International Journal of Internet, Broadcasting and Communication Vol.10 No.2 25-30 (2018)
http://dx.doi.org/10.7236/IJIBC.2018.10.2.5

IJIBC 18-2-5

Wine Quality Classification with Multilayer Perceptron

Garima Agrawal[1], Dae-Ki Kang[2†]

Department of Ubiquitous Computing, Dongseo University
garima.rga@gmail.com, dkkang@dongseo.ac.kr

Abstract

This paper is about wine quality classification with multilayer perceptron using the deep neural network. Wine complexity is an issue when predicting the quality. And the deep neural network is considered when using complex dataset. Wine Producers always aim high to get the highest possible quality. They are working on how to achieve the best results with minimum cost and efforts. Deep learning is the possible solution for them. It can help them to understand the pattern and predictions. Although there have been past researchers, which shows how artificial neural network or data mining can be used with different techniques, in this paper, rather not focusing on various techniques, we evaluate how a deep learning model predicts for the quality using two different activation functions. It will help wine producers to decide, how to lead their business with deep learning. Prediction performance could change tremendously with different models and techniques used. There are many factors, which, impact the quality of the wine. Therefore, it is a good idea to use best features for prediction. However, it could also be a good idea to test this dataset without separating these features. It means we use all features so that the system can consider all the feature. In the experiment, due to the limited data set and limited features provided, it was not possible for a system to choose the effective features.

Keywords: a *Deep neural network, multilayer perceptron; wine classification, tanh activation.*

1. Introduction

Quality of beverages is the very important part that seriously impacts the consumers which indirectly affects the manufactures. As per Troost, G., Wine is a complex beverage, and there are factors for the quality, and that depends on grape variety, edaphoclimatic conditions and enological practices [1]. So, this means that many features are interlinked. Wine quality broadly encompasses both extrinsic and intrinsic factors [2]. It is well known that the wine quality concept involves a great complexity and it has shown to include several sub-components and the difficulty faced by the wine industry in establishing quality produces varying approaches for its evaluation when tasting wine [3]. The wine industry shares marketing academics' uncertainty about the nature of product quality. Some suggest that the absence of faults [4] is a crucial determinant of quality.

In this paper, we have implemented the deep neural network to analyze that how accurately the deep neural network can predict. We are using multi-layer perceptron model for classification of wine

Manuscript Received: Feb. 20, 2018 / Revised: Mar. 3, 2018 / Accepted: Mar. 15, 2018
Corresponding Author: dkkang@dongseo.ac.kr
Tel:+82-51-320-1724, Fax: +82-51-327-8955
Department of Computer Engineering, Dongseo University, Busan, Korea

quality. From the experimental results, we show how different activation functions can produce mixed results.

2. Previous Research

Many previous studies show different approaches for classification of wine grade. Like Gaussian process regression and multi-task learning. Yeo et al [5] found that advanced machine learning technique has the potential for prediction of wine price.

We keep learning that wine quality varies because of many factors. One of the same views was presented by Ashenfelter [6], he mentioned that the quality depends on the type of grapes used. He proved it by a price equation which he derived using several factors. He found climate change and expert opinion are the key options when deciding the wine price.

Ribeiro et al. [7] have predicted wine vinification using data mining tools. They figured out good accuracies in all the techniques.

Lee et al. [8] also proposed a decision tree based method to predict the wine quality. They compared their approach using three machine learning approaches such as SVM, Bayes Net, and Multi-Layer Perceptron and they found their proposed method is better compared to other mentioned approaches.

Many Data mining technologies have been used to classification problems, and it has been applied to wine quality as well. In contrast, the aim of machine learning methods is like other applications are to create models from data to predict wine quality.

Yesim Er*1, Ayten Atasoy1 used SVM, K-NN AND Random Forest with cross-validation and percentage split test method [9]. They found that for each classification model, the results vary whenever test mode is changed.

In 1991, Wine dataset was contributed to UCI repository. There were 178 instances including 13 different attributes like alcohol, magnesium to classify three cultivars from Italy [10]. Principal component analysis (PCA) have been used too [11]. Another experiment had been used in wine aroma chromatograms which were measured with a Fast GC Analyser [12]. In the final study, three more classification methods were compared such as Linear Discriminant Analysis, Radial Basis Function Neural Networks, and Support Vector Machines (SVM) as per their performance in a two-stage architecture.

Cortez et al. [13] presented three types of models; the first one is a neural network. The Second is multiple regression, and last is support vector machine to measure the quality and the taste of wines. Shanmuganathan's [14] stressed on doing the prediction of the various temperatures on wine crops and quality. However, Chen et al. [15] were more focused to improve the taste and quality of wine by gathering the end users feedback. To conclude this problem, they choose to apply hierarchical clustering of wine to have more appropriate results which can provide an improved framework for wine classification.

3. Dataset

In this research paper, we will be using Portugal region Vinho Verde's wine samples dataset to forecast the wine quality. They were in best ten exporting countries in 2005. This country had 3.17% of the market share in 2005 [16]. The government has appointed The Vinho Verde Region's

Viticulture Commission (CVRVV) to certify, promote, defend, control and ensuring the authenticity and quality of all Vinho Verde Denomination of Origin wines. On the other hand, they also have a research and testing center improve, develop and modernize the winemaking abilities with the aim to produce the best quality wines to compete in the global market [17] including crops of wines. This center is known as Amândio Galhano Wine Center (EVAG), they are focused in four areas:

- Grape production
- Certified plants/rootstock
- Research
- Wine tourism

The authorization process for wine is little complicated. It can be evaluated by doing physicochemical test method and by doing sensory tests [18]. On a regular basis physicochemical test used to describe density, alcohol, or pH values check, while sensory tests depend on mainly on human experts. The taste is not fully understood by the human senses [19]. Hence, wine classification is a very hard task. Both tests are different from each other. Additionally, the physicochemical and sensory test results have very complicated relationships and are still not fully understood [20].

We have selected the dataset from https://archive.ics.uci.edu/ml/datasets/Wine+Quality [21]. In the same repository, two datasets were deposited, red, and white wine. In Contrast red wine dataset is smaller and has 1599 instances, and white wine dataset contains 4898 instances. Both datasets have the same number of 11 features, and the model predicts 12^{th} feature. The 11 attributes are inputs for objective tests (e.g., pH values, alcohol, density, etc.), and the 12^{th} attributed is predicted as per the sensory data. Wine quality is ranged between 0 (very bad) and 10 (very excellent). However, on comparing the wine samples, we found that it has normal wines more than the bad or best quality. We found out that dataset has only six classes samples, those are 3,4,5,6,7,8.

Both wine datasets have exactly same features, pH defines the acidic level, or basic wine is on a scale from 0 (very acidic) to 14 (very basic). Most wines are between 3-4 on the pH scale. Chloride is the amount of salt in the wine. Alcohol is the percent alcohol content of the wine.

A prestige-class wine contents an equilibrium constituent to set up the manufacturing process such as sugar fraction in sugarcane, which narrates with the proportion of alcohol. Additional following harmony specimen is diverse formulae of acid, which changes the perception and tartness of wine. Tannins resemble the acidity of the wine [22].

4. Methodology

We want to predict wine quality in very simple steps. Since this dataset is not very large, we cannot implement a very large deep neural network. So, we decided to build very simple 5 step model, which is shown in Figure:1

Software and Library: We experimented R with Keras and Tensorflow.

Model: We have built a deep neural network using multilayer perceptron model with multiple hidden layers.

Compilation: We have compiled the model using categorical cross-entropy and mean squared error (MSE) with Adaptive Moment Estimation (ADAM) optimizer.

Fitting Model: We tried to fit our model with different epochs and batch sizes, and validation split.

Evaluate Model: We tested our model and analyzed prediction based on different epochs and batch sizes using MSE and ADAM optimizer on different epoch and batch sizes with constant validation split

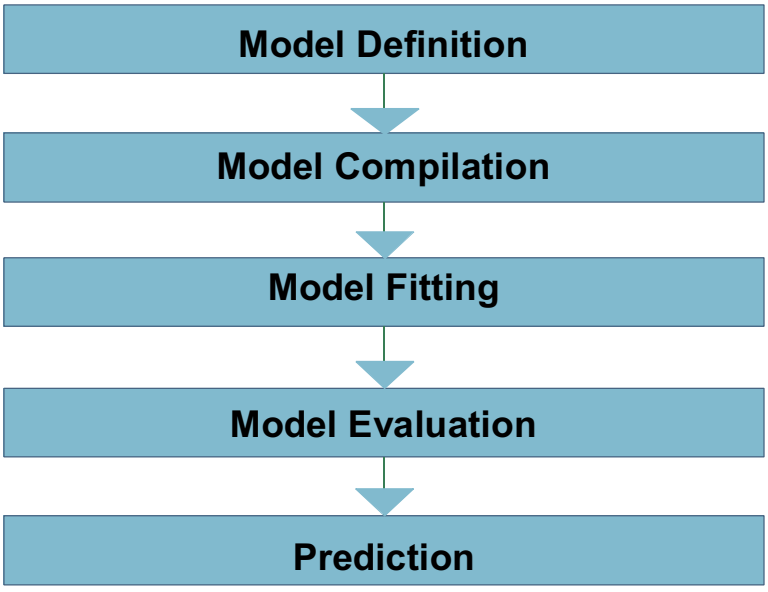

Figure 1: Five step life cycle of neural network

5. Conclusion

For this five-step model, we analyzed how the results vary whenever we use different activation function.

- When using with Rectified Linear Unit (RELU) activation on red wine dataset, we get the 53% validation accuracy while using tanh we have seen 52% validation accuracy with 200 epochs and 66 batch size with validation split at 0.15.
- When using with RELU activation on white wine dataset, we get the 48% validation accuracy while using tanh we have seen the 53% validation accuracy with 200 epochs and 66 batch size with validation split at 0.15. We evaluated the same model with different epochs and batch sizes.

Results from the experiments suggest us to conclude that data gets overfitted or under fitted when increasing the epochs and batch sizes. While using different activation function does not impact much.

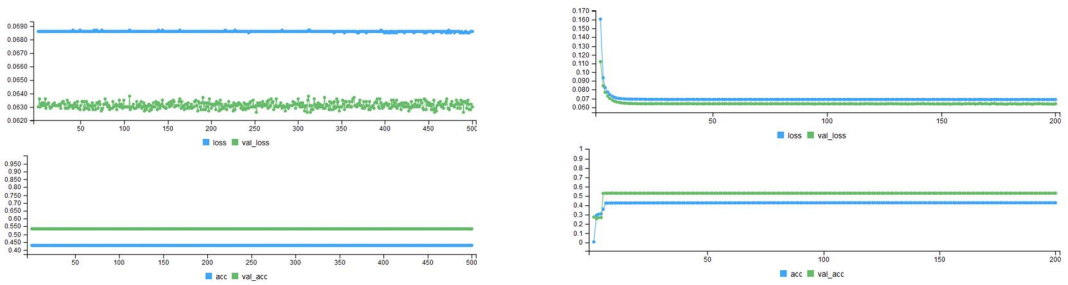

Figure 2. White wine dataset with tanh (left) and white wine dataset with RELU activation function (right)

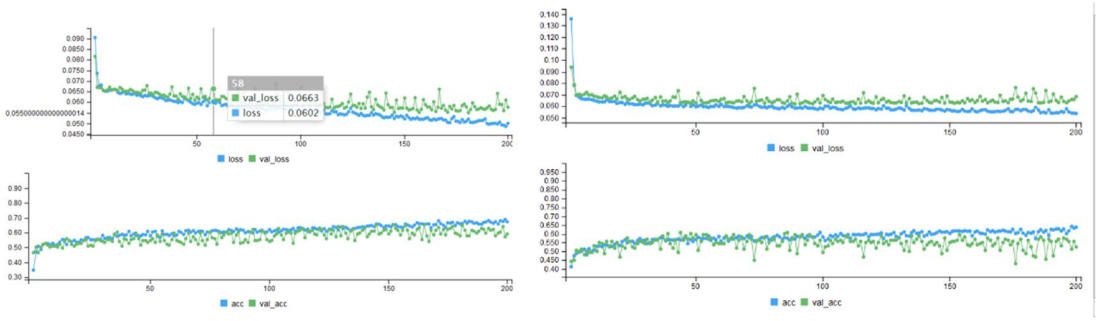

Figure 3. Red wine dataset with tanh (left) and red wine dataset with RELU activation function (right)

References

[1] G. Troost, *Tecnologı´a del vino*. Barcelona, Spain: Omega S.A, 1985.

[2] V. Jover, L. Montes, and F. Fuentes, *Measuring perceptions of quality in food products: the case of red wine*, 2004.

[3] S. Charters, and S. Pettigrew, "The Dimensions of Wine Quality," *Food quality and preference*, Vol. 18, No. 7, pp. 997-1007, 2007.
DOI: https://doi.org/10.1016/j.foodqual.2007.04.003

[4] C. Wright, Wine wizards. *The Qantas Club Magazine,* Spring 2001.

[5] M. Yeo, T. Fletcher, and J. Shawe-Taylor, "Machine Learning in Fine Wine Price Prediction," *Journal of Wine Economics*, Vol. 10, No. 2, pp.151-172, 2015.
DOI: https://doi.org/10.1017/jwe.2015.17

[6] O. Ashenfelter, "Predicting the Quality and Prices of Bordeaux Wine," *Journal of Wine Economics*, Vol. 5, No. 1, pp.40-52, 2010.
DOI: https://doi.org/10.1111/j.1468-0297.2008.02148.x

[7] J. Ribeiro, J. Neves, J. Sanchez, M. Delgado, J. Machado, and P. Novais, "Wine Vinification Prediction using Data Mining Tools.," in *Proc. Computing and Computational Intelligence,* Tbilisi, the Republic of Georgia, pp. 78-85. June 2009.

[8] S. Lee, J. Park, and K. Kang, "Assessing Wine Quality using a Decision Tree.", in *Proc. IEEE International Symposium on Systems Engineering (ISSE),* pp. 176-178. IEEE, September 2015.

[9] Y. Er, A. Atasoy1, "The Classification of White Wine and Red Wine According to Their Physicochemical Qualities," *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 4, Special Issue No. 1, pp. 23-26, 2016.
DOI: https://doi.org/10.18201/ijisae.265954

[10] A. Asuncion, and D. Newman, *UCI Machine Learning Repository,* University of California, Irvine, 2007.

[11] Center for Machine Learning and Intelligent Systems. *http://www.ics.uci.edu/~mlearn/MLRepository.html*

[12] S. Kallithraka, I. S. Arvanitoyannis, P. Kefalas, A. 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*, Vol. 73, No. 4, pp.501-514, 2001.
DOI: https://doi.org/10.1016/s0308-8146(00)00327-7

[13] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, J. Reis, "Modeling Wine Preferences by Data Mining from Physicochemical Properties," Department of Information Systems/R&D Centre Algoritmi, University of Minho, 4800-058 Guimarães, Portugal, *Viticulture Commission of the Vinho Verde Region (CVRVV)*, 4050-501 Porto, Portugal.

[14] S. Shanmuganathan, P. Sallis, and A. Narayanan, "Data Mining Techniques for Modeling Seasonal Climate Effects on Grapevine Yield and Wine Quality," in *Proc. IEEE International Conference on Computational Intelligence Communication Systems and Networks,* pp. 82-89, July 2010.

[15] B. Chen, C. Rhodes, A. Crawford, and L. Hambuchen, "Wineinformatics: Applying Data Mining on Sensory Wine Reviews Processed by the Computational Wine Wheel," in *Proc. IEEE International Conference on Data Mining Workshop*, pp. 142-149, Dec. 2014.

[16] FAO. FAOSTAT - Food and Agriculture Organization agriculture trade domain statistics. *http://faostat.fao.org/site/535/DesktopDefault.aspx?PageID=535*, July 2008.

[17] Introducing a refreshing world Vinho Verde Press Kit 2016. *http://winesofvinhoverde.com/wp-content/uploads/2015/11/VV_PressKit_2016.pdf*.

[18] S. Ebeler. Flavor Chemistry - Thirty Years of Progress, chapter *Linking Flavor Chemistry to the Sensory Analysis of Wine*, pp. 409-422. Kluwer Academic Publishers, 1999.

[19] D. Smith and R. Margolskee, "Making Sense of Taste," *Scientific American*, Special issue, Vol. 16, No. 3, pp.84-92, 2006.
DOI: http://doi.org/10.1038/scientificamerican0301-32

[20] A. Legin, A. Rudnitskaya, L. Luvova, Y. Vlasov, C. Natale, and A. D'Amico, "Evaluation of Italian Wine by the Electronic Tongue: Recognition, Quantitative Analysis, and Correlation with Human Sensory Perception," Analytica Chimica Acta, Vol. 484, No. 1, pp. 33-34, 2003.
DOI: https://doi.org/10.1016/S0003-2670(03)00301-5

[21] UCI Machine Learning Repository, Wine quality data set.
https://archive.ics.uci.edu/ml/datasets/Wine+Quality.

[22] H. K. Kubach, *Wine Grape Suitability, and Quality in a Changing Climate,* An Assessment of Adams County, Pennsylvania, 1950.

[23] S. Lee, H. Kim, H. Seok, and J. Nang, "Comparison of Fine-Tuned Convolutional Neural Networks for Clipart Style Classification," *International Journal of Internet, Broadcasting and Communication*, Vol. 9, No. 1, pp. 9-17, 2017.
DOI: http://doi.org/10.7236/IJIBC.2017.9.4.1