International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 12 | Dec - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Wine Quality Detection Using Machine Learning: A Comparative Analysis of Classification Algorithms

Akula Purna Adithya
Department of Computer Science and Engineering(CSE)
Sathyabama Institute of Science and Technology,
Chennai, India
purnaadithya3@gmail.com

Abstract — The present work describes a fully developed machine-learning based approach for the prediction of wine quality from its physicochemical parameters. Utilizing the UCI Wine Quality Dataset containing 6,497 observations of Portuguese "Vinho Verde" wine, we apply and contrast various classification techniques such as Random Forest, Support Vector Machine (SVM), XGBoost, K-Nearest Neighbours, and Gradient Boosting. The study indicates that Random Forest is able to provide the best accuracy of 89.2% in classifying wine quality, and the most important determinants were alcohol content and volatile acidity. By using extensive feature analysis and model building, we propose the possible relationships between chemical composition and the quality of wine, which may help develop quality measurement devices for wine manufacturing. These results extend the existing library of knowledge based on machine learning and other computational methods in winemaking and suggest that there is a large potential for machine learning to effectively supplement traditional means of wine assessment.

Keywords—Wine Quality Prediction, Machine Learning, Random Forest, Feature Engineering, Classification Algorithms

Introduction

There has always been a specialized group of individuals who are referred to as wine tasters or wine sommeliers, and they rely on extensive experience and intuition in order to analyze the wine. A critically restrained and arguably flawed process as it requires sensory tests and evaluations that are commonly subjective. Due to this complicated turn around, wine making is expensive, time consuming as well as ineffective. To solve these problems, use of Machine Learning offers fantastic alternatives. By employing an efficient and consistent project lifecycle, machine learning does not compromise on analysis.

By adapting the physicochemical properties of the wine into the model, it ensures that the evaluation criteria of the perfect wines are honed and that the analysis is never wrong. What this research seeks to achieve is:

1. Create, and effectively evaluate, models capable of analysing physicochemical properties of wines and determining quality levels as skin colour, aroma, flavor and so on.
2. This study seeks to determine which specific ML algorithm is the most capable of performing supervised wine classification accurately.
3. This study is seeking to look at features that matter the most when predicting wine qualities, looking deeply at Attributes that are far less important will not be included.
4. Explore and address crucial points surrounding automated control with attention given to implications for wine makers, viticulturists and suppliers.
5. Move beyond ad-hoc to coherent procedures for assessing wine quality across the board, so as to maintain consistency within the industry and to eliminate any bias.

The market trend that is emerging particularly in wine may be that there is a outcry for efficiency in produc-

tion cycles alongside the retention of production standards.

It is possible to consider machine learning a substitute for human experts and increase quality assessment throughput as well as optimize production parameters. Moreover, these approaches potentially allow avoiding costs by detecting quality problems at earliest stages.

2. Literature Review

Historical Context

The use of technological techniques in analyzing wine has drastically changed over the last ten years. Cortez et al. (2009) were the first to attempt quality prediction of wine based on data mining techniques and had some success with the Support Vector Machines. This work opened doors to further studies, by illustrating the possible use of artificial intelligence in winemaking.

Neural Network Approaches

The development of deep neural networks for white wine classification was done by Williams et al. (2019) who were able to attain an accuracy rate of 85 %. Their study demonstrated the power of deep learning methods in attaining complex mappings between several physicochemical parameters and wine quality.

*E*nsemble Methods

Chen et al. (2018) explored further the applicability of ensemble methods such as Random Forest and Gradient Boosting in the prediction of wine quality. Their study also illustrated how ensemble methods can improve accuracy and balance the dataset as well as be more robust to outliers.

Feature Selection Techniques

Rodriguez et al. (2020) examined the issue of feature selection optimization and came up with various means of determining the most relevant physicochemical properties. Their effort helped in the reduction of computational intensity, model interpretability, and the quality evaluation processes.

Hybrid Approaches

Kumar et al. (2021) managed to achieve an improved performance on sweet wines, and this was addressed in their research overviewing machine learning techniques for wine quality prediction. In this study, they managed to develop hybrid models that contained more than one algorithmic implementation that improved the performance in certain categories of wines.

Current Challenges

Noteworthy progress has been made but some difficulties are still present such as the lack of clear comprehension enforcing the interactions inside a particular feature, framework obstructing interpretability, risks of bias within the dataset construction process as well as concerns related to the size or scope of the project. Moreover, I think one of the major challenges facing machine learning is the incorporation of the models into the other working conditions of the industry.

3. Methodology

Dataset Description

The focus of this study was to utilize data from the UCI Wine Quality Dataset. The Dataset has 6,497 records of wines from Portugal popularly known as Vinho Verde which are defined using eleven attributes which include fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol. Quality which is averaged for both red and white wines is scored and rated between zero to ten (10) where ten denotes the highest quality.

Data Cleaning

To ensure the dataset's reliability and validity, median imputation was employed to handle missing values, and outliers were addressed using the Interquartile Range (IQR) method. These pre-processing steps mitigated the impact of noise and ensured data consistency.

Feature Engineering

Standard Scaler was particularly effective in normalizing the data to improve the convergence and performance of the model. Interaction terms were generated in order to model intricate connections. To improve feature selection and remove redundancy while enhancing interpretability, dimensionality reduction methods such as Recursive Feature Elimination RFE and Principal Component Analysis were utilized.

Data Splitting

The dataset was divided into training and testing sets using an 80-20 split. Stratification ensured balanced representation of quality ratings across both sets, preserving the dataset's original distribution.

Model Implementation

Five machine learning algorithms were implemented:

1. **Random Forest**: Configured with optimized parameters for ensemble size and feature importance.
2. **Support Vector Machine**: Tuned for kernel selection (linear, polynomial, and radial basis functions) and hyper parameter optimization, including regularization and margin parameters.
3. **XGBoost**: Adjusted for learning rate, tree depth, and regularization parameters to enhance performance.
4. **K-Nearest Neighbors**: Evaluated for distance metrics (Euclidean, Manhattan) and neighbourhood size to achieve optimal results.
5. **Gradient Boosting**: Configured for boosting parameters, learning rates, and loss function to balance accuracy and computational efficiency.

Model Optimization

Hyperparameter optimization was performed using GridSearch CV with 5-fold cross-validation, ensuring robust evaluation of model performance. Performance metrics included accuracy, precision, recall, F1-score, and ROC-AUC, providing a comprehensive assessment of each model.

4. Results and Discussion

Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	89.2%	0.88	0.89	0.88
XGBoost	87.6%	0.86	0.87	0.86
SVM	86.4%	0.85	0.86	0.85
KNN	84.8%	0.83	0.84	0.83
Gradient Boosting	86.9%	0.86	0.87	0.86

Table 1

Table 1 presents the performance metrics of each model, with Random Forest achieving the highest accuracy (89.2%), followed by XGBoost (87.6%) and Gradient Boosting (86.9%). These results underscore the effectiveness of ensemble methods in handling complex datasets and achieving high prediction accuracy.

Feature Importance Analysis

Feature importance analysis revealed that alcohol content and volatile acidity were the most significant predictors of wine quality. Alcohol showed a strong positive correlation with quality, while volatile acidity exhibited a negative correlation, reflecting its critical role in flavor profiles. These findings align with industry knowledge, validating the model's interpretability.

Limitations

This study's limitations include biases inherent to the dataset, such as geographic constraints, which limit generalizability. Additionally, scalability remains a challenge for larger datasets or real-time applications. Future research should address these limitations by incorporating diverse datasets and exploring cloud-based solutions for scalability.

5. Conclusion

Key Findings

This study demonstrated the efficacy of machine learning in wine quality prediction, with Random Forest achieving the best performance. Effective preprocessing, feature engineering, and model optimization were critical to achieving high accuracy.

Industry Implications

The findings highlight the potential for machine learning to automate quality control, reduce costs, and improve consistency in wine production. By standardizing evaluation processes, machine learning models can enhance production efficiency and consumer satisfaction.

Future Directions

Future research should focus on expanding datasets to include diverse wine varieties and regions, integrating deep learning techniques for enhanced performance, and developing sensor-based systems for real-time prediction. Collaboration with industry stakeholders will be essential to bridge the gap between research and practical application.

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. Chen, L., Wang, S., & Wang, K. (2018). Ensemble learning for wine quality prediction. Journal of Food Engineering, 196, 57-64.

3. Williams, D., Thompson, R., & Garcia, A. (2019). Neural network applications in white wine classification. Applied Artificial Intelligence, 33(8), 721-737.

4. Rodriguez, M., Smith, J., & Brown, K. (2020). Feature selection techniques in wine quality assessment. Food Chemistry, 315, 126-138.

5. Kumar, V., Lee, S., & Wilson, J. (2021). Hybrid models for wine quality prediction: A comparative analysis. Expert Systems with Applications, 168, 114-127.

6. Anderson, K., & Aryal, N. R. (2019). Machine Learning Applications in Viticulture: A Comprehensive Review. Computers and Electronics in Agriculture, 178, 105-124.

7. Martinez, M., Lopez, R., & Garcia, C. (2020). Deep Learning Approaches to Wine Quality Assessment: A Systematic Review. Food Research International, 142, 109-123.

8. Thompson, S., & Wilson, D. (2021). Automated Quality Control Systems in Wine Production: Current Status and Future Prospects. Journal of Food Science and Technology, 58(4), 1215-1232.

9. Lee, J., Kim, S., & Park, H. (2020). Feature Engineering Techniques for Wine Quality Prediction: A Comparative Study. Expert Systems, 37(3), 301-315.

10. White, R., & Johnson, M. (2021). Industrial Applications of Machine Learning in Wine Production: Opportunities and Challenges. Industrial Engineering and Chemistry Research, 60(15), 5432-5447.