A Deep Neural Network Approach to Predict the Wine Taste Preferences

Sachin Kumar, Yana Kraeva, Radoslava Kraleva and Mikhail Zymbler

Abstract In this study, deep neural networks are developed to evaluate its performance over wine data set from UCI repository. The data set consists of white and red wine samples from Portugal. Previous studies claimed that Support Vector Machine (SVM) outperformed the simple ANN and Multiple Regression (MR) on wine data set. We trained different neural networks models with different hidden layers and activations to understand if it is possible to achieve better accuracy. It is found that deep learning approach is able to provide better prediction accuracy than SVM even on a smaller data set.

Keywords Wine data · Neural networks · Feature selection · SVM

1 Introduction

Wine tasting is one of the popular leisure pursuits and a form of enjoyment among people of all age. Portugal is one of the top ten wine exporters in the northwest regions [1]. Its Vinho Verde wine is a popular brand being exported in northwest regions. On the other hand, it a world known fact that alcohol consumption and consuming bad or low-quality wine could be dangerous for health. Therefore, certification and quality assessment are one of the parameters that wine manufacturers and exports must follow.

In order to perform wine certification, several physicochemical and sensory examinations have to be done [2]. In the prior test, chemical properties of the wine, i.e., density of alcohol and pH values have to be measured in the laboratories whereas the second one relies on human expertise. As taste of anything is one of the least understandable by human senses [3] and wine is one of them. It is not wrong to say

S. Kumar (✉) · Y. Kraeva · M. Zymbler
Department of Computer Science, South Ural State University, Chelyabinsk, Russia
e-mail: sachinagnihotri16@gmail.com

R. Kraleva
Department of Informatics, South West University "Neofit Rilski", Blagoevgrad, Bulgaria

V. K. Solanki et al. (eds.), *Intelligent Computing in Engineering*,
Advances in Intelligent Systems and Computing 1125,
https://doi.org/10.1007/978-981-15-2780-7_120

that wine classification is a complex job [4]. In addition, it is hard to fully understand the relationship between laboratory tests and human sensory analysis [5].

Present world is equipped with high technology and advancement of generating and maintaining huge, massive, and complex databases. This huge database consists of important and crucial information which can be used for improved decision-making after proper analysis [6]. Machine Learning (ML) techniques have a big role to play in this improved decision-making process [7]. ML is efficient to extract the crucial and hidden information from the massive data. ML techniques can be broadly divided into supervised and unsupervised learning. In supervised learning, enough samples with respective class values are available to develop a trained model which can be tested further on new samples. In case, when class values are not available, unsupervised learning comes into play. It groups the data into homogeneous segments and further learning can be done to understand the samples in different groups. The right selection of ML techniques must be done in order to perform fruitful analysis. This selection usually depends on the nature of data under study. For continuous data with numeric values, prediction using regression methods, i.e., linear/multiple regressions are the preferred choice. However, these methods are based on certain assumptions with the choice and relationship between dependent and independent variables. Also, when the dimensions or the number of attributes in the data is large then these methods may not provide efficient results. ML techniques such as SVM and ANN could be better options in such cases [8, 9]. Both the techniques are quite powerful and have a proven record in learning from continuous data (linear and nonlinear features) and predicting good results. However, certain things must be taken into consideration such as feature selection and model selection. All features in the data set might not be useful for analysis; hence it is important to discard those attributes that contribute very less or nothing in the prediction [10]. Further, model selection for better accuracy relies on selection and adjustment of certain parameters such as number of neurons in hidden layer in ANN and kernel parameters for SVM [11]. However, several mechanisms are available to perform such activities but every decision support systems need to be more accurate and precise.

ML techniques are one of the best options to predict the quality of wine based on the data obtained from laboratory tests. There are some studies mentioned in the literature that used ML techniques to predict the wine quality. However, they used small sample obtained from laboratory tests. UCI repository is a well-known online data repository system [12]. In 1991, first time the wine data set was submitted in UCI repository which contains only 178 samples with 13 chemical attributes. This data set was used widely to test with ML techniques and also used as a benchmark data set for the further ML analysis. Sun et al. [13] used feed-forward ANN to predict the geographical origins of wine. The data they used consists of 170 sample and they mentioned to achieve 100% prediction accuracy. Further, ANN was again used [14] to predict sensory features from a wine data set. They only used 36 samples and achieved 6% error rate. Then 56 samples from the laboratory tests were used to characterize Italian wine [15]. These studies attempted well but the data sample used in these studies was quite small. Also, the ML techniques are widely known to

work for relatively large amount of data. Then only, one can rely on the outcomes produced by the ML techniques.

In this study, we are providing a case study on analyzing wine data using deep learning technique. The results are compared with the existing results provided in [4]. The deep learning model simply outperforms the SVM approach used in [4] and provides better accuracy. The model can be utilized to predict the test score for each new sample and would assist in decision-making process for wine producers and consumers as well.

In the next Sect. 2, the materials and methods used are discussed. In Sect. 3, experimental results and discussion have been provided which is followed by conclusion in Sect. 4.

2 Materials and Methods

2.1 Data Set Description

The wine data set available at UCI online repository database consists of 11 attributes, i.e., fixed acidity (tartaric acid g/dm^3), volatile acidity (acetic acid g/dm^3), citric acid (g/dm^3), residual sugar (g/dm^3), Sodium chlorides (g/dm^3), free sulfur dioxide (mg/dm^3), total sulfur dioxide (mg/dm^3), density (g/cm^3), pH, potassium sulfates (g/dm^3), and alcohol (vol. %) amount. The target attribute to be predicted is the quality scores on a range of 1–10 where 0 indicates the worst quality and 10 indicates the best. The total number of instances for red wine is 1599 and for white wine 4898.

Figure 1 shows the distribution of samples among all the wine scores on a scale of 0–10. Table 1 provides a statistical description of the attributes for both red wine and white wine samples in the data set. In Fig. 2, the heat map for both the red and white wine data set is provided. It was found that for red wine, alcohol and sulfates have the most positive correlation with quality and the most negative correlation is

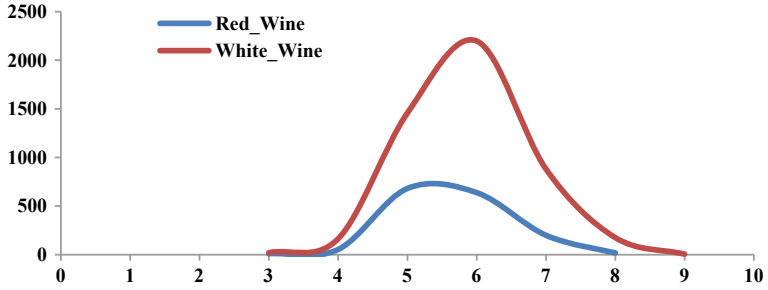

Fig. 1 Distribution of samples among all the wine scores on a scale of 0–10

Table 1 Statistical distribution of red and white wine samples

	Red wine samples				White wine samples			
	Min.	Max.	Mean	StdDev	Min.	Max.	Mean	StdDev
Fix. acid.	4.6	15.9	8.32	1.74	3.8	14.2	6.85	0.84
Vol. acid.	0.12	1.58	0.53	0.18	0.08	1.1	0.28	0.10
Citr. acid	0	1	0.27	0.19	0	1.66	0.33	0.12
Residual sugar	0.9	15.5	2.54	1.41	0.6	65.8	6.39	5.07
Chlorides	0.01	0.61	0.08	0.05	0.009	0.346	0.046	0.022
Free sulf. oxd.	1	72	15.8	10.46	2	289	35.31	17.007
Tot. sulf. oxd.	6	289	46.47	32.9	9	440	138.36	42.49
Density	0.99	1.01	0.99	0.002	0.987	1.039	0.994	0.003
pH	2.74	4.01	3.31	0.154	2.272	3.82	3.188	0.151
Sulfates	0.33	2	0.66	0.17	0.22	1.08	0.49	0.114
Alcohol	8.4	14.9	10.42	1.06	8	14.2	10.514	1.231

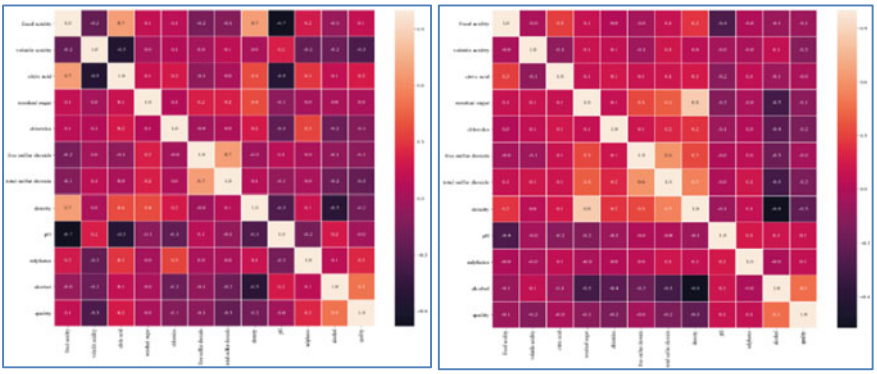

Fig. 2 Heat map showing correlation of attributes for red wine data (left figure) and white wine data (right figure)

with volatile acidity. For wine data set, alcohol has the most positive correlation with quality and density has the most negative correlation.

2.2 Data Preprocessing

Previous studies considered all the attributes available in the data set for analysis. A good data analysis approach recommends a thorough study of data and its attributes.

Procedure1 : Sequential forward floating search (SFFS)
Input : $U = Complete\ set\ of\ attributes$
Output : $X = set\ of\ selected\ most\ significant\ attributes$
Initialization : $j = 0, X_j = [Null],\ Y_j = U$

```
While True {
  y = max_{i∈Y_j} J(X_j ∪ {i})          //selection of best attributes from the subset
  X_{j+1} = X_j ∪ {y}
  j = j + 1
  x = max_{i∈X_j} J(X_{j+1} − {i})      //selecting least important feature
  While J(X_{j+1} − x) > J(X_k) {
    X_j = X_{j+1} − {x}
    j = j − 1
    x = max_{i∈X_j} J(X_j − {i})
    }
  }
```

All the attributes in the data set may or may not be relevant for the analysis. Therefore, a preprocessing must be performed prior to begin analysis. In order to select the most relevant attributes for analysis, the Sequential Forward Floating Search (SFFS) [16] method has been used. The procedure of SFSS is illustrated in procedure 1.

SFFS starts with an empty set of relevant features to be selected from the universal set U. X_j and Y_j represent selected and remaining features, respectively, and lowercase x and y represent single attribute value. $J(X_j)$ is the feature evaluation function. It measures the appropriateness of the attribute based on its characteristics. Further, it keeps adding the relevant features to the X_j, until the stopping criteria are fulfilled.

2.3 Support Vector Machines (SVM)

In this study, we have used Multiclass Support Vector Machine (MSVM) technique to train the model. The description of MSVM is briefly explained as follows: Initially, Support Vector Machine (SVM) has been designed and widely used for binary classification in which only two classes have to be predicted. Further, due to the demand of classification and prediction problem for more than two classes motivated the design of a new SVM which can handle the multiclass scenario. In 2001, some classification models approaching multiclass classification issue were proposed [17, 18]. The problem with these models had large computation complexity. Another study [19] addressed the issue of large computational cost and proposed a simplified version of multiclass SVM. In general, the multiclass SVM can be considered as One Against One (OAO-SVM), One Against All (OAA-SVM), and Direct Acyclic Graph (DAG

SVM) for classification of multiple classes. In this study, one against all SVM is considered for analysis.

2.4 Deep Neural Network

A deep learning model can be defined as a hierarchical structure in which each layer applies a linear transformation to the preceding layer. Assuming, a sample input data $D \in R^{n \times d}$, with n training samples of d dimensions or attribute values. The number of neurons in input layer is based on the number of dimensions in the sample data (d in this case). These input values are multiplied by the corresponding weights and supplied to each of the neurons in hidden layers. Further, the output of these hidden layers will again multiply with another weight matrix for next hidden layer and so on. On each hidden layer, there is an activation function that decides which neurons output should be forwarded to the input of next layer [20]. The popular activation functions in practice are sigmoid, tanh, and ReLU. Sigmoid accepts a real-valued single number and provides the output within the range [0, 1], whereas tanh provides output within the range $[-1, 1]$. ReLU is an abbreviation of Rectified Linear Unit. ReLU is the default activation function introduced in several machine learning libraries recently. ReLU provides an output x for a positive real-valued number otherwise 0.

Furthermore, the concept of backpropagation to reduce the prediction error has been used. The error can be calculated using difference between the actual and predicted output. Further, this error is propagated backward and weights on each hidden layer can be modified accordingly.

3 Results and Discussion

Experimental analysis has been performed on Anaconda Spyder 3.6 version. Scikit learn, pandas, keras with tensor flow as backend, etc. python packages have been used to perform the analysis. First SVM analysis has been performed on both red and white wine samples. Further, deep learning approach was used.

SVM model achieved the MSE 0.38. In order to evaluate the performance of deep neural networks, we created four networks with different hidden layers and trained for different number of iterations. All networks were trained with a batch size of 200 samples and a learning rate $\alpha = 0.001$. It can be seen in Figs. 3 and 4 that MSE (Mean Squared Error) were reduced with more training iterations. We used two popular activations functions sigmoid and ReLU. However, these functions have shown different performance for white and red wine samples. With red wine samples, neural networks with 1 and 2 hidden layers and sigmoid activation achieved lowest MSE (0.29), whereas with ReLU activation, network with 3 hidden layers achieved lowest MSE (0.22). With white wine samples, the network with 4 hidden layers and

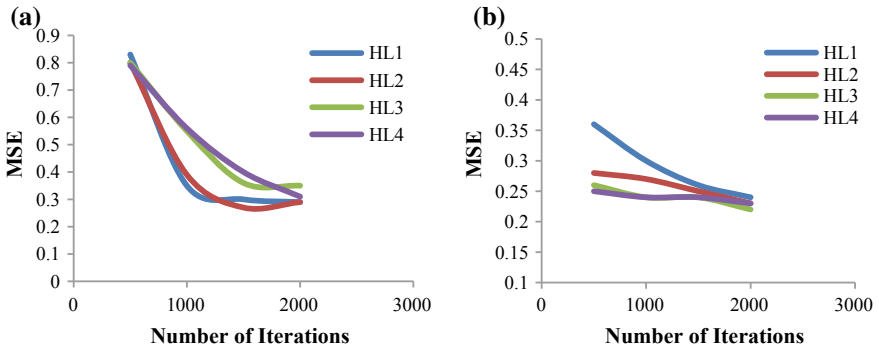

Fig. 3 MSE using deep neural networks for red wine samples with different hidden layers **a** using sigmoid **b** using ReLU

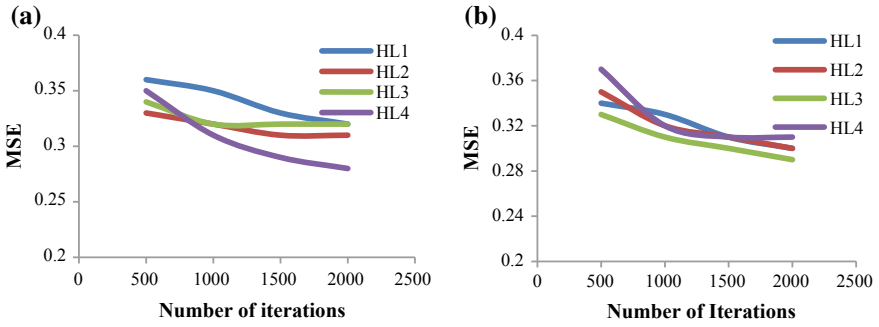

Fig. 4 MSE using deep neural networks for white wine samples with different hidden layers **a** using sigmoid **b** using ReLU

sigmoid activation achieved the lowest MSE (0.28), whereas the network with 3 hidden layers and ReLU activation achieved the lowest MSE (0.29).

Therefore, it has been observed in the analysis that it is possible to achieve better accuracy with deep neural architecture; however, it is difficult to correctly identify the appropriate number of hidden layers in the network. Therefore, it is suggested to try several networks with different hidden layers and activation functions and it must be trained sufficient number of times to reduce the MSE.

4 Conclusion

The study investigated the wine test preference analysis using deep neural network on the available data on UCI repository. Earlier studies mentioned that SVM was the best model to achieve the highest accuracy and minimum error in comparison to

Multiple Regression, and simple ANN model. In this study, we created deep neural network by increasing the number of hidden layers. It was found that MSE was slightly reduced in newly trained networks with multiple hidden layers. Although, for red and white wine samples the networks with different hidden layers performed differently but the results were better than SVM. Therefore, it is recommended to use different hidden layer network with different activation functions. Also, network must be trained sufficient number of times to achieve the lowest MSE.

Acknowledgements This work was financially supported by the Ministry of Education and Science of Russian Federation (government order 2.7905.2017/8.9).

References

1. FAO (2008) FAOSTAT—food and agriculture organization agriculture trade domain statistics, July 2008. http://faostat.fao.org/site/535/DesktopDefault.aspx?PageID=535
2. Ebeler S (1999) Linking flavour chemistry to sensory analysis of wine. Flavor chemistry—thirty years of progress. Kluwer Academic Publishers, New York, pp 409–422
3. Smith D, Margolskee R (2006) Making sense of taste. Sci Am (Special Issue) 16(3):84–92
4. Cortez P, Cerdeira A, Almeida F, Matos M, Reis J (2009) Modeling wine preferences by data mining from physicochemical properties. Decis Support Syst 47(4):547–553. Elsevier
5. Legin A, Rudnitskaya A, Luvova L, Vlasov Y, Natale C, D'Amico A (2003) Evaluation of Italian wine by the electronic tongue: recognition, quantitative analysis and correlation with human sensory perception. Anal Chim Acta 484(1):33–34
6. Turban E, Sharda R, Aronson J, King D (2007) Business intelligence, a managerial approach. Prentice-Hall, Upper Saddle River
7. Witten IH, Frank E (2005) Data mining: practical machine learning tools and techniques with java implementations, 2nd edn. Morgan Kaufmann, San Francisco
8. Rumelhart D, Hinton G, Williams R (1986) Learning internal representations by error propagation. Parallel distributed processing: explorations in the microstructures of cognition, vol 1. MIT Press, Cambridge, pp 318–362
9. Smola A, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199–222
10. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3(7–8):1157–1182
11. Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning: data mining, inference and prediction. Springer, New York
12. UCI repository. https://archive.ics.uci.edu/ml/datasets.html. Accessed 12 May 2018
13. Sun L, Danzer K, Thiel G (1997) Classification of wine samples by means of artificial neural networks and discrimination analytical methods. Fresenius' J Anal Chem 359(2):143–149
14. Vlassides S, Ferrier J, Block D (2001) Using historical data for bioprocess optimization: modeling wine characteristics using artificial neural networks and archived process information. Biotechnol Bioeng 73(1)
15. Wine data set link: http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/. Accessed 12 May 2018
16. Pudil P, Novovicˇová J, Kittler J (1994) Floating search methods in feature selection. Pattern Recognit Lett 15(11):1119–1125
17. Crammer K, Singer Y (2001) On the algorithmic implementation of multiclass kernel-based vector machines. J Mach Learn Res 2:265–292
18. Wang Z, Xue X (2014) Multi-class support vector machine. In: Ma Y, Guo G (eds) Support vector machines applications, pp 23–48

19. Bottou L et al (1994) Comparison of classifier methods: a case study in handwriting digit recognition. In: International conference on pattern recognition, pp 77–87
20. Knerr S, Personnaz L, Dreyfus G (1990) Single-layer learning revisited: a stepwise procedure for building and training a neural network. In: Fogelman J (ed) Neurocomputing: algorithms, architectures and applications. Springer, Berlin