A prediction model to identify the wine quality using Linear regression model of Machine Learning

Prince Anuragi, Kumari Shivani, Yash Pramod Bhoyar, Vidyanand Sahu

Nidhi Lal

Department of Computer Science Engineering

IIIT Nagpur

[prince.anuragi@cse.iiitn.ac.in](mailto:prince.anuragi@cse.iiitn.ac.in), [shivani.jaiswal@cse.iiitn.ac.in](mailto:shivani.jaiswal@cse.iiitn.ac.incom) [vidyanand1999@gmail.com](mailto:vidyanand1999@gmail.com), [nidhi.lal@cse.iiitn.ac.in](mailto:nidhi.lal@cse.iiitn.ac.in)

***Abstract*-*Machine Learning is at the heart of modern computational statistics. It allows us to predict results on the basis of a large dataset by prediction a sufficing algorithm and it has done quite well for prediction and statistical analysis. In this paper we will propose better quality results for wine quality. This paper proposes the potential better results on the basis of features that are relevant to the study making results more accurate and reasonable. Further, we will do comparative study between the previous work on the topic and our proposed work.***

**Keywords-Linear Regression, Algorithm, Machine Learning, RMSE**

1. INTRODUCTION

Wine is an alcoholic beverage made from the juice of grapes and its popularity has increased significantly in last decade. In its basic form, wine making is a natural process that requires very little human intervention. Nature provides everything that is needed to make wine, “it is up to humans to embellish, improve, or totally eradicate what nature has provided, to which anyone with extensive wine tasting experience can attest”. In this paper we will use Machine Learning to predict better quality of wine based on its physicochemical features and for this we will be using Linear Regression Model for Model Fitting [2]. **Linear Regression can be considered a Machine Learning algorithm that allows us to map numeric inputs to numeric outputs, fitting a line into the data points. Therefore, to predict better quality of wine we will be considering Linear Regression Algorithm for the Study of wine quality [1].** Wines are categorized using a number of different methods based on extrinsic and intrinsic quality dimensions [4]. For this project, we will be using the**Wine Dataset** from **UC Irvine. All the features of dataset are based on** physicochemical tests and have real values [6]. The Quality of wine differs for different age groups so, for proper measure of wine quality we will be considering people with high involvement and for sake of brevity we will consider physicochemical properties and chemical composition of the wine. Wine quality is directly linked to the quality of raw materials and methods which are used to grow those grapes [6].

2.PREVIOUS WORK

In their previous work [5], they have stated the problem of quality assessment as a regression model. For this classification they have considered the UC Irvine Dataset of White Wine from the Region in ­­­southern European country. Wine certification is usually considered to be assessed by physiochemical properties of the wine but on contrary the taste of wine and it’s quality is understood by human senses. Thus, it makes hard for us to classify the wines as there are really complex factors that differ taste of wine and relation among the physicochemical properties and sensory analysis still not fully explained. Progress in Information Technologies have made it easy to store complex and big data.

All stored data contains valuable patterns and trends which can be used for further prediction and decision making after optimizing it further.

A regression approach should be modeled to conserve order of grades. Performance of regression is usually measured as mean absolute deviation (MAD) and regression models can be used to differentiate different regression models, with ideal to have 1.0 score. In previous works that they’ve used neural networks (NNs) and more recently support vector machines (SVMs) for sophisticated analysis to extract knowledge. Better results were achieved with SVM with increasing performances, over NN and MR techniques. The overall accuracy they’ve obtained are 64.3% (Error Tolerance=0.5) and 86.8% (Error Tolerance=1.0). As dataset have seven classes the results obtained are better than to be expected from random classifiers. As SVM fitting is resources hungry on compute power, but desired results can be achieved in conceivable time. The results that they produced were relevant to wine science domain, helping to understand on how physicochemical characteristics affects the final quality of the wine. The data-driven approach is done by objective test and can be further integrated into a decision driven support system, improving accuracy and quality of oenologist performance.

3.PROPOSED WORK

In this work we used the Linear Regression Model of Machine Learning to improve the RMSE and Root Mean Square Errors for better quality prediction of Red Wine. The Wine Dataset that we have used is from UC Irvine . The classes are ordered and not balanced (e.g. there are many more normal wines than excellent or poor ones). Dataset consists of physiochemical (i.e. sulphates, citric acid, etc.) and sensory (quality) variable and these data aren’t correlated and hence we will be taking only those features which will be considered as more related to wines taste, aroma and color as human sensory might depending upon quality. For formatting proper data we will be taking features that have correlation

Outlier detection algorithms could be used to detect the few excellent or poor wines. Also, we are not sure if all input variables are relevant.

4.RESULTS AND DISCUSSION

These results go beyond previous reports, showing that \_\_

Contrary to the findings of \_\_ we did not find \_\_

Because of this potential limitation, we treat \_\_

Regarding the limitations of \_\_, it could be argued that \_\_

5.CONCLUSIONS

REFERENCES

1. Younger, Mary Sue. *Handbook for linear regression*. Vol. 1. North Scituate, MA: Duxbury Press, 1979.
2. Kutner, Michael H., Christopher J. Nachtsheim, John Neter, and William Li. *Applied linear statistical models*. Vol. 5. Boston: McGraw-Hill Irwin, 2005.
3. Myers, Raymond H., and Raymond H. Myers. *Classical and modern regression with applications*. Vol. 2. Belmont, CA: Duxbury press, 1990.
4. Charters, Steve, and Simone Pettigrew. "The dimensions of wine quality." *Food Quality and Preference* 18, no. 7 (2007): 997-1007.
5. Cortez, Paulo, Juliana Teixeira, António Cerdeira, Fernando Almeida, Telmo Matos, and José Reis. "Using data mining for wine quality assessment." In *International Conference on Discovery Science*, pp. 66-79. Springer, Berlin, Heidelberg, 2009.
6. Cortez, Paulo, António Cerdeira, Fernando Almeida, Telmo Matos, and José Reis. "Modeling wine preferences by data mining from physicochemical properties." *Decision Support Systems* 47, no. 4 (2009): 547-553.
7. Lund, Steven T., and Joerg Bohlmann. "The molecular basis for wine grape quality-a volatile subject." *Science* 311, no. 5762 (2006): 804-805.
8. De Orduna, Ramon Mira. "Climate change associated effects on grape and wine quality and production." *Food Research International* 43, no. 7 (2010): 1844-1855.