



ISSN: 2454-9940



**INTERNATIONAL JOURNAL OF APPLIED
SCIENCE ENGINEERING AND MANAGEMENT**

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

Supervised Machine Learning for Wine Quality Prediction: A Comparative Study of Feature Sets and Wine Types

Mr. Potharaju Srinivas

Assistant Professor, Department of CSE,
Malla Reddy College of Engineering for Women.,
Maisammaguda., Medchal., TS, India

Abstract—

in current years, maximum of the industries selling their merchandise primarily based totally at the first-class certification they acquired at the merchandise. The conventional manner of assessing the product first-class is time consuming, however with the invent of device studying strategies the techniques has come to be greater green and ate up less time than before. In this paper we've got explored, a few of the device studying strategies to evaluate the first-class of wine primarily based totally at the attributes of wine that relies upon on first-class. We have used white wine and purple wine first-class dataset for this studies work. We have used distinct function choice approach which include genetic algorithm (GA) primarily based totally function choice and simulated annealing (SA) primarily based totally function choice to test the prediction overall performance. We have used distinct overall performance degree which includes accuracy, sensitivity, specificity, positive predictive cost, bad predictive cost for comparison the usage of distinct function units and distinct supervised device studying strategies. We have used nonlinear, linear and probabilistic classifiers. We have determined that function choice-primarily based totally function units capable of offer better prediction than thinking about all of the functions for overall performance prediction.

I. INTRODUCTION

In current years there may be a modest growth within side the wine intake because it has been determined that wine intake has a high quality correlation to the coronary heart price variability [1]. With the growth within side the intake wine industries are searching for options to provide suitable first-class wine at much less cost. Different wines have one of kind purposes. Although maximum of the chemical substances are equal for one of a kind form of wine primarily based totally at the chemical tests, the amount of every chemical substances have one of a kind degree of awareness for one of a kind form of wine. These days it's miles really vital to categories one of a kind wine for first-class assurance [2]. In the beyond because of loss of technological assets it become hard for maximum of the industries to categories the wines primarily based totally at the chemical evaluation because it takes lot of time and additionally need extra money. These days with the appearance of the machine gaining knowledge of strategies it's miles viable to categories the wines as well as it's miles viable to determine out the significance of every chemical evaluation parameters

within side the wine and which one to disregard for discount of cost. The overall performance assessment with different function units may even assist to categories it in a greater distinctive way. In this paper A clever technique is proposed by thinking about genetic algorithm (GA) primarily based totally function choice as nicely as simulated annealing-primarily based totally function choice thinking about the nonlinear classifiers, linear classifiers and probabilistic classifiers to expect the first-rate in purple wine in addition to the white wine. The shape of the paper is prepared as follows: Section II affords the beyond paintings associated with this field. Section 3 describes approximately the methodologies used for these studies paintings. Section four describes approximately the end result of function choice in addition to the end result of classification. Section five describes approximately the end and destiny paintings.

II. RELATED WORKS

In the beyond few tries were made to apply exceptional system mastering strategies and characteristic choice techniques to the wine dataset. Err and at soy proposed a way to classify the excellent of wines the usage of 3 exceptional classifier such as guide vector machines, Random woodland and k-nearest neighborhood. They have used main aspect analysis for characteristic choice and that they discovered exact end result the usage of Random woodland set of rules [3]. Chen et al proposed techniques that will be expecting the grade of wine the usage of the human savory critiques. They have used hierarchical clustering technique and affiliation rule set of rules to method the critiques and are expecting the wine grade and that they discovered an accuracy of 85.25% while predicting the grade [4]. Appalasamy et al proposed a way to are expecting wine excellent primarily based totally on physiochemical take a look at data. They have mentioned that category technique enables to improve the excellent of wine all through the

production [5]. Beltrán et al proposed a method to categorize the wine primarily based totally on aroma chromatograms and that they have used PCA for dimensionality discount and wavelet rework for characteristic extraction and classifiers including neural network, linear discriminant analysis and aid vector system and discovered that aid vector system with wavelet transforms carry out higher than other classifiers [6]. Thacker et al. used analytical hierarchy process (ahp) to rank the attributes after which used exceptional system studying classifiers including aid vector system and random wooded area and that they discovered accuracy of 70.33% the usage of random wooded area and 66.54% the usage of SVM [7]. Reddy and Govindarajulu used a person centric clustering method to propose the product. They have used crimson wine information set for the survey purpose. They have allotted relative vote casting to the attributes primarily based totally at the literature review. Then they assigned weight to the attributes the usage of Gaussian Distribution Process. They judged the quality primarily based totally at the person desire group [8]. The above beyond work influenced us to strive exceptional characteristic choice set of rules as well as exceptional classifiers to examine the overall performance metrics.

III. METHODOLOGIES

The waft chart of the proposed method is proven in the Fig. 1. A. Data Collection the wine records set is publicly to be had within side the database of UCI. The datasets are associated with purple and white variations of the Portuguese "Vinho Verde" wine. This records set consists of the physiochemical variables in addition to sensory variables; altogether there are 12 attributes [9]. We have used genetic set of rules (GA)-primarily based totally characteristic units for characteristic selection. Pledsoe first supplied an adaptive optimization seek method is known as genetic set of rules and Holland mathematically supplied the genetic set of rules-primarily based totally method via way of means of getting inspiration from Darwin's idea of evolution. A variable is referred to as a gene. A chromosome is not anything however a series of gene. An initialization is accomplished randomly via way of means of the usage of populace of chromosome. The fine of the chromosomes is evaluated consistent with a predefined health function. High performance chromosomes are used to provide the offspring. The genetic operators which include mutation and crossover are used to shape the offspring. In this system the chromosomes are competing with every different and the fittest one survives at the end. The

premiere answer comes after a sequence of iterative computations. [10, 11].

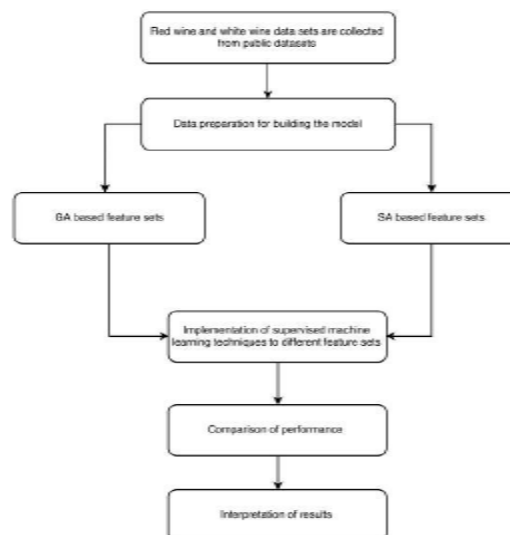


Fig. 1. Flow chart of proposed approach we have extensively utilized simulated annealing-primarily based totally function set for function choice. The broadly used combinatorial optimization approach is known as simulated annealing. It is one of the maximum famous seek algorithms. This approach used probabilistic method to locate the nearby optima that ultimately discover a higher solution [12]. This approach is broadly used for function choice approach. The simulated set of rules technique is cited below. It runs primarily based totally at the range of training. If the range of training is n then it runs nth times. In every run j, the subset of the function for the jet elegance is found. All the jet elegance styles are taken into one elegance and different sample belong to the different elegance even as comparing the modern-day string. This process allows providing the capabilities which classify styles as belonging to elegance j or now no longer elegance j [13]. After deciding on the capabilities by the use of simulated annealing (SA) and genetic set of rules (GA), we've got carried out the facts units into numerous classifiers and evaluate the overall performance parameters.

B. Performance Measure Metrics The parameters used to examine the overall performance and validations of classifier are as follows: accuracy, sensitivity, specificity, high-quality predictive cost (ppv), poor predictive cost (npv). The sensitivity is described because the ratio of authentic positives to the sum of authentic positives and fake negatives. The specificity is described because the ratio of authentic negatives to the sum of fake positives and

authentic negatives. In our studies we have used the Positive predictive cost and poor predictive cost to test the existing and absent of 1 kind of wine. So, the ppv is the chance that the only kind of wine is present given a high-quality take a look at end result and npv is the chance that the one kind of wine is absent given a poor take a look at end result [14]. Accuracy is described because the ratio of variety of correct predictions made to the overall prediction made and the ratio is improved with the aid of using a hundred to make it in phrases of percentage.

IV. RESULTS AND DISCUSSION

We have divided the records into corporations inclusive of teach records and check records. We skilled every classifier primarily based totally at the skilled records and expects the strength of classifier at the check records. So, every classifier capin a position to reveal all of the overall performance metrics inclusive of accuracy, sensitivity, specificity, PPV, and NPV primarily based totally at the check records. We have implemented all of the type strategies to the GA primarily based totally decreased function units for 2 styles of wine as well as SA primarily based totally decreased function units for 2 styles of wine to measures the overall performance parameter with recognize to every classifier. We separated every overall performance measures with recognize to GA and SA units and plot the column plot for better visualization. The effects of every overall performance degree with recognize to 2 function units are proven within side the Fig. 2, 3, 4, 5, and 6 respectively for purple wine and 7, 8, 9, 10, eleven for white wine.

A. Comparison of Accuracy for red wine

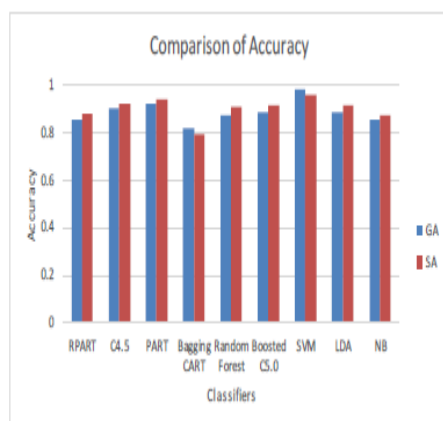


Fig. 2. Comparison of Accuracy on PCA and RFE sets

Fig. 2 display that SVM classifier indicates most accuracy amongst all of the classifiers. It is done

higher with the SA primarily based totally function sets. The accuracy of SVM classifier with SA function set determined to be 95. 23%. B. Comparison of Sensitivity for purple wine Fig. three indicates the sensitivity plot of all of the classifiers with two unique function sets. The plot indicates SVM has the highest sensitivity in comparison to others and it turned into determined to be 0.9717 with the SA primarily based totally function sets

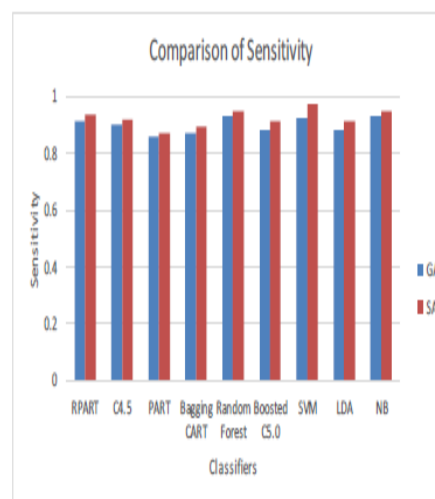


Fig. 3. Comparison of Sensitivity on PCA and RFE sets

B. Comparison of Specificity for red wine

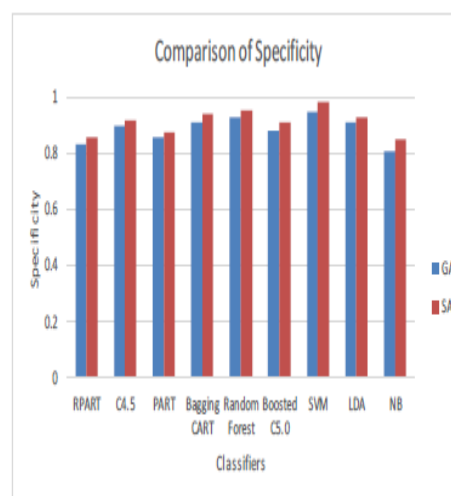


Fig. 4. Comparison of Specificity on PCA and RFE sets

Fig. four suggests that SVM classifier suggests most specificity amongst all of the classifiers. It is completed higher with the SA primarily based totally

characteristic sets. The specificity of SVM classifier with SA characteristic set observed to be 0.9835

Comparison of PPV for red wine

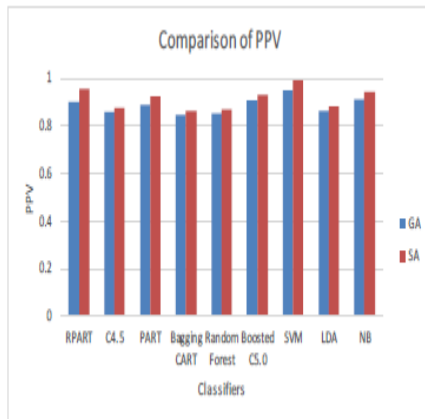


Fig. 5. Comparison of PPV on PCA and RFE sets

Fig. five indicates the PPV plot of all of the classifiers with two one of a kind characteristic sets. The plot indicates SVM has the highest PPV in comparison to others and it became observed to be 0.9912 with the SA primarily based totally characteristic sets

C. Comparison of NPV for red wine

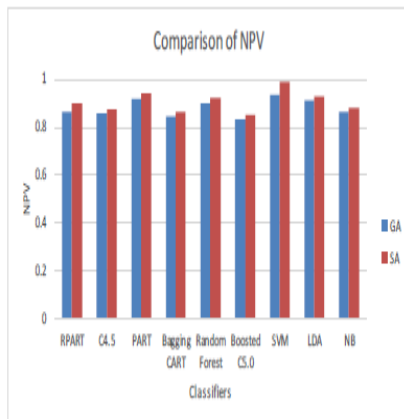


Fig. 6. Comparison of NPV on PCA and RFE sets

Fig. 6 suggests that SVM classifier suggests most NPV amongst all of the classifiers. It is executed higher with the SA primarily based totally characteristic sets. The NPV of SVM classifier with SA characteristic set discovered to be 0.9907

D. Comparison of Accuracy for white wine

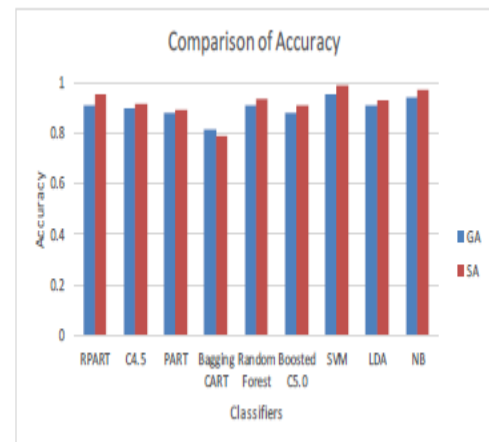


Fig. 7. Comparison of Accuracy on PCA and RFE sets

Fig. 7 indicates that SVM classifier indicates most accuracy amongst all of the classifiers. It is completed higher with the SA primarily based totally characteristic sets. The accuracy of SVM classifier with SA characteristic set discovered to be 98. 81% for white wine information set.

E. Comparison of Sensitivity for white wine

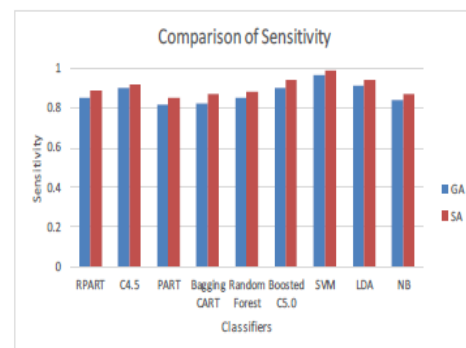


Fig. 8. Comparison of Sensitivity on PCA and RFE sets

Fig. eight indicates the sensitivity plot of all of the classifiers with two one-of-a-kind function units. The plot indicates SVM has the highest sensitivity as compared to others and it became determined to be 0.9934 with the SA primarily based totally function units for white wine information set.

F. Comparison of Specificity for white wine

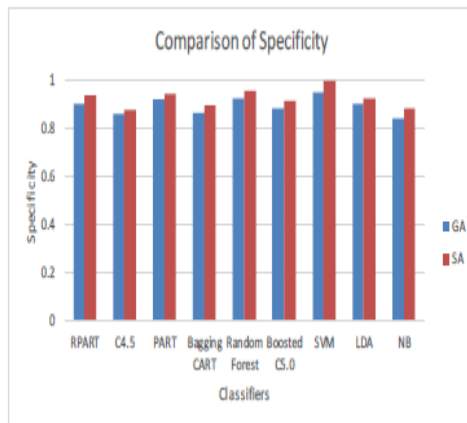


Fig. 9. Comparison of specificity on PCA and RFE sets

Fig. nine indicates that SVM classifier indicates most specificity amongst all of the classifiers. It is completed higher with the SA primarily based totally function sets. The specificity of SVM classifier with SA function set discovered to be 0.9956 for white wine statistics set.

G. Comparison of PPV for white wine

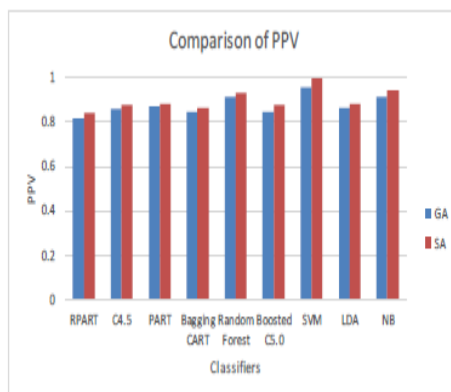


Fig. 10. Comparison of PPV on PCA and RFE sets

Fig. 10 suggests the PPV plot of all of the classifiers with two unique function units. The plot suggests SVM has the highest PPV as compared to others and it turned into discovered to be 0.9987 with the SA primarily based totally function units for white wine records set.

H. Comparison of NPV for white wine

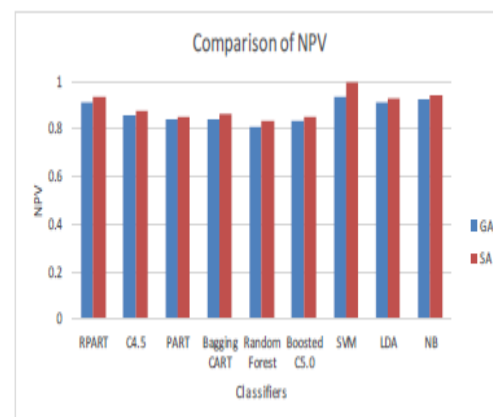
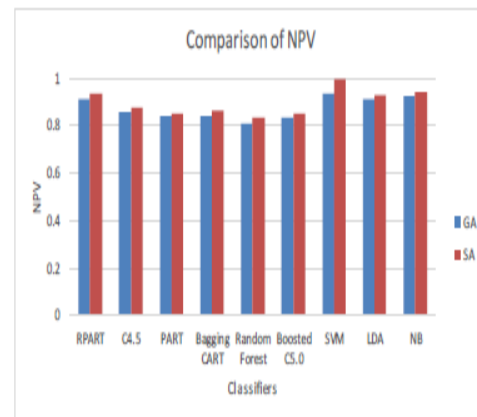


Fig. 11. Comparison of NPV once and RFE sets

Fig. eleven indicates that SVM classifier indicates most NPV amongst all of the classifiers. It is finished higher with the SA primarily based totally function sets. The NPV of SVM classifier with SA function set determined to be 0.9992 for white wine records set. The above plots display the overall performance metrics evaluation of extraordinary sort of wines primarily based totally at the metrics parameters such as accuracy, sensitivity, specificity, ppv and npv on two extraordinary function sets. The end result indicates that the SVM classifier plays higher for each sort of records sets. Especially it's far appearing higher in SA primarily based totally function sets. Although it's far easy to mention primarily based totally on our end result that Simulated Annealing is higher set of rules for function choice in comparison to genetic algorithm based function choice method, but the end result ought to be extraordinary for different datasets in addition to it can be extraordinary for larger datasets. Similarly primarily based totally on our end result we are able to say that SVM classifier is best, however in realistic lot of different parameters additionally come into

photograph that would alternate the scenario completely. This evaluation will provide a clean concept approximately the vital attributes for the prediction of excellent in addition to it saves lot of money and time for the industries.

V. CONCLUSION AND FUTURE WORK

This paper cited approximately capacity of genetic set of rules in addition to simulated annealing set of rules for characteristic choice in addition to the potentials of the classifiers to are expecting accurately primarily based totally on the brand new characteristic units. The characteristic choice set of rules furnished a clean concept approximately the significance of the attributes for prediction of excellent, which become time eating and steeply-priced while performed within side the conventional way. We have additionally as compared the overall performance metrics of linear, nonlinear, and probabilistic primarily based totally classifiers and it become determined that these classifiers executed nicely with the brand new characteristic units. We have determined that the SA primarily based totally characteristic units executed higher than the GA primarily based totally characteristic units. We have additionally determined that the SVM classifier executed higher as compared to all different classifiers for crimson wine and white wine statistics units. We have determined accuracy starting from 95.23% to 98.81% with extraordinary characteristic units. In destiny we are able to strive different overall performance measures and different system gaining knowledge of strategies for higher assessment on results. This evaluation will assist the industries to are expecting the excellent of the extraordinary kind of wines primarily based totally on positive attributes and additionally it'll beneficial for them to make true product within side 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*