Volume 9, Issue 3, March – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR2134

Red Wine Quality Prediction using Machine Learning

[1]B. Jhansi Vazaram; [2]D. Shiva Sankar; [3]M. Lokesh; [4]M. Mallikarjuna
[1]Head of the Department M. Tech, Ph. D. , [2,3,4]Students,
AI Department, Narasaraopeta Engineering College, Narasaraopeta, A.P, India

Abstract:- The objective of this study aimed to create a model to forecast the quality of red wine by examining its physicochemical attributes. Various factors affect the precision of quality prediction in red wine analysis. This paper presents a computational intelligence approach employing machine learning methods. Specifically, the Random Forest Classifier, Naive Bayes Algorithm, and Support Vector Machine were applied. Using these machine learning techniques and the provided information, it becomes possible to predict the quality of a given red wine sample.

Keywords:- *Red Wine, Naive Bayes Algorithm,Support Vector Machine, Quality Prediction and Random Forest Classifier.*

I. INTRODUCTION

A rapidly expanding area of artificial intelligence called machine learning enables computers to automatically learn from experience and get better over time without the need for explic it programming. With the use of machine learning, computers can now analyse enormous amounts of data, identify patterns and trends, and utilise that knowledge to forecast the future or guide decisions. Machine learning speech recognition and tailored suggestions have many practical applications. As machine learning advances, it is expected to have a big impact on a lot of businesses and aspects of everyday life.

The goal of applying machine learning to estimate the quality of red wine is to improve processing efficiency and accuracy.

Using past data and prediction algorithms, machine learning models may be taught to identify trends and predict the chance that a claim will be accepted or rejected.

Computers can learn on their own from historical data thanks to machine learning, a cutting-edge field of research.

Machine learning uses a variety of techniques to create mathematical models and produce predictions based on previously gathered data or information.

The goal of using machine learning to win equality prediction is to improve processing accuracy and efficiency. There are a few machine learning software programmes available for building this system.

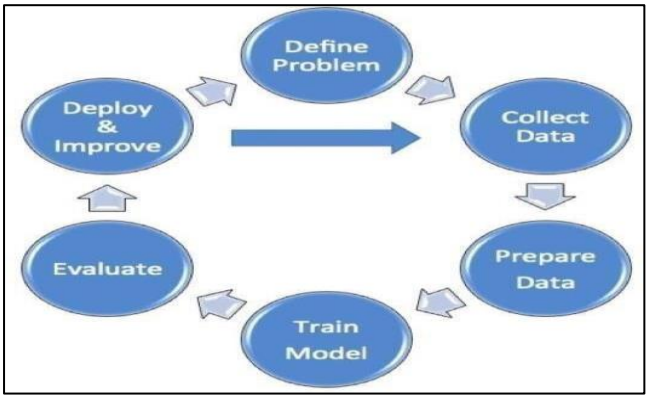

Fig 1: 5 Steps Involved in Model

II. LITERATURE SURVEY

The extant literature was reviewed to acquire therequisiteunderstanding of several concepts relatedto the current application of our model. Some ofthe most important conclusions came from those. the study of red wine quality assessment through arange of techniques, such as chemical analysis, machine learning, and sensory assessment. It would discuss the advantages and dis advantagesof each approach and highlight any knowledgegaps.

An explanation of the components that determine the quality of red wine and its chemical composition. It would discuss the effects ofvarious ingredients on red wine's flavour, aroma,and colour, such as acids, sugars, and phenolics. The many machine learning methods that havebeen used to forecast rely on the physicochemical properties of red wine and its quality. It wouldddiscuss the benefits and drawbacks of each algorithm and sugar, chlorides, free sulphur dioxide, total Sulphur dioxide, density, pH, sulphates, alcohol, quality,fixed acidity, and volatile acidity.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR2134

Table 1: Dataset

pH	Sample's pH
sulphates	Sample's sulphates
alcohol	Sample's alcohol
quality	Sample's quality

The dataset includes 12 columns: citric acid, residual highlight the key performance metrics that are used to evaluate each one's effectiveness.

The methods utilised to determine which featuresare most helpful in predicting the quality of red wine and to optimise the model's parameters for optimal results. Along with discussing the challenges and limitations of current methods, it would offer suggestions for the future.

Using information from the dataset, we try topredict the quality of a sample of red wine in this study.

We have obtained accuracy of up to 80% with Random Forest Classifier, 72% with Naive Bayes Algorithm, and approximately 57% with Support Vector Machine. The model is trained using thesemachine learning algorithms.

III. MATERIALS AND METHODOLOGY

Our model is suggested based on the following characteristics. To be able to forecast anything, we must have a dataset on the Quality of red wine, which we acquired from the Kaggle website.

Table 2: Description

Column	Description
fixed acidity	Sample's fixed acidity
Volatile acidity	Sample's volatile acidity
Citric acid	Sample's citric acid
Residual sugar	Sample's residual sugar
Chlorides	Sample's chloride
Free sulfur dioxide	Sample's free sulfur dioxide
Total sulfur dioxide	Sample's total sulfur dioxide
Density	Sample's density

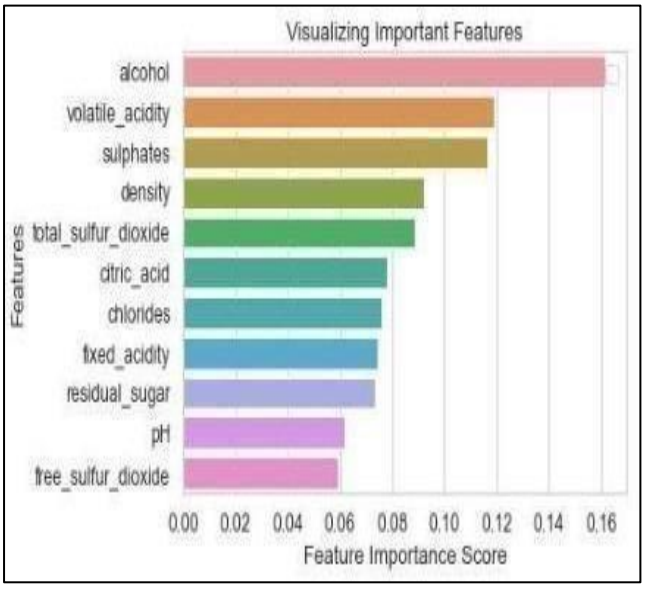

Fig 2: Features Visualizations

The quality values across the ranges are depicted in the below diagram.

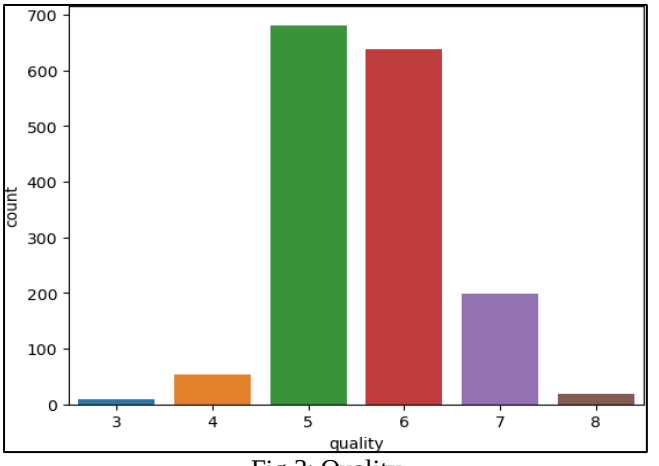

Fig 3: Quality

Pre-processing is the process of making changesto our data before presenting it to the algorithm. A process that transforms uncleaneddata from raw data. Comparatively, data are nevercollected in aformat suitable for analysis when theyare obtained from multiple sources. Because of a certain machine learning technique, data must bein a certain format. For instance, null values must behandled from the original raw data set in order toapply the Random Forest method because the RandomForest algorithm does not allow null values. To get the greatest results, a data collection shouldbe structured to allow for the utilisation of numerous machine learning technique.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR2134

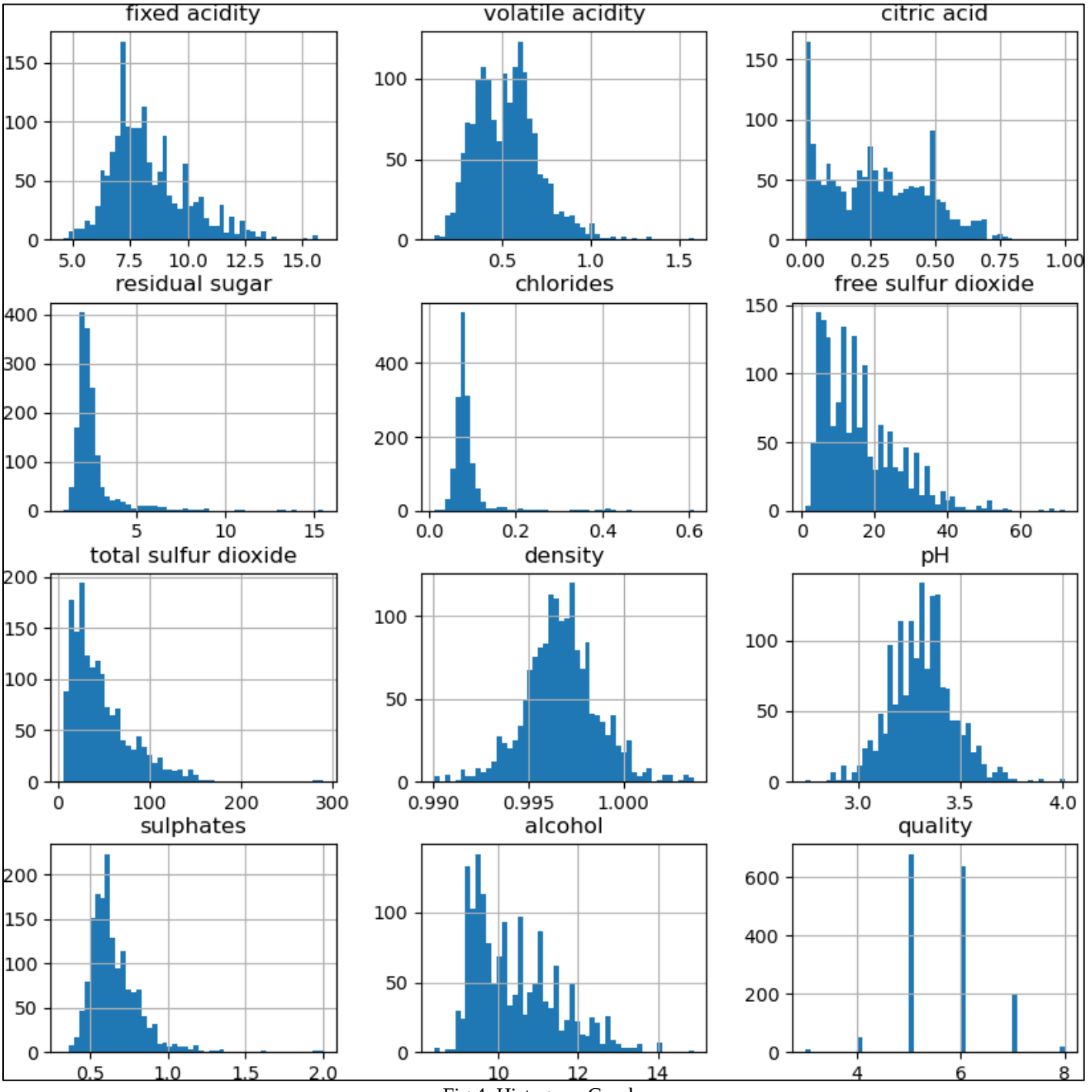

Fig 4: Histogram Graph

It is eventually found that there are certain outliers inthe data. We used a boxplot to provide a clear overview of the outliers for each column. A cleaned dataset was createdafter the unnecessary data was eliminated using percentiles.

We distinguished between good and poor quality feature values in the cleaned datasets. We used LabelEncoder() to translate the good and bad values into binary values of 0 and 1. It is discovered that bad becomes the binary value 0 and good becomes the binary value 1. We deleted the quality attribute since it is regardedas a goal variable and split the dataset into 70% fortraining and 30% for testing before running the models.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR2134

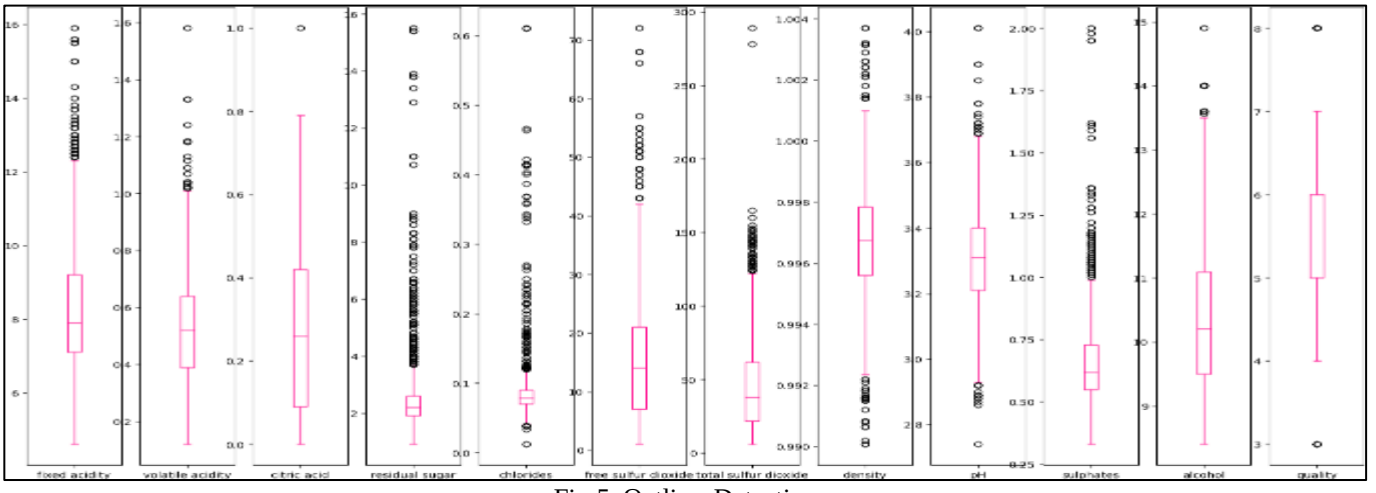

Fig 5: Outliers Detection

The technologies we're using for this include the Support Vector Machine, Naive Bayes Algorithm,and Random Forest Classifier. A random forest classifier model is used for training.Using the sklearn accuracy score, we are evaluating both the model's and the result's predicted accuracy.

The Naive Bayes Algorithm and Support Vector Machine were also applied in the same manner.

➢ *These are the Outcomes:*

Algorithm of Naive Bayes, 71.7%, 56.8% of support vector machines.

80.2% for Random Forest Classifier

We found that, out of the other two models, the Random Forest Classifier performs the best because of its accuracy. As a result, we thought that this model was themost reliable for determining the quality of the red winesample.

IV. CONCLUSION AND FUTURE SCOPE

Predicting whether a red wine will be good or badwas the study's objective, and machine learning techniques were employed to achieve this. The investigation showed a significant improvement inthe models' performance, and we discovered that the random forest classifier had a higher accuracy than the support vector machine and naïve bayes approaches. As predicting the quality of red wine was our aim, we chose a random forest classifier model.

On the other hand, future studies can concentrate on investigating several other deep learning uses. Better methods from machine learning and other domains can be applied to **improve** the methodology that can converge the modifications and perform multiple frame works.

REFERENCES

[1]. Paulo Cortez, António Cerdeira, Fernando Almeida, Telmo Matos, &José Reis. Modelingwinepreferences by data mining from physicochemical properties. DecisionSupport Systems, 47(4), 547- 553

[2]. Edelmann, Andrea , et al. "Rapid Method for theDiscrimination of RedWine Cultivars Based on Mid- Infrared Spectroscopy of Phenolic Wine Extracts." Journal of Agricultural & FoodChemistry49.3(2001):1139-1145.

[3]. Zhang Shiling, Xu Ruimin.Formation and prevention of volatileacidin wine. New Rural Technology,2008 (06): 81-82

[4]. Dahal, K., Dahal, J., Banjade, H., Gaire, S., 2021. Prediction of Wine Quality Using Machine Learning Algorithms. Open J. Stat. 11,278– 289

[5]. Rish, I., 2001. An Empirical Study ofthe Naïve Bayes Classifier. IJCAI 2001 Work Empir Methods Artif Intell 3.

[6]. Moreno, Gonzalez-Weller, Gutierrez, Marino, Camean, Gonzalez and Hardisson. (2007) “Differentiation of two Canary DO red wines according to their metal content from inductively coupled plasma optical emission spectrometry and graphite furnace atomic absorption spectrometry by usingProbabilistic Neural Networks”. Talanta72 263–268.