Classification of Sugar-Cane Varieties' Disease Resistance using Bagging and Boosting Machine Learning Algorithms

R. Shalini

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

Abstract— *Sugarcane, a vital agricultural produce, is susceptible to diseases that can severely impact crop quality and yield. Early identification and effective mitigation of sugarcane diseases are crucial for successful crop management. Disease outbreaks can lead to significant financial losses for farmers as they can devastate entire crop fields. To combat this issue, researchers are exploring the application of Artificial Intelligence (AI) techniques, particularly Bagging and Boosting algorithms in Machine Learning (ML), to analyze agricultural data and prevent crop damage caused by various factors, with diseases being a major concern.*

The research is prompted by the rapid evolution of sugarcane disease classes and the lack of disease diagnostic and recognition skills among farmers. By leveraging Bagging and Boosting algorithms, sugarcane farmers can gain valuable insights into disease identification and prediction, enabling them to take timely preventive measures. Integrating AI and ML techniques in sugarcane cultivation can enhance disease management strategies, safeguard crop productivity, and contribute to sustainable agricultural practices.

I. INTRODUCTION

Sugarcane sicknesses are a colossal wellspring of concern and risk for ranchers, since they can monetarily affect sugarcane result and creation, in the event that they are not recognized as soon a possible [5]. In the event that the development of a particular yield declines, it adversely affects the economy. Supportability underway is fundamental, as is asset productivity for seeds, water, soil, and manures. On the off chance that these harvests are annihilated while creating, horticultural produce alongside the expectation of keeping up with nature of these yields will lose a portion of their seriousness. Since the sugarcane infections are inescapable, it is basic to perceive and analyze them. Sugarcane plant sicknesses are an enormous logical subject of study that spotlights on the illness' natural elements. Plant illness location and determination has shown to be intriguing and needs specific thought [7]. Sanitation is influenced by plant infections and the illnesses are especially destructive to limited scope ranchers who depend on sufficient result to get by. Identifying these contaminations at beginning phases will bring about better sugarcane creation, and will help the two ranchers and clients. Early location of sugarcane diseases is used to execute protection estimates to limit extra damage.

II. ENSEMBLE TECHNIQUES

Troupe techniques include making an outfit of base classifiers and consolidating their expectations to pursue the last grouping choice. This segment examines famous outfit techniques, including stowing, supporting, and stacking. Every strategy is made sense of exhaustively, featuring its extraordinary qualities, preparing interaction, and outfit mix methodologies [1].

Troupe characterization offers a few benefits over single classifiers. This part investigates the advantages of troupe procedures, like superior precision, heartiness against exceptions and clamor, and better speculation to inconspicuous information [4][6]. The fundamental standards behind these advantages are examined, giving a more profound comprehension of why gathering classifiers frequently outflank individual classifiers.

III. METHODOLOGY

The primary objective of this paper is to predict the disease resistance levels of sugarcane varieties using machine learning algorithms. By leveraging this data, researchers and agricultural practitioners can develop predictive models to classify sugarcane varieties into different disease resistance categories. This classification can aid farmers in selecting the most suitable varieties for their specific regions and implement targeted disease management strategies, ultimately contributing to improved crop productivity and financial sustainability in sugarcane cultivation.

Troupe classifiers join different base classifiers to improve expectation precision and power. Sacking and Helping are broadly utilized outfit strategies that have shown momentous outcome in different spaces. This segment presents the idea of troupe characterization, features the meaning of Stowing and Supporting calculations, and layouts the targets of this near study.

3.1 Bagging Calculation

Packing (Bootstrap Collecting) is an outfit classifier calculation that creates numerous subsets of the first preparation information through bootstrap examining [4]. This part makes sense of the Packing calculation's preparation interaction, gathering mix procedure, and how it lessens overfitting. The advantages and constraints of Packing are examined, giving a complete comprehension of its key qualities.

3.2 Boosting Calculation

Supporting is another famous group classifier calculation that successively fabricates serious areas of strength for a by underscoring the misclassified examples during preparing. This segment digs into the Supporting calculation's iterative preparation process, the idea of frail students, and the troupe blend system [4][8]. The benefits and restrictions of Helping are featured, revealing insight into its remarkable highlights.

IV. EXPERIMENTAL RESULTS

The investigations have been coordinated by using Python programming tongue. The Python Scikit-learn is a pack for data portrayal, gathering and portrayal. The dataset used in this study contains information related to sugarcane crops and their susceptibility to diseases. The dataset follows a randomized block design with 180 rows and 5 columns, representing a well-structured experimental setup [9]. The data was collected to classify 45 varieties of sugarcane into three categories: resistant, intermediate, and susceptible, based on their disease resistance levels.

The standard dataset is distributed two sets one for preparing (70%) and one more set for testing (30%). The experimental results on the Malignant Melanoma dataset, using various AI algorithms for classification, are shown in the table-1 and also same shown in the figure-1 are as follows:

Table-1
Sugarcane disease prediction results

Algorithm	Accuracy	Precision	Recall
Bagging	91.43	91.5	91.5
Ada Boosting	93.42	93.5	93.5

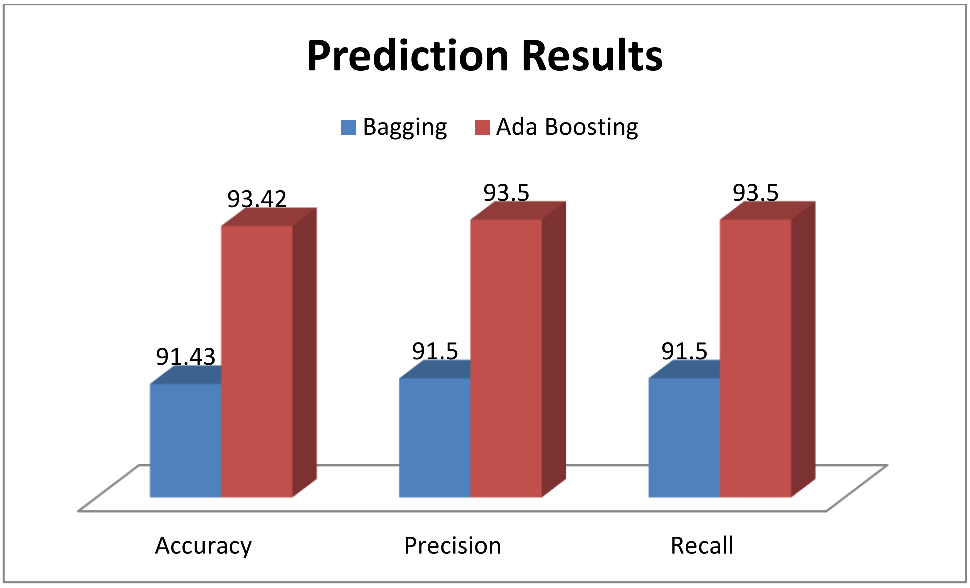

Figure-1: Sugarcane disease prediction results

Figure-1 presents the results of the sugarcane disease prediction using two machine learning algorithms: Bagging and Ada Boosting. The performance metrics evaluated for both algorithms include Accuracy, Precision, and Recall.

4.1 Bagging Results:

- Accuracy: The Bagging algorithm achieved an accuracy of 91.43%. Accuracy represents the proportion of correctly predicted outcomes (both true positives and true negatives) to the total number of samples.
- Precision: The precision of the Bagging model was measured at 91.5%. Precision indicates the ability of the model to correctly identify true positives among all the samples predicted as positive.
- Recall: The recall score for the Bagging model was also 91.5%. Recall, also known as sensitivity or true positive rate, measures the ability of the model to correctly identify positive samples among all the actual positive samples.

4.2 Ada Boosting Results:

- Accuracy: The Ada Boosting algorithm outperformed Bagging with an accuracy of 93.42%. This indicates that the Ada Boosting model had a higher proportion of correct predictions compared to the total number of samples.
- Precision: The precision achieved by the Ada Boosting model was 93.5%. The Ada Boosting model showed a high precision, meaning it accurately identified a significant proportion of true positive cases.
- Recall: The recall score for the Ada Boosting model was 93.5%, indicating its ability to effectively identify positive samples among all the actual positive cases. This high recall suggests the model's capability to capture a large portion of positive instances.

V. DISCUSSION

The experimental results demonstrate the effectiveness of both Bagging and Ada Boosting algorithms in predicting sugarcane disease resistance levels. Both models achieved high accuracy, precision, and recall scores, indicating their capability to accurately classify the sugarcane varieties into resistant, intermediate, and susceptible categories.

The Ada Boosting algorithm showed superior performance compared to Bagging, with higher accuracy, precision, and recall values. The sequential nature of Ada Boosting allows it to emphasize the misclassified instances in each iteration, leading to a more robust and accurate model. This is particularly crucial in the context of sugarcane disease prediction, as misclassification of disease resistance levels can significantly impact disease management strategies and crop yield.

The high precision values indicate that both models minimized the number of false positives, which is vital for avoiding unnecessary disease control measures for non-affected varieties. Additionally, the high recall scores highlight the models' ability to effectively capture positive instances, enabling the identification of disease-resistant varieties that can be prioritized in plant breeding and cultivation practices.

VI. CONCLUSION

In this study, we explored the application of two powerful machine learning algorithms, Bagging and Ada Boosting, for predicting sugarcane disease resistance levels. The experimental results demonstrated the effectiveness of both algorithms in accurately classifying sugarcane varieties into resistant, intermediate, and susceptible categories. Additionally, the Ada Boosting algorithm exhibited superior performance compared to Bagging, indicating its potential as a robust predictive tool for sugarcane disease resistance classification.

The high accuracy, precision, and recall scores obtained from both models highlight their ability to make accurate predictions and minimize false positives and false negatives. Such performance is critical in the context of sugarcane disease management, as misclassifications could lead to improper disease control measures and impact crop productivity.

It is worth noting that while both Bagging and Ada Boosting algorithms performed well in this study, the choice of algorithm depends on specific requirements, such as computational resources and model interpretability. Further investigations on larger and diverse datasets are warranted to confirm the generalizability of these results and to explore other potential machine learning approaches.

REFERENCES

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

[2] G Ravi Kumar, K Venkata Sheshanna and G Anjan Babu, “Sentiment analysis for airline tweets utilizing machine learning techniques”, International Conference on Mobile Computing and Sustainable Informatics, PP:791-799, Publisher:Springer, Cham, 2020

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

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

[5] Li, Y.-R.; Yang, L.-T. Sugarcane agriculture and sugar industry in China. Sugar Tech. 2015, 17, 1–8

[6] N. Michael, "Artificial Intelligence - A Guide to Intelligent Systems",2nd edition, Addison Wesley, 2005.

[7] Pereira, S.C.; Maehara, L.; Machado, C.M.M.; Farinas, C.S. Physical–chemical–morphological characterization of the whole sugarcane lignocellulosic biomass used for 2G ethanol production by spectroscopy and microscopy techniques. Renew. Energy 2016, 87, 607–617

[8] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Reading, MA: Addison-Wesley, 2005.

[9] UCI machine learning repository. http://archive.ics.uci. edu/ml/