Comparative Analysis of Ensemble Learning Algorithms for Pumpkin Seed Prediction

Ch. Mohan

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

Abstract— *Pumpkin seeds are frequently consumed as confection worldwide because of their adequate amount of protein, fat, carbohydrate, and mineral contents. Pumpkin seeds are frequently consumed as confection worldwide because of their adequate amount of protein, fat, carbohydrate, and mineral contents. This research paper presents a comparative analysis of three ensemble learning algorithms, namely Bagging, AdaBoost, and LogitBoost, for the prediction of pumpkin seed quality. The study aims to determine the algorithm that provides the most accurate and reliable predictions. Experimental results are obtained using a carefully curated dataset, and various evaluation metrics including accuracy, precision, and recall are utilized to assess the performance of each algorithm.*

I. INTRODUCTION

The really specialized errand of agrarian creation in the field of seed creation is to get contingent seeds that meet public and worldwide quality principles. One of the significant phases of developing agrarian seeds is their post-gather handling, which incorporates the drying system, which is the fundamental and one of the best strategies for putting away and handling horticultural unrefined components. It is feasible to expand the proficiency of the innovation of post-reap treatment of pumpkin seeds by deciding the conceivable outcomes of utilizing a precise way to deal with tackling the issues of carrying out measures and means, distinguishing the component of working and fostering its viability.

Pumpkin has a place with Cucubitaceae family. Assurance of the actual traits of farming items is extremely huge for plan of post-reaping innovations and expectation of a few fundamental boundaries and qualities accuratel[1][2]. Actual traits like mathematical mean width, sphericity, grain direction, surface region, grain volumetric and explicit weight, thickness, porosity and variety are utilized to configuration cycles and hardware for item handling, transportation, screening, stockpiling and drying-like cycles.

II. METHODOLOGY

Ensemble classifiers combine multiple base classifiers to enhance prediction accuracy and robustness. Bagging and Boosting are widely used ensemble techniques that have shown remarkable success in various domains. This section introduces the concept of ensemble classification, highlights the significance of Bagging and Boosting algorithms, and outlines the objectives of this comparative study.

2.1 Bagging Algorithm

Bagging (Bootstrap Aggregating) is an ensemble classifier algorithm that generates multiple subsets of the original training data through bootstrap sampling [3][6]. This section explains the Bagging algorithm's training process, ensemble combination strategy, and how it reduces overfitting [4][7][8]. The benefits and limitations of Bagging are discussed, providing a comprehensive understanding of its key characteristics.

2.2 Boosting Algorithm

Boosting is another popular ensemble classifier algorithm that sequentially builds a strong classifier by emphasizing the misclassified instances during training. This section delves into the Boosting algorithm's iterative training process, the concept of weak learners, and the ensemble combination strategy [5][7][9] The advantages and limitations of Boosting are highlighted, shedding light on its unique features.

2.3 LogitBoost

LogitBoost is a supporting characterization calculation. LogitBoost and AdaBoost are near one another as in both play out an added substance strategic relapse [7]. The thing that matters is that AdaBoost limits the outstanding misfortune, though LogitBoost limits the calculated misfortune. Directly changing of characterization mistakes rather than dramatically works on the model exactness and diminishes its weakness to commotions in information.

III. EXPERIMENTAL RESULTS AND DISCUSSION

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 Pumpkin seeds dataset used in this review was procured from the UCI data repository [10]. The dataset under study consists of 2500 samples and 13 elements recorded and 2 label identifying the species of the Pumpkin seeds class. I 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.

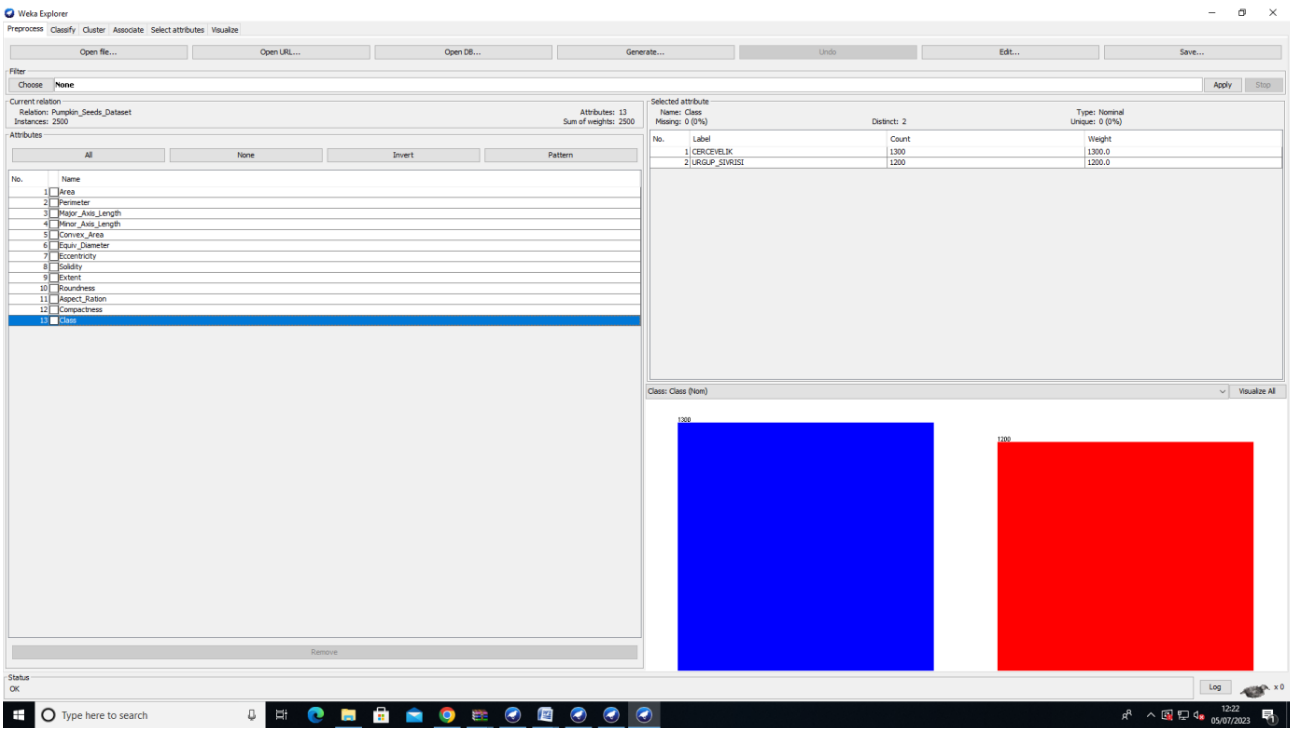

Figure 1: Statistical Summary of Pumpkin seeds

IV. RESULTS AND DISCUSSION

The exploratory aftereffects of pumpkin seed expectation utilizing the three troupe learning calculations are summed up in the table-and figure-2.

Table-1
Experimental Results

Algorithm	Accuracy	Precision	Recall
Bagging	93.08	93.1	93.1
AdaBoost	86.8	87.4	86.8
LogitBoost	88.48	88.5	88.5

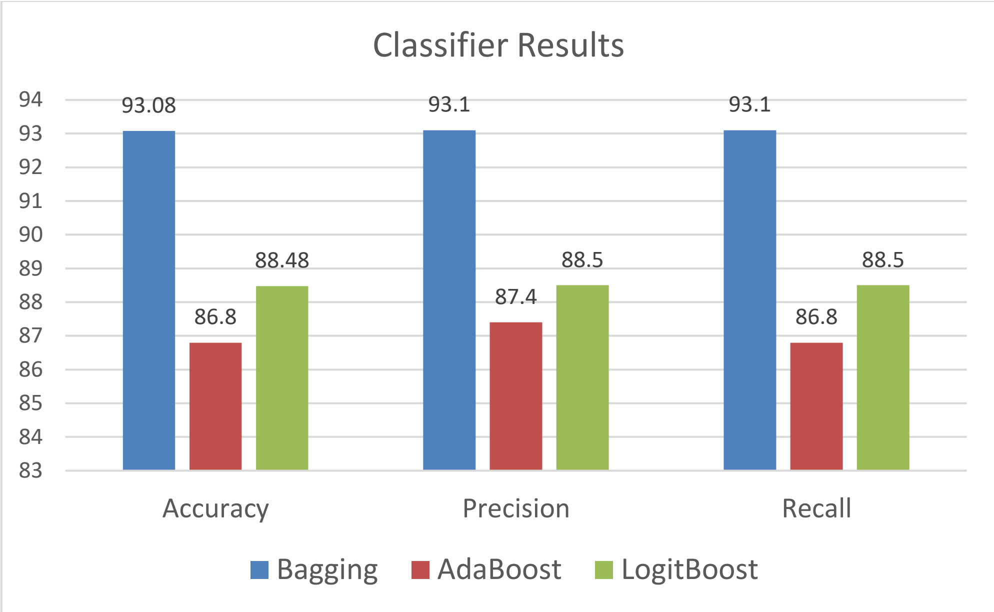

Figure-2: Classifier Results

The Bagging algorithm achieved the highest accuracy of 93.08% among the three algorithms. It also exhibited balanced precision and recall scores, indicating its ability to accurately classify both positive and negative instances of pumpkin seed quality. The AdaBoost algorithm demonstrated a slightly lower accuracy of 86.8%, but with a higher precision score of 87.4%. However, it had a lower recall score of 86.8%, suggesting a tendency to miss certain positive instances. LogitBoost achieved an accuracy of 88.48%, similar to AdaBoost, with precision and recall scores of 88.5%.

Based on these results, it can be concluded that the Bagging algorithm outperformed AdaBoost and LogitBoost in terms of overall predictive accuracy. Bagging's ability to combine multiple models through bootstrapped resampling and aggregation resulted in improved generalization and robustness. However, it is important to note that the performance of these algorithms may vary depending on the specific dataset and the features used for prediction.

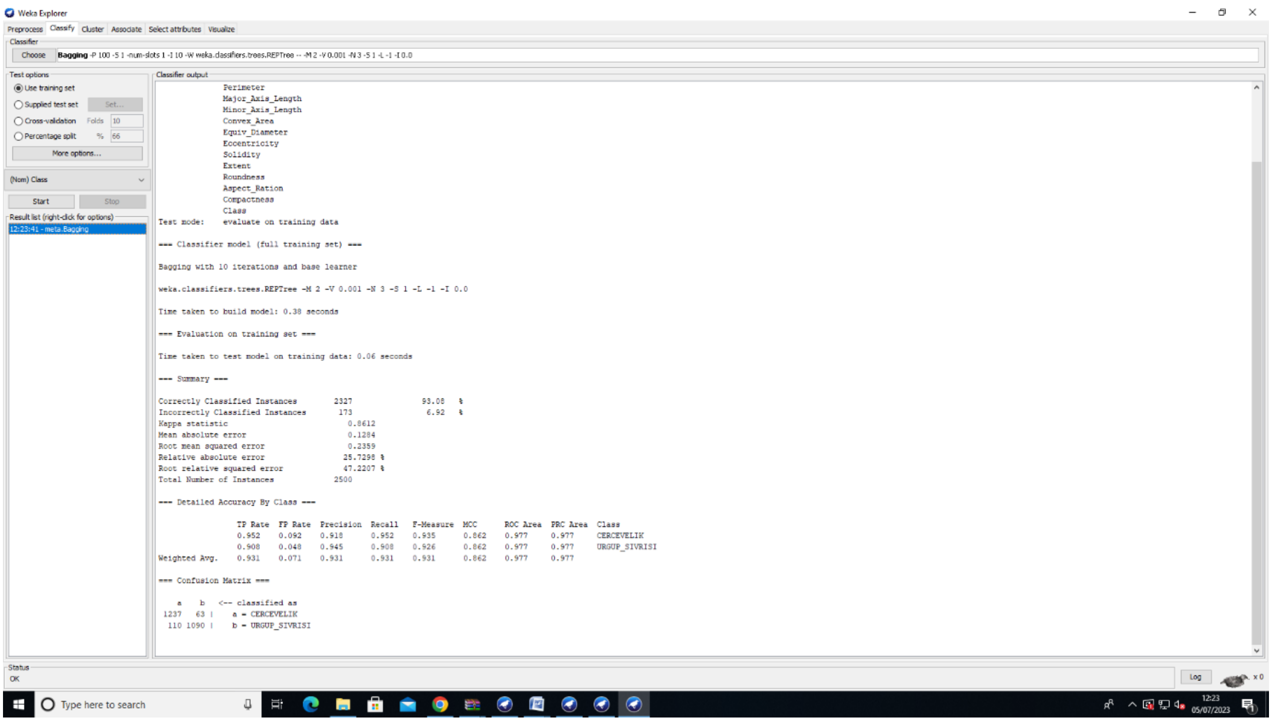

Figure-3: Experimental Results of Bagging

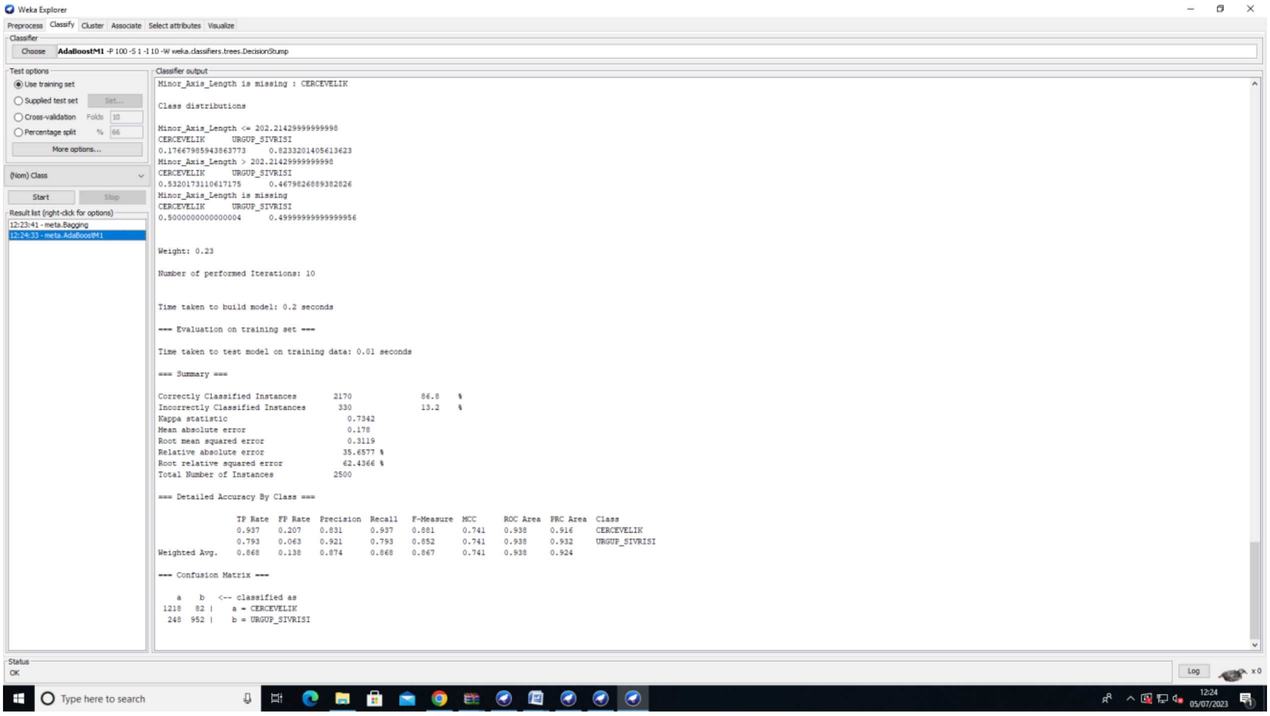

Figure-4: Experimental Results of AdaBoost

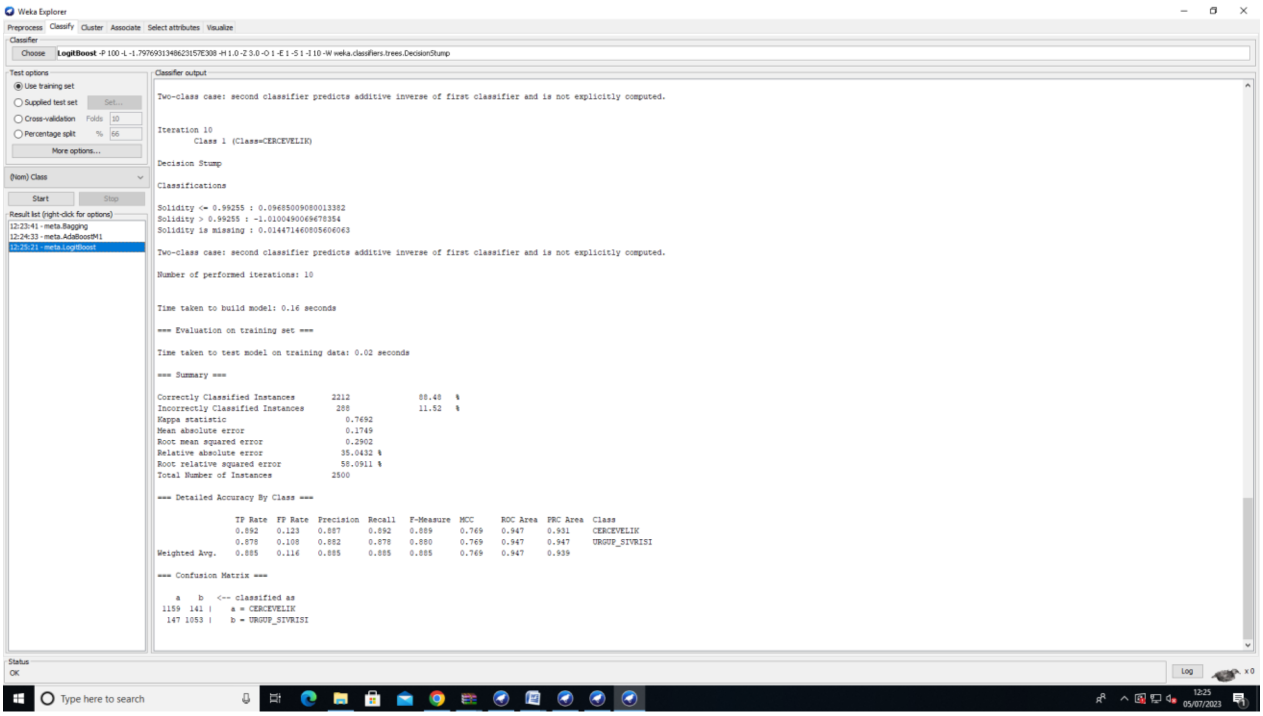

Figure-5: Experimental Results of LogitBoost

V. CONCLUSION

Overall, this study contributes to the field of agricultural research by evaluating and comparing the performance of different ensemble learning algorithms for pumpkin seed prediction. The findings can be utilized by seed producers, farmers, and researchers to enhance seed quality control and optimize pumpkin crop yields.

Further research can explore other ensemble learning algorithms and incorporate additional features to enhance the predictive performance for pumpkin seed quality prediction. Additionally, investigating the interpretability of the ensemble models and identifying the most important features for accurate predictions could provide valuable insights for pumpkin seed quality assessment.

REFERENCES

[1] Ardabili AG, Farhoosh R, Khodaparast MHH (2011). Chemical composition and physicochemical properties of pumpkin seeds (Cucurbita pepo subsp. pepo var. Styriaka) Grown in Iran. Journal of Agricultural Science and Technology 13:1053-1063.

[2] Bwade KE, Aliyu B (2012). Investigations on the effect of moisture content and variety factors on some physical properties of pumpkin seed (Cucurbitaceae spp). International Journal of Engineering, Business and Enterprise Applications (IJEBEA) 3(1):20-24.

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

[4] 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.

[5] 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

[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] 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

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

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