JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Review on Wine Quality Testing using Machine Learning

Prof.Swati R. Khokale, Bhushan Bisane, Diksha Bharambe, Divya Bhamare, Harshada Pawar

***Abstract*-Wine classification is a difficult task since taste because it is the least understood to the human senses, wine classification is challenging. Because sensory analysis is currently performed by human tasters, a subjective method, a good wine quality forecast can be very valuable in the certification phase. An automatic predictive system can be incorporated into a decision support system to improve performance speed and quality. A feature selection method can also aid in analysing the impact of analytical tests. If it is determined that certain input variables are very relevant in predicting wine quality, this information can be used to improve wine quality because some variables can be modified throughout the production process.**

Keywords – Machine Learning, SVM, Random ForestClassifier, K-means, Matplotlib, sk-learn, pandas, wine quality, data analysis.

1.Introduction

The quality of wine is extremely essential to both consumers and the manufacturing industries. Product quality certification is helping businesses increase revenue. Nowadays, wine is a widely consumed beverage all over the world, and enterprises rely on product quality certification to boost their market worth. Previously, product quality testing was done at the conclusion of the manufacturing process, which is a time consuming procedure that necessitates a lot of resources, such as the requirement for numerous human experts to assess product quality, making this process quite expensive. Every person has an opinion about the test, so determining the wine's quality based on human specialists is a difficult undertaking. There are various points that can be used to forecast wine quality, however not all of them are relevant for improved prediction. Wine is an alcoholic beverage prepared from grapes that has been fermented without the addition of sugars, acids, enzymes, water, or other nutrients. The sugar in the grapes is consumed by yeast, which turns it to ethanol and carbon dioxide. Distinct grape varietals and yeast strains make different styles of wine. These variances are the result of complex interactions between the grape's biochemical development, fermentation reactions, terroir, and the manufacturing method. Rice wine and fruit wines such as plum, cherry, pomegranate, and elderberry are examples of non-grape wines. Since many years, wine has been made. One of the most important aspects in this context is wine quality assessment, which can be utilised for certification. This form of quality certification aids in ensuring the quality of wine on the market. Wine characteristics will determine the quality of the wines. With the proliferation of wine labels in recent years, it has become increasingly difficult to distinguish between high-quality wines. So many significant aspects, such as chemical, scientific, and technical factors, go into making good wine. Machine learning approaches have gotten a lot of attention in the last few years in almost every industry. Most machine learning algorithms are capable of producing very accurate results, prompting many data scientists to use them in predictive analytics. Here we considered only red wine and white wine analysis. Red wine is made from dark coloured grape varieties. The actual colour of the wine can differ from violet, typical of young wines, through red for aged wines, to brown for older red wines. Most purple grape juice is actually greenish-white; the red colour derives from anthocyan pigments contained in the grape's skin; exceptions are the relatively rare teinturier kinds, which have red meat and generate red juice. White wine is made by fermenting the non-colored grape pulp. The grapes used to make white wine are usually green or yellow in color. White wine can be made from dark-skinned grapes if the winemaker is cautious not to let the skin taint the wort during the pulp-juice separation. One of the most difficult aspects of analysing wine kind and quality is achieving absolute accuracy in a short amount of time. This can be

accomplished through the use of machine learning algorithms.

2.Literature Survey

This paper examines industry-based doctoral dissertation research in a professional computing doctoral program for full time working professionals through the lens of different machine learning algorithms. This paper examines industry-based doctoral dissertation research in a professional computing doctoral program for full time working professionals through the lens of different machine learning algorithms.The research provides insights into differences in machine learning algorithm categorization. [1].

There have been many studies in which researchers have attempted to classify student attentiveness. Many of these approaches depended on a qualitative analysis and lacked any quantitative analysis.Focused on bridging the gap between qualitative and quantitative approaches to classify student attentiveness.Focused on bridging the gap between qualitative and quantitative approaches to classify student attentiveness. [2]

As a subfield of Artificial Intelligence (AI), Machine Learning (ML) aims to understand the structure of the data and fit it into models, which later can be used in unseen data to achieve the desired task. ML has been widely used in various sectors such as in Businesses, Medicine, Astrophysics, and many other scientific problems. Inspired by the success of ML in different sectors, here, we use it to predict the wine quality based on the various parameters. [1] Among various ML models, we compare the performance of Ridge Regression (RR), Support Vector Machine (SVM), Gradient Boosting Regressor (GBR), and multi-layer Artificial Neural Network (ANN) to predict the wine quality. Multiple parameters that determine the wine quality are analyzed. Our analysis shows that GBR surpasses all other models' performance with MSE, R, and MAPE of 0.3741, 0.6057, and 0.0873 respectively. This work demonstrates, how statistical analysis can be used to identify the components that mainly control the wine quality prior to the production. This will help wine manufacturer to control the quality prior to the wine production.[3]

Certification and quality assessment are crucial issues within the wine industry. Currently, wine quality is mostly assessed by physicochemical (e.g alcohol levels) and sensory (e.g. human expert evaluation) tests. In this paper, we propose a data mining approach to predict wine preferences that is based on easily available analytical tests at the certification step. A large dataset is considered with white vinho verde samples from the Minho region of Portugal. Wine quality is modeled under a regression approach[1], which preserves the order of the grades. Explanatory knowledge is given in terms of a sensitivity analysis, which measures the response changes when a given input variable is varied through its domain. Three regression techniques were applied, under a computationally efficient procedure that performs simultaneous variable and model selection and that is guided by the sensitivity analysis. The support vector machine achieved promising results, outperforming the multiple regression and neural network methods. Such model is useful for understanding how physicochemical tests affect the sensory preferences. Moreover, it can support the wine expert evaluations and ultimately improve the production. [4]

According to experts, wine quality is checked with its smell, flavor and color but we are not a wine experts. Here's the use of Machine Learning comes . In this article , we will focus on Wine Quality Prediction on the basis of given features. Also every industry need to prove product quality to promote their product so quality check is important. Firstly, we import necessary library for this model. Numpy will be used for making the mathematical calculations more accurate, pandas will be used to work with file formats like csv, xls etc. and sklearn (scikit-learn) will be used to import our classifier for prediction.from sklearn.model_selection import train_test_split is used to split our dataset into training and testing data. from sklearn import preprocessing is used to preprocess the data before fitting into predictor, [2] or converting it to a range of -1,1, which is easy to understand for the Machine Learning Algorithms. from sklearn import tree is used to import our decision tree classifier, which we will be using for prediction. HarvardX's Data Science Professional Certificate program covers several steps in a data science project, such as data wrangling, data exploration and visualization, probability and statistics, R language, Rmarkdown, and machine learning. This capstone project briefly applies each of these concepts in a real world case study. The School of Information and Computer Science at the University of California Irvine (UCI) maintains a machine learning repository used by the machine learning community for analysis of algorithms. This project uses the wine quality data set of white and red wines of vinho verde. The literal translation for vinho verde is "green wine", which means "young wine" with wine being released three to six months after the grapes are harvested.2 Vinho verde is a reference to the wine produced in the northern region of Portugal and follows the regulations set by the Comissão de Viticultura da Região dos Vinhos Verdes ("Wine Commission of the Vinho Verde Region"). [5]

We use a deep learning method, restricted Boltzmann machine, for nonlinear system identification. The neural model has deep architecture and is generated by a random

search method. The initial weights of this deep neural model are obtained from the restricted Boltzmann machines. To identify nonlinear systems, we propose special unsupervised learning methods with input data. The normal supervised learning is used to train the weights with the output data.[6]

Recent developments in information systems as well as computerization of business processes by organizations have led to a faster, easier and more accurate data analysis. Recent developments in information systems as well as computerization of business processes by organizations have led to a faster, easier and more accurate data analysis. Machine learning techniques make it possible to deduct meaningful further information from those data processed by data mining. [7]

By analyzing the physicochemical tests samples data of red wines from the north of Portugal, I was able to create a model that can help industry producers, distributors, and sellers predict the quality of red wine products and have a better understanding of each critical and up-to-date features. I have found that the Model 3 — Random Forest-based feature sets performed better than others. In general, using Model 3 as our best model for prediction, I determined four of the features as the most influential: volatile acidity, citric acid, sulphates, and alcohol. To be more specific, high-quality wines seem to have lower volatile acidity, higher alcohol, and medium-high sulphate values. Meanwhile, lower-quality wines tend to have low values for citric acid.[8]

3.Proposed System

The structure and behaviour of a system is defined by the system architecture, which is a conceptual design. An architecture description is a formal description of a system that is organised in a way that allows for reasoning about the system's structural attributes. It specifies the system components or building blocks and provides a framework for procuring products and developing systems that will operate together to implement the overall system. Kaggle was used to collect the data. Data that is fed into the machine learning module. The wine data set is used, and it is then preprocessed and converted. Apply several techniques to the ruduced data collection. The algorithm Random Forest is utilised.The trained module is now ready for output and prediction.

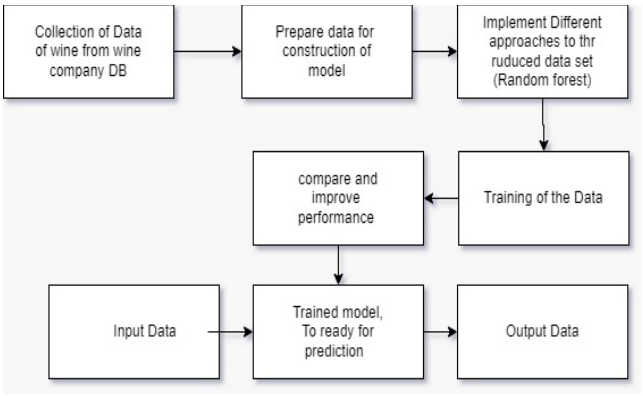

Fig. Proposed System

4. Conclusion

Based on the bar graphs shown, we conclude that not all input features are necessary and affect the data; for example, the bar plot versus quality and residual sugar shows that residual sugar is mild and does not vary dramatically as quality increases. So, because this characteristic isn't as important as others like alcohol and citric acid, we can skip it during feature selection. We used several algorithms to classify the wine quality, including 1) Random forest 2) SVM 3)Decision tree.We were able to attain a maximum accuracy of 88 percent utilising random forest. SVC is 83 percent accurate, whereas decision tree is 77 percent accurate.

References

[1] P. iCortez, iA. iCerdeira, iF. iAlmeida, iT. iMatos iand iJ. iReis. iModelling iwine ipreferences iby idata imining fromiphysicochemical iproperties. In iDecision iSupport iSystems, iElsevier, i47(4):547-553, I 2020

[2] https://archive.ics.uci.edu/ml/datasets/wine+quality
[3]https://www.researchgate.net/publication/221612614_Using_Data_Mining_for_ Wine_Quality_Assessment

[4] B. Chen, C. Rhodes, A. Crawford, and L. Hambuchen,“Wine informatics: applying data mining on wine sensory reviews processed by the computational wine wheel,” IEEE International Conference on Data Mining Workshop, pp. 142- 149, Dec. 2019.

[5] 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 ,2019

[6] 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.

[7] P.Appalasamy and A. Mustapha,(2012) "Classification based Data Mining Approach for Quality Control in Wine Production", Volume 12 (6), Journal of Applied Sciences.

[8] Shen Yin,(2019) "Research Article: Quality Evaluation Based on Multivariate Statistical Methods", Article ID 639652, 10 pages.

[9] DimitrijaAngelkov,(2018) "Data mining of wine quality", Volume 45, August 31.

[10] Y. Subba Reddy and Prof. P. Govindarajulu,(2017),"An Efficient User Centric Clustering Approach for Product Recommendation Based on Majority Voting: A Case Study on Wine Data Set", IJCSNS International Journal of Computer Science and Network Security, VOL.17 No.10.

[11] Bernad Chen,(2018) ―Wine Informatics-Association rule based classification‖, 11th December 2016. Miami, FL, USA.

[12] Yesim Er and AytenAtasoy,(2016) "The Classification of White Wine and Red Wine According to Their Physicochemical Qualities".

[13] Zhang Lingfeng and Feng Feng,(2017) "Wine quality identification based on data mining research",12th International Conference on Computer Science and Education (ICCSE).

[14] Nao Wariishi and Brendan Flanagan,(2015) "Sentiment Analysis of Wine Aroma", 4th International Congress on Advanced Applied Informatics.

[15] Yogesh Gupta,(2017) "Selection of important features and predicting wine quality using machine learning techniques", Procedia Computer Science, vol.125.

[16] AytenAtasoy,(2017) ―Classification of white wine and red wine according to their Physicochemical Qualities.

[17] Yunhui Zeng1 , Yingxia Liu1 , Lubin Wu1 , Hanjiang Dong1. "Evaluation and Analysis Model of Wine Quality Based on Mathematical Model ISSN 2330- 2038 E-ISSN 2330-2046,Jinan University, Zhuhai,China.

[18] Paulo Cortez1, Juliana Teixeira1, Ant´onio Cerdeira2."Using Data Mining for Wine Quality Assessment".

[19] Yesim Er*1 , Ayten Atasoy1. "The Classification of White Wine and Red Wine According to Their Physicochemical Qualities",ISSN 2147-67992147-6799,3rd September 2016.

[20] Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J., 2009. Modeling wine preferences by data mining from physicochemical properties. Decis. Support Syst. 47, 547–553. https://doi.org/10.1016/j.dss.2009.05.016