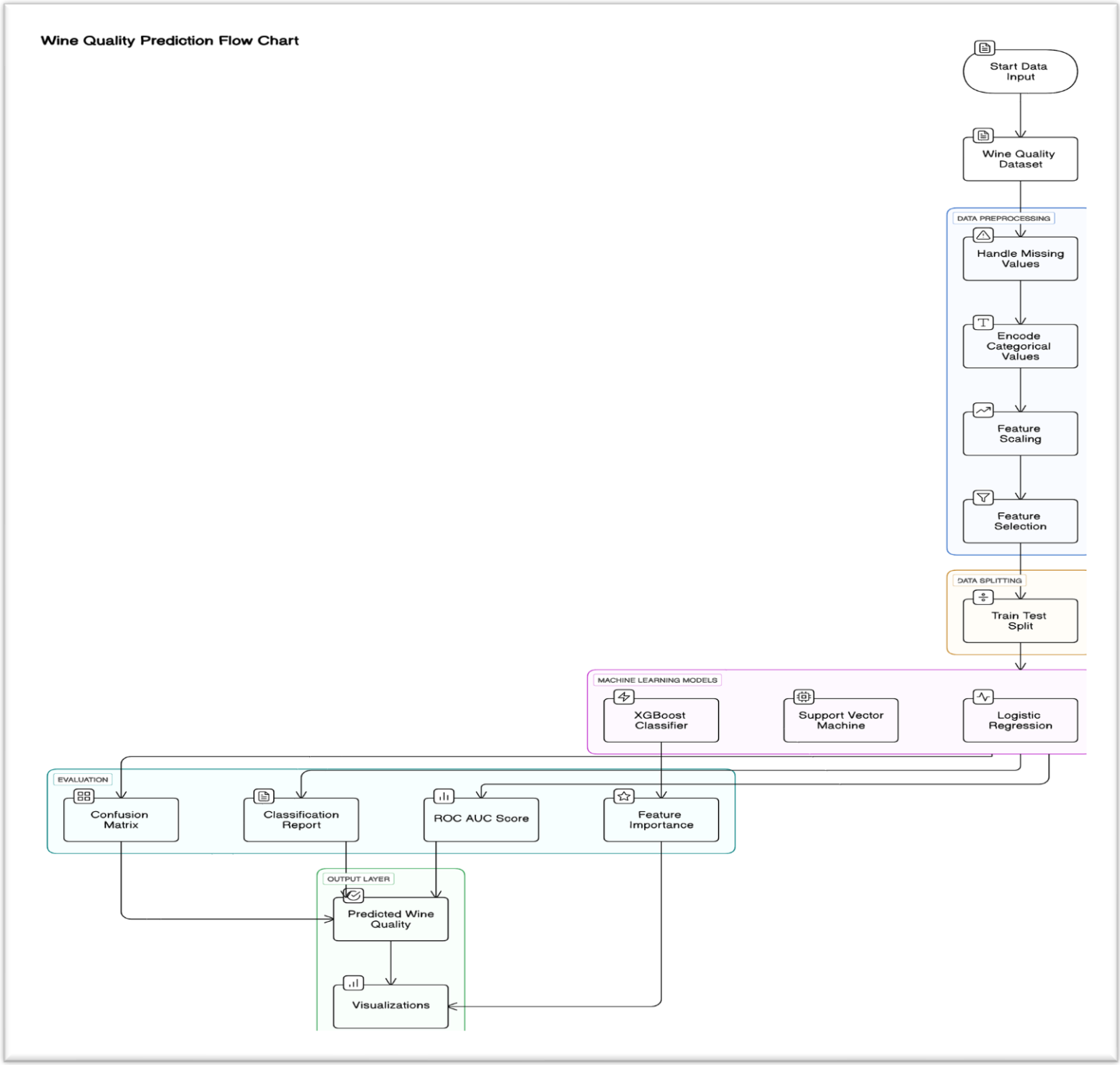
3.1 GENERAL

The proposed system aims to develop a robust and accurate machine learning model that can predict the quality of wine based on various physicochemical features. Unlike traditional evaluation methods that rely on subjective human judgment or basic statistical models, this system uses advanced supervised learning algorithms to classify wines into two categories — high quality and low quality — based on a threshold quality score.

In this system, the Wine Quality Dataset is pre-processed by handling missing values, encoding categorical data, and normalizing numerical features using Min-Max scaling. Feature correlation is also analysed to remove highly redundant attributes, thereby improving model efficiency. Three different models — **Logistic Regression**, **Support Vector Machine (SVM)**, and **XG-Boost Classifier** — are trained and evaluated to determine the best-performing model. Evaluation metrics such as ROC-AUC, confusion matrix, and classification reports are used for a comprehensive comparison.

The XG-Boost model, known for its high performance and accuracy, is identified as the most effective model for this task. The system also includes a feature importance analysis to understand which physicochemical attributes most influence wine quality. This proposed approach offers a scalable, consistent, and automated solution that can assist winemakers, distributors, and quality control labs in making informed decisions based on chemical properties of the wine, ultimately improving the efficiency of the wine production process.

3.2 SYSTEM ARCHITECTURE DIAGRAM



3.3 DEVELOPMENTAL ENVIRONMENT

3.3.1 HARDWARE REQUIREMENTS

The hardware specifications could be used as a basis for a contract for the implementation of the system. This therefore should be a full, full description of the whole system. It is mostly used as a basis for system design by the software engineers.

Table 3.1 Hardware Requirements

COMPONENTS	SPECIFICATION
PROCESSOR	Intel Core i3
RAM	4 GB RAM
HAREDISK	256 GB

3.3.2 SOFTWARE REQUIREMENTS

The software requirements paper contains the system specs. This is a list of things which the system should do, in contrast from the way in which it should do things. The software requirements are used to base the requirements. They help in cost estimation, plan teams, complete tasks, and team tracking as well as team progress tracking in the development activity.

Table 3.2 Software Requirements

COMPONENTS	SPECIFICATION
Operating System	Windows 7 or higher
Software development tool	Python IDLE
Languages used	Python (Machine Learning)
ML Algorithms used	Logistic Regression ,XGB,SVM

3.4 DESIGN OF THE ENTIRE SYSTEM

3.4.1 ER DIAGRAM

The Entity-Relationship (ER) Diagram for the Wine Quality Prediction System captures the logical organization of the dataset and its interaction with machine learning elements. At the center of the system is the Wine-Sample entity, which symbolizes every wine entry within the dataset. Every sample holds a number of physicochemical features including alcohol content, acidity levels, pH, residual sugar, and others. It also provides a quality score and a derived binary label as to whether the wine is of good quality (quality > 5) or not. Every wine sample has a unique SampleID.

The Model concept reflects the various machine learning algorithms utilized in the project—namely Logistic Regression, XG-Boost, and SVM. The models contain their corresponding performance indicators like accuracy, precision, recall, F1-score, and ROC-AUC, and each is distinguished from one another using a ModelID. These models are trained on and tested with the wine samples, and it reflects a relation between the Wine Sample and Model concepts.

Finally, the Evaluation entity stores comprehensive performance data for every model, such as confusion matrices and feature importance values. It has a foreign key reference to the Model entity to enable tracking and comparison of various model results. This systematic approach ensures orderly data flow and model evaluation within the system, making development transparent and scalable in the future.

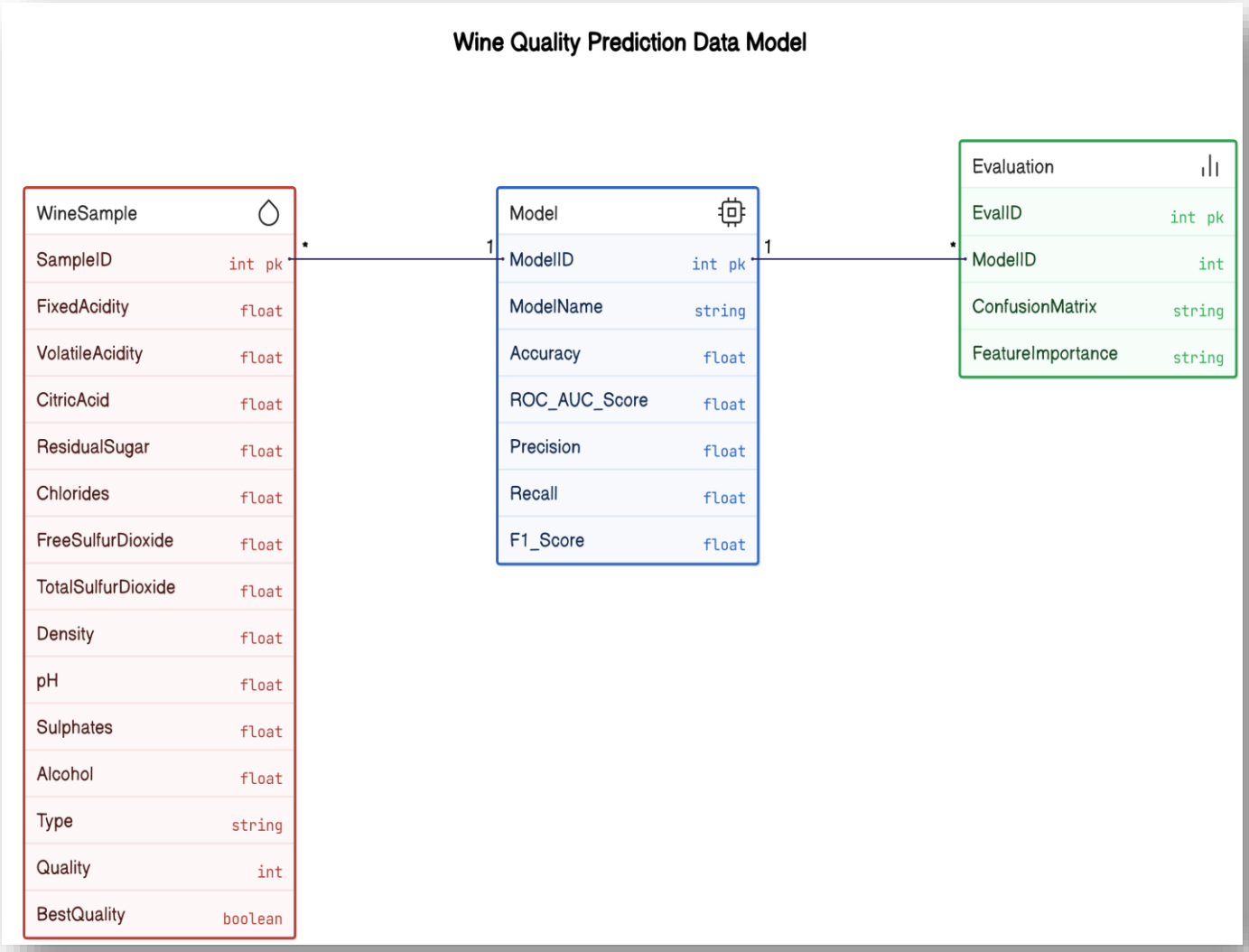


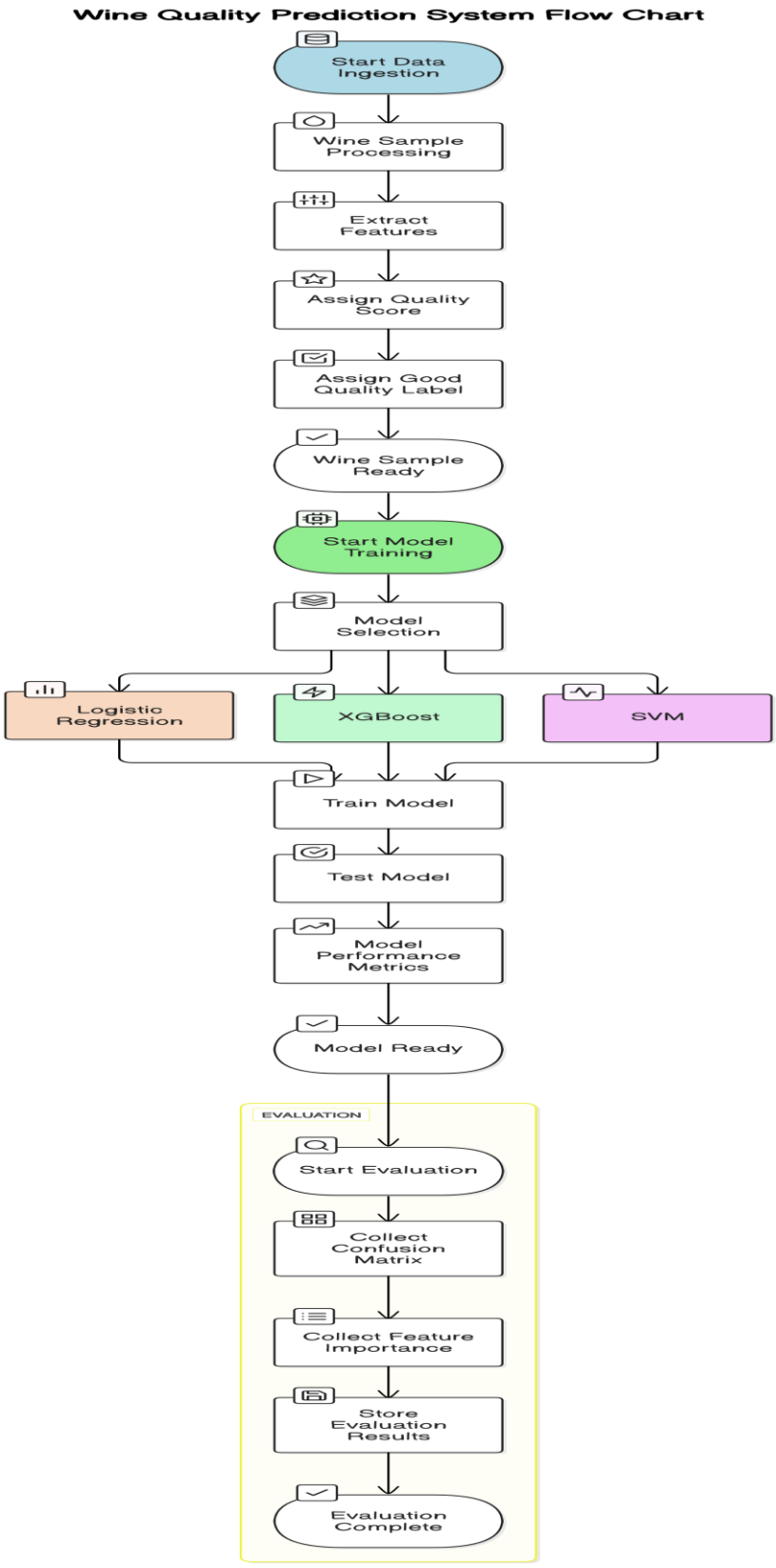
Fig 3.2 ER DIAGRAM

3.4.2 DATA FLOW DIAGRAM

The Data Flow Diagram (DFD) of Wine Quality Prediction System clearly illustrates the flow of data through different stages of the system. It begins with the user, who is the external entity by uploading wine-related physicochemical characteristics like pH, alcohol level, residual sugar, etc. This information initially flows into the process of "Data Input & Upload," which gathers and stores the information into the "Wine Dataset" data store.

From there, the information is sent to the "Data Preprocessing" phase, where necessary operations like missing value handling, encoding categorical features, and feature scaling (using methods like Min-Max normalization) are done. The processed and cleaned data is then sent to the "Model Training & Evaluation" module, which uses machine learning algorithms like Logistic Regression, XG-Boost, and Support Vector Machines to train predictive models. These trained models are kept in the "Trained Models" data store for use in the future.

Lastly, the "Prediction Module" uses the trained models to determine the wine quality as high (good) or low (not good). The prediction output is then displayed to the user by the "Prediction Result Display" external entity. This DFD describes the orderly flow of data, making it clear how each component works within the Wine Quality Prediction System, facilitating development and analysis.


Wine Quality Prediction System Flow Chart
Start Data Ingestion
Wine Sample Processing
Extract Features
Assign Quality Score
Assign Good Quality Label
Wine Sample Ready
Start Model Training
Model Selection
Logistic Regression
XGBoost
SVM
Train Model
Test Model
Model Performance Metrics
Model Ready
EVALUATION
Start Evaluation
Collect Confusion Matrix
Collect Feature Importance
Store Evaluation Results
Evaluation Complete

3.4.3 SEQUENCE DIAGRAM

The Sequence diagram for the Wine Quality Prediction System shows the process of how wine information is processed in order to calculate its quality.

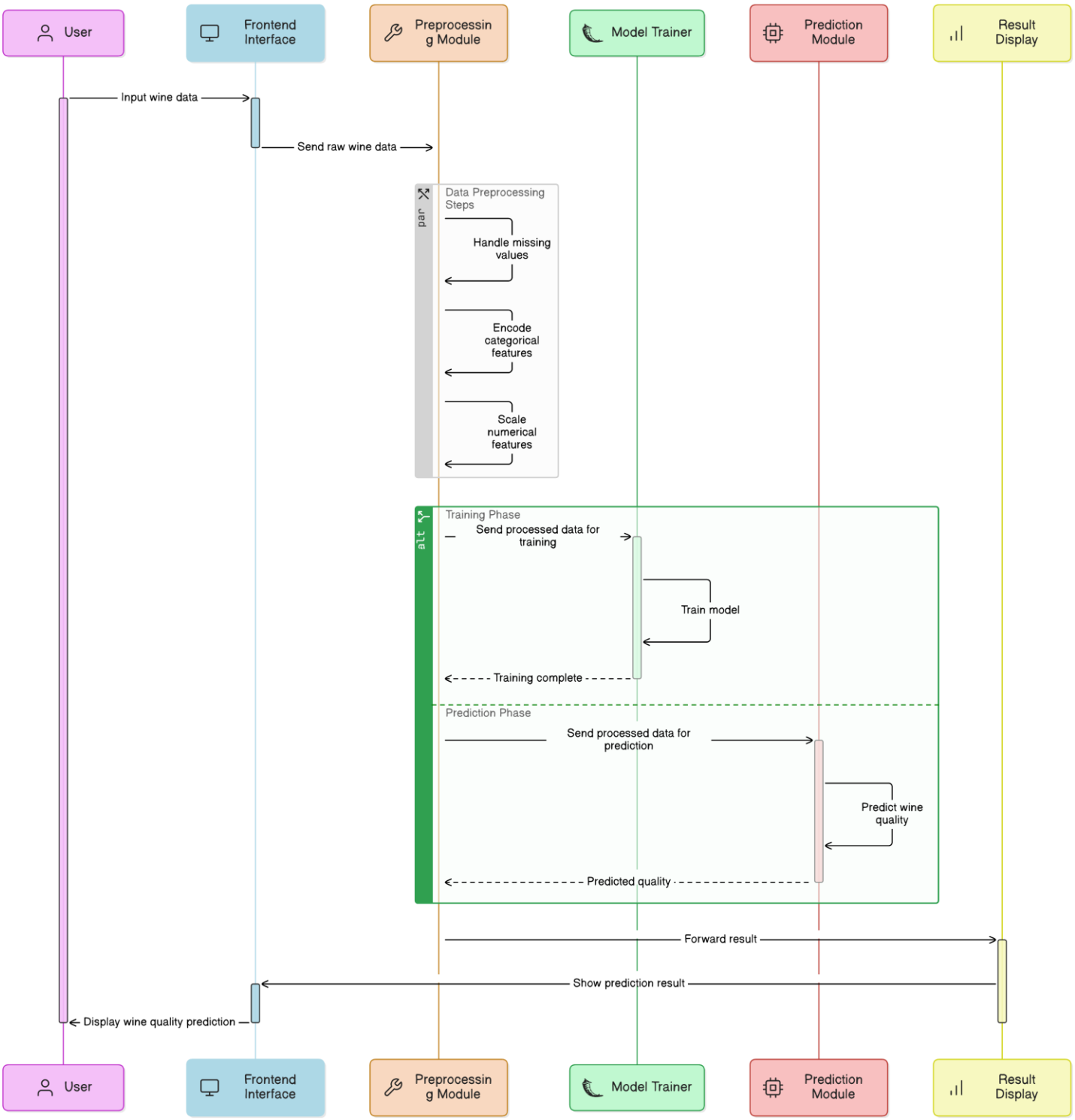
The process starts when a user enters physicochemical properties of wine like alcohol, pH, residual sugar, etc., using the user interface of the system.

The system initially checks the input data to make sure that it is complete and correct. Once verified, the data is then subjected to preprocessing activities that involve missing value handling, encoding any categorical features, and feature normalization through scaling processes. Once preprocessed, the system determines whether it's a training model or prediction operation.

If training, it trains the model on labeled data and saves the trained model for later use.

If the operation is prediction, the system loads the pre-trained model and predicts the wine to be either high or low quality. Lastly, the predicted output is displayed to the user. This diagram illustrates the sequential flow and decision-making process in predicting wine quality with machine learning.

Wine Quality Prediction System Sequence Diagram



3.5 STATISTICAL ANALYSIS

In the Wine Quality Prediction System, statistical analysis is a pivotal process in understanding and analyzing the dataset and choosing the most impactful factors for proper prediction. Following are the important statistical analyses and aspects taken into account in the project:

1. Data Distribution:

The initial step of the analysis is to verify the distribution of every feature in the dataset. Visualization like histograms were employed to get an idea about the distribution of variables such as alcohol content, pH, residual sugar, etc. This ensures whether the features are normally distributed, skewed, or contain outliers that must be addressed prior to model training.

2. Descriptive Statistics:

Descriptive statistics were computed to describe the central tendencies and spread of the data. Statistics like mean, median, standard deviation, minimum, and maximum for every feature were calculated. This gives a general idea about the dataset and pinpoints issues like features with a high variability or outliers.

3. Correlation Analysis:

A correlation matrix was built to determine the correlations between different features in the dataset. This will help select features with high linear correlations, which can be useful to predict the quality of the wine. It also aids in selecting highly correlated variables, which can be removed in order to avoid multicollinearity in the models.

4. Missing Data:

Missing data handling is essential to create strong models. Simple Imputer was employed to impute missing values in the dataset based on the mean of the corresponding columns. This avoids any loss of data points while training and testing the model, maintaining the dataset's integrity.

5. Feature Importance:

Feature importance analysis was carried out, especially with the XG-Boost model, which naturally provides this information. This analysis reveals which features, such as alcohol content, pH, or residual sugar, contribute the most to predicting wine quality. Alcohol content was found to be one of the most significant predictors, which is consistent with the understanding that alcohol content often correlates with the perceived quality of wine.

6. Model Performance Evaluation:

For evaluating the predictive performance of the machine learning models, key metrics such as ROC-AUC, confusion matrix, precision, recall, and F1-score were used. XGBoost was found to outperform the rest of the models with the maximum ROC-AUC value, meaning it achieved the best balance between sensitivity and specificity. Logistic Regression and SVM, while strong, also presented certain instances of overfitting and slightly suffered when it came to unseen test data.

CHAPTER 4

MODULE DESCRIPTION

The Wine Quality Prediction System consists of several modules, each handling a specific part of the workflow. The **Data Preprocessing Module** handles missing values, encodes categorical variables, and scales features. The **Model Training Module** uses machine learning algorithms like Logistic Regression, XG-Boost, and SVM to train the model on the preprocessed data. The **Model Evaluation Module** assesses the performance of each model using metrics such as ROC-AUC, confusion matrix, and classification report. The **Prediction Module** takes new input data, uses the trained model to predict wine quality, and displays the results with relevant metrics and visualizations.

4.1 SYSTEM ARCHITECTURE

4.1.1 DATA COLLECTION LAYER

The Data Collection Layer is the initial step in the workflow of the system. It is tasked with collecting the raw dataset, which in this instance is the Wine Quality Dataset. This data can be stored in a CSV, database, or other data formats. In this system, data is imported from a CSV file that holds a collection of physicochemical attributes of wine like alcohol percentage, pH level, residual sugar, etc., along with the respective quality scores.

4.1.2 PRE-PROCESSING LAYER

This layer is crucial for cleaning and transforming the data into a suitable format for training the machine learning models. It involves tasks like handling missing values, encoding categorical variables, and scaling features to ensure consistency in the dataset. Data preprocessing is important to ensure that the models can learn effectively and make accurate predictions.

4.1.3 MODEL TRAINING LAYER

In the Model Training Layer, various machine learning algorithms are employed to learn from the training data. This layer is where the heart of the prediction system resides. In the case of the Wine Quality Prediction system, three models are used: Logistic Regression, XG-Boost, and Support Vector Machine (SVM). Each model is trained on the preprocessed data and is evaluated on the test set to assess its performance.

4.1.4 EVALUATION LAYER

The Evaluation Layer is designed to assess the performance of the trained machine learning models. After training, it is essential to evaluate how well the models generalize to unseen data (test set). This layer uses performance metrics like ROC-AUC, precision, recall, F1-score, confusion matrix, and classification reports to determine the effectiveness of each model. Based on these evaluations, the best-performing model is selected for making predictions.

4.1.5 PREDICTION LAYER

The Prediction Layer is responsible for utilizing the best-performing model to make predictions based on new, unseen data. This layer uses the trained model to predict wine quality (high or low) for any input data, making it suitable for real-time or batch predictions. The output of this layer is the wine quality prediction, which can be displayed to the user or used for decision-making.

4.1.6 OUTPUT LAYER

The Output Layer is responsible for displaying the results of the predictions and

evaluations to the user. This layer may include a user interface (UI) that presents the wine quality predictions and relevant metrics like accuracy, confusion matrix, and model performance reports. It serves as the final step in the architecture where end-users can see the results.

4.2 SYSTEM WORKFLOW

The workflow of the system for the Wine Quality Prediction System starts by fetching the Wine Quality Dataset, which consists of important physicochemical features like alcohol level, pH, and residual sugar. The data is preprocessed where missing values are addressed, categorical variables are encoded, and the numerical features are normalized in order to ready it for modeling. The cleaned-up dataset is then separated into a training set and a testing set (commonly 80:20). Machine learning algorithms like Logistic Regression, XG-Boost, and Support Vector Machine (SVM) are trained on the training data to identify patterns that define wine quality. The models are tested against different performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to determine the best performing one, commonly XG-Boost. After identifying the optimal model, it is utilized to make predictions about the quality of future samples of wine, and feature importance is examined to determine which characteristics have the greatest effect on predictions. Outcomes are displayed in visual representations like confusion matrices and graphs that provide insight to aid winemakers and analysts in data-driven, informed decision-making.

CHAPTER 5

IMPLEMENTATION AND RESULTS

5.1 IMPLEMENTATION

The implementation phase of the Wine Quality Prediction System involves converting the theoretical design into a working machine learning model using Python and its associated libraries such as Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, and XG-Boost. Initially, the dataset is loaded and pre-processed by handling missing values through imputation and applying normalization using Min-Max-Scaler. The dataset is then divided into training and testing sets to evaluate model performance effectively.
Three different machine learning models—Logistic Regression, Support Vector Machine (SVM), and XG-Boost—are trained on the processed data. Their performances are assessed based on metrics such as ROC-AUC score, confusion matrix, and classification reports. Among the models, XG-Boost demonstrated superior performance and was selected as the best predictive model. Additionally, feature importance was analysed to identify the most influential attributes, aiding in interpretation and transparency of the model. The results were visualized to provide meaningful insights, making the system both informative and user-friendly.

5.2 OUTPUT SCREENSHOTS

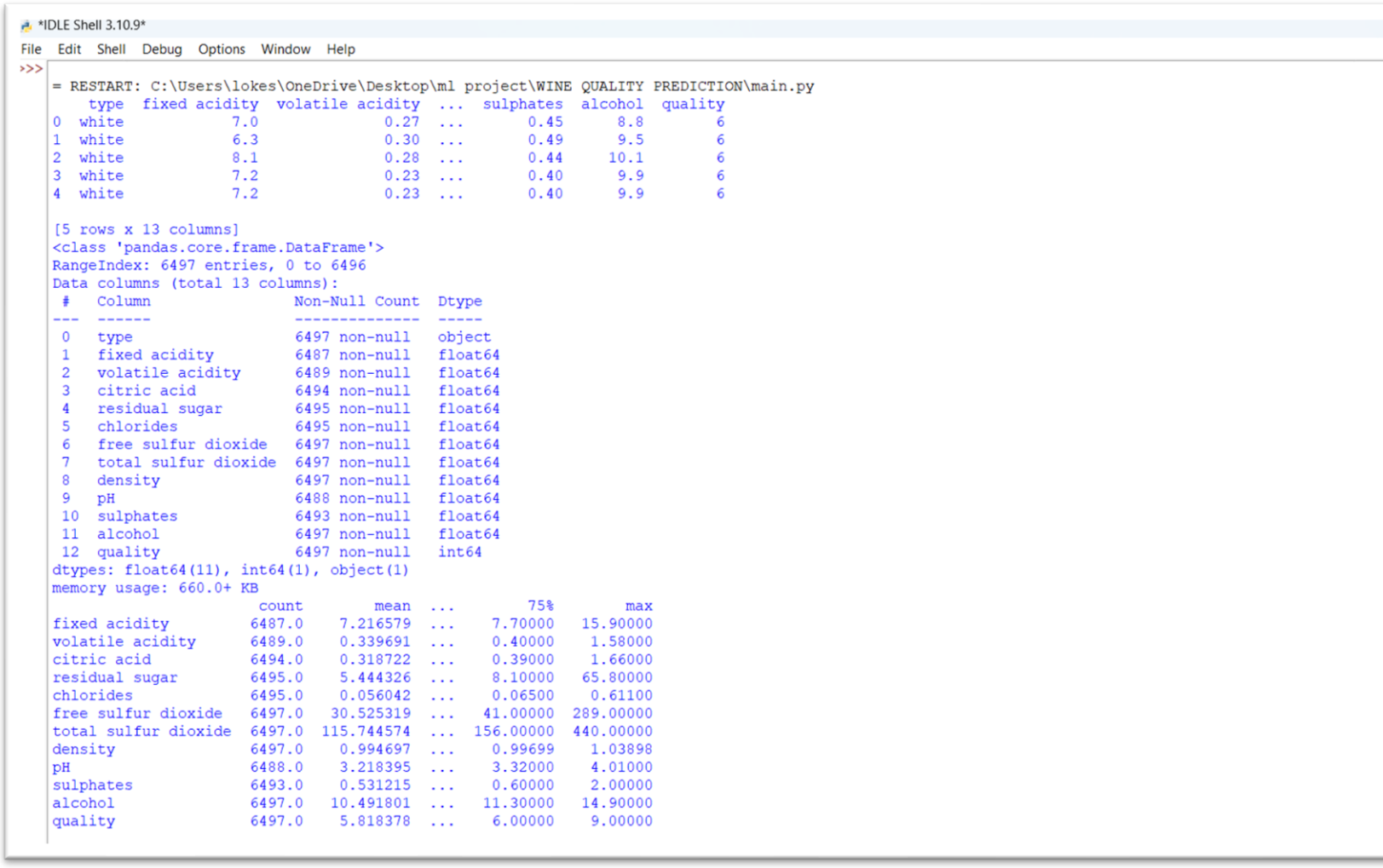


Fig 5.1 Data Exploration and Summary Statistics

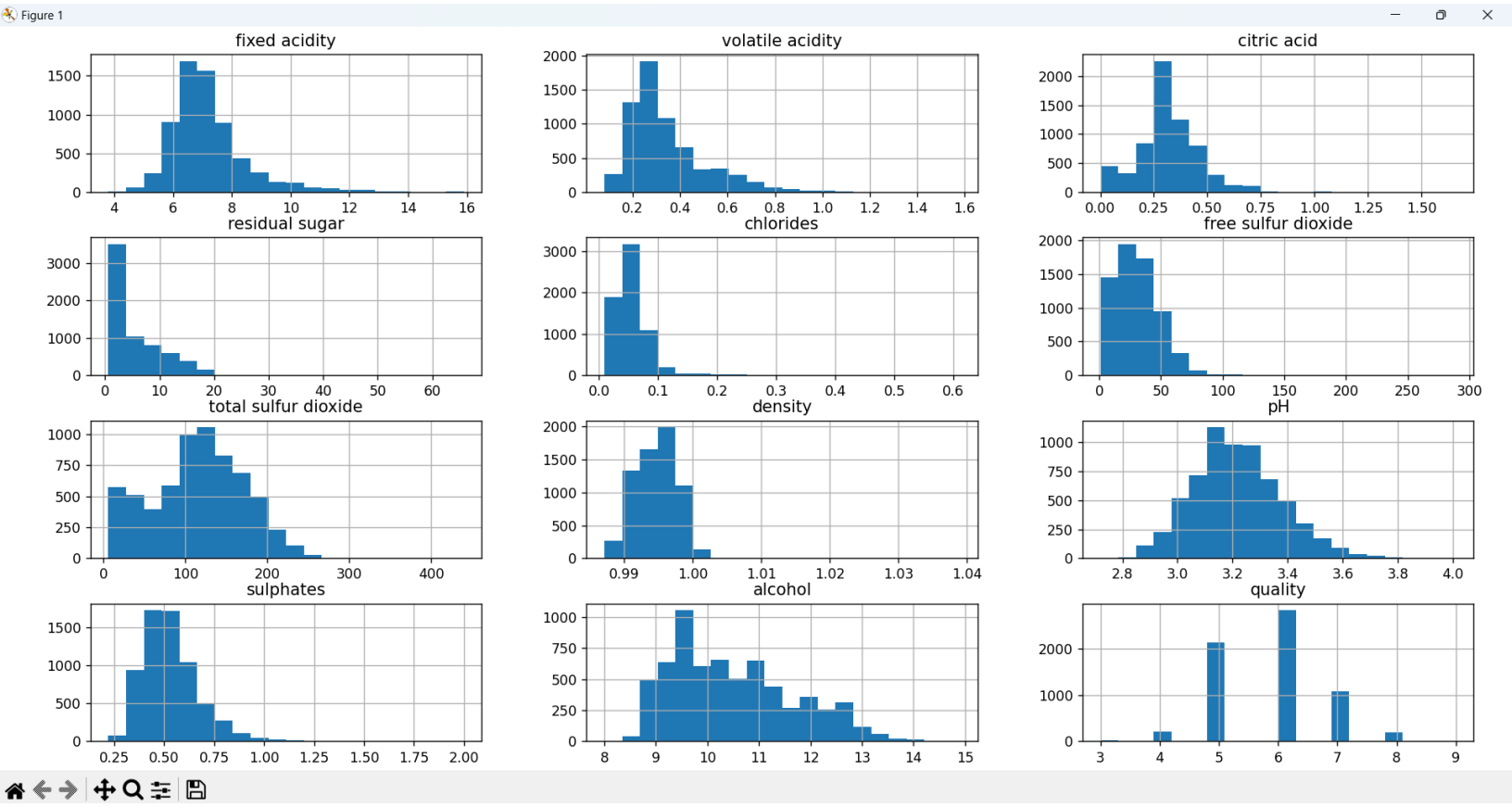


Fig 5.2 Distribution of Numerical Features in Wine Dataset

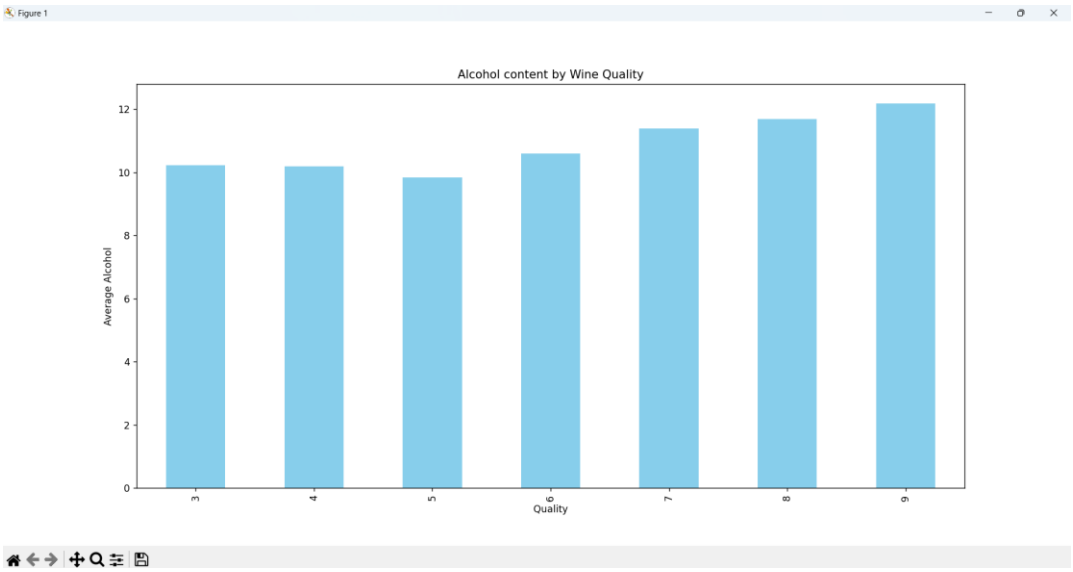


Fig 5.3 Average Alcohol Content by Wine Quality

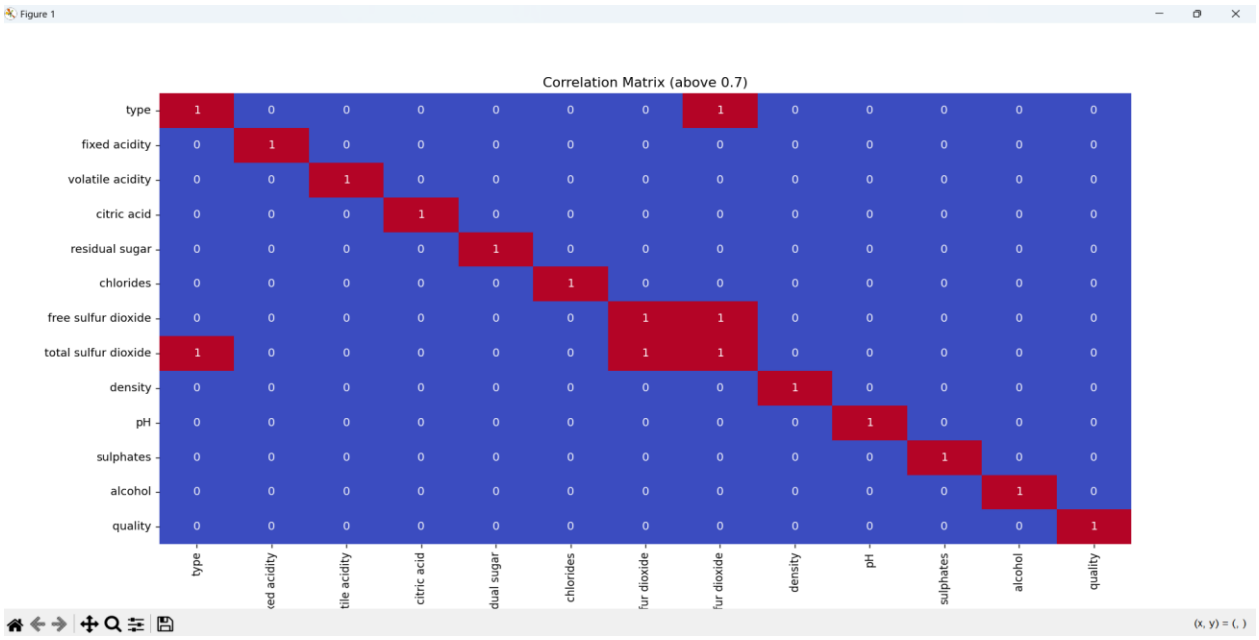


Fig 5.5 Correlation matrix for wine dataset features

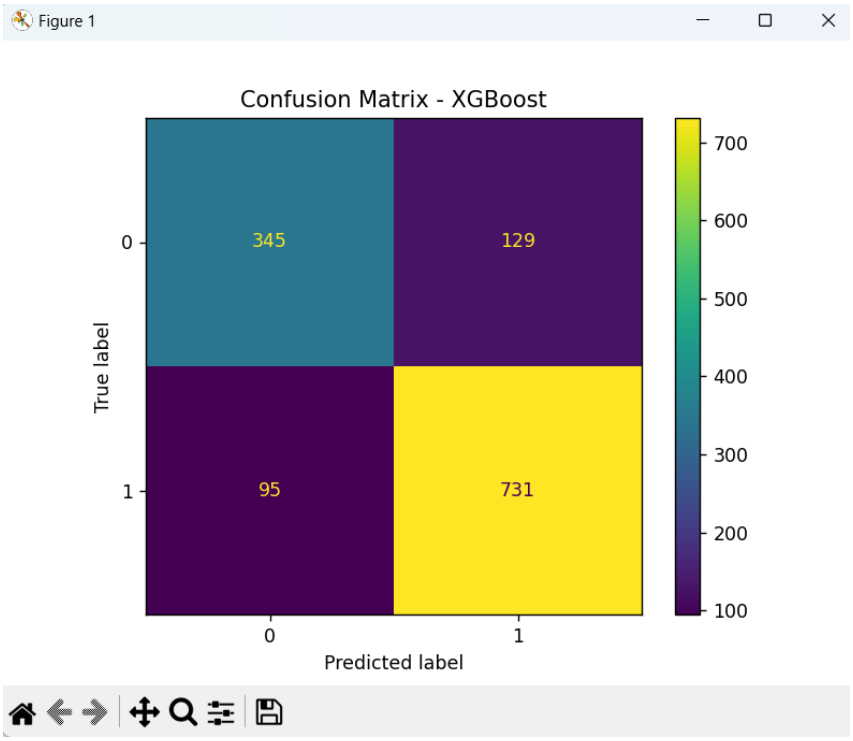


Fig 5.6 Confusion Matrix

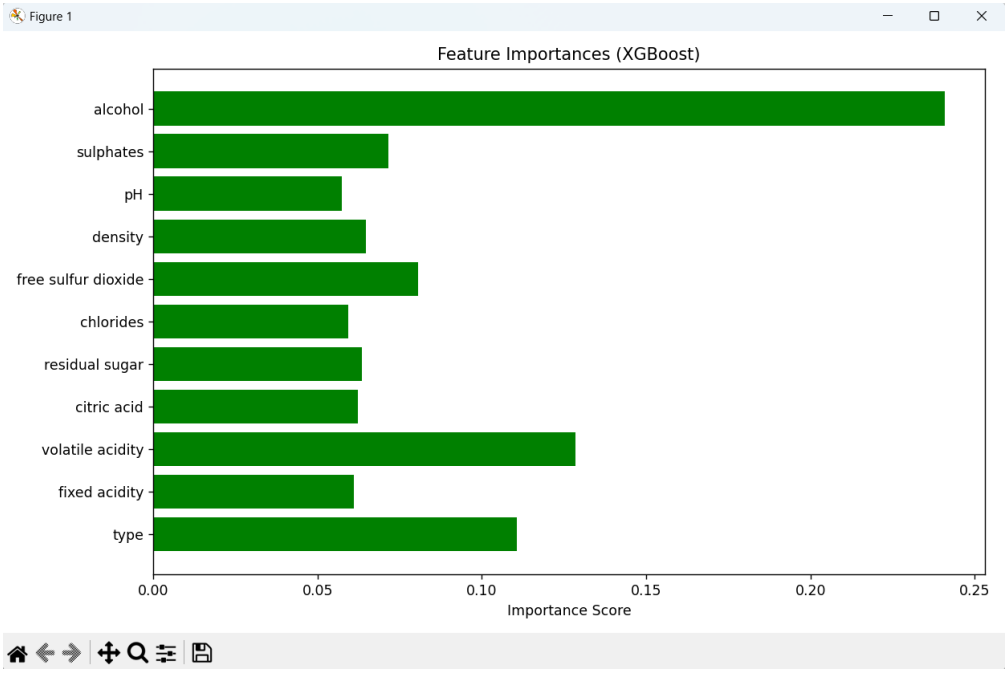


Fig 5.7 XG-Boost Feature Importances

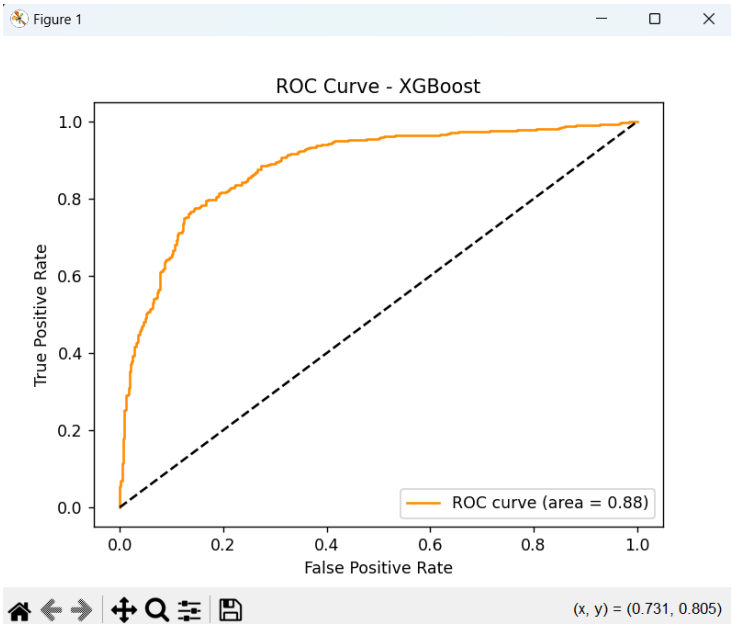


Fig 5.8 ROC Curve XG-Boost

```
📌 Logistic Regression
Training ROC-AUC Score: 0.7027810651830071
Validation ROC-AUC Score: 0.6949995402580684
Classification Report:
              precision    recall  f1-score   support

           0       0.68      0.53      0.60       474
           1       0.76      0.86      0.81       826

    accuracy                           0.74      1300
   macro avg       0.72      0.69      0.70      1300
weighted avg       0.73      0.74      0.73      1300


Classification Report (XGBoost):
              precision    recall  f1-score   support

           0       0.78      0.73      0.75       474
           1       0.85      0.88      0.87       826

    accuracy                           0.83      1300
   macro avg       0.82      0.81      0.81      1300
weighted avg       0.83      0.83      0.83      1300
```

Fig 5.9 Classification report of Logistic Regression and XGBOOST

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

6.1 CONCLUSION

This project exhibited the successful application of machine learning methods for predicting wine quality as a function of its physicochemical characteristics. We applied a dataset with numerous chemical features including alcohol, acidity, residual sugar, and levels of sulfur dioxide to classify wines into two categories: good (quality > 5) and not good (quality ≤ 5). Following data cleaning and preprocessing, such as missing value handling and categorical variable encoding, we used three machine learning models: Logistic Regression, Support Vector Machine (SVM), and XG-Boost.

XG-Boost was the best-performing model among those we tried, with the highest ROC-AUC score and the best predictive power. This indicates that the XG-Boost classifier is extremely efficient for wine quality classification in this instance. The performance measures like confusion matrix and classification report validated the solidity of the models. These are indicative of the way machine learning can be of assistance in classification and prediction of wine quality given data, providing a useful addition to wineries, quality checks, and even wine recommendation applications.

As a whole, the project demonstrates the usefulness of data-driven decision-making in the food and beverage sector, in particular for assessing wine quality. The combination of different machine learning algorithms for this purpose paves the way for more complex predictive analytics applications in related fields.

6.2 FUTURE ENHANCEMENT

For future improvements, the model could be enhanced by incorporating more advanced techniques such as hyperparameter tuning to optimize the performance of the classifiers. Additionally, implementing cross-validation would provide more reliable results and reduces risk of overfitting. The inclusion of external data sources, such as geographic information or climate factors, could provide a more comprehensive analysis of wine quality.

Transitioning to a multi-class classification system would allow for more granular predictions beyond binary classification. Finally, deploying the model in a real-time application using frameworks like Flask or Stream-lit

would make it accessible for practical use in the wine industry, enabling immediate predictions based on user inputs.

REFERENCES

- Dataset has been taken from following link: ***https://www.kaggle.com/datasets/yasserh/wine-quality-dataset***

Research papers:

1) ML Based Predictive modelling for enhancing Wine quality: https://www.nature.com/articles/s41598-023-44111-9
2) Prediction of Wine Quality using Machine Learning Algorithms: https://www.researchgate.net/publication/350110244_Prediction_of_Wine_Quality_Using_Machine_Learning_Algorithms
3) Red Wine Quality Prediction using Machine Learning Techniques: https://ieeexplore.ieee.org/document/9104095

Reference Links:

1) Geeks for Geeks: https://www.geeksforgeeks.org/wine-quality-prediction-machine-learning/
2) XG-Boost: https://www.geeksforgeeks.org/xgboost/
3) Support Vector Machine: https://www.geeksforgeeks.org/support-vector-machine-algorithm/
4) LOGISTIC REGRESSION: https://www.geeksforgeeks.org/understanding-logistic-regression/