Performance Comparative Analysis of Decision Tree and Logistic Regression Algorithms for Dry Beans Prediction

Avarpu Sneha

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

Abstract— *Dry beans are the most generally developed palatable vegetable harvest around the world, with high hereditary variety. Crop creation is firmly affected by seed quality. Thus, seed characterization is significant for both showcasing and creation since it helps fabricate economical cultivating frameworks. The accurate prediction of dry beans can greatly benefit agricultural practices by enabling effective crop management. In this study, we compared the performance of two popular machine learning algorithms, Decision Tree and Logistic Regression, for dry beans prediction. Experimental results showed that the Decision Tree algorithm achieved an accuracy of 95.81%, with precision and recall values of 95.9% and 95.8% respectively. On the other hand, Logistic Regression achieved an accuracy of 92.79%, with precision and recall values of 92.8%. The Decision Tree algorithm outperformed Logistic Regression in all metrics, showcasing its ability to accurately classify dry beans instances. These findings contribute to the growing body of knowledge in machine learning for agricultural applications, providing valuable insights for optimizing crop management strategies*.

I. INTRODUCTION

Individuals eat dry beans, which are a kind of vegetable that is self-pollinated. Beans are a critical harvest on a worldwide scale and are well known with the two ranchers and customers. Dry beans represent almost 50% of the grain vegetables consumed straight by people in most of agricultural nations [1][2]. An arrangement of value control ensures that endorsed seed meets public and worldwide quality benchmarks. For most of food items, visual attributes are the essential measure utilized by customers while going with buying choices [4]. Like other vegetable species, normal beans show the most variety as far as development designs, actual elements (size, shape, and concealing), development, and capacity to develop and adjust [11]. Arranging and ordering bean seeds physically is a tedious interaction. Moreover, this technique is wasteful and dreary, especially while working with enormous creation volumes. Human examiners are generally accountable for actually looking at unrefined components, and smoothing out the auditors' findings is troublesome. These contemplations reaffirm the significance of true estimation frameworks. Thus, programmed evaluating and order strategies are required.

II. CLASSIFICATION

Characterization is a regulated AI strategy where the model attempts to foresee the right name of a given info information. In order, the model is completely prepared utilizing the preparation information, and afterward it is assessed on test information prior to being utilized to perform forecast on new concealed information [7].

III. METHODOLOGY

A wide range of sorts of order procedures have been proposed in writing that incorporates Decision Trees, Naive Bayesian techniques, Neural Networks, Logistic Regression, SVM and KNN and so on. In this paper, we assess the exhibition of the Logistic Regression and Decision Tree for predicting the dry beans.

3.1 Logistic Regression

Strategic relapse is an estimation used to foresee a twofold result: either something occurs, or doesn't. This can be displayed as Yes/No, Valid/Bogus. Autonomous factors are broke down to decide the double result with the outcomes tending to be categorized as one of two classifications [7][8]. The free factors can be all out or numeric, however the reliant variable is dependably clear cut. Composed this way:

$P(Y=1|X)$ or $P(Y=0|X)$

It ascertains the likelihood of ward variable Y, given free factor X. This can be utilized to compute the likelihood of a word having a good or regrettable underlying meaning (0, 1, or on a scale between). Or on the other hand it tends to be utilized to decide the article contained in a photograph (tree, bloom, grass, and so on), with each item given a likelihood somewhere in the range of 0.

3.2 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 [3] [10]. 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 [5][6]. 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][8][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.

IV. 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 dry beans dataset used in this review was procured from the UCI data repository [12]. The dataset under study consists of 13611 samples and 17 elements recorded and 7 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 are shown in the figure1.

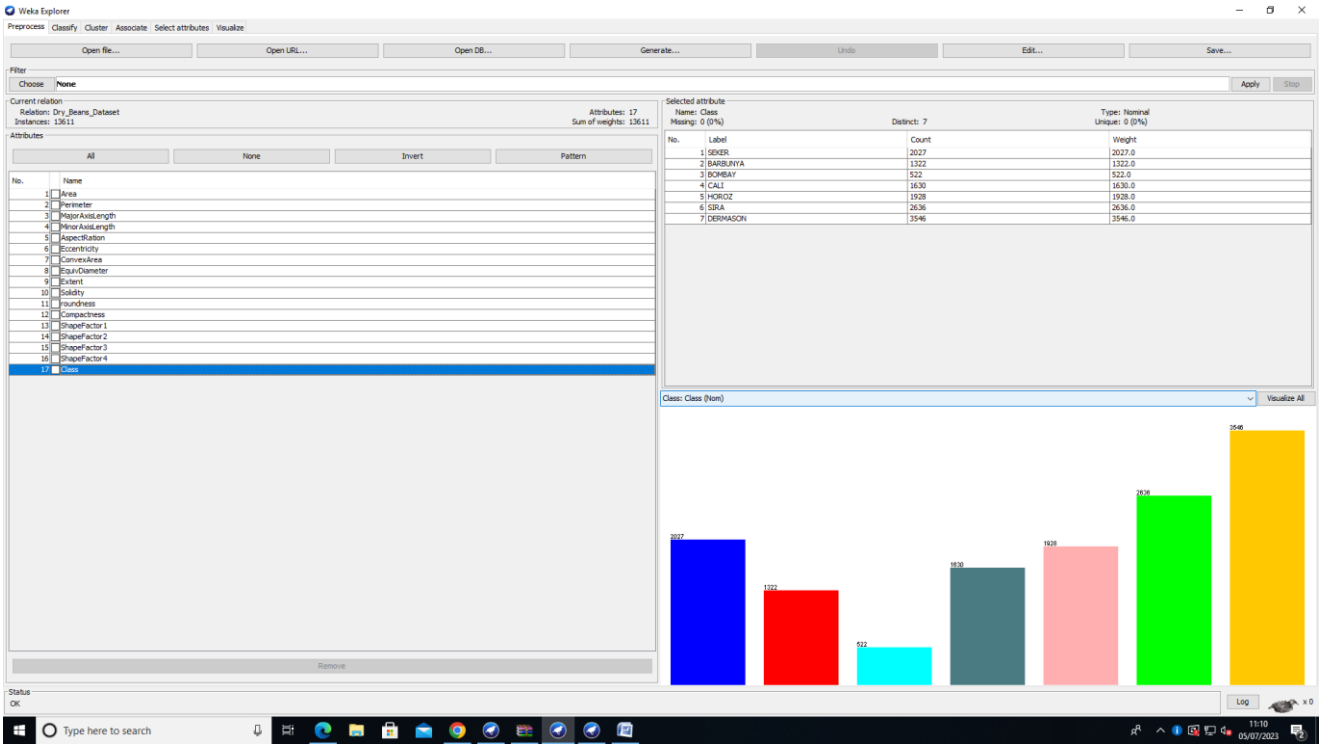

Figure-1: Descriptive statistics of the dry bean dataset.

The experimental results for predicting dry beans using two popular machine learning algorithms, Decision Tree and Logistic Regression, are presented in Table-1 and figure-2. These algorithms were evaluated based on three performance metrics: Accuracy, Precision, and Recall.

Table-1
Experimental Results

Algorithm	Accuracy	Precision	Recall
Decision Tree	95.81	95.9	95.8
Logistic Regression	92.79	92.8	92.8

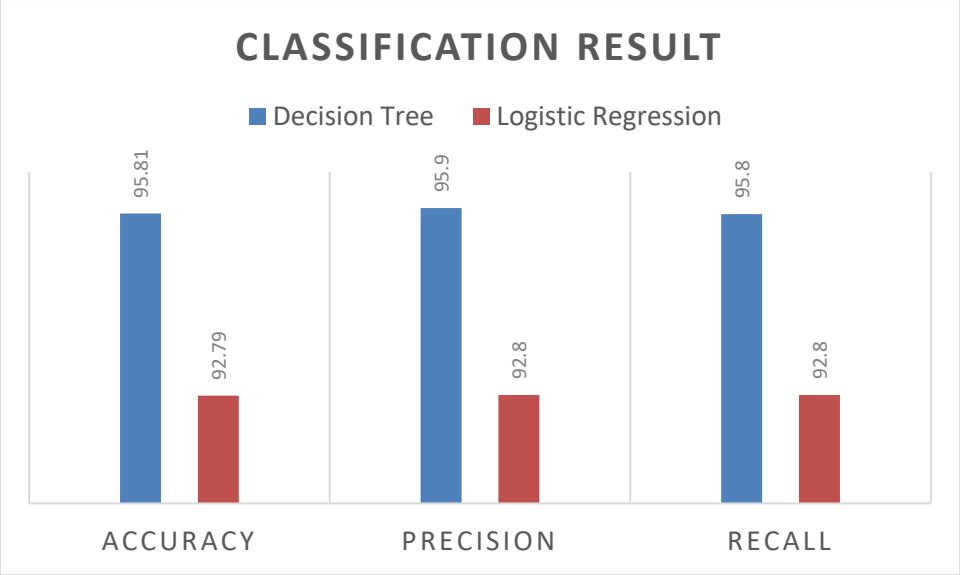

Figure-2: Performance of classifiers

The Decision Tree algorithm achieved an impressive accuracy of 95.81%. This indicates that the model accurately predicted the class labels of the dry beans dataset for a significant portion of the instances. Additionally, the Precision and Recall values for Decision Tree were both high, with precision at 95.9% and recall at 95.8%. This suggests that the Decision Tree algorithm performed well in correctly classifying the positive instances while minimizing false positives and false negatives.

On the other hand, the Logistic Regression algorithm achieved a slightly lower accuracy of 92.79%. Although the accuracy is not as high as that of the Decision Tree, it still demonstrates a reasonable level of predictive performance. The Precision and Recall values for Logistic Regression were also similar, with precision and recall both at 92.8%. This indicates that the Logistic Regression model performed consistently in correctly classifying the positive instances, although not as effectively as the Decision Tree algorithm. The experimental results screen shots of the two classifiers are shown from figure-3 to figure-4.

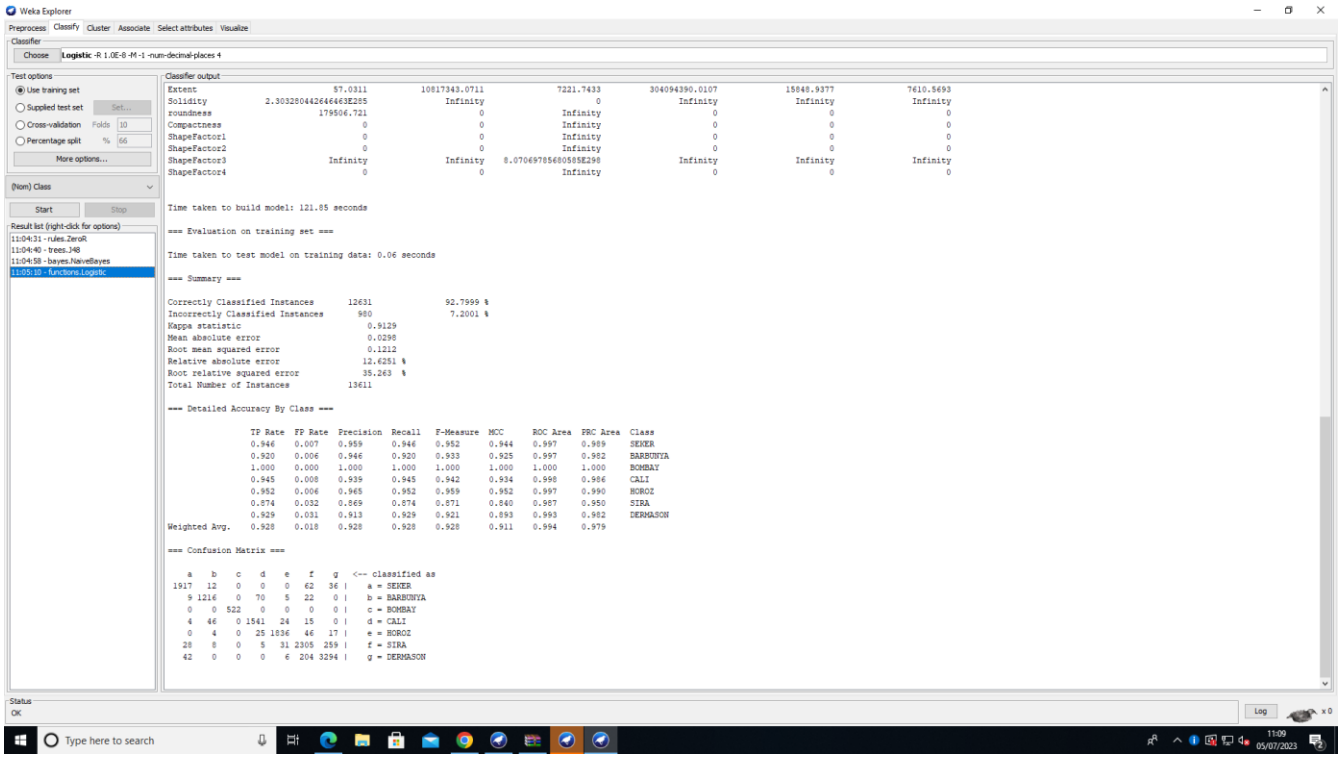

Figure-3: Experimental results of Logistic Regression

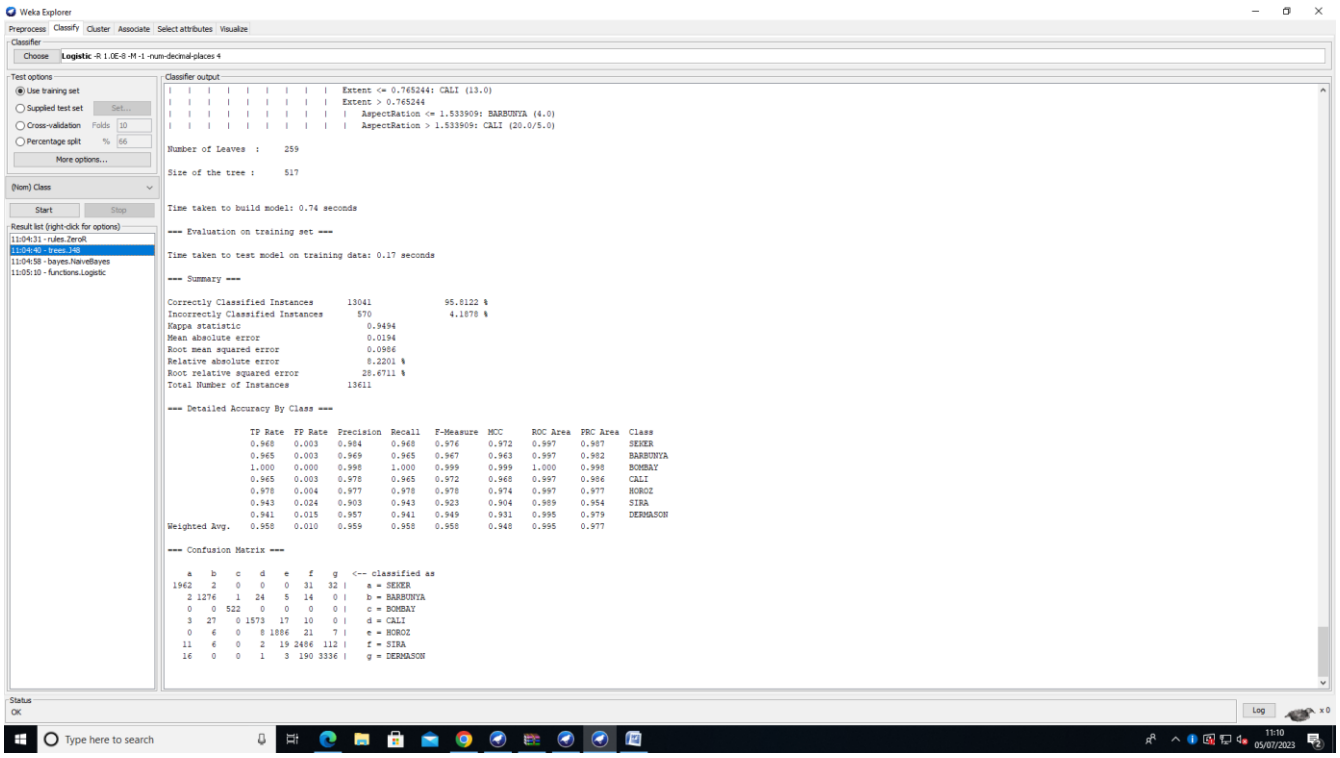

Figure-4: Experimental results of Decision Tree

Comparing the two algorithms, it is evident that the Decision Tree outperformed Logistic Regression in terms of accuracy, precision, and recall. The Decision Tree algorithm's ability to capture complex relationships and make accurate predictions based on the input features contributed to its superior performance. However, it is important to note that the performance of the algorithms can vary depending on the specific characteristics of the dataset and the preprocessing steps applied.

These results highlight the potential of using machine learning algorithms for dry beans prediction. With a high accuracy rate, the Decision Tree algorithm proves to be a promising choice for accurately classifying dry beans instances. However, further analysis and experimentation may be required to explore the robustness of these algorithms across different datasets and to evaluate their performance against other advanced algorithms.

V. CONCLUSION

In conclusion, the experimental results demonstrate the effectiveness of both Decision Tree and Logistic Regression algorithms for dry beans prediction. The Decision Tree algorithm exhibited superior performance in terms of accuracy, precision, and recall, making it a strong candidate for further research and implementation in real-world scenarios. These findings contribute to the growing body of knowledge in machine learning for agricultural applications, potentially aiding in crop management and optimizing agricultural practices.

REFERENCES

[1] A. A. S. Palilo, B. A. Majaja, and B. Kichonge, “Physical and mechanical properties of selected common beans (Phaseolus vulgaris L.) cultivated in Tanzania,” Journal of Engineering, vol. 2018, Article ID 8134975, 9 pages, 2018.

[2] C. Larochelle, E. Katungi, and Z. Cheng, “Household consumption and demand for bean in Uganda,” Determinants and implications for nutrition security, vol. 23, 2016.

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

[4] F. Maalouf, S. Ahmed, and Z. Bishaw, “Chapter 6 - faba bean,” in The Beans and the Peas, Wood head Publishing, Sawston, UK, 2021.

[5] 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.

[6] 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

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

[8] 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.

[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] S. Mamidi, M. Rossi, D. Annam et al., "Investigation of the domestication of common bean (Phaseolus vulgaris) using multilocus sequence data," Functional Plant Biology, vol. 38, no. 12, pp. 953–967, 2011.

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