Bhavya AG. 2020, 8:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Wine Quality Prediction Using Different Machine Learning Techniques

Bhavya AG
Department of Computer Science and Engineering
Cambridge Institute of Technology
K.R. Puram Bangalore,India
bhavyaga59@gmail.com

***Abstract*- In recent years, most of the industries promoting their products supported the standard certification they received the products. The normal manner of assessing product quality is time overwhelming, but with the invention of machine learning techniques, the processes has become a lot of economical and consumed less time than before. In this paper we've explored, a number of machine learning techniques to assess the standard of wine supported the attributes of wine that depend on quality. We've used white wine and red wine quality dataset for this analysis. We've used a completely different feature choice technique like a genetic algorithmic rule (GA) based feature selection and simulated annealing (SA) based feature choice to envision the prediction performance. We've used completely different performance measure such as accuracy, sensitivity, specificity, positive predictive value, negative predictive value for comparison using different feature sets, and different supervised machine learning techniques. We've got used nonlinear, linear, and probabilistic classifiers. We've got found that feature selection-based feature sets able to provide better prediction than considering all the features for performance prediction.**

***Keywords*- machine learning; feature selection; naïve bayes; svm; accuracy; wine quality.**

I. INTRODUCTION

The intrinsic characteristics (visual, taste, smell), environmental characteristics (climate, region, site), and management practices (viticulture practice), additionally physicochemical ingredients (acid, pH, etc.) are the factors of interest in assessing the standard of Wine. Data processing techniques in predicting wine quality are in progress, with some promising leads to the domain. Physicochemical and sensory tests are crucial in Wine certification. It's the routine follow in physicochemical laboratory tests, to characterize wine by determination of density, alcohol, or pH values, however, sensory tests believe mainly on human experts. Wine classification is a difficult task as taste is the least understood of the human senses. The link between physicochemical and sensory analysis are advanced to grasp. Within the food business, additionally, to the food quality analysis, machine learning techniques have conjointly been applied in the classification of wine quality. Machine learning methods provide a way to build models from the information of known category labels to predict the standard of a wine. In previous days Wine was considered a luxury item. Today, it's standard and enjoyed by a large form of individuals. Professional wine reviews offer insights on wines available in large quantities annually. A scientific manner is required to utilize that large amount of reviews to profit wine customers, distributors, and makers. No two persons choose the wine alike even they taste the wine at the same time whereas having the ability to share and find all the identical attributes. Experience helps a lot and hinders the

International Journal of Science, Engineering and Technology

An Open Access Journal

taster. Thus assessing the quality of wine relies only on the taster's experience and sense is a big process. In recent years there's a modest increase within wine consumption because it has been found that wine consumption has a positive correlation to the heart rate variability [1]. With the rise within the consumption, wine industries are trying to find alternatives to provide good quality wine at less cost. Different wines have different functions.

Though most of the chemicals are the same for various types of wine supported on the chemical tests, the number of every chemical has a completely different level of concentration for various types of wine. Currently, it's extremely vital to classify different wines for quality assurance [2]. Within the overdue tp lack of technological resources, it becomes difficult for most industries to classify the wines based on chemical analysis as it takes a lot of time and also needs more money. These days with the advent of the machine learning techniques it is possible to classify the wines as well as it is possible to figure out the importance of each chemical analysis parameter in the wine and which one to ignore for reduction of cost.

II. RELATED WORKS

In the past few attempts have been made to use different machine learning approaches and feature selection techniques to the wine dataset. Er and Atasoy et al. [3] proposed a method to classify the quality of wines using two different classifiers such as support vector machines, naive Bayes.

They have used principal component analysis for feature selection and they found good results using the Random forest algorithm. Chen et al,[4] proposed an approach that will predict the grade of wine using the human savory reviews. They have used a 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. Appalasamy et al. [5]. 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. Beltrán et al.[6] 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. Thakkar et al.[7], used the 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 an accuracy of 70.33% using random forest and 66.54% using SVM.

III. SYSTEM ARCHITECTURE

The system architecture is the conceptual design that defines the structure and behaviour of a system. An architecture description is a formal description of a system, organized in a way that supports reasoning concerning the structural properties of the system. It defines the system parts or building blocks and provides an idea from that products may be procured, and systems developed, which will work along to implement the overall system. The System architecture is shown below.

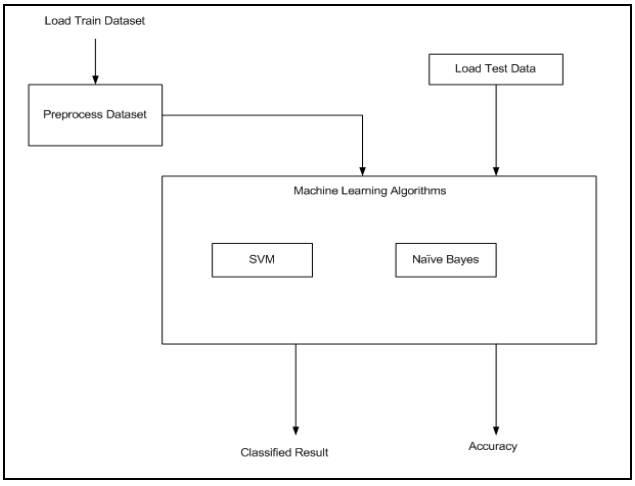

Fig. 1 System Architecture.

The input Design is the method of converting the user-oriented inputs into the computer-based format. The goal of designing input data is to form automation as easy and free from errors as possible. Providing a good input design for the application easy data input and selection features are adopted. The input design requirements like user-friendliness, consistent format and interactive dialogue for giving the correct message and facilitate for the user at right time are also considered for the development of the project. Input design is a part of overall system design which needs very careful attention. Usually, the gathering of input data is the most costly a part

International Journal of Science, Engineering and Technology

An Open Access Journal

of the system, that has to be the route through a variety of modules. It is the purpose wherever the user able to send the information to the destination machine along with known IP address; if the IP address is unknown then it may risk to error. Quality output is one, that meets the requirements of the end-user and presents the information clearly. In any system results of processing are communicated to the users and to other systems through outputs. It's the most significant and direct source of information to the user. Efficient and intelligent output improves the system's relationship with source and destination machine. Outputs from computers are required primarily to get the same packet that the user has sent instead of the corrupted packet and spoofed packets. They are also used to provide a permanent copy of these results for later consultation.

1. Existing System

In Existing, the algorithm went with the Decision Tree, KNN approach for wine quality prediction.

2. Proposed System

In a proposed system, an intelligent approach is projected by considering genetic algorithm (GA) based feature selection yet as simulated annealing-based feature selection considering the nonlinear classifiers, linear classifiers, and probabilistic classifiers to predict the quality in red wine yet as the white wine. Proposed with Random forest and SVM Classifiers gives the best result compared to the existing model.

3. Algorithms

3.1. Naïve Bayes algorithm

Naïve Bayes model is simple to make and significantly helpful for terribly giant information sets. Alongside simplicity, Naive Bayes is understood to beat even extremely subtle classification ways.

3.2. SVM Classifier Algorithm

SVM is a statistical learning-based solver. Statistical is mathematics of uncertainty. It aims at gaining knowledge, making decisions from a set of data. If we plot the data about the X and Y axis and to classify it, we see that there are many hyperplanes that can classify it. But to choose which one is the best or correct solution is a very difficult task. For removing this type of problem SVM used. There are many linear classifiers or hyperplanes that separate the data, but only to choose those which one has a maximum margin. The reason behind this if we use a hyperplanes for classification it might be closer to one data set compared to other. So we use the concept of maximum margin classifier or hyperplanes.

IV. RESULTS AND DISCUSSION

We have divided the data into two groups such as train data and test data. we tend to trained every classifier supported the trained data and predict the facility of classifier on the test data. So, each classifier ready to show all the performance metrics like accuracy, supported the test data. we've applied all the classification techniques to the genetic algorithm based mostly reduced feature sets for 2 varieties of wine in addition as simulated annealing based mostly reduced feature sets for 2 varieties of wine to measures the performance parameter with regard to every classifier. we tend to separated every performance measures with respect to genetic algorithm and simulated annealing sets and plot the column plot for higher visualization. The results of each performance measure with respect to two features sets are shown in figure 2, 3 respectively for red wine and figure 4,5 for white wine.

1. Comparison of accuracy for Naïve Bayes red wine

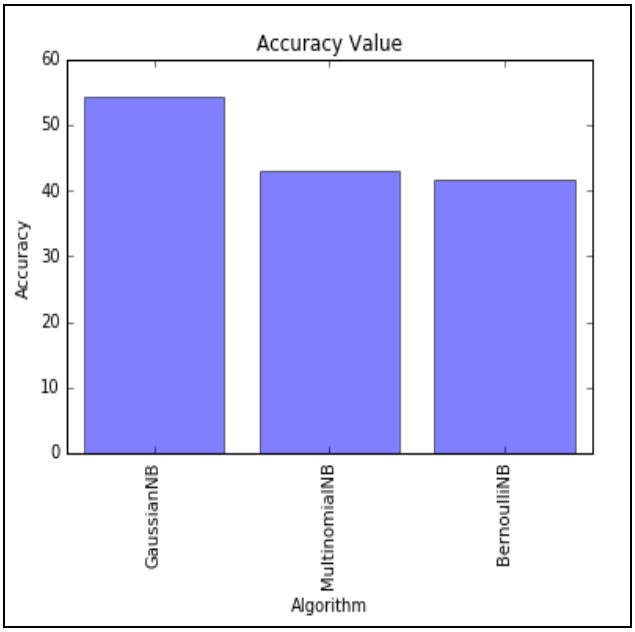

Fig.2 comparison of accuracy-Naive Bayes red wine.

Fig.2 shows that Gaussian Naive Bayes classifier shows maximum accuracy among all the classifiers. It is performed better with confirmation technique based feature sets.

International Journal of Science, Engineering and Technology

An Open Access Journal

2. Comparison of accuracy for SVM red wine

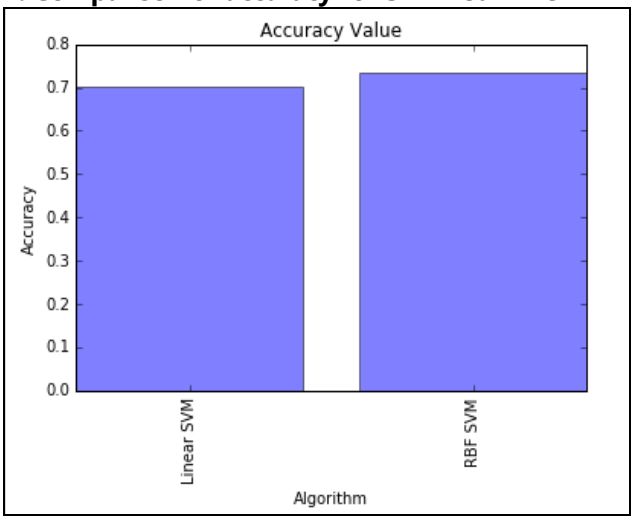

Fig.3 Comparison of accuracy-SVM red wine.

Fig.3 shows that RBF (Radial Basis Function) SVM classifier shows maximum accuracy among all the classifiers. It is performed better with confirmation technique- based feature sets.

3. Comparison of accuracy for Naïve Bayes white wine

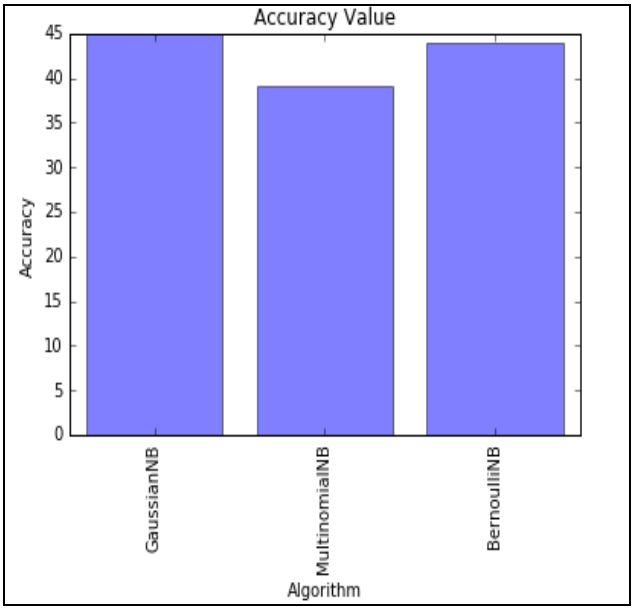

Fig.4 comparison of accuracy-Naive Bayes white wine.

Fig.4 shows that Gaussian Naive Bayes classifier shows maximum accuracy among all the classifiers. It is performed better with confirmation technique based feature sets.

4. Comparison of accuracy for SVM white wine

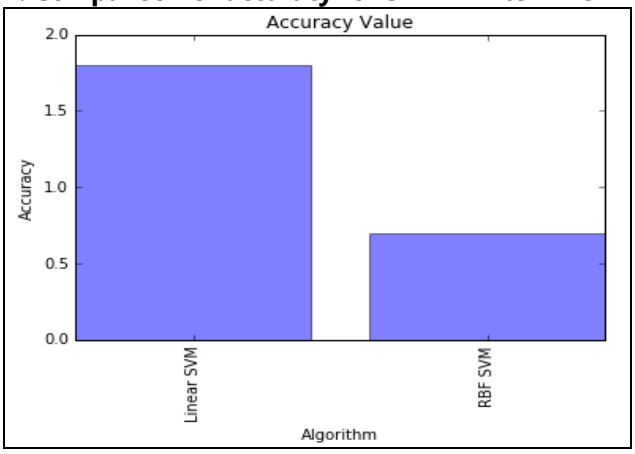

Fig.5 comparison of accuracy-SVM white wine.

Fig.5 shows that Linear SVM classifier shows maximum accuracy among all the classifiers. It is performed better with confirmation technique-based feature sets.

V. CONCLUSION

For each classification model, we analyzed how the results vary whenever test mode is changed. The study includes the analysis of classifiers on both red and white wine data set. Different classifiers like k-nearest-neighbourhood, random forests, and support vector machines are evaluated on datasets. Results from the experiments lead us to conclude that Random Forests Algorithm performs better in classification task as compared against the support vector machine, and k-nearest neighbourhood. The feature selection algorithm provided a clear idea about the importance of the attributes for prediction of quality, which was time-consuming and expensive when done in the traditional way.

We 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. We have found that the SA based feature sets performed better than the GA based feature sets. We have also found that the SVM classifier performed better compared to all other classifiers for red wine and white wine data sets.

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 recommendataions" , Inrtenational 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] Alejandra Urtubia, J Ricardo Perez-Correa, Alvaro Soto, and Philippe Pszcz´olkowski. "Using data mining techniques to predict industrial wine problem fermentations". Food Control, 18(12):1512–1517, 2007.

[10] Appalasamy P, 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):598–601, 2012.

[11] Bernard Chen, Christopher Rhodes and Aaron Crawford, "Wineinformatics: Applying Data Mining on Wine Sensory Reviews Processed by the Computational Wine Wheel", 2014 IEEE International Conference on Data Mining Workshop.

[12] Charif Haydar, Anne Boyer "A New Statistical Density Clustering Algorithm based on Mutual Vote and Subjective
Logic Applied to Recommender Systems" UMAP'17, July 9-12, 2017, Bratislava, Slovakia.

[13] Gongzhu Hu, Tan Xi, Faraz Mohammed, "Classification of Wine Quality with Imbalanced Data", 2016 IEEE.

[14] Hamid Parvin, Hoseinali Alizadeh, Behrouz Minati," A Modification on K-Nearest Neighbor Classifier" GJCST, Vol.10 Issue 14 (Ver.1.0) November 2010.

[15] Kavuri N. C. and Madhusree Kundu. "ART1 Network: Application in Wine Classification", International Journal of Chemical Engineering and Applications, Vol. 2, No. 3, June 2011.

[16] Kunal Thakkar et al, (IJCSIT),"AHP and Machine Learning Techniques for Wine Recommendation" International Journal of Computer Science and Information Technologies, Vol. 7 (5), 2016, 2349-2352.

[17] Linqi Song, Cem Tekin, Mihaela van der Schaar "Clustering Based Online Learning in Recommender Systems: A Bandit Approach", SongTekin ICASSP2014.

[18] Maria J Latorre, Carmen Garcia-Jares, Bernard Medina, and Carlos Herrero, "Pattern recognition analysis applied to classification of wines from Galicia (northwestern Spain) with the certified brand of origin", Journal of Agricultural and Food Chemistry, 42(7):1451–1455, 1994.

[19] Paritosh Nagarnaik, Prof. A. Thomas, "Survey on Recommendation System Methods", IEEE SPONSORED 2ND INTERNATIONAL CONFERENCE ON ELECTRONICS AND COMMUNICATION SYSTEM (ICECS 2015).

[20] Paulo Cortez, Antonio Cerdeira, Fernando Almeida, Telmo Matos, and Jos´e Reis. "Modeling wine preferences by data mining from physicochemical properties". Decision Support Systems, 47(4):547–553, 2009.

[21] S Kallithraka, IS Arvanitoyannis, P Kefalas, An 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.

International Journal of Science, Engineering and Technology

An Open Access Journal

[22] Soude Fazeli, Hendrik Drachsler, Marlies Bitter-Rijpkema, Francis Brouns, Wim van der Vegt, and Peter B. Sloep, "User-centric Evaluation of Recommender Systems in Social Learning Platforms: Accuracy is Just the Tip of the Iceberg", IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES.

[23] Umutoni Nadine, Huiying Cao, Jiangzhou Deng, "Competitive Recommendation Algorithm for E-commerce" 2016, 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD).

[24] Zhibo Wang, Jilong Liao, Qing Cao, Hairong Qi, Zhi Wang, "Friendbook: A Semantic-based Friend Recommendation System for Social Networks", IEEE TRANSACTIONS ON MOBILE COMPUTING 2015.