



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 5, May 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118

Novel Approach for Crop Yield Prediction Using Machine Learning Approach

Santoshi Malusare¹, Sneha Tirth²

P.G. Student, Department of Computer Engineering, Trinity Engineering College, Pune, India¹

Professor, Department of Computer Engineering, Trinity Engineering College, Pune, India²

ABSTRACT: Agriculture relies heavily on the ability to predict crop yields. When it comes to crop yield, there are a number of factors at play. This research is focused on developing cost effective methods for predicting crop yields using available parameters like irrigation, fertilizer, and temperature. Sequential forward FS, sequential backward elimination FS, correlation-based FS, random forest variable importance, and the variance inflation factor algorithm are among the five Feature Selection (FS) algorithms discussed in this study. In general, machine learning techniques are well-suited to a specific region, so they greatly assist farmers in predicting crop yields. Crop prediction can be improved by using a new FS technique called modified recursive feature elimination (MRFE). With the help of a ranking algorithm, the MRFE technique identifies and prioritizes the most important features in a dataset.

KEYWORDS: Feature Selection methods, Machine Learning, Performance Metrics, Crop prediction

I. INTRODUCTION

The data analysis is the cleansing and modelling process for the purpose of finding useful information and conclusions. It is a process of analysis, extraction and prediction of meaningful data from enormous data to extract patterns. This process is used by companies to useful information about the raw data of their customers. In the field of agriculture this analysis may also be used. Most farmers relied on their long-standing experiences in the field with specific crops to expect higher returns in the following harvest period. Mostly due to inappropriate irrigation or crop selection, or the crop yield sometimes is lower than expected. Farmers insist on the need for an effective mechanism for predicting and improving crop growth and Most agricultural research focuses on biological mechanisms to detect and improve crop growth.

The result of the crop yield depends primarily on parameters like crop variety, seed type and environmental parameters such as sunlight (temperature), soil (ph), water (ph), plenty of plants and moisture. The best crops to have greater harvest rates and net crop yield can be predicted by analysing land and atmosphere in a particular area. The farmers will be helped by that prediction.

For suitable soil-type, temperature, water level, space depth, pH soil, season, fertiliser and months, choose appropriate crops for your farm.

II. RELATED WORK

MRFE technique mostly useful in Agriculture Field. The accurate prediction of crop yield can help governments and authorities to have strategic decision in policy making and also helpful for farmers. The focus on enhancing the productivity without considering the ecological impacts of the input resources has resulted into environmental degradation. For crop prediction, the MRFE approach finds the most appropriate attributes. Eight soil properties (N, P, K, Zn, Cu, Fe, Mn, and EC) and two climatic factors (seasons and rainfall) are selected as essential features in the proposed MRFE. N, P, and K are macronutrients found in soil that aid in crop growth and quality. Soil micronutrients Zn, Cu, Fe, and Mn are important in photosynthesis and respiration. Altmann et al. [15] proposed an improved RF model with the PIMP measure for FS. The PIMP ranking measure and Gini importance were compared to find that the PIMP-RF model significantly outperformed the Gini-RF model. Kursu and Rudnicki [16] described the Boruta FS technique, and the Boruta package provided their algorithm a convenient interface, with the Madalon data set being used for their experimental analysis. Ruß and Kruse [17] proposed a novel FS technique for wheat yield prediction with two regression models, support vector regression (SVR) and the regression tree (Reg tree), for a comparison. Darst et al. [18] compared the RF and RF-RFE in terms of the selection of variables, and concluded that the latter was not likely to scale to high-dimensional data. Hsieh et al. [19] used the RFE algorithm to select key features that impact rice blast disease (RBD). Their work analyzed climatic data collected over five years.

The characteristics comparison of the proposed MRFE technique with existing FS techniques such as sequential forward feature selection (SFFS), Boruta, and RFE. FS technique called the MRFE to overcome the limitations of the RFE. The efficiency of the MRFE is analyzed, following the results of the experiments. After the features are selected, classification algorithms take the lead in the prediction process. In much of the research, a single prediction model (such as the kNN [14], NB [24], DT [25], and SVM [26]), along with an ensemble prediction model (like the RF [27]), and Bagging [28] techniques have been used to classify crop prediction. Each algorithm displays prediction characteristics of its own. However, there is a need to find the classifier that works best with the proposed FS technique for crop prediction. Therefore, this work analyzes the performance of each classifier with the proposed MRFE technique to predict the most suitable crops for specific land areas

III. METHODOLOGY

MRFE technique mostly useful in Agriculture Field. The accurate prediction of crop yield can help governments and authorities to have strategic decision in policy making and also helpful for farmers. The focus on enhancing the productivity without considering the ecological impacts of the input resources has resulted into environmental degradation. For crop prediction, the MRFE approach finds the most appropriate attributes. Eight soil properties (N, P, K, Zn, Cu, Fe, Mn, and EC) and two climatic factors (seasons and rainfall) are selected as essential features in the proposed MRFE. N, P, and K are macronutrients found in soil that aid in crop growth and quality. Soil micronutrients Zn, Cu, Fe, and Mn are important in photosynthesis and respiration. Random forests are a machine learning algorithm. The basic premise of the algorithm is that building a small decision-tree with few features is a computationally process. The algorithm works as follows: for each tree in the forest, we select a bootstrap sample from S where $S(i)$ denotes the i th bootstrap. We then learn a decision-tree using a modified decision-tree learning algorithm. The algorithm is modified as follows: at each node of the tree, instead of examining all possible feature-splits, we randomly select some subset of the features $f \in F$: Where, F is the set of features. The node then splits on the best feature in f rather than F . In practice f is much, much smaller than F . Deciding on which feature to split is oftentimes the most computationally expensive aspect of decision tree learning. By narrowing the set of features, we drastically speed up the learning of the tree.

IV. EXPERIMENTAL RESULTS

This section examines the experimental results of the implemented model. The new decision-making model can be implemented here, and simulation can be carried out. The tests would have higher accuracy, and the consistency would be more dependable. Therefore, the overall success of the approach proposed in this section shall be measured. The outcomes and viability of the suggested solution are calculated and contrasted with the parameters of the calculation.

The experimental result evaluation, we have notation as follows:

TP: True positive (correctly predicted number of instance)

FP: False positive (incorrectly predicted number of instance),

TN: True negative (correctly predicted the number of instances as not required)

FN false negative (incorrectly predicted the number of instances as not required),

On the basis of this parameter, we can calculate four measurements

Accuracy = $\frac{TP+TN}{TP+FP+TN+FN}$

Precision = $\frac{TP}{TP+FP}$

Recall = $\frac{TP}{TP+FN}$

F1-Measure = $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

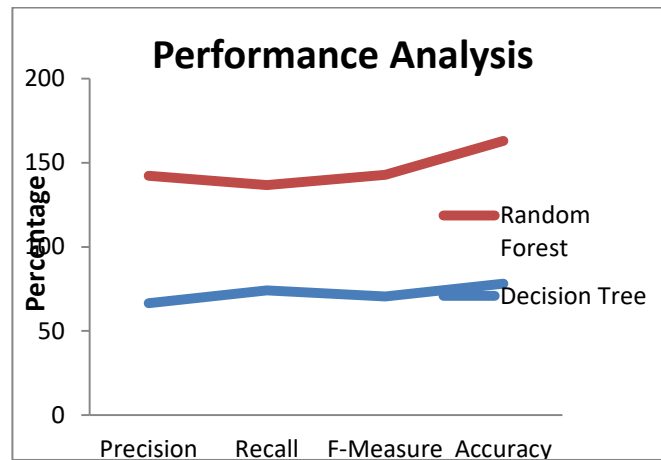


Figure 2. Performance analysis

The AROC curve represents the sensitivity and specificity performance. Figure 4 represents the performance of AROC, which is 0.9677. Here, from Figure 4, the value of AROC is 0.9677, which indicates the classifier distinguishes the class perfectly. From the results obtained, it is clear that the proposed method outperforms other existing methods.

By using the decision optimization model and Discrete Hybrid DBN-VGG Classification, the best crop was selected, depending upon the rank obtained by the crop as shown in Table 1. When compared to others, wheat, rice, and millets have a higher probability of achieving high yield than others.

1) Performance Evaluation of the Modified RFE Varying RF Parameter: To optimize the performance of the proposed MRFE technique, the parameter of the RF classifier is fine-tuned to identify the most important features. The range of the RF parameter, *ntree* and *mtry*, is varied and the performance evaluated to determine the importance of the features chosen.

2) Performance Evaluation of the Modified RFE Varying Ranking Method: The ranking method ranks features in order, from the best to the worst. It eliminates irrelevant features in the process, greatly impacting the FS process.

3) Performance Evaluation of Modified RFE Technique Using Various Classifiers: This evaluation identifies the best classifier to predict a suitable crop for cultivation

V. CONCLUSION

Predicting a suitable crop for cultivation is critical to agriculture. In this work, the MRFE, a novel approach, has been proposed for selecting salient features using a permutation crop data set and a ranking method to identify the most suitable crop for a particular region. Experiments were conducted to evaluate the efficiency of the proposed MRFE technique using the kNN, NB, DT, SVM, RF, and bagging classification techniques to predict the most suitable crops for cultivation. Soil and environmental factors were considered for an analysis of the crop prediction process. The results indicate that the MRFE with the bagging technique classifier gives better crop prediction ACC than the MRFE with other classifiers. The performance of the MRFE technique for the crop data set was assessed and compared with existing techniques like the SFFS, Boruta, and RFE. Furthermore, the suitability of the proposed MRFE technique was evaluated using three benchmark data sets. The results show that the proposed MRFE technique outperforms the others. Nevertheless, the MRFE technique needs performance-wise improvements before it can be used in large feature data sets.

REFERENCES

1. Baidar, T. Rice Crop Classification and Yield Estimation Using Multi-Temporal Sentinel-2 Data: A Case Study of Terai Districts of Nepal. 2020. Available online: <https://www.semanticscholar.org/paper/Rice-crop-classification-and-yield-estimation-using-Baidar/bf0c1e646caf560db7d19567c8ee10fec0ec3372> (accessed on 21 February 2020).
2. Nielsen, D.C.; Vigil, M.F.; Benjamin, J.G. Evaluating decision rules for dryland rotation crop selection. *Field Crops Res.* **2011**, *120*, 254–261. [CrossRef]

3. Lavanya, K.; Jain, A.V.; Jain, H.V. Optimization and Decision-Making in Relation to Rainfall for Crop Management Techniques. In Information Systems Design and Intelligent Applications; Springer: Berlin/Heidelberg, Germany, 2019; pp. 255–266.
4. Tamsekar, P.; Deshmukh, N.; Bhalchandra, P.; Kulkarni, G.; Hambarde, K.; Husen, S. Comparative Analysis of Supervised Machine Learning Algorithms for GIS-Based Crop Selection Prediction Model. In Computing and Network Sustainability; Springer: Berlin/Heidelberg, Germany, 2019; pp. 309–314.
5. Lagos-Ortiz, K.; Medina-Moreira, J.; Alarcón-Salvatierra, A.; Morán, M.F.; del Cioppo-Morstadt, J.; Valencia-García, R. Decision Support System for the Control and Monitoring of Crops. In Proceedings of the 2nd International Conference on ICTs in Agronomy and Environment, Guayaquil, Ecuador, 22–25 January 2019; pp. 20–28.
6. Das, S.; Sengupta, S. Feature Extraction and Disease Prediction from Paddy Crops Using Data Mining Techniques. In Computational Intelligence in Pattern Recognition; Springer: Berlin/Heidelberg, Germany, 2020; pp. 155–163.
7. Fegade, T.K.; Pawar, B. Crop Prediction Using Artificial Neural Network and Support Vector Machine. In Data Management, Analytics and Innovation; Springer: Berlin/Heidelberg, Germany, 2020; pp. 311–324.
8. Hammer, R.G.; Sentelhas, P.C.; Mariano, J.C.Q. Sugarcane Yield Prediction through Data Mining and Crop Simulation Models. *Sugar Tech.* **2020**, *22*, 216–225. [[CrossRef](#)]
9. Chaudhari, R.; Chaudhari, S.; Shaikh, A.; Chiloba, R.; Khadtare, T. Soil Fertility Prediction Using Data Mining Techniques. *Mukt. Shabd. J.* **2020**, *9*.
10. Champaneri, M.; Chachpara, D.; Chandvidkar, C.; Rathod, M. Crop Yield Prediction Using Machine Learning. Available online: https://www.researchgate.net/publication/340594772_CROP_YIELD_PREDICTION_USING_MACHINE_LEARNING (accessed on 4 April 2020).
11. Feng, K.; Tian, J. Forecasting reference evapotranspiration using data mining and limited climatic data. *Eur. J. Remote Sens.* **2021**, *54*, 363–371. [[CrossRef](#)]
12. Bhojani, S.H.; Bhatt, N.J.N.C. Wheat crop yield prediction using new activation functions in neural network. *Neural Comput. Appl.* **2020**, *32*, 13941–13951. [[CrossRef](#)]
13. Garg, H. Neutrality operations-based Pythagorean fuzzy aggregation operators and its applications to multiple attribute group decision-making process. *J. Ambient Intell. Humaniz. Comput.* **2020**, *11*, 3021–3041. [[CrossRef](#)]
14. Samorn, N.; Yotha, N.; Srisilp, P.; Mukdasai, K. LMI-Based Results on Robust Exponential Passivity of Uncertain Neutral-Type Neural Networks with Mixed Interval Time-Varying Delays via the Reciprocally Convex Combination Technique. *Computation* **2021**, *9*, 70. [[CrossRef](#)]
15. Alharbi, A.; Equbal, K.; Ahmad, S.; Rahman, H.U.; Alyami, H. Human Gait Analysis and Prediction Using the Levenberg-Marquardt Method. *J. Healthc. Eng.* **2021**, *18*, 2021.
16. Layona, R.; Suharjito, S.; Tunardi, Y.; Tanoto, D. Optimization of land suitability for food crops using neural network and swarm optimization algorithm. *Int. Rev. Comput. Softw.* **2016**, *11*, 1. [[CrossRef](#)]
17. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [[CrossRef](#)]



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details