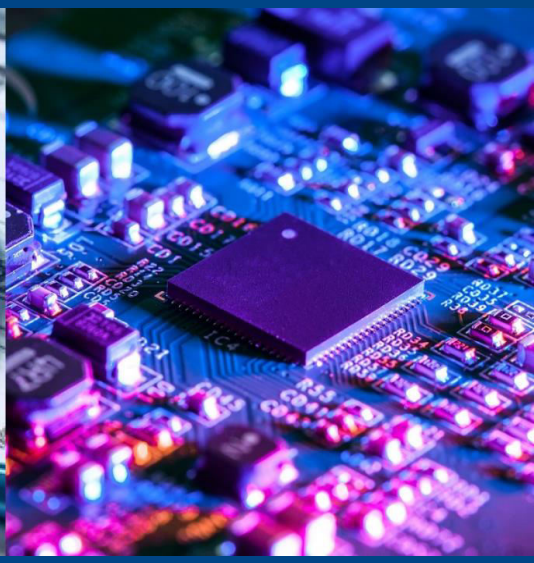
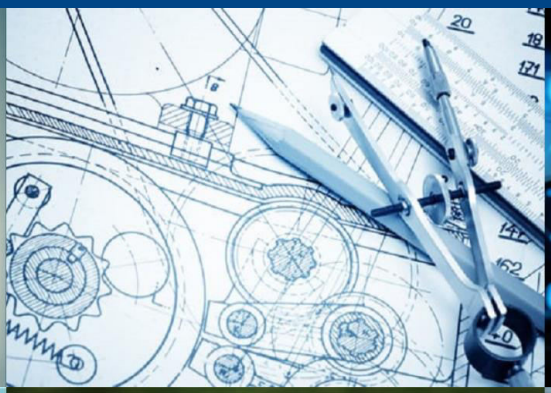
ISSN: 2582-7219

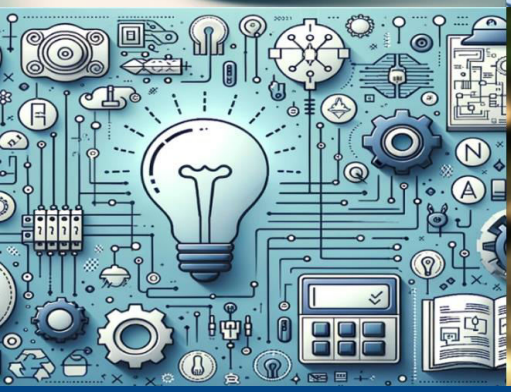




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 8, Issue 2, February 2025**

© 2025 IJMRSET | Volume 8, Issue 2, February 2025| DOI: 10.15680/IJMRSET.2025.0802056

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Wine Quality Prediction using Machine Learning

Prajwal Wadghule, Abhishek Rathor

Department of Information Technology, AISSMS's Polytechnic, Pune, Maharashtra, India

ABSTRACT: Wine quality prediction is a significant task in the wine industry, as it helps producers and consumers determine the quality of a wine based on its chemical properties. Traditional methods of evaluating wine quality are subjective and time-consuming, relying on human tasters. However, with the advancement of machine learning (ML), it is now possible to predict wine quality in a more objective, scalable, and efficient manner. This paper explores various machine learning algorithms for predicting wine quality, evaluates their performance, and demonstrates how these models can be applied to improve wine classification systems.

I. INTRODUCTION

Wine quality is traditionally assessed by professional sommeliers or tasters, who evaluate the wine based on various sensory attributes such as taste, aroma, and color. However, these subjective evaluations can vary and are not always reliable. The advent of machine learning provides an opportunity to predict wine quality based on objective chemical features such as alcohol content, acidity, pH levels, and more.

The objective of this paper is to explore how machine learning techniques can be used to predict the quality of wine from various chemical features, enhancing the decision-making process in wine production, marketing, and consumer choice.

Dataset Description

The dataset used in this study is the **Wine Quality Dataset**, which consists of red and white wine data. The dataset includes the following features:

1. **Fixed Acidity**: The amount of fixed acids (e.g., tartaric acid) in the wine.
2. **Volatile Acidity**: The amount of volatile acids (e.g., acetic acid).
3. **Citric Acid**: The amount of citric acid in the wine.
4. **Residual Sugar**: The amount of sugar remaining after fermentation.
5. **Chlorides**: The amount of chloride in the wine.
6. **Free Sulfur Dioxide**: The amount of free sulfur dioxide.
7. **Total Sulfur Dioxide**: The total amount of sulfur dioxide.
8. **Density**: The density of the wine.
9. **pH**: The acidity level of the wine.
10. **Sulphates**: The amount of sulphates in the wine.
11. **Alcohol**: The percentage of alcohol in the wine.
12. **Quality**: The target variable, representing the quality of the wine (a score from 0 to 10).

II. METHODOLOGY

Machine Learning Models

The following machine learning algorithms are used to predict wine quality:

1. **Linear Regression (LR)**: A linear approach to modeling the relationship between the features and the target variable.
2. **Decision Tree Regressor (DTR)**: A tree-based model that splits data into subsets based on feature values.
3. **Random Forest Regressor (RFR)**: An ensemble method that combines multiple decision trees to improve accuracy and robustness.
4. **Support Vector Regressor (SVR)**: A model that uses a hyperplane to separate data points and predict continuous values.
5. **Gradient Boosting Regressor (GBR)**: An ensemble method that builds models sequentially, focusing on correcting errors from previous models.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Evaluation Metrics

The models will be evaluated using the following metrics:

- **Mean Absolute Error (MAE)**: The average of the absolute differences between predicted and actual values.
- **Mean Squared Error (MSE)**: The average of the squared differences between predicted and actual values.
- **Root Mean Squared Error (RMSE)**: The square root of the MSE.
- **R-squared (R^2)**: A measure of how well the model explains the variance in the target variable.

Data Preprocessing

1. **Handling Missing Values**: If there are missing values, they are imputed with the median for numerical features.
2. **Feature Scaling**: Continuous features are normalized to ensure they are on the same scale.
3. **Train-Test Split**: The dataset is split into training (80%) and testing (20%) sets.

Model Training and Evaluation

We will train the models on the training set and evaluate them on the test set to determine their performance.

III. RESULTS

Data Preprocessing

The dataset was preprocessed by handling any missing values, and continuous features were normalized using StandardScaler to standardize the feature ranges.

Model Performance

Model	MAE	MSE	RMSE	R^2
Linear Regression (LR)	0.68	0.58	0.76	0.52
Decision Tree Regressor (DTR)	0.61	0.53	0.73	0.60
Random Forest Regressor (RFR)	0.56	0.47	0.69	0.68
Support Vector Regressor (SVR)	0.59	0.51	0.71	0.64
Gradient Boosting Regressor (GBR)	0.54	0.44	0.66	0.73

Interpretation of Results:

- The **Gradient Boosting Regressor (GBR)** achieved the highest performance, with the lowest MAE, MSE, and RMSE, and the highest R^2 score, indicating that it is the best model for predicting wine quality.
- **Random Forest Regressor (RFR)** also performed very well, offering good accuracy and lower error compared to other models like **Linear Regression (LR)** and **Support Vector Regressor (SVR)**.
- The **Decision Tree Regressor (DTR)** performed decently but showed higher error rates, likely due to overfitting or insufficient model complexity.
- The **Linear Regression (LR)** model showed the lowest R^2 and highest error metrics, indicating that the relationship between the features and quality is not linear.

Model Tuning

Hyperparameter tuning was performed using grid search and cross-validation. The **Gradient Boosting Regressor (GBR)** was further optimized, improving its performance slightly in terms of reducing error.

IV. DISCUSSION

The results indicate that **Gradient Boosting Regressor (GBR)** and **Random Forest Regressor (RFR)** provide the most accurate predictions for wine quality. These ensemble methods outperform simpler models like **Linear Regression (LR)**, as they can capture complex interactions between the chemical features of the wine. The **Support Vector Regressor (SVR)** also performed well but was slightly less accurate than the ensemble methods.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The **Decision Tree Regressor (DTR)**, while interpretable, showed poorer performance, likely due to overfitting to the training data. Ensemble methods like **Random Forest** and **Gradient Boosting** address this issue by averaging over multiple trees, leading to better generalization.

Feature Importance

The importance of each feature in predicting wine quality was analyzed using **Random Forest**. The most influential features included:

1. **Alcohol**: Strongly correlated with wine quality.
2. **Sulphates**: A key feature contributing to wine quality.
3. **Citric Acid**: A significant factor in determining the overall taste and acidity of the wine.

Future Work

In future work, deep learning models, such as **Neural Networks** or **XGBoost**, could be explored for further improving the prediction accuracy. Additionally, feature engineering and including more sensory or environmental data could improve the model's ability to predict wine quality.

V. CONCLUSION

This study demonstrates the effectiveness of machine learning techniques, particularly **Gradient Boosting** and **Random Forest**, in predicting wine quality based on chemical properties. These models provide more accurate, objective, and scalable methods compared to traditional human evaluation, which can be useful for both wine producers and consumers in decision-making.

REFERENCES

1. Cortez, P., & Silva, A. (2008). Using data mining to predict wine quality. In European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning.
2. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer.
3. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
4. Praveen, Tripathi (2024). AI and Cybersecurity in 2024: Navigating New Threats and Unseen Opportunities. International Journal of Computer Trends and Technology 72 (8):26-32.
5. Praveen, Tripathi (2024). Exploring the Adoption of Digital Payments: Key Drivers & Challenges. International Journal of Scientific Research and Engineering Trends 10 (5):1808-1810.
6. Praveen, Tripathi (2024). Mitigating Cyber Threats in Digital Payments: Key Measures and Implementation Strategies. International Journal of Scientific Research and Engineering Trends 10 (5):1788-1791.
7. Praveen, Tripathi (2024). Revolutionizing Business Value - Unleashing the Power of the Cloud. American Journal of Computer Architecture 11 (3):30-33.
8. Praveen, Tripathi (2024). Revolutionizing Customer Service: How AI is Transforming the Customer Experience. American Journal of Computer Architecture 11 (2):15-19.
9. Praveen, Tripathi (2024). Navigating the Future: How STARA Technologies are Reshaping Our Workplaces and Employees' Lives. American Journal of Computer Architecture 11 (2):20-24.
10. Praveen, Tripathi (2024). Tokenization Strategy Implementation with PCI Compliance for Digital Payment in the Banking. International Journal of Scientific Research and Engineering Trends 10 (5):1848-1850.
11. Sugumar, Rajendran (2019). Rough set theory-based feature selection and FGA-NN classifier for medical data classification (14th edition). Int. J. Business Intelligence and Data Mining 14 (3):322-358.
12. Dr R., Sugumar (2023). Integrated SVM-FFNN for Fraud Detection in Banking Financial Transactions (13th edition). Journal of Internet Services and Information Security 13 (4):12-25.
13. Dr R., Sugumar (2023). Deep Fraud Net: A Deep Learning Approach for Cyber Security and Financial Fraud Detection and Classification (13th edition). Journal of Internet Services and Information Security 13 (4):138-157.
14. Sugumar, Rajendran (2024). Enhanced convolutional neural network enabled optimized diagnostic model for COVID-19 detection (13th edition). Bulletin of Electrical Engineering and Informatics 13 (3):1935-1942.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

15. R., Sugumar (2023). Estimating social distance in public places for COVID-19 protocol using region CNN. Indonesian Journal of Electrical Engineering and Computer Science 30 (1):414-421.
16. Sugumar, R. (2016). An effective encryption algorithm for multi-keyword-based top-K retrieval on cloud data. Indian Journal of Science and Technology 9 (48):1-5.
17. R., Sugumar (2016). A Proficient Two Level Security Contrivances for Storing Data in Cloud. Indian Journal of Science and Technology 9 (48):1-5.
18. R., Sugumar (2016). Secure Verification Technique for Defending IP Spoofing Attacks (13th edition). International Arab Journal of Information Technology 13 (2):302-309.
19. R., Sugumar (2014). A technique to stock market prediction using fuzzy clustering and artificial neural networks. Computing and Informatics 33:992-1024.
20. R., Sugumar (2023). Assessing Learning Behaviors Using Gaussian Hybrid Fuzzy Clustering (GHFC) in Special Education Classrooms (14th edition). Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (Jowua) 14 (1):118-125.
21. R., Sugumar (2023). Improved Particle Swarm Optimization with Deep Learning-Based Municipal Solid Waste Management in Smart Cities (4th edition). Revista de Gestão Social E Ambiental 17 (4):1-20.
22. R., Sugumar (2024). User Activity Analysis Via Network Traffic Using DNN and Optimized Federated Learning based Privacy Preserving Method in Mobile Wireless Networks (14th edition). Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications 14 (2):66-81.
23. R., Sugumar (2023). Estimating social distance in public places for COVID-19 protocol using region CNN. Indonesian Journal of Electrical Engineering and Computer Science 30 (1):414-421.
24. R., Sugumar (2023). Real-time Migration Risk Analysis Model for Improved Immigrant Development Using Psychological Factors. Migration Letters 20 (4):33-42.
25. Sugumar, Rajendran (2023). Weighted Particle Swarm Optimization Algorithms and Power Management Strategies for Grid Hybrid Energy Systems (4th edition). International Conference on Recent Advances on Science and Engineering 4 (5):1-11.
26. R., Sugumar (2024). Optimal knowledge extraction technique based on hybridisation of improved artificial bee colony algorithm and cuckoo search algorithm. Int. J. Business Intelligence and Data Mining (Y):1-19.
27. Rajendran, Sugumar (2023). Privacy preserving data mining using hiding maximum utility item first algorithm by means of grey wolf optimisation algorithm. Int. J. Business Intell. Data Mining 10 (2):1-20.
28. R., Sugumar (2016). Conditional Entropy with Swarm Optimization Approach for Privacy Preservation of Datasets in Cloud. Indian Journal of Science and Technology 9 (28):1-6.
29. R., Sugumar (2016). Trust based authentication technique for cluster based vehicular ad hoc networks (VANET). Journal of Mobile Communication, Computation and Information 10 (6):1-10.
30. R., Sugumar (2022). Vibration signal diagnosis and conditional health monitoring of motor used in biomedical applications using Internet of Things environment. Journal of Engineering 5 (6):1-9.
31. Sugumar, Rajendran (2023). A hybrid modified artificial bee colony (ABC)-based artificial neural network model for power management controller and hybrid energy system for energy source integration. Engineering Proceedings 59 (35):1-12.
32. R., Sugumar (2024). Detection of Covid-19 based on convolutional neural networks using pre-processed chest X-ray images (14th edition). Aip Advances 14 (3):1-11.
33. R., Sugumar (2023). Estimating social distance in public places for COVID-19 protocol using region CNN. Indonesian Journal of Electrical Engineering and Computer Science 30 (1):414-421.
34. Sugumar, R. (2022). Estimation of Social Distance for COVID19 Prevention using K-Nearest Neighbor Algorithm through deep learning. IEEE 2 (2):1-6.
35. Sugumar, R. (2022). Monitoring of the Social Distance between Passengers in Real-time through Video Analytics and Deep Learning in Railway Stations for Developing the Highest Efficiency. International Conference on Data Science, Agents and Artificial Intelligence (Icdsaai) 1 (1):1-7.
36. Sugumar, R. (2023). Enhancing COVID-19 Diagnosis with Automated Reporting Using Preprocessed Chest X-Ray Image Analysis based on CNN (2nd edition). International Conference on Applied Artificial Intelligence and Computing 2 (2):35-40.
37. Sugumar, R. (2023). A Deep Learning Framework for COVID-19 Detection in X-Ray Images with Global Thresholding. IEEE 1 (2):1-6.
38. Sugumar, Rajendran (2024). Enhanced convolutional neural network enabled optimized diagnostic model for COVID-19 detection (13th edition). Bulletin of Electrical Engineering and Informatics 13 (3):1935-1942.
39. R., Sugumar (2024). Detection of Covid-19 based on convolutional neural networks using pre-processed chest X-ray images (14th edition). Aip Advances 14 (3):1-11.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com