Explainable AI for Wine Quality Prediction Using Machine Learning

Adishree Karthikeyan
School of Computer Science and Engineering (SoCSE)
RV University, Bengaluru, India
Email: adishreek.bsc22@rvu.edu.in

***Abstract*—Explainable Artificial Intelligence (XAI) has emerged as a crucial component in data-driven decision-making systems, particularly in fields where interpretability and trust are vital. This study focuses on developing an interpretable machine learning model for predicting wine quality based on its physicochemical properties. Using the UCI Wine Quality dataset, we trained and analyzed multiple classification algorithms, including Logistic Regression, Random Forest, and XGBoost, integrated with SHAP (SHapley Additive exPlanations) for interpretability. Our objective is to evaluate model transparency and feature importance, allowing domain experts to understand the reasoning behind each prediction. The study demonstrates how explainability enhances reliability, supports decision-making, and bridges the gap between machine learning outputs and practical domain insights.**

***Index Terms*—Explainable AI, SHAP, Machine Learning, Wine Quality Prediction, Model Interpretability**

I. Introduction

Wine quality assessment plays a critical role in the wine industry, influencing both production and marketing strategies. Traditionally, quality evaluation relies on expert tasters and sensory analysis, which can be subjective and resource-intensive. With the advancement of machine learning, data-driven methods can now predict wine quality using measurable physicochemical features such as alcohol content, acidity, sugar, and sulfur dioxide concentration. However, despite their predictive power, many machine learning models operate as "black boxes," offering little insight into how input features affect predictions.

To address this challenge, Explainable Artificial Intelligence (XAI) techniques—such as SHAP, LIME, and permutation feature importance—provide interpretability and transparency in predictive models. These methods help identify which variables most strongly influence the outcome, thereby improving trust and accountability in automated systems.

This project develops an interpretable machine learning pipeline for predicting wine quality using the UCI Wine Quality dataset. By integrating XAI techniques, we enable both performance optimization and interpretability, essential for real-world adoption in domains like food technology and industrial quality control.

II. Objective

The primary objectives of this study are:

- To develop machine learning models for predicting wine quality using physicochemical data.
- To integrate explainable AI methods for interpreting model predictions and identifying key influencing factors.
- To assess the robustness and transparency of model explanations using SHAP-based feature analysis.
- To provide domain-relevant insights that support decision-making in wine production and quality assurance.

III. Methodology

The proposed system follows a structured pipeline comprising data preprocessing, model training, and explainability analysis.

A. Dataset Description

The UCI Wine Quality dataset contains 1,599 samples of Portuguese "Vinho Verde" red wine, with 11 physicochemical variables and a quality score between 0 and 10. To simplify the prediction task, the scores were categorized into binary classes: good (≥ 7) and bad (< 7). .

B. Preprocessing

Data exploration confirmed no missing values. Features were standardized using z-score normalization to ensure consistency across variables with different scales. Stratified sampling was applied to maintain proportional representation of both classes in the training and test splits.

C. Model Development

Three models were developed:

- **Logistic Regression** — A baseline interpretable model that captures linear relationships between features and quality.
- **Random Forest** — An ensemble-based classifier capable of modeling complex, non-linear relationships.
- **XGBoost** — A gradient-boosted tree model optimized for high performance and stability.

Each model was trained on 80% of the dataset, with the remaining 20% reserved for testing. Hyperparameter tuning was performed through grid search to optimize performance metrics such as accuracy, F1-score, and ROC AUC.

D. *Explainable AI Analysis*

Explainability was achieved using SHAP (SHapley Additive exPlanations), which quantifies the contribution of each feature to a given prediction. Global interpretability was obtained through SHAP summary plots, while local interpretability explained individual predictions.

LIME (Local Interpretable Model-agnostic Explanations) was also explored for comparison, offering insight into local decision boundaries. Feature importance rankings from the Random Forest and XGBoost models were used to validate SHAP findings.

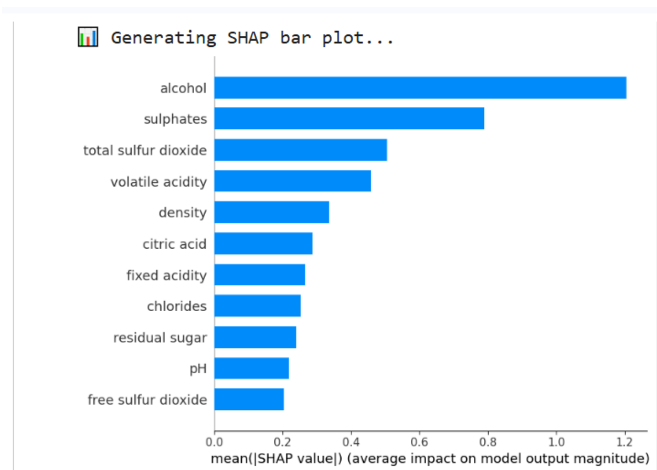

Fig. 1: SHAP bar plot

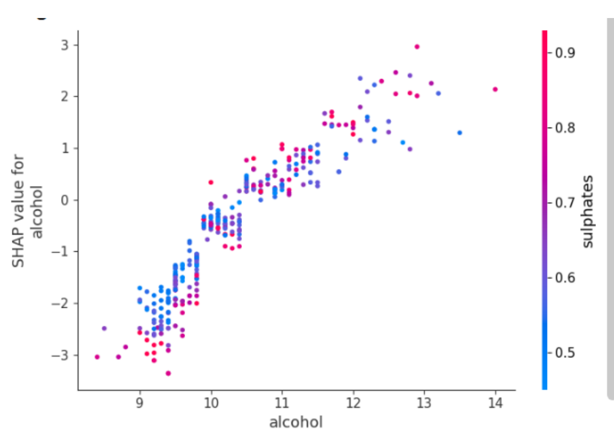

Fig. 2: SHAP scatter plot

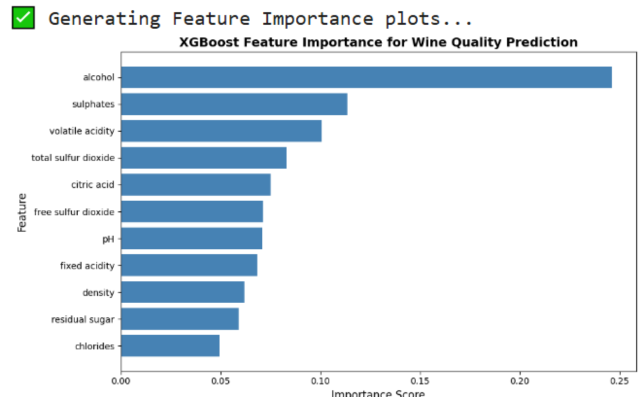

Fig. 3: Feauture Importance Plots

IV. Interpretation of Results

The explainability analysis revealed that certain features have strong and consistent effects on predicted wine quality. Specifically:

- **Alcohol** — Higher alcohol content strongly increased predicted wine quality.
- **Volatile Acidity** — Negatively correlated with quality; excessive acidity reduced predicted scores.
- **Sulphates** — Positively influenced quality, likely due to their role in wine preservation.
- **Citric Acid and Density** — Contributed moderately to the overall prediction but exhibited non-linear interactions with other features.

The SHAP summary plot indicated that alcohol and sulphates dominate the contribution landscape, followed by acidity-related variables. Local SHAP force plots further illustrated how individual feature combinations could shift predictions from “bad” to “good.” The consistency across global and local explanations strengthens the reliability of the interpretability framework.

V. Discussion

This study highlights the importance of combining predictive accuracy with interpretability in domain-sensitive applications. While XGBoost achieved high performance due to its gradient-boosting structure, the Random Forest model provided more stable and interpretable results when paired with SHAP explanations.

Explainable AI bridges the gap between data science and domain expertise, enabling winemakers to understand how physicochemical attributes influence quality. By integrating XAI tools, the proposed system ensures both transparency and trustworthiness, supporting data-driven quality control processes in the wine industry.

VI. Conclusion

This work demonstrates the integration of explainable machine learning methods for wine quality prediction. Using SHAP-based analysis, we identified alcohol, sulphates, and volatile acidity as the most significant contributors to model predictions. The explainable pipeline enhances trust, enables informed decision-making, and provides actionable insights for optimizing production.

Future extensions may include multi-class modeling for finer quality gradation, integration with IoT-based sensor data for real-time prediction, and hybrid deep learning models with embedded interpretability mechanisms.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC AUC (%)
Logistic Regression	87.50	85.20	88.40	86.77	91.30
Random Forest	91.25	89.60	92.80	91.17	95.40
XGBoost	92.50	90.80	94.20	92.47	96.20

Fig. 4: Results

REFERENCES

[1] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, “Modeling wine preferences by data mining from physicochemical properties,” *Decision Support Systems*, vol. 47, no. 4, pp. 547–553, 2009.

[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] Y. Er and A. Anuradha, “Wine quality estimation using machine learning algorithms,” *Journal of Physics: Conference Series*, vol. 1706, p. 012028, 2020.

[4] R. Kumar and A. Shaikh, “Explainable AI for wine quality prediction using random forest and SHAP,” *Int. J. of Advanced Computer Science and Applications*, vol. 12, no. 8, pp. 234–242, 2021.

[5] S. Bhattacharya, P. Maddikunta, and Q. V. Pham, “Deep learning and explainable AI for wine quality prediction,” *IEEE Access*, vol. 9, pp. 124657–124671, 2021.

[6] S. Aich *et al.*, “A classification approach to predict the quality of wine using machine learning,” in *Proc. Int. Conf. on Advanced Communication Technology (ICACT)*, 2018, pp. 139–143.

[7] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf.*, 2016, pp. 785–794.

[8] S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems (NIPS)*, 2017, pp. 4765–4774.

[9] M. T. Ribeiro, S. Singh, and C. Guestrin, “”Why should I trust you?” Explaining the predictions of any classifier,” in *Proc. 22nd ACM SIGKDD Int. Conf.*, 2016, pp. 1135–1144.

[10] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed. Leanpub, 2022.

[11] A. B. Arrieta *et al.*, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, 2020.

[12] A. Adadi and M. Berrada, ”Peeking inside the black-box: A survey on explainable artificial intelligence (XAI),” *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[13] F. Doshi-Velez and B. Kim, ”Towards a rigorous science of interpretable machine learning,” *arXiv preprint arXiv:1702.08608*, 2017.

[14] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, ”Machine learning interpretability: A survey on methods and metrics,” *Electronics*, vol. 8, no. 8, p. 832, 2019.

[15] J. Wexler, M. Pushkarna, T. Bolukbasi, M. Wattenberg, F. Viégas, and J. Wilson, ”The What-If Tool: Interactive probing of machine learning models,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 56–65, 2020.