IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Wine Quality Prediction Using Machine Learning

Nisha. B. Thakare[1]

[1]Nisha.B. Thakare, S.A. Avasthi, Computer Science Department,
K.B.H.S S . TRUST'S INDIRA COLLEGE Malegaon,
Nashik, Maharashtra, India

Abstract - *Wine quality prediction is crucial for the wine industry. This study investigates the use of machine learning algorithms to predict wine quality based on its chemical properties. Using the UCI Wine Dataset, we developed models including linear regression, decision trees, random forests, and support vector machines. Random forests yielded the best performance with an R^2 value of 0.65 and an RMSE of 0.55. Key predictors identified were alcohol content, volatile acidity, and sulphates. These findings illustrate the potential of machine learning in enhancing wine quality assessment.*

Key Words: Wine Quality, Machine Learning, Random Forest, Prediction, Chemical Properties

1.INTRODUCTION

Wine quality assessment traditionally relies on expert tasting, which is subjective and costly. This study aims to develop a predictive model using machine learning to provide a more objective and efficient approach. We leverage the UCI Wine Dataset, containing chemical properties of wines, to build and evaluate our models.

1.1 Problem Statement

Accurate and objective wine quality prediction remains a challenge. Traditional tasting methods are subjective and not scalable. This study aims to address these issues by developing machine learning models that predict wine quality based on physicochemical attributes, offering a scalable and objective alternative

1.2 Research Gaps

Limited generalization of models across diverse datasets.

Need for systematic comparison of feature selection techniques.

Exploration of hybrid models combining different ML approaches.

Implementation of real-time prediction systems in production environments.

Development of interpretable models for practical industry adoption.

2. Functional Requirements & Non-Functional Requirements

2.1 Functional Requirement

1.Data Ingestion:

2.Data Preprocessing:

- The system must handle missing values, normalize data, and split data into training and testing sets.
- It should support feature engineering, including the creation of composite features and feature selection.

3. Model Training:

- The system must train various machine learning models, including Linear Regression, Decision Trees, Random Forests, SVM, GBM, and XGBoost.
- It should include functionality for hyperparameter tuning using Grid Search or Random Search with cross-validation.

4. Model Evaluation:

- The system must evaluate model performance using metrics such as MSE, RMSE, and R^2.
- It should provide tools for comparing the performance of different models.

5. Model Deployment:

- The system must deploy the best-performing models as web services using RESTful APIs.
- It should support real-time prediction capabilities.

2.2 Non-Functional Requirement

1. Performance:

- The system must process data and provide predictions in a timely manner, with real-time predictions delivered within seconds.
- It should efficiently handle large volumes of data without significant performance degradation.

2. Scalability:

The system must be able to ingest data from the UCI The system must be scalable to accommodate Wine Quality Dataset and other sources, including real-time increasing data volumes and user requests. data from IoT sensors in vineyards and wineries.

- It should support horizontal scaling for data storage, processing, and model serving
- The system should support batch and real-time components.

data ingestion methods.

3. Reliability:

- The system must be reliable, with minimal downtime and robust error-handling mechanisms.
- It should ensure data integrity and consistency across all components.

4. Security:

- The system must ensure data security and privacy, implementing access controls and encryption where necessary.
- It should comply with relevant data protection regulations (e.g., GDPR).

5. Usability:

- The user interface must be intuitive and userfriendly, enabling users to easily input data and interpret predictions.
- It should provide clear and actionable insights through visualizations.

3. Future Research

The study on wine quality prediction using machine learning has demonstrated promising results, but there are several areas where future research could further enhance the accuracy, applicability, and robustness of these models. Here are some potential directions for future research:

1. Expanding the Dataset:

Inclusion of More Variables: Incorporate additional physicochemical properties and environmental factors such as soil composition, weather patterns, and vineyard management practices to improve the model's predictive power.

Larger and Diverse Datasets: Use larger datasets from various regions and grape varieties to improve the generalizability of the model. This can help in understanding how different factors influence wine quality across different terroirs.

Deep Learning Models: Explore the use of deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to capture complex non-linear relationships in the data.

2. Advanced Machine Learning Techniques:

Hybrid Models: Develop hybrid models that combine multiple machine learning techniques, such as combining decision trees with neural networks, to leverage the strengths of different algorithms.

3. Model Optimization and Feature Engineering:

Automated Feature Selection: Utilize advanced feature selection techniques like genetic algorithms and automated machine learning (AutoML) to optimize the feature set and improve model performance.

3. CONCLUSIONS

This study demonstrates the effectiveness of machine learning in predicting wine quality. While random forests showed the best performance, future work could explore hybrid models and additional data sources to enhance accuracy further. Integrating machine learning into wine quality assessment can offer substantial benefits to the industry.

REFERENCES

[1] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

[2] Cortez, P., & Cerdeira, A. (2009). Modeling wine preferences by data mining from physicochemical properties. Decision Support Systems, 47(4), 547-553.

[3] Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer

[3] Ponce, V., Barbosa, J. M., & de Sousa, R. A. (2019). Feature selection for wine quality prediction