Wine Vinification Disposition Prophecy Using Machine Learning Methodology

[1] Divya Reji [2] Pauline Paul

[1] PG Scholar, Amal Jyothi College of Engineering, Kanjirapally, Kerala

[2] Asst. Professor, Amal Jyothi College of Engineering, Kanjirapally, Kerala

[1] divyareji98@gmail.com , [2] paulinepaul@amaljyothi.ac.in

Abstract: **Wine: one of the most sophisticated and posh drinks has posed itself as a favorite among all of us humans. But when it comes to its quality, we get cautious. Here there exists a wine-maker who wishes to launch a new wine. To predict its quality, we dive into its chemical factors. For that matter, we have few factors like acetic acid content, sweetness etc. to know the quality of our sample. As a solution, we have created a machine learning model using random forest approach to predict whether the quality is worth it or not.**

Keywords: Wine Quality, Wine Vinification, Random Forest, Machine Learning, Quality of wine, Wine Quality Prediction

I. FOREWORD

Maynard Amerine once said that, it is easier to detect than define quality of wine. There are many types of wine in the world. Grape wine is the most exquisite one. It is an alcoholic drink made from fermented grapes. Yeast combined with sugar present in the grapes and transforms it to ethanol and carbon dioxide releasing heat in the process alongside. Different types of grapes and types of yeasts are major contributors in different styles of wine produced. Apart from tasting, one can know the quality via its chemical properties too.

Machine Learning plays a crucial role here. It is one of the most promising fields. It takes into account a relating dataset, apply few of the machine learning algorithms to train a model, test the model and then acquire the accuracy score which gives the solution of the problem.

(a) Necessities:

- Google Colab Notebook
- Kaggle dataset
- Suitable Algorithm

(b) RANDOM FOREST:

Random Forest is a popular machine learning algorithm that comes under supervised learning. It is used for both Classification and Regression problems. Based on the concept of ensemble learning, it yields a accuracy score to provide the solution of the problem.

VERSITALITY:

- It ensures best accuracy, robustness and ease of use than other algorithms.
- Provides two straightforward methods for feature selection viz, mean decrease impurity and mean decrease accuracy.
- It prevents the problem of overfitting.

II. LITERATURE RESEARCH

[1] This paper proposes an assessment using the least absolute shrinkage and LASSO and random forest. Our data analysis is based on a real wine dataset provided by a wine firm in Greece. For this purpose, we employ the LASSO method. Additionally, the random forest method is used and its findings are taken consideration with those derived by four different methods, LDA, CART, k-nearest neighbours (kNN) and support vector machines, and ten-fold cross-validation method. The results of our analysis show that the statistical technique of this algorithm actually improves the accuracy of the prediction wine quality, up to almost 95%.

[2] This paper allows to build a model with user interface which predicts the wine quality. Random forest algorithm is used in determining wines' quality whose correctness would further be escalated using KNN which makes our model dynamic. Result

DOI: 10.5281/zenodo.6369829
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

of this proposed model is used to determine the wines' quality on a scale of Good, Average or Bad.

[3] In this paper, a hybrid model consisting of two classifiers at least, e.g., RF and SVM, is proposed for wine quality prediction. To evaluate the performance of the proposed hybrid model, experiments also made on the wine datasets to show the merits of the hybrid model.

III. PROCEDURE

In this prediction system, we use Google Colab, a Jupyter notebook that runs in the cloud, making it easy to set up, access, and share. It helps visualize maps, charts using third-party Python packages. It helps serve modelling to develop a model and acquire solution to the problem.

Figure 1: Google Colab

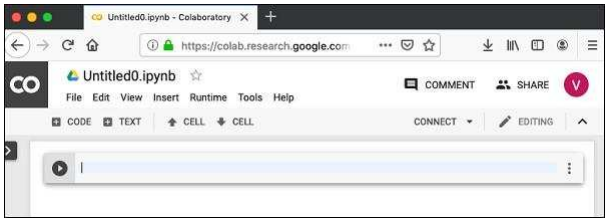

Figure 2: Colab Notebook First Look

(a) RECORDS:

It is group of data that will be used for prediction of the quality of wine. Every column represents a chemical factor of the sample wine and each row exhibits a value from the data set to analyse wine quality. The data is a csv format which shows the dataset of the quality prediction of wine.

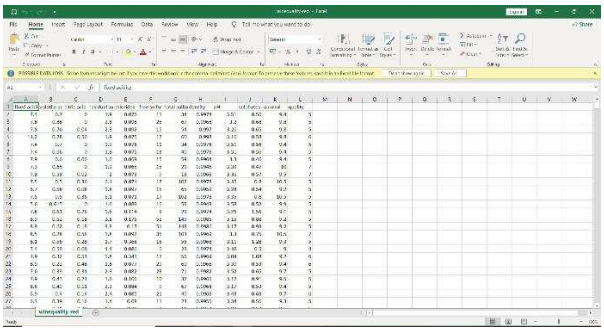

Figure 3: Dataset

IV. ENFORCEMENT

Firstly, open your Google Colab notebook through your google email ID and then mount your drive to load the dataset into the book. For mounting the drive, we need to write two lines of code:

from google.colab import drive

drive.mount('/content/drive')

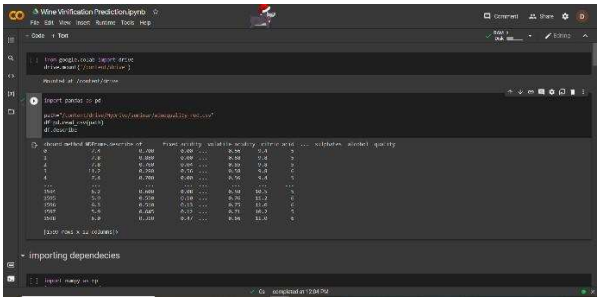

Figure 4: Imported the dataset into Colab

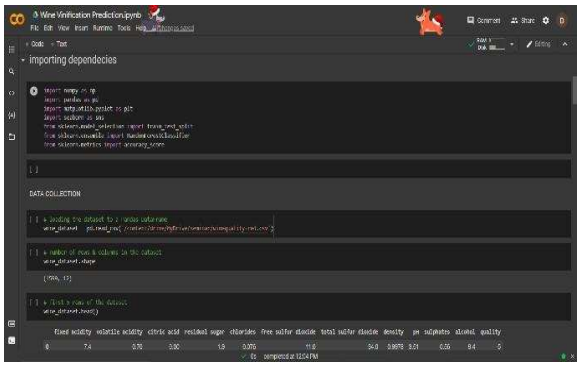

Figure 5: Imported the required dependencies

DOI: 10.5281/zenodo.6369829
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

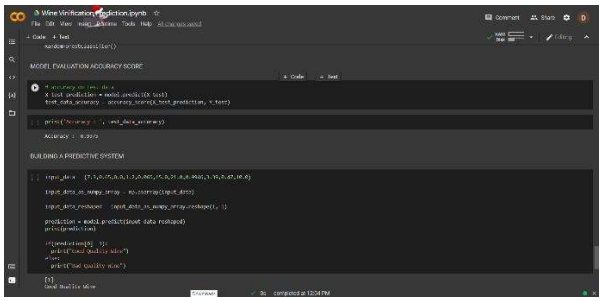

Figure 6: Predicting the quality

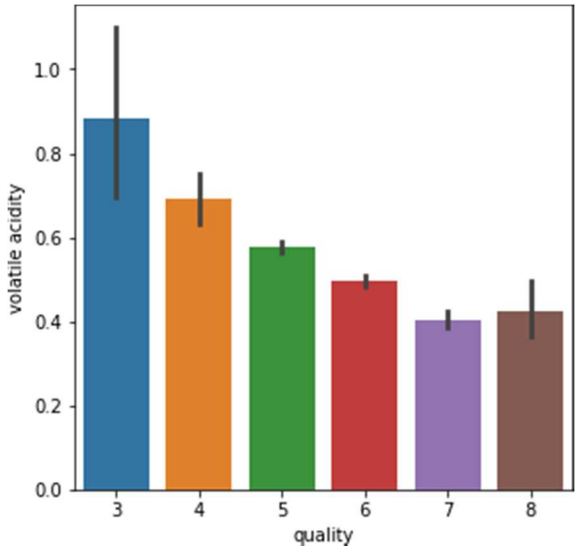

Figure 7: Quality VS Volatile Acidity of the Wine

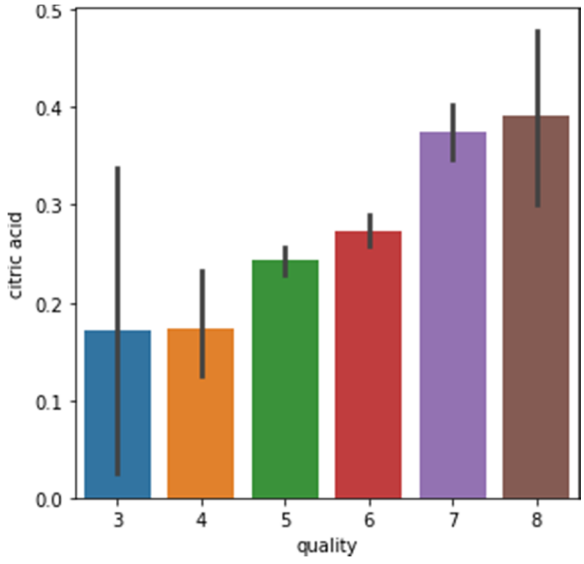

Figure 8: Quality VS Citric Acidity of the Wine

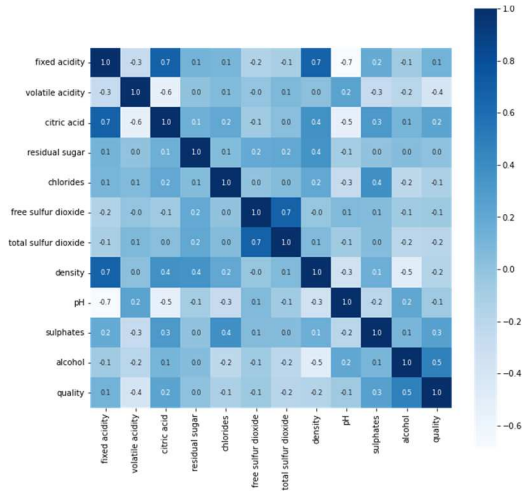

Figure 9: Heatmap describing the various chemical compositions in the wine

V. OUTCOMES

Through the above execution, we now can evidently see that our model has given the accuracy score of 93.75%, stating that it will be a good fit for determining the disposition of wine. When an input data set from the test data set was tested using this model, it showed the result to be a good quality indicating the company produced a good production.

VI. SYNOPSIS

In recent time, people's interests have also increased in the wine industry, which itself is an add-on for the demanded growth in this market. The above marketers are also adopting creative and innovative ways to increase the wine quality and therefore increase its sales. Therefore, wine quality certification plays a prominent role in all processes in this area, and it definitely needs human help to test wine. This paper goes through the use Random Forest algorithm- a classifier algorithm. Implemented using this dataset, its output is taken. Results showed that using this procedure, through experimenting these 1599 samples in-total comprising of 11 chemical composition factors. This work deduces the routine manual work of finding it and thereby avoiding any errors while executing it.

DOI: 10.5281/zenodo.6369829
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

VII. FUTURE ENRICHMENT

For future improvements, a broader data set may be used for such work and other machine learning techniques may be explored. Moreover, it can also be expanded from this analysis to include feature development methods.

VIII. CITATIONS

[1] S. Aich, A. A. Al-Absi, K. Lee Hui and M. Sain, "Prediction of Quality for Different Type of Wine based on Different Feature Sets Using Supervised Machine Learning Techniques," 2019 21st International Conference on Advanced Communication Technology (ICACT), 2019, pp. 1122-1127.

[2] Mahima, Gupta U., Patidar Y., Agarwal A., Singh K.P. (2020) Wine Quality Analysis Using Machine Learning Algorithms. In: Sharma D.K., Balas V.E., Son L.H., Sharma R., Cengiz K. (eds).

[3] Chiu T.HY., Wu CW., Chen CH. (2021) A Hybrid Wine Classification Model for Quality Prediction. In: Del Bimbo A. et al. (eds) Pattern Recognition. ICPR International Workshops and Challenges. ICPR 2021.

[4] Er Y (2016) The classification of white wine and red wine according to their physicochemical qualities. Int J Intell Syst Appl Eng 4(1):23–26

[5] S. Kumar, K. Agrawal and N. Mandan, "Red Wine Quality Prediction Using Machine Learning Techniques," 2020 International Conference on Computer Communication and Informatics (ICCCI), 2020, pp. 1-6.

DOI: 10.5281/zenodo.6369829
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam