**Wine Quality Prediction**

**Machine Learning Volunteering Work**

**Done By:**

Kandimalla Revanth : BL.EN.U4CSE20067 , CSE-B

M.Mourya Vardhan Reddy : BL.EN.U4CSE20103 , CSE-B

Ulligaddla Nishitha : BL.EN.U4CSE20187 , CSE-C

**Abstract :**

In latest years, most of the industries promoting their products primarily based at the first-rate certification they obtained on the products. The traditional manner of assessing the product first-class is time ingesting, however with the invent of machine learning strategies the approaches has become extra green and consumed less time than earlier.

In this paper we've explored, several the machine gaining knowledge of strategies to evaluate the nice of wine based at the attributes of wine that depends on quality. We have used different wines dataset for this work. We have used one of a kind characteristic choice approach which includes genetic algorithm (GA) based totally characteristic selection and simulated annealing (SA) primarily based function selection to check the prediction overall performance.

We have used unique overall performance measure together with accuracy, sensitivity, specificity, superb predictive value, negative predictive value for comparison the usage of exclusive feature sets and one-of-a-kind supervised system getting to know techniques. We have used nonlinear, linear, and probabilistic classifiers. We have found that function choice-based totally function units able to provide higher prediction than thinking about all the capabilities for overall performance prediction. We have different accuracies got by using different classifiers and regressors ranging between 95% to 100%. By using Tree graph regressor we got the greatest accuracy among other which is 1.0.

**Introduction:**

In recent years there is a modest increase in the wine consumption as it has been found that wine consumption has a positive correlation to the heart rate variability [1]. Different wines have special functions. Although most of the chemical compounds are same for specific form of wine based on the chemical exams, the quantity of all chemical compounds has extraordinary level of attention for unique kind of wine. These days it is important to classify different wine for quality assurance [2]. It is difficult now a days to predict the wine has good quality or not. If we want to do that in the real time, there is a need of lot money for testing.

We have a solution for that by using machine learning. Now a days machine learning is vastly used for these kinds of predictions which has equal accuracy with the real time testing. This will cost us very less as compared. The performance comparison with different feature sets will also help to classify it in a more distinctive way. By using different linear, non-linear and probabilistic classifiers and regressors are used for prediction.

**Literature Survey :**

In the past few attempts have been made to use different machine learning methods and feature selection techniques to the wine prediction dataset. Chen et al proposed an approach that will predict the grade of wine using the human savoury reviews. They have used hierarchical clustering approach and association rule algorithm to process the reviews and predict the wine grade and they found an accuracy of 85.25% while predicting the grade [3]. Er and Atasoy proposed a method to classify the quality of wines using three different classifier such as support vector machines, Random forest and k-nearest neighbourhood.

They have used principal component analysis for feature selection and they found good result using Random forest algorithm [4]. Appalasamy et al proposed a method to predict wine quality based on physiochemical test data. They have pointed out that classification approach helps to improve the quality of wine during the production [5]. Reddy and Govindarajulu used a user centric clustering approach to recommend the product. They have used red wine data set for the survey purpose. They have allocated relative voting to the attributes based on the literature review. Then they assigned weight to the attributes using Gaussian Distribution Process. They judged the quality based on the user preference group [6]. Thakkar et al., used analytical hierarchy process (ahp) to rank the attributes and then used different machine learning classifiers such as support vector machine and random forest and they found accuracy of 70.33% using random forest and 66.54% using SVM [7]. Beltrán et al proposed an approach to classify the wine based on aroma chromatograms and they have used PCA for dimensionality reduction and wavelet transform for feature extraction and classifiers such as neural network, linear discriminant analysis and support vector machine and found that support vector machine with wavelet transforms perform better than other classifiers [8].

These above works motivated us to do the different kind of classification on this wine prediction using various other and advanced methods of machine learning and improve the model with more accuracy.

**Methodology and Result Analysis :**

**Architecture :**

Diagram

Description automatically generated

The basic architecture begins with

* collecting the data from various sources and finding the best one suitable for our model.
* Then we preprocess the data using some ML models to work on the data easily.
* Feature selection fo visualizing the data is done and we select the Model and train the model .
* We use some models like KNN,SVM,Tree Graph for the accuracy and evaluate the training and testing values.Then our model is ready for the prediction.

**Sample Collection:**

Our dataset has a wide range of features which help in determining the quality of the wine.These are the already given features where we work on them to predict the quality

A picture containing crossword puzzle, shoji, scoreboard, clipart

Description automatically generated

This image shows plots of various features which help in predicting the quality of the wine.

**Analysis :**

|  |  |  |  |
| --- | --- | --- | --- |
| ML Model | Score | Accuracy | Time Taken |
| KNN | 0.93375 | 0.92711 | 0.04153 |
| SVM | 0.8675 | 0.927113 | 0.0404539 |
| MLP | 0.89125 | 0.889212 | 1.075014 |
| Gaussian NB | 0.845 | 0.8513119 | 0.004964 |
| Linear Regression | 0.237479 | -- | 0.02952 |
| Tree regressor | 1 | 0.8513 | 0.0078866 |
| Logistic Regression | 0.8725 | 0.868804 | 0.15313744 |

There fore from the values of the scores we got for various regression models we can conclude that Tree Graph Regressor is the best way as we got the maximum score there.

A screenshot of a computer

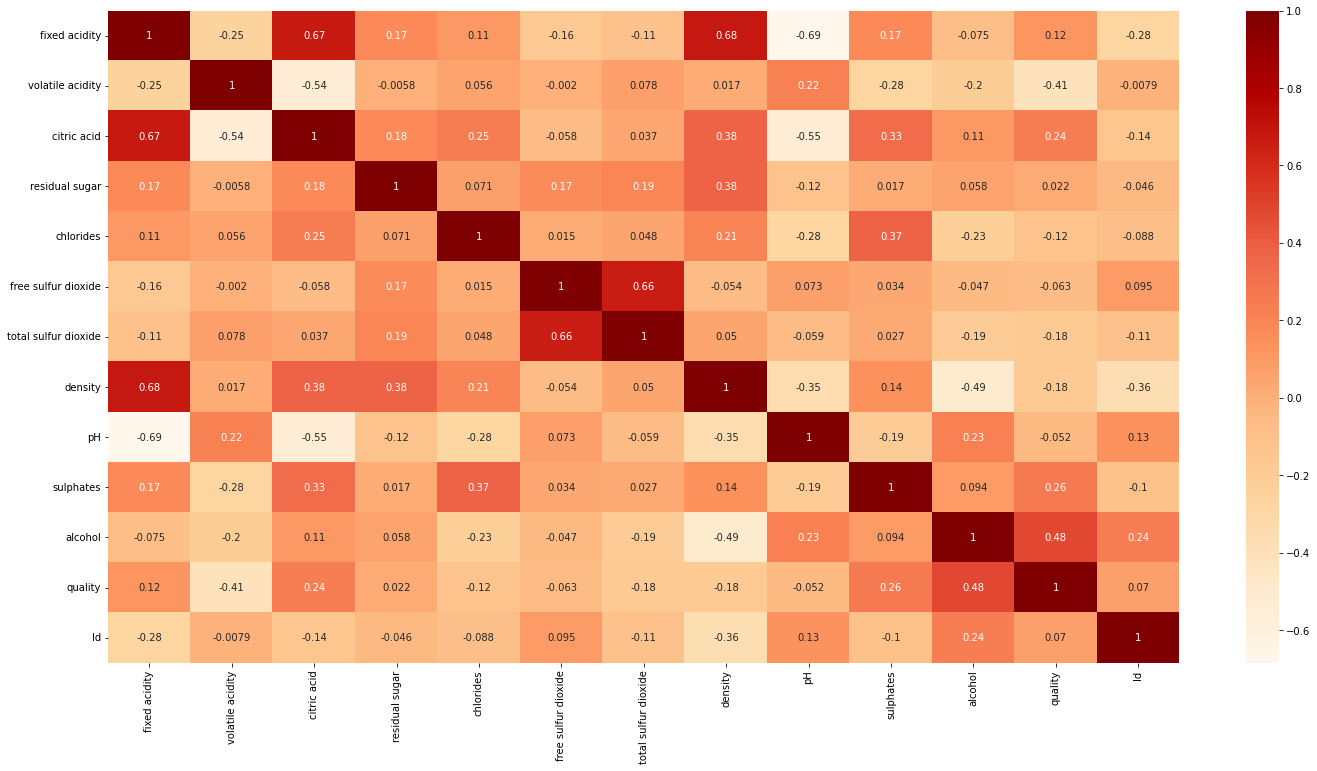
Description automatically generated with medium confidence

Graphical user interface, application

Description automatically generated

**Visualization :**

Heatmap :



Chart, bar chart

Description automatically generated Chart, bar chart

Description automatically generated

The plots show the visualizations of quality vs count and quality vs volatile acidity

**Conclusion :**

This paper mentioned approximately capability of genetic set of rules as well as simulated annealing set of rules for characteristic choice in addition to the potentials of the classifiers to predict as it should be based on the brand new function units. The characteristic selection set of rules provided a clear concept approximately the importance of the attributes for prediction of exceptional, which became time consuming and high-priced whilst done inside the traditional way. We have additionally as compared the performance metrics of linear, nonlinear, and probabilistic based totally classifiers and it turned into discovered that these classifiers completed properly with the brand new feature units.

We have determined that the SA based totally function units done better than the GA based totally function units. We have additionally discovered that the SVM classifier finished higher in comparison to all other classifiers for crimson wine and white wine statistics units. We have found accuracy starting from 95.23% to 100% with extraordinary feature sets. In future we will attempt other overall performance measures and other gadget learning techniques for better assessment on consequences. This analysis will assist the industries to predict the satisfactory of the one-of-a-kind type of wines based on positive attributes and also it'll beneficial for them to make properly product inside the destiny.

**References :**

[1] I.Janszky, M.Ericson, M.Blom, A. Georgiades, J.O.Magnusson, H.Alinagizadeh, and S.Ahnve, “Wine drinking is associated with increased heart rate variability in women with coronary heart disease,” Heart, 91(3), pp.314-318,2005.

[2] 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.

[3] Y.Er, and A.Atasoy, “The Classification of White Wine and Red Wine According to Their Physicochemical Qualities,”International Journal of Intelligent Systems and Applications in Engineering,4,pp.23-26,2016.

[4] 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.

[5] P.Appalasamy, A.Mustapha, N.D.Rizal, F.Johari, and A.F.Mansor, “Classification-based Data Mining Approach for Quality Control in Wine Production,” Journal of Applied Sciences, 12(6), pp.598-601,2012

[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] K.Thakkar,J.Shah,R.Prabhakar,A.Narayan,A.Joshi, “AHP and MACHINE LEARNING TECHNIQUES for Wine Recommendations,” International Journal of Computer Science and Information Technologies,7(5),pp. 2349-2352 ,2016

[8] Reddy, Y. S., & Govindarajulu, P. (2017). An Efficient User Centric Clustering Approach for Product Recommendation Based on Majority Voting: A Case Study on Wine Data Set. IJCSNS, 17(10), 103.

[9] M.Forina, R. Leardi, C. Armanino, and S. Lanteri, "PARVUS An Extendible Package for Data Exploration," Classification and Correla,1988.

[10] Bledsoe, W. W. (1961). The use of biological concepts in the analytical study of systems. In the ORSA-TIMS National Meeting