International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers

***[1]N Sudha Laxmaiah, [2]Konda Shireesha, [3]Bandaru Prathima**
[1]Assistant Professor, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, India
[2,3]Students, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, India

ARTICLEINFO

Article History:
Accepted: 20 April 2023
Published: 08 May 2023

Publication Issue
Volume 10, Issue 3
May-June-2023

Page Number
63-68

ABSTRACT

Agriculture is the heart of many countries and soil is the main important element of agriculture. There are different soil kinds and each kind has different features for different crops. In this field, now a day's different methods and models are used to increase the quantity of the crops. So the main purpose of this system is to create a model that helps farmers to know which crop should take in a particular type of soil. In this system, we are using machine learning techniques which help to suggest the crops according to soil classification or soil series. The model only suggests soil type and according to soil type it can suggest suitable crops. In this, different classifiers are used and according to that the model suggests the crop.

Keywords : Soil Series, Machine Learning, Data Mining, Chemical Features, Prediction

I. INTRODUCTION

Agriculture is the main source of the Indian Economy. India is also known as a country of farmers. In India 50% man force is involved in agriculture activities. In Agriculture the soil is the main and basic thing [1]. But now also the farmers are using the traditional method. Because of the traditional method farmers did not get satisfactory results means the quantity of crops is not increasing. To increase the quantity of crops need good quality of soil. So soil testing is done. Soil testing is important rather than the main task of farming. The production and quality of crops totally depend on the soil. Soil testing is essential because it gives information on all nutrients which present in the soil such as Ca (Calcium), K (Potassium), and N (Nitrogen). Farmers in India, especially many regions in Maharashtra state faces drought due to which their crop and yielding are getting degraded. They don't have any idea about the availability of nutrients in their field. They use their own experience to plow the crop which has very less success ratio. Due to less success ratio they are unable to pay their loan amount sanctioned for their crop. In unsuccessful for their repayment of the loan amount they attempt to suicide. It is the main reason for increasing suicide cases [2]. To help the farmers to decide the crop to be plow for their benefits we motivated to build this system. This

dataset consists of the available nutrient for farmers' soil and rainfall for a particular region. Based on nutrients value, our system predicts soil type. According to the soil type system predicts a list of crops that can grow in a particular soil. Hence the yield of the crop increases, as well as the farmer, earn more money with this new method. We create the system with the help of advanced technology. We use machine learning to create the system. Machine learning concentrates on the creation of computer programs that can access data and use it to learn from that. Machine learning allows building models from sample data and give the ability to take decision automatically according to past experiences [2].

II. RELATED WORK

In this section, we review some of the significant works done in the agriculture field for crop prediction.
In [3], the authors concentrated on the use of applications of data mining techniques in the agricultural field. As data mining is a new rising technology so authors also study and examined the problem of forecasting agricultural productivity. The authors discussed the main objective of this work was finding the desired data models that give high accuracy and high generality in terms of the yield forecasting capabilities. For these different types of data mining techniques were judged on different data sets by authors.
In [4], the authors proposed a method named Crop Selection Method (CSM). The authors specify the proposed method to help to find out the crop selection problems and also help to raise the net yield rate of crop over seasons and help to get maximum economic growth. The authors discussed the different influencing parameters which can be used for crops by different forecasting models. The authors also specify the machine learning and different methods of machine learning. Theproposed Crop Selection Method classified crops as seasonal crops, whole year crops, short time plantation crops, long time plantation crops.
In [5], the authors suggested a smart way to forecast the crop yield and also suggested the ideal climatic factors to maximize the crop yield. In this the multivariate polynomial regression, support vector machine regression, and random forest models were used to forecast the crop yield per acre. The suggested method also uses yield and weather data which was collected by the author from the United States Department of Agriculture. The author also used the Mean Absolute Error (MAE), Root MeanSquared Error (RMSE), median absolute error and R-Square values to compare between multivariate polynomial regression, support vector machine regression and random forest.
In [6], the authors used the sliding window non-linear regression method to suggest crop yield and price by evaluating patterns from the past data. The authors studied several districts of the state of Tamilnadu. The authors proposed a system designed in such way that it suggests the best crop choices for a farmer to do farming. The designed system had done demand level classification. The demand level classification means the demand for the crops is predicted by classifying the dataset of change in the market prices of crops. In the proposed system the authors also did the text to speech conversion.
In [7], authors have proposed an ideal system called AgroConsultant, which aims to help the Indian farmers to decide about which crop to grow depending on the season, geographical location, and soil characteristics. For making such a system the authors used a machinelearning algorithm like the random forest, K-Nearest Neighbors (K-NN), Decision Tree, Neural Network. The proposed system also includes the map visualization feature and rainfall predictor.
In [8], the authors created an innovative structure named as eXtensible Crop Yield Prediction Framework (XCYPF). This framework provides the selection of crop and also dependent and independent

variables and datasets for crop yield forecasting. The authors discussed how this framework is adaptable and expandable. The authors used rainfall data and surface temperature for crop yield forecasting for rice and sugarcane crops. The authors used the new methodology which combines the use of vegetation indices.

In [9], authors designed a crop yield prognosis model (CRY) which works on an adaptive cluster approach. It uses the Bee Hive modeling approach to study and classify the crop based on crop growth pattern and yield. The authors also described the Bee Hive Cluster which was used because it describes the agricultural datasets which help to decide crop growth and the Bee Hive Cluster is considered as heterogeneous data cluster which maintained as repositories. The authors studied the Bee Hive which had better performance than others. This Bee Hive algorithm is used for getting patterns and the graphs were plotted to describe yield variation.

In [10], authors built a cloud-based agricultural framework which helps the Indian farmers and agricultural industries to get useful information about agriculture. This framework provides soil classification and crop yield forecasting. The authors use a hybrid support vector machine for soil classification and customized artificial neural network to forecast wheat yield. Also the authors used Amazon S3 to store agricultural data and they used the Heroku cloud.

III. PROPOSED SYSTEM

For the purposes of predicting crop yield and making crop recommendations, we will use the Random Forest Regression and Random Forest Classifier. The following list of steps describes the step-by-step data technique used in this work:

A.Importing the required library

B.Dataset is loaded from the csv file

C.Data is split into training and test sets

D.Using the random forest algorithm

E.Build the model

Research revealed that the random forest algorithm outperformed other algorithms in terms of accuracy. In comparison to other approaches like neural networks, it is less prone to overfitting, yields more accurate predictions, and is also computationally less expensive.

Crop Yield Prediction: In order to construct a single, effective model, we use a technique called Random Forest Regression which combines different decision tree models. Bootstrap aggregation, often known as bagging, is the method used to create the decision trees. In this procedure, subsets of the data are chosen at random, and the decision tree is then fitted to each subset. The final model is created by averaging the output from each tree. The features and target variables (crop yield) are first isolated from the dataset once it has been loaded from the csv file. Following that, the data is divided into training and testing datasets. Since models that require numerical input cannot use categorical variables directly, one hot encoding is used to represent the categorical variables (state, district, season, and crop) as numerical values in the model. The parameters: state, district, season, crop, and area in hectares are then taken into consideration by a random forest regression. Based on these variables, our algorithm builds decision trees and generates a forecast of the crop's production at that specific location. The combined forecasts of the separate trees form the basis of the prediction. The random forest regression model is an ensemble learning technique that combines the forecasts of various decision trees, assisting in increasing the model's accuracy and lowering the variance of the forecasts. Because the target variable is continuous (i.e., the output is a real value) and we wanted to predict the value of the output based on the input features, Random Forest Regression was used in this situation.

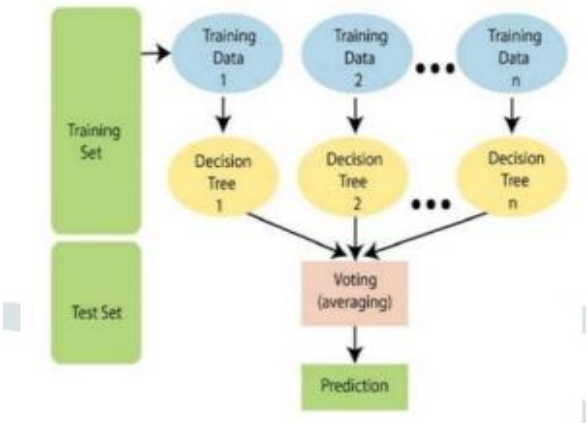

Fig 1 : Random forest algorithm

Crop Recommendation: With the help of many stated factors, we use the Random Forest Classifier to select crops based on user input for nitrogen, phosphorous, potassium, temperature, humidity, pH, and rainfall. The dataset (a CSV file) was then read. The dataset is split into training and testing datasets after being partitioned into labels and features. The Random Forest Classifier is then applied once the data has been divided. It constructs a large number of decision trees using randomly selected input features and data, and it produces a class that corresponds to the mean prediction (regression) or classification of all the individual trees. A random subset of the characteristics is selected at each node in the decision tree, and the best split point is established using an impurity metric (such as the Gini index or entropy) of the classes at the node. The impurity measure is minimized in order to select the ideal split point. After the trees have been constructed, predictions can be made by using the mean prediction (regression) or the mode of the classes (classification). A few tuning parameters for the Random Forest algorithm include the amount of features to take into account at each split point, the number of trees to construct, and the minimal number of samples needed at a leaf. Because we want to categorize the data into a particular group and the target variable is categorical (i.e., the output is discrete), we use the Random Forest Classifier in this situation.

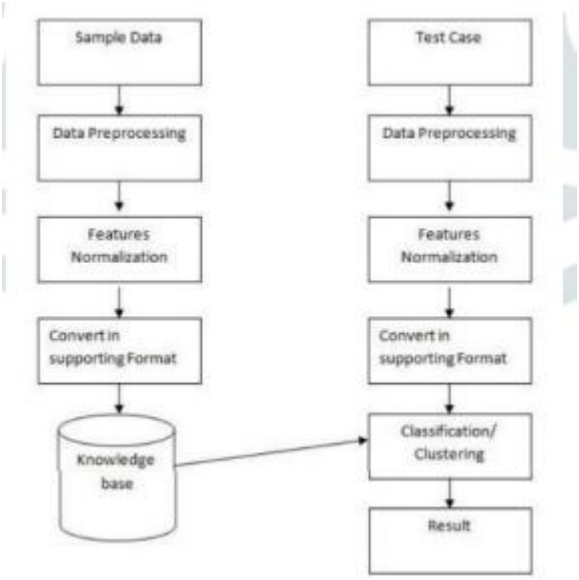

Fig 2 : Flowchart of prediction model

IV. RESULTS

The pre-processing method in our project goes through a feature selection procedure and chooses appropriate features. The random forest classifier is then given the selected feature. In this manner, the random forest algorithm categorizes the traits, forecasts if the crop is appropriate for the agricultural land, and outputs the crop yield. Evaluation is done on the results of the random forest algorithm method.

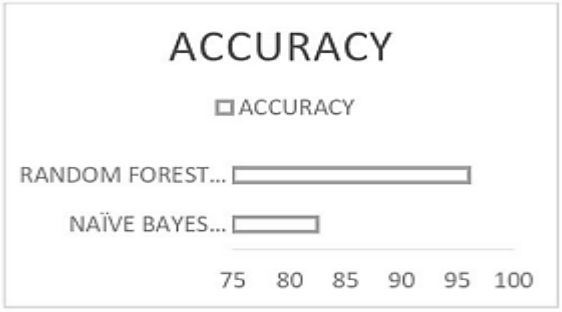

Fig 3 : Accuracy comparison

When compared to the accuracy level of the existing system, the accuracy level for the proposed system is higher. Given that it uses a bagging technique, the random forest algorithm has a high accuracy level. While decision trees are a boosting technique that increase the complexity of the model and cause it to underfit the training data, this technique reduces the complexity of the model that overfits the data.

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented a few algorithms that analyze soil compositions to forecast crop yields, recommend the most suitable crop options based on datasets, and provide clear and accurate information corresponding to soil composition. In terms of crop production prediction and crop recommendation, it has been found that combining random forest regression with random forest classifier techniques yields the best accurate results. Huge efforts have been made recently to increase crop productivity and forecasting. The development of precise crop growth estimation models employing data and communication technologies will benefit agricultural stakeholders and workers, boost employee accountability for food imports and exports, and improve food security. The crop dataset is analyzed using a random forest classifier. Finally, it can be said that the Random Forest Algorithm can be a real method of increasing crop yield with greater accuracy in agricultural areas. The crop dataset is predicted using a hybrid combination of machine learning algorithms in the future. In order to anticipate crop disease and the best crop for the soil, image datasets are also analyzed and added to the framework.

VI. REFERENCES

[1]. S. P. Raja, B. Sawicka , Z. Stamenkovic and G. Mariammal, "Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers," in IEEE Access, vol. 10, pp. 23625- 23641,2022,doi: 10.1109/ACCESS.2022.3154350.

[2]. M. Rashid, B. S. Bari, Y. Yusup, M. A.Kamaruddin and N. Khan, "A ComprehensiveReview of Crop Yield Prediction Using MachineLearning Approaches With Special Emphasis onPalm Oil Yield Prediction," in IEEE Access, vol.9, pp.6340663439,2021, doi:10.1109/ACCESS.2021.3075159.

[3]. D. J. Reddy and M. R. Kumar, "Crop YieldPrediction using Machine Learning Algorithm,"2021 5th International Conference on IntelligentComputing and Control Systems (ICICCS), 2021,pp.14661470, doi:10.1109/ICICCS51141.2021.94 32236.

[4]. N. Suresh et al., "Crop Yield Prediction UsingRandom Forest Algorithm," 2021 7th International Conference on Advanced Computing and Communication Systems

[5]. Agarwal, Sonal & Tarar, Sandhya. (2021). A HYBRID APPROACH FOR CROP YIELD PREDICTION USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS. Journal of Physics: Conference Series. 1714. 012012. 10.1088/1742- 6596/1714/1/012012

[6]. 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.963970,doi:10.1109/ICICV50876.2021.9388479.

[7]. M. Kavita and P. Mathur, "Crop YieldEstimation in India Using Machine Learning,"2020 IEEE 5th International Conference onComputing Communication and Automation(ICCCA),2020,pp.220224,doi:10.1109/ICCCA49 541.2020.9250915.

[8]. Y. J. N. Kumar, V. Spandana, V. S. Vaishnavi, K. Neha and V. G. R. R. Devi, "Supervised Machine learning Approach forCrop Yield Prediction in Agriculture Sector," 2020 5th International Conference on Communication and Electronics Systems (ICCES),2020,pp.736741,doi:10.1109/ICCES48766.2020.9137868.

[9]. P. S. Nishant, P. Sai Venkat, B. L. Avinash and B. Jabber, "Crop Yield Prediction based on

Indian Agriculture using Machine Learning," 2020 International Conference for EmergingTechnology(INCET),2020,pp.14,doi:10.1109/INCET49848.2020.9154036

[10]. R. Reshma, V. Sathiyavathi, T. Sindhu, K. Selvakumar and L. SaiRamesh, "IoT based Classification Techniques for Soil ContentAnalysis and Crop Yield Prediction," 2020 Fourth International Conference on I-SMAC (IoTin Social, Mobile, Analytics and Cloud) (I- SMAC), 2020, pp. 156-160, doi: 10.1109/I-SMAC49090. 2020.9243600.

Cite this article as :

N Sudha Laxmaiah, Konda Shireesha, Bandaru Prathima, "Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 3, pp. 63-68, May-June 2023.
Journal URL : https://ijsrst.com/IJSRST523102112