International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Data Mining Techniques for Detection and Prediction of Crop Diseases and Pests

Simrandeep Kaur, M.Tech CSE
Dr.Gurpreet Singh, Professor in CSE
Er. Kajal, Assistant Professor in CSE
St. Soldier Institute of Engg. & Technology, Near NIT, Jalandhar

Abstract: Agriculture research is rapidly growing, due to advancement of technologies and upcoming challenges. It has been proven to be leading role in improving the overall growth rate of any country. Especially in India, there is a dire need todo extensive research for better productivity in agriculture. To improve the growth rate of agriculture, researchers of this domain used different data mining techniques to solve agriculture related problems. Data mining approaches such as classification helps to predict the crops diseases, production and loss. It supports farmer while taking right decisions. This research focuses on prediction of loss due to various insect. We analyze the damages by using well-known classifiers such as AdaBoostM1, Proposed Technique AdvC4.5, PART, Ridor and design Ensemble Models of above mentioned classifiers which gave better results as compared to classifiers.

Keywords: AdvC4.5, PART, DM

1. Introduction: Agricultural impact on economy is very high in Asian countries. Almost 65% of the total population of Asian countries was directly connected to agriculture. The current GDP of India in agriculture is 20.19% which is low in respect of available resources and requirement. The services sector is the largest sector of India. Gross Value Added (GVA) at current prices for the services sector is estimated at 96.54 lakh crore INR in 2020-21. The services sector accounts for 53.89% of total India's GVA of 179.15 lakh crore Indian rupees. With GVA of Rs. 46.44 lakh crore, the Industry sector contributes 25.92%. While Agriculture and allied sector share 20.19%. However, government is struggling for improvement, but there is still need for extensive research and finding best ways to utilize the resources. Every country has its own agriculture challenges for production, due to changes in some major resources like soil, weather, water and light. But some common issues also exist like identifying disease, making decision on their condition (spray decision) and so on. From last decade, data mining techniques make a tremendous change in yield production by predicting the future trends and upcoming threats to crops. Data mining techniques provides very fruitful results to overcome the challenges, such as identification of diseases, help to make right decisions, prediction for upcoming challenges and yield production.

In this paper, research scope is limited to analyzing the loss of crops, due to disease or growth of insects. The correct prediction of loss in crops helps farmer to apply some suitable action to overcome the issues and maintain its expected production. To analyze the damages, very well-known data mining techniques will be applied such as AdaBoostM1, Proposed Technique AdvC4.5, PART, Ridor as discussed in Methodology section. We design different ensemble models

based on above mentioned techniques to improve accuracy. However, ensemble models improve the accuracy of weak classifiers, but there is still need to analyze and apply some hybrid approaches of evolutionary algorithms and data mining for better results which is our future research direction discussed in Conclusion and Future Research Direction section. The dataset we used for our experiment is grass grub damage which contains 329 records, 7 attributes.

2. Data Mining Techniques

Data mining functionalities are used to specify the kind of patterns to be found in data mining tasks, which can be classified into two categories predictive mining and descriptive mining. Predictive mining works on the concept of inference to make predictions/forecasts. Descriptive mining describes the general properties of the data. Data mining functionalities include:

- Characterization and Discrimination
- Mining Frequent Patterns, Associations, and Correlations
- Classification and Prediction
- Cluster Analysis

3. Proposed Method

The AdvC4.5 algorithm is a famous algorithm in Data Mining. The AdvC4.5 algorithm acts as a Decision Tree Classifier. AdvC4.5 is a data mining algorithm and it is used to generate a decision tree. The AdvC4.5 algorithm is very helpful to generate a useful decision, that is based on a sample of data. When we generate the decision trees with the help of AdvC4.5 algorithm, then it can be used for classification of the dataset, and that is the main reason due to which AdvC4.5 is also known as a statistical classifier.

4. Flow Chart

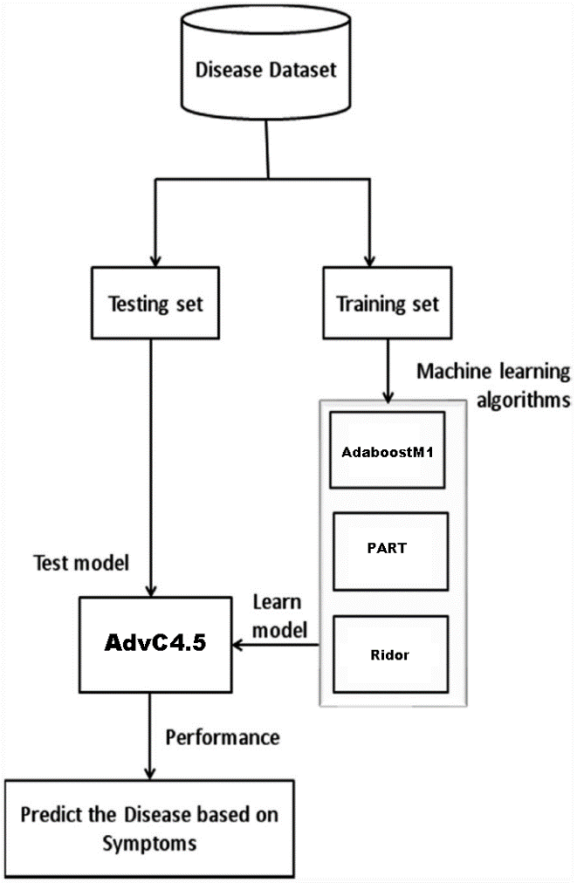

Fig. 1.1: Flow Chart of Proposed Technique Advc4.5

5. Dataset Used

Table 1.1: Dataset Used for prediction of crop diseases

Temp	Crop Category	Crop Name	Pest Category	Pest Name	Disease Category	Disease Name
15	maize	Adlay (Coix lacryma-jobi)	bacterial antagonists of plant diseases	Acidovorax citrulli (antagonist)	Ascomycota - Dothideomycetes	Alternaria (anamorphic genus)
18	maize	Barley (Hordeum vulgare)	bacterial antagonists of plant diseases	Acinetobacter lwoffii (antagonist)	Ascomycota - Dothideomycetes	Alternaria alstroemeriae
25	maize	Chenopodium (crop)	bacterial antagonists	Actinomadura rubra (antagonist)	Ascomycota - Dothideomycetes	Allophoma tropica

			of plant diseases			
33	maize	Digitaria (crop)	bacterial antagonists of plant diseases	Actinoplanes (genus - antagonists)	Ascomycota - Dothideomycetes	Alternaria (anamorphic genus)

6. Results

Table 1.2: Shows the accuracy with parameters CCI, ICI, ER of Algorithms AdaBoostM1, AdvC4.5, PART, Ridor

Algorithms	AdaboostM1	AdvC45	PART	Ridor
CCI	29.48	44.37	33.13	36.17
ICI	70.51	55.62	66.86	63.82
ERROR RATE	99.87	92.96	98.26	128.34

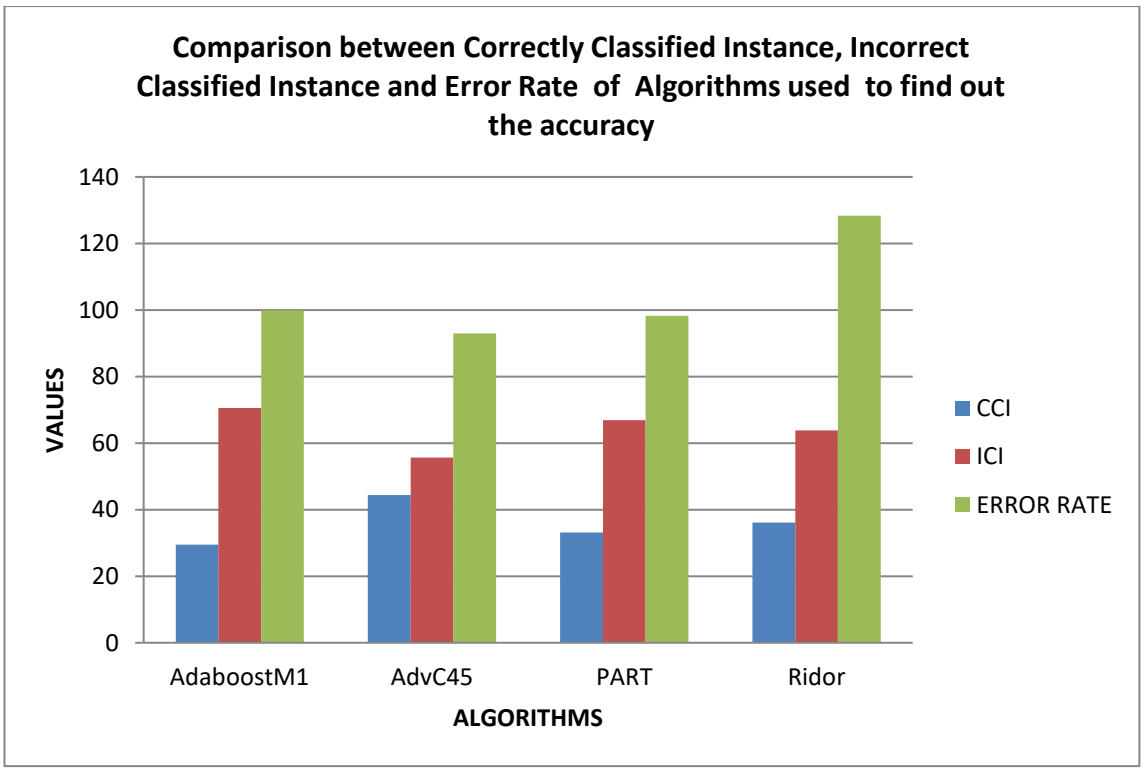

Fig. 1.2: Comparison between Correctly Classified Instance, Incorrect Classified Instance and Error Rate of Algorithms used to find out the accuracy

Table 1.3: Shows the Analysis Between TP RATE, FP RATE, PRECISION, RECALL AND F-MEASURE of AdaBoostM1, AdvC4.5, PART, Ridor

Parameters	AdaboostM1	AdvC45	PART	Ridor
TP RATE	1	1	0.918	0.309
FP RATE	1	0.72	0.841	0.147
PRECISION	0.295	0.367	0.313	0.469
RECALL	1	1	0.918	0.309
F-MEASURE	0.455	0.537	0.467	0.377

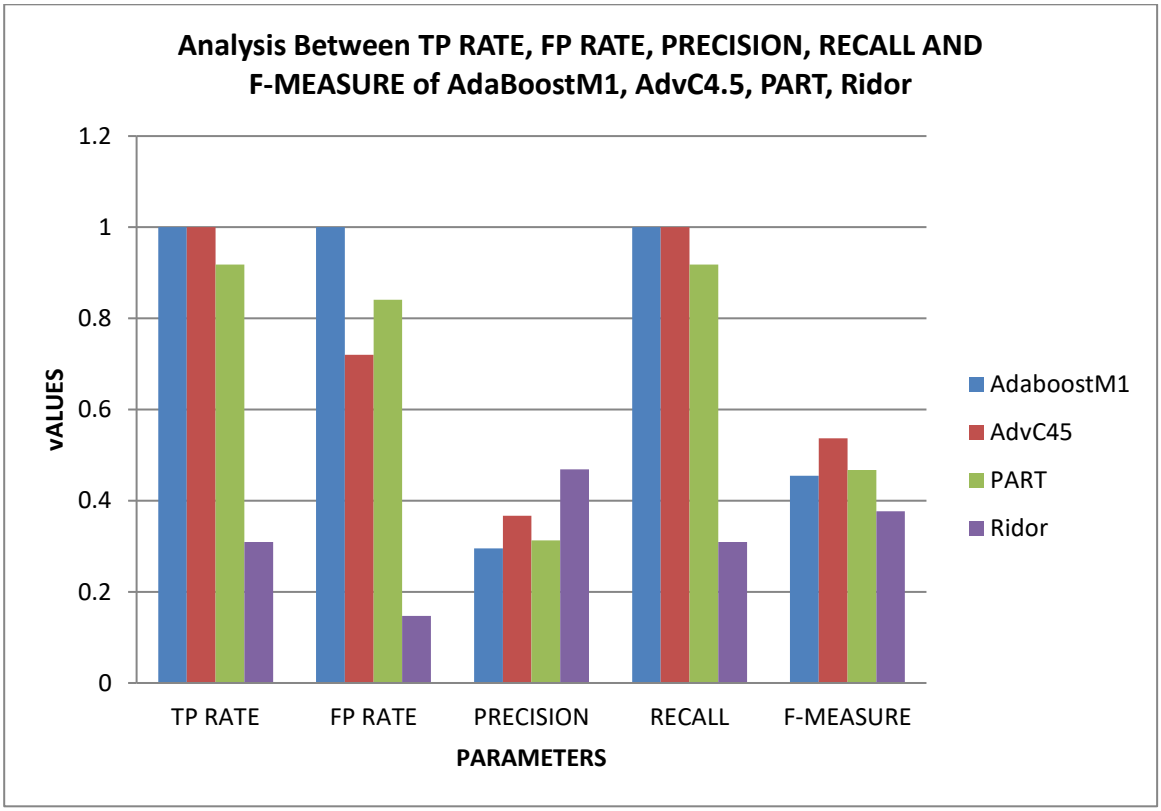

Fig. 1.3: Analysis Between TP RATE, FP RATE, PRECISION, RECALL AND F-MEASURE of AdaBoostM1, AdvC4.5, PART, Ridor

Using the proposed model, results achieved are better, obtained 44.37% accuracy and other algorithms obtained 29.48%, 33.13%, 36.17%.

7. Conclusion

In this research, we did extensive analysis of different data mining classifiers on different feature sets to predict the crop diseases. The classifiers we used are: **AdaBoostM1, AdvC4.5, PART, Ridor**. Advc4.5 performed better than other classifiers for correctly classified instances and less Incorrect classified instances .Results can be further improved by applying some hybrid approaches of data mining and evolutionary algorithms. The evolutionary algorithms help to achieve global optima without getting stuck in local optima. Researchers are trying to solve many real life problems by deep learning. Deep Learning techniques and hybrid approaches (evolutionary & data mining) can solve the crops related problem which is our future research direction.

References

[1] Vanitha CN, Archana N, Sowmiya R, "Agriculture Analysis Using Data Mining And Machine Learning Techniques", 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS).

[2] Suman Avdhesh Yadav, Biswa Mohan Sahoo, Smita Sharma, Lipsa Das, "An Analysis of Data Mining Techniques to Analyze the Effect of Weather on Agriculture", 2020 International Conference on Intelligent Engineering and Management (ICIEM).

[3] Mursal Dawodi. Jawid Ahmad Baktash, Tomohisa Wada, "Data-Mining Opportunities in E-Government: Agriculture Sector of Afghanistan", IEEE, 2019.

[4] R.Aarthi, D.Sivakumar, "An Enhanced Agricultural Data Mining Technique for Dynamic Soil Texture Prediction", Third International Conference on Computing and Network Communications (CoCoNet'19).

[5] Ahsan AbdUllah, Stephen Brobst, Muhammad Umer, "Data Mining Pest, Pesticide and Metrological Data", Fifth International Workshop 011 Artificial Intelligence in Agriculture, Cairo, Egypt, 2004.

[6] Hassina Ait Issad, Rachida Aoudjit, Joel J.P.C. Rodrigues, "comprehensive review of Data Mining techniques in smart agriculture", Elsevier, 2019.

[7] Umair Ayub, Syed Atif Moqurrab, "Predicting Crop Diseases Using Data Mining Approaches: Classification", IEEE, 2018.

[8] Tamil Selvi M , Jaison B , "Adaptive Lemuria: A progressive future crop prediction algorithm using data mining", Elsevier 2021.