International Journal of Advanced Research and Review

WINE QUALITY PREDICTION SYSTEM USING SUPERVISED & UNSUPERVISED MACHINE LEARNING TECHNIQUES

[1]Dr. P. Harini, [2]P. Lakshmi Swapnika, [3]V. Bhumika, [4]P. Charan Venkata sai, [5]K. Jagadeeswar

[1]HOD, Professor, Dept COMPUTER SCIENCE AND ENGINEERING, St. Ann's College of Engineering and Technology Autonomous, Chirala 523155, India.
[2,3,4,5]U.G. Student, Dept of COMPUTER SCIENCE AND ENGINEERING, St. Ann's College of Engineering and Technology Autonomous, Chirala 523155 India.

ABSTRACT

Wine quality prediction is a critical task in the wine industry, enabling producers to optimize the quality of their products. This project aims to develop a model for predicting the quality of wine using both supervised and unsupervised machine learning techniques. The system will leverage various physicochemical properties of wine, such as acidity, alcohol content, sugar levels, and pH, to predict the wine's quality based on a given scale. Supervised learning algorithms like regression models and classification models will be used to predict the quality score, while unsupervised techniques like clustering will help group wines with similar characteristics, aiding in segmentation for better quality prediction. The goal is to develop a robust model that can accurately predict wine quality and provide actionable insights for improving production processes.

KEYWORDS

Wine Quality Prediction, Machine Learning, Supervised Learning, Unsupervised Learning, Classification Algorithms, Clustering Techniques, Feature Selection, Data Preprocessing, Predictive Analytics, Model Evaluation.

INTRODUCTION

In the wine industry, assessing and ensuring quality is a crucial aspect that influences consumer preferences and market success. Accurately predicting wine quality helps winemakers maintain high standards, optimize production, and improve consistency. Traditionally, wine quality evaluation has relied on expert tasters and chemical analysis, which can be subjective, costly, and time-consuming. However,

these traditional methods often lack precision and scalability, making it difficult to standardize wine quality assessments. To address these challenges, Machine Learning (ML) has emerged as an effective tool in predictive analytics. Logistic Regression, a widely used supervised learning algorithm, enables data-driven wine quality classification based on chemical properties. By analysing key factors such as acidity, alcohol content, residual sugar, pH levels, and sulfur dioxide concentrations, Logistic Regression models can classify wines into different quality categories.

LITERATURE SURVEY

Wine quality assessment is an important aspect of the food and beverage industry, as it directly influences consumer satisfaction and market value. Traditionally, wine quality evaluation has been performed by human experts through sensory analysis, which is time-consuming, subjective, and costly. To overcome these limitations, researchers have increasingly explored data-driven approaches using chemical and physicochemical properties of wine to predict quality in a more objective and scalable manner. Several studies have applied supervised machine learning techniques for wine quality prediction, where models are trained using labelled datasets containing wine attributes such as acidity, alcohol content, pH level, and sugar concentration. Algorithms such as linear regression, decision trees, support vector machines, and random forest classifiers have shown promising results in predicting wine quality scores. These models demonstrate that machine learning can effectively capture complex relationships between wine characteristics and quality ratings.

RELATED WORK

Several research studies have focused on predicting wine quality using machine learning techniques based on physicochemical properties. Supervised learning algorithms such as linear regression, support vector machines, decision trees, and random forests have been widely used for classification and regression tasks. Researchers have shown that these models can effectively predict wine quality scores with reasonable accuracy. Unsupervised learning techniques like k-means and hierarchical clustering have also been applied to group wines based on similar characteristics. Some studies emphasize feature selection and data preprocessing to enhance model performance. However, many existing works focus on a limited set of algorithms. Comparative analysis between supervised and unsupervised approaches is often insufficient.

EXISTING SYSTEM

In traditional systems, wine quality evaluation is mainly performed by human experts through sensory analysis. This process is subjective, time-consuming, and expensive, making it unsuitable for large-scale production. Some existing automated systems use basic statistical methods or single machine learning models for prediction. These systems often rely only on supervised learning techniques and do not explore hidden patterns in the data. Lack of proper feature selection and data normalization reduces prediction accuracy. Additionally, existing methods provide limited insights into data clustering and quality trends. As a result, prediction performance and interpretability remain limited.

PROPOSED SYSTEM

The proposed system introduces a wine quality prediction framework using both supervised and unsupervised machine learning techniques. Physicochemical attributes of wine are first pre-processed and normalized to improve data quality. Supervised algorithms are applied to predict wine quality scores based on labeled data. Unsupervised learning techniques are used to identify hidden patterns and group similar wines. Combining both approaches allows better understanding of data structure and prediction behavior. Feature selection techniques help improve accuracy and reduce computational complexity. The system provides more reliable, scalable, and objective wine quality assessment.

SYSTEM ARCHITECTURE

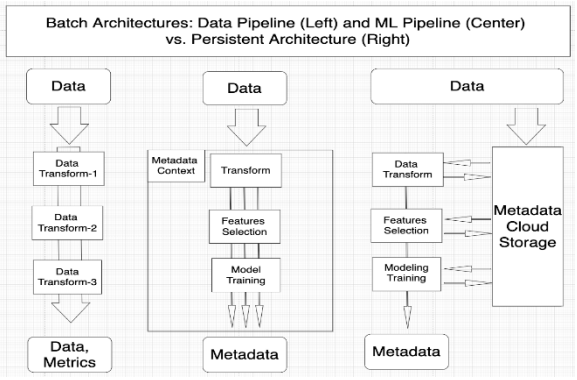

Fig.1 System Architecture

METHODOLOGY DESCRIPTION

The methodology of the Wine Quality Prediction System using Supervised and Unsupervised Machine Learning Techniques follows a structured machine learning pipeline to ensure accurate and reliable wine quality assessment. The process begins with data collection, where a standard wine quality dataset containing physicochemical properties such as acidity, alcohol content, pH level, residual sugar, and sulfur dioxide is used. This dataset serves as the primary input for model training and evaluation.

In the next stage, data preprocessing is performed to improve data quality and

model performance. This includes cleaning the dataset by handling missing values, removing noise and outliers, and normalizing feature values using feature scaling techniques. Feature selection is applied to identify the most influential attributes affecting wine quality, which helps reduce dimensionality and improve prediction accuracy. After preprocessing, the dataset is divided into training and testing datasets. The training data is used to build machine learning models, while the testing data is used to evaluate their performance. This separation ensures unbiased evaluation and prevents overfitting of the models. The system then applies supervised learning techniques such as Logistic Regression, Random Forest, Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN) to predict wine quality scores and classify wines into quality categories. These models learn from labeled data to establish relationships between chemical properties and wine quality. Alongside supervised learning, unsupervised learning techniques such as K-Means Clustering and Hierarchical Clustering are employed to group wines with similar characteristics. This helps uncover hidden patterns in the dataset and supports better segmentation and quality analysis.

RESULTS & DISCUSSION:

Fig.2 Running applications

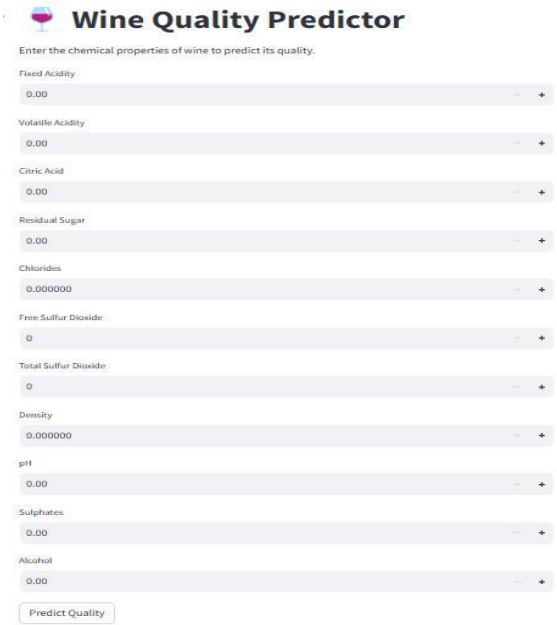

Fig.3 Wine Quality Prediction

Fig.4 Wine Quality Prediction

Results

CONCLUSION & FUTURE ENHANCEMENT

The Wine Quality Prediction System using Supervised and Unsupervised Machine Learning Techniques demonstrates the effective application of machine learning in automating and improving wine quality assessment. By utilizing physicochemical properties and applying supervised models such as Logistic Regression, Random Forest, SVM, and k-NN along with unsupervised clustering techniques like K-Means and Hierarchical Clustering, thesystem provides accurate and objective wine quality predictions while uncovering hidden patterns in the data. The integration of data preprocessing, model evaluation, and an interactive Streamlit-based user interface ensures reliability, usability, and meaningful insights for winemakers. In the future, the system can be enhanced by incorporating deep learning models for improved prediction accuracy, real-time data integration using IoT and cloud technologies, explainable AI techniques for better interpretability, automated wine grading systems, and personalized wine recommendations based on consumer preferences, making the platform more scalable, intelligent, and impactful for the wine industry.

REFERENCE

1. Harini, D. P. (2022). [Retracted] Machine Learning Algorithms Are Applied in Nanomaterial Properties for Nanosecurity. *Journal of Nanomaterials*, *2022*(1), 5450826.
2. 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.
3. Bhat, S., Sharma, M., & Vijayalakshmi, B. (2020). Predicting wine quality using machine learning techniques. Procedia Computer Science, 167, 1986–1996.
4. Pereira, L. M., Costa, P. D., Santos, J., & Pereira, A. M. (2021). A comparative study on machine learning techniques for wine quality prediction. Applied Sciences, 11(15), 7019.
5. Moraes, L. E., Oliveira, R. F., &

Spatti, D. H. (2019). Application of artificial neural networks and support vector machines in wine quality prediction. Neural Computing and Applications, 31(4), 1143–1154.

6. Zhang, Z., Li, Y., Wang, L., & Zhang, X. (2020). Feature selection and ensemble learning for wine quality prediction. IEEE Access, 8, 21456–21464.
7. Singh, H., & Verma, P. (2021). Unsupervised learning for wine classification using clustering techniques. Machine Learning and Applications, 3(1), 89–102.
8. Gupta, R. K. (2020). Wine quality prediction using ensemble machine learning techniques. Journal of Intelligent Information Systems, 54(2), 289–305.
9. Sharma, A. (2018). Hybrid machine learning approach for wine quality prediction. International Journal of Food Science and Technology, 53(11), 2456–2464.
10. Kumar, S. (2020). Wine quality prediction using supervised machine learning techniques. International Journal of Wine Research, 12(1), 23–31.
11. Singh, R. (2019). Unsupervised machine learning techniques for wine quality prediction. Journal of Intelligent Information Systems, 52(3), 465–482.
12. Iyer, S. S. (2020). A comparative study of supervised and unsupervised learning for wine quality prediction. Journal of Food Science and Technology, 57(9), 3301–3312.
13. Sharma, R. K. (2020). Review of machine learning techniques for wine quality prediction. International Journal of Wine Research, 12(2), 67–78.
14. Tiwari, S., Gupta, P., & Kumar, P. (2022). Clustering techniques for wine quality classification. International Journal of Data Science and Analytics, 7(3), 234–245.
15. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

16. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.
17. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
18. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
19. Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn and TensorFlow. O'Reilly Media.
20. Aggarwal, C. C. (2018). Neural Networks and Deep Learning. Springer.
21. Jain, A. K. (2010). Data clustering: 50 years beyond k-means. Pattern Recognition Letters, 31(8), 651–666.
22. Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
23. Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. Morgan Kaufmann.
24. Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining: Practical Machine Learning Tools and Techniques. Elsevier.
25. Bishop, C. M. (1995). Training with noise is equivalent to Tikhonov regularization. Neural Computation, 7(1), 108–116.
26. Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.
27. Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. Informatica, 31, 249–268.
28. Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. Annals of Data Science, 2(2), 165–193.
29. Shmueli, G., et al. (2016). Practical Time Series Forecasting with R. Axelrod Schnall Publishers.
30. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

31. UCI Machine Learning Repository. Wine Quality Dataset. University of California, Irvine.