

Agricultural Assistance Using Machine Learning Techniques

Rekha Sundari Mantripragada*

Department of Information Technology, Anil Neerukonda Institute of Technology and Sciences (ANITS), Bheemunipatnam, Visakhapatnam, Andhra Pradesh

Abstract

Using machine learning techniques in smart farming is gaining momentum worldwide. The main goal is to achieve better production by predicting right crop considering present conditions of weather and soil. The climatic changes that are being uncertain results in reduced yield when the farmers follow traditional way of growing crops. The features of soil and conditions of the weather change time to time and this criteria when concentrated leads to precision farming. The study implements machine learning techniques to predict the right crop for cultivation, expecting better yield, taking into account the changes in the weather and soil every time, as and when the farmer aims for growing a fresh crop.

Keywords: Crop Prediction • Millets • Random forest • feature selection

Introduction

Agriculture is the backbone of Indian economy, but unfortunately also the field that faces problems of low income, low technological development, and less encouragement to grow wide variety of crops. Farmers are only habituated to grow crops following the ancestral history than analyzing situations from other side.

New generations had brought abundant technological revolution and proved that nothing is impossible with the help of technology. With this technology growth and development in social network media the thought process of people turned towards building a new and healthy environment with good eating and drinking habits. The great shift in peoples' living and working style triggered a positive change in agricultural field too. Demand to various types of cereals other than rice and wheat has been the part of the market. People started placing various types of as the part of their menu replacing regular food such as rice and wheat. As people are not forced into a physically strained work culture preference to low carbohydrate and rich nutritional food became the part of their daily food culture.

Millets are 3 to 5 times nutritionally superior to rice and wheat in terms of proteins, minerals and vitamins. This is only one example to show that the time for producers to concentrate on the lifestyle of consumers and their upcoming interests has raised and this way of analysing will give raise to many economically viable options for growing crops. Nowadays consumers find many items as a solution to their lifestyle disorders how tasteless and costly there are it's the smartness of the farmers to realize the need of the hour. There are many crops that grow in poor soil conditions with less water availability and crops that withstand high temperatures also exist.

Pandhe A, et al. [1] used random forest algorithm to train the models so as to attain accurate prediction and considered 5 climatic parameters precipitation, Temperature, cloud cover, vapour pressure and wet day frequency for training the model. Cross validation a resampling procedure used to evaluate machine learning models the accuracy of model was stated as 87% but the precision recall confusion matrix were nowhere discussed. Priya S, et al. [2] predicted

***Address for Correspondence:** Rekha Sundari Mantripragada, Department of Information Technology, Anil Neerukonda Institute of Technology and Sciences (ANITS), Bheemunipatnam, Visakhapatnam, Andhra Pradesh, E- mail: mr.sundari@gmail.com

Copyright: © 2022 Mantripragada RS. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Date of Submission: 02 June, 2022, Manuscript No. JCSB-22-68534; **Editor Assigned:** 04 June, 2022, PreQC No. P-68534; **Reviewed:** 18 June, 2022, QC No. Q-68534; **Revised:** 23 June, 2022, Manuscript No. R-68534; **Published:** 30 June, 2022, DOI: 10.37421/0974-7230.2022.15.417

the occurrence of the pest in cotton crops based on weather factors using various clustering algorithms. The work analyzed how the changes in weather change the infection of the pest from higher to lower level. Ozdarici OK S, et al. [3] in the study explained the performance of RF and maximum likelihood classification method to crop classification through pixel based and parcel based approaches by considering thematic maps. The work concluded that RF method with parcel based approach is a reliable method to generate crop maps with good accuracy for agricultural lands.

Malathi B, et al. [4] in the study predicted how the role of various components like growth rate of area, production and yield effect the growth of millets. The study concluded that yield plays a major role in the growth of millets than area and integration. The work also summarized the need of encouragement for the growth of millets due to their nutrient superiority and quality. Nevavuori P, et al. [5] in their study on data sets with images in growth phase, with numerical parameters like thermal time and sowing date applied CNN a deep learning methodology to construct a model for crop yield prediction. Gopal M, et al. [6] in their work proposed a multiple linear regression - ANN model for crop yield prediction of paddy crop. Results show that the hybrid models out performance the commercial models.

Dataset and Methodology

Dataset

The datasets comprises the soil specific attributes and the production quantity of the crop that was produced by crop previously yielded in that soil. The attributes that were taken into consideration were:

- Rainfall
- Soil Temperature
- pH value of the soil
- Soil moisture
- Crop data
- Crop Production

The parameters above play a major role to extract the quality of the soil

Soil is a complex, living, changing and dynamic component of the ecosystem. It is an important variable that can be degraded or can be wisely managed. Learning Soil moisture variations help us to figure out and predict changes in surface temperature, precipitation, drought, flood, and the impacts of future climate changes. This is an important variable in the climate system. Intelligence about the soil moisture content and how it changes is a key for the design and management of agricultural system.

Soil Temperature also assists in the prediction of plant growth. It affects the plant growth indirectly by influencing the water and nutrient uptake there by affecting root growth. Low temperatures lower the traffic from the root to the shoot and vice versa. Each plant reacts contrastingly to soil temperature, contingent upon its scope of resistance of soil temperatures and its temperature ideal. The objective is to keep up soil temperatures as near to the yield's ideal temperature as could be expected under the circumstances, and to keep varieties in soil temperature inside a harvest's scope of resilience.

Soil pH is the major indicator of soil health. It perturbs crop yields, crop suitability, and plant nutrient availability and is the combined measure of soil acidity and alkalinity. Adding proper amount of nitrogen fertilizer, liming and by following best cropping practices a good pH value that improvise soil health can be attained. Soils with pH more no than 7.0 are viewed as basic soils. Micronutrient insufficiencies, for example, iron inadequacy, are regular in these dirt. Yields developed in soils with pH level lower than 5.5 may show poisonous quality side effects of metals (for example iron, manganese) and lacks of different supplements, for example, magnesium.

Liming the dirt is generally prescribed in such soils. The perfect soil pH go for most harvests is somewhere in the range of 5.8 and 6.5, a range where most supplements are accessible for the yields to take them in. Increased intensities of rainfall events and increased rainfall totals would increase leaching rates in well-drained soils with high infiltration rates and would cause temporary flooding or water-saturation, resulting in reduced organic matter decomposition in many soils or depressional sites. This may affect a significant proportion of especially the better soils that would also give rise to greater amounts and frequency of runoff on soils in sloping terrain, with sedimentation down slope and worse downstream.

Step by step procedure for crop recommendation system is as shown below:

The Figure 1 provides a brief understanding of the crop recommendation system. Many recommender systems have been proposed in the literature, but the recommendation system in Figure 1 suggests the farmer the best crop to be yielded based on the history of the data that is existing in the database [7,8]. The pattern that matches to the soil pattern of the new farmer is found in the

database and the best crop among them is suggested to the farmer. The below system provides many advantages compared to other recommender systems by using data analyses techniques like classification and prediction [9]. First the new soil pattern is classified to the existing soil patterns and then among them the best is recommended based on the crop yield and profit gained and is mainly:

The parameters of the soil: The parameter of the soil where the crop is to be predicted are estimated by taking the samples to the lab and is supplied as test data to the classifier.

Collecting dataset: The dataset logs contain historical data about the different crops yielded under various soil parameters, rainfall concentration and their production rate.

Preprocessing: The Dataset collected has some noisy data, so the data is first preprocessed. Preprocessing includes, scaling of values, removing noisy data and outliers and encoding of categorical data. As it is easy to spot anomalies in a graph than in numbers the quality of data is identified by making basic plots of all the considered features. Figure 2 indicates the anomalies in rainfall, Figure 3 anomalies in production; Figure 4 indicates anomalies in pH and Figure 5 anomalies in temperature. To detect the outliers we used the interquartile range method and removed the outliers and extreme values from the dataset. To remove noisy values we used numeric cleaner that identifies values that are too larger or too closer to a certain value and replaces them with the default value.

Classifying new soil pattern to the existing based on random forest algorithm: Random Forest (RF) algorithm is a supervised learning algorithm that is best suitable for classification and prediction. The RF algorithm looks similar to decision tree that uses a tree like graph to depict the possible ways of arriving at a solution, but the working of the RF algorithm is entirely different from the decision trees. The RF algorithm instead of using all the data available in the training data set as decision trees randomly selects only some data points and constructs multiple uncorrelated trees from the data. The DT algorithm selects the root node and child nodes through a process of calculating parameters like information gain and entropy but the Rf algorithm does all the selections randomly. The most advantage of RF algorithm is it

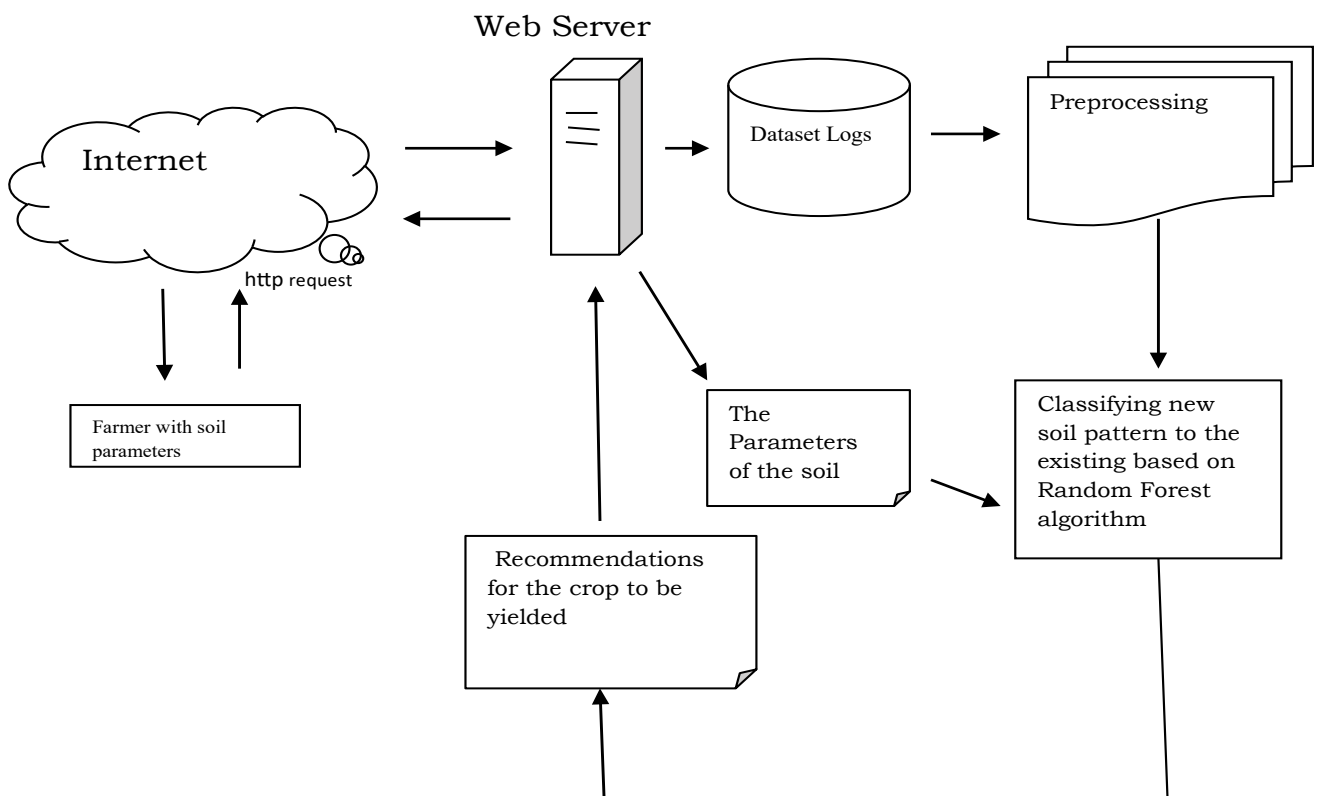


Figure 1. Crop recommendation system.

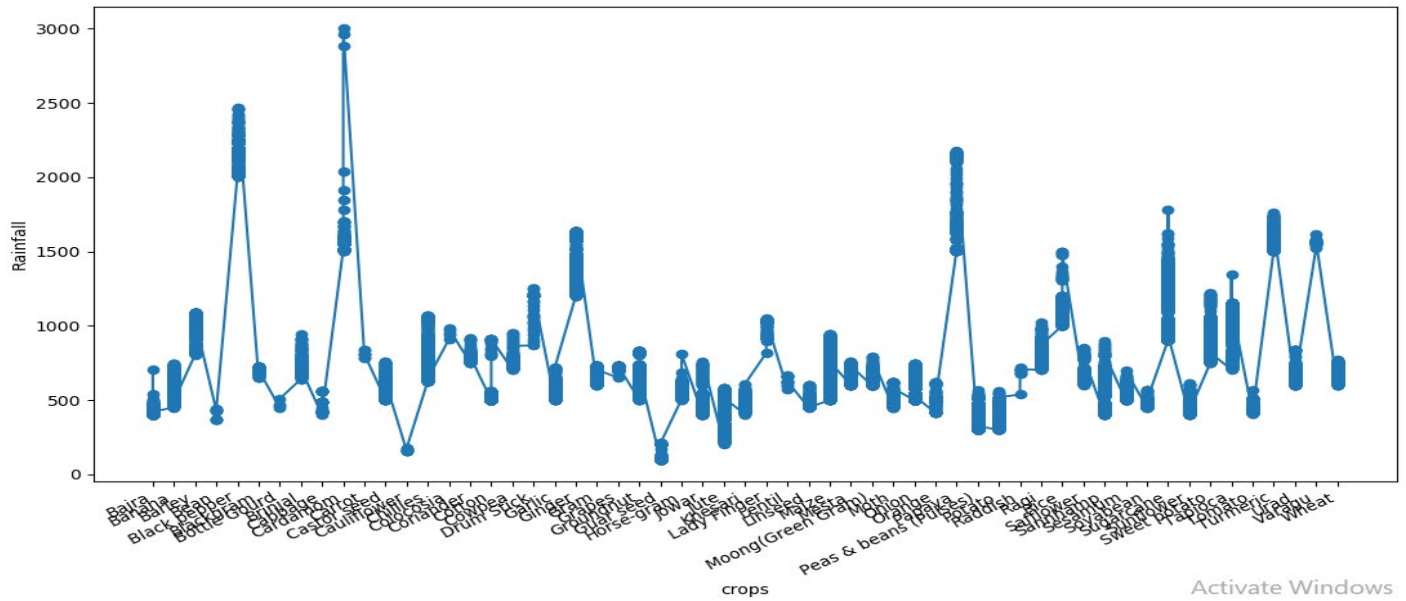


Figure 2. Anomalies in rainfall.

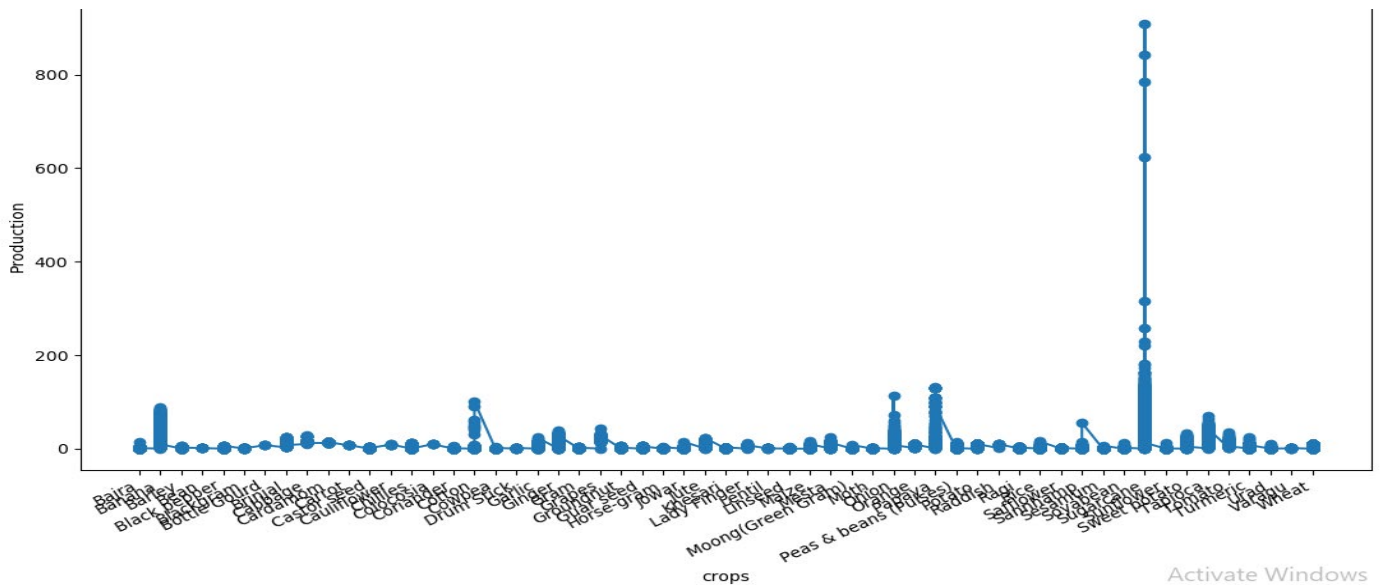


Figure 3. Anomalies in production.

doesn't suffer from the problem of over fitting as the decision trees. In DT with agricultural data we have the problem of over fitting as there is no commonality among the data. Each crop will have different values for all of its features i.e. temperature production etc.

DT will grow to the maximum size when it tries to fit the training data closely. This feature of DT may result in good accuracy but effects prediction time negatively as depth of the tree is high. This implies that the noise and arbitrary vacillations in the training data is learned as concepts by the model. The issue is that these concepts are applied on test data and adversely sway the models capacity. RF selects samples randomly and constructs different trees for each sample. This reduces the height of the tree and increases number of uncorrelated trees increasing the accuracy and decreasing the prediction time. The RF algorithm is also popular in feature ranking as it is easily calculated using mean decrease impurity or mean decrease accuracy method. From the trained data set the feature importance can be calculated.

In order to quantify the usefulness of all the variables in the entire random forest, relative importance of the variables is calculated. The importance returned in Skicit-learn represent how including a particular variable improves the prediction. A number of tricks in the Python language, namely

list, comprehension, zip, sorting and argument unpacking helps to find the importance of variables in the dataset. If any variable with less importance is resulted, it can be removed to improve the performance. But here as we have required importance for all the columns, they are considered.

The Random Forest Algorithm proceeds in two steps:

1. Construction of Random Forest
2. Prediction

Construction of random forest:

- a) Randomly select "K" sample data points from total "m" data points where $k \ll m$.
- b) For each datapoint randomly select the root node.
- c) Split the remaining features into child nodes.
- d) Repeat the b to c steps until "K" data points are completed.

Prediction:

- a) Considers the test data and use the rules of each randomly created

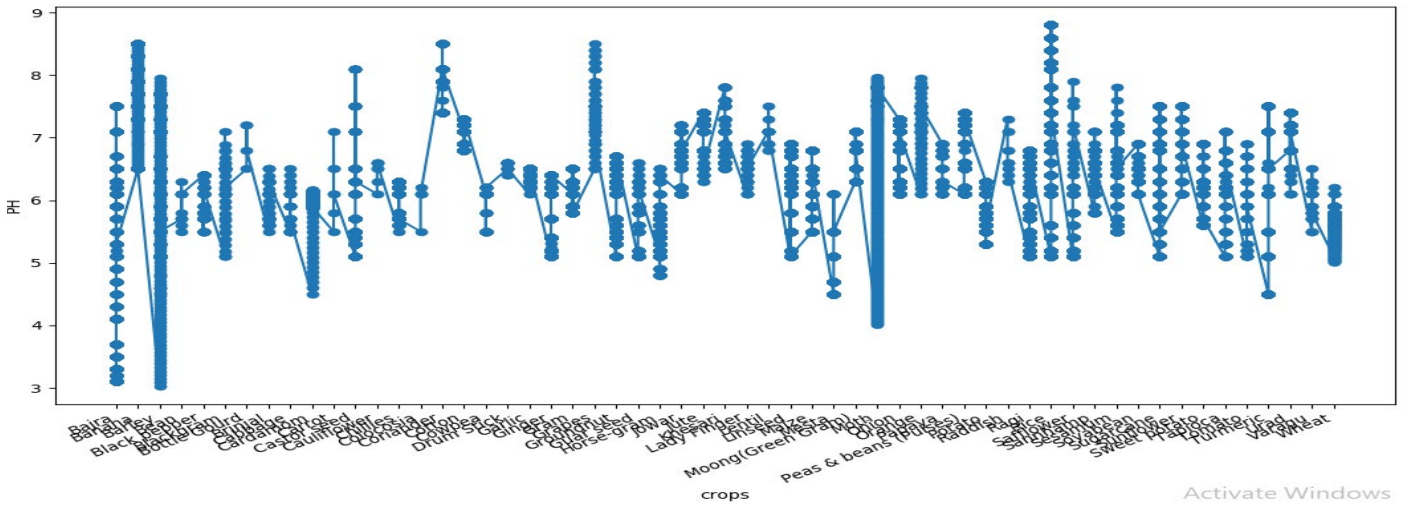


Figure 4. Anomalies in pH.

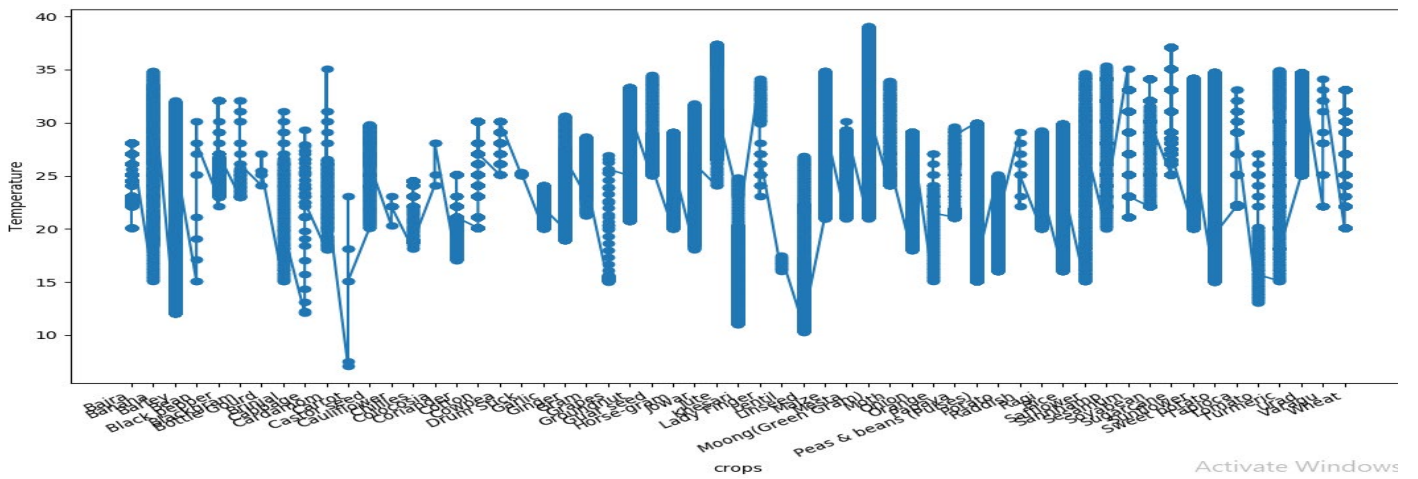


Figure 5. Anomalies in temperature.

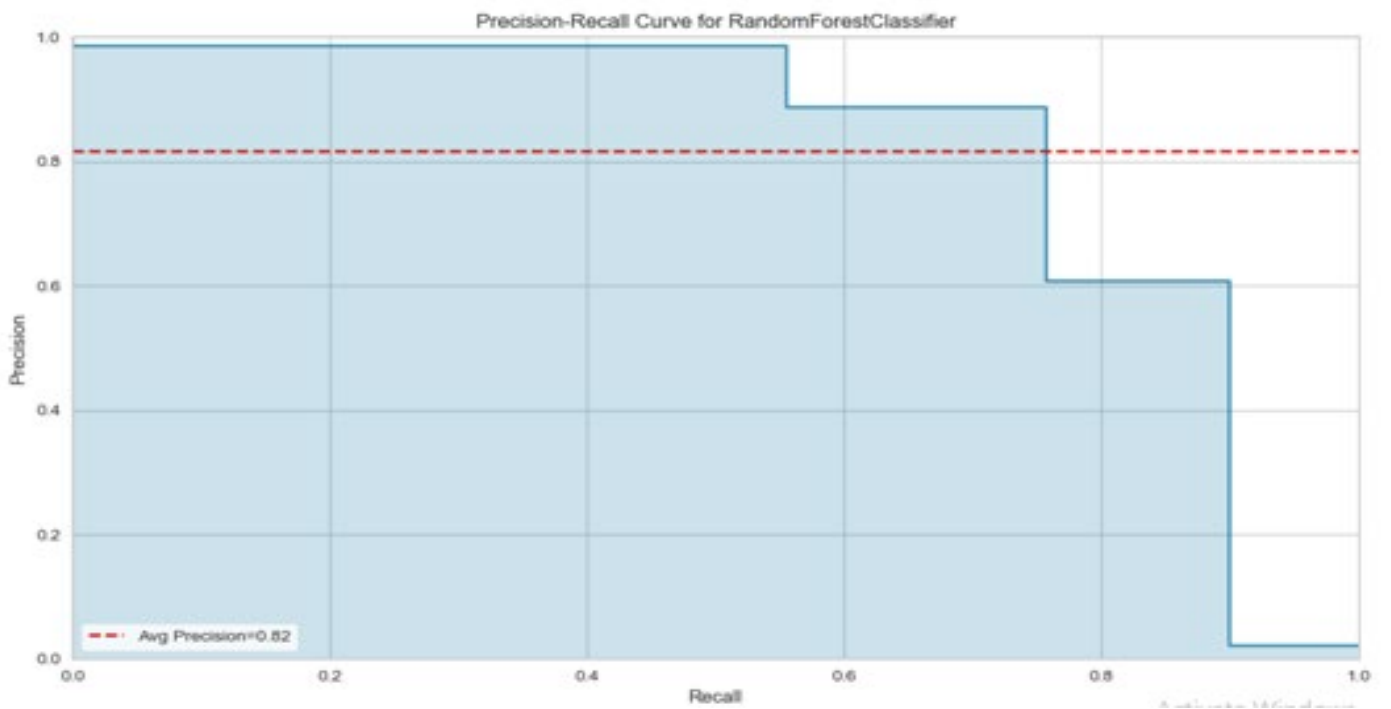


Figure 6. Precision recall curve for random classifier.

decision tree to predict the class variable and stores the predicted class (target).

b) Calculate the votes for each predicted class.

c) Consider the high voted predicted class as the final prediction from the random forest algorithm.

Recommendations for the crop to be yielded

Considering voting to arrive at the final conclusion is a good feature to arrive at a solution while some trees arrive at right answer and the other may arrive at wrong answer considering them as a group with the system of voting the algorithm predicts the right answer. The predicted crop that had highest voting, matching with the soil parameters entered by the farmer, is recommended as the best crop to be grown with good productivity [10].

Results and Discussion

The two important steps that make Random Forest algorithm popular are Bagging and Random featureness. Precision-Recall curves are a metric used to evaluate a classifier's quality, particularly when classes are very imbalanced. The precision-recall curve shows the tradeoff between precision, a measure of result relevancy, and recall, a measure of how many relevant results are returned. A large area under the curve in Figure 6 represents both high recall and precision, the best case scenario for a classifier, showing a model that returns accurate results for the majority of classes it selects.

Conclusion

In the proposed work Random Forest Algorithm is implemented on the data collected for crop prediction. As our future work we aim to collect the features of the soil by using IOT devices with sensors that measure moisture and temperature so that we can predict the crop according to the changing

features of the crop instantly without facing the difficulty of taking the samples to the lab. As RF algorithm is best suited for these predictions we focus on adding other features like electrical conductivity and texture of soil to improve the predictive accuracy of the crop to be yielded.

References

1. Pandhe, Abhijeet, Praful Nikam and Vijay Pagare. "Crop yield prediction based on climatic parameters." *Methodol* 6 (2019).
2. Shanmuga, Priya S and M. Abhinaya. "Feature selection using random forest technique for the prediction of pest attack in cotton crops." *Int J Pure Appl Math* 118 (2018): 2899-2903.
3. Ok, Asli Ozdarici, Ozlem Akar and Oguz Gungor. "Evaluation of random forest method for agricultural crop classification." *Eur J Remote Sens* 45 (2012): 421-432.
4. Malathi, B., Chari Appaji, G. Rajender Reddy and K. Dattatri, et al. "Growth pattern of millets in India." *Indian J Agric Res* 50 (2016): 382-386.
5. Nevavuori, Petteri, Nathaniel Narra, and Tarmo Lipping. "Crop yield prediction with deep convolutional neural networks." *Comput Electron Agric* 163 (2019): 104859.
6. Gopal, P.S. Maya and R. Bhargavi. "A novel approach for efficient crop yield prediction." *Comput Electron Agric* 165 (2019): 104968.
7. Konaté, Jacqueline, Amadou G. Diarra and Seydina O. Diarra, et al. "Syragri: A recommender system for agriculture in Mali." *Information* 11 (2020): 561.
8. Eitzinger, Anton, James Cock and Karl Atzmanstorfer. "GeoFarmer: A monitoring and feedback system for agricultural development projects." *Comput Electron Agric* 158 (2019): 109-121.
9. Kamilaris, Andreas and Francesc X. Prenafeta-Boldú. "Deep learning in agriculture: A survey." *Comput Electron Agric* 147 (2018): 70-90.
10. Gupta, Neeraj, Saurabh Gupta and Mahdi Khosravy. "Economic IoT strategy: the future technology for health monitoring and diagnostic of agriculture vehicles." *J Intell Manuf* 32 (2021): 1117-1128.

How to cite this article: Mantripragada, Rekha Sundari. "Agricultural Assistance Using Machine Learning Techniques." *J Comput Sci Syst Biol* 15 (2022): 417.