Comparing the Effectiveness of Multilayer Perceptron and Support Vector Machine for Wine Quality Classification

Aida Hallaci
Aida.Hallaci@city.ac.uk

Abstract

This study compares two supervised machine learning models, Multilayer Perceptron (MLP) and Support Vector Machines (SVM), for predicting wine quality using a publicly available dataset. The accuracy of the MLP model was found to be 84%, meanwhile the SVM model obtained an accuracy of 92%.

1. Introduction

This paper aims to implement two machine-learning models to identify chemical priorities and predict the quality of wines based on them. The models that are going to be implemented on this study are SVM AND MLP, by comparing them, we tend to contribute in the field of the winery for its quality.
The wine industry places great importance on product quality certification, a key factor for consumers and manufacturers. However, the traditional method of quality testing at the end of production can be time-consuming and expensive, as it requires multiple human experts and can lead to subjective opinions.[1] Kumar et.al. (2018) compared linear regression, neural networks, and support vector machines to predict wine quality based on 11 physicochemical characteristics. Their study found that SVM was the most effective technique for wine quality predictions and using selected predictors rather than all predictors improved precision [2].

1.1. Multilayer Perceptron (MLP)

The Multilayer Perceptron (MLP) is an artificial neural network architecture that mimics the organisation of the human brain. It consists of multiple hidden layers between the input and output. Each neuron in the hidden layer is connected to the next layer through weights that are updated during the learning phase until the error is below a certain threshold.[3]

MLP model architecture consists of inputs nodes ('N') referred to as the input layer, and it has an 'X' number of hidden layers, each of which is made up of o certain number of nodes. Depending on the task, one output layer consists of one or more output nodes. The network of MLP is fully connected, where each node in the previous layer is connected to each node in the nest layer. The feed-forward is the process of processing inputs at each node and then passing them to the next node until it generates the output. After the production is forwarded, we calculate the error by taking the difference between the output and actual values. Then, we send this error back to each mode and the corresponding weight so we can reduce the error. This process is called backpropagation. Those two processes continue several times until we get a minimal error.

One of the disadvantages that the model of MLP has, which also occurred in our project, is the process of tuning hyperparameters because it is a time-consuming procedure, and choosing the wrong parameters, can lead to overfitting or underfitting.

1.2. Support Vector Machines (SVM)

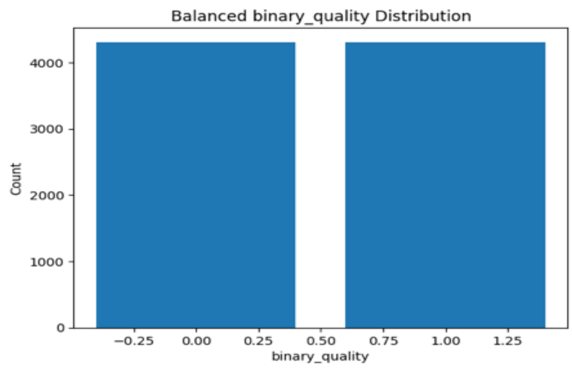
Support Vector Machine (SVM) is a powerful machine learning method used for classification tasks in various text and image classification applications. It has attractive features and promising empirical performance due to its ability to handle limited samples and data dimensionality [4]. One limitation of Support Vector Machines is that their performance can be highly dependent on the choice of kernel function and its associated parameters. The choices should be made carefully because, as we know, gamma is a hyperparameter that has a crucial impact on the shape of the decision boundary.The advantage of SVM over MLP in this study ,is the fact that SVM model is generally better suited for binary classification as our model is. Another benefit of SVM in this paper is that our dataset is a small one,and it is less vulnerable to overfitting when compared to MLP.

2. Dataset

This dataset is obtained from the UCI Machine Learning Repository and contains a total of 6497 instances or rows, with 1599 instances for red wine and 4898 instances for white wine. The main goal of this dataset is to utilise the physicochemical characteristics of the wine to make predictions about its quality. The dataset contains 12 features or attributes, which include sensory data like quality ratings on a scale of 0 to 10 and physicochemical characteristics like acidity, pHCA, alcohol content, and residual sugar.

2.1 Initial Data Analysis

Our research project aimed to evaluate the quality of a dataset.To accomplish this, we transformed the original quality column from a multiclass variable with a range of 0 to 10 into a binary column using a threshold of 6. Essentially, any score less than 6 was given a label of 0, while scores equal to or greater than 6 were assigned a label of 1. During our initial data exploration, we noticed that the quality column was imbalanced, with most instances labelled 0.



To address this issue, we applied the Synthetic Minority Over-sampling Technique (SMOTE) to the dataset, which generated synthetic instances of quality score 1. This helped to increase the proportion of quality scores of 1 in the dataset and reduce the class imbalance. Before using SMOTE, the quality column contained 4,311 instances with a score of 0 and only 1,009 instances with a score of 1. Additionally, we combined the red and white wine datasets at the beginning of our analysis, which resulted in a larger dataset with a total of 6,497 instances.
We verified that the dataset did not contain any missing values. We then addressed the class imbalance issue in the quality column by applying SMOTE, which increased the proportion of quality scores of 1. To normalise the features and prevent any feature from dominating the analysis, we applied StandardScaler. The final pre-processed dataset was used for analysis and modelling.

3. Methodology

1.The wine quality dataset was prepared using a methodology that split the data into a testing and training set. More specifically, 20% of the original dataset was held back as a testing set, and the remaining 80% was used for model selection and training.
2. We used the grid search technique to determine the best-performing models for the SVM and MLP algorithms. This method includes searching through a range of hyperparameters to determine the best combination that results in the highest model performance..
3. We searched over different combinations of hyperparameters using a 10-fold cross-validation approach. Based on the result performed by both MLP and SVM, we choose the best findings.

4.Also for both of models to get a better view of the performance of our classification methods we used ROC curve and confusion matrix

2.1. Architecture used for the MLP

Our study used a Multi-Layer Perceptron (MLP) with a flexible architecture that allows a grid search over a range of hyperparameters to identify the highest model performance. We implemented the GridSearchCV function on the training data to search systematically for the optimal hyperparameters so the model generalises well to unseen data. The hyperparameters being tuned for this model are hidden layer size, activation function, the solver algorithm, the learning rate, the L2 regularization parameter(alpha), and the momentum parameter. Two activation functions were considered for the hidden layers model: Tanh and ReLu. A Tanh function was used because is often utilised for classification tasks which is beneficial for our model, while the ReLu s commonly used as the default activation function for the hidden layers in many neural network architectures.[5]
The maximum number of iterations for the training set is set to 1000 after we tried several iterations, this is done to prevent overfitting and to optimise the weights of the neural network.
The L2 regularization parameter, determined by the alpha hyperparameter in this model, is utilized to prevent overfitting by adding a penalty term to the loss function[6],and is assigned with two values 0.0001 and 0.05. Additionally, we also utilized k-fold cross-validation to validate the model's performance on different subsets of the training data,

2.2. Architecture used for SVM

Our study aimed to evaluate the performance of the support vector machine (SVM) algorithm using four different kernels: polynomial (2,3,4) degree, sigmoid, RBF, and linear.
Choosing the correct kernel is crucial when using the SVM algorithm. The kernel is like a unique function that transforms our data from a low-dimensional space to a higher-dimensional space. So, picking the appropriate kernel is crucial since it affects how our data will be processed because, based on that transformation, it tells how well our SVM model will perform.

a. Our search space of three hyperparameters: kernel function, the gamma value, and the parameter C.
b. We detected that the regularization parameter ‘C’ strongly impacts determining how well our model can perform. As shown in the table, we tested a range of values to define the optimal one.
c. To analyse the performance of the models, we utilised the technique of 10-fold cross-validation.
d. To fit the SVM model we used the ‘RandomizedSearchCV()’ function, so we can evaluate different combinations of hyperparameters.

2.4 Result and Discussion

The hyperparameters for SVM and MLP models were tuned through a grid search to select the best models for predicting wine quality. After we defined using the default hyperparameters for both models, 10-fold cross-validation was performed on the training data through GridSearchCV, so the process of training our model could produce the optimal performance. Based on our result for the SVM model, we detected that the parameter ‘C’ strongly impacts determining how well the trained model we built performs. Since increasing C increases the accuracy, we can say that the classes require a complex decision boundary to be classified. As the table.3 shows the SVM model's accuracy ranged from 54% to 90%, which indicates that small changes that we did in the hyperparameters (C and gamma) resulted into significant changes in the model accuracy; this makes the model of SVM very sensitive. Whereas the MLP model results seem, to vary slightly every time we run it because the model depends on random initialization of the weights. As we observed from the table.3, the accuracy of MLP varies between 65% to 83%.

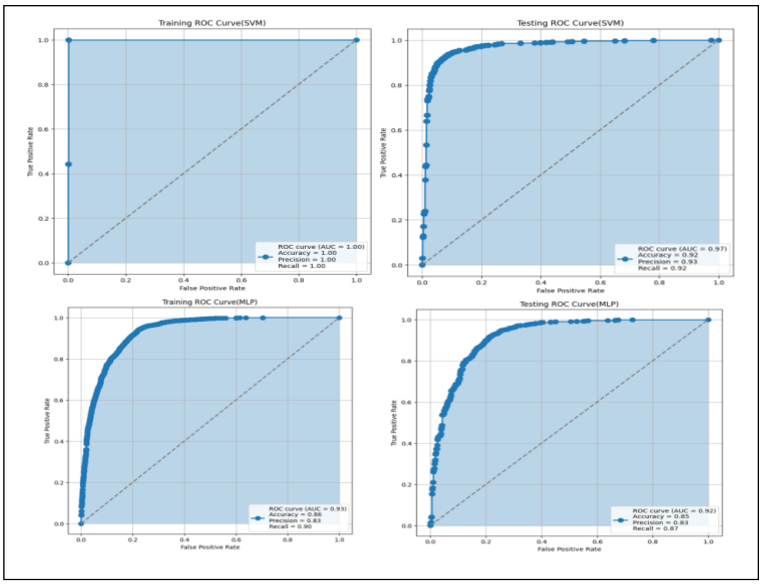
Kernel	Degree	C	Gamma	Accuracy
Poly	3	0.001	-	0.584
Poly	4	0.001	-	0.549
Poly	2	0.01	-	0.624
Poly	3	0.01	-	0.68
Poly	4	0.01	-	0.628
Poly	2	0.1	-	0.688
Linear	-	0.001	-	0.767
Linear	-	0.01	-	0.774
Linear	-	0.1	-	0.778
Linear	-	1	-	0.778
Linear	-	10	-	0.779
RBF	-	10	1	0.904
RBF	-	1	1	0.9
RBF	-	10	0.1	0.86
RBF	-	1	0.1	0.829
RBF	-	0.1	0.1	0.805
RBF	-	10	0.001	0.784
RBF	-	0.1	0.01	0.78
RBF	-	1	0.001	0.777
RBF	-	10	0.01	0.774
RBF	-	1	0.01	0.773

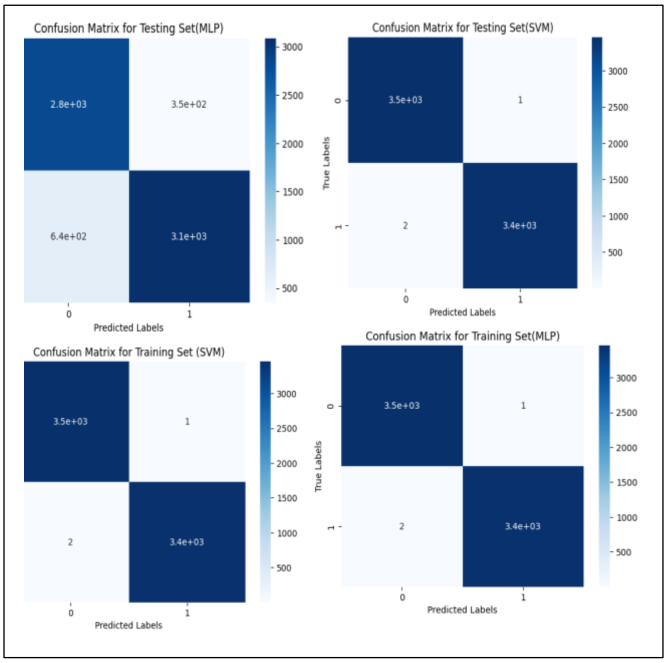
Hidden Layer	Activation	Learning Rate	Momentum	Solver	Accuracy
(15, 15, 15)	Tanh	Constant	0.1	Adam	0.828
(15, 15, 15)	Tanh	Constant	1	SGD	0.793
(15, 15, 15)	Tanh	Constant	1	Adam	0.834
(6, 5)	Tanh	Adaptive	0.1	SGD	0.774
(6, 5)	Tanh	Adaptive	0.1	Adam	0.807
(6, 5)	Tanh	Adaptive	1	SGD	0.658
(6, 5)	Tanh	Adaptive	1	Adam	0.807
(6, 9)	Tanh	Adaptive	1	SGD	0.648
(6, 9)	Tanh	Adaptive	1	Adam	0.804
(5, 5)	relu	constant	0.1	sgd	0.762
(5, 5)	relu	constant	0.1	adam	0.801
(5, 5)	relu	constant	1	sgd	0.753
(5, 5)	relu	constant	1	adam	0.798

Table.3 Hyperparameters Tuning

3. Algorithm Comparison

We analysed the confusion matrix and ROC curve for training and testing data to provide more information about the models. Studying the confusion matrix of SVM for the training set, we recognised that the misclassification rate is 3/6,897 = 0.0435, or 4.35%, which is very low, indicating that the model correctly classifies a significant part of the samples.





On the other hand misclassification rate of MLP is (352 + 639) / (2821 + 352 + 639 + 3085) = 0.1776 or 17.76% , compared to SVM has a higher probability of classifying the data inaccurately. Meanwhile, even in the testing set, the SVM appears to outperform the MLP, with a misclassification rate of 8.4% for SVM and 15.2% for MLP.

Based on the confusion matrix, during the testing process, it indicates that among the total 1725 samples, there were 700 true positives (correctly predicted high-quality wines) and 760 true negatives (correctly predicted low-quality wines). Nonetheless, there were also 151 false negatives (predicted low-quality but were high-quality) and 114 false positives (predicted high-quality but were low-quality). Meanwhile, the confusion matrix for the SVM model displayed that out of 1725 total samples, there were 778 true positives and 811 true negatives, but it also had 73 false negatives and 63 false positives. Based on these statements, it can be said that SVM models performed better.

In the industry of wine quality, it is very important for winemakers to predict its quality accurately. Based on that information, winemakers tend to improve their production process. Another reason why rating the quality of wines accurately is to increase consumer confidence and drive sales. In the case of wine quality, it is very important to minimise false positives (predicting a higher quality when it is lower); because otherwise, it can lead to much damage to the winery's reputation and financial loss; at the same time, minimise false negatives (predicting a lower quality when it is higher), because it could cause the winery the chance to sell. Achieving a balance between minimising false positives and false negatives is essential in accurately predicting wine quality.

In addition, to visualize the performance of the classifiers, we plotted a Receiver Operating Characteristic (ROC) curve. The ROC curve and AUC were plotted for the SVM model, and we observed that during the training process (Figure 2), the AUC was equal to 1.00, which is perfect. However, this does not indicate overfitting because we utilized cross-validation to prevent overfitting. Also, the test accuracy was 92%, which indicates that SVM model is performing well in both training and test process, without the phenomena of overfitting. Examining the test ROC plot for our wine quality prediction models, we saw that both models have high AUC values, with the MLP model is 0.92 and the SVM achieving an AUC of 0.97. In the context of wine quality prediction, SVM is a better model for minimizing false positives, and MLP model has a lower false positive rate However, SVM model performs better overall because of its larger AUC value, which improves the ability to differentiate between false and true classifications.

4. Conclusion

In this study we use two models- Multilayer Perceptron(MLP) and Support Vector Machine to forecast the wine quality and determine whether it is required to manufacture it or not. The findings from both of the machine learning models lead to the conclusion that SVM model outperforms MLP as a model for predicting the wines quality. In terms of accuracy value, misclassification rate performed way better than MLP.

In the future we can expand this code in all data classes, for the binary column which originally was a multiclass variable ranging from 0 to 10,and we can adapt our method to handle multi-class classification models.

Reference:

[1] M. Tariq et al., "Predicting Wine Quality using Multilayer Perceptron and Support Vector Machine," Bachelor Thesis, Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden, 2020. [Online]. Available: https://www.diva-portal.org/smash/get/diva2:1574730/FULLTEXT01.pdf. [Accessed: 29-Mar-2023].
[2] S. Kumar, A. Kumar, and A. K. Singh, "Selection of important features and predicting wine quality using machine learning techniques," International Journal of Computer Science and Mobile Computing, vol. 7, no. 6, pp. 121-129, June 2018. [Online]. Available: https://www.researchgate.net/publication/322347075_Selection_of_important_features_and_predicting_wine_quality_using_machine_learning_techniques. [Accessed: 29-Mar-2023].
[3] Singhal, A., & Sharma, D. K. (2023). Voice signal-based disease diagnosis using IoT and learning algorithms for healthcare. In Implementation of Smart Healthcare Systems using AI, IoT, and Blockchain. Harvard style: Singhal, A. and Sharma, D.K., 2023. Voice signal-based disease diagnosis using IoT and learning algorithms for healthcare. In Implementation of Smart Healthcare Systems using AI, IoT, and Blockchain.
[4] Hasan, K.M.A. and Karim, M.A. (2015) Data classification using support vector machine. In: 2015 2nd International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), Dhaka, Bangladesh, 21-23 May 2015. IEEE, pp. 1-
[5] shiksha.com. (2023). ReLU and Sigmoid Activation Function. [online] Available at: https://www.shiksha.com/online-courses/articles/relu-and-sigmoid-activation-function/ [Accessed 13 Apr. 2023].
[6]McCaffrey, J. (2017, October 5). Neural Network L2 Regularization Using Python. Visual Studio Magazine. https://visualstudiomagazine.com/articles/2017/09/01/neural-network-l2.aspx