International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Analyzing the Effectiveness of Ensemble and Classical Machine Learning Techniques in Wine Quality Prediction

Mukesh Rani[1] and Sunil Kumar[2]

[1]Department of Chemistry, Govt PG College , Hisar
[2]Department of AI and Data Science, Guru Jambheshwar University of Science and Technology.

Abstract
Wine quality prediction is of high interest to food chemists owing to its significance for quality assurance and consumer satisfaction. Conventional tasting techniques used to assess wine quality today are unsophisticated, expensive and time-consuming with requiring semi-skilled personnel for expert sensory analysis. Therefore, the present study aims to provide a comparative study of ensemble and classical machine learning techniques for the prediction of wine quality from 12 physicochemical properties derived from wine samples. The study includes classical machine learning models such as Logistic Regression, Decision Tree, K-Nearest Neighbors, and Support Vector Machine, along with bagging-based ensemble models including Random Forest and Extra Trees, and boosting-based ensemble models such as XGBoost, LightGBM, and CatBoost. In this paper, we have also used some preprocessing techniques such as KNN imputation, power transformation, feature standardization and SMOTE to evaluate these nine machine learning algorithms. The experimental results reveal that the Random Forest classifier performed best, with 82.73% accuracy and an ROC-AUC score of 0.8869. In this paper, we have also made efforts to measure the computational performance of these learning techniques, such as memory usage, CPU usage, execution time and time complexity. The results demonstrate that ensemble methods effectively predict wine quality and support automated quality assessment in the wine industry.

Keywords: Artificial Intelligence, Wine Quality Prediction, Machine Learning, Logistic Regression, Decision Tree, K-Nearest Neighbors, and Support Vector Machine, Random Forest, Extra Trees, XGBoost, LightGBM, CatBoost etc.

1. Introduction

Wine quality assessment is essential within the food chemistry and beverage industry as it directly influences product reliability, consumer satisfaction, and market value [1]. Wine quality is usually evaluated by experts based on sensory features including aroma, acidity, taste, and alcohol concentration. However, manual sensory analysis is unsophisticated, expensive and time-consuming. This may also require semi-skilled analysts and even can produce inconsistent results due to human variations [2]. Hence, in the last decades there has been an increasing interest in developing automated and intelligent wine quality prediction systems.

Over the last decades, developments and advancements in artificial intelligence and machine learning have made it possible to automatically predict wine quality from physicochemical properties of wine samples. By identifying hidden patterns and nonlinear relationships among chemical parameters, machine learning algorithms can develop accurate, efficient, automated, and intelligent prediction systems [3]. The recent studies in the literature have shown that classical machine learning models such as Logistic Regression, Decision Tree, K-Nearest Neighbors and Support Vector Machine exhibit good performance in prediction and classification tasks [4-8]. Similarly, the ensemble methods such as bagging-based ensemble models (Random Forest, Extra Trees) and boosting-based ensemble models (XGBoost and LightGBM and CatBoost) are also used extensively in designing the accurate, efficient, automated, and intelligent prediction and classification solutions for different nature of problems in science, engineering and other sectors [9-12]. The ensemble learning methods are increasingly popular because they improve the accuracy of classifiers, alleviate overfitting issues, and can efficiently deal with complex nonlinear datasets [9]. In addition, sophisticated data preprocessing techniques such as KNN imputation, power transformation, feature standardization and Synthetic Minority Oversampling Technique (SMOTE) help these machine learning models to become more robust and improve classification performance on imbalanced datasets [13].

Cortez et al. [2] presented one of the most commonly used wine quality datasets and showed how data mining techniques could be utilized in predicting wine quality. The work presented in this paper is inspired by recent advancements in artificial intelligence for developing automated, intelligent, and accurate classification and prediction systems. This paper presents the comparative performance evaluation of ensemble and classical machine learning techniques for wine quality prediction based on various physicochemical properties in the wine samples. The work in this paper explores the performance of nine machine learning models. Logistic Regression, Decision Tree, K-Nearest Neighbors, and Support Vector Machine are used as classical models. Random Forest and Extra Trees are used as bagging-based ensemble models, while XGBoost, LightGBM, and CatBoost are used as boosting-based ensemble models. In this paper, we have also used some preprocessing techniques such as KNN imputation, power transformation, feature standardization and SMOTE (for handling class imbalance in dataset) to assists these machine learning models to improve their performance. The primary objectives of this research are:

1. To develop AI-based wine quality prediction models.
2. To compare the performance of nine machine learning algorithms.
3. To apply advanced preprocessing and feature engineering techniques.
4. To evaluate models using multiple performance metrics.
5. To analyze computational complexity and resource utilization.
6. To identify the most effective algorithm for wine quality prediction.

The performance of these machine learning models in the context of predicting the quality of wine is evaluated using several statistical and computational metrics: accuracy, precision, recall, F1-score, ROC-AUC, along with execution time (ET), CPU utilization during training (CUL), and memory consumption during training (MML).

The rest of the research comprises five sections: the related work in the domain of predicting the quality of wine using machine and deep learning algorithms are explicated in Section 2. The methodology adopted

in this paper to analyze the effectiveness of ensemble and classical machine learning techniques in wine quality prediction is presented in Section 3. Section 4 describes the experimental results and Section 5 concludes the research work.

2. Related Work

Wine quality prediction has received a lot of attention among researchers due to its importance in food chemistry, industrial quality control, and intelligent decision-making systems. Currently, wine quality assessment is based on expert sensory analysis, but this practice is highly subjective, expensive and time-consuming. To address these challenges, researchers have increasingly applied machine learning and ensemble learning approaches for the automation of quality assessment of wines in wine industry.

One of the oldest and most influential studies was performed by Cortez et al. [2]. They designed the Wine Quality dataset and applied data mining techniques to make predictions about wine quality based on physicochemical properties. Their study showed that machine learning algorithms are capable of modelling the problems for predicting wine quality. In this way, this study set the path for many subsequent studies in the same domain. Random Forest algorithm was proposed by Breiman with aim to improve the classification accuracy by learning many decision trees [3] and combining these results through bagging. Due to its robustness against overfitting and noisy data, Random Forest has become one of the most popular ensemble methods for wine quality prediction [14]. Extra Trees (Extremely Randomized Trees) is another ensemble machine learning algorithm that makes predictions by creating many independent decision trees [15]. It is very similar to Random Forest but adds more randomness in terms of node splitting and data selection, leading to faster training times while at the same time tending to prevent overfitting on noisy data. XGBoost was developed by Chen and Guestrin [10]. It is a highly scalable gradient boosting framework widely known for its strong predictive performance in machine learning competitions. This is a boosting ensemble algorithm that uses the mechanism of gradient boosting with various regularization techniques in order to improve both the classification accuracy and computational efficiency. LightGBM was also introduced as an extremely efficient boosting algorithm based on histogram learning and leaf-wise growth strategy [11]. This approach achieves better predictive performance with substantially lower training time; making it applicable to large datasets such as wine quality datasets. CatBoost was developed to handle the prediction bias and the categorical features issues of boosting algorithms [12]. This model utilizes ordered boosting approaches and has shown to perform effectively for classification tasks in many real-world applications, such as food quality prediction systems.

Dahal et al. [16] performed a comparative analysis of different machine learning algorithms (Ridge Regression (RR), Support Vector Machine (SVM), Gradient Boosting Regressor (GBR), and multi-layer Artificial Neural Network (ANN)) for wine quality prediction and reported that ensemble learning methods are better at both accuracy and classification than traditional machine learning models. Their study also highlighted the importance of preprocessing and feature selection techniques to achieve better predictive performance. Jain et al. [17] presented a machine learning-based predictive framework for wine quality analysis using physicochemical attributes. This study showed that Random Forest and XGBoost performed better than other traditional machine learning approaches in predictive performance. An ensemble learning-based wine quality prediction approach was presented by Zeng [18]. He demonstrated that the stacking ensemble models gave accuracy of around 87%, outperforming individual

machine learning algorithms. The study also highlighted the effectiveness of ensemble learning in improving prediction reliability and industrial applicability.

In recent studies, the researchers have also explored deep learning approaches for wine quality prediction. Di and Yang [19] proposed a one-dimensional Convolutional Neural Network (1D-CNN) model for wine quality prediction. This model captures correlations among physicochemical attributes and outperforms in terms of both accuracy and classification than traditional machine learning models. Their study showed that deep learning methods are getting more attentions of the researchers in order to design intelligent solutions for the prediction of wine quality. A recent study [20] highlighted the importance of feature engineering and machine learning algorithms in the context of wine quality prediction with imbalanced datasets. In this study, the authors showed that Support Vector Machine and ensemble methods perform well in the context of wine quality prediction with imbalanced datasets.

Although previous studies reported partially good prediction accuracies but most of them mainly concentrated on the statistical performance metrics, such as accuracy and ROC-AUC, while neglecting the computational efficiency metrics like CPU utilization, memory consumption, and execution time. Furthermore, very few studies have comprehensively compared both classical and ensemble learning models under unified preprocessing and feature engineering frameworks. Therefore, this study fills these research gaps by conducting a detailed comparison of nine machine learning algorithms based on predictive as well as computational performance metrics for wine quality prediction.

3. Methodology

The methodology used in this paper for analyzing the effectiveness of ensemble and classical machine learning techniques in wine quality prediction is presented with help of Figure 1. The complete methodology into five layers. The first layer is related to dataset collections as well as to analysis the dataset though studying its features. Thereafter, preprocessing layer includes statistical analytics techniques to improve dataset quality such as missing value handling, power transformation, feature scaling, feature engineering, and SMOTE balancing. In the model layer, four classical machine learning and five ensemble learning models are employed for predicting the quality of the wine. In evaluation layer, different statistical and computational performance metrics are used to shows the effectiveness of ensemble and classical machine learning techniques. The last layer (output layer) shows the effectiveness of ensemble and classical machine learning techniques in context of predicting the quality of wine through data visualizations and also provides some useful insights.

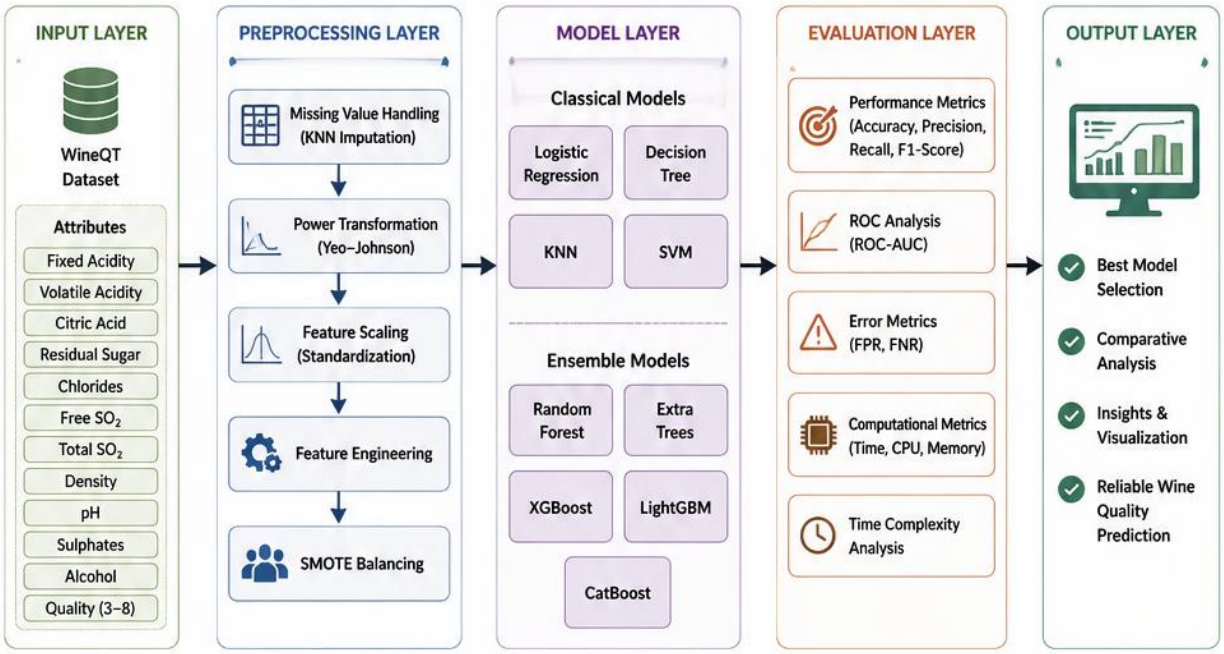

Figure 1: Proposed Methodology Layout for Wine Quality Prediction Using Machine Learning Models

3.1 WineQT Dataset Description

3.1.1 Dataset Source

The dataset used in this research is the WineQT dataset [21], which is available at Kaggle platform. The dataset consists of red wine samples described by their physicochemical properties together with quality scores assigned based on sensory analysis. The dataset is widely used in machine learning and data mining research for wine quality prediction and classification tasks.

3.1.2. Dataset Characteristics

The WineQT dataset consists of the 1,143 wine samples with 13 attributes including physicochemical properties and research-generated wine quality labels as shown in Table 1. Each sample represents a red wine instance described by several chemical features that influence wine quality.

Table 1. Overview of the WineQT Dataset Characteristics

Parameter	Description
Total Samples	1143
Total Features	13
Dataset Type	Structured Tabular Dataset
Prediction Type	Binary Classification
Data Category	Physicochemical Wine Data

3.1.3. Input Features Description

The dataset includes multiple physicochemical attributes that significantly influence wine quality prediction. The features included in the dataset are described in Table 2.

Table 2. Description of Features in the WineQT Dataset

Feature	Description
Fixed Acidity	Concentration of non-volatile acids present in wine
Volatile Acidity	Amount of acetic acid responsible for vinegar taste
Citric Acid	Citric acid concentration contributing freshness
Residual Sugar	Remaining sugar after fermentation
Chlorides	Salt concentration in wine
Free Sulfur Dioxide	Free SO_2 concentration preventing oxidation
Total Sulfur Dioxide	Total SO_2 concentration in wine
Density	Density of wine samples
pH	Acidity or alkalinity level
Sulphates	Sulfate concentration acting as antimicrobial agent
Alcohol	Alcohol percentage present in wine
Id	Unique identifier for each sample
Quality	Wine quality score

3.1.4. Target Variable Transformation

The original wine quality scores ranged from 3 to 8, where higher values represent better wine quality. To make prediction easier and boost classification performance, the multiclass quality labels were converted into binary classes as follows:

Quality $\geq 6 \rightarrow$ Good Wine (1)
Quality $< 6 \rightarrow$ Bad Wine (0)

This kind of binarization improves the performance of machine learning algorithms to perform efficient classification between high-quality and low-quality wine samples.

3.2 Dataset Challenges

The WineQT dataset presents several challenges for machine learning-based prediction systems, including:

i. **Class Imbalance:** Some quality classes contain fewer samples than others.
ii. **Nonlinear Relationships:** Physicochemical properties exhibit complex nonlinear interactions.
iii. **Feature Correlation:** Several chemical attributes are highly correlated.
iv. **Noise and Variability:** Wine quality evaluation may contain subjective variations.

To address these challenges, advanced preprocessing techniques such as KNN Imputation, Power Transformation, Feature Engineering, Standardization, and SMOTE balancing were applied in this study.

3.2 Data Preprocessing

Data preprocessing is an essential step in machine learning, since the raw datasets can contain missing values, inconsistent distribution, redundant information and imbalanced class labels. Proper preprocessing techniques on raw datasets improves the data quality, enhances model learning capability,

and increases prediction accuracy. Additionally, the WineQT dataset was preprocessed by applying of some advanced preprocessing techniques in this study to achieve a successful wine quality prediction. Preprocessing framework involves missing values handling, feature scaling, power transformation, feature engineering and class balancing. In this paper, we applied several advanced preprocessing techniques such as missing value handling, feature scaling, power transformation, feature engineering, and class balancing to prepare the WineQT dataset for effective wine quality prediction.

3.2.1 Handling Missing Values using KNN Imputation [22]

K-Nearest Neighbors (KNN) Imputation was used to handle missing values in the dataset. KNN Imputation performs model-based estimation where the nearest neighbors are identified through Euclidean distance and used to estimate the missing value.

The Euclidean distance between two samples is calculated as:

$$D(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Where:

- x_i and y_i represent feature values of two samples
- n denotes the total number of features

The missing value is estimated using the average values of the nearest neighboring samples.

3.2.2 Feature Standardization [23]

To normalize feature distributions and to improve algorithm convergence, we applied the StandardScaler technique for feature scaling. Standardization maps features to zero mean and unit variance distributions, in order to avoid forcing any attribute with a greater range of values to dominate the learning process.

The standardization formula is given as:

$$Z = \frac{X - \mu}{\sigma}$$

Where:

- X represents the original feature value
- μ denotes the mean of the feature
- σ represents the standard deviation

Standardization is particularly important for distance-based and optimization-based algorithms such as KNN and SVM.

3.2.3 Power Transformation [24]

A Power Transformation using Yeo–Johnson method is applied to reduce skewness and enhance normality of features. The transformation stabilizes variance and increases the performance of models with non-normally distributed features.

The Yeo–Johnson transformation is defined as:

$$Y(\lambda) = \begin{cases} \frac{(Y+1)^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ log(Y+1), & \lambda = 0 \end{cases}$$

Where:

- Y is the original feature value
- λ is the transformation parameter

This preprocessing step helps improve classification performance by reducing data skewness and enhancing feature distribution.

3.2.4 Feature Engineering

Feature engineering was performed to generate additional informative attributes from existing physicochemical features. These engineered features ensure that machine learning models can capture hidden relationships between different wine characteristics.

The generated features include:

1. **Alcohol-Sulphates Interaction**

$$Alcohol\_Sulphates = Alcohol \times Sulphates$$

2. **Acidity Ratio**

$$Acidity\_Ratio = \frac{FixedAcidity}{VolatileAcidity}$$

3. **Sulfur Balance**

$$Sulfur\_Balance = \frac{FreeSO_2}{TotalSO_2}$$

4. **Density-pH Interaction**

$$Density\_pH = Density \times pH$$

5. **Alcohol Density**

$$Alcohol\_Density = \frac{Alcohol}{Density}$$

These engineered attributes improve feature representation and contribute to enhanced predictive capability.

3.2.5 Class Balancing using SMOTE [13, 25]

The WineQT dataset is highly imbalanced — some wine qualities have far fewer samples than others. In order to solve this problem, SMOTE (Synthetic Minority Over−Sampling Technique) was used. Synthetic minority class samples are generated by SMOTE in order to balance the distribution between classes and thus improve classification performance.

The SMOTE equation is expressed as:

$$X_{new} = X_i + \delta(X_{zi} - X_i)$$

Where:

- X_i represents a minority class sample
- X_{zi} denotes one of its nearest neighbors
- δ is a random number between 0 and 1

SMOTE helps reduce model bias toward majority classes and improves recall and F1-score for minority samples.

3.3 Machine Learning Algorithm Selection

3.3.1 Logistic Regression [26]

Logistic Regression predicts probabilities using the sigmoid function.

Equation

$$P(Y = 1) = \frac{1}{1 + e^{-z}}$$

Where:

$$z = w_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n$$

3.3.2 Decision Tree [8,11]

Decision Tree recursively splits data based on feature conditions.

1) Entropy Equation

$$Entropy(S) = -\sum_{i=1}^{n} p_i log_2 p_i$$

2) Information Gain

$$IG(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

3.3.3 K-Nearest Neighbors (KNN) [27]

KNN classifies samples based on nearest neighbors.

Distance Equation

$$D(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

3.3.4 Support Vector Machine (SVM) [6]

SVM finds the optimal hyperplane maximizing margin.

Hyperplane Equation

$$w^T x + b = 0$$

Optimization Objective

$$\min \frac{1}{2} ||\ w\ ||^2$$

3.3.5 Random Forest [3,6]

Random Forest is an ensemble of decision trees.

Prediction Equation

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(x)$$

Where:

- $T_i(x)$ = Output of ith tree
- N = Number of trees

3.3.6 Extra Trees Classifier [28]

Extra Trees introduces random feature splits for improved generalization.

Gini Index

$Gini = 1 - \sum_{i=1}^{n} p_i^2$

3.3.7 XGBoost [10]

XGBoost uses gradient boosting optimization.

Objective Function

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y_i}) + \sum_{k=1}^{K} \Omega(f_k)$$

3.3.8 LightGBM [11]

LightGBM uses histogram-based learning and leaf-wise splitting.

Gain Formula

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}\right]$$

3.3.9 CatBoost [12, 29]

CatBoost handles categorical encoding efficiently using ordered boosting.

Gradient Update

$$\mathrm{F_m(x)} = \mathrm{F_{m-1}(x)} + \eta \mathrm{h_m(x)}$$

Where:

- $F_m(x)$ = Updated model
- η = Learning rate
- $h_m(x)$ = Weak learner

4. Experimental Setup and Results

4.1 Experimental Setup

In this paper, the experimental study is carried out using a system equipped with an Intel Core i5 processor operating at 1.30 GHz with 16 GB RAM and Windows 11 operating system. The implementation of all ensemble and classical machine learning techniques is done using Python 3.10 in Jupyter Notebook environment.

4.2 Evaluation Metrices

The brief description of evaluation metrices used in this paper for analyzing the effectiveness of ensemble and classical machine learning techniques in wine quality prediction is presented in Table 3. These evaluation metrices used in this paper are motivated from the metrices used in [30, 31].

Table 3. Performance Evaluation Metrics Used for Wine Quality Prediction

Metric	Mathematical Formula	Description
Accuracy	$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$	Measures the overall correctness of predictions
Precision	$\text{Precision} = \frac{TP}{TP + FP}$	Measures correctly predicted positive samples among predicted positives
Recall (Detection Rate)	$\text{Recall} = \frac{TP}{TP + FN}$	Measures the ability to correctly identify actual positive samples
F1-Score	$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Harmonic mean of precision and recall
False Positive Rate (FPR)	$\text{FPR} = \frac{FP}{FP + TN}$	Measures negative samples incorrectly classified as positive
False Negative Rate (FNR)	$\text{FNR} = \frac{FN}{FN + TP}$	Measures positive samples incorrectly classified as negative
ROC-AUC Score	$\text{ROC-AUC} = \int_0^1 TPR(FPR)\, d(FPR)$	Evaluates the classification capability of the model
Execution Time	$\text{Execution Time} = T_{training} + T_{prediction}$	Measures total training and prediction time
CPU Usage	$\text{CPU Usage} = \frac{CPU_{used}}{CPU_{total}} \times 100$	Represents processor utilization during execution
Memory Usage	$\text{Memory Usage} = \text{Memory}_{\text{allocated}} - \text{Memory}_{\text{free}}$	Indicates memory consumption during model execution

Where:

- **TP** = True Positive
- **TN** = True Negative
- **FP** = False Positive
- **FN** = False Negative
- **TPR** = True Positive Rate
- $T_{training}$ = Training time
- $T_{prediction}$ = Prediction time

4.3 Experimental Results

Figure 2 represents the comparative statistical performance analysis of machine learning models for wine quality prediction. Figures 3, 4, and 5 present the comparative computational analysis of execution time, memory usage, and CPU utilization of the machine learning models used for wine quality prediction, respectively.

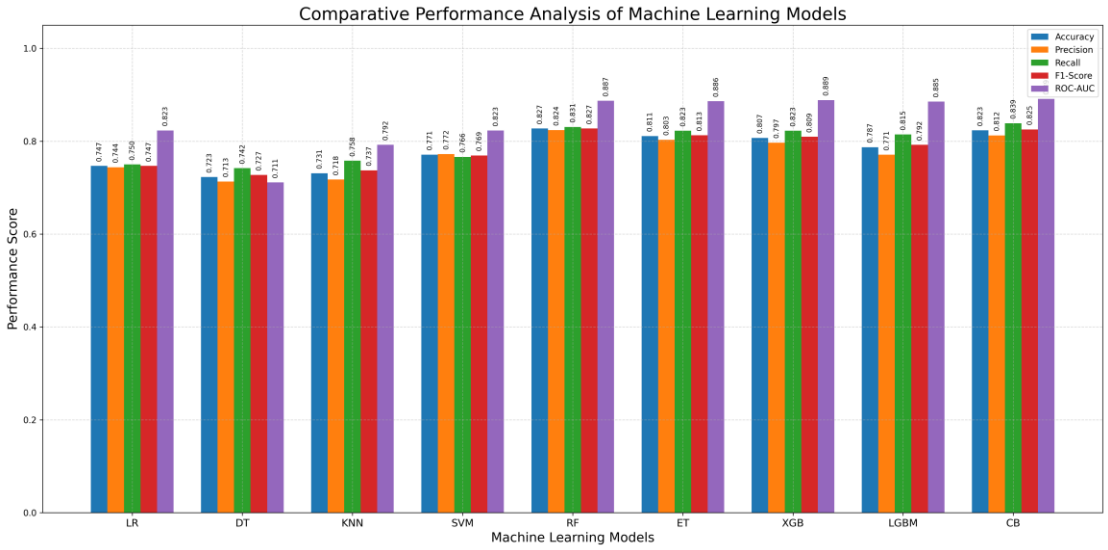

Figure 2. Comparative Statistical Performance Analysis of Machine Learning Models for Wine Quality Prediction

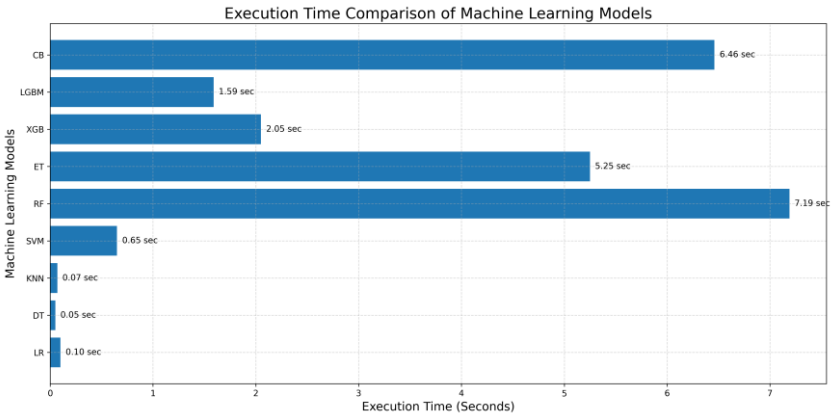

Figure 3. Comparative Computational Performance Analysis related to execution time of Machine Learning Models for Wine Quality Prediction

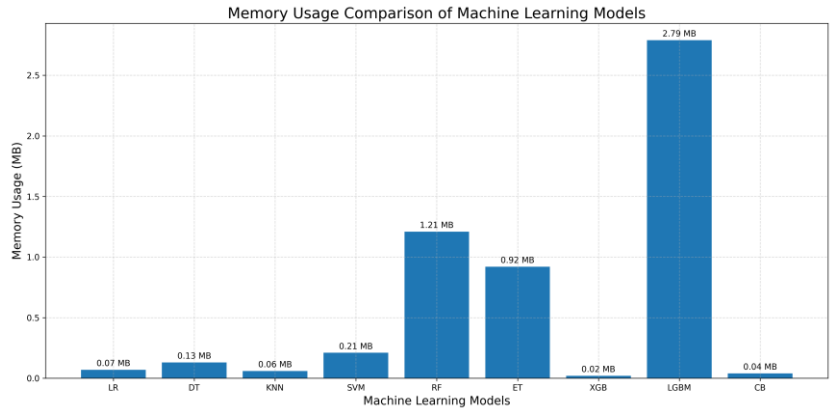

Figure 4. Comparative Computational Performance Analysis related to memory usages of Machine Learning Models for Wine Quality Prediction

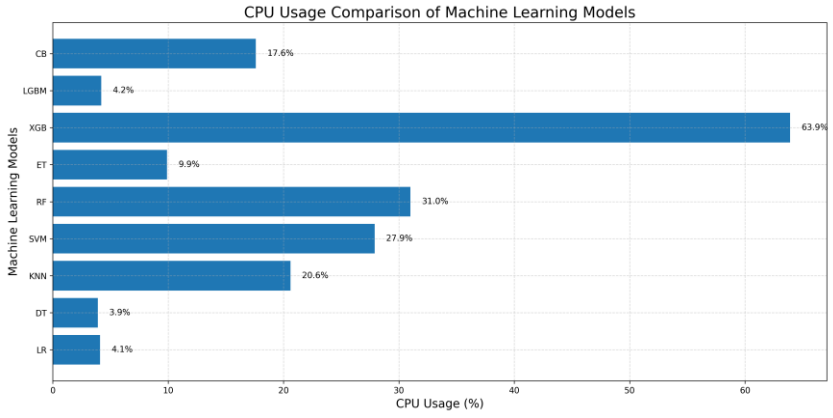

Figure 5. Comparative Computational Performance Analysis related to CPU usages of Machine Learning Models for Wine Quality Prediction

5. Conclusion

In this study, five ensemble and four classical machine learning algorithms are used for predicting the quality of red wine. The experimental results showed that Random Forest yielded the highest prediction accuracy of 82.73%, whereas CatBoost yielded the highest ROC-AUC score at 0.8911 indicating a high classification capability. The other ensemble methods (XGBoost, LightGBM, and Extra Trees) also outperformed classical machine learning algorithms likely due to their ability to model nonlinear relationships and reduce overfitting as well. The future work will focus on exploring the effectiveness of other machine and deep learning algorithms in context of red wine quality prediction. Moreover, future work will also focus on considering the multiple large-scale and complex wine datasets containing a greater number of features and diverse data characteristics for the study.

References

1. Shahrajabian, M. H., & Sun, W. (2024). Assessment of wine quality, traceability and detection of grapes wine, detection of harmful substances in alcohol and liquor composition analysis. Letters in Drug Design & Discovery, 21(8), 1377-1399.
2. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. Decision support systems, 47(4), 547-553.
3. Breiman, L. Random Forests. Machine Learning, 45(1), 5-32.
4. Kumar, A., Gupta, R., Kumar, S., Dutta, K., & Kumar, R. (2025). Intelligent Intrusion Detection System Using Improved Osprey Optimization and Stacked Ensemble Learning for IoT-Based Healthcare Systems. Security and Privacy, 8(6), e70121.
5. Vermani, K., Noliya, A., Kumar, S., & Dutta, K. (2023). Ensemble Learning Based Malicious Node Detection in SDN-Based VANETs. Journal of Information Systems Engineering & Business Intelligence, 9(2).
6. Adhikary, K., Bhushan, S., Kumar, S., & Dutta, K. (2022). Evaluating the performance of various SVM kernel functions based on basic features extracted from KDDCUP'99 dataset by random forest method for detecting DDoS attacks. Wireless Personal Communications, 123(4), 3127-3145.
7. Adhikary, K., Bhushan, S., Kumar, S., & Dutta, K. (2019). Evaluating the performance of various machine learning algorithms for detecting DDoS attacks in VANETs. International Journal of Control Automation, 12(5), 478-486.
8. Adhikary, K., Bhushan, S., Kumar, S. & Dutta, K. (2020). Decision tree and neural network based hybrid algorithm for detecting and preventing DDoS attacks in VANETS. Int. J. Innov. Technol. Explor. Eng, 9, 669-675.
9. Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.
10. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
11. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

12. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems, 31.
13. Elreedy, D., & Atiya, A. F. (2019). A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. Information sciences, 505, 32-64.
14. Khan, R., Goyal, A., Kanyal, H. S., Parashar, D., Sharma, S. K., & Iqbal, M. (2026). Improved machine learning framework with feature engineering and SHAP analysis for predicting wine quality. Discover Applied Sciences, 8(1), 27.
15. Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine learning, 63(1), 3-42.
16. Dahal, K. R., Dahal, J. N., Banjade, H., & Gaire, S. (2021). Prediction of wine quality using machine learning algorithms. Open Journal of Statistics, 11(2), 278-289.
17. Jain, K., Kaushik, K., Gupta, S. K., Mahajan, S., & Kadry, S. (2023). Machine learning-based predictive modelling for the enhancement of wine quality. Scientific Reports, 13(1), 17042.
18. Zeng, Q. (2022, December). Prediction of Wine Quality Using Ensemble Learning Approach of Machine Learning. In 2022 International Conference on mathematical statistics and economic analysis (MSEA 2022) (pp. 770-774). Atlantis Press.
19. Di, S., & Yang, Y. (2022). Prediction of red wine quality using one-dimensional convolutional neural networks. arXiv preprint arXiv:2208.14008.
20. Zaza, S., Atemkeng, M., & Hamlomo, S. (2023, October). Wine feature importance and quality prediction: A comparative study of machine learning algorithms with unbalanced data. In International Conference on Safe, Secure, Ethical, Responsible Technologies and Emerging Applications (pp. 308-327). Cham: Springer Nature Switzerland.
21. Yasser, H. (2021). Wine quality dataset. Kaggle. https://www.kaggle.com/datasets/yasserh/wine-quality-dataset
22. Fritz, M. (2023). Decision tree classification with missing values (Doctoral dissertation, Technische Universität Wien).
23. Ozsahin, D. U., Mustapha, M. T., Mubarak, A. S., Ameen, Z. S., & Uzun, B. (2022, August). Impact of feature scaling on machine learning models for the diagnosis of diabetes. In 2022 International Conference on Artificial Intelligence in Everything (AIE) (pp. 87-94). IEEE.
24. Zwanenburg, A., & Löck, S. (2026). Location and Scale-Invariant Power Transformations for Transforming Data to Normality. Machine Learning, 115(3), 34.
25. Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. BMC bioinformatics, 14(1), 106.
26. LaValley, M. P. (2008). Logistic regression. Circulation, 117(18), 2395-2399.
27. Kramer, O. (2013). K-nearest neighbors. In Dimensionality reduction with unsupervised nearest neighbors (pp. 13-23). Berlin, Heidelberg: Springer Berlin Heidelberg.
28. Zegaar, A., Ounoki, S., & Telli, A. (2024). Machine learning for groundwater quality classification: A step towards economic and sustainable groundwater quality assessment process. Water Resources Management, 38(2), 621-637.
29. Kahraman, A. (2025). Machine learning techniques for improved prediction of cardiovascular diseases using integrated healthcare data. Frontiers in Artificial Intelligence, 8, 1694450.

30. Vakili, M., Ghamsari, M., & Rezaei, M. (2020). Performance analysis and comparison of machine and deep learning algorithms for IoT data classification. arXiv preprint arXiv:2001.09636.
31. Mittal, S., Rajput, P., & Subramoney, S. (2021). A survey of deep learning on cpus: opportunities and co-optimizations. IEEE Transactions on Neural Networks and Learning Systems, 33(10), 5095-5115.