ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

QUALITY PREDICTION FOR WINE USING MACHINE LEARNING

Dr.T.Narashimha Rao, V.Kalyan[2], V.Sri Bhavana [3], Md.Afreen Fathima [4], M.Rohith Phanendra [5], [1]Professor, [2,3,4,5]Students Dept. of Computer Science & Engineering, Dhanekula Institute of Engineering and Technology, A.P., India

ABSTRACT

Quality assessment is a key factor for the wine industry, where the aim is to meet consumers' needs/demands and in these days the consumption of wine is very common to all. So it became important to investigate quality of wine before its consumption to preserve human health. We are employing an ANN model to predict the wine quality based on several features in project. Qualities predicted on a scale of 3 to 8, with 3 indicating poor wine quality and 8 indicating good wine quality. Here the user provides input features of the wine then this model determines that given wine is good to consumable or not. By using ANN(artificial neural networks) we can improve accuracy of prediction and helps wine manufacture companies to control the quality prior to the wine production

INTRODUCTION

The most often used substance is wine and its worth is widely regarded in society. In today's competitive market, wine quality is always critical for buyers and, more significantly, for producers to boost income. Professionals usually judged the wine's quality at the ending of the manufacturing process. However, the production companies needs to invest money and effort in the process of production in order to reach that level. To generate good wine, production companies use several combinations of wine in order to appeal to a larger number of clients and meet their demands. But, wine quality validation was always done at the end of the process. If a substandard quality wine is produced as a result of changing wine production combinations, the production company's efforts, money, and time will be wasted. Everyone in the world having there own tastes, assessing a person's quality based on their preferences is challenging. Wine producers can know use this technologies to get an pre-result before the actual production. This technologies are very useful for the producers. Production companies might save money and time as a result of this technologies. Various efforts have made to

DOI : 10.36893.JK.2022.V12I01N01.540-546

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

estimate wine quality using available data since the growth of machine learning techniques over the previous decade . These wine quality data can be available from various sources like (UCL ML Repository, and Kaggle) And these Machine Learning approaches can provide a result prior to production, allowing production companies to trying out various combinations until they achieve a high-quality wine. Changing the combinations a greater number of times may result in numerous flavours and, eventually, a new brand. As a result, determining the basic parameters that affect wine quality is critical.

LITERATURE SURVEY

International Journal of Computer Science and Information Technologies published AHP and MACHINE LEARNING TECHNIQUES for Wine Recommend. They utilised the analytical hierarchy technique to rank the variables and then used several machine learning classifiers like support vector machine and Random Forest to classify the red-wine-dataset and assign weights to them to solve the problem.The result that obtained after analysis of the data is used to recommend a wine to individuals. They discovered that random forest was used by 70.33 percent and SVM was used by 66.54 percent.

A Case Study on Wine Data To propose the wine, they used a user-centric clustering approach. They used the red wine data set for the poll. They allocated relative votes to the attributes based on the research findings. Then they assigned weights to the attributes using Gaussain Distribution process. They assessed the quality based on the preferences of the users.

For this study, they took white wine & red wine quality datasets. They have used different feature selection technique such as genetic algorithm (GA) based feature extraction and simulated annealing (SA) based feature extraction to check the prediction performance. They have also compared the performance metrics of linear, nonlinear, and probabilistic based classifiers and it was found that these classifiers performed well with the new feature sets. And they have found that the SA based feature sets performed better than the GA based feature sets. They have also found that the SVM classifier performed better compared to all other classifiers for red wine and white wine data sets

EXISTING SYSTEM:

In the existing system they used to get the red wine dataset from the UCI Repository. Random Forest and Support Vector Machine, and Naive Bayes algorithms

DOI : 10.36893.JK.2022.V12I01N01.540-546

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

were used. since 70% and 30% of the dataset is divided into training and testing, respectively. Based on the training set outcomes, the best of the three techniques is predicted. The existing system looks into accuracy, precision, misclassification error, F-score, recall, and specificity to determine the best of three. With an accuracy of 67.25 percent, Support Vector Machine was chosen to be the best method. When used on RStudio software to estimate red wine quality, Random Forest finished in second with an accuracy of 65.83 percent, and the Nave Bayes algorithm came in third with an accuracy of 55.91 percent.

PROPOSED SYSTEM

With this proposed system, we want to improve the accuracy of wine quality prediction. We are using the dataset from the UCI machine learning repository [11] to conduct the research. The dataset for this red wine data contains 1599 occurrences with 12 features. We're creating an Artificial Neural Network model here. The dataset is then split into two parts: 70 percent training and 30 percent testing for the training and testing phase of the model. Then the user gives 11 input features to the model and the model predicts a value in between the range of 0 and 1. The threshold value is 0.5 in this case. If the predicted value less than the threshold value, the wine combination result a poor quality. If the predicted value exceeds then threshold value, the wine combination result a good quality. In most datasets, artificial neural networks outperform other mathematical models, The class of the test data set can be predicted fast and easily by ANN. It can also predict multi-class. Artificial neural networks boost accuracy and performance.

METHODOLOGY & IMPLEMENTATION

To conduct the research, we are utilising data from the UCI machine learning library. Fixed acidity, citrus acid, volatile acidity, residual sugar, chlorides, thickness, free sulphur dioxide, absolute sulphur dioxide, pH, alcohol, and sulphates are among the 1599 instances with 12 variables in the dataset for red wine data. We have a quality attribute in the red wine dataset with values ranging from 3 to 8. The Bad Wine is indicated by 3,4,5, whereas the Good Wine is indicated by 6,7,8. The values of the quality property are converted into Bad and Good Strings. They're then converted to a numerical format (0s and 1s). The value '0' indicates that a bad wine will come from a combination of attributes in the dataset and the value '1' indicates that a good wine will come from a combination of attributes in

DOI : 10.36893.JK.2022.V12I01N01.540-546

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

the dataset. The dataset is then split into two parts: 70 percent training and 30 percent testing.

After that we have to build an Artificial Neural Network model.

ARTIFICAL NEURAL NETWORK

The input layer, output layer, and hidden layer/s are the three layers that make up an ANN. There must be a connection between the input layer nodes and the hidden layer nodes, as well as between each hidden layer node and the output layer nodes. The data is taken from the network by the input layer. The raw data from the input layer is received by the hidden layer, which processes it. The obtained value is then transferred to the output layer, which will also analyse and send the information from the hidden layer. And weights are present in between this layers the weights assigned randomly from 0-1 range. After initialize the weights for input and hidden layers, calculate the activation function for middle layer. Calculate the error predict value with error loss. If we are getting more loss in that epoch go with another epoch update the weights and do back propagation.

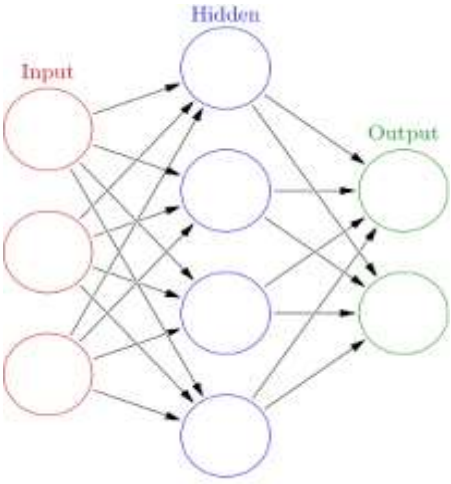


Fig 1: Structure of Artificial Neural Networks

Testing data is used to train the ANN Model. And the Testing dataset is used to validate the ANN Model and used to find the Accuracy, Precision, F1-Score. Then the user gives 11 input features to the model and the model predicts a value between 0 and 1. The threshold value is 0.5 in this case. If the predicted value less than the threshold value, the wine combination result a poor quality. If the predicted value exceeds the threshold value, the wine combination result a good quality.

Measures to Calculate Performance

To determine the model's effectiveness and efficiency, performance measurements are utilised. Using the testing dataset, we can assess the ANN model's performance in terms of accuracy, precision, and F1-Score.

Accuracy: The value predicted when the sum of True Positive and True Negative

DOI : 10.36893.JK.2022.V12I01N01.540-546

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

values of a confusion matrix is divided by the sum of True Positive, False Positive, False Negative, and True Negative values.

Accuracy=(True Positive + True Negative) / (True Positive + False Positive + False Negative + True Negative)

Precision is the result of dividing True Positive by the confusion matrix's sum of True Positive and False Positive values.

Precision = True Positive / (True Positive + False Positive)

F-Score: The F1 Score is calculated by multiplying Recall and Precision by the confusion matrix's sum of Recall and Precision. The result is then divided by two.

F1 Score =2 * (Recall * Precision) / (Recall + Precision)

The structure of the ANN is shown in Fig 1. Fig 2 shows a glimpse of information from the red wine dataset, which has 1599 values and 12 features. The confusion matrix of the red wine dataset for the training set utilising ANN the training set contains of 1199 records and 12 features is shown in Fig 3. The error matrix of the red wine dataset for testing set utilising ANN the testing dataset contains of 400 records and 12 features and they are randomly chosen from the original dataset is shown in Fig 4. Fig 5: As the epochs go on, the model accuracy of the training and testing sets improves. Fig 6 As the epochs go on, the model loss of the training and testing sets decrease. Fig 7 presents the red wine dataset table for the ANN training set. Figure 8 presents a table of red wine datasets for ANN testing.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1599 entries, 0 to 1598
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   fixed acidity         1599 non-null   float64
 1   volatile acidity      1599 non-null   float64
 2   citric acid           1599 non-null   float64
 3   residual sugar        1599 non-null   float64
 4   chlorides             1599 non-null   float64
 5   free sulfur dioxide   1599 non-null   float64
 6   total sulfur dioxide  1599 non-null   float64
 7   density               1599 non-null   float64
 8   pH                    1599 non-null   float64
 9   sulphates             1599 non-null   float64
 10  alcohol               1599 non-null   float64
 11  quality               1599 non-null   int64
dtypes: float64(11), int64(1)
memory usage: 150.0 KB
```

Fig 2: snapshot of information of red wine dataset.

```
Error matrix of red wine dataset for testing set using Artifical Neural Networks

[[418 122]
 [126 533]]
```

Fig3: A screenshot of the red wine dataset's error matrix for the training set using Artificial Neural Networks.

```
Error matrix of red wine dataset for testing set using Artifical Neural Networks

[[144  60]
 [ 48 148]]
```

Fig4: A screenshot of the red wine dataset's error matrix for the testing set using Artificial Neural Networks.

DOI : 10.36893.JK.2022.V12I01N01.540-546

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

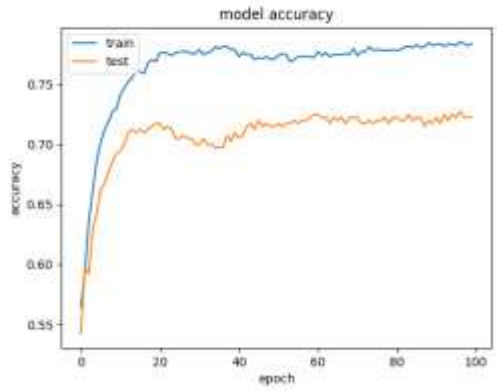


Fig 5: Snapshot of model accuracy of training and testing sets by increasing epochs in ANN.

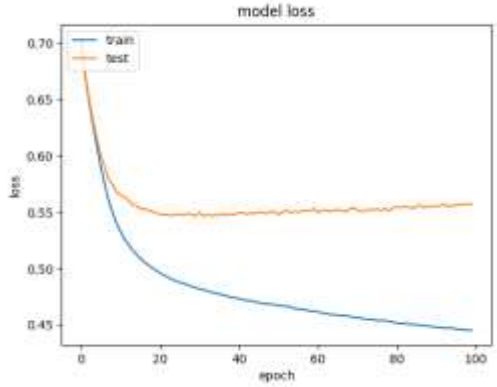
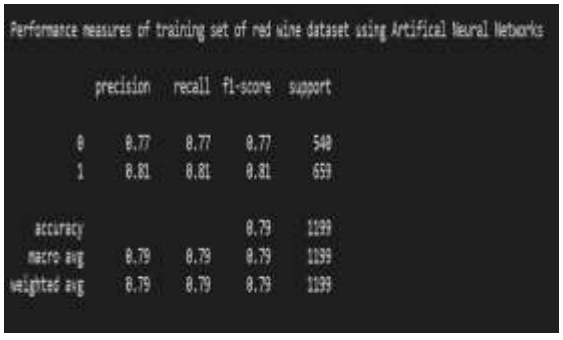
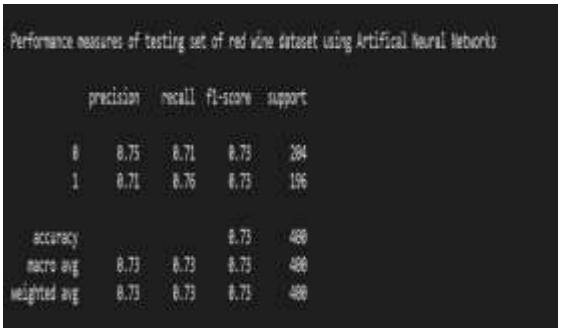


Fig 6: Snapshot of model loss of training and testing sets by increasing epochs in ANN.


Performance measures of training set of red wine dataset using Artifical Neural Networks

	precision	recall	f1-score	support
0	0.77	0.77	0.77	540
1	0.81	0.81	0.81	659
accuracy			0.79	1199
macro avg	0.79	0.79	0.79	1199
weighted avg	0.79	0.79	0.79	1199

Fig 7: shows the table of red wine dataset for training set using ANN.


Performance measures of testing set of red wine dataset using Artifical Neural Networks

	precision	recall	f1-score	support
0	0.75	0.71	0.73	284
1	0.71	0.76	0.73	196
accuracy			0.73	480
macro avg	0.73	0.73	0.73	480
weighted avg	0.73	0.73	0.73	480

Fig 8: shows the table of red wine dataset for testing set using ANN.

CONCLUSION

The wine industry's quality assessment is a critical aspect, and the industries are always modifying their feature combinations in order to attract more customers and meet their wants. As a result, the ANN model was used to predict wine quality. Changes feature combinations to predict wine quality. This investigation resolves accuracy, precision, and F-score. Because the training dataset and testing dataset both contain 70% and 30% of the data from the original dataset, the results show that the Artificial Neural Network gives an accuracy 73 percent when used to predict red wine quality and guides wine manufacturing companies in quality control prior to production. Better algorithms, which combine the greatest qualities of all current data mining techniques and enhance accuracy, can be created in the future

DOI : 10.36893.JK.2022.V12I01N01.540-546

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

REFERENCES

[1] 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.

[2] S. Ebeler, "Linking Flavour Chemistry to Sensory Analysis of Wine," in Flavor Chemistry, Thirty Years of Progress, Kluwer Academic Publishers, 1999, pp. 409-422.

[3] 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.

[4] A. Asuncion, and D. Newman (2007), UCI Machine Learning Repository, University of California, Irvine, [Online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.html

[5] S. Kallithraka, IS. Arvanitoyannis, P. Kefalas, A. El-Zajouli, E. Soufleros, and E. Psarra, "Instrumental and sensory analysis of Greek wines; implementation of principal component analysis (PCA) for classification according to geographical origin," Food Chemistry, 73(4): 501-514, 2001.

[6] N. H. Beltran, M. A. Duarte-MErmound, V. A. S. Vicencio, S. A. Salah, and M. A. Bustos, "Chilean wine classification using volatile organic compounds data obtained with a fast GC analyzer," Instrum. Measurement, IEEE Trans., 57: 2421-2436, 2008.

[7] S. Shanmuganathan, P. Sallis, and A. Narayanan, "Data mining techniques for modelling seasonal climate effects on grapevine yield and wine quality," IEEE International Conference on Computational Intelligence Communication Systems and Networks, pp. 82-89, July 2010.

[8] 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.

[9] 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.

[10] 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.

[11]UCI Machine Learning Repository, Wine quality data set, [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Wine+Quality.

[12]J. Han, M. Kamber, and J. Pei, "Classification: Advanced Methods," in Data Mining Concepts and Techniques, 3rd ed., Waltham, MA, USA: Morgan Kaufmann, 2012, pp. 393-443.

[13]W. L. Martinez, A. R. Martinez, "Supervised Learning" in Computational Statistics Handbook with MATLAB, 2nd ed., Boca Raton, FL, USA: Chapman & Hall/CRC, 2007, pp. 363-431.

DOI : 10.36893.JK.2022.V12I01N01.540-546