IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204019 |

Wine Quality Prediction Using Artificial Intelligence and Machine Learning

Prof. Rahul. D. Dhongade[1], Abhay Mule, Onkar Padole[2], Anurag Sasane[3], Rushikesh Vairat[4]

Prof, Department of Computer Engineering AISSMS Polytechnic, Pune, India[1]

Students, Department. of Computer Engineering., AISSMS Polytechnic, Pune, India[2,3,4]

ABSTRACT: To Predict the wine Quality is a difficult task. Wine Quality is a very important factor for the consumers as well as the wine industry. The traditional method is time-consuming expensive and repetitive. In this time of data science, Artificial Intelligence (AI) & Machine Learning (ML) are likely to replace human intervention in different sectors. Machine Learning (ML) is used in various Industries like medicine, Astrophysics and it will replace many more industries. Artificial Intelligence (AI) & Machine Learning (ML) aims to learn the structure of data and fit it into models, which latterly can be used on unseen data to negotiate a particular task. In this age of Artificial Intelligence (AI) & Machine Learning (ML), we can make decisions to predict the wine quality on the basis of some features/variables. We can use it for to predict wine quality through the "eleven physicochemical properties" of wine. The work utilizes the Python programming language for the prediction of wine quality. We are using Machine learning algorithms are used to detect good or bad wine quality. Among various machine learning methods, we are using the Extremely Randomized Trees (Extra Trees) ML method and it will perform perfectly for the prediction. This will be easier to access for anyone. In this way can make simple wine quality predictions.

KEYWORDS: Machine Learning, Artificial Intelligence, ML methods, wine quality

I. INTRODUCTION

Wine is the most common beverage global sphere, and the quality of wine is also an important factor for the consumers, as well as for the wine industries. It will also have potential benefits for the heart. In the past decade, the consumption of wine was considered a royal commodity. The production of wine has increased as per the global demand and nowadays it has a wide range of customers. In the past decade wine quality is determined at the end of the production, and it required more time and money. Consumption of wine is increasing nowadays not for only fun but also for the human heart.Machine learning prediction is now easy with the different algorithms and ML models. Researching can make the right prediction using the right tools. Python language has great benefits in analyzing the data set. It will also help to work easier. The correct set of parameters is essential for better prediction. Machine learning algorithm to predict the wine quality through the features/variables. Machine learning models are useful for predicting the wine quality. Machine learning algorithms can predict the values accurately. Every person has their own opinion about the taste, it's difficult to identify quality according to a person's taste.

II. RELATED WORK

[1] Dahal, K., Dahal, J., Banjade, H. and Gaire, S. used algorithms like Ridge regression, support vector machine, gradient boosting regressor, and Artificial neural network and found out that the Gradient Boosting showed better execution than others. If we increase the training datasets, at that point we get the benefits of prediction performance of artificial neural network. They utilized a few datasets with restricted values, focusing on the red wine dataset only.

[2] Perpetual Opoku Agyemang utilized the strategies of Ordinal logistic regression and multinomial Logistic Regression and found out that the ordinal logistic regression model minimized the whole of squared errors. It failed the parallel line test so the multinomial logistic regression technique was used. The multinomial logistic regression model was less delicate to exceptions and the resultant output observed was the exactness rate for the red wine is 60.3% and white wine is 53.8%. We can say that precision exceptionally is too low using these two models as compared to the other machine learning models.

[3] Gongzhu Hu, Tan Xi, Faraz Mohammed, and Huaikou Miao utilized machine learning algorithms like Decision Tree, Adaboost, and Random Forest. According to their findings, a statistical method known as the Synthetic Minority

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204019 |

Oversampling Strategy (SMOTE) is successful in adjusting datasets by expanding the number of cases. It is observed that SMOTE was applied to these three ML models and can be connected to different dataset(s) and nourish it to other ML models.

[4] F.Fan, J.Li, G.Gao, and C. Ma authors used machine learning methods like fitting analysis, variance analysis, and Q cluster analysis. It was watched by them that the optimized model of principal component examination can't obtain correct evaluation results only by the grape and the physical-chemical pointers. These models utilize a very large number of successful and dependable raw information and related data to solve the issue step by step.

[5] Using an SVM classifier as their machine learning technique, Satyabrata Aich, Ahmed Abdulhakim Al-Absi, Kueh Lee Hui, and Mangal Sain were able to achieve better results than any other classifier when it came to both red and white wine data sets. The precision levels shifted from 95.23% to 98.81% over different feature sets. Limitations that can be watched are that authors utilized a small dataset and tested it utilizing SVM as it were, having the opportunity they might have computed the execution of the other available ML models as well.

[6] Sunny Kumar, Kanika Agrawal, and Nelshan Mandan proposed to utilize ML methods like the Naïve Bayes algorithm, Support Vector Machine, and Random Forest and found that the accuracy obtained for the Naïve Bayes algorithm, Support Vector Machine, Random Forest are 55.91% and 55.89%, 67.25% and 68.64%, 65.83% and 65.46% individually and concluded that they utilized only red wine dataset which has imbalanced data that's why exactness is low.

III. METHODOLOGY

A.*Problem Statement*

Predicting wine quality on the premise of physicochemical properties by using 6498 wine samples from the wine dataset which has the fundamental features that could affect the quality of wine by analyzing Extremely randomized trees (Extra trees) ML method and selecting the precise one for our web application.

B.*Proposed System Overview*

This paper mainly focuses on the execution of wine quality analysis using Artificial Intelligence and Machine Learning.
In this proposed system architecture, we have the following modules:

1. Importing Dependencies - We have to import all the dependencies, basic modules, libraries, and packages that we will need for this project. We are importing Python libraries and packages - numpy, pandas, train_test_split, cross_val_score, ExtraTreesClassifier, and pickle.
2. Load Wine Dataset - We have loaded the wine dataset, which includes information on 11 physicochemical properties, as well as the wine type and quality. The dataset comprises 6,498 samples of wine, and we utilized the quality column, which ranges from 0 to 10, for making predictions.
3. Data Analysis - A technique to collect, transfigure, and organize data to make future predictions, and make informed data-driven opinions. We are analyzing the number of rows, columns, datatype, statistical measures, values for each class, etc.
4. Exploratory Data Analysis and Preprocessing - To identify implicit outliers, we can use box plots or distribution plots. If we detect outliers in certain attributes, removing them could enhance the accuracy of our model. However, this will depend on the impact of these outliers on our results. We should also check if there are any missing values and whether the class distribution is balanced. To handle missing values, we can fill them with the mean value of the corresponding trait using the mean() function. To address imbalanced classes, we can use the SMOTE oversampling function, which generates new features from minority classes.
5. Train-test-split and assessing and opting the ML model - A train-test split is a technique that involves dividing your dataset into two subparts:- a training set and a testing set. The training set is used to train the machine learning model, while the testing set is used to estimate its performance. By doing this, you can train your model on a subset of the data and also test its accuracy on data that it has not seen ahead. In our project, we are assessing the performance of the Extra Trees machine learning model and we have chosen the most accurate one to be used in our web application.

All this processing will be done at the backend. The user will see a modern webpage with device responsiveness, input fields for inputting wine properties, and a result page.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204019 |

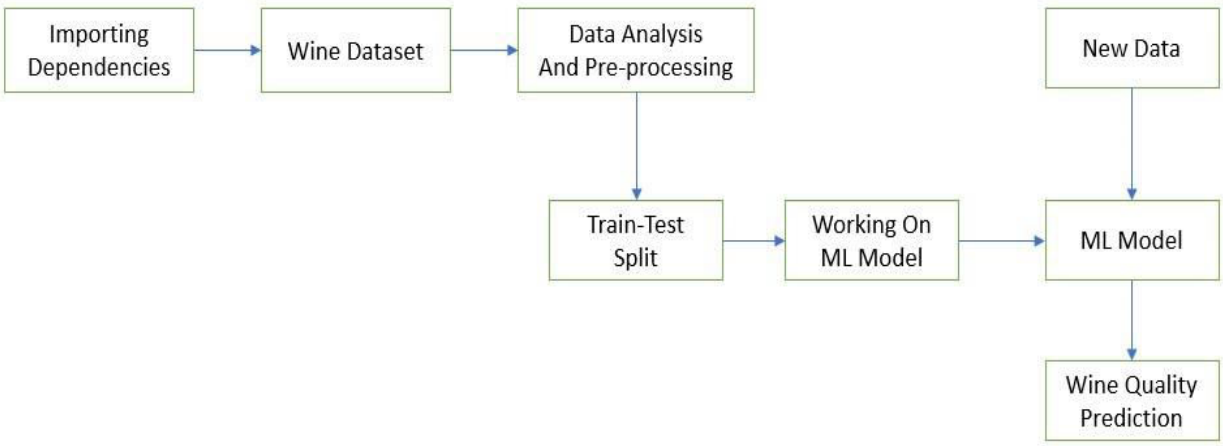

Figure 1: Proposed System Architecture

C.*ML Models*

Three ensemble machine learning methods were used in our study. A model was built using each method and applied to the data. The basic ideas of the three methods are given here.

- Extra Trees (Extremely Randomized Trees).
  It is an ensemble-supervised machine-learning method that uses decision trees. Unlike other tree-based methods like Random Forest, it creates a large number of decision trees using randomly selected features and samples, making it less prone to overfitting.

IV. SIMULATION RESULTS

Table 1 presents the ML models analyzed, and based on the results, Extra Trees is the most accurate model for this dataset. Consequently, the application will utilize this model to predict a user's wine quality based on their data.

Sr. No.	ML Model	Accuracy (%)	CV Score
1	Extra Trees	89.53	81.71

Table 1: Evaluation of Performance of used ML Model.

Figure 2: GUI of the web app.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204019 |

The page of our prediction application is displayed in Figure 2. It is a contemporary web application built using the Flask framework and designed to be device-responsive, ensuring optimal viewing on various devices including mobiles, tablets, desktops, and laptops. The input fields have also been equipped with validation capabilities.

The application's backend comprises several components, including data analysis, preprocessing, model evaluation, and selection, as discussed in Section III. The front end, on the other hand, is built using Flask framework, HTML, CSS, and Bootstrap.

To obtain a quality prediction, the user must input 11 physicochemical properties of wine, and the application will generate a prediction based on this input.

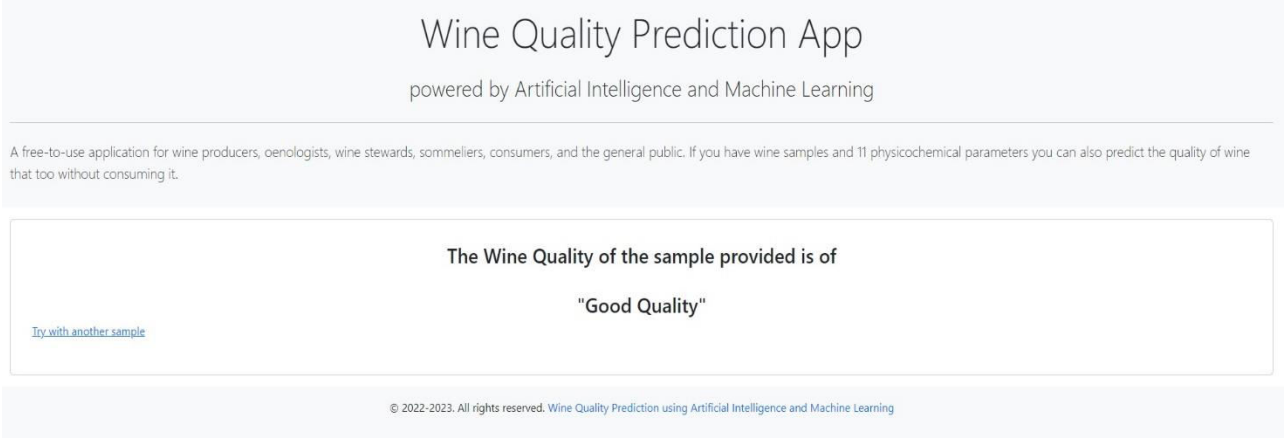

Figure 3: Predicted results of user's input.

The outcome of the wine quality for the user's provided data is illustrated in Figure 3. The application has indicated it as "Good Quality," although it could be subject to variation based on the entered data.

V. CONCLUSION AND FUTURE WORK

We presented an analysis of wine quality on the dataset of 6497 red and white wine samples of "Vinho Verde" wine by using one ensemble of machine learning methods. The Synthetic Minority Oversampling Technique (SMOTE) was used to efficiently solve the problems created by imbalanced data. In conjunction with SMOTE, the Extra trees classifier outperformed the other machine learning models in terms of accuracy and CV score and it is going to be used by the prediction system. This web app can be deployed on any cloud service platform so that it can be accessed by anyone from anywhere. It might be beneficial for the global wine market, enology students, and consumers as well. Various performance measures and other machine learning techniques can be applied for better comparison of results. Better algorithms can be developed which involves the combination of best features of all other data mining techniques. We are also working on the next version where the user can upload his dataset and can select from various trained ML models.

REFERENCES

[1] "Prediction of Wine Quality Using Machine Learning Algorithms" written by K. R. Dahal, J. N. Dahal, H. Banjade, S. Gaire, published by Open Journal of Statistics, Vol.11 No.2, 2021.
[2] P.Opoku, A. (2010)'Modelling the preference of wine quality using logistic regression techniques based on physiochemical properties', YOUNGSTOWN STATE UNIVERSITY.
[3] G. Hu, T. Xi, F. Mohammed, and H. Miao, "Classification of wine quality with imbalanced data,"publised in 2016 by IEEE (ICIT), 2016
[4] F.Fan, Jianping Li, G.Gao and C.Ma, "Mathematical model application based on statistics in the evaluation analysis of grape wine quality,"12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2015,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204019 |

[5]T.M.Getangali, Sowjanya M Y, Rohith S N, Paper on "prediction of Wine quality using Machine Learning" 2021 published by JETIR.ORG

[6]https://www.kaggle.com/datasets/yasserh/wine-quality-dataset

[7]https://scikit-learn.org/

[8] https://www.geeksforgeeks.org/wine-quality-prediction-machine-learning/

[9]S. Kumar, K. Agrawal and N. Mandan, "Red Wine Quality Prediction Using Machine Learning Techniques," in 2020 published by (ICCCI), 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details