A Comparative Study in Classification of Dry Beans through Multilayer Perceptrons and Support Vector Machines

Vighnesh Mane

vighnesh.mane@city.ac.uk

Abstract

The primary goal of this work is to compare and contrast the performance of two different models, namely the multilayer perceptron and the support vector machine. For example, the SMOTE is used to turn imbalanced data into balanced data so that the models can be improved in their correctness. Many hyperparameters are then applied to these models using random search and grid search approaches to find the optimal values for these parameters. For the multilayer perceptron and SVM, the outcomes of training and testing are shown by means of the confusion matrix and the receiver operating characteristic (ROC) curves.

1. Introduction

The studies of the information technology field are bringing about a sea change in the world of agriculture. Dry beans are one of the most significant crops farmed all over the world, and they are one of the most widely planted crops. It takes a significant amount of effort and time on the part of the human to categorize dry beans; however, it can be done relatively quickly and easily by using specific dimension values of the dry beans, which can then be classified into models such as multilayer perceptron and support vector machine, respectively. A significant amount of effort and time on the part of humans will be saved, particularly by those employed in the agricultural industry. When it comes to agricultural expenses, seed is a critical input, and dried beans play a critical role in food technology [2]. Private companies, primarily in the United States, are responsible for the vast bulk of dry bean seed technology operations [2].

The primary goal of this work is to compare and contrast the performance of two different models, namely the multilayer perceptron and the support vector machine. In order to classify the dry beans according to their 16 dimensions, these models will be used.

2. Data

This data set is taken from UCI machine learning repository. Approximately 13000 instances are contained within the system, as well as 16 different features. A total of 7 classes are included in the data collection, and they are as follows: "SEKER", "BARBUNYA," "BOMBAY", "Cali", "HOROZ", "DERMASON", "CALI", "HOROZ”, "SIRA" and "DERMASON."

2.1 Exploratory Data Analysis

This data set is obtained for UCI machine learning repository. Firstly, the data set was imported and after importing as 13611 instances and 16 features was obtained respectively. Then the operation to check null values was performed on the data. Fortunately, no null value was obtained from the data set. After that the unique values of different classes was obtained which can be seen in Fig. 1. Then distribution of the dry bean parameters was done which is depicted in Fig. 2. The observation of this was Numerous characteristics exhibit skewness and outliers in their distributions; these points may resemble a distinct class of dried beans. The distribution of each characteristic among the dry bean classes was investigated. Fig.3 depicts the

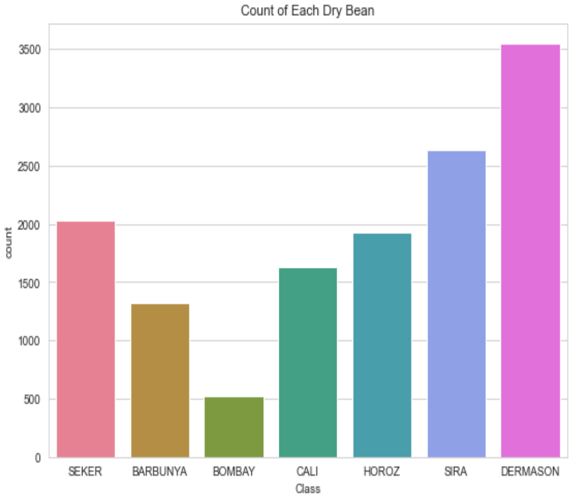

Fig. 1 Count of dry bean classes

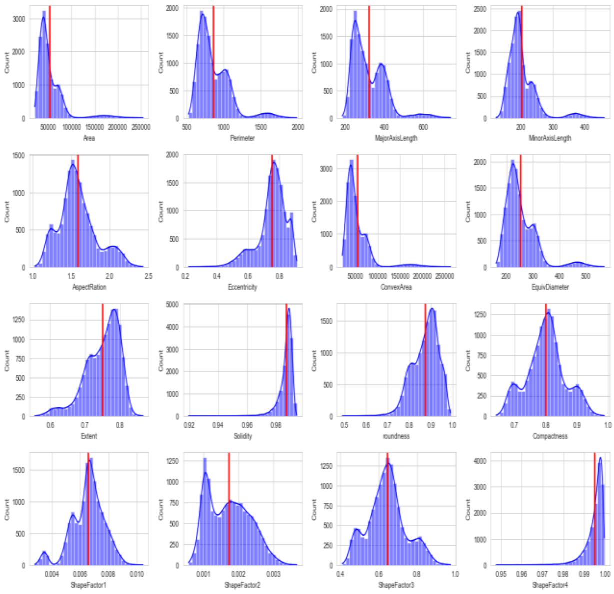

Fig. 2 Distribution of dry bean parameters

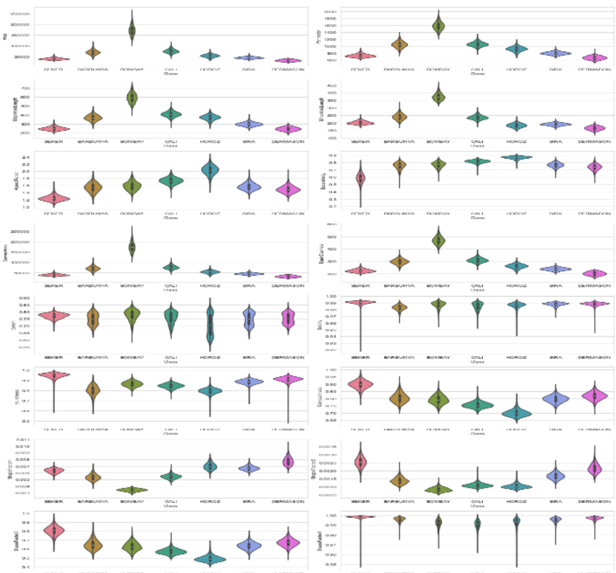

Fig. 3 Distribution(violin) of dry bean parameters

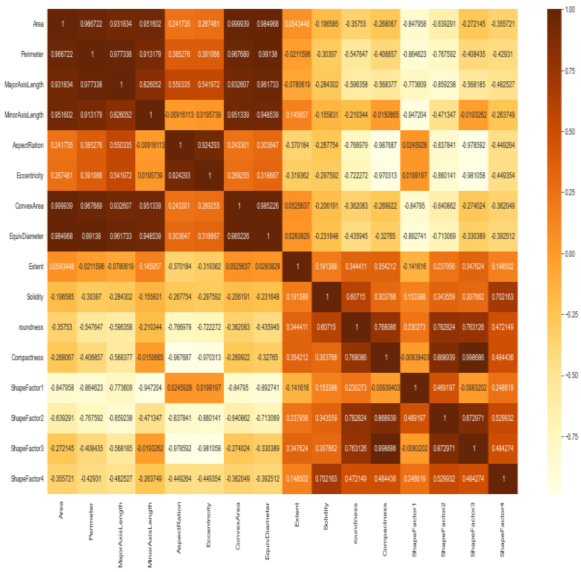

Fig. 4 Correlation matrix

the distribution of the dry bean parameters in which the observation obtained was the Bombay class is significantly different from the other classes; it has a wider area and perimeter and can be easily differentiated from the other classes by its Minor Axis Length and Shape Factor 1 values. Dermason class shares some characteristics with Seker class and some with Sira class. It may be challenging to appropriately label this class. Certain characteristics have a highly skewed distribution with lengthy tails. Further to improve the accuracy the data was scaled using min max scaler technique and to then balance the data the SMOTE technique was applied. Fig. 4 shows the correlation matrix in which many correlations are formed.

3. Summary of Algorithms

3.1 Multilayer Perceptron

MLPs are computer systems that can learn events from examples and determine how reactions to occurrences in the environment are generated [2]. It is determined by the outputs of the input values after they have passed through neurons and been multiplied by the weights of neurons. The MLP architecture is comprised of three layers: an input layer, hidden layers, and an output layer [6]. It is a feedforward neural network having one or more layers between the input layer and the output layer, and it is a feedforward neural network. The data flow occurs solely in one direction from the input layer to the output layer, which is why it is referred to as a feedforward network [5]. Table. 1 shows the advantages and disadvantages of the multilayer perceptron.

MLP Advantages	MLP Disadvantages
• Among its many advantages is that its parameters can be easily modified using a process known as the backpropagation algorithm, which has been shown to be quite effective in practise. This is a significant advantage of the multilayer perceptron. • The capacity to learn non-linear models is a valuable asset. Real-time learning of models is enabled by this feature.	• It is necessary to fine-tune many hyperparameters, including the number of hidden neurons, layers, and iterations. • The loss function of an MLP with hidden layers is non-convex when more than one local minimum occurs. Because of this, random weight initializations may have a significant impact on validation accuracy.

3.1 Support Vector Machine

Kernel-based support vector machines (SVM) are a powerful tool for classifying and forecasting data. It is more general than other machine learning techniques. SVM has a strong theoretical foundation and delivers more accurate results than other algorithms in many situations [2]. SVM is based on the premise that every dataset with a larger dimensionality can be linearly separated. When it comes to datasets with several dimensions, SVM is an excellent choice [5].

SVM Advantages	SVM Disadvantages
• Among its many advantages is that its parameters can be easily modified using a process known as the backpropagation algorithm, which has been shown to be quite effective in practise. This is a significant advantage of the multilayer perceptron. • The capacity to learn non-linear models is a valuable asset. Real-time learning of models is enabled by this feature.	• It is necessary to fine-tune many hyperparameters, including the number of hidden neurons, layers, and iterations. • The loss function of an MLP with hidden layers is non-convex when more than one local minimum occurs. Because of this, random weight initializations may have a significant impact on validation accuracy.

4. Hypothesis Statement

In this study, we will evaluate the effectiveness of a support vector machine (SVM) with a multilayer perceptron for the purposes of classification (MLP). According to the findings of research paper [2], the SVM model outperforms the MLP model in terms of accuracy. Consequently, it is possible that the accuracy of the SVM model will improve for the purposes of this paper. It is possible for the models to have poor performance because the data set is uneven. It has been demonstrated in a study report [7] that the performance of models can be enhanced by using scaling and SMOTE approaches. Although the SVM model may take longer to run than the MLP model, the MLP model may be more useful when performing complex procedures.

5. Methodology

The MLP and SVM models are up against one other in this investigation. Both the training and testing portions of the datasets were created from the same set of data. Most of the information went to training, while only 20 percent was given to testing for evaluation. Following the creation of the baseline MLP model and the assignment of the default parameters, the data was scaled using the min max scaler technique in order to boost its accuracy even more. Data imbalance necessitated use of the SMOTE algorithm. After that, a random search was conducted on that MLP model with randomly generated parameters to see what results were returned. This was followed by a 5-fold grid search on the model, then the cross validation. The model was given a variety of hyperparameters, such as learning rate, optimizer, batch size, epochs, etc., to execute grid search. The divided data was fed into the SVM model for the purposes of training the model. The one versus the rest classifier was utilised because there were more than two classes in the dataset. The model was then subjected to a random grid search as an additional step. Grid search parameters were then applied in the form of hyperparameters such as gamma and kernels.

6. Choice of parameters and experimental results

The first model was the MLP model, this model is useful to perform the complex tasks in neural networks. Firstly, the baseline was created with the input size of 16, hidden layer was 30, dropout was 0.3 and F. relu activation was selected. Also, Adam optimizer was used. Then the model was trained and tested using default parameters but the accuracy was on approximate 19 percent because which the data used for model was scaled and then the model went through the training and testing process one more time and the accuracy obtained was 85.5 percent for training and 84.9 percent for testing. Further to make the data balance SMOTE technique was applied but the training and testing accuracy remained almost same after applying SMOTE. Then further after applying the random grid search and the grid search the accuracy for both training and testing further increased to 93 and 89 percent respectively. For the SVM model the baseline model was created. The SVM model was directly applied to the sampled data. The SVM model gave the accuracy of 91.9 percent for the training side and the 89.9 percent for the testing side. After applying random grid search and the grid search with various hyper parameters this model provided the accuracy of 91.52 percent and 87.63 percent for the training and testing side along with the cross validation of 5-fold.

7. Analysis and critical evaluation of results

The Fig. 5 shows the convolution matrix of training and testing side of the MLP and SVM models respectively. As seen in the Fig. 5 most of the classes are closer to 1 as the SMOTE techniques was applied. This shows that the performance of the both the models good as both the models received the accuracy of the more than 90 percent and also the SVM model received the accuracy of the more than 90 percent which suggests the when classifying the dry beans using certain dimension the MLP model is better as compared to the SVM model and the main disadvantage of the SVM model is that it is time consuming but the results of the SVM model are the accurate as it is just below the MLP model. Fig. 6 show the receiver operating characteristic curve for training and testing side of MLP and SVM models respectively. As shown all the parameters in the ROC figure 6 generates best results as they tend to locate near the 1 so the performance of the models is best and it assesses the accuracy of the model's predictions regardless of the threshold used for categorization.

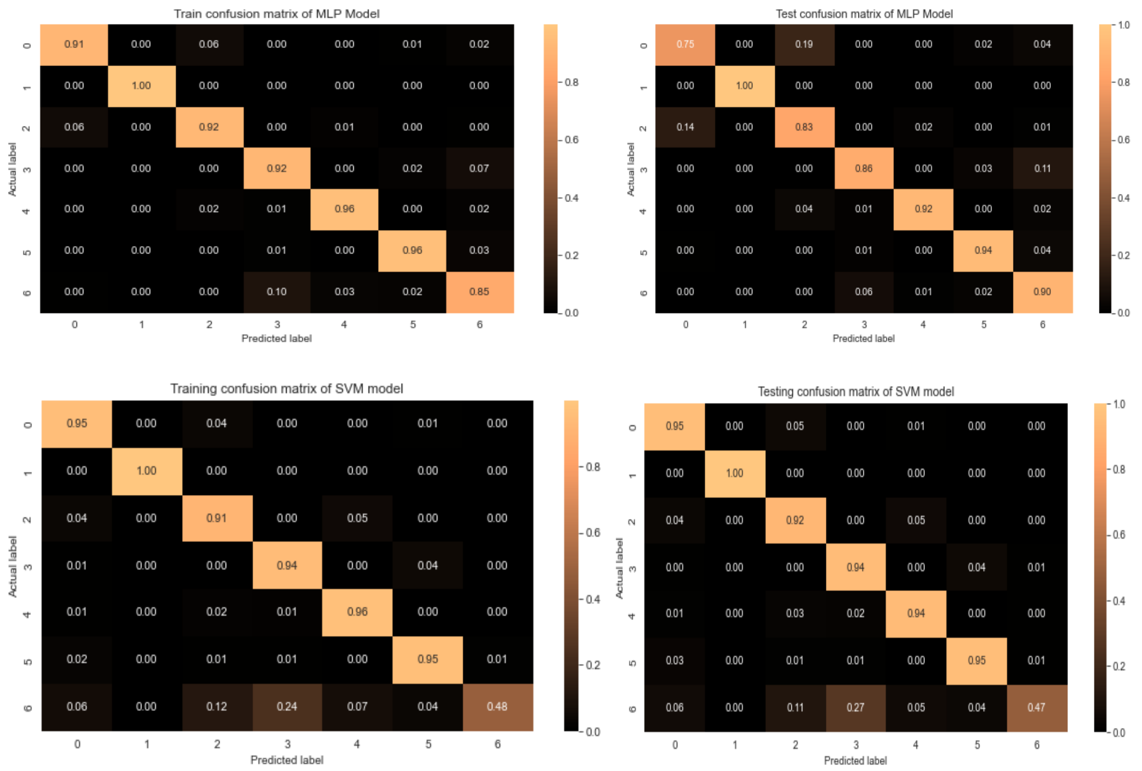

Fig. 5 Convolution matrix for training and testing of MLP and SVM models

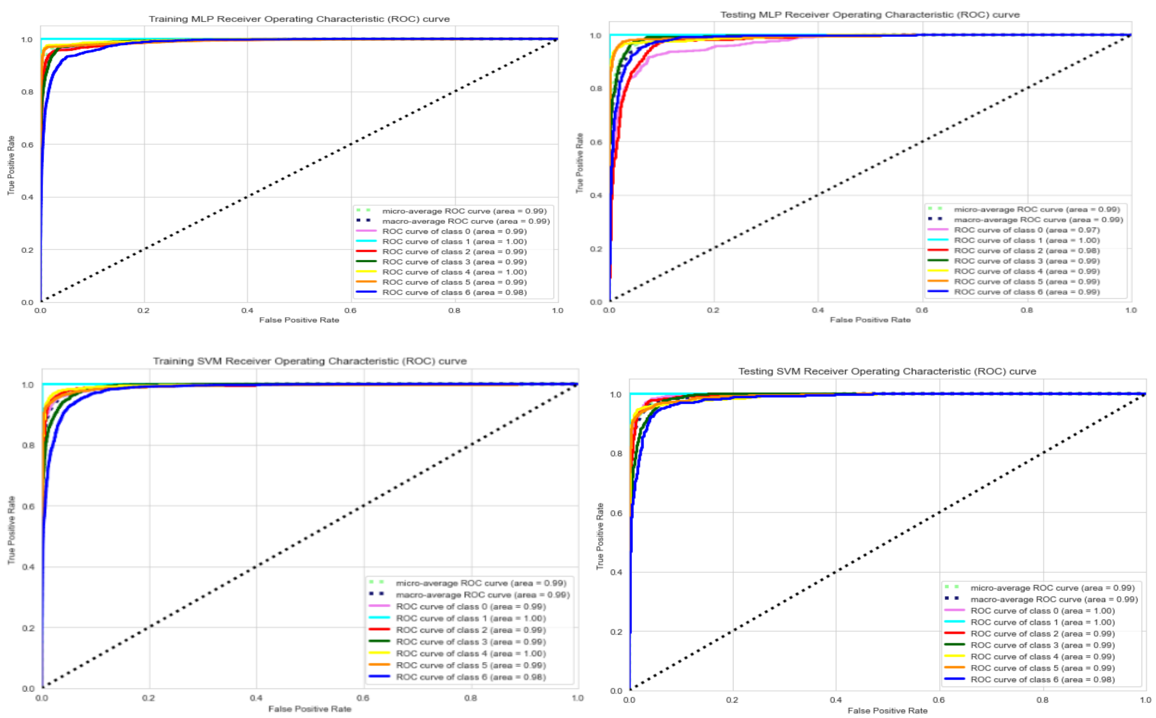

Fig. 6 ROC for training and testing of the MLP and SVM models

8. Conclusion

MLP and SVM models were compared along with each other on the basis of various methodologies. It was observed that the final accuracy of the MLP was higher than the SVM model also in terms of performing the operations the MLP took lesser time as compared to the SVM. Also, as per the hypothesis statement we predicted that the SVM will outperform the MLP model but as the complex task was introduced to SVM model it started consuming more time. Although it got an accuracy which is just below the MLP model but it consumed huge amount of time. This fails the hypothesis statement but SVM only got outperformed by MLP marginally.

Lessons learned are that there might be the possibility of those two models can differ from each other marginally. Also, for future work the training loss curve for the models will be considered which will further help in observing the models from a different perspective.

9.References

[1] Słowiński, Grzegorz. *Dry Beans Classification Using Machine Learning*. http://ceur-ws.org/Vol-2951/paper3.pdf. Accessed 8 May 2022.

[2] Koklu, M. and I AliOzkan. "Multiclass Classification of Dry Beans Using Computer Vision and Machine Learning Techniques - ScienceDirect." *Multiclass Classification of Dry Beans Using Computer Vision and Machine Learning Techniques - ScienceDirect*, www.sciencedirect.com, 0 July 2020, https://www.sciencedirect.com/science/article/pii/S0168169919311573?via%3Dihub.

[3] S. S. Chawathe, "Classification of Dry Beans Using Image Features," 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2021, pp. 0950-0956, doi: 10.1109/UEMCON53757.2021.9666732.

[4] C. H. Mendigoria et al., "Seed Architectural Phenes Prediction and Variety Classification of Dry Beans (Phaseolus vulgaris) Using Machine Learning Algorithms," 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC), 2021, pp. 01-06, doi: 10.1109/R10-HTC53172.2021.9641554.

[5] L. Abhishek, "Optical Character Recognition using Ensemble of SVM, MLP and Extra Trees Classifier," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-4, doi: 10.1109/INCET49848.2020.9154050.

[6] Y. Atahan, A. Elbir, A. Enes Keskin, O. Kiraz, B. Kirval and N. Aydin, "Music Genre Classification Using Acoustic Features and Autoencoders," 2021 Innovations in Intelligent Systems and Applications Conference (ASYU), 2021, pp. 1-5, doi: 10.1109/ASYU52992.2021.9598979.

[7] S. Haldar, R. Mukherjee, P. Chakraborty, S. Banerjee, S. Chaudhury and S. Chatterjee, "Improved Epilepsy Detection method by addressing Class Imbalance Problem," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018, pp. 934-939, doi: 10.1109/IEMCON.2018.8614826.

[8] Dataset link - https://archive.ics.uci.edu/ml/datasets/Dry+Bean+Dataset

Appendix

Glossary

Confusion Matrix - Tables like as the "confusion matrix" are generally applied the accuracy of the model (or "classifier") on a testing dataset for which the true information was available. Nevertheless, even if the confusion matrix is straightforward to grasp, some of the underlying terminology may provide a problem for some people.

Cross-Validation - Models can be tested against non-overlapping subsets of data that were not included in the training sample and see how well they adapt to new data.

Epoch - A complete training loop across the whole information, ensuring that each instance has been examined once before being incorporated to the model. There are N training cycles in a single period, and N is the total value of instances.

Backpropagation- The most commonly used algorithm for gradient descent on neural networks. It begins with a forward pass that evaluates and maintains the ultimate output values for each node. It is then determined in a backwards run along the graph the second derivative for every parameter's inaccuracy.

Baseline- A model that serves as a benchmark for evaluating the performance of a more complicated model.

Cross-entropy- Log Loss may be generalised to multiclass classification issues. To put it another way, cross-entropy measures how different two probabilities distributions are.

Early stopping- A technique of normalisation in which the model training is terminated before the training error reaches a predetermined level. As loss on a validation dataset increases (i.e., generalisation performance degrades), early stopping ends model training.

Batch- Model training samples used in one cycle

Feed Forward neural network- non-cyclic or recursive neural connections

Hyperparameters- The parameters which you may fine-tune as you train a model over time. Hyperparameters include, for example, the batch size, learning rate.

Hidden layer- An artificial layer between the input and output layers of a neural network (the prediction). As a training tool, activation functions (such as ReLU) may be found in the hidden layers

Learning rate- Gradient descent scalars are scalars that are used to train models. The gradient descent technique doubles the learning rate by the gradient for each iteration. The gradient step is the name given to the finished product.

SoftMax- A multiclass classification model function that delivers probability for every conceivable class.

Optimizer- The gradient descent algorithm as it has been performed explicitly.

Perceptron- One layer of a perceptron is used as a binary classifier in a single layered neural network. Sustained learning is used to categorise the data it receives from many sources.

Support Vector Machine- An effort is made by the neural network to identify a hyperplane that separates data points from distinct classes by the greatest margin possible in the feature space.

Multilayer Perceptron- An input, hidden, and output layer feedforward neural network. Backpropagation and other error-correcting algorithms are used to train the network.

Implementation details

In this study two models were compared with each other. Firstly, the baseline model for the created for the MLP model with the input layers, hidden layers and the output layers. After scaling the dataset for the MLP model the accuracy was improved. Then the MLP model was introduced to the SMOTE and then further the MLP model was introduced to the hyperparameters for tuning. Then further the certain performance parameter of the MLP model was taken.

Similarly, for the SVM model the baseline was first created. Then the model was applied to the scaled data. Further SMOTE technique was introduced to the SVM model. Further the SVM model was tuned using various parameters. At last it was observed that the MLP model just marginally outperformed the SVM model.