International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Crop Yield Prediction based on Indian Agriculture using Machine Learning

Ashish Maurya[1], Govind Yadav[2], Chandrabhan Yadav[3]

[1]Ashish Maurya CSE department ITM Gida Gorakhpur
[2]Govind Yadav CSE department ITM Gida Gorakhpur
[3]Chandrabhan Yadav CSE department ITM Gida Gorakhpur

Abstract - Fertilizer use is typically under the limited control of farmers. For the farmers to achieve higher yields and reduce fertilizer loss, competent guidance is required for the best use of these fertilizers. Additionally, there is a connection between rainfall volume and nutrient loss for various fertilizer applications after each rainfall event. Rainfall that is moderate and falls at the right moment can help nutrients penetrate the soil's rooting zone and dissolve dry fertilizer. However, too much rain can increase the possibility of runoff and the pace at which nutrients like nitrogen (N) which is quint essential, phosphorus (P), and potassium (K) which are crucial, manganese (Mn), and boron (B) that are present in the soil. This research presents nutrient recommendations using an updated iteration of the random forest algorithm which is based on time-series data to forecast the required quantity of nutrients for various crops by examining rainfall patterns and crop fertility. The method suggested in this study, comes in handy for improving soil fertility by providing nutrients recommendations for optimum conditions for crop growth and reducing leaching and runoff potential.

***Key Words*:** Machine Learning, NPK Prediction, Proper Yield, Indian Agriculture, Random Forest.

1. INTRODUCTION

Agriculture plays a very important role in national economic growth. Agriculture contributes 17-18% to India's GDP and ranks second worldwide in farm outputs. Plants need fertilizers and fertilizers replace the nutrients which crops take from the top layer of the soil. The absence of fertilizers can cause a drastic reduction in the volume of crop output.But fertilization requires precise action. Rainfall patterns and the amount of nutrients needed for a certain crop must be considered when using fertilizers. Machine learning is the current technology that can solve this problem by using available data for crop fertility and rainfall. Farmers can greatly benefit from the support of robust information about crops. The proposed model also uses a machine-learning algorithm (random forest regression algorithm with k-fold cross-validation technique) and takes two inputs from the user that are crop and location. After applying the algorithm, the model predicts the amount of nutrients required along with the best time to use fertilizers. The website is built using Python (language) to provide access on all platforms and can be shared among users.

2. RELATED WORK

A comprehensive study of the available literature presents a catalog of previous studies to address this issue. Machine Learning algorithm have been done and achive various result in agriculture fields.

The authors show in [1] “Predicting fertilizer” usage can assist farmers to attain a proper yield with little waste by preventing toxicity and deficiency in plants to some extent.

The Paper [2] “Makes use of fuzzy logic systems” that enable the reduction of fertilizer usage which results in an increase in crop productivity. Additionally, it shows that the enhanced efficiency of fertilizers is not sufficient for complications that can be caused by compaction. These issues can be prevented by improving the fertilizer recommendation which requires the establishment of a quantifiable relation under N and P for fertilizer usage,in terms of agricultural yield.

The Paper [3] “Crop yield prediction” It is a difficult task to predict crop yield due to stochastic rainfall patterns and also temperature variation. So we can apply different data mining techniques as propounded in state that nitrogen leaching is prone in areas that have no-till management and this may cause crop loss.

The Paper [4] “Survey on Crop yield using beehive clustering techniques as a prediction model for agricultural datasets”, Agricultural research around the world emphasizes the need for a reliable system to forecast and boost crop growth. There is a clear need for an prophetic modelling system with efficient predictive yield management methodology die to multi- dimensional variable metrics and also the lack of a prophetic modelling approach, which leads to crop yield loss.

Additionally the Paper [5] “Crop Yield prediction using data analytics and hybrid approach”, Agricultural data is constantly and massively generated. As a result, the age of bid data has arrived for agricultural data. Data collection via electronic devices is aided by smart technologies. We will analyze and mine this agricultural data in our project to obtain useful results using technologies such as data analytics and machine learning, and this information will be provided to

farmers in order to improve crop yield in terms of production and productivity.

Morever, the Paper [6] "Survey on Crop yield using KNN Algorithm on Indian Agricultural", Creating a software program framework for prognostication crop yield primarily based totally on weather and plantation information. The core of this scheme is an approach for unattended data partitioning for locating spatio-temporal trends in climate knowledge using kernel strategies, which provides strength in dealing with complicated data. For this purpose, a strong weighted kernel k-means formula with spatial constraints is given. The algorithm can efficiently manage noise, outliers, and auto-correlation in spatial data for successful and efficient data analysis.

Besides the Paper [7] "Proposed work on Data mining techniques for crop yield prediction", Predicting crop yields well ahead of harvest will help farmers and government agencies make more informed decisions about storage, sale, setting minimal assist rates, importing/exporting, and other activities. A detailed analysis of vast quantities of data derived from varied variables including soil quality, pH, EC, N, P, K, and so on. Since crop prediction requires many databases, this prediction method is an ideal candidate for data mining. We derive information from massive amounts of data using data mining.

The Paper [8] "Artificial Neural Network Rice crop yield prediction", The intention of this study was to use neural networks to predict rice manufacturing yield and to research about the factors that impact rice crop yield in Maharashtra, India. Data for 27 districts in Maharashtra state, India, were collected from publicly available Indian government records. For the Kharif season from 1998 to 2002, the parameters considered were common temperature, highest temperature, reference crop evapotranspiration, location, production, and yield.

Also the Paper [9] "Fuzzy logic-based Technique", Crop Yield prediction using Temperature and Rainfall parameters predicting through ARMA, SARIMA and ARMAX models: Agriculture is extremely important in India's economy. As a result, crop yield forecasting is an essential task for India's development. Crops are affected by a variety of weather conditions, including temperature and rainfall. As a result, when forecasting a crop's yield, it's important to take these factors into account.

Lastly, the Paper [10] "Support verctor machine learning on agriculture" it use ANNs for the prediction of rice yield in the districts of Maharashtra, India. They considered climatic factors namely (considering range) temperature, precipitation and reference crop evapotranspiration. The records were collected from Indian Government repository from 1998 to 2002. The generated outputs show that RF is an effective and different machine-learning method for crop yield predictions at regional and global scales for its high accuracy.

3. PROPOSED MODEL

In this study, a predictive model for the nutrients required for crops was obtained using random forest. Random forest regression with the k-fold cross-validation technique represents the model and the model with acceptable accuracy for the prediction is then obtained. A total of seven features have been used to evaluate the algorithm.

As shown in Fig 3.1, the algorithm requires input from the user (such as location and crop). The location is fed to the Weather API which will return certain characteristics (e.g. temperature, humidity, rainfall) and if there is a possibility of heavy rainfall, a precautionary message is displayed to the user, otherwise, the proposed algorithm is followed.

The flowchart of the methodology is presented below:

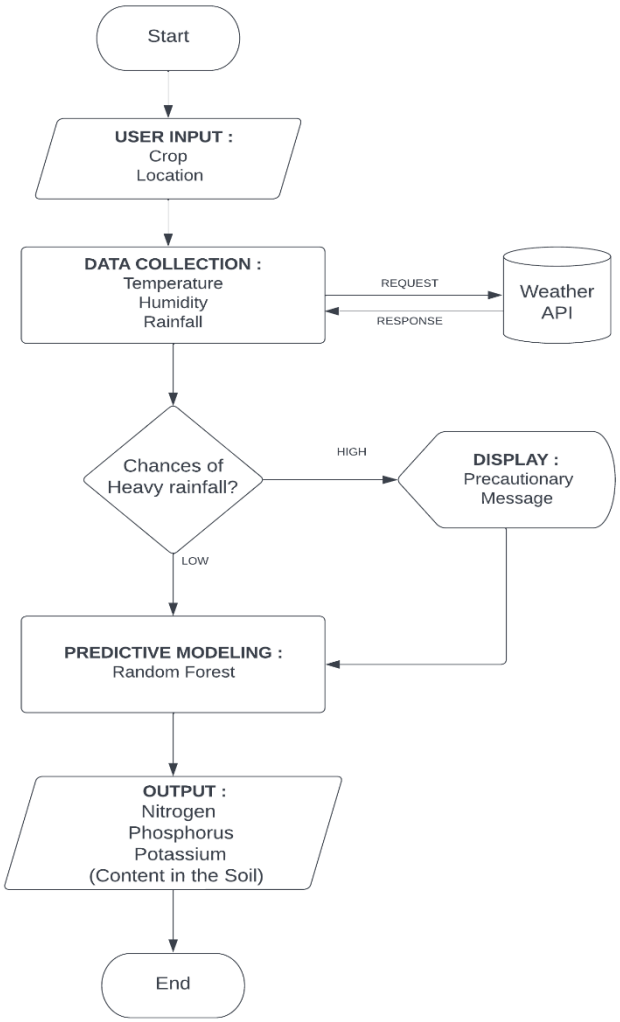

Figure 3.1: Flowchart of Proposed Model

ALGORITHM USED

Random forest (RF) is a collection of multiple decision trees that have variable hyper-parameters and are trained using varying subsets of data. In our project, we are going to take crop and location as input, and based on it, we will predict the value of N, P, and K. First, we will divide our dataset into the training and the test datasets, where training dataset is 80% of the original dataand the rest 20% is test data. Then we will create three different random forests of size 50(decision tree) for each N, P, and K and outputs the mean of the classes as the prediction of all the trees.

BEGIN:
Step 1: The dataset of size n=2200 is divided into training and test dataset (where the raining set is 80% and the test set is 20% that is training set=1,760 and the test set=240).
Step 2: Apply random forest regression to each N, P and K (Nitrogen, Phosphorus & Potassium) value with n estimators=50 (n estimators is the number of decision trees).
Step 3: Train the N label, P Label and K Label with the training dataset and dependent variable (Where the dependent variable is N for N Label, P for P Label and K for K Label).
Step 4: Each N Label, P Label and K Label generates a 50 decision tree as an output based on the training dataset.
END

Input Features

- Crop: rice, cotton, mango, orange, lentil,etc.
- Temperature: temperature measured in Celsius
- Humidity: measured relatively in percentages
- Rainfall:rainfall in mm

Output Features

- Label N: proportion of Nitrogen content in the soil
- Label P:proportion of Phosphorous content in the soil
- Label K: proportion of Potassium content in the soil

4. RESULTS ANALYSIS

Result analysis in our proposed system is an essential part of this research paper. By the analysis of results, we can compare that how much better this proposed system is performing. In result analysis we will see accuracy of different crop yield prediction that are predicted using our proposed system. We have taken datasets of 30 cases for result analysis.

Actual dataset contains eight features. All of the features are not useful for the proposed model. Therefore, a dimension reduction technique called feature selection is applied and seven features, then selected for evaluation.

	A	B	C	D	E	F	G
1	Crop	Temperature	Humidity	Rainfall	Label_N	Label_P	Label_K
2	rice	20.87974371	82.00274423	202.9355362	90	42	43
3	rice	21.77046169	80.31964408	226.6555374	85	58	41
4	rice	23.00445915	82.3207629	263.9642476	60	55	44
5	rice	26.49109635	80.15836264	242.8640342	74	35	40
6	rice	20.13017482	81.60487287	262.7173405	78	42	42
7	rice	23.05804872	83.37011772	251.0549998	69	37	42
8	rice	22.70883798	82.63941394	271.3248604	69	55	38
9	rice	20.27774362	82.89408619	241.9741949	94	53	40
10	rice	24.51588066	83.5352163	230.4462359	89	54	38
11	rice	23.22397386	83.03322691	221.2091958	68	58	38
12	rice	26.52723513	81.41753846	264.6148697	91	53	40
13	rice	23.97898217	81.45061596	250.0832336	90	46	42
14	rice	26.80079604	80.88684822	284.4364567	78	58	44
15	rice	24.01497622	82.05687182	185.2773389	93	56	36
16	rice	25.66585205	80.66385045	209.5869708	94	50	37
17	rice	24.28209415	80.30025587	231.0863347	60	48	39
18	rice	21.58711777	82.7883708	276.6552459	85	38	41
19	rice	23.79391957	80.41817957	206.2611855	91	35	39
20	rice	21.8652524	80.1923008	224.5550169	77	38	36
21	rice	23.57943626	83.58760316	291.2986618	88	35	40
22	rice	21.32504158	80.47476396	185.4974732	89	45	36
23	rice	25.15745531	83.11713476	231.3843163	76	40	43
24	rice	21.94766735	80.97384195	213.3560921	67	59	41
25	rice	21.0525355	82.67839517	233.1075816	83	41	43
26	rice	23.48381344	81.33265073	224.0581164	98	47	37
27	rice	25.0756354	80.52389148	257.0038865	66	53	41
28	rice	26.35927159	84.04403589	271.3586137	97	59	43
29	rice	24.52922681	80.54498576	260.2634026	97	50	41
30	rice	20.77576147	84.49774397	240.0810647	60	49	44

Figure -4.1: Customized Datasets For Crop Yield

Eco-Fertilization, a user-friendly system, has been implemented in the form of a website to provide cross-platform functionality and suggest appropriate timings and amount of nutrients required for an inputted crop with a heavy rainfall alert system (as shown in figures).

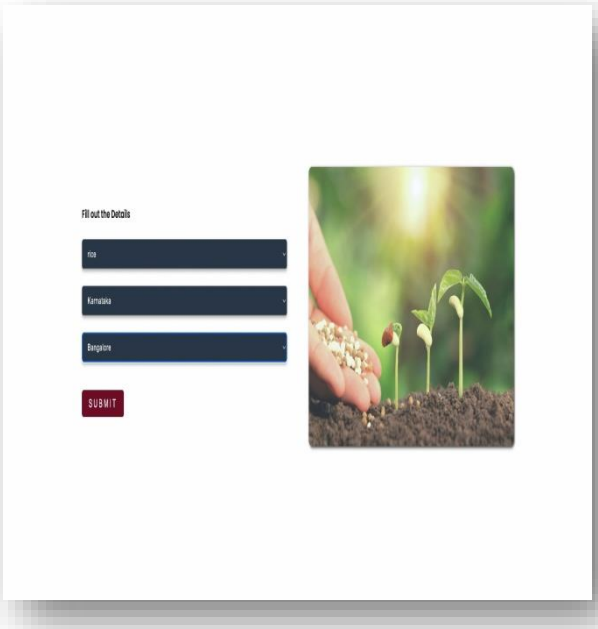

Figure-4.2: Details fill using the drop-down menu

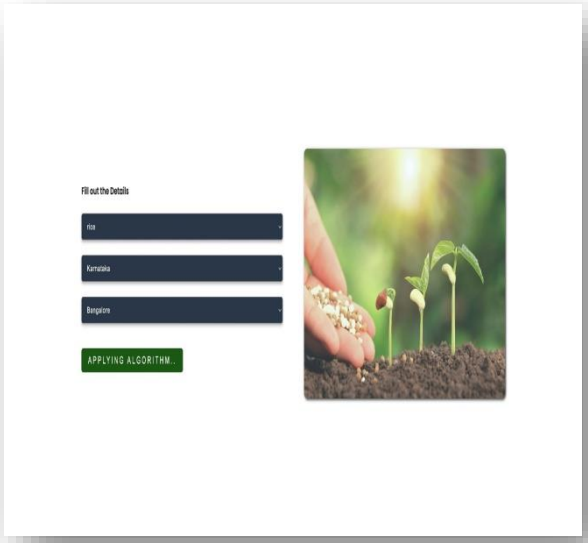

Figure-4.3: Applying Algorithm to input details

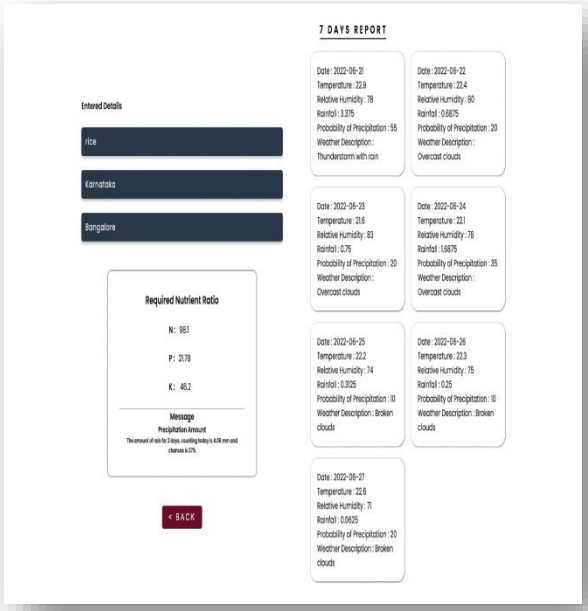

Figure-4.4: Output with seven days of weather forecasts & alerts/messages

5. CONCLUSIONS

This research aims to provide an intelligent and optimistic decision for the farm system to optimize fertilizer usage. The proposed system is able to achieve 92% of accuracy, which is quite good for any predictive model. It provides information about the use and the amount of nutrients required by the crops for satisfactory crop growth and production with respect to weather conditions. It provides weather alerts and messages. Alerts are displayed in the output of this application in case of bad weather conditions. The accuracy can be improved further with development in technologies.

REFERENCES

1] KrutikaHampannavar,Vijay Bhajantri,Shashikumar G. Totad "Prediction of Crop Fertilizer Consumption," Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),2018,PP.1-5.

[2] G. Prabakaran,D. Vaithiyanathan, Madhavi Ganesa, "Fuzzy decision support system for improving the crop productivity and efficient use of fertilizers," Computers and Electronics in Agriculture, vol-150, 2018, PP. 88-97.

[3] Shital Bhojani,Nirav Bhatt, "Data Mining Techniques for Crop Yield Prediction," Computers and Electronics in Agriculture, vol-6, 2018, PP. 357-358.

[4] Yulong Yin, Hao Ying,Huifang Zhen,Qingsong Zhang,YanfangXue,ZhenlingI, "Estimation of NPK requirements for rice production in diverse Chinese environments under optimal fertilization rate," Agricultural and Forest Meteorology, vol-279, 2019, PP. 1-6.

[5] Laura J.T. Hess,Eve-Lyn S. Hinckley,G. Philip Robertson,Pamela A. Matson, "Rainfall intensification increases nitrate leaching from tilled but not no-till cropping systems in the U.S. Midwest," Agriculture, Ecosystems & Environment, vol-290, 2020, PP. 1-10.

[6] Potnuru Sai Nishant,Pinapa Sai Venkat,Bollu Lakshmi Avinash,B. Jabber, "Crop Yield Prediction Based on Indian Agriculture using Machine Learning," 2020 International Conference for Emerging Technology (INCET), 2020, PP. 1-4

[7] Tony Yang,Kadambot H.M.,Siddique,Kui Liu, "Cropping systems in agriculture and their impact on soil health," Global Ecology and Conservation, vol-23, year, PP. 1-13.

[8] JánosKátai,ÁgnesOláhZsuposné,Magdolna Tállai,Tarek Alshaal, "Would fertilization history render the soil microbial communities and their activities more resistant to rainfall fluctuations? ," Ecotoxicology and Environmental Safety, vol-201,2020,PP.1-11.

[9] Usman Ahmed,Jerry Chun-Wei Lin, Gautam Srivastava,YoucefDjenouri, "A nutrient recommendation system for soil fertilization based on Evolutionary Computation," Computers and Electronics in Agriculture, vol-189, 2021, PP. 1-7.

[10] A.Hussein, Diogenes L. Antille, ShreevatsaKodur, GuangnanChen,Jeff N.Tullberg, "Controlled traffic farming effects on productivity of grain sorghum, rainfall and fertilizer nitrogen use efficiency," Journal of Agriculture and Food Research, vol-3, 2021, PP. 1-17.

[11] Zujiao Shi, Donghua Liu, Miao Liu, Muhammad Bilal Hafeez, Pengfei Wen, Xiaoli Wang, Rui Wang, Xudong Zhang, Jun Li, "Optimized fertilizer recommendation method for nitrate residue control in a wheat–maize double cropping system in dryland farming," Field Crops Research, vol-271, 2021, PP. 1-10.

[12] Janmejay Pant, R.P. Pant, Manoj Kumar Singh, Devesh Pratap Singh, Himanshu Pant, "Analysis of agricultural crop yield prediction using statistical techniques of machine learning," Materials Today: Proceedings, vol-46, 2021, PP. 1-10.

[13] Benny Antony, "Prediction of the production of crops with respect to rainfall.," Environmental Research, vol-202, 2021, PP. 1-5.

[14] Akash Manish Lad, K. Mani Bharathi, B. Akash Saravanan, R. Karthik, "Factors affecting agriculture and estimation of crop yield using supervised learning algorithms," Materials Today: Proceedings, 2022, PP. 1-10.

[15] Raves Akhtar, Shabbir Ahmad Sofi, "Precision agriculture using IoT data analytics and machine learning," Journal of King Saud University - Computer and Information Sciences, 2021, PP. 1-17.

[16] Saheed Garnaik, Prasanna Kumar Samant, Mitali Mandal, Tushar Ranjan Mohanty, Sanat Kumar Dwibedi, "Untangling the effect of soil quality on rice productivity under a 16-years long-term fertilizer experiment using conditional random forest," Computers and Electronics in Agriculture, vol-197,2022, PP. 1-10.

[17] RubbyAworka, LontsiSaadio Cedric, Wilfried Yves Hamilton Adoni,, "Agricultural decision system based on advanced machine learning models for yield prediction: Case of East African countries," Smart Agricultural Technology, vol-3, 2022, PP. 1-9.

[18] Senthil Kumar Swami Durai, Mary DivyaShamili, "Smart farming using Machine Learning and Deep Learning techniques," Decision Analytics Journal, vol-2, 2022, PP. 1-30.

[19] M.S. Suchithra, Maya L. Pai, "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters," Information Processing in Agriculture, vol-7, 2022, PP. 1-11.

[20] Krutika Hampannavar, Vijay Bhajantri, Shashikumar G. Totad "Prediction of fertilizer consumption can prevent the toxicity and deficiency in plants to certain extent and this can help farmers to get proper yield without much wastage" IEEE,DOI:10.1109/ICCUBEA.2018.8697827.

BIOGRAPHIES

Ashish Maurya is an undergraduate and his areas of interests are python, cloud, machine learning, natural language processing. He has done training in Python from Techsrijan consultancy services Pvt.Ltd.

Govind Yadav is an undergraduate and his areas of interests are machine learning, cloud computing, natural language processing, and data structure. He has has done training in Python.

Chandrabhan Yadav is an undergraduate and his areas of interests are machine learning, cloud computing, natural language processing, and data structure. He has has done training in Python..

.