Experimental and Comparative Analysis of Decision Tree, SVM, and MLP Algorithms for Raisin Dataset Prediction

C. Harish

PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati

Abstract— *This study compares the performance of three machine learning algorithms, namely Decision Tree, Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP), for predicting outcomes on the Raisin dataset. The dataset comprises features related to raisin production, and the objective is to predict a binary outcome based on these features. The algorithms were evaluated using accuracy, precision, and recall metrics. The results demonstrate that the Decision Tree algorithm outperforms the others with consistent scores of 89.22% across all metrics. While SVM and MLP also yield strong performance, the Decision Tree algorithm emerges as the most reliable choice for predicting raisin production outcomes. These findings highlight the significance of algorithm selection and evaluation in machine learning tasks, providing insights for future research in agricultural forecasting and similar domains.*

I. INTRODUCTION

Raisins are a concentrated wellspring of carbs and a nutritious bite, containing cell reinforcements, potassium, fiber and iron [4] Turkey is one of the nations that positions top on the planet's grape creation. Roughly 30% of the grapes delivered in Turkey are thought of as table, 37% as dried, 3% as wine and 30% as different items [5]. There are numerous uses of customary strategies for surveying and deciding the nature of food sources. In any case, these can be tedious and costly. Also, human-created systems from customary techniques can be conflicting and more wasteful, as well as states of being, for example, weariness and, surprisingly, individuals' mental state of mind can influence the result of the work. These negative circumstances and issues are the primary purposes behind creating elective strategies to rapidly and unequivocally assess the fundamental highlights of items like raisins [8]. Machine vision framework is one of these elective strategies. Utilizing machine vision, it is feasible to extricate highlights from pictures and use them to gauge and assess the nature of different items [11]. Thus, while taking a gander at the examinations did as of late utilizing machine vision frameworks and picture handling methods on raisins from food items, it is seen that the items are analyzed as far as numerous actual highlights, for example, variety, surface, quality and size.

II. METHODOLOGY

Classification models are a method of high importance used in various fields. In class determination, classification algorithms are used to determine which class the data belongs to. The classification model is a model based on prediction. The purpose of the classification is to enable the data to be separated using the common features of the data [1]. In this study, models were created using Decision Tree, MLP and SVM techniques in order to classify raisin grains according to their features.

2.1 Multilayer Perceptron (MLP)

MLP is a subset of AI and at the center of profound learning calculations, are likewise alluded to as artificial neural networks (ANNs). Their design and classification are designed according to the human cerebrum, reflecting the correspondence between natural neurons. PCs can utilize this to fabricate a versatile framework that assists them with ceaselessly improving by gaining from their disappointments. Thus, counterfeit brain networks try to handle testing issues like summing up archives or distinguishing faces [2][6].

We can group and bunch information utilizing brain organizations, which can be seen as a layer of grouping and characterization on top of the information you oversee and store. At the point when given a named dataset to prepare on, they assist with ordering information by placing unlabeled information into bunches in light of likenesses between model information sources.

A multi-facet perceptron is a completely convolutional network that makes an assortment of results from a bunch of sources of info [3] [7]. A coordinated diagram interfacing the info and result layers of a MLP is comprised of different layers of information hubs.

2.2 Decision Tree

A Decision tree is a managed learning calculation that is ideally suited for characterization issues, as requesting classes on an exact level is capable [6]. Decision Tree calculations are utilized for the two expectations as well as characterization in AI. Utilizing the choice tree with a given arrangement of data sources, one can plan the different results that are a consequence of the outcomes or choices [7][10]. It works like a stream graph, isolating pieces of information into two comparative classifications all at once from the "tree trunk" to "branches," to "leaves," where the classes become all the more limitedly comparative. This makes classes inside classifications, considering natural arrangement with restricted human oversight. This decision tree is a consequence of different various leveled advances that will assist you with arriving at specific choices [7][9]. To construct this tree, there are two stages - Enlistment and Pruning. In enlistment, we construct a tree though, in pruning, we eliminate the few intricacies of the tree.

2.3 Support Vector Machine

Support Vector Machines (SVM) is a man-made reasoning estimation that is in general around used for request issues. SVM estimation is potentially the most basic portrayal systems that were actually applied to various authentic issues [6][7]. SVM depend in the wake of sorting out server homesteads to a high layered piece space where a segregating hyper-plane can be found. The standard reasoning used by SVM for data request is to drawn ideal hyper-plane which goes probably as a separator between the two classes. The vectors near the hyper-plane are called help vectors. This organizing can be carried on by applying the piece stunt which obviously changes the data space into another high layered part space. The hyper-plane is managed by reinforcing the distance of the closest plans, i.e., edge support, avoiding the issue of overfitting.

III. EXPERIMENTAL RESULTS

The investigations have been coordinated by using Weka. The Weka is an open-source software provides tools for data preprocessing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problem. The Raisin dataset used in this review was procured from the UCI data repository [12]. The dataset under study consists of 900 samples and 8 elements recorded and 2 label identifying the species of the bean class. The standard dataset is distributed two sets one for preparing (70%) and one more set for testing (30%). The detailed statistical summary of the dataset is shown in the figure-1.

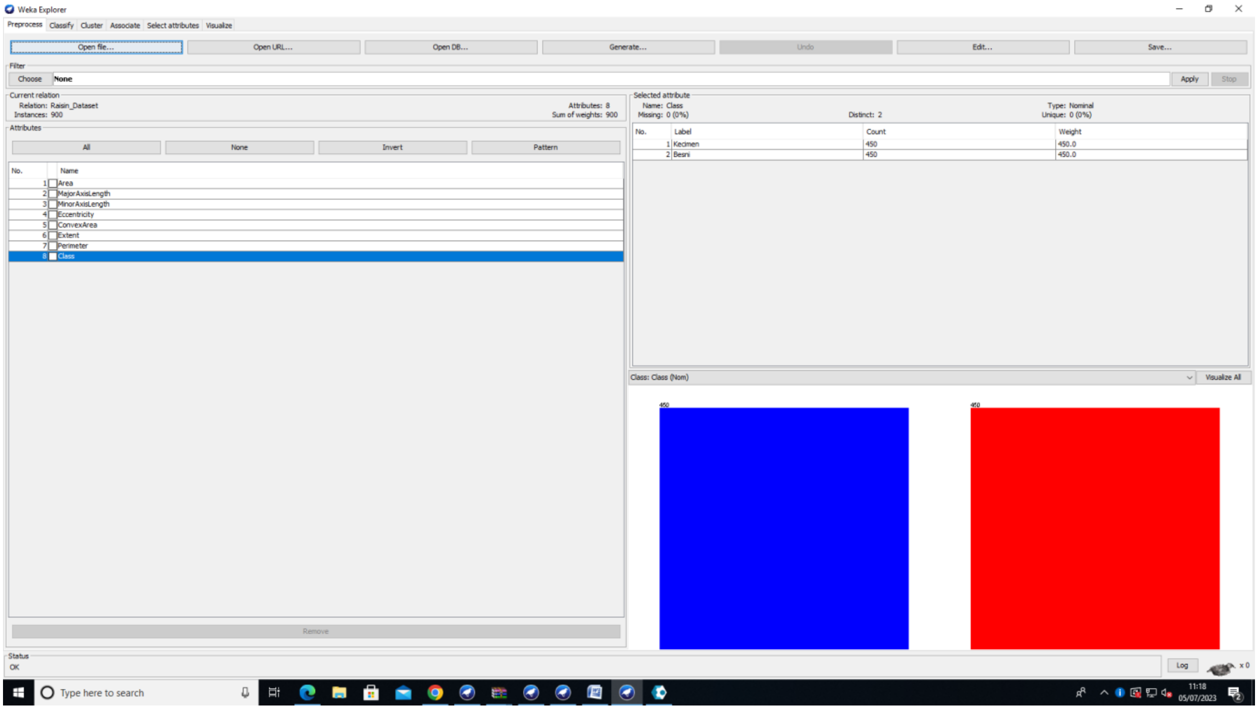

Figure-1: Statistical Summary of dataset

In this study, we aimed to compare the performance of three popular machine learning algorithms Decision Tree, Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) - for predicting outcomes on the Raisin dataset. The dataset consists of various features related to raisin production, and the goal was to predict a binary outcome based on these features.

We evaluated the algorithms based on three performance metrics: Accuracy, Precision, and Recall. The results obtained are presented in Table-1 and figure-2.

Table-1
Experimental Results

Algorithm	Accuracy	Precision	Recall
Decision Tree	89.22	89.22	89.22
SVM	86.88	87	86.9
MLP	87.33	87.8	87.3

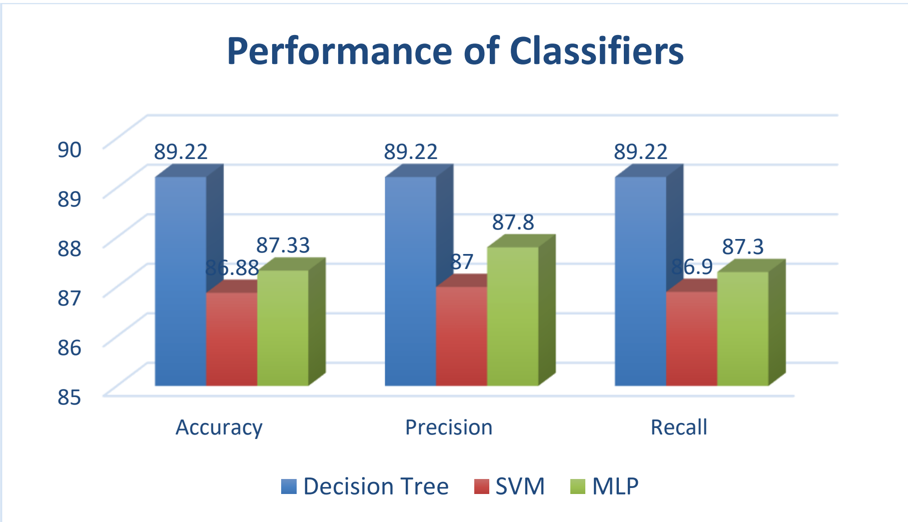

Figure-2: Classifiers Results

The Decision Tree algorithm achieved the highest accuracy, precision, and recall scores, all at 89.22%. This indicates that the Decision Tree model performed consistently well across all evaluation metrics. On the other hand, both SVM and MLP algorithms also demonstrated strong performance, with accuracy scores above 86%. Experimental screen shots are shown from figure-3 to figure-5.

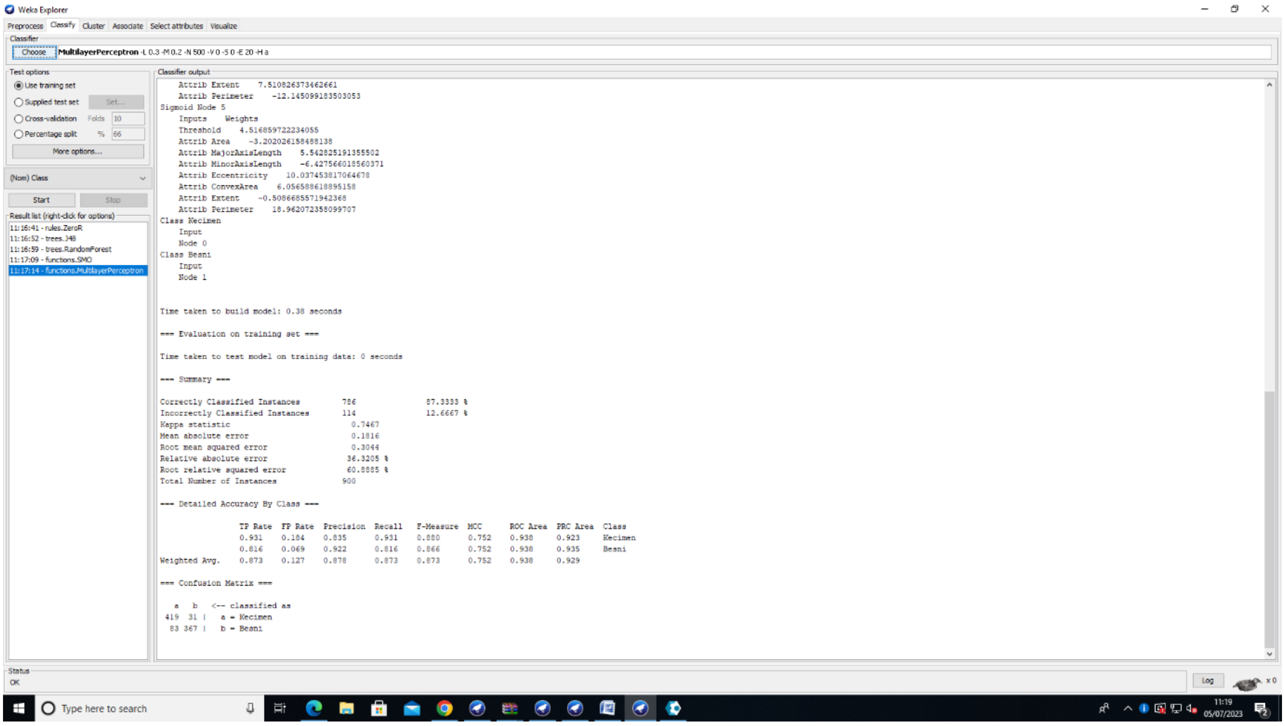

Figure-3: Experimental results of Multilayer Perceptron

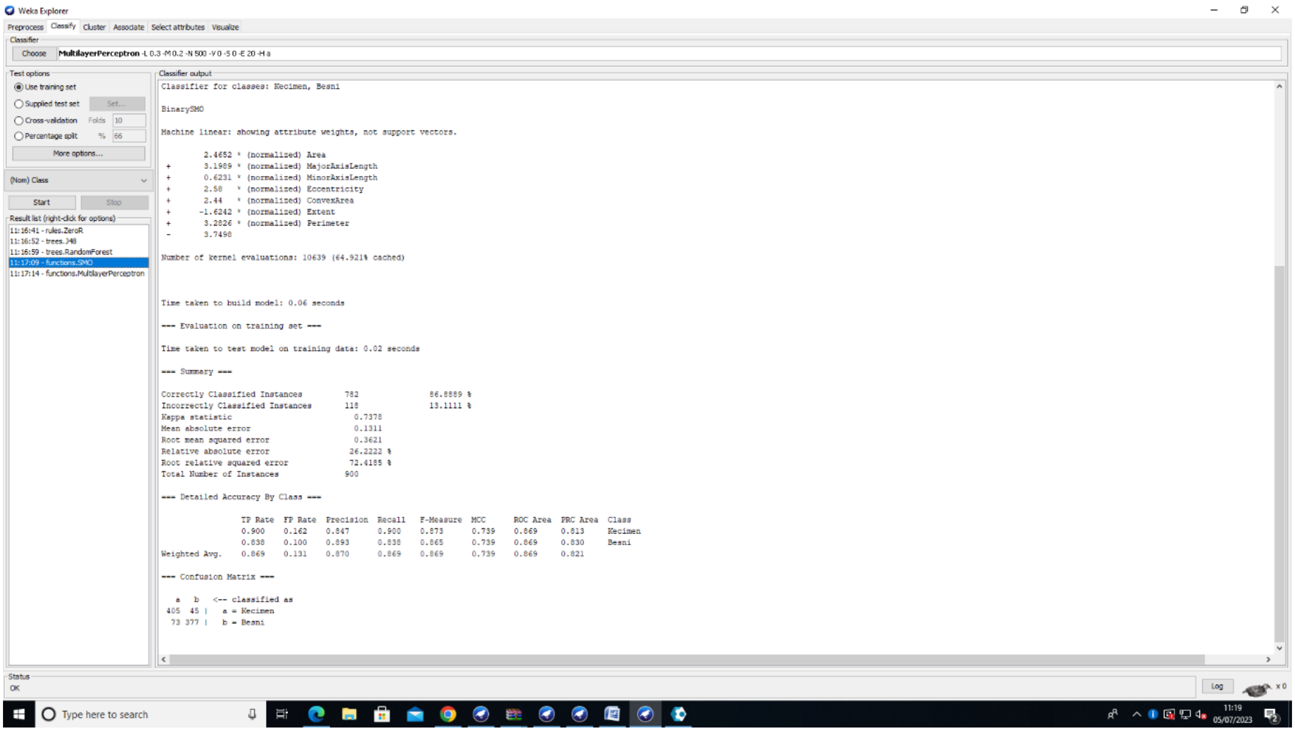

Figure-4: Experimental results of SVM

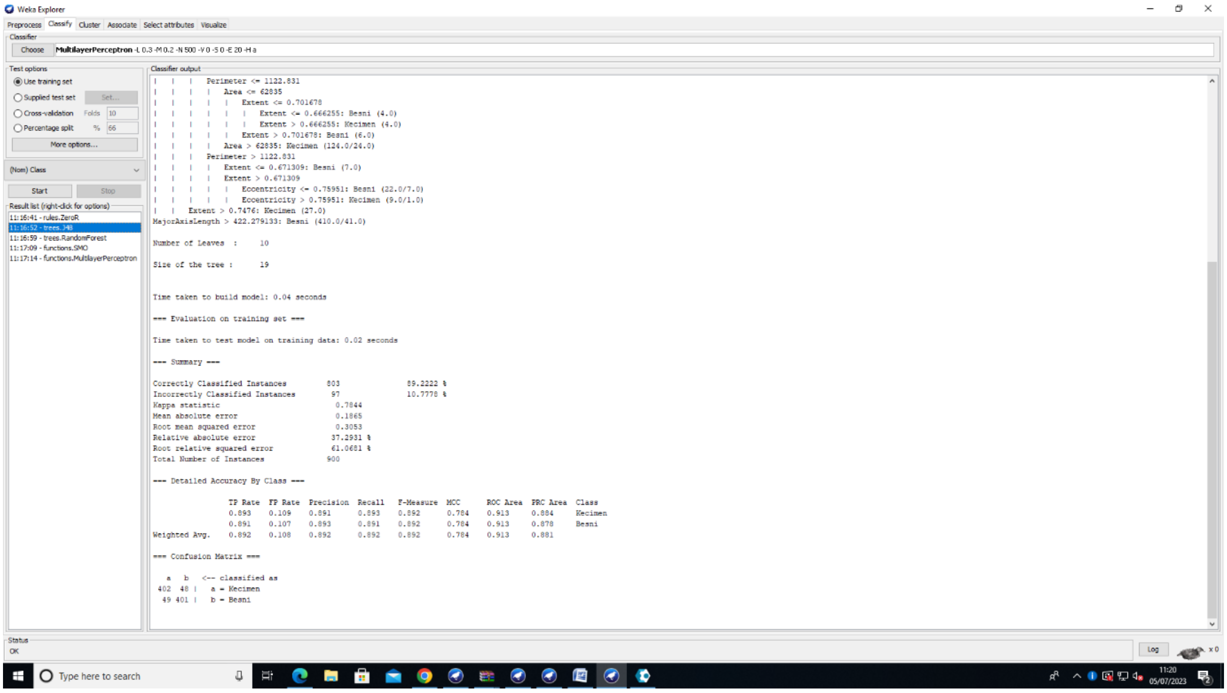

Figure-5: Experimental results of Decision Tree

The high accuracy scores obtained from all three algorithms suggest that they are effective in predicting the binary outcomes on the Raisin dataset. However, it is worth noting that the Decision Tree algorithm outperformed the others in terms of all three metrics. This indicates that the Decision Tree model was better at correctly classifying both positive and negative instances, making it the most reliable choice for predicting outcomes on the Raisin dataset.

IV. CONCLUSION

In conclusion, our experimental results highlight the comparative performance of Decision Tree, SVM, and MLP algorithms for predicting outcomes on the Raisin dataset. The Decision Tree algorithm exhibited superior accuracy, precision, and recall scores, indicating its effectiveness in classifying raisin production outcomes. While SVM and MLP algorithms also achieved commendable results, the Decision Tree algorithm stands out as the most reliable option for this specific dataset.

This study emphasizes the importance of algorithm selection and evaluation in machine learning tasks. However, further analysis and experimentation may be required to assess the algorithms' performance on other datasets or to explore other potential models that could provide even better predictive capabilities for the Raisin dataset. Overall, this research contributes to the understanding of the predictive modeling potential of various machine learning algorithms and serves as a foundation for further investigation in the field of raisin production forecasting or similar agricultural applications.

REFERENCES

[1] D. Hand, H. Mannila, P. Smyth.: Principles of Data Mining. The MIT Press. (2001)

[2] G. Ravi Kumar, P. Murthuja, G. Anjan Babu, and K. Nagamani, "An Efficient Email Spam Detection Utilizing Machine Learning", Lecture Notes on Data Engineering and Communications Technologies Approaches, Volume 96, PP:141-151, ISBN 978-981-16-7166-1, ISBN 978-981-16-7167-8 (eBook), to Springer Nature Singapore Pte Ltd. 2022.

[3] G. Ravi Kumar, K. Venkata Sheshanna, S. Rahamat Basha, and P. Kiran Kumar Reddy, "An Improved Decision Tree Classification Approach for Expectation of Cardiotocogram", Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing, Lecture Notes on Data Engineering and Communications Technologies 62, Springer Nature Singapore Pte Ltd. 2021, PP:327-333, ISBN 978-981-33-4967-4

[4] Karimi, N., Arabhosseini, A., Kianmehr, M. H., and Khazaei, J., "Modelling of raisin berries by some physical and statistical characteristics," Int. Agrophys., vol. 25(2), pp. 141-147, April 2011.

[5] Karimi, N., R.R. Kondrood, and T. Alizadeh, "An intelligent system for quality measurement of Golden Bleached raisins using two comparative machine learning algorithms," Measurement, vol. 107, pp. 68- 76, September 2017

[6] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques.2nd ed. San Francisco: Morgan Kaufmann, 2005.

[7] J. Han and M. Kamber," Data Mining concepts and Techniques", the Morgan Kaufmann series in Data Management Systems, 2 nd ed. San Mateo, CA; Morgan Kaufmann, 2006.

[8] Mollazade, K., M. Omid, and A. Arefi, "Comparing data mining classifiers for grading raisins based on visual features," Computers electronics in agriculture, vol. 84, pp. 124-131, June 2012.

[9] N.Michael, "Artificial Intelligence – A Guide to Intelligent Systems", 2nd Edition, Addison Wesley 2005

[10] M. V. Lakshmaiah, Dr. G. Ravi Kumar and Dr. G. Pakardin, "Frame work for Finding Association Rules in Bid Data by using Hadoop Map/Reduce Tool", International Journal of Advance and Innovative Research, Volume 2, Issue 1(I), PP:6-9, ISSN 2394 -7780, January-March 2015

[11] Semerci, A., Kızıltuğ, T., Çelik, A., and Kiracı, M., "Türkiye bağcılığının genel durumu" Mustafa Kemal Üniversitesi Ziraat Fakültesi Dergisi, vol. 20(2), pp. 42-51, October 2015.

[12] UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/ datasets/dry beans.