RESEARCH ARTICLE OPEN ACCESS

Prediction of Wine Quality Using ML

Ms. Puneetha M R[1], Ananya P[2], Ashritha M[3], Bhavana H S[4], Chethana B M[5]

Assistant Professor, Student, Student, Student, Student

Department of Computer Science and Engineering,

K S School of Engineering and Management, Bangalore, Karnataka, India

Abstract:

This project proposes a ML-based approach to predict wine quality. Leveraging a dataset containing wine features and their corresponding sample values, ML algorithms are utilized to learn the intricate relationships between input features and wine quality. Key features such as pH value, temperature, alcohol content, Brix, gravity, and fixed acidity are extracted using specific hardware instruments like pH meter, thermometer, and hydrometer. The Random Forest algorithm is utilized due to its effectiveness in handling complex datasets and providing robust predictions. By inputting user-provided data from these equipment's into the trained model, rapid and accurate predictions of wine quality are generated and compared with standard quality rating in numerical range. This method offers significant benefits for winemakers and enthusiasts, enabling early assessment of wine quality and facilitating decision-making processes during production. Overall, this study demonstrates the potential of ML techniques in enhancing wine quality prediction, thereby contributing to the advancement of winemaking practices.

Keywords — **Wine quality prediction, ML, Random Forest algorithm, pH level, temperature, Alcohol content. Brix. Gravity. Fixed acidity.**

1. INTRODUCTION

Predicting wine quality using ML has become a popular application, offering a sophisticated approach to assess and enhance wine's quality. With the abundance of data available from the vineyards and winemaking processes, ML algorithm can be implemented to analyze patterns and make predictions about the final product's quality. The primary aim of prediction of wine quality through ML is to create a dependable model that can precisely assess and display the quality of wines. It could be achieved by training the model on a dataset that contains the information about wine varieties and their physiochemical features which assist in determining the quality.

2. LITERATURE SURVEY

2.1 Wine Quality Prediction using ML Algorithm-2019

Traditional ML methods like linear regression caused inefficiency in the prototype due to overfitting, while standard lab values were not directly collected, limiting accuracy. Wine quality testing relied solely on quantitative analysis, neglecting qualitative factors like taste and aroma. Bridging the gap between quantitative and qualitative approaches is vital. Integrating AI and ML can combine chemical data with sensory evaluations, enhancing decision-making in wine production.

2.2 Analysis of ML Algorithm to predict Wine Quality-2021

This study utilizes ML to forecast wine quality systematically. It involves pre-processing the initial dataset, training algorithms, and assessing model

accuracies. However, the dataset is limited, and there's a lack of detail on interactions with wine experts and data origin. Contacting experts and adding more data could enhance model accuracy. Incorporating expert insights into feature selection and model development is crucial for robust predictions. Clarity on data collection methods and expert involvement ensures dataset representativeness & ML approach robustness.

2.3 Prediction of Wine Quality using different ML Techniques-2022

This paper excavates into using ML for wine's quality forecasting, aiming for precision over probability-based predictions. It critiques the prevalent approach for its reliance on probability-driven techniques, which lead to increased error rates. To address this, the paper suggests focusing on a carefully selected set of robust ML techniques to enhance prediction accuracy. By prioritizing precision over probabilistic dependencies, the research lays a foundation for improved product quality assessments beyond wine. This nuanced approach contributes to advancing ML applications in product quality assurance.

2.4 Prediction of Wine Quality using ML Algorithms-2022

The research paper addresses the challenge of forecasting wine quality using ML algorithms but highlights concern about the complexity of the algorithms, which can effect in reducing accuracy and require significant computational resources. To mitigate these issues, the paper suggests balancing model complexity and interpretability by considering alternative, streamlined algorithms. By focusing on simplifying the optimization process and selecting algorithms carefully, the research aims in improving the effectiveness and reliability of wine quality determination models in practical applications.

3. PROPOSED SYSTEM

The Wine quality determination system is developed to impose the ML techniques upon the data to assess and determine the quality of wine based on the important factors. Through a careful data pre-processing, the dataset is refined by handling missing values and label binarization is done for further analysis. The model is trained using ML techniques such as random forest, decision trees. Thus, it is trained to solve the complexity of the wine's quality prediction task and its accuracy is evaluated. This model is well designed by anticipating the enduring issues faced by wine maker's while manufacturing wine. It provides them insights regarding wine's quality and expertise to adapt to new data that may change with time.

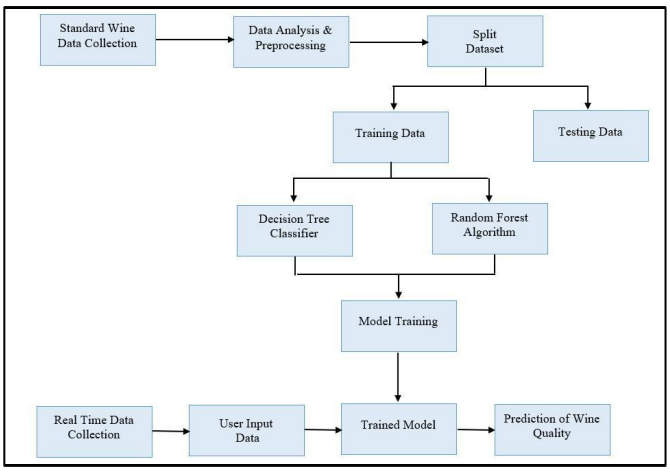

Fig 1. Workflow Architecture

3.1 Proposed Model

The accomplishment of the prediction system rely on the selection of algorithm, quality and relevance of the data, thorough evaluation, and optimization of the model. The proposed model developed involves several steps, from data collection and pre-processing to model training, deployment, and its evaluation.

3.2 Data Collection

The initial step is data collection, where information from measuring instruments is gathered to serve as input for the predictive model. This data should encompass a variety of features such as alcohol, pH, sugar content, and more, which play crucial roles in determining wine quality. Further a standard wine dataset is gathered with various wine characteristics to train the ML model. This dataset should represent a wide range of wine samples. This dataset undergoes the data pre-processing step which involves handling missing values, label binarization

and standardize the features, this prevents any feature from dominating the model training process.

3.3 Model Training

The standard wine dataset is split into training and testing sets to estimate the performance of the model. Random forest algorithm along with Decision trees is an appropriate algorithm which is selected and employed for the prediction task. It aligns with the nature of wine grade prediction problem and the characteristics of the dataset.

3.4 Wine Quality Prediction Task

After the prototype has undergone training and validation, it is time to provide the user-collected input to the system. Users can provide the input of specific wine characteristics into the prototype, and the algorithm will generate predictions regarding grade of the wine on basis of patterns from the training dataset. This interactive element allows users to obtain real-time predictions for specific wines they are interested in.

4. IMPLEMENTATION

The tools utilized for the implementation of our developed model are described in this section. Python is employed extensively for purposes like classification and regression in different domains due to its effectiveness in handling complex relationships within data.

4.1 Python

Python is a high-level general purpose programming language that is popular for its creativity, simplicity, and readability. Python emphasizes readability of the code and ease of use, makes it an exemplary choice for beginners and experienced developers alike. It contains an extensive standard library and an enormous third-party packages ecosystem, Python's syntax is clear and expressive, it allows the developers to put down concise and efficient code. It is an interpreted language, meaning that the code is executed line by line, facilitating rapid development and debugging.

4.2 Jupyter notebook

An open-source web application that focus on interactive creation and sharing of live code contained in documents, equations, narrative text and so on. Jupyter Notebooks provide an interactive computing environment, it helps the users to execute code in a gradual staged manner, view the results, and include explanatory text in a single document. This makes it a popular tool among data scientists, educators, and researchers for tasks such as ML, data analysis, and scientific research.

4.3 Libraries and Frameworks

- **Scikit-learn**: Provides a wide range of ML algorithms and data preprocessing tools, model selection, and evaluation.
- **Pandas**: Used for data manipulation and analysis, particularly helpful for handling tabular data.
- **NumPy**: Essential for numerical computations and array operations.
- **Matplotlib**: A comprehensive plotting library for developing static, interactive, and animated visualizations.

4.4 Random Forest Algorithm

It is composite learning algorithm in ML that operates with creation of numerous decision trees in the time of training and outputs the mode (classification) or mean prediction (regression) of the individual trees for user input data. That is also a versatile and powerful algorithm renowned for its exceptional precision and resilience. It is employed extensively for purposes like classification and regression in different domains due to its effectiveness in handling complex relationships within data.

5. CONCLUSION

In conclusion, the integration of ML for predicting wine quality marks a significant advancement for the industry. Through algorithms like Random Forest and comprehensive data analysis, accurate model has been developed, empowering winemakers to optimize production processes and meet consumer

preferences. Health considerations are also addressed, ensuring regulatory compliance and consumer well-being. This application of ML enhances quality control, fosters innovation, and promotes excellence in winemaking.

Overall, the combination of machine learning algorithms, real-time data collection, and specific hardware instruments represents a powerful toolset for improving wine quality prediction and production processes. Its advantages extend beyond the vineyard, offering societal benefits through enhanced agricultural practices and consumer satisfaction. As technology continues to evolve, further advancements in this field hold the promise of even greater efficiency, sustainability, and quality in wine production.

6. ACKNOWLEDGMENT

We convey our thanks to Ms. Puneetha M R, Assistant Professor, for all her help and support during the project's development. Her knowledge, support, and perceptive criticism have been invaluable in helping us to clarify our thinking and improve our strategy. Her steadfast passion and commitment are greatly appreciated since they have greatly aided in our endeavor's success.

7. REFERENCES

[1]. Devika Pawar, Aakanksha Mahajan, Sachin Bhoithe, “Wine Quality Prediction using ML Algorithms”, International Journal of Computer Applications Technology and Research Volume 8–Issue 09, 2019.

[2]. Nitin Khilari1, Pravin Hadawale, Hasan Shaikh, Sachine Kolase, “Analysis of ML Algorithm to predict Wine Quality” IRJET Dec 2021.

[3]. Prasanna M1, Kamalesh Kumar S, “Wine Quality Prediction using ML Techniques” (IJARSCT) February-2022

[4]. Swati R. Khokale, Bhushan, Diksha, Divya Bhamare, Harshada Pawar, "Review on Wine Quality Testing using ML", JETIR May-2022