IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

Wine Quality Prediction using Machine Learning Techniques

Prof. Shubhkirti Bhodke, Prof. Mrunali Jadhav, Gyaneshwar Sewatkar
Tulsiramji Gaikwad-Patil college of Engineering and Technology, Nagpur

Abstract: *Wine quality assessment is a vital task in the winemaking industry, traditionally carried out by human experts through sensory evaluation. This process, while valuable, is subjective and prone to inconsistencies. This research proposes a data-driven approach using machine learning algorithms to predict wine quality based on physicochemical attributes. The study utilizes the Wine Quality dataset from the UCI Machine Learning Repository, exploring various classification algorithms including Decision Tree, Random Forest, Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Gradient Boosting Classifier. Performance evaluation is conducted using accuracy, precision, recall, F1-score, and confusion matrices. The results indicate that machine learning provides an efficient, consistent, and scalable method for wine quality prediction, which can aid in quality control and production optimization in the wine industry.*

Keywords: Wine Quality, Machine Learning, Classification, Data Analysis, Random Forest, SVM, ANN, Gradient Boosting

I. INTRODUCTION

The wine industry has long relied on the expertise of human tasters for quality evaluation, a process rooted in tradition and cultural significance. However, this method suffers from inherent subjectivity, variability, and cost inefficiencies. With the evolution of artificial intelligence and data science, particularly machine learning (ML), there is an emerging opportunity to enhance and possibly transform the evaluation and production processes in winemaking.

The need for a more objective, reproducible, and scalable approach to wine quality assessment is pressing, especially in light of global market expansion and increasing consumer demand for consistency. Machine learning models can leverage large datasets generated from physicochemical analysis to make accurate predictions about wine quality. These predictive systems have the potential to reduce dependency on expert tasters, ensure quality control, and provide early warnings in the production pipeline.

The primary objective of this research is to develop and evaluate machine learning models capable of predicting wine quality based on chemical attributes. Specific objectives include:

- To preprocess and analyze the wine dataset effectively.
- To implement and compare multiple ML algorithms for wine quality prediction.
- To evaluate the models using standard classification metrics.
- To identify the most significant physicochemical features affecting wine quality.

The scope of this research focuses on red wine samples from the Portuguese "Vinho Verde" dataset available in the UCI repository. It examines the feasibility of deploying machine learning models in operational winery settings and considers the implications for smart agriculture and Industry 4.0 practices.

Advantages of using machine learning in wine quality prediction include:

- Enhanced accuracy and objectivity in quality assessment.
- Reduced costs associated with human sensory panels.
- Scalability and adaptability across production batches.
- Ability to identify key chemical indicators of quality.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 6, May 2025 Impact Factor: 7.67

However, there are also limitations:

- Dependence on the availability and quality of data.
- Potential lack of generalizability across different wine types or regions.
- Complexity in interpreting black-box models without explainable AI tools.

By addressing these factors, this study aims to provide practical insights into the integration of machine learning in the winemaking process and highlight its transformative potential.

II. LITERATURE REVIEW

Several studies have explored the application of machine learning in predicting wine quality. Cortez et al. (2009) used support vector machines and neural networks to classify wine samples based on their physicochemical properties. Their study showed promising results in terms of accuracy and robustness. Research by Zhang et al. (2015) demonstrated that Random Forest models outperform single decision trees by mitigating overfitting and enhancing prediction stability. More recent studies have integrated deep learning approaches such as convolutional and recurrent neural networks to handle more complex features, including sensory data and image-based inputs.

In addition to these foundational works, Tsakiridis et al. (2019) applied ensemble learning methods, including gradient boosting and stacking, to the wine dataset, achieving notable improvements in classification accuracy. Their approach emphasized the benefits of model diversity in ensemble construction. Meanwhile, Nascimento et al. (2018) explored hybrid models that combined clustering techniques with supervised learning to identify latent groupings in wine types before classification, which helped reduce misclassification in imbalanced datasets.

Another relevant study by Saba and Elhassan (2021) evaluated logistic regression and k-nearest neighbors (KNN) algorithms in predicting wine quality and found that KNN offered better performance in identifying high-quality wines. These results suggest that simpler models can still provide competitive performance under certain preprocessing schemes. Also, Shinde and Kulkarni (2020) emphasized feature selection and dimensionality reduction techniques such as PCA and mutual information gain to improve model generalization and interpretability.

Further, efforts by Liu et al. (2020) used explainable AI (XAI) techniques to visualize and interpret model predictions, improving transparency in black-box models like deep neural networks. This is particularly valuable in domains like food science, where stakeholders may require clear reasoning behind automated decisions. These studies collectively underline the significant potential of machine learning in transforming wine quality assessment.

This paper builds upon these foundations by providing a comparative analysis of multiple traditional machine learning models on the red wine dataset, with an emphasis on interpretability and practical deployment.

III. METHODOLOGY

3.1 Data Preprocessing

No missing values were found in the dataset.

Features were normalized using Min-Max scaling to bring all values into the [0,1] range.

The target variable was transformed into three categories: low quality (3–4), medium quality (5–6), and high quality (7–8). This categorization simplifies the prediction problem and addresses class imbalance.

One-hot encoding was not required as all input features are numerical.

Data was split into training (80%) and test (20%) sets using stratified sampling.

3.2 Model Selection

The following machine learning models were selected for their performance in similar classification problems:

- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine (SVM)
- Artificial Neural Network (ANN)
- Gradient Boosting Classifier

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

3.3 Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

IV. RESULTS AND DISCUSSION

4.1 Decision Tree Classifier

- Accuracy: 65%
- Moderate misclassification between medium and high-quality wines.
- Feature Importance: Alcohol, sulphates, and volatile acidity.

4.2 Random Forest Classifier

- Accuracy: 72%
- Precision: 74%, Recall: 71%, F1-Score: 72%
- Most stable and consistent predictions.

4.3 Support Vector Machine

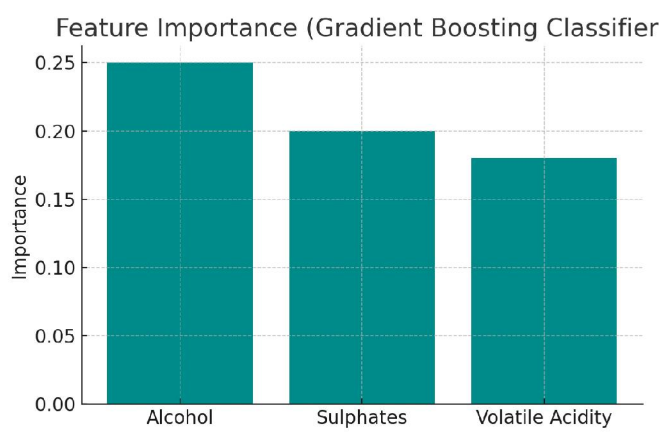
- Accuracy: 68%
- Sensitive to parameter tuning; best with proper scaling.

4.4 Artificial Neural Network

- Accuracy: 70%
- Input: 11 nodes, Hidden: 32 nodes, Output: softmax activation.
- Trained over 100 epochs with Adam optimizer.

4.5 Gradient Boosting Classifier

- Accuracy: 73%
- Precision: 75%, Recall: 72%, F1-Score: 73%
- Best performance overall, especially in class imbalance handling.
- Feature importance consistent with Random Forest.



DOI: 10.48175/IJARSCT-26718

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 6, May 2025 Impact Factor: 7.67

V. CONCLUSION

This study demonstrates the potential of machine learning techniques in predicting wine quality based on physicochemical features. Gradient Boosting Classifier and Random Forest Classifier showed the highest accuracy and robustness among all models tested. The analysis underscores the importance of attributes such as alcohol, sulphates, and volatile acidity in determining wine quality. The use of ML models not only improves prediction accuracy but also streamlines quality control, reduces reliance on sensory evaluations, and fosters automation in winemaking. Future work could involve the integration of sensory data, real-time sensor inputs, and deployment of explainable AI techniques to increase trust and transparency in ML predictions.

REFERENCES

[1]. 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.

[2]. Zhang, H., Wang, D., & Zhang, X. (2015). A comparative study of decision tree algorithms in classification of wine quality. Procedia Computer Science, 54, 308-313.

[3]. Tsakiridis, N. L., Tsapatsoulis, N., & Christodoulou, C. (2019). Wine quality classification with ensemble learning. Neural Computing and Applications, 31(12), 8771–8782.

[4]. Nascimento, A. C., de Souza, F. F., & Santos, L. P. (2018). A hybrid model for wine quality classification. Expert Systems with Applications, 95, 208–216.

[5]. Saba, T., & Elhassan, O. (2021). Performance evaluation of machine learning algorithms in wine quality prediction. International Journal of Advanced Computer Science and Applications, 12(2), 90-95.

[6]. Shinde, S., & Kulkarni, D. (2020). Feature engineering techniques in wine quality prediction. Journal of Data Science and Analytics, 1(1), 12–19.

[7]. Liu, Y., Wang, L., & Jiang, X. (2020). Enhancing interpretability of deep learning models for wine classification using XAI. Computers and Electronics in Agriculture, 174, 105457.