MAT JOURNALS

Journal of Web Development and Web Designing
Volume 4 Issue 1

A Comparative Study of Supervised Machine Learning Algorithms for Fruit Prediction

Razat Agarwal[1], Prof. Parth Sagar[2]
[1]UG Student, [2]Professor
[1,2]Department of Computer Engineering, RMD Sinhgad School of Engineering, Pune, Maharashtra, India
Email:*razatagarwal98@gmail.com*
DOI: *http://doi.org/10.5281/zenodo.2621205*

Abstract

In this paper, machine learning techniques have been applied for the fruit image classification and prediction over a large dataset. In the implemented work, five models have been developed and their performances are compared in predicting the fruit names. These models are based on five supervised learning techniques i.e., Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor (KNN), Naive Bayes, and Softmax. The experimental results show that Support Vector Machine algorithm performs the best for large datasets and also Support Vector Machine is the best for small datasets. The results also reveal that reduction in the number of fruits reduces the accuracy's of each algorithm.

Keywords: *Machine Learning, Classifier, Random Forest, SVM, KNN, Naive Bayes, Softmax*

INTRODUCTION

Predicting the fruits by the images has always been an attractive topic to researchers and companies. Among those popular methods that have been employed, Machine Learning techniques are very popular due to the capacity of identifying fruits from massive amounts of data that capture the under-lying variety of fruits. In this project, we applied supervised machine learning algorithms for fruit image classification and prediction.

Impact of many factors on the fruit images makes the fruit prediction a difficult and highly complicated task. In order to overcome such difficulties five models have been developed using supervised learning techniques; Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor (KNN), Naive Bayes, and Softmax, and their performances are compared in predicting the fruit images.

The scenario of a shopping mart is taken for fruit prediction where the fruit can be added to the bill only by showing it to the camera hence reducing the effort of the sales person of manually entering the details.

RELATED WORK

Correct prediction of fruits is of great importance for the owner of the shopping mart as incorrect prediction would lead to a loss of business. Many methods have been deployed for the same. Convolutional Neural Network based method was one of the first techniques to be used for fruit prediction [1]. Horea Muresan[2] in the past year applied Neural Network on the same dataset and created the dataset. Kumar [3] did acomparative study on the Stock Market Trend Prediction. In this research paper, the comparative study of the supervised machine learning algorithms using two different sizes of dataset has been proposed. The algorithms have been com-pared based upon the parameter: Size of the dataset. Accuracy has been computed for each algorithm.

RESEARCH DATA

The data used in this research paper has been collected from Kaggle. The data available has 65,429 RGB images comprising of 95 fruits. The image is pre-processed to train the model

faster. The data is converted to a smaller dataset with 18 fruits with 8,846 images in the training set and 2,961 images in the test set. The larger dataset has 95 fruits with 48,905 images in the training set and 16,421 images in the test set.

PROPOSED METHODOLOGY

The proposed architecture for the implemented work mainly consists of five steps: Loading and Preprocessing the data, ap-plying standard scaler, principal component analysis, training the models and result evaluation.

Loading and Preprocessing the Data

The data used in this research paper has been collected from Kaggle. The data available has 65,429 RGB images comprising of 95 fruits. Different variety of the same fruit is classified as different fruits with different names. The image is loaded and then converted into gray-scale to reduce the dimension of the image. Then the image is re-sized for faster training of model. Then the image is flatten to fit the image in a row of data in the table.

Standard Scaler

The flatten image goes through Standard Scaler () which standardize features by removing the mean and scaling to unit variance. This is done to increase the speed of training of the model.

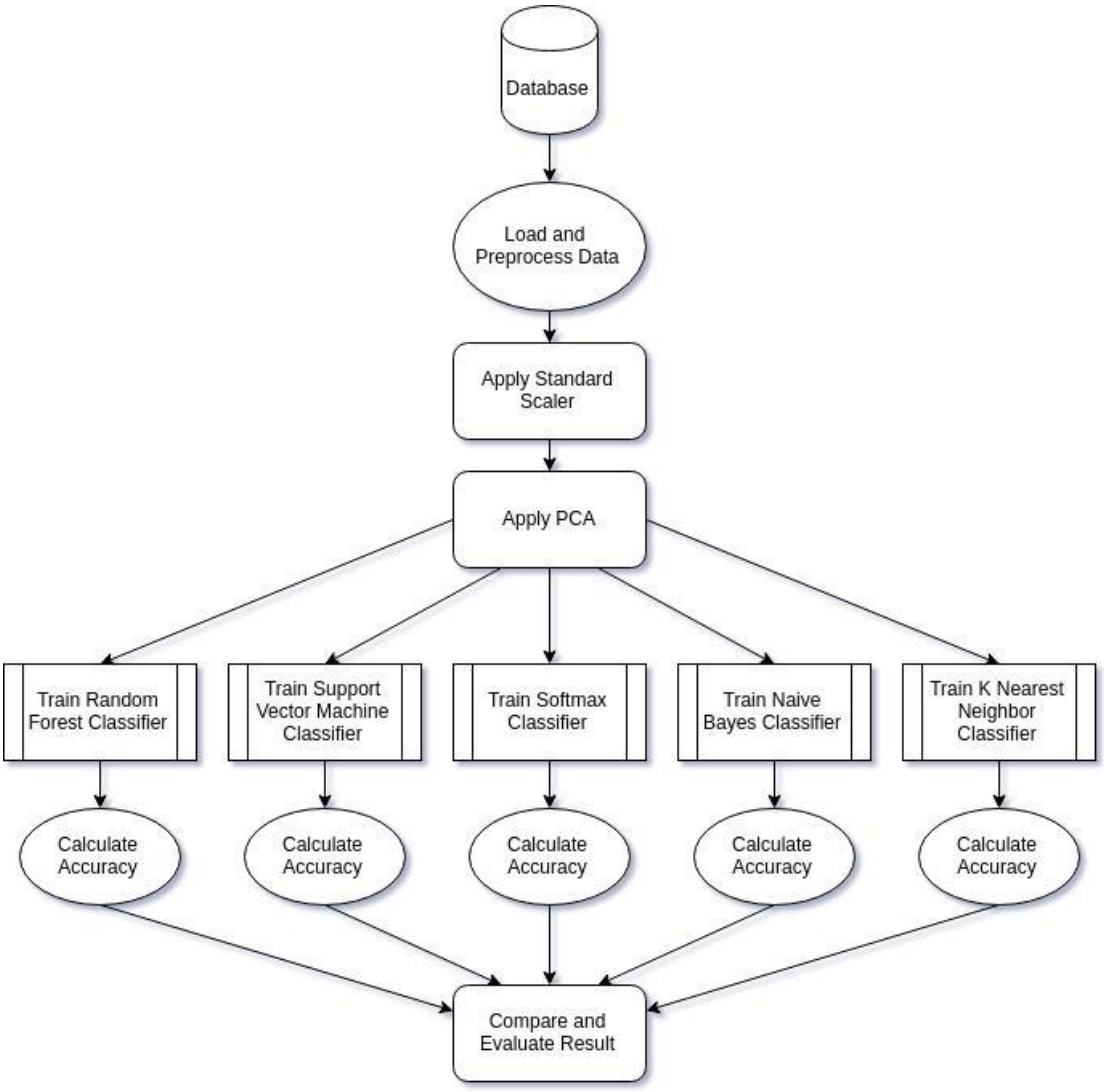

Figure 1: *Architecture of the System.*

Principal Component Analysis
Principal Component Analysis is done on the scaled im-age database to apply linear dimensionality reduction using Singular Value Decomposition of the data to project it to a lower dimensional space. The value of 0.95 for the number of component parameter has been used. It means that scikit-learn choose the minimum number of principal components such that 95% of variance is retained.

Training the model
The training set of images are used to train the five supervised machine learning models to predict the fruit label. The training is done using 48,905 images in the case of large dataset and 8,846 images in the case of small dataset.

Result Evaluation
The results are evaluated for every model using the accuracy to predict the test set images. The test set is of 16,421 images in the case of large dataset and 2,961 images in the case of small dataset. Accuracy is mathematically expressed using the equation given below, where TP is true positive, TN is true negative, FP is false positive and FN is false negative

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

EXPERIMENTAL RESULTS
The results evaluated for 18 fruits (small dataset) are shown in Table I and 95 fruits (large dataset) are shown in Table II.

Table 1: *Accuracy on Small Dataset.*

S. No.	Model	Accuracy
1	Random Forest	85.41%
2	Support Vector Machine	91.18%
3	Softmax	75.88%
4	Naive Bayes	70.98%
5	K Nearest Neighbor	89.09%

Table 2: *Accuracy on Large Dataset.*

S. No.	Model	Accuracy
1	Random Forest	87.77%
2	Support Vector Machine	93.28%
3	Softmax	61.87%
4	Naive Bayes	62.42%
5	K Nearest Neighbor	91.74%

Comparison of different algorithms for small dataset has been shown in Figure 2. It verifies our result that Support Vector Machine Classifier performs best for small dataset. Comparison for large dataset is shown in Figure 3. It verifies our result that Support Vector Machine Classifier performs best for large dataset as well

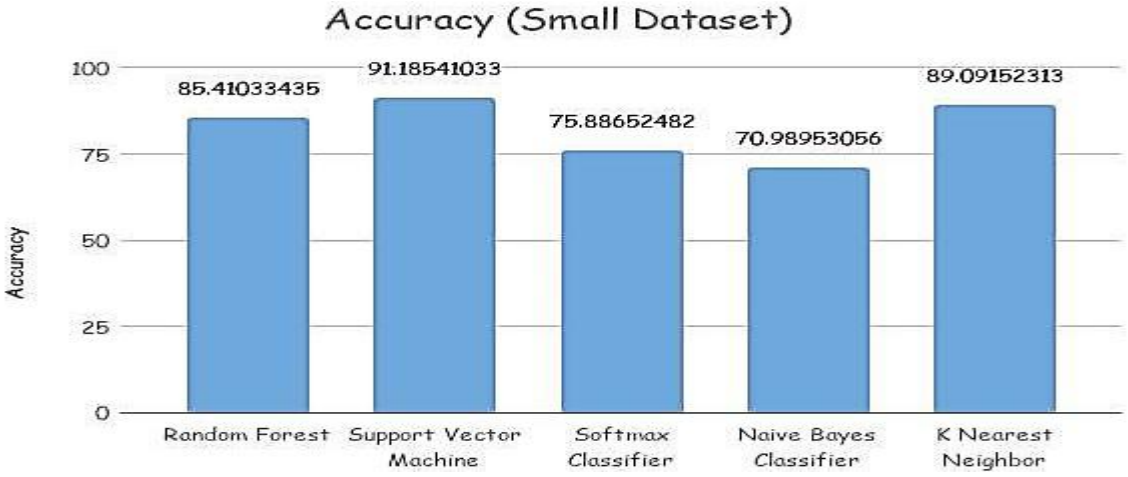

Figure 2: *Accuracy Comparison of Small Dataset.*

CONCLUSION

In this paper, Supervised machine learning algorithms SVM, Random Forest, KNN, Naive Bayes Algorithm, and Softmax Algorithm have been applied for the fruit image classifica-tion and prediction. The results reveal that for large dataset, Support Vector Machine Algorithm out performs all the other algorithms in terms of accuracy and when the size of the dataset is reduced to 20% of the original, then also Support Vector Machine Algorithm shows the best results in terms of accuracy. Also, reduction in the number of fruits reduces the accuracy of each algorithm in predicting the fruit labels.

FUTURE SCOPE

The fruit image prediction can be improved using Convolu-tional Neural Net-works. The accuracy of the work can also beincreased by adding more data of different or same fruits. The work can be expanded and can be evolved to classify multiple fruits in a single image by object detection on the image. The models can be made more accurate by implementing model evaluation and model selection algorithms.

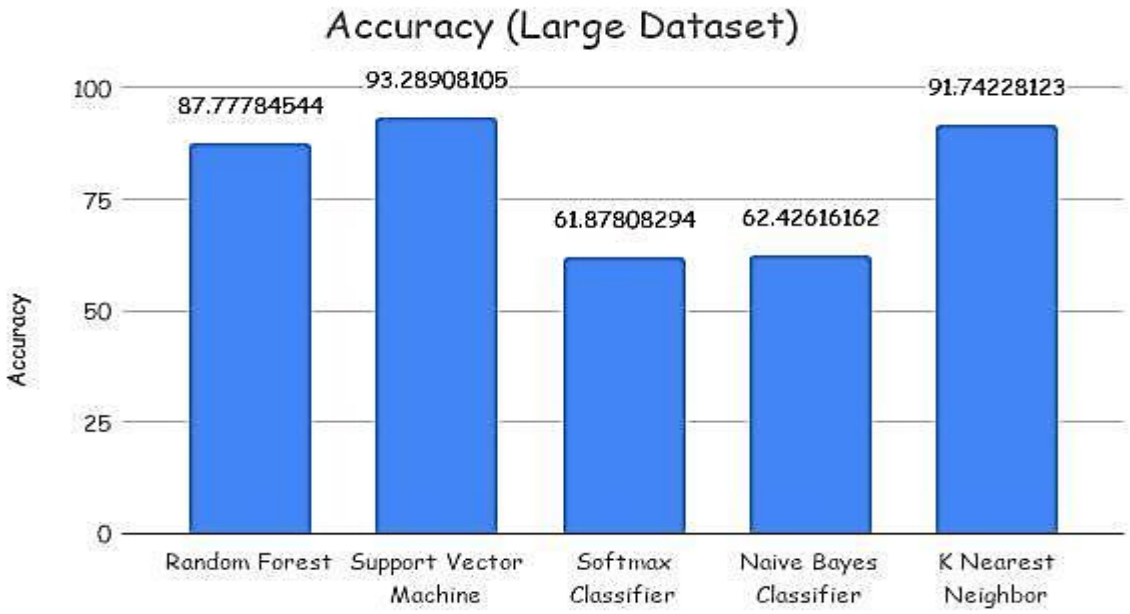

Figure 3: *Accuracy Comparison of Large Dataset.*

ACKNOWLEDGEMENT

We would like to thank all those who helped in making this research paper.

REFERENCES

1. Sa; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; McCool, C. DeepFruits: A Fruit Detection System Using Deep Neural Networks. Sensors 2016, 16, 1222. I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez and C. McCool, DeepFruits: A Fruit Detection System Using Deep Neural Networks, Sensors 2016, 16, 1222.
2. H. Muresan and M. Oltean, Fruit recognition from images using deep learning, Acta Univ. Sapientiae, Informatica Vol. 10, Issue 1, pp. 26-42, 2018.
3. I. Kumar, K. Dogra, C. Utreja and P. Yadav, A comparative study of supervised machine learning algorithms for stock market trend prediction, 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1003-1007.
4. B. Guo, S. R. Gunn, R. I. Damper and J.

D. B. Nelson, Customizing Kernel Functions for SVM-Based Hyperspectral Image Classification, IEEE Transactions on Image Processing, vol. 17, no. 4, pp. 622-629, April 2008.
5. R. Garreta and G. Moncecchi, Learning scikit-learn: Machine Learning in Python, ISBN - 978-1783281930.

Cite this article as: *Razat Agarwal, & Prof. Parth Sagar. (2019). A Comparative Study of Supervised Machine Learning Algorithms for Fruit Prediction. Journal of Web Development and Web Designing, 4(1), 14–18. http://doi.org/10.5281/zenodo.2621205*