Ambient Science, 2022: Vol. 09(3); 47-50
DOI:10.21276/ambi.2022.09.3.ga03

Published by: National Cave Research and Protection Organization, India

ambient SCIENCE
Vol. 09(3):47-50
Year 2022

GENERAL ARTICLE

Crop Prediction Using Machine Learning

Devaki Palaniappan*, Jayanthi Manoharan, Deepthi Pragalathan, Shilpashri Murugesan

Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, Tamil Nadu

Study Area: Coimbatore, Tamilnadu, India
Coordinates: 11°01′00.5″N 76°57′20.9″E

Key words: Support Vector Machine, Random Forest Classifier, Soil, Climate

Abstract

Agriculture plays the important role in human survival. In today's situation, agriculture is the most important field. Farmers often use fertilizers in random quantities. So, this affects the crop yield and causes soil acidification and damages the soil and top layer, and there may be a chance of selecting the wrong crops which will reduce their income. Our model suggests the right crop based on weather conditions by using a Support Vector Machine and Random Forest to grow in a particular farm based on parameters such as PH,(N,P,K)and the parameters such as rainfall, temperature, and humidity.

Introduction:

Agriculture plays a vital role in production. In most cases, the farmers are not well aware of what kind of crops they should grow on their farms. This is why we are focusing on finding out the best crop to grow in order to get a better yield. We use the dataset from Kaggle to create a machine-learning model for predicting the suitable crop to grow at a particular place. By predicting the suitable crop to be grown, we will help the farmers to decide on the crop yield. This will eradicate the problem of fertilizer and nutrient deficiency in fields occurring because of choosing the wrong crops which can scale down the production efficiency. Our model is to provide the farmers with an ideal prediction for growing crops taking into consideration the factors such as the composition of the soil, and environmental factors like temperature, rainfall, humidity Nitrogen (N), Phosphorous (P), and Potassium (K).

The main ideology of the research is around the concept of predicting the most suitable crop to be grown with the help of a machine learning algorithm. It predicts the accuracy of future production of eleven different crops such as rice, maize, chickpea, black gram, lentil, kidney beans, orange, apple, papaya, coconut, pigeon peas, moth beans, mung bean, pomegranate, mango, banana, grapes, watermelon, muskmelon, jute, cotton and coffee crops using different supervised machine learning methods and predict the most suitable crop. The dataset is required to train the model according to the actual values and then test the model for accuracy. Presently our farmers are not using analysis and technology, so there may be the chance of incorrect selection of crops for cultivation which will reduce their income. We aim to help farmers, and we have also made it more and more visual and so easier to operate with their inputs. This process is mainly laying the foundation for further evaluation through the addition of secondary factors, which mainly have an impact on deciding the crop to be grown on a particular field in a specific place.

Others' findings and opinions

Medar *et al*. (2019), Gandge & Yogesh (2017) proposed a system of crop yield prediction dealing with the geographical value that denotes parameters including crop name, Rainfall etc . By this method, an accurate crop can be predicted by using the climatic conditions of the respective field. The Random Forest and Decision tree are the two algorithms used for crop prediction. By the comparison of the two algorithms, Random forest gives the best accuracy. Higher accuracy gives a higher profit for the crop.

The method proposed by Kumar *et al*. (2020), Bondre *et al*. (2019) gives the right crop according to the given input values. The use of a Support Vector Machine (SVM) gives higher accuracy and efficiency. The datasets used for the paper are locality data and sample crop datasets. By use of this method, the right crop can be predicted with the help of parameters like Fertilizers (Nitrogen, Phosphorous and Potassium) and soil content like Ph values. In the final prediction, Fertilizers and soil content are suitable for some crops like Rice, radish, carrot, Maize, and Black gram.

The two distinct processes used Nigam *et al.* (2019), Rajak *et al*. (2017) for predicting the yield of sugarcane class are from figures about graph characteristics. The first process is Random Forest which implements learn. The Second one is Gradient Boosting tree implements based on the XG Boost library, after the separation of value for training and testing purposes. The parameters contain soil type, area, width, yield accuracy, sugarcane grade and type

*Corresponding Author: *devaki.p.cse@kct.ac.in*

ISSN- 2348 5191 (Print) & 2348 8980 (Electronic)

Ambient Science, 2022: Vol. 09(3); 47-50
DOI:10.21276/ambi.2022.09.3.ga03

and rainfall This paper shows work applied is accurate for Random forests which means decision-making in Sugar industries.

Jain *et al.* (2020) and Reddy *et al.* (2019) proposed crop yield prediction using attributes like LTTS, climate and soil attributes, NDVI and algorithms like Supervised machine learning (SML) algorithm. This method classifies the yield prediction into three categories, time duration of SCLC forecasting by soil and weather attributes, Support Vector Machine Regression (SVR) is estimated by NDVI, and by the use of SVR, NDVI is set as input for sugarcane crop. In the proposed system, we have predicted the crop yield by forecasting climatic conditions using time series and forecast NDVI using predicted climatic parameters and using NDVI to predict actual stage crop.

The whole sugarcane system deals with the complexity and variability of sugar production (Pudumalar *et al.*, 2016). The aim of this adaptation, dealing with the numerous objectives of distinct sectors and broader circles. The research of water, soil viscosity and sugarcane climatic conditions gives the most effective, profitable and reliable sugar production.

Garanayak *et al.* (2021) proposed a method, that predicted actual crops for farmers to cultivate in their fields based on attributes like soil features, soil types and crop type. With the help of this method, we can forecast the right crop under climatic conditions and in places like states and districts. Thus the system will give suitable crops and increased yield production.

A method proposed by Suresh *et al.* (2018) deals with the suggestion of suitable crops based on soil type, rainfall, and degrees with high accuracy. Random Forest, Naive Bayes and SVM algorithms are used in this technique. The highest accuracy of the model is 99.91% .

Model which is evolved for the cause of forecasting the crop yield using the data yield and weather parameters like average daily max and min temperature, humidity values in the morning and evening and rainfall in mm (Kamatchi *et al.*, 2019). The model has been developed using weather conditions. Using this forecasting model estimates the preharvest of sugarcane yield for particular districts that can be computed in advance before the predicted harvest. As the data which is used for developing the forecast model is of high accuracy, and high reliability.

Further, Suresh *et al.* (2021) again develop a model that helps in forecasting yield to help farmers to be better and take precautions against climate conditions. This system deals with numerical methods, extends linear regression to algorithms of machine learning like the random forest, based on data of climatic conditions (rainfall, temperature); gives satellite input as (EVI, NDVI, from MODIS). At the graph point, the sound during production will make the effects of weather on yields and leads to prediction. The outcome of this study that helps the surveys, refine models, and the parameters that were not past seem to be the same.

Kulkarni *et al.* (2018) and Doshi *et al.* (2018) proposed a system that deals with suitable crops based on soil parameters. Proposed for the crops like banana, paddy, cotton, sorghum, groundnut, pulses, vegetables, sugarcane, coriander and Drainage, Soil, Depth, Ph, Permeability, texture, Water holding, and Erosion were the parameters considered for prediction. Moreover, this method used distinct machine learning classifiers like support vector machine (SVM), ANN, Random Forest and Naïve Bayes for crop prediction which gives higher precision, thereby increasing productivity. Thus, this paper is helpful for farmers to increase production in agriculture, prevent soil degradation and reduce chemical use in crop production and also use water resources efficiently.

Distinct machine-learning algorithms are used for crop yield prediction (Dighe *et al.*, 2018). Moreover, the yield is estimated by the production of machine learning algorithms like root mean square error, support vector machine and random forest. Along with the prediction of machine learning algorithms, the right crop is to be cultivated at the right time using soil parameters and weather conditions.

The use of scale yield monitor yield prediction is evaluated (Paul *et al.*, 2015; Jeong *et al.*, 2016). The accuracy prediction classifies, predictive models into four distinct parameters. The overall system focuses on sugarcane yield prediction and climatic conditions, especially time. The model MLR and ANN give the higher predicted error value. At the time of testing, the ANN reduces the yield of sugarcane, and the RF predicted a higher yield.

Babu (2013) suggest a way to help farmers focus on vegetable cultivation for profitable cultivation in Sri Lanka, the demand for recent technologies in this sector is greater. Deep Learning, Machine Learning and Visualization were used. As the product application is developed. The users should make input their location to begin the process of prediction. Data pre-processing is initiated after receiving the location in the system. After the prediction, optimization for multiple sowing of crops according to the user input and requirements, visualizing cultivation and data production on graphs and maps will be given in the application. This paper also demonstrates the steps of model development, model testing and training.

Loganathan *et al.* (2019)'s work was to forecast worldwide and regional food security, the accurate yield forecasting is obtained. Machine learning algorithms will give an accurate estimate of wheat output for the farmer before the two months of crop maturation. The research has estimated the accurate agricultural production and the predictions using data such as weather, and temperature. The authors were able to estimate wheat production with around 75% accuracy before two months of the growing

Ambient Science, 2022: Vol. 09(3); 47-50
DOI:10.21276/ambi.2022.09.3.ga03

season. The project predicts the weather yield in India by weather features and raw crop features.

Proposed System

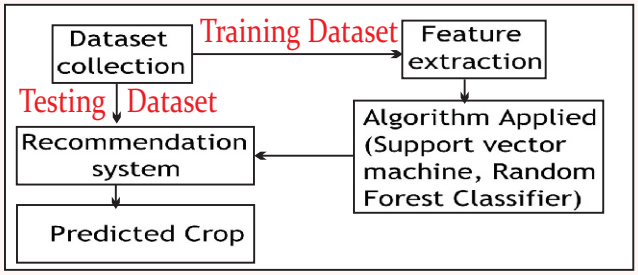

Figure-1: Block Diagram of Proposed System

There are five phases which are:

1) Dataset Collection
2) Pre-processing of data
3) Feature Extraction
4) Machine Learning Algorithm
5) Recommendation System

(1) **Dataset collection:** it consists of parameters like Nitrogen(N), Potassium(K), Phosphorous(P), PH value, temperature, Rainfall and Humidity. The dataset from the Kaggle has 2200 rows of data which have been taken from past historic data. This dataset includes 11 different crops such as apple, maize, cotton, mung bean, kidney beans, pigeon peas, moth beans, black gram, pomegranate, muskmelon, banana, mango, grapes, watermelon, lentil, rice, orange, coffee, coconut, chickpea, jute and papaya.

(2)**Pre-Processing:** the data pre-processing can be done in various stages, beginning with

1. Reading the collected dataset, the process continues to data cleaning.
2. The data which is obtained from different resources are in raw form.
3. This contains some missing, incomplete, inconsistent, and redundant data.
4. In preprocessing, the missing or incomplete data should be filtered and normalized

(3)**Feature Extraction:** this part involves the process of prediction and using most important attribute from the dataset. Through this process, redundant and unusable information is removed for the application of classifiers.

(4) **Methodology:** in this proposed system, we applied different Machine Learning algorithms like Support Vector Machine (SVM), Random Forest Classifier, Logistic Regression.

Support Vector Machine (SVM): it is a model or supervised machine learning algorithm which is used for classification. It is also used for regression and classification challenges.SVM is generally used as training data points in space which is categorized into groups by an intelligible gap. The input x denotes PH, temperature, humidity, Rainfall and (N,P,K) values, the output y denotes the suitable Predicted Crop.

We have applied the Support Vector Machine (SVM) approach in our model:

(i) After dividing data into test and train, we will use the support vector classifier class, which is denoted as svc in the library
(ii) SVM classifies the linearly separable data, then will see non-linear kernels. Then we can create SVM classification object
(iii) At last we fit our data, which is called to train the algorithm on training data.

The classifier has got an accuracy of 71%.

B. Random Forest Classifier: Random forest is a supervised learning technique. It consists of many numbers of decision trees for different subsets of the data work on a single subset instead of a large one. This improves the accuracy of modal prediction by several folds. It is usable even for large and varied datasets as it results with high accuracy in less time.

$$Gini=1-\sum_{i=1}^{c}(p_i)^2$$

This formula focuses on the probability and class to find the Gini of a node of each branch, determining, which branches are likely to occur. Here, pi denotes the relative frequency and c represents the no of classes. The random forest has 3 parameters like n tree-which denotes the n number of trees, and m-denotes how variables were taken at a split of node. The classifier has got an accuracy of 55%as shown in Table-1.

Output analysis

Table1: Accuracy of each Algorithm

Algorithm used	Accuracy(in %)
Support Vector Machine	71%
Random Forest	55%

Ultimately, this system deals with decision making which will help farmers to choose the right crop depending on soil parameters like Nitrogen, Phosphorous, Potassium, PH, Humidity, Temperature, and Rainfall. We will definitely get greater productivity of the crops and also get greater income. Here, the farmers can plant the right crop to get increased yield using suggestions also helps in the overall greater income or profitability of our country. This model will surely help farmers to produce crops with greater yields. This research will definitely give a deep insight into the different conditions affecting crop yield which can help farmers to reduce income losses and generate a greater yield. This model will contribute positively towards reducing environmental depletion and maintains a balance between the ecological health and continued agricultural growth of the land.

References:

Babu, S. (2013): A software model for precision agriculture for small and marginal farmers, pp. 352-355. In: *IEEE Global Humanitarian Technology Conference: South Asia Satellite*.

Devdatta, B.A. & Santosh M. (2019): Prediction of crop yield and fertilizer recommendation using machine learning algorithms. *Int. J. Eng. Appl. Sci. Tech*., 4(1):371-376

Dighe, Dighe, D., Joshi, H., Katkar, A., Patil, S. &

Doshi, Z., Sanghvi, D.J., Nadkarni, S., Agrawal, R. & Shah, N.(2018): Agro consultant: intelligent crop recommendation system using machine learning algorithms, pp. 1-6. In: *Fourth International Conference on Computing Communication Control and Automation*.

Gandge, Y. (2017): A study on various data mining techniques for crop yield prediction, pp. 420-423. In: *International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques*.

Garanayak, M., Sahu, G., Mohanty, S.N., Jagadev, A.K. (2021): Agricultural Recommendation System for Crops Using Different Machine Learning Regression Methods. *Int. J. Agric. Environ. Inf. Syst.*, 12(1):1-20

Jain, S. & Dharavath R. (2020): Machine Learning convergence for weather-based crop selection, pp 1-6. In: *IEEE International Students' Conference on Electrical, Electronics and Computer Science*.

Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E. & Kim, S.H. (2016): Random forests for global and regional crop yield predictions. *PLoS One*, 11(6):e0156571

Kamatchi, S.B. & Parvathi, R. (2019): Improvement of Crop Production Using Recommender System by Weather Forecasts. *Proc. Comp. Sci*., 165: 724-732.

Kokate, S. (2018): Survey of crop recommendation systems. *Int. Res. J. Eng. Tech*., 5:476-481

Kulkarni N., H, Srinivasan, G.N., Sagar, B.M &Cauvery, N.K.(2018): Improving crop productivity through a crop recommendation system using ensembling technique, pp. 114-119. In: *3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions*.

Kumar, Y.J.N., Spandana, V., Vaishnavi, V.S., Neha, K. & Devi, V.G.R.R. (2020): Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector, pp. 736-741. In: *5th International Conference on Communication and Electronics Systems*.

Loganathan, B.G., Mohan, E. & Siva Kumar, R. (2019): IoT based water and soil quality monitoring system. *Int. J. Mech. Eng. Tech*., 10(2): 537-541

Medar, R.V.S. & Rajpurohit, S. (2019): Crop yield prediction using machine learning techniques, pp. 1-5. In: *IEEE 5th International Conference for Convergence in Technology*.

Nigam, A., Saksham G., Archit, A. & Parul A. (2019): Crop yield prediction using machine learning algorithms, pp. 125-130: In: *Fifth International Conference on Image Information Processing*.

Paul, M., Vishwakarma, S.K. & Verma, A. (2015): Analysis of soil behaviour and prediction of crop yield using data mining approach, pp. 766-771. In: *International Conference on Computational Intelligence and Communication Networks*.

Pudumalar, S, Ramanujam, E., Rajashree, R.H., Kavya, C., Kiruthika, T. & Nisha, J. (2016): Crop recommendation system for precision agriculture, pp. 32-36. In: *Eighth International Conference on Advanced Computing*.

Rajak, R.K., Ankit, P., Mitalee, P., Pooja, S., Suresh, R. & Avinash, D. (2017): Crop recommendation system to maximize crop yield using machine learning technique. *Int. Res. J. Eng. Tech*., 4(12):950-953

Reddy, A.D., Bhagyashri D. & Aarti, W. (2019): Crop recommendation system to maximize crop yield in Ramtek region using machine learning. *Int. J. Sci. Res. Sci. Tech*., 6(1):485-489.

Suresh, A., Kumar, PG, & Ramalatha, M. (2018): Prediction of major crop yields of Tamilnadu using K-means and Modified KNN, pp. 88-93. In: *International Conference on Communication and Electronics Systems*.

Suresh, G., Kumar, S.A., Lekashri, S. & Manikandan, R. (2021): Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming. *Int. J. Mod. Agric*., 10(1):906-914.