Wine Quality and Type Prediction from Physicochemical Properties Using Neural Networks for Machine Learning: A Free Software for Winemakers and Customers

Nuriel Shalom Mor (✉ nuriel.mor@gmail.com)

Darca and Bnei Akiva, Israel

Research Article

Wine Quality and Type Prediction from Physicochemical Properties Using Neural Networks for Machine Learning: A Free Software for Winemakers and Customers

January 30, 2022

Nuriel S. Mor, Ph.D.
Darca and Bnei Akiva Schools, Israel

Tigabo Asras, Eli Gal, Tesfahon Demasia, Ezra Tarab, Nathaniel Ezekiel, Osher Nikapros, and Oshri Semimufar
Software Engineering Specialization, Harel Bnei Akiva School, Israel

Eva Gladky, Maria Karpenko, Daniel Sason, Danis Maslov, and Omri Mor
Chemistry Specialization, Hadash Darca School, Israel

Abstract. Quality assessment is a crucial issue within the wine industry. The traditional way of assessing by human experts is time consuming and very expensive. Machine learning techniques help in the process of quality assurance in a wide range of industries. The purpose of this study was to develop and offer a free software, for winemakers and customers in which they can easily provide the physicochemical properties of the wine and receive an accurate prediction of the anticipated quality and type of the wine. We used comprehensive datasets of 6497 examples, which contained physicochemical properties and appropriate quality. We combined these datasets, built and trained several neural networks models. We evaluated their performance and selected the best model. Wine quality estimations were modeled as a regression problem and wine type detection as a classification problem. The best model performed well for prediction of wine quality (root means squared error = 0.54) and type (f-score=0.99). With our free software, winemakers and customers can examine how a fine change in each physicochemical property could affect the quality of the wine. They could easily figure out the importance of each physicochemical property, and which one to ignore for reduction of cost. The process is very fast, accurate and does not require taste experts for sensory tests.
Keywords: *neural networks for machine learning, regression, classification, physicochemical properties, wine quality, performance metrics*

1 Introduction

Artificial Intelligence (AI) in the era of deep learning affects a wide range of industries. Even industries that previously did not make use of advanced technologies (Dahal et al., 2021; Mor & Dardeck, 2018; Mor, 2021). AI and more specifically neural networks for machine learning help to enhance the production in many different types of industries. Machine learning can help in the process of quality assurance. AI makes quality assurance more efficient and more accurate and saves expensive manpower use (Aich et al., 2018).

In recent years there is an increase in wine consumption, and wine industries strive to produce good quality wine at less cost (Kumar, 2020). Most of the chemicals are almost the same for different types of wine. However, the exact fine concentration of each chemical is different in different types of wines (Kumar, 2020).

For the purpose of quality assurance, It is very important to predict wine quality from its chemical properties. To maintain the quality with less cost, there is a need to know the relationship between the fine different chemicals' concentration and wine's quality (Gupta et al., 2020). In the traditional approach, taste experts were used. This task is very challenging, takes a lot of time and is expensive (Gupta et al., 2020).

Neural networks excel at predicting outcomes from a set of many features. Neural networks are designed to find very complex relationships between input features and output, and have the ability to model complex relationships between input and output (Mor & Dardeck, 2021). Relevant to our case, neural networks are suitable to learn the complex relationship between the exact fine concentration of each chemical in the wine (input) to its quality (output).

In order to develop an optimal solution for our problem, before constructing and training our model, we need to carefully select the most important wine's chemical features which their concentration might influence quality (Gupta, 2018).

Wine's quality is assessed by physicochemical and sensory tests. Physicochemical tests include features such as density, alcohol, and pH

values. Sensory tests are performed by human taste experts which is a very complex and expensive process. In our literature review, we found eleven consistent physicochemical properties common to almost all data sets designed to evaluate wines' quality (Cortez et al., 2009). The eleven physicochemical properties are: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol (Cortez et al., 2009).

The predominant fixed acids found in wines are tartaric, malic, citric, and succinic. Their respective levels found in wine can vary greatly but in general one would expect to see 1,000 to 4,000 mg/L tartaric acid, 0 to 8,000 mg/L malic acid, 0 to 500 mg/L citric acid, and 500 to 2,000 mg/L succinic acid. All of these acids originate in grapes with the exception of succinic acid, which is produced by yeast during the fermentation process (Sarkar et al., 2018).

Volatile acidity is a measure of the low molecular weight (or steam distillable) fatty acids in wine and is generally perceived as the odour of vinegar. Winemakers are usually most concerned with acetic acid, which accounts for more than 93% of steam distillable acids in wine. The average volatile acidity value for red table wines during this period is about 0.60 g/L. The average volatile acidity value for white table wines is about 0.43 g/L (Vilela, 2018).

Citric acid has many uses in wine production. Citric acid is a weak organic acid, which is often used as a natural preservative or additive to food or drink to add a sour taste to food. Citric acid is often added to wines to increase acidity, complement a specific flavor or prevent ferric hazes. It can be added to finished wines to increase acidity and give a "fresh" flavor. The disadvantage of adding citric acid is its microbial instability. Since bacteria use citric acid in their metabolism, it may increase the growth of unwanted microbes (Zhong et al., 2020).

Residual Sugar is from natural grape sugars leftover in a wine after the alcoholic fermentation finishes. It's measured in grams per liter. Residual sugar levels vary in different types of wine. In fact, many grocery store wines labeled as "dry" contain about 10 g/L of residual sugar. Noticeably sweet wines start at around 35 grams per liter of residual sugar and then go up from there (Wang & Peng, 2017).

Chlorides (sodium chloride) give the wine a salty flavor which may turn away potential consumers. Sodium chloride, commonly known as salt, is an ionic compound with the chemical formula NaCl. The maximum concentration of chlorides in wine is about 0.20 - 0.60 g/L (Vallone et al., 2021).

Sulfur dioxide (SO2) and its salts have been added during winemaking since the 17th century. SO2 and its sulfite salts, remain an essential winemaking additive as there is no one other additive that has the same dual properties of anti-oxidation and preservation. It remains a potentially adverse reaction causing and toxic product for wine consumers and winemakers in amounts greater than 10 mg/L, and that accordingly, should be handled with care. When SO2 is incorporated into a must or a wine, a fraction of it will react with sugars, or aldehydes (ethanal) or ketones. The remaining fraction, called free, is the one with the most important properties. SO2 Total = SO2 free + SO2 reacted. The most active fraction of free SO2 is called active SO2 and is composed of molecular SO2. During maturation and storage, concentrations of free SO2 values of 25 mg/L on red wine and 30 mg /L on white wine are recommended. An active SO2 concentration of 0.35 mg/L ensures a minimum protection, and a value of 0.6 mg/L maximum protection (Capece et al., 2020).

Density is the mass per unit volume of wine or must at 20°C. It is expressed in grams per milliliter, and denoted by the symbol 20°C. Alcohol is less dense than water, in fact the specific gravity of alcohol is approximately 0.8, or 20% less dense. So, as the yeast consumes the sugar in the wine, and converts it to alcohol, we are lowering the gravity of the must. After fermentation is complete, the specific gravity the wine should be at, or slightly less than 1.00 (Pickering et al., 1998).

pH is a scale used to specify the acidity or basicity of an aqueous solution (lower pH indicates higher acidity). The pH level of a wine ranges from 3 to 4 (Forino et al., 2020).

Sulfites, also commonly called sulfur dioxide, are chemical compounds that contain the sulfite ion. They are found naturally in a variety of food sources, including black tea, peanuts, eggs, and fermented foods. They are also used as a preservative in many foods. Sulfites are a food preservative

widely used in winemaking, thanks to their ability to maintain the flavor and freshness of wine. While they're found in many foods and beverages, they're particularly associated with a long list of side effects related to wine consumption, including the dreaded wine-induced headache (Roullier-Gall et al., 2017).

Alcohol is an organic compound that carries at least one hydroxyl functional group ($-OH$) bound to a saturated carbon atom. Wine can have anywhere between 5% and 23% Alcohol by Volume (ABV). The average alcohol content of wine is about 12%. This amount varies depending on the variety of wine, as well as the winemaker and their desired ABV (Vasiljevic et al., 2018).

These days with the progress of machine learning techniques, and advances in information technologies, it is possible to collect, store and process massive and highly complex datasets. It is possible to classify the wines and select the importance of each chemical parameter for its quality (Aich et al., 2018; Gupta, 2018).

The purpose of this study is to develop and offer an easy access open system for winemakers and customers in which they can easily provide the physicochemical properties of the wine and receive an accurate prediction of the anticipated quality of the wine.

Our goal is that with the help of our trained model, winemakers and customers could examine how a fine change in each component could affect the quality of the wine. They could easily figure out the importance of each physicochemical property in the wine and which one to ignore for reduction of cost. In addition, they could easily figure out the exact concentration of each physicochemical property needed to increase the quality of their wine and apply it accordingly.

Therefore, our first task is to find comprehensive datasets which contain wines' physicochemical properties and the appropriate quality. Our second task is to combine these datasets, build and train models, evaluate their performance and select the best model. Our third task is to provide easy access to the trained model so winemakers and customers could predict the anticipated quality of the wine based on its physicochemical properties. Such a system could help to enhance the production process and reduce the need for human expertise for product quality assurance.

2 Method

We used the Wine Quality Dataset from the UCI machine learning repository which contained two separate datasets for red wine and white wine (Cortez et al., 2009). We combined these two datasets, related to red and white vinho verde wine samples, from the north of Portugal. The datasets were designed to model wine quality based on physicochemical tests.

The datasets contain 6497 different wines' examples. For each example, levels of the eleven physicochemical properties were determined: fixed acidity (g(tartaric acid)/L), volatile acidity (g(acetic acid)/L), citric acid (g/L), residual sugar (g/L), chlorides(g(sodium chloride)/L), free sulfur dioxide (mg/L), total sulfur dioxide (mg/L), density (g/ml), pH , sulphates (g(potassium sulphate)/L), and alcohol (% vol.).The quality of the wines is a score between 1 and 10.

We used the Keras functional API to train a model to predict two outputs. We combined the two datasets to predict the wine quality and whether the wine is red or white solely from the attributes. We modeled wine quality estimations as a regression problem and wine type detection as a classification problem.

Our strategy was to examine different architectures and different numbers of hidden layers. We evaluated the performance of each model and selected the best model for deployment.

3 Experiments and Results

In order to yield the best results we had to perform preprocessing to deal with imbalanced data, as presented in figure 1. We removed classes with quality equal to 3, 4, 8 and 9 because they had very few observations, presented in figure 2. In addition we normalized the data by the mean and standard deviation of each property.
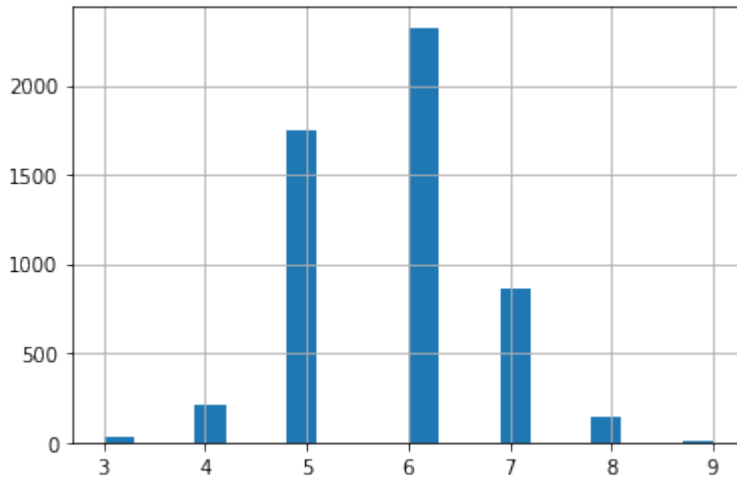
$$Xnorm = \frac{x - u}{\sigma} \tag{1}$$
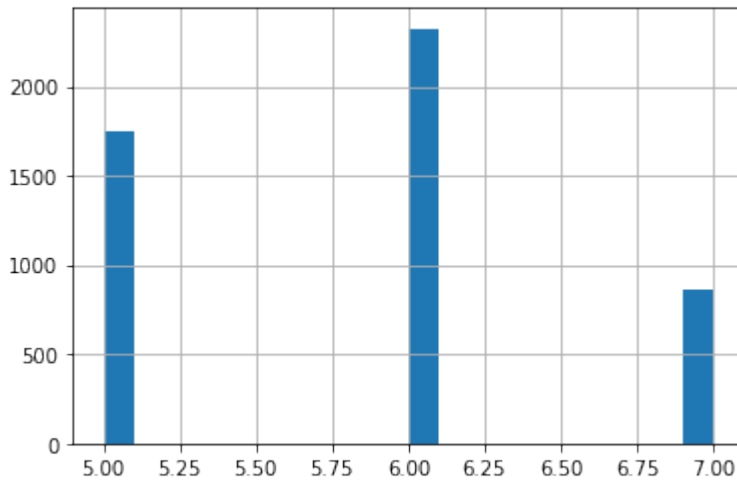
Figure 1: Imbalanced data - the quality of the wine

Figure 2: Balanced data - the quality of the wine

We examined several models.The neural network model with four layers and with relu activation function in the hidden layers, yielded the best results in the regression problem of quality, and in the classification problem of wine type detection.

Figure 3 presents the reduction of root mean squared error during training of wine quality prediction. Figure 4 presents the reduction in loss for wine

type prediction during training. Figure 5 presents the wine type confusion matrix. Figure 6 presents four performance metrics: accuracy, precision, recall and f1-score of the trained wine type model. Figure 7 is the scatter plot for wine quality.
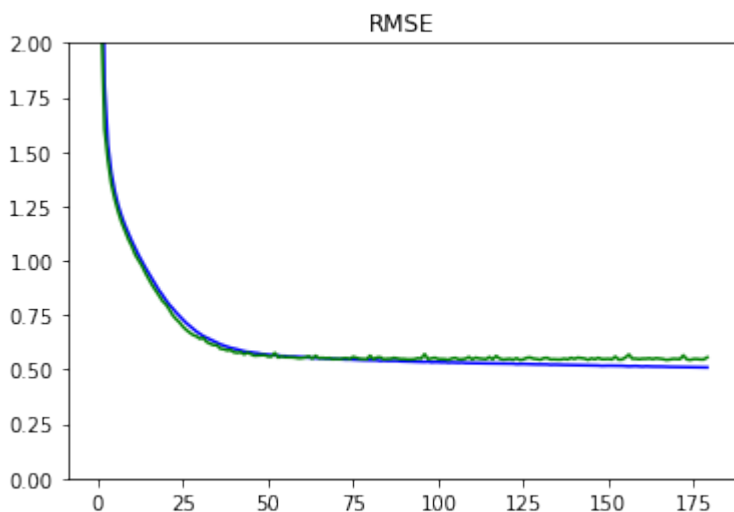

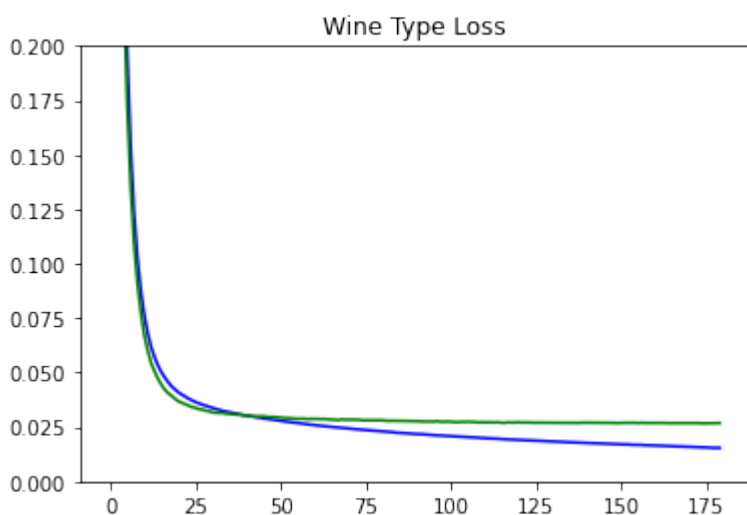Figure 3: Wine quality root mean squared error

Figure 4: Wine type loss

Figure 5: Wine type confusion matrix

```
               precision    recall  f1-score   support

       white       1.00      1.00      1.00       725
         red       1.00      0.99      0.99       262

    accuracy                           1.00       987
   macro avg       1.00      1.00      1.00       987
weighted avg       1.00      1.00      1.00       987

Classification accuracy: 0.996960
```
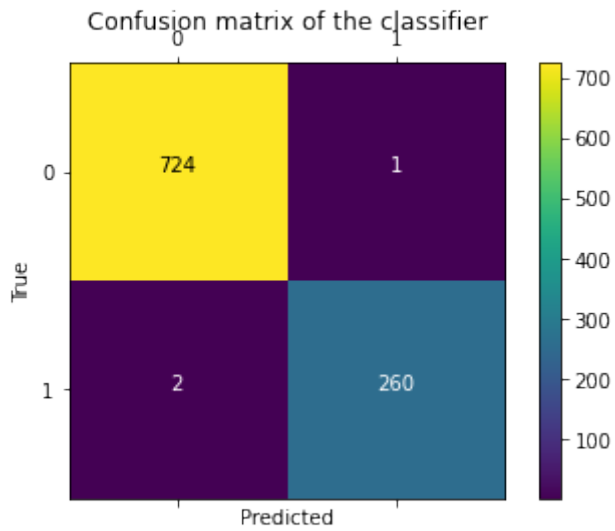
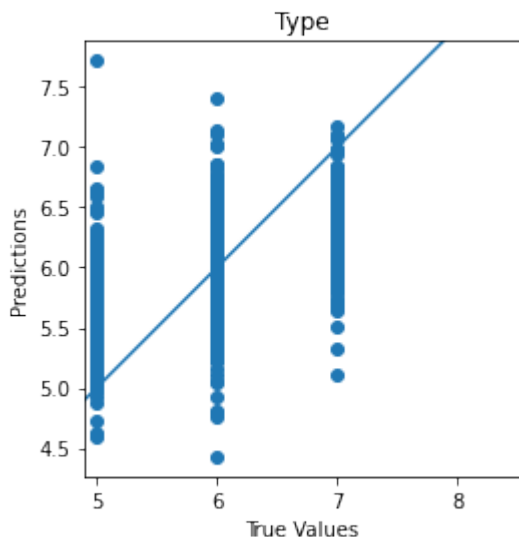Figure 6: Metrics classification report of wine type

9

Figure 7: Scatter plot wine quality

4 Conclusion

The model performs well for prediction of wine quality and type as obtained from the confusion matrix and the loss metrics. We are offering easy access to our trained model. Winemakers and customers can easily receive an accurate prediction of the anticipated quality and type of the wine from its physicochemical properties.

With our free software and access to the trained model,winemakers and customers can examine how a fine change in each physicochemical property could affect the quality of the wine. They could easily figure out the importance of each physicochemical property, and which one to ignore for reduction of cost. In addition, they could easily figure out the exact concentration of each physicochemical property needed to increase the quality of their wine and apply it accordingly. The process is very fast, accurate and does not require taste experts for sensory tests. Sensory tests performed by human taste experts are very complex, slow and expensive.

4.1 The Free Software

The trained model:
`https://drive.google.com/drive/folders/1-Mj5g6xRb200qJWfYnfs3M4cor9pYLrV?usp=sharing`

A file required for normalization of new examples:
`https://drive.google.com/file/d/1CSsnWOGO3WqPs2XEZqnegQzOLLZA721Y/view?usp=sharing`

Test the anticipated quality and type of your wine:
`https://colab.research.google.com/drive/1ASpdSpk02qSVypTx_nU4IL1jDODpAp6a?usp=sharing`

Notebook with preprocessing, training and evaluation:
`https://colab.research.google.com/drive/1XX6GidW1AqY5gBbdXsoRxQ84xnVcg6a9?usp=sharing`

Please send correspondence to Nuriel S. Mor, Ph.D. Software engineering program with an emphasis on AI, Darca and Bnei Akiva schools, Israel. Email: nuriel.mor@gmail.com

5   References

Aich, S., Al-Absi, A. A., Hui, K. L., Lee, J. T., & Sain, M. (2018, February). A classification approach with different feature sets to predict the quality of different types of wine using machine learning techniques. In 2018 20th International conference on advanced communication technology (ICACT) (pp. 139-143). IEEE.

Capece, A., Pietrafesa, R., Siesto, G., & Romano, P. (2020). Biotechnological approach based on selected Saccharomyces cerevisiae starters for reducing the use of sulfur dioxide in wine. Microorganisms, 8(5), 738.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. Decision support systems, 47(4), 547-553.

Cortez, P., Teixeira, J., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009, October). Using data mining for wine quality assessment. In International Conference on Discovery Science (pp. 66-79). Springer, Berlin, Heidelberg.

Dahal, K. R., Dahal, J. N., Banjade, H., & Gaire, S. (2021). Prediction of Wine Quality Using Machine Learning Algorithms. *Open Journal of Statistics, 11*(2), 278-289.

Forino, M., Picariello, L., Rinaldi, A., Moio, L., & Gambuti, A. (2020). How must pH affects the level of red wine phenols. LWT, 129, 109546.

Gupta, U., Patidar, Y., Agarwal, A., & Singh, K. P. (2020). Wine quality analysis using machine learning algorithms. In Micro-Electronics and Telecommunication Engineering (pp. 11-18). Springer, Singapor

Gupta, Y. (2018). Selection of important features and predicting wine quality using machine learning techniques. Procedia Computer Science, 125, 305-312.

Kumar, S., Agrawal, K., & Mandan, N. (2020, January). Red Wine Quality Prediction Using Machine Learning Techniques. In 2020 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-6). IEEE.

Mor, N. S., & Dardeck, K. L. (2021). Applying a Convolutional Neural Network to Screen for Specific Learning Disorder. *Learning Disabilities: A Contemporary Journal, 19*(2), 161-169.

Mor, N. S., & Dardeck, K. L. (2020). Applying Deep Learning to Specific Learning Disorder Screening. *arXiv preprint arXiv:2008.13525.*

Mor, N. S., & Dardeck, K. L. (2018). Quantitative Forecasting of Risk for PTSD Using Ecological Factors: A Deep Learning Application. *Journal of Social, Behavioral, and Health Sciences, 12*(1), 61-73. DOI:10.5590/JSBHS.2018.12.1.04.

Mor, N. S. (2021). Generating Photo-realistic Images from LiDAR Point Clouds with Generative Adversarial Networks. *arXiv preprint arXiv:2112.11245.*

Pickering, G. J., Heatherbell, D. A., Vanhanen, L. P., & Barnes, M. F. (1998). The effect of ethanol concentration on the temporal perception of viscosity and density in white wine. American journal of enology and viticulture, 49(3), 306-318.

Roullier-Gall, C., Hemmler, D., Gonsior, M., Li, Y., Nikolantonaki, M., Aron, A., ... & Schmitt-Kopplin, P. (2017). Sulfites and the wine metabolome. Food chemistry, 237, 106-113.

Sarkar, D., Bali, R., & Sharma, T. (2018). Analyzing Wine Types and Quality. In Practical Machine Learning with Python (pp. 407-446). Apress, Berkeley, CA.

Vallone, R. C., Martínez, L. E., Olmedo, F. G., & Sari, S. E. (2021). Effects of Salinity on Vineyards and Wines from Mendoza, Argentina. In Saline and Alkaline Soils in Latin America (pp. 161-176). Springer, Cham.

Vasiljevic, M., Couturier, D. L., & Marteau, T. M. (2018). Impact on product appeal of labeling wine and beer with (a) lower strength alcohol verbal descriptors and (b) percent alcohol by volume (% ABV): An experimental study. Psychology of Addictive Behaviors, 32(7), 779.

Vilela, A. (2018). Lachancea thermotolerans, the Non-Saccharomyces yeast that reduces the volatile acidity of wines. Fermentation, 4(3), 56.

Wang, B., & Peng, B. (2017). A feasibility study on monitoring residual sugar and alcohol strength in kiwi wine fermentation using a fiberoptic FTNIR spectrometry and PLS regression. Journal of food science, 82(2), 358-363.

Zhong, W., Chen, T., Yang, H., & Li, E. (2020). Isolation and selection of non-Saccharomyces yeasts being capable of degrading citric acid and evaluation its effect on kiwifruit wine fermentation. Fermentation, 6(1), 25.