WWW.IJRPR.COM
5818

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Advanced Techniques for Predicting Quality of Wine Based on ML Approaches

Waghole Pradnya [A], Narode Swapnil [b], Shinde Sudarshan [C], Shinde Pravin[d], Mrs. Lingampalli P. V. [e]

[a,b,c,d] Students, Rajiv Gandhi Colllege Of Engineering, Ahmednagar, Maharashtra
[e] Assistant Professor, Rajiv Gandhi College Of Engineering, Ahmednagar

ABSTRACT

This paper are useful for Decision Tree and Random Forests, a model for training, predicting, and evaluating wine quality dataset determines whether each wine sample is red or white and determines its quality, which can be low, medium, or high. Wine is a reduced alcohol-content beverage made from fermented grape and other fruit juices. Wine quality is determined by the vintage and wine's flavor. Like wine itself, the process of tasting it is old. Other than flavor, a variety of other elements or characteristics are taken into account while determining the wine's quality. The dataset used to examine "Wine Quality" shows how well wines (both white and red) perform depending on various physiochemical characteristics (fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulphur dioxide, total sulphur dioxide, density, pH, sulphates, and alcohol). Each wine pairing in the dataset has a quality score that ranges from 0 to 10 (from lowest to highest). This research will reveal some significant correlations between the chemical components of wine, such as its acidity and sugar concentration, and its quality, flavour.

Keywords: - Machine Learning , wine quality, Logistic regression, recall precision, random forest, KNN, SVM.

I. INTRODUCTION

Testing is a crucial component that ensures product quality and helps to improve product quality. Different kinds of businesses are embracing and implementing new technology nowadays to check and evaluate the quality of their products. Testing a product's quality with human knowledge is a costly and labor-intensive process that takes time to perform. This study examines various machine learning techniques for determining wine quality, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Ada Boost Classifier. Wine quality evaluation is one of the key elements that may be used for certification, and this type of quality certification aids in ensuring wine quality in the market. The input variables for the red wine data set are density, pH, sulphates, citric acid, residual sugar chlorides, free sulphur dioxide, total sulphur dioxide, fixed acidity, volatile, acidity,and,alcohol.Higher value indicates better wine quality.

The quality is rated on a scale of one to ten.

Machine learning algorithms are often categorized as supervised, unsupervised and semi supervised algorithms:

Supervised machine learning algorithms can use what they have learnt in the past to forecast future events. The learning algorithm creates an inferred function from the examination of a known training dataset in order to forecast the output values. After sufficient training, the system is capable of providing objectives for any new input. The learning algorithm can also check its output against the desired, correct output to identify mistakes and fix the model as necessary.

Unsupervised machine learning algorithms when the data used to train is neither categorised nor tagged, it is used. Unsupervised learning investigates how computers can use unlabeled data to derive a function that describes a hidden structure. Although the system is unable to determine the proper output, it examines the data and is able to infer hidden structures from unlabeled data using datasets.

Semi-supervised machine learning algorithms Since they use both labelled and unlabelled data for training—typically a small quantity of labelled data and a big amount of unlabelled data—they fall between supervised and unsupervised learning. This technique enables systems to significantly increase learning accuracy. Semi-supervised learning is typically used when the collected labelled data calls for knowledgeable and pertinent resources to train it / learn from it. Otherwise, collecting unlabeled data usually doesn't call for extra resources.

II. LITERATURE SURVEY

Sowmya D el. et.author examines how physical and chemical characteristics, such as the amount of alcohol, chlorides, sulphates, and other ingredients, affect the quality of wine in this paper. This study uses a variety of physicochemical factors to examine the different types and qualities of wine. Red and white wine samples were used to produce two datasets [1].

Nitin Khilari el.et This study demonstrates how the results for each categorization model fluctuate depending on the test style. A component of the research is the analysis of classifiers using red wine datasets. The results are explained using the proportion of properly detected cases, precision, recall, and F measure[2].

Prasanna M el. et. In this study author can assist businesses in establishing standards for determining if a wine is of good or low quality. Using a variety of ML approaches and a Bigdata framework, we have created a model to forecast the wine quality conditions. From the provided dataset, this varied algorithm provides us with the best and most accurate results [3].

Dragana Radosavljevi et.el.author proposed that, a technique for classifying wines based on their measurable physical and chemical attributes is described. A collection of 6497 wine samples from Portugal's northwest are used in the experiment. The classification was performed on both the entire dataset as well as the two subgroups that contained information specifically about white and red wine. Each dataset is divided into two categories [4].

III. DATA MINING TECHNIQUES

Linear Regression

A supervised learning method called linear regression studies the fixed acidity to forecast wine quality based on independent factors. This establishes the link between the dataset variables and the prediction. Regression models differ in the sort of association that is assessed between dependent and independent variables as well as the quantity of independent variables.

$$s = \frac{(\sum \beta)(\sum \alpha^2) - (\sum \alpha)(\sum \alpha\beta)}{n(\sum \alpha^2) - (\sum \alpha)^2}$$

$$t = \frac{n(\sum \alpha\beta) - (\sum \alpha)(\sum \beta)}{n(\sum \alpha^2) - (\sum \alpha)^2}$$

Logistic Regression

When the inference is categorical, the logistic regression method is applied. This procedure is used to determine a wine's alcohol content and residual flavors, which determine whether the wine is sweet, sour, salty, or bitter.

$$s = e\hat{}(x_0 + x_1 * t)/\left(1 + e\hat{}(x_0 + x_1 * t)\right)$$

Naïve Bayes Classifier

The Bayesian classifier is an ML model that is trusted and utilised for classification-related tasks. This is one of the Bayesian classification algorithms that is entirely based on the Bayes theorem, i.e., the probability that A will occur when B does so may be calculated by multiplying the likelihood that B will occur when A does so by the likelihood that A will occur. After that, this is split by the likelihood of B.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

IV. PROPOSED WORK

Figure 1 shows the general organizational layout of the proposed Wine quality grading scheme. the Kaggle dataset for wine quality [25]. In order to assume that

the machine learning models would be more accurate, regression analysis and classification approaches were applied. All data were fed into three distinct classifiers

to divide the input data into excellent quality and bad quality. In order to reuse the system in the event of information loss and avoid failure, the files reserved here are done in a redundant manner. Due to the fact that the data is spread across numerous workstations, the HDFS supports parallel processing.

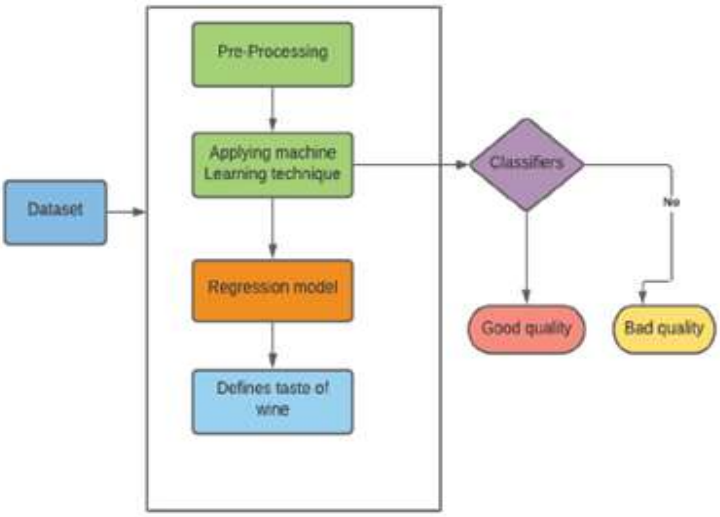

Fig. Wine Quality model

V. CONCLUSION

Components of a wine, such as its aroma, flavour, and colour, can help determine its quality. Using machine learning techniques, it is possible to anticipate the wine's quality and assist businesses in determining if a given bottle is of good or low quality. Using a variety of ML approaches and a Bigdata framework, we have created a model to forecast the wine quality conditions. From the provided dataset, this varied algorithm provides us with the best and most accurate results. Different assessment measures, such as precision, recall, accuracy, and confusion matrix, will be used to distinguish between the algorithm's results. We conclude that based on the findings above, Random Forest Classifier and Logistic Regression provide the highest and best accuracy of 99% to forecast the Quality of wine.

References

[1]. Sowmya D , Sayyed Johar "ANALYZING WINE TYPES AND QUALITY USING MACHINE LEARNING TECHNIQUES" Vol. 4, Issue 3, ISSN No. 2455-2143, Pages 519-529 Published Online July 2019 in IJEAST (http://www.ijeast.com).

[2] Nitin Khilari*, Pravin Hadawale, Hasan Shaikh, Sachin Kolase "Analysis of Machine Learning Algorithm to predict Wine Quality" Online ISSN : 2394-4099 doi.org/10.32628/IJSRSET229235

[3]. Prasanna M and Kamalesh Kumar S Wine Quality Prediction using ML Techniques and KNIME Volume 2, Issue 1, February 2022 DOI: 10.48175/IJARSCT-2526

[4] Dragana Radosavljevi "A DATA MINING APPROACH TO WINE QUALITY PREDICTION" International Scientific Conference "UNITECH 2019" – Gabrovo 320.