International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 1246-1252 www.ijaem.net ISSN: 2395-5252

Wine quality prediction

Swapnil Darade, Nilesh Korade

Submitted: 25-06-2021 Revised: 04-07-2021 Accepted: 07-07-2021

ABSTRACT: On wine quality datasets, several data mining algorithms have already been used to evaluate wine attributes such as quality or class. Not only does the quality of wine depend on the amount of alcohol, it also depends on different qualities, these attributes change with time and so the quality of wine is also refined. Deep learning methods are used in this study to evaluate certain attributes. It is important to determine the quality of the wine and to classify it into different categories on the basis of a quality evaluation. Data mining is the right way to accomplish this by analyzing the data collection, since it collects valuable information.

KEYWORDS – wine quality, pre-processing, classifier algorithm, random forest etc.

I. INTRODUCTIONS

Wine quality is very significant because the consumption of low-quality wine has a detrimental impact on human health. The wine industry is a big field where people are deceived in terms of wine quality. The manual approach to identifying and marking the quality of wine at various levels of quality is time-consuming and not precise. There is a need for a method that can distinguish the product at various levels of quality by taking into account its characteristics such as alcohol, malic acid, ash, ash alkalinity, magnesium, total phenols, flavanoids, non-flavanoid phenols, proanthocyanins, dilute wine colour strength, Hue, OD280/OD315, proline.It can be labelled at various quality levels by evaluating the quantity of these attributes present in the wine. The use of data mining classification algorithms is the best approach to this problem

OBJECTIVE:

Our main objective is to predict the wine quality using deep learningthrough python programming language.A large dataset is considered and wine quality is modelled to analyze the quality of wine through parameters like fixed acidity, volatile acidity etc. All these parameters will be analyzed through machine learning algorithms like random forest classifier algorithm which helps to rate the wine of scale 1-10 or bad-good.It can support the wine expert evaluations and ultimately improve the productions.

RELATED WORK OR LITERATURE SURVEY

[1]"Classification of Wine Quality withImbalanced Data"Author:Gongzhu Hu, Tan Xi, Faraz Mohammed,Huaikou MiaoTo classify wine into various categories of quality, we propose a data analysis approach. Our study used a data collection of white wines from 4898 observations collected from the Minho region in Portugal.We used the Synthetic Minority Over- Sampling Technique (SMOTE) to over-sample the minority class because the frequency of events in the data set was imbalanced with around 93 percent of the observations from one group.The balanced information was used to model a classifier that classifies A wine of high quality, normal quality, and three grades Bad quality, of poor quality

[2] Wine Quality Detection through Machine Learning Algorithms

Author: Akanksha Trivedi, Ruchi SehrawatDepartment of Computer Science and Engineering, GGSIPU, USIC&T, Dwarka, New Delhi, India

Machine learning is one of the emerging areas of research. Many algorithms of data mining have already been used on wine quality dataset to analyze the wine attributes such as quality or class. The quality of wine is not only based on the quantity of alcohol but it also depends on various attributes, these attributes changes with time and so the quality of wine also refines. In this report, machine learning techniques are utilized to analyze those attributes. Firstly data pre processing takes place i.e. making data appropriate for the models that are built for prediction

[3] Wine Quality Prediction Using Data Mining

Author:Shruthi P Department of CSE ATME College of Engineering Mysuru, INDIA

Certifying the quality of food items is the country's main concern. It is recommended that the people of the country use only quality-assured goods. It is also important to apply the same thing

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 1246-1252 www.ijaem.net ISSN: 2395-5252

to the wine industry. It is important to determine the quality of wine and to classify it into various categories on the basis of a quality assessment.Thecorrect approach to this is data mining, as it collects valuable information by analysing the data collection.

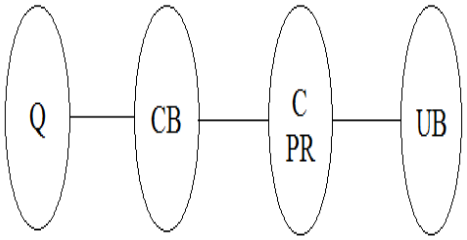
[4] Prediction of Quality for Different Type of Wine based on Different Feature Sets Using Supervised Machine Learning Techniques
Author: SatyabrataAich*, Ahmed Abdulhakim Al-Absi**, Kueh Lee Hui***, and Mangal Sain****

Most companies have been marketing their products in recent years, based on the quality certification they received for the products. The conventional method of measuring the quality of the product is time-consuming, but the methods have become more efficient and consumed less time than before with the advent of machine learning techniques.We also discussed some of the machine learning strategies in this paper to measure the quality of wine based on the quality-dependent attributes of wine. For this research paper, we used white wine and red wine consistency datasets.

[5] Red Wine Quality Prediction Using Machine Learning Techniques
Author: Sunny Kumar, Kanika Agrawal, Nelshan Mandan

People are seeking to lead a luxurious life nowadays. They prefer to use the stuff either for their regular basis or for their show off. The consumption of red wine these days is very popular for all. In order to maintain human health, it has become important to evaluate the quality of red wine before its consumption. Therefore, this analysis is a step towards predicting the quality of red wine using its different attributes.The data set is taken from sources and added to techniques such as Random Forest, Support Vector Machine, and Naïve Bayes. Different measures are determined and the outcomes are compared between the training set and the test set and, accordingly, the best of the three strategies are estimated based on the results of the training set.

II. MATHEMATICAL MODELING



Where,
Q = read the dataset
CB = preprocess
C = apply classifier algorithem
PR = preprocess request evaluation
UB = predict outcome

Set Theory
1) Let S be as system which input image
S = {In, P, Op,Φ}
2) Identify Input In as
In = {Q}
Where,
Q = User entered input image(dataset)
3) Identify Process P as
P = {CB, C, PR}
Where,
CB = Pre-process
C = apply classifier algorithm
PR = Pre-process request evaluation
4) Identify Output Op as
Op = {UB}
Where,
UB = Predict outcome
Φ=Failures and Success conditions.

Failures:
1. Huge database can lead to more time consumption to get the information.
2. Hardware failure.
3. Software failure.

Success:
1. Search the required information from available in Datasets.
2. User gets result very fast according to their needs.

Space Complexity:
The space complexity depends on Presentation and visualization of discovered patterns. More the storage of data more is the space complexity.

Time Complexity:
Check No. of patterns available in the datasets= n

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 1246-1252 www.ijaem.net ISSN: 2395-5252

If (n>1) then retrieving of information can be time consuming. So the time complexity of this algorithm is $O(n^n)$.

Above mathematical model is NP-Complete.

EXISTING SYSTEM AND DISADVANTAGES

This paper presents an integrated model using machine learning algorithm to predict in using past data available. The existing system is implemented and detect wine quality using naive bays algorithm.

PROPOSEDSYSTEM AND ADVANTAGES

It give insights of the dependency of target variables on independent variables using deep learning techniques to determine the quality of wine because it gives the best outcomes for the assurance of quality of wine.The dependent variables is quality rating other variables i.e alcohol , sulphur etc. are assumed to be predictors or independent variables.

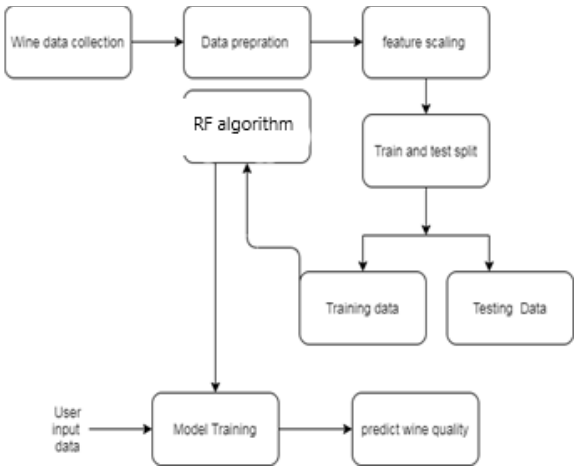


Figure: Advance System Architecture

Advantages:

1) Secure and efficient system.
2) Improve wine quality prediction.

Result:

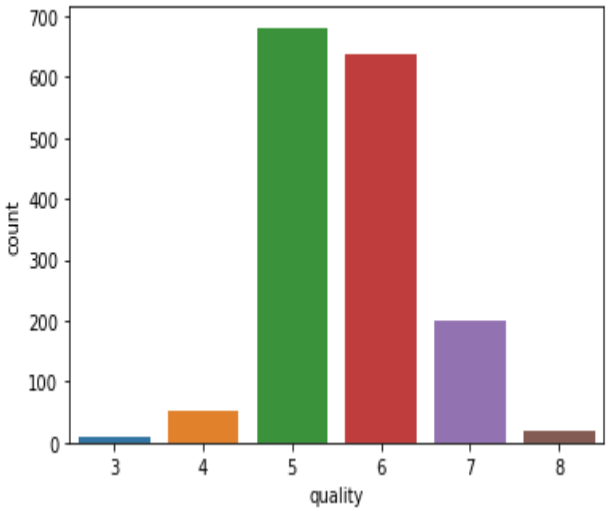


Fig :- Quality of wine count.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 1246-1252 www.ijaem.net ISSN: 2395-5252

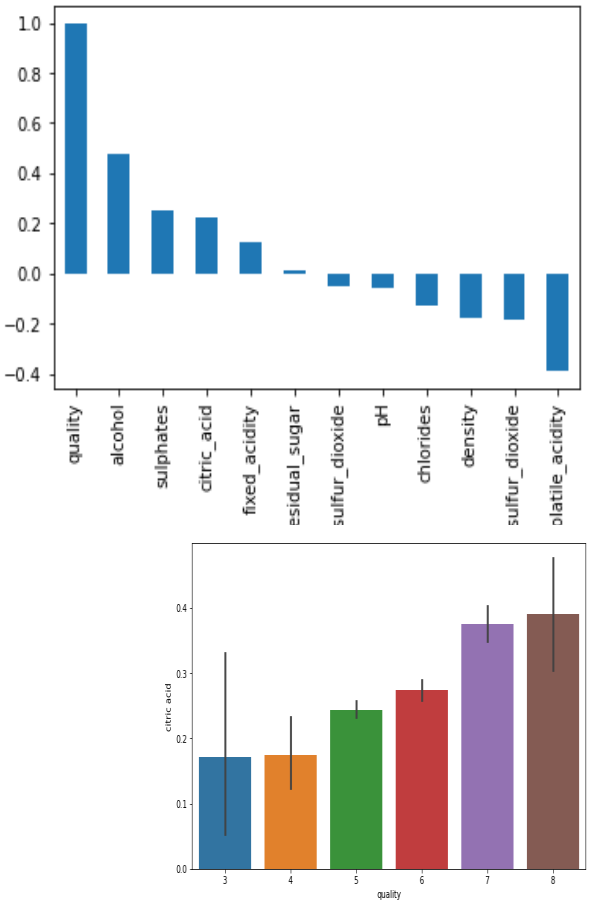


Fig:- Quality of Alcohol

***Accuracy Decision Tree And Random Forest Algorithm**

Regression Model	precision	recall	F1-score
Decision tree	0.9275	0.9229	0.9245
Random Forest	0.9661	0.9687	0.9660

```
+----------------+--------------------+--------------------+--------------------+
| Regressor Model |     Precision      |       Recall       |      F1Score       |
+----------------+--------------------+--------------------+--------------------+
|  Decision Tree  | 0.9275119525941894 | 0.9229166666666667 | 0.9245859696662925 |
|  Random Forest  | 0.9661947165672622 |      0.96875       | 0.9660315070376045 |
+----------------+--------------------+--------------------+--------------------+
```

<u>0.92</u> is accuracy ofDecision Tree algorithm.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 1246-1252 www.ijaem.net ISSN: 2395-5252

***Random Forest Algorithm**

	precision	recall	F1-score	support
0	0.91	0.97	0.94	273
1	0.74	0.43	0.54	47

	precision	recall	f1-score	support
0	0.91	0.97	0.94	273
1	0.74	0.43	0.54	47
accuracy			0.89	320
macro avg	0.82	0.70	0.74	320
weighted avg	0.88	0.89	0.88	320

<u>0.91</u> is accuracy of Wine quality prediction as compare to CNN by using random forest algorithm.

***SVM(Support Vector Machine) Algorithm.**

	precision	recall	F1-score	support
0	0.88	0.98	0.93	273
1	0.71	0.26	0.37	47

	precision	recall	f1-score	support
0	0.88	0.98	0.93	273
1	0.71	0.26	0.37	47
accuracy			0.88	320
macro avg	0.80	0.62	0.65	320
weighted avg	0.86	0.88	0.85	320

<u>0.88</u>is SVM accuracypredicted by using SVM algorithm.

***MLP (multilayer perceptron):**

```
from sklearn.neural_network import MLPClassifier

# creating the model
model = MLPClassifier(hidden_layer_sizes = (100, 100), max_iter = 150)

# feeding the training data to the model
model.fit(x_train, y_train)

# calculating the accuracies
print("training accuracy :", model.score(x_train, y_train))
print("testing accuracy :", model.score(x_test, y_test))
```

```
training accuracy : 0.9307756463719766
testing accuracy : 0.75
```

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 1246-1252 www.ijaem.net ISSN: 2395-5252

```
print(classification_report(y_test, y_pred))

print(confusion_matrix(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.80      0.74      0.77       204
           1       0.75      0.81      0.78       196

    accuracy                           0.78       400
   macro avg       0.78      0.78      0.77       400
weighted avg       0.78      0.78      0.77       400

[[151  53]
 [ 37 159]]
```

	precision	recall	F1-score	support
0	0.80	0.74	0.77	204
1	0.75	0.81	0.78	196

Training Accuracy	0.930775646
Testing Accuracy	0.75000000

MLPhas 0.93 which is highest as compare to other algorithm.
Comparison of accuracy.

	MLP	**Random forest**	**Decision Tree**	**SVM**
Accuracy	**0.93**	**0.91**	**0.92**	**0.88**

MLP(multilayer perceptron) is highest accuracy with 0.93

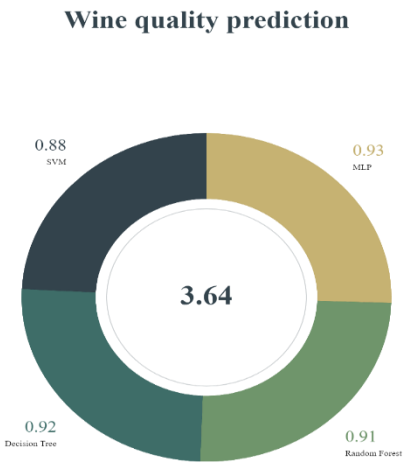


Fig:- Pie Chart Of Accuracy

III. CONCLUSION

The most popular technique today is data mining, which The archives are used for investigation. It looks at the data and produces the yield required. It helps to play the sound test on the market with the advancement in innovation, thereby helping the customer.This paper addressed the MLP algorithm's potential as well as the simulated annealing algorithm for feature selection, as well as the classifier's potential to accurately predict based on the new feature sets. A clear idea of the significance of the attributes for quality prediction was provided by the feature selection algorithm.

IV. FUTURE WORK

In the future, it is possible to create better algorithms that require the combination of all other data mining techniques with the best features. If such modifications in the hyper plane and balanced tree technique are used along with the necessary probability, much better precision can be observed.

REFERENCES

[1]. Sáenz-Navajas, M. P., Campo, E., Sutan, A., Ballester, J., &Valentin, D., "Perception of wine quality according to extrinsic cues: The case of Burgundy wine consumers," Food Quality and Preference, 27(1), 2013, pp.44-53.

[2]. Simple guide to confusion matrix terminology. (2014, March). Retrieved April 2018, from data school: http://www.dataschool.io/simple-guide-toconfusion- matrix-terminology/.

[3]. Ertas, G., "Detection of high GS risk group prostate tumors by diffusion tensor imaging and logistic regression modelling," Magnetic resonance imaging, 2018.

[4]. Cho, J., Kim, H., Gebreselassie, A. L., & Shin, D., "Deep neural network and random forest classifier for source tracking of chemical leaks using fence monitoring data," Journal of Loss Prevention in the Process Industries, Elsevier, 2018.

[5]. B. Chen, C. Rhodes, A. Crawford, and L. Hambuchen, "Wineinformatics: applying data mining on wine sensory reviews processed by the computational wine wheel," IEEE International Conference on Data Mining Workshop, pp. 142-149, Dec. 2014.

[6]. K. Agrawal and H. Mohan, "Text Analysis: Techniques, Applications and Challenges," presented in 2019 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, Tamil Nadu, India, 2019.

[7]. K. Agrawal and H. Mohan, "Cardiotocography Analysis for Fetal State Classification Using Machine Learning Algorithms," 2019 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, Tamil Nadu, India, 2019, pp. 1-6.