Performance Evaluation of Machine Learning and Neural Network Algorithms for Wine Quality Prediction

Harika Kakarala
Department of CSE,
SRM University-AP, India
harika_kakarala@srmap.edu.in

Asish Karthikeya Gogineni
Department of CSE,
SRM University-AP, India
asishkarthikeya_gogineni@srmap.edu.in

Thadi Venkata Satya Murty
Department of CSE,
SRM University-AP, India
satyamurty_thadi@srmap.edu.in

Srilatha Tokala
ACT Lab, Department of CSE,
SRM University-AP, India
srilatha_tokala@srmap.edu.in

Murali Krishna Enduri
ACT Lab, Department of CSE,
SRM University-AP, India
muralikrishna.e@srmap.edu.in

Satish Anamalamudi
Networking Lab, Department of CSE,
SRM University-AP, India
satish.a@srmap.edu.in

***Abstract*—The assessment of wine quality is of paramount importance to both consumers and the wine industry. Recognizing its impact on customer satisfaction and business success, companies are increasingly turning to product quality certification to enhance sales in the global beverage market. Traditionally, quality testing was conducted towards the end of the manufacturing process, resulting in time-consuming and resource-intensive procedures. This approach involved the engagement of multiple human experts to evaluate wine quality, leading to high costs. Moreover, since taste perception is subjective and varies among individuals, relying solely on human specialists for assessing wine quality presents significant challenges. Our research focuses on advancing the quality of wine prediction by leveraging diverse characteristics of wine. We applied various feature selection techniques and explored machine learning algorithms to identify the optimal combination of parameters for accurate wine quality prediction. This approach reduces the time and costs associated with traditional quality assessment methods and provides a more standardized and consistent evaluation process. Our findings contribute to the advancement of wine industry practices, enabling businesses to make informed decisions and deliver high-quality products that meet consumer expectations.**

***Index Terms*—feature selection, machine learning, wine quality, quality assessment**

I. Introduction

Over the past few years, machine learning and neural networks revolutionized various industries, and the wine industry is no exception. Wines, with their diverse flavors and historical significance, have traditionally been evaluated subjectively in terms of quality. Nevertheless, the advent of sophisticated analytical techniques and the availability of extensive data have presented fresh opportunities for objectively assessing and predicting wine quality.

Wine is generally categorized or differentiated based on color, smell, and taste. However, it is feasible to ascertain whether a wine possesses high or low quality. There are various features that were involved in the dataset that we downloaded. Because the qualities are chemical, let's go through them now for a better grasp of the process. The fixed acidity property of wine is attributed to its primary components, namely tartaric, succinic, citric, and malic acids. The Volatile Acidity feature characterizes the presence of gaseous acids in the wine. The residual sugar quantity remains after its fermentation. Citric acid, a weak organic acid with limited occurrence, is primarily found in citrus fruits. The chloride content quantifies the salt levels in wine, while free sulfur dioxide is employed to prevent oxidation and microbial degradation [1]. Sulfates help in preserving the freshness of wine by safeguarding it against germs and oxidation. [2].

The acidity of a wine is measured using the pH scale. Using the type feature we can predict the color of the wine. The range of the target variable quality is from 1 to 10. This research holds great significance in its ability to improve the wine production process. Historically, winemakers have heavily relied on their expertise and sensory evaluation to evaluate wine quality [3]. Although this approach is valuable, it is constrained by subjectivity and human limitations in processing extensive data. By harnessing the power of machine learning and neural networks, we can construct models that concurrently consider multiple factors, enabling objective and consistent quality predictions. Furthermore, the implementation of wine quality prediction models grounded in machine learning and neural networks can have far-reaching implications for the entire wine industry. Producers stand to gain considerable advantages from these predictive models, as they can identify pivotal factors that impact wine quality, optimize production processes, and ensure consistent quality across different batches. Similarly, consumers can reap the benefits of enhanced precision in wine quality predictions, enabling them to make well-informed decisions when selecting wines that align with their preferences.

To fulfill these goals, this paper extensively explore the

foundational principles of machine learning and neural networks, elucidating their application in predicting wine quality [4]. Various algorithms and methodologies commonly employed in this field will be examined, delving into their strengths and limitations. Furthermore, relevant datasets will be analyzed, followed by data pre-processing and the training of predictive models [5]. The performance of these models will be rigorously evaluated and compared, offering valuable insights into their effectiveness and potential practical applications.

A. Applications

- **Quality Control**: The prediction of wine quality enables the evaluation of its excellence prior to bottling and distribution to consumers. This can help winemakers identify potential issues with the wine and make adjustments to improve its quality [6].
- **Pricing**: The utilization of wine quality prediction enables the estimation of a wine's market value by forecasting its quality. This can help winemakers set prices that accurately reflect the quality of their products, and maximize their profits [7].
- **Vintage Prediction**: The prediction of wine is predicted by a specific vintage, considering weather patterns and other environmental factors. This can help winemakers plan their production processes and make adjustments as needed to optimize the quality of their wine [8].

II. Related Recent Work

Numerous studies have concentrated on the utilization of conventional machine learning algorithms like decision trees, random forests, and support vector machines to predict wine quality [2]. Ioannis Athanasiadis et al. (2021) employed a decision tree-based methodology to classify wines into distinct quality categories by considering chemical composition and sensory attributes. Their findings showcased encouraging accuracy in predicting wine quality [9]. In addition to the traditional machine learning techniques, researchers have increasingly embraced neural networks for wine quality prediction. Deep learning models, including CNN and RNN, have demonstrated their efficiency in capturing intricate relationships and patterns present in wine datasets. For instance, Shengnan Di et al. (2022) employed a CNN architecture to analyze images of wine labels and accurately predict their respective quality ratings [10].

Moreover, researchers have investigated feature selection and dimensionality reduction for wine quality prediction. Marianthi Basalekou et al. (2023) employed PCA to reduce the dimensionality of input features, leading to enhanced prediction accuracy and computational efficiency. Their study highlighted the benefits of dimensionality reduction in wine quality prediction [11]. To characterize wines based on their geographical origin, Principal Component Analysis (PCA) was conducted and reported. The study analyzed 33 Greek wines, considering their physical and chemical attributes. To obtain a better knowledge of the purpose, algorithm types used, and various tactics for predicting wine quality, numerous researchers have produced relevant studies in international and national research articles, as well as chapters.

It is important to mention that certain studies have embraced ensemble learning approaches [9], which involve combining multiple models or predictions to attain improved accuracy and robustness. Ensemble methods, such as bagging and boosting, have demonstrated the potential in enhancing the overall performance of wine quality prediction models. Data mining techniques are employed to predict wine quality. To make wine quality predictions, machine learning techniques like a variety of applications use the information to create models. The UCI store was supplied with an informational index in 1991 with 178 entries with estimated amounts of 13 distinctive chemical compounds, including alcohol and magnesium. This was done to purchase three cultivars from Italy. Due to its convenient separability, this dataset has been extensively utilized as a benchmark for novel information mining classifiers[12].

III. Methodology

This section offers a detailed overview of the approach employed for wine quality prediction using machine learning and neural networks.

A. Random Forest Classifier model

It is an extensively recognized model that merges the concepts of ensemble learning and decision trees. It finds extensive application in classification tasks, by predicting the quality of the wine [13]. The fundamental concept behind this model is to incorporate randomness in both the data and feature selection procedures.

It has the ability to handle complex datasets, resistance to overfitting, feature importance estimation, anomalies or outliers detection, and robustness to noise making it an attractive choice for many applications [14]. When working with large datasets, training a Random Forest with a substantial number of decision trees can incur significant computational costs. Using Eq. 1 we calculated the random forest classifier for the model.

$$RFfi_i = \frac{\Sigma_j norm f_{ij}}{\Sigma_{j \in allfeatures, k \in alltrees} normfi_{jk}} \tag{1}$$

B. Logistic Regression model

A statistical model called logistic regression was created especially for binary classification tasks. A particular kind of regression analysis is used to calculate the likelihood of a binary outcome by taking into account one or more independent variables as shown in Eq. 2. This model is particularly well-suited for scenarios where the dependent variable is binary or dichotomous, indicating 0 or 1. The primary objective of the model is to assess the likelihood of a given instance belonging to a specific class [15]. This model enables researchers and practitioners with diverse statistical backgrounds to easily grasp and utilize it. Dealing with missing data can pose

challenges that require the application of suitable imputation techniques.

$$P = \frac{e^{a+bx}}{1 + e^{a+bx}} \tag{2}$$

C. K-Nearest Neighbors Classifier (KNN)

KNN is also called a Lazy learner algorithm, which is non-parametric in nature. It employs proximity as the basis for making classifications or predictions regarding the grouping of a given data point [16]. The algorithm stores all available instances and determines the classification of new data points by measuring their similarity to existing data points. This estimation assesses the probability of a particular group from its nearest neighbors is in Eq. 3 we can calculate the distances between the points. It functions based on the principle of similarity, determining the prediction for a new instance by considering the majority vote or averaging its k nearest neighbors.

$$d(x, y) = \sqrt{\Sigma_{i=1}^{n}(x_i - y_i)^2} \tag{3}$$

D. Support Vector Classifier (SVC)

SVC is primarily utilized for classification tasks of supervised learning algorithms. It falls under the umbrella of machine learning algorithms and is built upon the principle of support vectors.

The SVC algorithm operates by establishing a hyperplane or a group of hyperplanes within a high-dimensional space. These hyperplanes serve the purpose of delineating distinct classes of data points [17]. The primary goal of SVC is to identify the hyperplanes that maximize the margin between the classes, thereby enabling efficient handling of datasets [18]. Using Eq. 4 we calculate the decision function from the model.

One of the notable strengths of SVC lies in its ability to effectively handle datasets characterized by intricate decision boundaries. It is very sensitive in hyperparameter selection, particularly the regularization parameter and the choice of the kernel function. Optimal hyperparameters require careful tuning, as improper settings can lead to issues such as overfitting or underfitting of the model.

$$f(x) = w^T + b \tag{4}$$

E. Gaussian Naive Bayes Model

It is a popular model rooted from the Bayes theorem. By assuming conditional independence among the features given the class label, the Naive Bayes model simplifies computation and exhibits computational efficiency [19]. When predicting the class of an instance, the model calculates the probability based on the observed features. It utilizes Baye's theorem shown in Eq. 5 to compute the conditional probability of the class given the features. It uses a small training dataset to judge the parameters, making it efficient in terms of computational resources. In cases where features are highly correlated, this assumption can result in suboptimal performance. Therefore, caution should be exercised when dealing with such situations to ensure the model's effectiveness.

$$P(b|a) = \frac{P(a|b) * P(b)}{P(a)} \tag{5}$$

F. eXtreme Gradient Boosting (XGBoost)

A well-liked and effective model is XGBoost, extensively utilized for both classification and regression tasks. It is built on a gradient boosting architecture and is a member of the ensemble learning family [20]. XGBoost integrates multiple weak predictive models, commonly decision trees, in order to generate a strong and precise final prediction using Eq. 6. It is also used for ranking, anomaly detection, and user-defined prediction problems.

$$f(x) \approx f(a) + f^{'}(a)(x - a) + 0.5(f^{''}(a)(x - a)^2) \tag{6}$$

G. Multi Layer Perceptron (MLP) Classifier

An artificial neural network created specifically for classification tasks is used by the MLP classifier. It operates as a feedforward neural network, comprising multiple layers of interconnected nodes known as neurons.

In the MLP classifier, a dense network structure is formed by each neuron that is linked to the subsequent layer neuron. The neurons within the hidden layers employ nonlinear activation functions to transform the weighted sum of inputs received from the preceding layer. This enables the network to effectively learn intricate patterns and relationships present in the data [6]. Throughout the training process, the MLP classifier adjusts the weights and biases associated with its neurons using a technique called backpropagation [21]. This involves propagating the error between predicted and actual outputs backward through the network.

IV. Dataset Statistics

This study employs three distinct Wine datasets to conduct comprehensive experiments as shown in Table I.

TABLE I
DATASETS STATISTICS

Datasets	Total Observations	Number of 0s	Number of 1s
White & Red Wine	6497	5220	1277
Red Wine	1599	1382	217
White Wine	4898	3838	1060

These datasets serve as a valuable resource for the experimental analysis and evaluation of wine quality. The utilization of multiple datasets allows for a robust and thorough investigation of wine quality prediction algorithms and their performance across different types of wines.

A. Steps to perform

Import all the necessary packages of Python like Pandas, Numpy, seaborn, matplotlib, and SKlearn for analyzing the dataset Later on, we need to perform data-preprocessing steps

- **Feature Selection** It is the process used for reducing the input variables utilized in your model by exclusively considering pertinent data and eliminating irrelevant noise within the data [22]. There are 2 types of feature selection models namely supervised models and unsupervised models since we used classification algorithms

we perform only supervised feature selection models [23]. In this research, we used Variance Inflation Factor (VIF) feature selection technique, which evaluates the multicollinearity among predictor variables in the dataset and drops features based on high VIF values. If the VIF value is greater than 10 then we will drop the attribute from the dataset accordingly [24], [25].

- **Data Standardization** It is a standard scaler class from the sklearn.preprocessing module in sci-kit-learn employed to standardize or normalize numerical features within a dataset
  Overall the utilization of a standard scaler ensures that the features are suitably scaled, resulting in enhanced model performance, more precise variable comparisons, and improved interpretability of the outcomes [26].
- **Lable Binarization** It is also known as binary transformation used to convert the target variables into binary form, it is performed only on the target variable, not on the feature variables. In this research, we ensured that if the target variable i.e., if the quality is $\geq$7 then we labeled it as 1, if it is $\leq$6 then it is labeled as 0.

B. Performance Measures

In this research, we four different performance measures namely accuracy, precision, recall, and f1-score are used. For defining the equations used, we have taken four different variables namely True Positives represented as A, True Negatives represented by B, False Positives represented as C, and finally, False Negatives represented by D.

Accuracy (A): The most generally used performance metric is accuracy. It focuses on the correct classifications and does not directly take into account the misclassifications represented by false positives and false negatives. Using Eq. 7, we calculate the accuracy measure.

$$Accuracy = \frac{A+B}{A+C+B+D} \tag{7}$$

Precision (P): Precision is a performance measure commonly employed in machine learning and statistics, specifically in binary classification tasks. The calculation method for precision is shown in Eq. 8.

$$Precision = \frac{A}{A+C} \tag{8}$$

Precision evaluates the precision of positive predictions, reflecting the model's performance when it identifies instances as positive. It measures the model's capability to minimize false positives.

Recall (R): Recall serves as an essential performance measure frequently employed in machine learning and binary classification tasks. It assesses the model's capability to accurately detect the positive instances from the overall instances in the dataset as indicated in Eq. 9.

$$Recall = \frac{A}{A+D} \tag{9}$$

Recall emphasizes the comprehensiveness of positive predictions, illustrating the model's effectiveness in capturing all positive instances within the dataset. It quantifies the model's capacity to minimize false negatives.

F1-Score (F): The F1-score, frequently employed in machine learning and binary classification tasks, serves as a popular performance measure as shown in Eq. 10.

$$F1\text{-}Score = 2 * \frac{P * R}{P + R} \tag{10}$$

A high F1 score shows that the model is proficient at both accurately forecasting positive instances and capturing all positive examples within the dataset. It also shows that the model has successfully balanced precision and recall.

V. Experimental Results

In this section, we initiate the pre-processing stage of our analysis, which involves the utilization of a feature selection technique as mentioned previously. The objective of this technique is to eliminate unnecessary data from the dataset, enhancing the efficiency and effectiveness of our subsequent analysis. Specifically, we employed the Variance Inflation Factor (VIF) to assess the attributes' collinearity and identify variables that exhibit high multicollinearity.

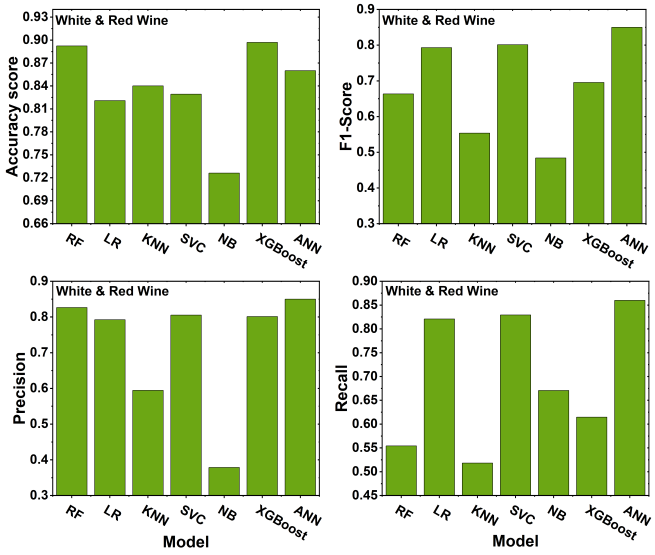

Fig. 1. Bar plots for Accuracy Score, F1-Score, Precision, and Recall for White and Red Wine dataset.

In Fig. 1, several machine learning classifiers were evaluated based on their performance metrics. ANN model performs well compared to all the remaining models for all three performance measures. Whereas in accuracy measure, the XGBoost model outperforms well. Naive Bayes gives lowest performance value in terms of accuracy, precision, and f1-scores. In recall measure, the KNN model gives the lowest value.

These findings indicate that the random forest classifier and XGBoost exhibit superior accuracy compared to the multilayer

perceptron model. Moreover, logistic regression, SVC, and the multilayer perceptron model show promising f1-scores, with the multilayer perceptron model achieving the highest score. The recall scores of logistic regression and SVC were found to be on par with those of the multilayer perceptron model. These observations highlight the varying performance of the classifiers and provide insights into their efficiency for wine quality prediction tasks using the white and red wine dataset.

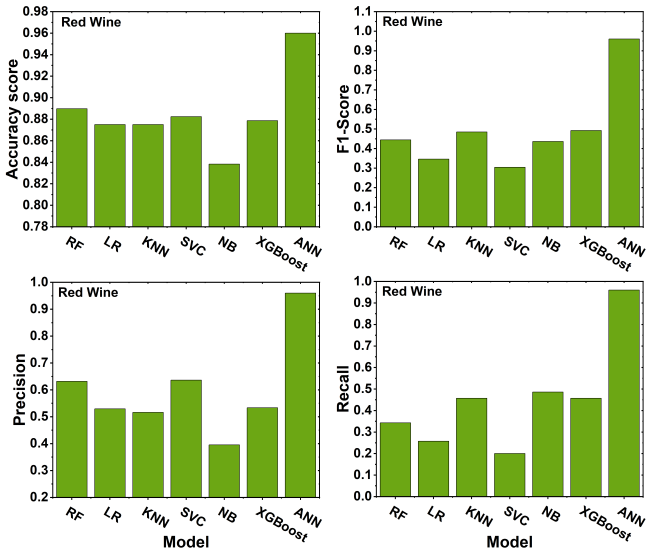

Fig. 2. Bar plots for Accuracy Score, F1-Score, Precision, and Recall for Red Wine dataset.

In Fig. 2, we observed that multilayer perceptron (ANN model) outperformed other models across various performance measures. The random forest classifier and XGBoost demonstrated competitive accuracy rates, while logistic regression, KNN, and gaussianNB exhibited relatively lower performance. The multilayer perceptron model's high precision and recall rates further validate its effectiveness in accurately predicting wine quality using the red wine dataset.

In Fig. 3, the performance is compared with the white wine dataset. We observe that the multilayer perceptron (ANN) model exhibits superior performance in predicting the quality of the wine. The lowest performance is given by Naive Bayes in terms of accuracy score and precision. In terms of recall and f1-score, the SVC model gives the lowest performance.

VI. Conclusion and Future Work

This research paper presents an investigation into a diverse set of established algorithms for wine quality prediction and conducts a comprehensive comparative analysis to evaluate their performance. While this study provides valuable insights into wine quality prediction using various ML algorithms, for further exploration and improvement. To improve the prediction model accuracy and generalizability, one should first consider expanding the datasets to include regions, wine varieties, and production techniques. The inclusion of additional

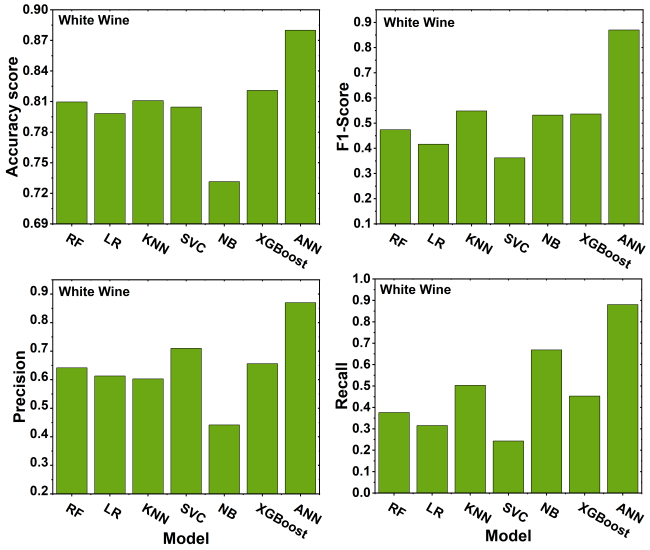

Fig. 3. Bar plots for Accuracy Score, F1-Score, Precision, and Recall for White Wine dataset.

features such as grape variety, region of origin, or weather conditions during the growing season could provide valuable information for improving the predictive models. Furthermore, the development of models that offer interpretability and shed light on the underlying factors influencing wine quality would be highly advantageous for both winemakers and consumers. Lastly, exploring advanced techniques such as ensemble learning can further enhance the predictive performance of the algorithms.

References

[1] P. Giménez-Gómez, M. Gutiérrez-Capitán, J. M. Rıos, F. Capdevila, A. Puig-Pujol, and C. Jiménez-Jorquera, "Microanalytical flow system for the simultaneous determination of acetic acid and free sulfur dioxide in wines," *Food Chemistry*, vol. 346, p. 128 891, 2021.

[2] Y. Gupta, "Selection of important features and predicting wine quality using machine learning techniques," *Procedia Computer Science*, vol. 125, pp. 305–312, 2018.

[3] G. D. Nelson, "Red and white wine data analysis-predict quality of wine," 2020.

[4] S. Aich, A. A. Al-Absi, K. L. Hui, J. T. Lee, and M. Sain, "A classification approach with different feature sets to predict the quality of different types of wine using machine learning techniques," in *2018 20th International conference on advanced communication technology (ICACT)*, IEEE, 2018, pp. 139–143.

[5] P. Dhaliwal, S. Sharma, and L. Chauhan, "Detailed study of wine dataset and its optimization," *Int. J. Intell. Syst. Appl.(IJISA)*, vol. 14, no. 5, pp. 35–46, 2022.

[6] K. Dahal, J. Dahal, H. Banjade, and S. Gaire, "Prediction of wine quality using machine learning algorithms," *Open Journal of Statistics*, vol. 11, no. 2, pp. 278–289, 2021.

[7] L. S. Lockshin and W. Timothy Rhodus, "The effect of price and oak flavor on perceived wine quality," *International journal of wine marketing*, vol. 5, no. 2/3, pp. 13–25, 1993.

[8] I. Horowitz and L. Lockshin, "What price quality? an investigation into the prediction of wine-quality ratings," *Journal of Wine Research*, vol. 13, no. 1, pp. 7–22, 2002.

[9] I. Athanasiadis and D. Ioannides, "A machine learning approach using random forest and lasso to predict wine quality," *International Journal of Sustainable Agricultural Management and Informatics*, vol. 7, no. 3, pp. 232–251, 2021.

[10] S. Di, "Prediction of red wine quality using one-dimensional convolutional neural networks," *arXiv preprint arXiv:2208.14008*, 2022.

[11] M. Basalekou, P. Tataridis, K. Georgakis, and C. Tsintonis, "Measuring wine quality and typicity," *Beverages*, vol. 9, no. 2, p. 41, 2023.

[12] J. Liang, X. Zhao, D. Li, F. Cao, and C. Dang, "Determining the number of clusters using information entropy for mixed data," *Pattern Recognition*, vol. 45, no. 6, pp. 2251–2265, 2012.

[13] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, J. Akinjobi, *et al.*, "Supervised machine learning algorithms: Classification and comparison," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 48, no. 3, pp. 128–138, 2017.

[14] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert systems with applications*, vol. 134, pp. 93–101, 2019.

[15] T. G. Nick and K. M. Campbell, "Logistic regression," *Topics in biostatistics*, pp. 273–301, 2007.

[16] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.

[17] S. Aich, A. A. Al-Absi, K. L. Hui, and M. Sain, "Prediction of quality for different type of wine based on different feature sets using supervised machine learning techniques," in *2019 21st International Conference on Advanced Communication Technology (ICACT)*, IEEE, 2019, pp. 1122–1127.

[18] C. Gold and P. Sollich, "Model selection for support vector machine classification," *Neurocomputing*, vol. 55, no. 1-2, pp. 221–249, 2003.

[19] A. H. Jahromi and M. Taheri, "A non-parametric mixture of gaussian naive bayes classifiers based on local independent features," in *2017 Artificial intelligence and signal processing conference (AISP)*, IEEE, 2017, pp. 209–212.

[20] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[21] T. Windeatt, "Ensemble mlp classifier design," *Computational Intelligence Paradigms: Innovative Applications*, pp. 133–147, 2008.

[22] P. Bhardwaj, P. Tiwari, K. Olejar Jr, W. Parr, and D. Kulasiri, "A machine learning application in wine quality prediction," *Machine Learning with Applications*, vol. 8, p. 100 261, 2022.

[23] V. Kumar and S. Minz, "Feature selection: A literature review," *SmartCR*, vol. 4, no. 3, pp. 211–229, 2014.

[24] T. A. Craney and J. G. Surles, "Model-dependent variance inflation factor cutoff values," *Quality engineering*, vol. 14, no. 3, pp. 391–403, 2002.

[25] R. D. Kothawade, *Wine quality prediction model using machine learning techniques*, 2021.

[26] M. S. Gal and D. L. Rubinfeld, "Data standardization," *NYUL Rev.*, vol. 94, p. 737, 2019.