Journal of Advanced Zoology
ISSN: 0253-7214
Volume 45 Issue S-4 Year 2024 Page 113-117

Random Forest Classifier For Crop Prediction Based On Soil Data

Kranti G. Sapkal[1*], Avinash B. Kadam[2]

[1 *] *Research Scholar, Sant Gadage Baba Amravati University, Amravati, India.*
Email: krantisapkal2006@gmail.com
[2] *Assistant Professor, Sant Gadage Baba Amravati University, Amravati, India.*
Email: avinashkadam28@gmail.com

***Correspondence Author:** Kranti G. Sapkal*
**Research Scholar, Sant Gadage Baba Amravati University, Amravati, India.*
Email: krantisapkal2006@gmail.com

Abstract

Agricultural development is crucial to feed the growing population. Most farmers tend to cultivate the crops which will give the more economical benefits besides checking the suitability of the crop according to the soil conditions. Use of technology in the agricultural sector leads the sustainable improvements in the agricultural production. Machine learning approach to suggest the suitable crop based on the soil parameters can help the farmers to cultivate the crops accordingly and can produce more yield. In this paper Random Forest Classifier is used to train the Machine Learning model on soil dataset using Python. Model performance is evaluated using confusion matrix and classification report having precision, recall and F1 score. Model accuracy achieved is 99% without parameter tuning.

Keywords: Agriculture, Machine Learning, Random Forest, Confusion Matrix.

1. INTRODUCTION

Machine learning techniques are used in different sectors, like product recommendation, entertainment, education, agriculture and so on [1]. From past few years, agriculture has been using IoT and machine learning techniques. Machine learning is a very useful technology for recommending the crop to cultivate during the season based on various parameters. The ultimate view point of ML is to automate the data analysis process with the help of algorithms that are enabled with continuous learning skill [2]. Hence ML refers to the set of techniques meant to deal with huge data, collected from IoT sensors in the most intelligent way in order to derive actionable insights [3]. Machine Learning algorithms are mainly classified as supervised, unsupervised and reinforcement learning algorithms [4].Use of specific ML algorithm is totally depends on the type of problem and data available. The study aims to implement the Random Forest classifier in python on the dataset containing 22 varieties of crops. The model's performance is calculated under two criterions- Entropy and Gini Index. The aim of this is model is to suggest crop for cultivation for the particular soil type and climatic conditions more accurately.

2. LITERATURE SURVEY

Champaneri, Mayank implemented a model for predicting the crop yield in advance of its harvest would help the policy makers and farmers for taking appropriate measures for mar-keting and storage. Random forest is the most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks [5]. Jeevan Nagendra Kumar, Y. et al. implemented a system to predict crop production from the collec-tion of past data. Using data mining techniques crop yield is predicted. Random Forest algo-rithm is used for predicting the best crop yield as output. In agriculture field, the crop yield prediction is mostly appropriate[6]. B S, Anisha, and Ramakanth P. Kumar implemented smart agriculture for maximizing agricultural farm water supplies, crop prediction, and wild animal prevention. Depending on the level of soil moisture, the system can be used to turn the water sprinkler on / off, thereby making the process easier to use[7]. Kalimuthu, M., P. Vaishnavi, and M. Kishore used Naive Bayes, a supervised learning algorithm to predict the crop at high accuracy. Using seed data of crop is used with the appropriate parameters like temperature, humidity and moisture content, which helps the crops to achieve a successful growth. In addition the authors developed the software, a mobile application for Android is being developed[8]. Yamaç et al. evaluated the performance of deep learning (DL), artificial neural network (ANN) and k-nearest neighbour (kNN) models to estimate field capacity (FC) and permanent wilting point (PWP) using four combinations of soil data. The DL, ANN and KNN models are compared with the previous published pedotransfer functions (PTF)[9]. Kavita, and Pratistha Mathur presented research shows several existing models that consider elements such as temperature, weather condition, performing models for the effective crop yield prediction. In the experimental study they showed the combination of ML with the agricultural domain field for improving the advancement in crop prediction. Most of the existing models utilized neural networks, random forests, KNN regression tech-niques for CYP and a variety of ML techniques were also used for best prediction[10]. The research work done by Pant J et.al. shows the different machine algorithms are used to pre-dict crop yield in India. Researchers have used the data set for making prediction for four primary crops such as potatoes, rice, wheat and maize. The decision tree Regressor achieves highest accuracy to predict crop yield[11]. M. Keerthana et al. have implemented a system for crop yield prediction from formerly collected data. This has been settled with usage of some of the machine learning techniques. In this study Ensemble of Decision Tree Regressor with AdaBoost Regressor is used to predict the outcome with increased accuracy rate[12]. D. J. Reddy and M. R. Kumar explored various ML techniques utilized in the field of crop yield estimation and provided a detailed analysis in terms of accuracy using the techniques. The research shows several existing models that consider elements such as temperature, weather condition, performing models for the effective crop yield prediction[13]. Paper presented by S. Vaishnavi et al. depicts many Machine Learning techniques have been used to analyse the agriculture parameters. Proper prediction of crops can be informed to ag-riculturists in time basis[14].

3. METHODOLOGY

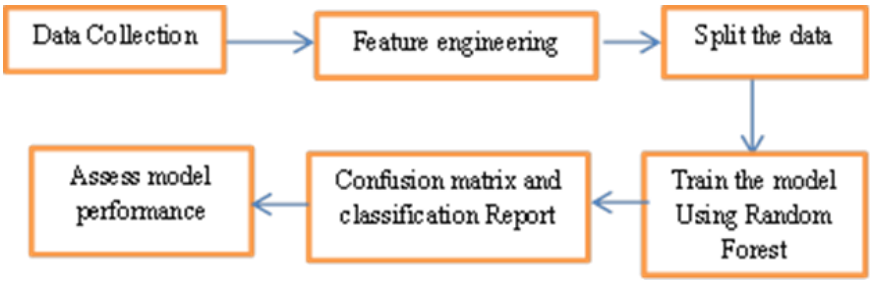

Figure 1. Workflow of the proposed model

3.1. Implementation of Machine Learning Algorithm on Dataset

Random Forest Classifier

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classifications and Regression problems [15]. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression [16]. Random forests are an ensemble method, meaning they combine predictions from other models. Each of the smaller models in the random forest ensemble is a decision tree. Random forest contains multiple decision trees and compute the average to improve the accuracy [17].The

final out-put depends on the maximum votes of predictions from each tree, instead of relying on one decision tree[18] One of the reasons for its popularity as a machine learning, approach is that it can handle the issue of overfitting and accuracy can be increased by using more trees.[19].
Figure2 depicts the feature importance of the seven features. Less important feature can be omitted to improve the accuracy of the model.

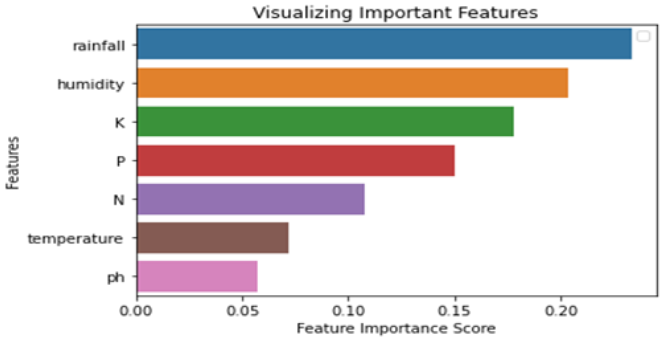

Figure2. Feature importance for Random Forest classifier

3.2. Confusion Matrix and Classification Report

Confusion Matrix gives the matrix of frequency of true negatives, false negatives, true positives and false positive [20] and calculated using the actual labels of test datasets and predicted values as shown in Figure3.

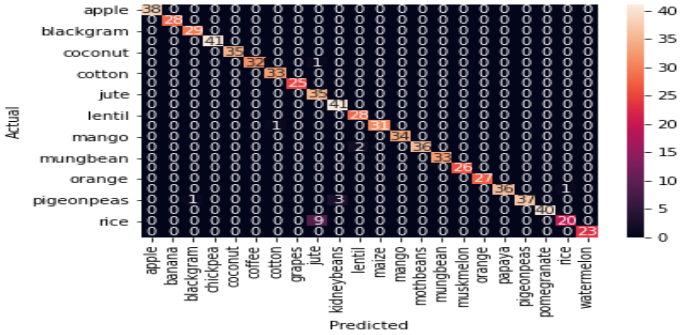

Figure 3. Confusion Matrix

Classification Report is a metric used for evaluating the performance of a classification algorithm's predictions. It gives three things: Precision, Recall and f1-score of the model.
Precision refers to a classifier's ability to identify the number of positive predictions which are relatively correct. It is calculated as the ratio of true positives to the sum of true and false positives for each class [21].

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

Recall is the capability of a classifier to discover all positive cases from the confusion matrix. It is calculated as the ratio of true positives to the sum of true positives and false negatives for each class [21].

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

F1 score is a weighted harmonic mean of precision and recall, with 0.0 being the worst and 1.0 being the best. Since precision and recall are used in the computation, F1 scores are often lower than accuracy measurements [21].

$$\text{F1 Score} = \frac{2*PR}{(P+R)}$$

Accuracy is the number of correct predictions divided by the total number of predictions [21][22][23].

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

Tabel1. Classification report of the model.

Crop	precision	recall	f1-score	support
apple	1.00	1.00	1.00	38
banana	1.00	1.00	1.00	28
blackgram	1.00	1.00	1.00	29
chickpea	1.00	1.00	1.00	41
coconut	1.00	1.00	1.00	35
coffee	1.00	1.00	1.00	33
cotton	1.00	1.00	1.00	33
grapes	1.00	1.00	1.00	25
jute	0.85	1.00	0.92	35
kidneybeans	1.00	1.00	1.00	41
lentil	1.00	1.00	1.00	28
maize	1.00	1.00	1.00	32
mango	1.00	1.00	1.00	34
mothbeans	1.00	1.00	1.00	38
mungbean	1.00	1.00	1.00	33
muskmelon	1.00	1.00	1.00	26
orange	1.00	1.00	1.00	27
papaya	1.00	1.00	1.00	37
pigeonpeas	1.00	1.00	1.00	41
pomegranate	1.00	1.00	1.00	40
rice	1.00	0.79	0.88	29
watermelon	1.00	1.00	1.00	23
accuracy				
macro avg	0.99	0.99	0.99	726
weighted avg	0.99	0.99	0.99	726

4. CONCLUSION

The study carried out for suggesting the suitable crop for the particular land that can help farmers to grow crops more efficiently by using Random Forest Classifier. Soil dataset from Kaggle repository is used to develop the model. Model gives the maximum accuracy of 99 % without tuning any parameters. Confusion matrix and classification report of Precision, Recall, F1 score is used to evaluate the performance of model. In the future, new data from the fields can be collected to get soil content and incorporate other machine learning algorithms to classify more varieties of crops.

REFERENCES

1. Mekonnen, Yemeserach et al. 2020. "Review—Machine Learning Techniques in Wireless Sensor Network Based Precision Agriculture." Journal of The Electrochemical Society.
2. Van Klompenburg,Thomas, Ayalew Kassahun, and Cagatay Catal. 2020. "Crop Yield Prediction Using Machine Learning: A Systematic Literature Review." Computers and Electronics in Agriculture 177: 105709.
3. Elijah, Olakunle et al. 2018. "An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges." IEEE Internet of Things Journal 5(5).
4. Nishant, Potnuru Sai, Pinapa Sai Venkat, Bollu Lakshmi Avinash, and B. Jabber. 2020. "Crop Yield Prediction Based on Indian Agriculture Using Machine Learning." 2020 International Conference for Emerging Technology, INCET 2020: 5–8.
5. Champaneri, Mayank, and Information Technology. 2020. "Crop Yield Prediction Using Regression Model." International Journal of Innovative Technology and Exploring Engineering 9(10): 269–73.

6. Jeevan Nagendra Kumar, Y. et al. 2020. “Supervised Machine Learning Approach for Crop Yield Prediction in Agriculture Sector.” Proceedings of the 5th International Conference on Communication and Electronics Systems, ICCES 2020 (Icces): 736–41.
7. B S, Anisha, and Ramakanth P. Kumar. 2020. “Real-Time Monitoring of Agricultural Land with Crop Prediction and Animal Intrusion Prevention Using Internet of Things and Machine Learning at Edge.” Proceedings of CONECCT 2020 - 6th IEEE International Conference on Electronics, Computing and Communication Technologies.
8. Kalimuthu, M., P. Vaishnavi, and M. Kishore. 2020. “Crop Prediction Using Machine Learning.” Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020 (Icssit): 926–32.
9. Yamaç, Sevim Seda, Cevdet Şeker, and Hamza Negiş. 2020. “Evaluation of Machine Learning Methods to Predict Soil Moisture Constants with Different Combinations of Soil Input Data for Calcareous Soils in a Semi Arid Area.” Agricultural Water Management 234(December 2019).
10. Kavita, and Pratistha Mathur. 2021. “Satellite-Based Crop Yield Prediction Using Machine Learning Algorithm.” 2021 Asian Conference on Innovation in Technology, ASIANCON 2021 (Iciccs): 1466–70.
11. Pant, J., Pant, R. P., Kumar Singh, M., Pratap Singh, D., & Pant, H. (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. Materials Today: Proceedings, 46, 10922–10926. https://doi.org/10.1016/J.MATPR.2021.01.948
12. M. Keerthana, K. J. M. Meghana, S. Pravallika and M. Kavitha, "An Ensemble Algorithm for Crop Yield Prediction," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 963-970, doi: 10.1109/ICICV50876.2021.9388479.
13. D. J. Reddy and M. R. Kumar, "Crop Yield Prediction using Machine Learning Algorithm," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1466-1470, doi: 10.1109/ICICCS51141.2021.9432236.
14. S. Vaishnavi., M. Shobana., R. Sabitha. and S. Karthik., "Agricultural Crop Recommendations based on Productivity and Season," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 883-886, doi: 10.1109/ICACCS51430.2021.9441736.
15. Jeevan Nagendra Kumar, Y. et al. 2020. “Supervised Machine Learning Approach for Crop Yield Prediction in Agriculture Sector.” Proceedings of the 5th International Conference on Communication and Electronics Systems, ICCES 2020 (Icces): 736–41.
16. Kumar, V., Kumar, R., Kumar, S., & Jorvekar, P. P. (2021). Agriculture Soil Analysis for Suitable Crop Prediction. PP, Agriculture Soil Analysis for Suitable Crop Prediction (May 24, 2021).
17. Elavarasan, D., & Durairaj Vincent, P. M. (2020). Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications. IEEE Access, 8, 86886–86901. https://doi.org/10.1109/ACCESS.2020.2992480
18. Pande, Shilpa Mangesh, et al. "Crop recommender system using machine learning approach." 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2021.
19. Bhat, Showkat Ahmad, and Nen-Fu Huang. "Big data and ai revolution in precision agriculture: Survey and challenges." IEEE Access 9 (2021): 110209-110222.
20. Sharma, Robin. "Artificial Intelligence in Agriculture: A Review." 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2021.
21. Rao, M. S., Singh, A., Reddy, N. V. S., & Acharya, D. U. (2022). Crop prediction using machine learning. Journal of Physics: Conference Series, 2161(1). https://doi.org/10.1088/1742-6596/2161/1/012033
22. Suresh, N., Ramesh, N. V. K., Inthiyaz, S., Priya, P. P., Nagasowmika, K., Kumar, K. V. N. H., Shaik, M., & Reddy, B. N. K. (2021). Crop Yield Prediction Using Random Forest Algorithm. 2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021, 279–282. https://doi.org/10.1109/ICACCS51430.2021.9441871
23. Geetha, V., Punitha, A., Abarna, M., Akshaya, M., Illakiya, S., & Janani, A. P. (2020, July 3). An Effective Crop Prediction Using Random Forest Algorithm. 2020 International Conference on System, Computation, Automation and Networking, ICSCAN 2020. https://doi.org/10.1109/ICSCAN49426.2020.9262311