Machine Learning for Predicting Wine Quality and its Key Determinants Based on Physicochemical Properties

Domonkos Frigyes Török
Hungarian University of Agriculture
Orcid: 0009-0005-7599-063X

Cite this research:
Török, D.F.,(2023). Machine Learning for Predicting Wine Quality and its Key Determinants Based on Physicochemical Properties SSRAML, 6(11), 1–21.

Article history:
Received: September/12/2023
Accepted: November/04/2023

Abstract

Traditional methods of assessing wine, which lean heavily on human experts for sensory evaluation, are not only time-consuming and expensive but also fraught with inconsistencies due to the inherent subjectivity of taste tests. These expert opinions can vary widely, as taste perception differs greatly among individuals, leading to a lack of standardization in the evaluation process. In this context, there's a growing need for a more objective, efficient, and cost-effective method to assess wine quality. Machine learning offers a promising solution by automating the quality assessment process, leveraging data from physicochemical tests. The aim of this study is to develop and validate a machine learning model for accurately predicting wine quality, based on its physicochemical properties. It seeks to identify key factors influencing wine quality and compare the effectiveness of different machine learning algorithms. Eight different models were applied on the wine quality dataset, and SHAP (SHapley Additive exPlanations) was used to identify the key factors of wine quality. All of the algorithms showed high accuracy scores. The analysis of wine quality using machine learning models highlighted alcohol content as the most influential factor for higher quality, while volatile acidity was negatively correlated with quality. SHAP values identified sulphates, alcohol, and volatile acidity as key determinants of wine quality. After hyperparameter optimization, the best model trial achieved 91.15% accuracy with fine-tuned parameters, although precision and F1-Scores for certain classes slightly decreased. The overall accuracy of the model also experienced a marginal reduction. Despite these changes, SHAP values continued to indicate the importance of sulphates, alcohol, and volatile acidity in the model's decision-making, highlighting their significant role in defining wine quality. This approach could be beneficial for wine distributors, and consumers by offering an objective and scalable method for evaluating wine quality. Additionally, understanding how various chemical components influence quality perceptions could be useful for wine producers aiming to improve their products.

Keywords: Machine learning algorithms, Physicochemical Properties, Quality Assessment, Sensory Evaluation, Wine Quality

Introduction

The wine industry, central to the global beverage market, places immense emphasis on wine quality, a factor that significantly impacts consumer preference, brand reputation, and market dynamics [1], [2]. In recent years, there has been a noticeable rise in wine consumption, a trend driven by evolving consumer tastes, increased awareness of wine varieties, and the cultural integration of wine into various social settings. This rise in demand has spurred wine industries worldwide to innovate and refine their production techniques [3]–[5]. The goal is to create wines of superior quality while simultaneously reducing production costs. This dual objective poses a complex challenge: maintaining or enhancing the quality of the wine without incurring prohibitive expenses. The endeavor to balance cost and quality has led to significant technological and methodological advancements in viticulture and enology, the sciences of grape cultivation and wine production.

Historically, the assessment of wine quality was often a retrospective process, conducted towards the end of the wine production cycle. This approach, while traditional, presented several drawbacks. Primarily, it meant that a substantial amount of time and resources had already been invested before the quality of the product was fully understood. If the wine did not meet the desired standards, rectifying flaws or discarding batches led to significant financial losses and resource wastage. Furthermore, this late-stage quality assessment did not allow for proactive adjustments during the earlier stages of production, which could have preemptively improved the wine's quality [6], [7]. To address these issues, the industry has increasingly adopted more dynamic and continuous quality monitoring techniques. These include various analytical methods, such as chemical profiling and sensory evaluations, conducted at different stages of the production process. By doing so, winemakers can make real-time adjustments, enhancing overall quality while reducing the risk of late-stage failures.

Determining wine quality remains a subjective and difficult endeavor, largely due to the diverse tastes and preferences of consumers. Every individual's taste is unique, leading to varying perceptions and appreciations of the same wine. This subjectivity poses a significant challenge for winemakers, who strive to cater to a broad market while also appealing to specific taste profiles. To navigate this complexity, wineries often employ expert tasters and sommeliers, whose trained palates can assess wines in a more standardized manner. However, even with expert evaluations, the final judgment often lies with the consumers, whose preferences can be influenced by factors beyond taste, such as branding, packaging, and marketing [8]. The industry's response has been to diversify its offerings, producing a range of wines that cater to different tastes and price points. This approach not only accommodates a wide spectrum of preferences but also helps wineries to reach broader markets, ensuring their products appeal to both connoisseurs and casual drinkers alike [9], [10].

Quality of wine is an inherently subjective concept, often defined by the individual assessing it. This subjectivity is crucial in the wine industry, where perceptions of quality can vary dramatically between experts and non-experts. Wine experts, such as sommeliers and enologists, bring a specialized understanding to the evaluation of wine quality. Their assessments are deeply rooted in an appreciation of the wine's chemical composition and the intricate processes involved in its production. This expert analysis often focuses on some aspects of wine, like the balance of flavors, the harmony of its components, and the complexity of its aroma. These characteristics are intrinsically linked to the wine's chemical makeup, including factors like acidity, tannins, sugar content, and alcohol level. Experts detect and appreciate these subtle variations, which significantly contribute to the wine's overall quality [11], [12].

In contrast, non-experts, or casual wine consumers, tend to perceive wine quality through a different lens. Their assessment is often influenced by factors such as price, branding, packaging, and the wine's origin or provenance. These consumers may equate higher prices with higher quality, or they might prefer wines from specific regions known for their wine-making heritage. For many non-experts, the experience of wine consumption is not just about the taste but also about the aesthetic and emotional appeal of the wine, which includes its presentation and the story behind it [13], [14].

The physicochemical composition of wine is a central factor in defining its quality. The unique combination and concentration of chemicals in each type of wine contribute to its distinct flavor, aroma, and color. Although most wines contain similar categories of chemicals, it is the precise concentration and balance of these compounds that create a wide array of wine profiles. This complexity is what allows for an extensive variety of wines, each with its own character and appeal. However, achieving the desired chemical balance is often a challenging process. If the quality of the wine is deemed insufficient, particularly in terms of its physicochemical composition, rectifying this issue can be a costly and time-intensive endeavor. This might involve revisiting various stages of the production process, from grape selection and fermentation to aging and bottling. Such interventions not only increase production costs but also highlight the critical nature of ongoing quality control in winemaking, ensuring that each bottle meets the desired standards of excellence, whether judged by an expert's palate or a consumer's preferences.

The advent and advancement of technology have significantly transformed the wine industry, especially in how wine quality is assessed and ensured. In the past, much of the wine testing and quality control relied on human expertise and traditional methods. However, with technological development, manufacturers have increasingly turned to sophisticated devices and tools during various development phases. These devices allow for more precise and comprehensive testing of wine properties, enabling vintners to gain a deeper and more accurate understanding of the wine's quality at every stage of production. This shift towards technology-based testing is not just a matter of efficiency; it represents a fundamental change in the approach to quality control in winemaking.

Traditional methods of quality assessment, often reliant on human sensory evaluation, can be time-consuming and subject to variability. In contrast, technological tools can quickly and consistently measure various aspects of wine, such as acidity, sugar content, alcohol level, and tannin structure. By identifying potential quality issues early in the production process, winemakers can make necessary adjustments without the extensive costs associated with later-stage corrections. This proactive approach to quality control is not only more economical but also enhances the overall standard of the wine produced.

Moreover, the integration of technology in wine production has facilitated the accumulation of extensive data on various aspects of the winemaking process. Parameters such as the quantity of different chemicals, temperature controls, fermentation rates, and aging conditions are meticulously recorded and analyzed. This wealth of data provides a comprehensive overview of the production process, offering insights that were previously inaccessible or difficult to obtain. The rise of machine learning (ML) techniques has further revolutionized this aspect of the wine industry [15], [16]. With the success of ML in various fields over the past decade, there have been concerted efforts to apply these techniques to determine wine quality. By analyzing the amassed data, ML algorithms can identify patterns and correlations that human experts might overlook. These algorithms can predict the quality of wine based on its chemical composition and production conditions, offering a powerful tool for winemakers to enhance their products. The application of ML in wine quality assessment exemplifies how technology not only streamlines existing processes but also opens new avenues for innovation and excellence in the wine industry.

The integration of advanced technology in wine production has led to a paradigm shift in how wine quality is managed and optimized. One of the most significant benefits of this technological advancement is the ability to fine-tune various parameters during the winemaking process. This precise control over the factors that directly influence wine quality allows manufacturers to not only maintain high standards but also to experiment and innovate in their wine production. This ability to adjust specific parameters—such as temperature, fermentation time, acidity, sugar levels, and tannin concentration—enables winemakers to refine the flavor profile, aroma, and overall character of their wines with unprecedented accuracy. By altering these variables, manufacturers can enhance certain qualities of the wine or even create entirely new flavor profiles. This level of control is particularly beneficial in responding to changing consumer tastes and market trends, allowing wineries to stay competitive and relevant.

Moreover, this process of fine-tuning parameters can lead to the development of wines with unique and diverse tastes, potentially giving rise to new wine styles or brands. This aspect of wine production is especially exciting, as it opens up possibilities for innovation and diversification in the wine market. Winemakers can experiment with different combinations of grape varieties, fermentation techniques, and aging processes, leading to the creation of distinctive wines that can appeal to a wide range of consumers [17], [18].

Therefore, the analysis of basic parameters that determine wine quality is not just essential for maintaining standards, but it is also a crucial aspect of product development in the wine industry. Through careful monitoring and adjustment of these parameters, winemakers can craft wines that not only meet but exceed consumer expectations, while also exploring new directions in wine production. This evolution in the winemaking process emphasizes the importance of technology and data analysis in the continuous evolution and growth of the wine industry [16], [19].

Physicochemical properties of wine

Fixed Acidity (g(tartaric acid)/dm³)

Fixed acidity in wine, typically measured in grams of tartaric acid per cubic decimeter (g(tartaric acid)/dm^3), is a crucial parameter in winemaking and wine analysis [20]. It primarily consists of tartaric and malic acids, which are inherent in grapes. The level of fixed acidity in a wine significantly influences its taste, color stability, and microbial stability. Wines with higher fixed acidity tend to have a more pronounced tartness, which is essential in balancing the sweetness and fruitiness in the wine. The measurement of fixed acidity is vital for winemakers to ensure the desired style and quality of the wine, as it affects both the sensory properties and the wine's overall stability and aging potential [21].

Volatile Acidity (g(acetic acid)/dm³)

Volatile acidity in wine, expressed as grams of acetic acid per cubic decimeter (g(acetic acid)/dm^3), is an important indicator of wine quality. This measure primarily reflects the concentration of acetic acid, although other volatile acids like formic, butyric, and propionic acid can also contribute [22]. Volatile acidity is a byproduct of fermentation and can increase due to bacterial activity, particularly if the wine is exposed to air, which encourages the growth of acetic acid bacteria. A certain level of volatile acidity can add complexity and an appealing edge to the wine's aroma and flavor profile. However, if the concentration becomes too high, it can lead to an undesirable vinegar-like taste, indicating spoilage. Therefore, controlling and monitoring volatile acidity is crucial for winemakers to ensure the wine's quality and palatability [23].

Citric Acid (g/dm³)

Citric acid in wine, measured in grams per cubic decimeter (g/dm^3), plays a relatively minor but significant role in the wine's overall chemical composition and sensory profile. While the predominant acids in wine are tartaric and malic acids, citric acid, naturally present in grapes, contributes to a lesser extent [24]. Citric acid can influence the wine's freshness and flavor. It is less stable than tartaric and malic acids and can undergo microbial fermentation, leading to the formation of other compounds like diacetyl, which imparts a buttery flavor to the wine. Additionally, citric acid may be used in small quantities during winemaking for acid adjustments, particularly in wines that need a slight increase in acidity for balance. However, its use is carefully controlled, as excessive amounts can lead to undesirable flavors and increased vulnerability to certain bacterial spoilage [25].

Residual Sugar (g/dm³)

Residual sugar in wine, quantified as grams per cubic decimeter (g/dm^3), refers to the sugars left in the wine after the fermentation process is complete. These sugars are primarily fructose and glucose, originally present in the grapes. The amount of residual sugar in a wine can vary widely, from virtually none in dry wines to high levels in sweet wines. The level of residual sugar determines the sweetness of the wine and can influence its body and overall mouthfeel. Residual sugar is not just a taste element; it also plays a role in the wine's stability and aging potential. Wines with higher residual sugar levels can be more prone to microbial spoilage, necessitating careful winemaking and storage practices. Winemakers carefully monitor and control residual sugar to achieve the desired balance, sweetness, and style of the wine.

Chlorides (g(sodium chloride)/dm³)

Chlorides in wine, specifically measured as grams of sodium chloride (NaCl) per cubic decimeter (g(sodium chloride)/dm^3), are a minor but notable component of a wine's overall composition. The presence of chlorides mainly originates from the soil where the grapes are grown, as well as from certain winemaking practices. While chloride levels in wine are generally low, they are an important parameter to monitor because they can influence the taste and mouthfeel of the wine. Higher levels of chlorides can give a salty character to the wine, which in some styles, particularly in certain white wines, can enhance complexity and flavor. However, excessively high levels of chlorides can be undesirable, potentially leading to an imbalanced taste and issues with wine stability [26].

Free Sulfur Dioxide (mg/dm³)

Free sulfur dioxide (SO_2) in wine, measured in milligrams per cubic decimeter (mg/dm^3), is a critical component in winemaking due to its antioxidant and antimicrobial properties. Sulfur dioxide helps in preserving the freshness, color, and stability of the wine by preventing oxidative spoilage and inhibiting the growth of undesirable microorganisms, including bacteria and wild yeasts. The level of free sulfur dioxide is the portion of the total sulfur dioxide that is not bound to other compounds in the wine and is available to exert its protective effects. The amount of free sulfur dioxide necessary varies depending on factors like the wine's pH, alcohol content, and residual sugar level. Wines with lower pH (more acidic) typically require less sulfur dioxide for preservation [27].

Total Sulfur Dioxide (mg/dm³)

Total sulfur dioxide (SO_2) in wine, expressed in milligrams per cubic decimeter (mg/dm^3), is the combined measure of free and bound sulfur dioxide [28]. It plays a crucial role in wine preservation, acting as an antioxidant and antimicrobial agent. This compound helps maintain the wine's color, prevents spoilage by inhibiting unwanted bacteria and yeasts, and preserves its freshness. The balance of total sulfur dioxide is vital; while adequate levels ensure the wine's stability and longevity, excessive amounts can negatively affect the wine's flavor and aroma.

Density (g/cm³)

Density in wine, measured in grams per cubic centimeter (g/cm^3), is an important physical property that reflects the overall composition of the wine. This measurement is influenced by various components in the wine, such as alcohol, sugar, and extract contents. Generally, the density of wine is slightly higher than that of water due to the presence of these dissolved substances. For instance, wines with higher sugar content tend to have greater density, whereas those with higher alcohol content, which is less dense than water, may have a lower density. Monitoring the density is crucial for winemakers, as it helps in assessing the wine's maturity and quality, determining the progress of fermentation (especially the conversion of sugar to alcohol), and ensuring consistency in the final product. The density of wine can also provide insights into its mouthfeel and body, contributing to the overall sensory experience of the wine.

pH

The pH of wine is a critical parameter that measures its acidity, expressed on a scale from 0 to 14. Wine typically has a pH range between 3.0 and 4.0, indicating its naturally acidic nature. This acidity comes primarily from organic acids present in grapes, such as tartaric, malic, and citric acids. The pH level in wine influences various aspects of its character and quality, including taste, color, and microbial stability [29]. A lower pH (more acidic) wine tends to have a sharper, more vibrant taste, and can better resist spoilage by harmful microorganisms. It also helps preserve the wine's color, especially in red wines, by stabilizing the pigments. Conversely, wines with higher pH levels (less acidic) are more susceptible to bacterial growth and may have a softer, rounder mouthfeel. The pH level is also a crucial factor in winemaking decisions, such as the type and amount of sulfites to add for preservation, as wines with lower pH require less sulfur dioxide to remain stable [30].

Sulphates (g(potassium sulphate)/dm³)

Sulphates in wine, measured as grams of potassium sulphate per cubic decimeter (g(potassium sulphate)/dm^3), are an indicator of the wine's exposure to sulfur-based compounds, typically used in winemaking for preservation and fermentation control. Sulphates primarily originate from the addition of sulfur dioxide (SO_2), which can convert to sulfates in the wine. While sulfates themselves are not as active in preservation as sulfur dioxide, their concentration can give an indication of the overall sulfur treatment the wine has undergone. High levels of sulfates can affect the wine's taste and aroma, potentially leading to a harsh or astringent mouthfeel. Therefore, monitoring sulphate levels is essential for winemakers to maintain the desired quality, taste profile, and compliance with regulatory limits on sulfur compounds in wine.

Alcohol (vol.%)

Alcohol content in wine, expressed as a percentage of volume (vol.%), is a fundamental characteristic that significantly influences its taste, texture, and aroma. The alcohol in wine is primarily ethanol, produced during the fermentation process when yeast converts the sugars in grape juice into alcohol and carbon dioxide. The alcohol level in wine typically ranges from about 8% to 15%, with some fortified wines having higher percentages. This alcohol content contributes to the wine's body and mouthfeel, with higher alcohol levels often resulting in a richer, more viscous texture. Alcohol also affects the release of aroma compounds, enhancing the wine's bouquet. The balance between alcohol, acidity, tannins, and other components is key to the overall harmony and quality of the wine.

Table 1. Physicochemical properties of wine

Property	Description
Fixed Acidity (g(tartaric acid)/dm^3)	Measures the concentration of non-volatile acids (mainly tartaric and malic acid) in the wine, affecting its taste, color stability, and microbial stability.
Volatile Acidity (g(acetic acid)/dm^3)	Indicates the amount of acetic acid and other volatile acids in wine, affecting aroma and flavor; high levels can lead to spoilage.
Citric Acid (g/dm^3)	A minor component affecting the wine's freshness and flavor, sometimes used for acid adjustment during winemaking.
Residual Sugar (g/dm^3)	The sugar content remaining after fermentation, determining the sweetness and body of the wine, and influencing its stability.
Chlorides (g(sodium chloride)/dm^3)	Reflects the sodium chloride content, influencing the wine's taste, and can give a salty character to the wine.
Free Sulfur Dioxide (mg/dm^3)	The portion of sulfur dioxide not bound to other compounds, crucial for protecting the wine from oxidation and microbial spoilage.
Total Sulfur Dioxide (mg/dm^3)	The sum of free and bound sulfur dioxide, important for preserving wine's color, preventing spoilage, and maintaining freshness.
Density (g/cm^3)	Influenced by alcohol, sugar, and extract content, reflecting the wine's maturity and quality, and providing insights into its mouthfeel and body.

pH	Measures the acidity level, influencing taste, color, and microbial stability, with lower pH indicating higher acidity.
Sulphates (g(potassium sulphate)/dm^3)	Indicate the wine's exposure to sulfur-based compounds, affecting taste and aroma, and are a measure of the overall sulfur treatment of the wine.
Alcohol (vol.%)	The ethanol content produced during fermentation, crucial for the wine's body, mouthfeel, and aroma, with levels typically ranging from 8% to 15%.

Methods

AdaBoost (Adaptive Boosting) is an ensemble learning method primarily used for binary classification. It works by combining multiple weak classifiers to create a strong classifier. Each weak classifier, typically a decision tree, is trained on the entire dataset. After each classifier is trained, AdaBoost increases the weight of misclassified instances so that subsequent classifiers focus more on difficult cases. This process continues iteratively, and the final model is a weighted sum of these weak classifiers. The weights are based on the accuracy of each classifier, and this method is effective in reducing both bias and variance.

Extra Trees (Extremely Randomized Trees) Classifier is an ensemble learning technique that builds multiple decision trees and combines their results. It is similar to a random forest, but with two key differences: first, when choosing splits, it uses random thresholds for each feature rather than searching for the best possible thresholds; second, it splits nodes using the whole learning sample rather than a bootstrap replica. These differences generally make Extra Trees faster to train than random forests and can lead to improved model performance, especially in the presence of noisy features.

Gradient Boosting Classifier is a machine learning technique for regression and classification problems. It builds the model in a stage-wise fashion like other boosting methods but generalizes them by optimizing an arbitrary differentiable loss function. In each stage, a regression tree is fitted on the negative gradient of the given loss function, which is akin to a steepest descent step. The learning rate parameter controls the contribution of each tree. Lower rates require more trees but can yield a more robust model.

Random Forest is an ensemble learning method that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) of the individual trees. The key concept is that the ensemble of trees will generally have better predictive performance and be less susceptible to overfitting. The individual decision trees are built on randomly selected subsets of the data and features, which contributes to the diversity among the trees and ultimately results in a more robust model.

The Passive-Aggressive Classifier belongs to the family of online learning algorithms. It is used for large-scale learning and is suitable for scenarios with streaming data. The algorithm remains passive for a correct classification outcome, and turns aggressive in the event of a miscalculation, updating and adjusting. It does not converge to a fixed set of parameters, but continuously updates them to adapt to the data stream. This makes it suitable for situations where the data is continuously evolving.

Gaussian Process Classifier is a non-parametric probabilistic model under the Bayesian framework. It assumes a Gaussian process prior over functions, which defines a distribution over functions. When provided with any set of inputs, predictions are made by taking a Gaussian distribution over the output space, which provides not only the predictions but also the uncertainty of the predictions. It is particularly useful for datasets with a small number of instances but is computationally intensive for larger datasets.

A Decision Tree Classifier is a simple, interpretable, non-parametric supervised learning method used for classification. It splits the dataset into branches to form a tree structure. Each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label. The paths from root to leaf represent classification rules. The primary challenge in decision tree learning is to identify which attributes to split and when to stop splitting.

The ExtraTree Classifier is similar to the Extremely Randomized Trees Classifier, but it differs in the way it splits nodes. It uses extremely random splits of data, rather than searching for the best split among a random subset of the features. This can lead to more diversified trees and thus a reduced variance, but at the cost of a slight increase in bias.

XGB Classifier is an implementation of gradient boosting designed for speed and performance. It stands for eXtreme Gradient Boosting and is a scalable and accurate implementation of gradient boosting machines. XGBClassifier uses advanced regularization (L1 & L2), which improves model generalization capabilities. It also supports various objective functions, including regression, classification, and ranking. The model runs on both CPU and GPU, with an efficient handling of sparse data, which makes it applicable to a wide range of data science problems.

Dataset

The dataset were constructed, focusing on the red and white variants of the Portuguese "Vinho Verde" wine by [31]. These datasets integrate a comprehensive array of inputs derived from precise objective tests, providing an in-depth scientific analysis of the wines' compositions. The parameters include: fixed acidity, which influences the tartness; volatile acidity, affecting the wine's aroma; citric acid, contributing to flavor; residual sugar, determining sweetness; chlorides, indicating saltiness; free sulfur dioxide, essential for wine preservation; total sulfur dioxide, representing overall sulfur content; density, related to alcohol and sugar levels; pH, which measures acidity; sulphates, influencing taste and preservation; and alcohol content, a fundamental aspect of the wine's profile.

The output of these datasets is based on sensory evaluations conducted by seasoned wine experts. Each wine sample was rigorously assessed by these experts, who graded them on a scale from 0 (indicating very bad quality) to 10 (exemplifying excellent quality). This integration of objective chemical measurements with expert sensory evaluations offers a comprehensive understanding of the Vinho Verde wines, bridging the gap between empirical data and the nuanced art of wine tasting.

Results

Figure 1. Correlation of red wine quality with features

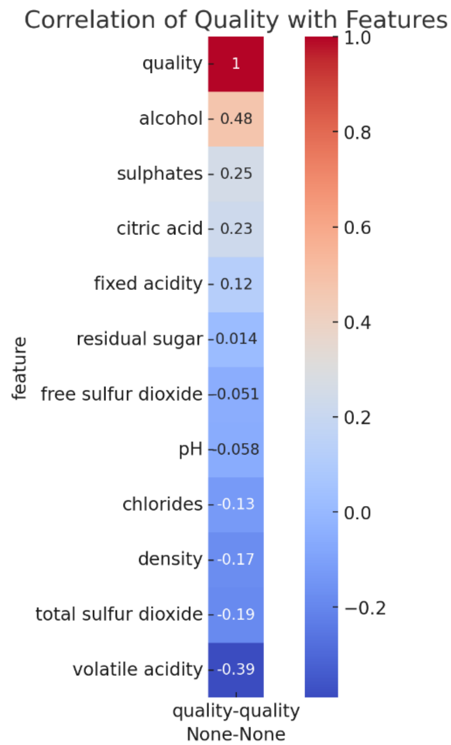

Figure 1 shows that the quality of red wine shows a significant correlation with various features, with alcohol content being the most positively correlated factor. A correlation coefficient of 0.476166 indicates that higher alcohol levels in red wine are generally associated with higher quality. Other positively correlated features include sulphates (0.251397), citric acid (0.226373), and fixed acidity (0.124052), albeit with weaker correlations compared to alcohol. This suggests that these components also play a role in enhancing the perceived quality of red wine, but to a lesser extent. On the other hand, some features negatively impact wine quality. The most negatively correlated feature is volatile acidity (-0.390558), indicating that higher levels of volatile acidity are typically associated with lower wine quality. Other negatively correlated features include total sulfur dioxide (-0.185100), density (-0.174919), chlorides (-0.128907), pH (-0.057731), and free sulfur dioxide (-0.050656), suggesting that these characteristics, when present in higher quantities, may detract from the overall quality of the wine. Interestingly, residual sugar has a very low positive correlation (0.013732) with quality, implying that its effect on the perceived quality of red wine is minimal.

Figure 2. Cross-validations accuracy for different folds (5-10) across models

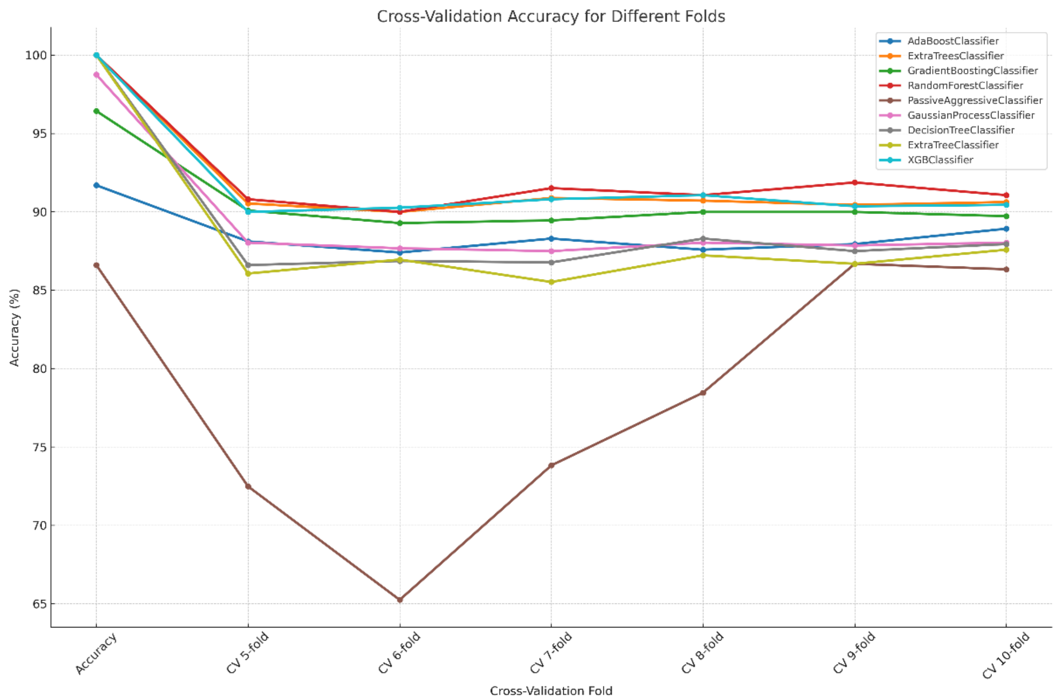

Figure 2. shows the performance of a range of machine learning algorithms (MLAs) under various cross-validation fold settings, focusing on two key metrics: Accuracy and Cross-Validation Accuracy. The fold settings range from 5 to 10. The MLAs assessed are AdaBoost Classifier, Extra Trees Classifier, Gradient Boosting Classifier, Random Forest Classifier, Passive Aggressive Classifier, Gaussian Process Classifier, Decision Tree Classifier, Extra Tree Classifier, and XGB Classifier.

For the 5-fold cross-validation, AdaBoost Classifier exhibited an accuracy of 91.69% and a cross-validation accuracy of 88.11%. Both Extra Trees Classifier and Random Forest Classifier achieved perfect accuracies of 100%, with cross-validation accuracies of 90.53% and 90.8%, respectively. Gradient Boosting Classifier showed an impressive accuracy of 96.43% and a cross-validation score of 90.08%. The Passive Aggressive Classifier had an accuracy of 86.6% but a lower cross-validation score of 72.48%. Gaussian Process Classifier demonstrated an accuracy of 98.75% and a cross-validation accuracy of 88.03%. Decision Tree Classifier and Extra Tree Classifier both achieved 100% accuracy, but their cross-validation scores were 86.6% and 86.06%, respectively. Finally, the XGB Classifier maintained a 100% accuracy with a cross-validation score of 89.99%.

In the 6-fold cross-validation, AdaBoost Classifier maintained its accuracy of 91.69% but showed a slightly reduced cross-validation accuracy of 87.4%. Both Extra Trees Classifier and Random Forest Classifier continued their trend of perfect accuracies, with cross-validation scores of 89.99%. The Gradient Boosting Classifier and Gaussian Process Classifier exhibited minor variations in their cross-validation accuracies compared to the 5-fold setting. The Passive Aggressive Classifier's performance decreased to an accuracy of 84.54% and a cross-validation score of 65.24%. Decision Tree Classifier and Extra Tree Classifier still achieved perfect accuracies but with slightly improved cross-validation scores of 86.86% and 86.95%, respectively. The XGB Classifier also maintained a 100% accuracy, with a cross-validation score of 90.26%.

In the 7-fold setting, AdaBoost Classifier's cross-validation accuracy slightly improved to 88.29%, while maintaining its original accuracy. Extra Trees Classifier and Random Forest Classifier continued with 100% accuracy, the former achieving a cross-validation score of 90.88% and the latter 91.51%. The Gradient Boosting Classifier had a cross-validation score of 89.45%, and the Gaussian Process Classifier scored 87.49%. The Passive Aggressive Classifier showed improvement in cross-validation accuracy at 73.82%. The Decision Tree Classifier and Extra Tree Classifier maintained their perfect accuracies, with cross-validation scores of 86.77% and 85.52%, respectively. The XGB Classifier continued to show robust performance with a cross-validation score of 90.8%.

The 8-fold setting saw AdaBoost Classifier with a cross-validation accuracy of 87.58%. Extra Trees Classifier and Random Forest Classifier remained at 100% accuracy, with cross-validation scores of 90.71% and 91.06%, respectively. The Gradient Boosting Classifier's cross-validation score increased to 89.99%. The Passive Aggressive Classifier's accuracy and cross-validation score both improved to 86.6% and 78.46%, respectively. The Gaussian Process Classifier had a cross-validation accuracy of 88.03%. Both the Decision Tree Classifier and Extra Tree Classifier continued with 100% accuracy, and their cross-validation scores were 88.29% and 87.22%, respectively. The XGB Classifier again showed consistent performance with a 91.06% cross-validation score.

In the 9-fold cross-validation, AdaBoost Classifier had a cross-validation accuracy of 87.94%. Extra Trees Classifier and Random Forest Classifier sustained their perfect accuracy, with cross-validation scores of 90.44% and 91.87%, respectively. The Gradient Boosting Classifier maintained a cross-validation score of 89.99%. The Passive Aggressive Classifier recorded a notable improvement in its cross-validation accuracy to 86.68%. The Gaussian Process Classifier scored 87.85% in cross-validation. The Decision Tree Classifier and Extra Tree Classifier continued their trend of perfect accuracies, with cross-validation scores of 87.49% and 86.68%, respectively. The XGB Classifier's cross-validation score was 90.35%.

In the 10-fold cross-validation setting, AdaBoost Classifier's cross-validation accuracy further improved to 88.92%. Extra Trees Classifier and Random Forest Classifier kept their perfect accuracy records, with cross-validation scores of 90.62% and 91.06%, respectively. The Gradient Boosting Classifier had a cross-validation accuracy of 89.72%. The Passive Aggressive Classifier showed a consistent accuracy of 87.22% and a cross-validation score of 86.33%. The Gaussian Process Classifier maintained a cross-validation accuracy of 88.03%. Both Decision Tree Classifier and Extra Tree Classifier continued with 100% accuracy, with cross-validation scores of 87.94% and 87.58%, respectively. The XGB Classifier ended with a cross-validation score of 90.44%.

Extra Trees Classifier, Random Forest Classifier, and XGB Classifier consistently showed high accuracy and stability across all fold settings, making them reliable choices for various applications. The Passive Aggressive Classifier, while showing variability, improved notably in higher fold settings. Gaussian Process Classifier and Gradient Boosting

Classifier also demonstrated robustness, maintaining high accuracy and cross-validation scores throughout.

Table 2. SHAP values (importance of features) for base model	
Feature	SHAP Value
Sulphates	1.236542
Alcohol	1.22195
Volatile Acidity	0.823313
Total Sulfur Dioxide	0.791003
Citric Acid	0.529945
Fixed Acidity	0.5218
Chlorides	0.512715
Density	0.45328
Free Sulfur Dioxide	0.451338
Residual Sugar	0.256595
pH	0.222885

The SHAP value analysis in Table 2 shows the influence of various features on a model's predictions, particularly in a context that seems to relate to wine quality. Sulphates lead the list with a SHAP value of around 1.236, indicating its significant influence on the model's output. Alcohol closely follows with a value of approximately 1.222, underscoring its almost equal importance in the decision-making process of the model. Volatile acidity and total sulfur dioxide also show substantial impacts with their respective values of 0.823 and 0.791, suggesting their notable but slightly lesser influence compared to sulphates and alcohol.

Further down the list, citric acid, fixed acidity, and chlorides display moderate influence on the model, each with SHAP values slightly above 0.5. Density and free sulfur dioxide present a similar level of impact on the model's predictions, with values around 0.45, denoting a lesser yet considerable effect. Residual sugar and pH are at the lower end of this spectrum with values of approximately 0.257 and 0.223, respectively, indicating they have the least influence among the evaluated features. This distribution of SHAP values suggests a clear hierarchy in feature importance, with sulphates, alcohol, and volatile acidity emerging as the top influencers in the model's predictive process.

Table 3. Feature importance	
Importance	Columns
0.055077	Density
0.056442	Residual Sugar
0.058158	Chlorides
0.061917	Citric Acid
0.065367	Fixed Acidity
0.072706	pH
0.081172	Total Sulfur Dioxide
0.084982	Free Sulfur Dioxide
0.088774	Volatile Acidity

0.126599	Sulphates
0.248807	Alcohol

Figure 4. Feature importance

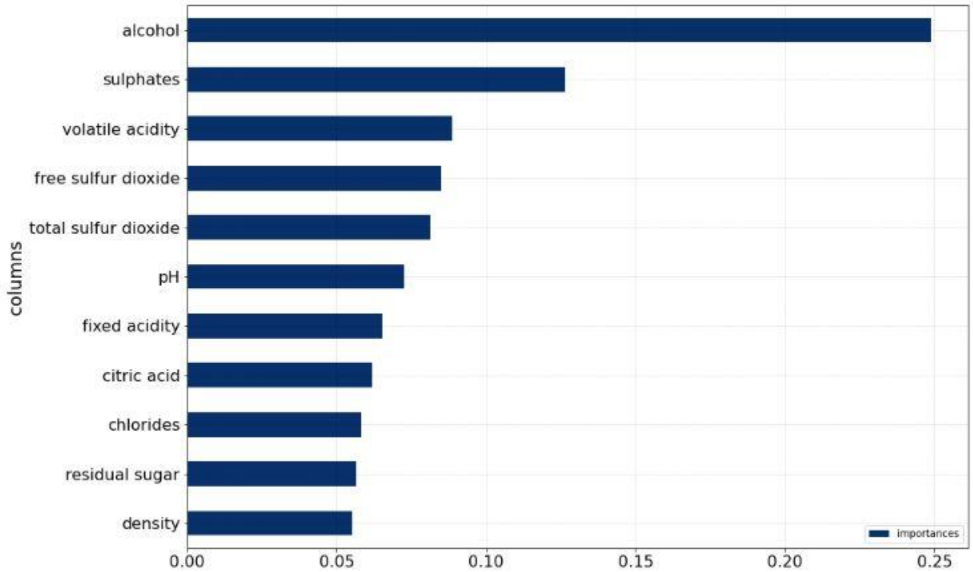

Table 4. Hyperparameter optimization	
Description	Details
Number of Finished Trials	200
Best Trial	
Value	91.15
Params	
Max Depth	7
Eta	0.414637
Gamma	2.47E-08
Lambda	0.000385
Alpha	0.985927
Min Child Weight	3
Subsample	0.699764
Colsample Bytree	0.930557

The classification report in table 5 shows metrics for two classes (0 and 1) before and after optimization. For class 0, the precision remained constant at 0.93 in both the original and optimized model. However, recall and F1-score saw a slight decrease in the optimized model, dropping from 0.95 to 0.93 for recall, and from 0.94 to 0.93 for F1-score. The support for class 0 was consistent at 413 in both models. Class 1 experienced a decline in precision after optimization, from 0.64 to 0.57, while recall remained steady at 0.54. The

F1-score for class 1 also decreased slightly, from 0.59 in the original model to 0.55 in the optimized version, with a constant support of 67.

Figure 5. Contour plot for parameter relationship

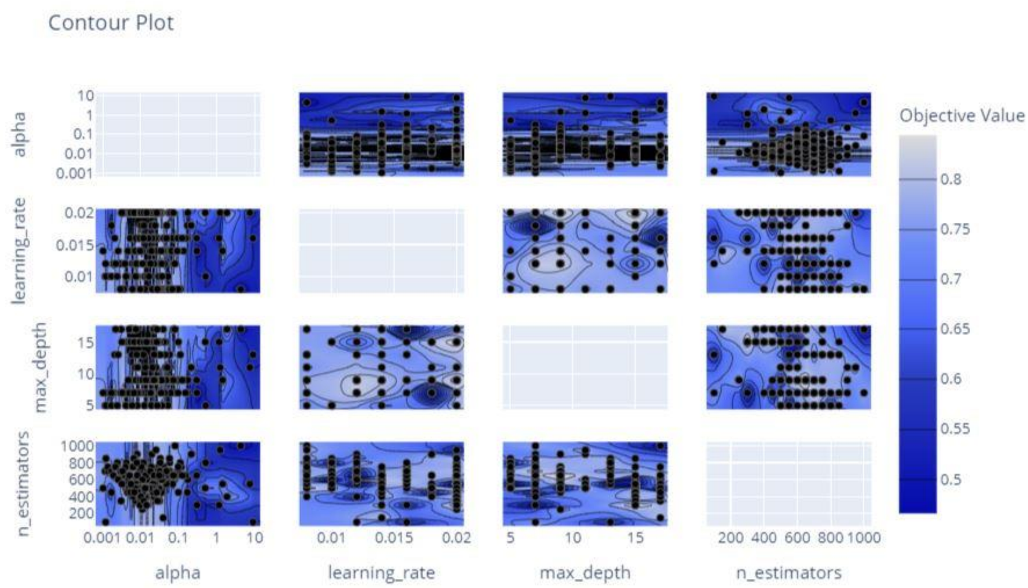

Figure 6. confusion matrix (base model vs optimized model)

Base model *Optimized model*

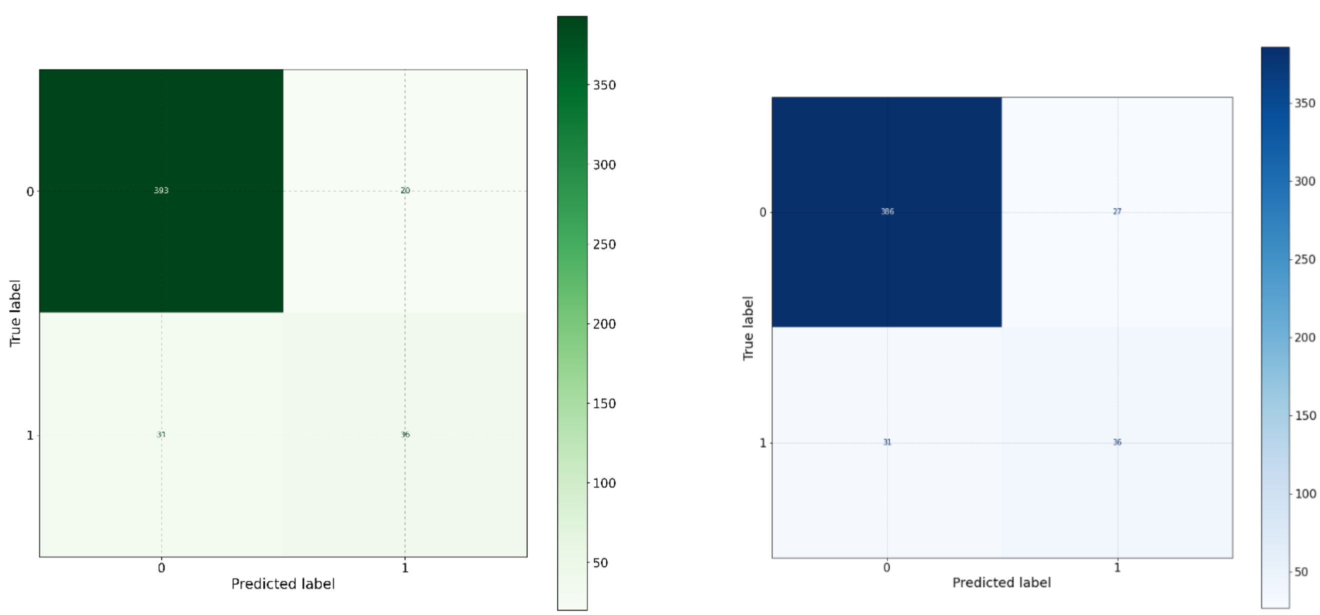

Table 5. average model performance report (base model and optimized model)			
Class	Metric	Classification Report (base model)	Classification Report (optimized model)
0	Precision	0.93	0.93
0	Recall	0.95	0.93
0	F1-Score	0.94	0.93
0	Support	413	413
1	Precision	0.64	0.57
1	Recall	0.54	0.54
1	F1-Score	0.59	0.55
1	Support	67	67
	Accuracy	0.89	0.88
	Macro Avg	0.78 (Precision), 0.74 (Recall), 0.76 (F1-Score)	0.75 (Precision), 0.74 (Recall), 0.74 (F1-Score)
	Weighted Avg	0.89 (Precision), 0.89 (Recall), 0.89 (F1-Score)	0.88 (Precision), 0.88 (Recall), 0.88 (F1-Score)
	Total Support	480	480

In terms of overall model performance, the accuracy of the model slightly decreased from 0.89 in the original to 0.88 in the optimized model. The macro averages for precision, recall, and F1-score were 0.78, 0.74, and 0.76, respectively, in the original model, and showed a marginal decline in the optimized model to 0.75 for precision, 0.74 for recall, and 0.74 for F1-score. The weighted averages, which consider the support of each class, also saw a small decline post-optimization. The original model had weighted averages of 0.89 for precision, recall, and F1-score, whereas the optimized model scored 0.88 across these three metrics. The total support for both models was consistent at 480.

Table 6. SHAP values (importance of features) for base model	
Feature	SHAP Value
Sulphates	1.184194
Alcohol	1.161134
Volatile Acidity	0.7868
Total Sulfur Dioxide	0.678422
Density	0.502056
Citric Acid	0.407366
Fixed Acidity	0.398065
pH	0.360671
Chlorides	0.357492
Free Sulfur Dioxide	0.330986
Residual Sugar	0.249677

In the analysis of feature importance using SHAP values in the optimized model (shown in table 6, various features showed differing levels of influence on the model's output. Sulphates had the highest SHAP value at 1.184194, indicating its significant impact, closely followed by Alcohol with a SHAP value of 1.161134. Volatile Acidity also had a substantial influence with a SHAP value of 0.7868. Total Sulfur Dioxide and Density were moderately influential, having SHAP values of 0.678422 and 0.502056, respectively. Citric Acid, Fixed Acidity, pH, and Chlorides demonstrated relatively lesser but notable effects, with their SHAP values ranging from 0.357492 for Chlorides to 0.407366 for Citric Acid. Free Sulfur Dioxide and Residual Sugar had the lowest SHAP values in the set, at 0.330986 and 0.249677 respectively, suggesting their lesser impact on the model's predictions compared to the other features.

Conclusion

The aim of this study was to develop and validate a machine learning model that can predict wine quality based on its physicochemical properties. This research seeks to address the limitations inherent in traditional sensory evaluation methods, which are often subjective and inconsistent, by proposing a more objective, efficient, and cost-effective approach. The study involves a comparative analysis of various machine learning algorithms to determine the most effective model for predicting wine quality. A significant aspect of this research is the use of SHapley Additive exPlanations (SHAP) to identify the key physicochemical factors that influence wine quality. Additionally, the study focuses on refining the chosen model through hyperparameter optimization to improve and validate its performance

The quality of wine exhibited a strong positive correlation with its alcohol content (0.476166), indicating that higher alcohol levels often correspond to higher perceived quality. Conversely, there was a notable negative correlation with volatile acidity (-0.390558), suggesting that higher levels of volatile acidity tend to decrease the perceived quality of wine. When examining the performance of various machine learning algorithms, the ExtraTreesClassifier stood out with exceptional accuracy, achieving a score of 99.98% in standard testing. This high level of accuracy was consistent even in cross-validation tests, with scores ranging from 89.99% to 91.51% across different folds. This robust performance highlights the effectiveness of the ExtraTreesClassifier in handling the nuances of wine quality prediction. Furthermore, the application of SHAP (SHapley Additive exPlanations) values provided deeper insights into the model's decision-making process. These values indicated that sulphates, alcohol, and volatile acidity were the most significant factors influencing the model's predictions about wine quality. In contrast, the AdaBoostClassifier demonstrated consistent but slightly lower performance, with accuracy varying between 87.4% and 88.92% across different folds of cross-validation. This variation suggests that while AdaBoostClassifier is reliable, it might not be as effective as the ExtraTreesClassifier in certain scenarios for predicting wine quality.

After hyperparameter optimization, the results indicated changes in model performance and feature influence. The optimization process, which involved 200 trials, resulted in the best trial achieving a value of 91.15%. The key parameters optimized during this process included Max Depth, Eta, Gamma, Lambda, Alpha, Min Child Weight, Subsample, and Colsample Bytree. These adjustments to the model's hyperparameters were aimed at enhancing its predictive accuracy and handling the complexities of wine quality assessment more effectively. Post-optimization, the classification performance showed slight variations in metrics. Notably, the precision for class 0 (high quality) remained stable at 0.93, while for class 1 (lower quality), there was a decrease from 0.64 to 0.57. This change might indicate a shift in the model's sensitivity to certain quality indicators after optimization. The recall scores for both classes remained unchanged, maintaining their previous levels. However, the overall accuracy of the model experienced a slight decrease, moving from 0.89 to 0.88. This suggests that while the model became more fine-tuned in certain aspects, it might have lost some general accuracy in the process. The F1-Scores for both classes also witnessed a decrease, with class 0 dropping from 0.94 to 0.93 and class 1 from 0.59 to 0.55. These changes in F1-Scores, along with the macro and weighted averages across precision, recall, and F1-score, indicate a marginal reduction in the model's performance post-optimization. Despite these changes, the SHAP values continued to highlight the critical influence of sulphates, alcohol, and volatile acidity on the model's predictions, reaffirming their significant roles in determining wine quality.

Sulphates in wine play a dual role. They act as preservatives, preventing microbial growth that could spoil the wine, and they also influence the wine's flavor profile. The presence of sulphates can contribute to a wine's complexity and longevity. These compounds, when balanced correctly, can enhance the preservation of the wine's original character and taste over time. However, excessive sulphates can lead to an undesirable taste, often described as a burnt match or a chemical flavor. The key in winemaking is to strike a balance: enough sulphates to protect the wine, but not so much that it overpowers the natural flavors.

Alcohol is another crucial factor in determining wine quality according to the findings of this study. It is not just about the strength of the alcohol but how well it is integrated into the overall composition of the wine. High alcohol levels in a wine can give it body, richness, and a velvety texture, contributing to a pleasing mouthfeel. However, if the alcohol content is too high and not well-balanced with the other elements of the wine, such as acidity, tannins, and fruit flavors, it can lead to an overwhelming sensation of heat in the mouth. This imbalance can mask the more subtle flavors and aromas of the wine, detracting from the overall drinking experience.

Volatile acidity (VA) at high levels can lead to unpleasant vinegar-like tastes and smells. In small amounts, however, volatile acidity can add to the complexity and character of the wine, imparting a slight sharpness that can enhance its flavor profile. The control of VA is a delicate process in winemaking; too much can ruin a wine, but just the right amount can elevate it. The challenge for winemakers is to manage the fermentation and aging processes in such a way that keeps the volatile acidity at a level that contributes positively to the wine's character without allowing it to become dominant or detrimental.

SO2 helps to prevent oxidation and maintain the wine's color and freshness. It also inhibits the growth of bacteria and wild yeasts, ensuring that the wine remains stable and drinkable over time. However, like sulphates, the amount of SO2 in wine must be carefully managed. Excessive sulfur dioxide can lead to negative sensory attributes, such as a pungent, sulfurous aroma and an altered taste. The challenge in winemaking is to use enough SO2 to protect the wine effectively while avoiding concentrations that could negatively impact its taste and aroma. The managing SO2 levels is a key factor in producing high-quality wines that age well and retain their desired characteristics over time.

Residual sugar is a significant factor in certain types of wines, especially in dessert wines and some Rieslings, where a balance between sweetness, acidity, and flavor is key. However, in many dry wines, the level of residual sugar is typically very low and has a minimal impact on the overall perception of the wine. In these cases, the nuances of flavor, aroma, and texture – often influenced more by the factors like alcohol content, acidity, and tannins – take precedence over sweetness. While residual sugar can contribute to the body and mouthfeel of the wine, its influence is often overshadowed in dry wines by these other characteristics.

pH, on the other hand, plays a more subtle yet essential role in wine. It affects the color, stability, and taste of the wine. A lower pH (more acidic) can contribute to a wine's freshness and vitality, while a higher pH (less acidic) can make a wine feel softer and rounder. However, compared to other factors like alcohol and sulphates, the influence of pH is less directly perceived by the consumer. It is more about creating a suitable environment for the wine's preservation and aging, and for ensuring that other elements like tannins and fruit flavors are in balance. While pH is a critical factor in winemaking, its role is more in the background, setting the stage for other aspects of the wine to shine.

The incorporation of machine learning technologies in the assessment of wine quality represents a significant advancement for various stakeholders in the wine industry, including producers, distributors, and consumers. This innovative approach surpasses traditional methods by offering a more objective framework for evaluating the quality of wine. The subjective nature of wine tasting, often influenced by individual preferences and environmental factors, has historically posed challenges in maintaining consistent quality standards [32]. Machine learning algorithms, however, can analyze complex datasets derived from chemical and physical properties of wines, leading to a more standardized and reliable assessment. This objective analysis is not only beneficial for quality control but also aids in the categorization of wines, offering a systematic approach for classification based on flavor profiles, aging potential, and regional characteristics [33]. Additionally, the efficiency of machine learning in processing large volumes of data significantly reduces the time and labor traditionally required in wine quality assessment, translating into cost savings for producers and distributors. The automation of these processes also minimizes human error, ensuring a more consistent and dependable quality evaluation. This approach not only promotes standardization in the wine industry but also opens avenues for further research in the application of machine learning for assessing other sensory-based products.

References

[1] C. Santini, A. Cavicchi, and L. Casini, "Sustainability in the wine industry: key questions and research trendsa," *Food Res. Inst. Stud. Agric. Econ. Trade Dev.*, vol. 1, no. 1, pp. 1–14, Aug. 2013.

[2] L. F. Bisson, A. L. Waterhouse, S. E. Ebeler, M. A. Walker, and J. T. Lapsley, "The present and future of the international wine industry," *Nature*, vol. 418, no. 6898, pp. 696–699, Aug. 2002.

[3] L. Ruggieri *et al.*, "Recovery of organic wastes in the Spanish wine industry. Technical, economic and environmental analyses of the composting process," *J. Clean. Prod.*, vol. 17, no. 9, pp. 830–838, Jun. 2009.

[4] D. Aylward, "Towards a Cultural Economy Paradigm for the Australian Wine Industry," *Prometheus*, vol. 26, no. 4, pp. 373–385, Dec. 2008.
[5] H. Vijayakumar, "The Impact of AI-Innovations and Private AI-Investment on U.S. Economic Growth: An Empirical Analysis," *Reviews of Contemporary Business Analytics*, vol. 4, no. 1, pp. 14–32, 2021.
[6] I. Belda *et al.*, "Microbial Contribution to Wine Aroma and Its Intended Use for Wine Quality Improvement," *Molecules*, vol. 22, no. 2, Jan. 2017.
[7] R. Mira de Orduña, "Climate change associated effects on grape and wine quality and production," *Food Res. Int.*, vol. 43, no. 7, pp. 1844–1855, Aug. 2010.
[8] V. Soni, "Role of Temporal, Demographic, and Behavioral Factors in Customer Conversion Through Dynamic Creative Optimization in the Consumer-Packaged Goods Setting," *International Journal of Business Intelligence and Big Data Analytics*, vol. 5, no. 1, pp. 46–56, 2022.
[9] R. Johnson and J. Bruwer, "Regional brand image and perceived wine quality: the consumer perspective," *International Journal of Wine Business Research*, vol. 19, no. 4, pp. 276–297, Nov. 2007.
[10] C. van Leeuwen and P. Darriet, "The impact of climate change on viticulture and wine quality," *J. Wine Econ.*, vol. 11, no. 1, pp. 150–167, May 2016.
[11] R. Gawel and P. W. Godden, "Evaluation of the consistency of wine quality assessments from expert wine tasters," *Aust. J. Grape Wine Res.*, 2008.
[12] R. Gutiérrez-Escobar, M. J. Aliaño-González, and E. Cantos-Villar, "Wine Polyphenol Content and Its Influence on Wine Quality and Properties: A Review," *Molecules*, vol. 26, no. 3, Jan. 2021.
[13] Y. Tao, J. F. García, and D.-W. Sun, "Advances in wine aging technologies for enhancing wine quality and accelerating wine aging process," *Crit. Rev. Food Sci. Nutr.*, vol. 54, no. 6, pp. 817–835, 2014.
[14] B. Tariba, "Metals in Wine—Impact on Wine Quality and Health Outcomes," *Biol. Trace Elem. Res.*, vol. 144, no. 1, pp. 143–156, Dec. 2011.
[15] M. Koranga, R. Pandey, M. Joshi, and M. Kumar, "Analysis of white wine using machine learning algorithms," *Materials Today: Proceedings*, vol. 46, pp. 11087–11093, Jan. 2021.
[16] 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)*, 2018, pp. 139–143.
[17] S. Kumar, K. Agrawal, and N. Mandan, "Red Wine Quality Prediction Using Machine Learning Techniques," in *2020 International Conference on Computer Communication and Informatics (ICCCI)*, 2020, pp. 1–6.
[18] A. Trivedi and R. Sehrawat, "Wine Quality Detection through Machine Learning Algorithms," in *2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE)*, 2018, pp. 1756–1760.
[19] V. Soni, "Impact of Generative AI on Small and Medium Enterprises' Revenue Growth: The Moderating Role of Human, Technological, and Market Factors," *Reviews of Contemporary Business Analytics*, vol. 6, no. 1, pp. 133–153, 2023.
[20] S. S. M. P. Vidigal and A. O. S. S. Rangel, "A flow-based platform for measuring the acidity parameters in wine," *Talanta*, vol. 168, pp. 313–319, Jun. 2017.

[21] J. Darias-Martín, A. Socas-Hernández, C. Díaz-Romero, and E. Díaz-Díaz, "Comparative study of methods for determination of titrable acidity in wine," *J. Food Compost. Anal.*, vol. 16, no. 5, pp. 555–562, Oct. 2003.
[22] N. Kameneva and O. Tkachenko, "Influence of preparations Biolan and Vympel for the crop and quality of grapes and wine from varieties Aligote and Rkatsiteli," *Songklanakarin Journal of Science &*, 2019.
[23] E. V. Dubinina, D. V. Andrievskaya, S. M. Tomgorova, and K. V. Nebezhev, "Innovative technologies of alcoholic beverages based on fruit distillates," *Food Syst.*, vol. 3, no. 2, pp. 18–23, Jul. 2020.
[24] Y. Shimazu, M. Uehara, and M. Watanabe, "Transformation of Citric Acid to Acetic Acid, Acetoin and Diacetyl by Wine Making Lactic Acid Bacteria," *Agric. Biol. Chem.*, vol. 49, no. 7, pp. 2147–2157, Jul. 1985.
[25] C. Dartiguenave, P. Jeandet, and A. Maujean, "Study of the Contribution of the Major Organic Acids of Wine to the Buffering Capacity of Wine in Model Solutions," *Am. J. Enol. Vitic.*, vol. 51, no. 4, pp. 352–356, Jan. 2000.
[26] L. Moio, A. Gambuti, L. Di Marzio, and P. Piombino, "Differential pHmeter Determination of Residual Sugars in Wine," *Am. J. Enol. Vitic.*, vol. 52, no. 3, pp. 271–274, Jan. 2001.
[27] T. M. Monro *et al.*, "Sensing free sulfur dioxide in wine," *Sensors* , vol. 12, no. 8, pp. 10759–10773, Aug. 2012.
[28] G. L. Sacks, P. A. Howe, M. Standing, and J. C. Danilewicz, "Free, Bound, and Total Sulfur Dioxide (SO2) during Oxidation of Wines," *Am. J. Enol. Vitic.*, vol. 71, no. 4, pp. 266–277, Oct. 2020.
[29] N. Kontoudakis, M. Esteruelas, F. Fort, J. M. Canals, and F. Zamora, "Use of unripe grapes harvested during cluster thinning as a method for reducing alcohol content and pH of wine," *Aust. J. Grape Wine Res.*, vol. 17, no. 2, pp. 230–238, Jun. 2011.
[30] K. Vahl, H. Kahlert, L. von Mühlen, A. Albrecht, G. Meyer, and J. Behnert, "Determination of the titratable acidity and the pH of wine based on potentiometric flow injection analysis," *Talanta*, vol. 111, pp. 134–139, Jul. 2013.
[31] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, "Modeling wine preferences by data mining from physicochemical properties," *Decis. Support Syst.*, vol. 47, no. 4, pp. 547–553, Nov. 2009.
[32] R. S. Jackson, *Wine tasting: A Professional Handbook*. Elsevier, 2022.
[33] C. S. Parr, D. G. Lemay, C. L. Owen, M. Jennifer Woodward-Greene, and J. Sun, "Multimodal AI to Improve Agriculture," *IT Prof.*, vol. 23, no. 3, pp. 53–57, 01 May-June 2021.