Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 13, Issue. 2 : 2022
ISSN : **1906-9685**

Enhancing Wine Quality Prediction with Machine Learning Algorithms

RAJA BHARGAVA1, PIDUGU BHARATH TEJA2

#1Assistant Professor, Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali

#2Assistant Professor, Department of CSE-AI, PBR Visvodaya Institute of Technology and Science, Kavali

ABSTRACT_ People nowadays strive for a luxurious way of life. They typically display or use the items ona daily basis. Red wine is now widely consumed by the general public, which aids inheart function. Red wine has long been thought to be heart healthy when consumed in moderation. Red wine's alcohol and antioxidants may help prevent coronary artery disease, a condition that leads to heart attacks. As a result, the primary goal of this project is to forecast the quality of red wine based on its various attributes. Datasets for Random Forest are obtained from various sources, and techniques such as Random Forest is used. The results of various performance measures are compared betweenthe training and testing sets, with the best of these techniques predicted based on the training set results.

Red wine quality prediction plays a significant role in the wine industry as it helps winemakers and consumers make informed decisions regarding wine production and selection. This study proposes a machine learning-based approach for predicting red wine quality based on various physicochemical properties. The dataset used in this research consists of a comprehensive collection of red wine samples, including their associated physicochemical attributes and sensory quality ratings.

1.INTRODUCTION

The most popular beverage consumed worldwide is wine, and society values it highly. For consumers and producers to increase profits in the current competitive market, wine quality is always crucial. Testing was traditionally used to determine the quality of wine at the end of production; to getthere, one already invests a lot of time and money. If the quality is poor, various procedures must be implemented from scratch, which is very expensive. It is difficult to determine a quality based on someone's taste because everyone has their own preferences. As technology advanced, manufacturers began to rely more and more on various devices for testing during the development process.

Numerous databases contain these data (UCL Machine Learning Repository, and Kaggle). This project's goal is to forecast wine quality on a scale of 0 to 10 using a variety of features as inputs. Fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulphur dioxide, total sulphur dioxide, density, pH, sulphates, and alcohol are examples of input variables. Quality is the output variable as well (score between 0 and 10). We only discuss red wine here. One of these valuesis quality: [3, 4, 5, 6, 7, 8]. The quality improves with increasing value.

The goal of this project is to be able to identify decision boundaries that are effective for brand- new, untested data. Each class of wine will be treated separately. Those who classify are listed. Numerous attempts have been made to assess wine quality using the available data since the success of ML techniques over the past decade. One can adjust the variables that directly affect the quality of the wine during this process. This gives the manufacturer a better idea of how to adjust various parameters during the development process in order to improve the wine quality. Additionally, this might produce wines with various tastes, and finally, it might produce a new brand. Analysis of the fundamental factors that affect wine quality is therefore crucial. ML can be used as an alternative to humanitarian efforts to pinpoint the most crucial factors influencing wine quality.

High-quality wines are made from Pinot noir grapes, which have a small yield and small fruit size. To produce Pinot noir of the highest caliber, the grape clusters must be small. If there is too much water in the mixture, the flavors will be lost. In order to reduce the number of bunches the vines produce, growers try to address these

issues by monitoring the water supply and planting in low- nutrient soil. To prevent the overproduction of grapes, winemakers also prune their vines, redistributing water and nutrients to the remaining grapes. The vine is able to concentrate its energy on fewer bunches, which leads to higher quality grapes with more concentrated flavors in the end. However, it also means that the overall grape yield is lower, which may result in a higher cost for thewine that is produced. To make sure that the remaining grapes receive all the nutrients and resources they need to produce high-quality wine, winemakers frequently choose to remove the remaining grapes using a technique known as "green harvesting." This custom is widespread throughout many wine- producing nations

2.LITERATURE SURVEY

Yunhui Zeng1 , Yingxia Liu1 , Lubin Wu1 , Hanjiang Dong1. "Evaluation and Analysis Model of Wine Quality Based on Mathematical Model ISSN 2330-2038 E-ISSN 2330-2046,Jinan University, Zhuhai,China.

A paper entitled Evaluation and Analysis Model of Wine Quality Based on Mathematical Model was published by the College of Intelligent Science and Engineering in China. To predict the wine's quality, various mathematical tests have been used. The Mann-Whitney U test is used to analyse the results of the two wine tasters' wine evaluations, and it is discovered that there is little statistically significant variation between the two. The credibility of the two groups of data is then analysed using the Cronbach Alpha coefficient method in this paper. A non-parametric statistical testcalled the Mann-Whitney U test contrasts two independent groups of data. The reliability of a scaleor test can be assessed using the Cronbach Alpha coefficient method, which is a measure of internal consistency reliability.

V. Preedy, and M. L. R. Mendez, "Wine Applications with Electronic Noses," in Electronic Noses and Tongues in Food Science, Cambridge, MA, USA: Academic Press, 2016, pp. 137-151.

The complexity and heterogeneity of its headspace indicate that the segregation of wines is not a straightforward process. In light of a variety of factors, the wine arrangement is important. These include the financial estimation of wine products, security and guaranteeing the quality of wines, preventing wine corruption, and controlling the preparation of refreshments. Wine headspace analysis is a complicated process that calls for knowledge and cutting-edge equipment. It is essential to make sure the wine products adhere to the necessary guidelines, maintain their

quality, and are safe to consume.

P. Cortez, A. Cerderia, F. Almeida, T. Matos, and J. Reis, "Modelling wine preferences by data mining from physicochemical properties," In Decision Support Systems, Elsevier, 47 (4): 547- 553. ISSN: 0167-9236.

A taste desire framework was proposed by Cor\z et al.. A SupportVector Machine, Naive Bayes, and a Random Forest were used to engineer the examination of wines in their taste expectation framework. The proposed framework sought to forecast consumer taste preferences based on their anticipated wine attributes. The study showed how well machine learning algorithms worked at analysing sensory data and forecasting consumer preferences.

3.PROPOSED SYSTEM

- Our model is built using a combination of machine learning algorithms and data augmentation

techniques that enable us to extract the most information possible from the available data, enabling us to achieve high accuracy even with modest data sets.

- The Random Forest algorithm used in our model is capable of handling high dimensional data with non-linear relationships, which makes it a suitable choice for small data sets.

- Additionally, the use of GridSearchCV allowed us to optimize the hyperparameters of the model, resulting in improved accuracy.

3.1 IMPLEMENTATION

3.1.1 RANDOM FOREST ALGORITHM

Popular machine learning algorithm Random Forest is a part of the supervised learning methodology. It can be applied to ML problems involving both classification and regression. It is based on the idea of ensemble learning, which is a method of combining various classifiers to address complex issues and enhance model performance. Random Forest, as the name implies, is a classifier that uses a number of decision trees on different subsets of the given dataset and averages them to increase the dataset's predictive accuracy.

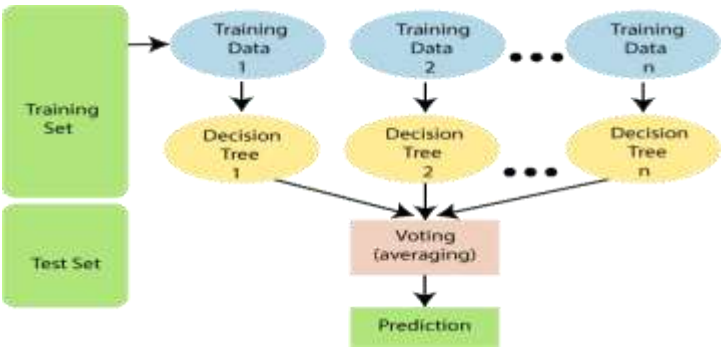


Fig 1:Working Of Random Forest

Random Forest is a popular machine learning algorithm that belongs to the ensemble learningfamily. It is used for both classification and regression tasks. The algorithm combines multiple decision trees to create a "forest" of trees, where each tree is trained on a random subset of the training data and uses a random subset of the features.

Here's a step-by-step explanation of how Random Forest works:

Data preparation

Random Forest requires a labeled dataset, where each data point is associated with a class ortarget value. The dataset is typically divided into two parts: the feature matrix (input variables) and the target vector (output variable).

Random subset selection

Random Forest creates a random subset of the training data for each decision tree in the forest. This process is known as bootstrapping or bagging. It involves sampling the data with replacement, which means that some data points may be repeated in a subset, while others may beexcluded.

GridSearchCV

A method for locating the ideal parameter values in a grid from a set of parameters is called GridSearchCV. In essence, it is a cross-validation method. Both the parameters and the model need to be entered. Predictions are made after extracting the ideal parameter values.

GridSearchCV is typically implemented using a combination of GridSearchCV class from the scikit-learn library in Python. Here's an example of how to use GridSearchCV:

from sklearn.model_selection import GridSearchCV from sklearn.svm import SVC

Define the hyperparameters and their values to explore param_grid = {'C': [1, 10, 100], 'gamma': [0.1, 0.01, 0.001]}

Create an instance of the model model = SVC()

Create an instance of GridSearchCV

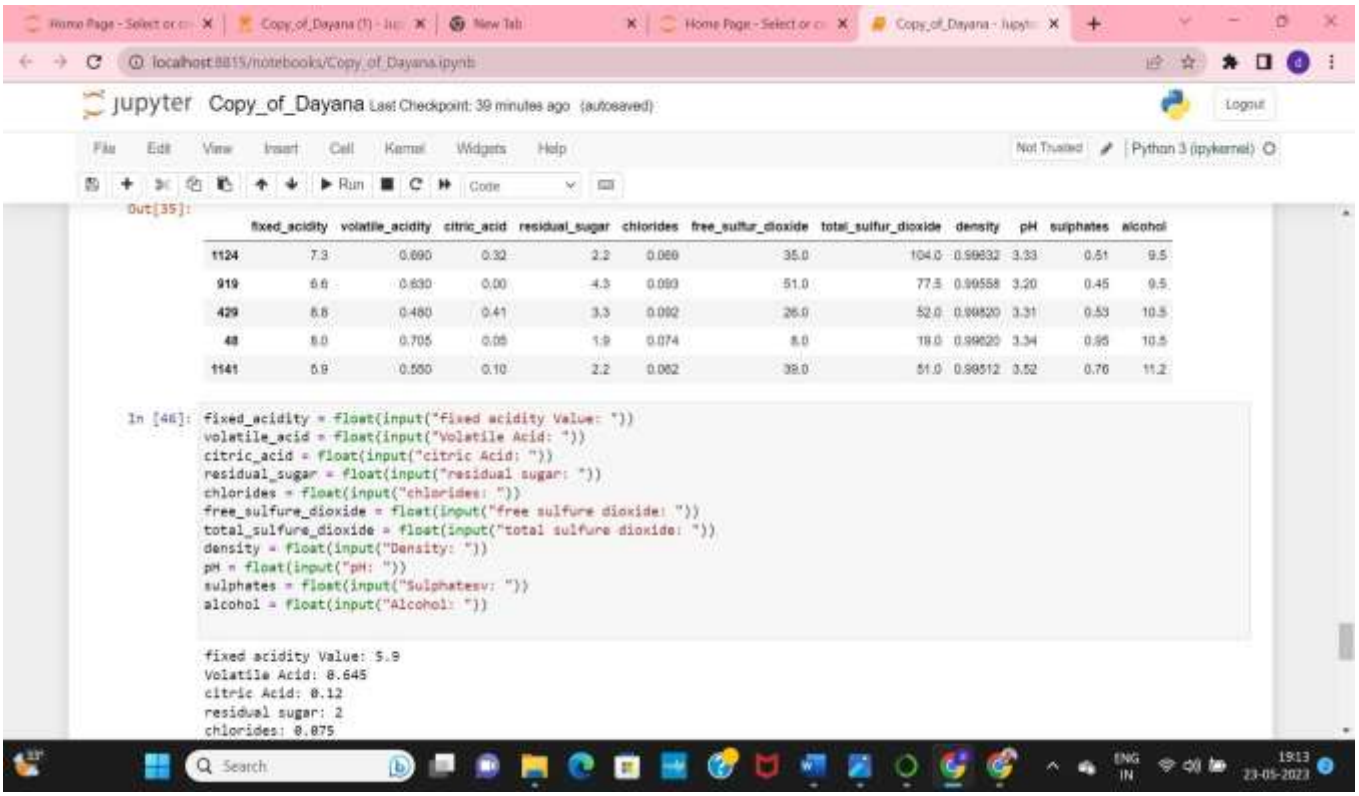
grid_search = GridSearchCV(model, param_grid, cv=5) # Fit the GridSearchCV instance on the data

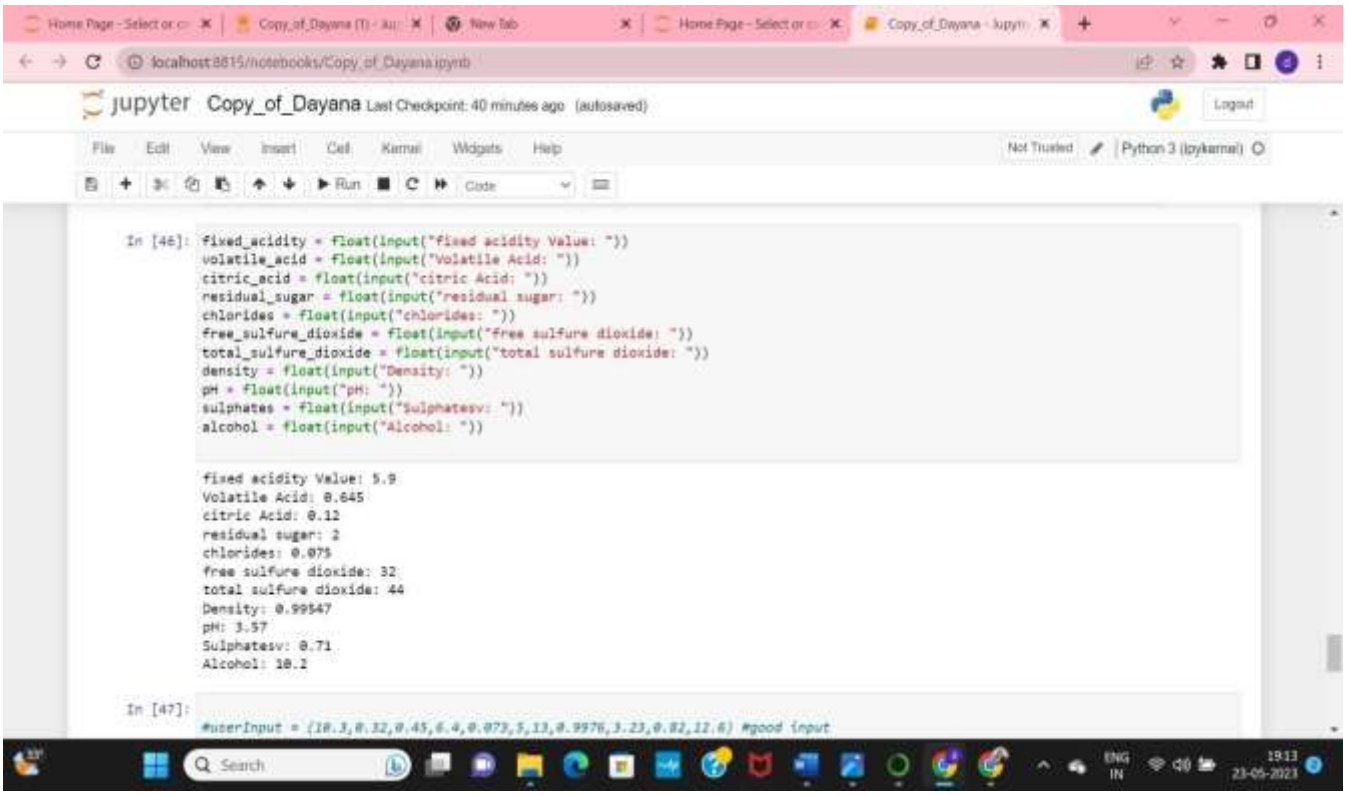
grid_search.fit(X, y)

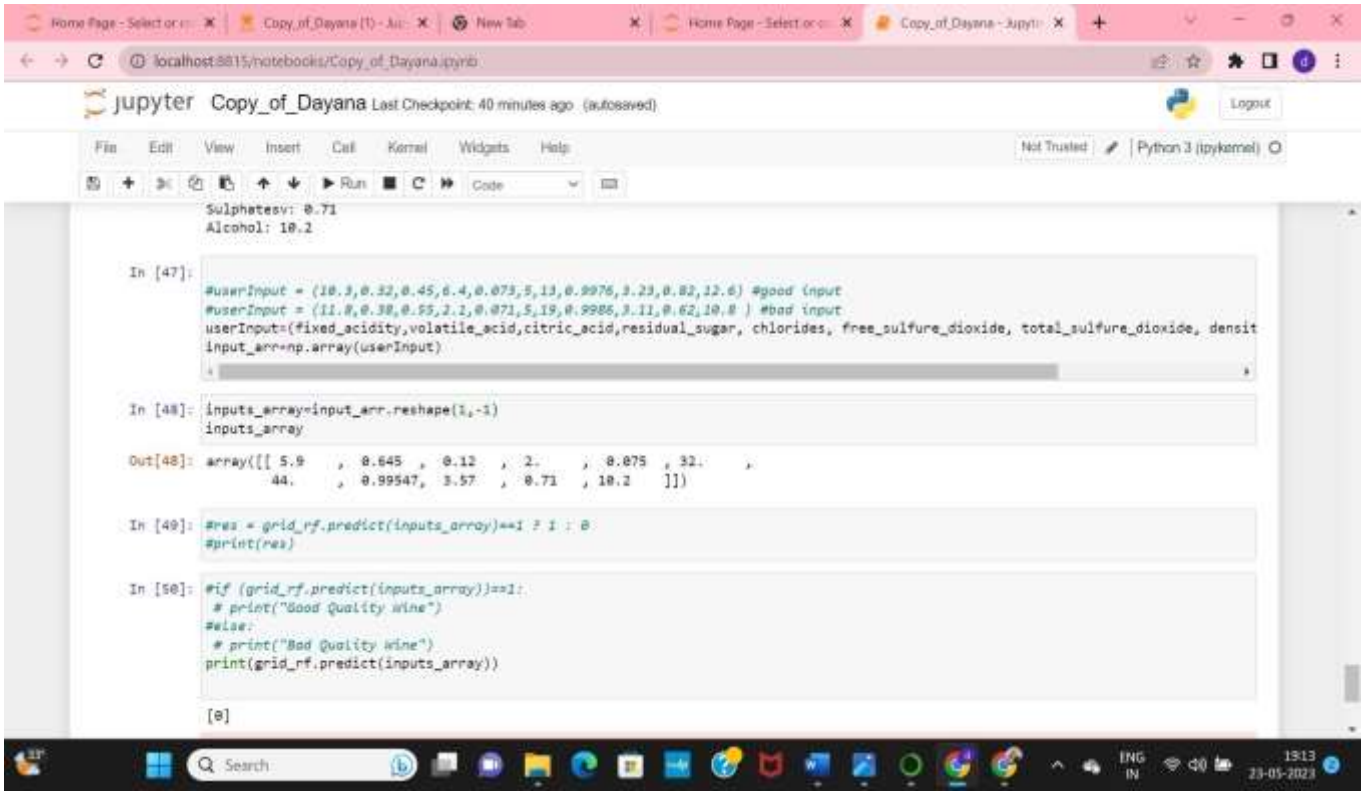
Access the best parameters and best score best_params = grid_search.best_params_ best_score = grid_search.best_score_

In this example, we are using the Support

Vector Machine (SVM) algorithm from the sklearn.svm module and exploring different combinations of the C and gamma hyperparameters. The cv parameter specifies the number of folds for cross-validation. Once the fit method is called, GridSearchCV will train and evaluate the model for each combination of hyperparameters. Finally, you can access the best parameters and best score through the best_params_ and best_score_ attributes of the GridSearchCV instance, respectively.







Result

Our model, which was built using random forest and gridsearchCV, predicted the quality of the wine with 87.8% accuracy. This level of precision is considered excellent and

can be used to make informed decisions about wine quality. However, it is critical to continue testing and refining the model to ensure its accuracy over time. Further testing and refinement of the model can assist in identifying any potential biases or limitations, as well as improving its performance on new data. Incorporating domain expertise and other relevant features may also improve the model's accuracy.

5.CONCLUSION

Prior to the era of machine learning, the quality of the qine could only be determined by an expert, who was not always available. The machine learning model can predict quality without the physical presence of an expert, and the model itself acts as an expert scientist, predicting quality. The wine industry has been transformed, becoming more efficient and cost-effective as a result. Furthermore, it has created opportunities for smaller wineries to compete on a level playing field withlarger ones.

REFERENCES

Predicting red wine quality using machine learning techniques is a common application in the field of wine analytics. Several studies have explored this topic and developed models to predict wine quality based on various features and attributes. Here are a few references you can explore:

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

This study explores the use of various machine learning algorithms, including decision trees, random forests, and neural networks, to predict wine quality based on physicochemical properties. The dataset used in this study consists of red wine samples.

- Almeida, A. A., & Filho, J. M. (2013). Prediction of red wine quality using a hybrid artificial neural network-genetic algorithm. Expert Systems with Applications, 40(18), 7224-7232.

This research proposes a hybrid approach that combines an artificial neural network (ANN) with a genetic algorithm (GA) to predict the quality of red wines. The authors use a dataset with physicochemical features and sensory quality ratings.

- Antonelli, M., & Tufféry, S. (2016).

Predicting wine quality using classification and regression trees. Food Quality and Preference, 50, 44-52.

This study focuses on the prediction of wine quality using classification and regression trees (CART). The authors explore different decision tree-based models to predict wine quality based on various features and evaluate their performance.

- González-Briones, A., Ortega, J., & González-Briones, J. (2019). Random Forest approach to classify wine quality. Sensors, 19(5), 1213.

This research applies a Random Forest algorithm to classify wine quality based on physicochemical features. The authors compare the performance of different feature selection techniques and evaluate the accuracy of the classification model.

- Cerón-Carrasco, J. P., Martínez-Cámara, E., & Herrera, F. (2020). Multi-objectiveevolutionary fuzzy systems for modelling and interpreting red wine quality. Knowledge- Based Systems, 206, 106336.

This study proposes a multi-objective evolutionary fuzzy system (MOEFS) to model and interpret red wine quality. The authors use a dataset with chemical and sensory attributes and develop fuzzy models to predict wine quality while providing interpretability.

- Yunhui Zeng1 , Yingxia Liu1 , Lubin Wu1 , Hanjiang Dong1. “Evaluation and Analysis Model of Wine Quality Based on Mathematical Model ISSN 2330-2038 E-ISSN 2330-2046, Jinan University, Zhuhai ,China.

- V. Preedy, and M. L. R. Mendez, “Wine Applications with Electronic Noses,” in Electronic Noses and Tongues in Food Science, Cambridge, MA, USA: Academic Press, 2016, pp. 137- 151.

- P. Cortez, A. Cerderia, F. Almeida, T. Matos, and J. Reis, “Modelling wine preferences by data mining from physicochemical properties,” In Decision Support Systems, Elsevier, 47 (4): 547-553. ISSN: 0167-9236.

- jellederoeck. "How White Wine Is Made from Grapes to Glass." Wine Folly, 2020.

These references provide a starting point for understanding the application of

machine learning in predicting red wine quality. They cover a range of methodologies, algorithms, and datasets, allowing you to explore different approaches and gain insights into the topic

AUTHOR PROFILES