International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 3, May-June 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Machine Learning Approach for Wine Quality Assessment

Mrs. Apeksha Nayak*, Harsh Tyagi**, Priyanshu Sharma***, Prashant Tyagi****
*(Ass. Prof., Department of Computer Science and Engineering, Shri Ram Group Of Colleges, Dr. A.P.J. Abdul Kalam Technical University
Email: Er.Apeksha2015@gmail.com)
**(Department of Computer Science and Engineering, Shri Ram Group Of Colleges, Dr. A.P.J. Abdul Kalam Technical University
Email: tyagivatsharsh1@gmail.com)
***(Department of Computer Science and Engineering, Shri Ram Group Of Colleges, Dr. A.P.J. Abdul Kalam Technical University
Email: priyanshusharma4301@gmail.com)
****(Department of Computer Science and Engineering, Shri Ram Group Of Colleges, Dr. A.P.J. Abdul Kalam Technical University
Email: prashanttyagi460@gmail.com)

Abstract:

This research presents a machine learning-based approach to predict the quality of red wine using its physicochemical properties. Traditional wine evaluation methods rely on expert tasters, which are often subjective, time-consuming, and inconsistent. To overcome these limitations, this study utilizes a publicly available red wine dataset and applies the Gradient Boosting Regressor algorithm for predictive modelling. The dataset includes key features such as alcohol content, volatile acidity, sulphates, and citric acid, all of which influence wine quality.

The data underwent preprocessing including normalization and correlation analysis to improve model accuracy. After training and evaluation, the Gradient Boosting model achieved an R^2 score of 0.79 and a Root Mean Squared Error (RMSE) of 0.48, indicating strong predictive performance. The model was further deployed into a user-friendly web application using Streamlit, allowing real-time input and prediction through an interactive interface.

This study demonstrates the potential of machine learning in automating quality control in the food and beverage industry. It also highlights how user-friendly deployment platforms like Streamlit can bridge the gap between technical machine learning models and practical end-user applications.

Keywords — Machine Learning, Wine Quality, Gradient Boosting, Regression, Streamlit, Data Science.

I. INTRODUCTION

Wine quality plays a vital role in the global beverage industry, influencing both consumer satisfaction and market value. Traditionally, wine quality is evaluated through human sensory analysis by trained professionals. While this method is insightful, it is inherently subjective, time- consuming, and prone to inconsistency. As the demand for scalable and automated quality control systems grows, machine learning has emerged as a reliable alternative to manual testing.

Machine learning techniques, particularly regression models, can analyze large datasets and uncover complex patterns that influence wine quality. These techniques enable objective predictions based on measurable physicochemical parameters such as alcohol content, pH, acidity, sulphates, and residual sugar. Such predictive capabilities allow producers and quality control experts to streamline evaluation processes and maintain product consistency.

II. RELATED WORK

In recent years, the application of machine learning in food and beverage quality assessment has gained substantial attention. Several studies have focused on predicting wine quality based on physicochemical properties, exploring various supervised learning algorithms to achieve accuratepredictions.

Cortez et al. (2009) used data from the Vinho Verde project and concluded that ensemble methods like Random Forest and Gradient Boosting outperformed traditional linear regression techniques. González Viejo et al. (2017) used Artificial Neural Networks, emphasizing alcohol and volatile acidity as important features. Pereira et al. (2020) compared classifiers and found Gradient Boosting to be the most accurate. Kotsiantis et al. (2013) highlighted the interpretability of decision trees.

III. METHODOLOGY/PROPOSED SYSTEM

This section outlines the dataset, preprocessing, model selection, and deployment.

A. Dataset Description

The dataset used is the Red Wine Quality Dataset from the UCI Repository, comprising 1,599 samples with 11 attributes including: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol. The target variable is quality, rated from 0 to 10.

B. Data Preprocessing

Preprocessing involved:

- No missing values
- Normalization using Standard Scaler
- Correlation analysis to find key features
- Train-test split using stratified sampling (80/20)

C. Model Selection and Training

Various regression algorithms were evaluated, and Gradient Boosting Regressor was selected for its accuracy and robustness. It is an ensemble method that builds weak learners sequentially, optimizing performance through error correction.

D. System Architecture and Deployment The system includes:

- Backend: Gradient Boosting model returning predicted quality
- Frontend: Streamlit interface for real- time prediction
- Data Layer: CSV data and pickled model files.

The model was serialized and deployed using Streamlit with input sliders, prediction output, and visualizations.

IV. RESULTS AND EVALUATION

Evaluation metrics on test data include:

- MAE: 0.34
- MSE: 0.23
- RMSE: 0.48
- R^2 Score: 0.79

These metrics indicate the model is accurate and reliable in predicting wine quality.

V. DISCUSSION

The model achieved a strong R^2 score of 0.79 with low RMSE, confirming good prediction accuracy. Feature importance analysis aligned with domain knowledge, highlighting alcohol and sulphates as key indicators of quality.

VI. CONCLUSION

This research demonstrates that Gradient Boosting Regressor can predict red wine quality effectively based on physicochemical attributes. The Streamlit-based deployment enables practical use by non-technical users, offering a scalable and objective solution to traditional wine assessment methods.

VII. FUTURE SCOPE

Future improvements include:

- Inclusion of white wine and beer datasets
- Cloud-based deployment
- Integration with IoT sensors
- Mobile app interface
- Use of deep learning models

- Inclusion of sensory data (e.g., aroma)

ACKNOWLEDGMENT

We express our gratitude to Mrs. Apeksha Nayak, our project supervisor, for her guidance and support throughout this study. We also thank our peers and the UCI Machine Learning Repository for resources and insights.

REFERENCES

[1]P. Cortez et al., 'Modeling wine preferences by data mining from physicochemical properties,' Decision Support Systems, vol. 47, no. 4, pp. 547–553, Oct. 2009.
[2]A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed., O'Reilly Media, 2019.
[3]Streamlit, 'Streamlit Documentation.' [Online].Available:https://docs.streamlit.io
[4]Scikit-learn, 'Scikit-learn Documentation.[Online]. Available: https://scikit-learn.org
[5]UCI Machine Learning Repository, 'Wine Quality Dataset.' [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Wine+Qualit

