

Farming Assistance for Soil Fertility Improvement and Crop Prediction using XGBoost

Mangesh Deshmukh^{1,*}, Amitkumar Jaiswar^{1,**}, Omkar Joshi^{1,***}, and Rajashree Shedge^{1,****}.

¹ Department of Computer Engineering, Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, India.

Abstract. In India most vital and widely practiced occupation is Agriculture and it plays a vital role in the development of our country. Soil properties, rainfall, temperature, humidity and soil pH are the factors on which agriculture is depended. In agriculture, the selection of the wrong crop may reduce crop production. Farmers should know which crops can be grown in their area. Machine Learning-based solutions are widely used in the agriculture sector. This proposed work is a recommendation system in which Machine Learning techniques are used to recommend best three crops based on soil and weather parameters. The top three crops are recommended because farmers may not have access to a particular crop if only one crop is recommended. Previous studies in this field have been done by using different Machine Learning algorithms such as Random Forest, KNN, Naïve Bayes, etc. In this proposed system XGBoost Machine Learning algorithm is used which gives better results than other algorithms. In addition, the system provides information about how to improve the soil for growing the desired crop and gives the weather forecast for next five days. As a result, this system will help farmers minimize their financial losses while also increasing crop productivity.

Keywords- Machine Learning, Crop Recommendation System, XGBoost, Weather Forecast

1 Introduction

Improving crop production is necessary in agriculture because 1.2 billion people depend on agriculture and country's 60 percent land is used for agriculture [1]. Farmers are unable to select the most suitable crops depending on soil and environmental factors, and the process of manually predicting the right crops has more often than not, resulted in failure. They only choose the crops which are generally grown in their neighborhood or trending crops. Initially studies were conducted to recommend suitable crop for sowing but in many of them only soil parameters were taken into consideration, not taking weather parameters into consideration can affect the result significantly. Some studies have used algorithm which are not generating accurate results. Also the Farmers don't have knowledge about soil nutrients which are present in the land. Applying nutrients in an inadequate amount to the soil and no proper rotation of crop may lead to reduction in soil fertility which will reduce the yield of the crop [2]. Bad weather condition also affects the crop production and farmer has to face huge loss.

The designed system was made taking the above problems into consideration. The proposed system uses Machine Learning which a game changer for agriculture sector. XGBoost machine learning algorithm is used for recommending the best three crops based on the soil and weather parameters. Also, for the farmers who don't

have data of their soil ingredients the system recommends best three crops that has the highest production in that season in their region. The designed system also gives recommendation about the soil improvements based on the parameters of soil and the desired crop that farmers want to grow. Also the system forecast weather for next five days which help farmers to take necessary precautions.

Further the paper is structured as follows: section 2 takes a glance at previous work done in crop recommendation system. In section 3 data extraction, pre-processing and model are discussed. Section 4 depicts the model accuracies and result analysis. Finally, In section 5 concluding remarks are mentioned.

2 Related Work

With more and more research been done in the field of agriculture and advanced Machine Learning, many Research and Studies were conducted over the past few years that provides valuable knowledge about the potential of modern day technologies in the agricultural sector.

Kevin Tom Thomas et.al [1] Crop Prediction work done using KNN with cross validation and has taken only soil parameters into consideration. Manhendira N. et.al [2] Crop Prediction is done by using Decision Tree algorithm and have taken both soil and weather parameters into consideration.

*email: man.des.rt18@rait.ac.in

**email: ami.jai.rt18@rait.ac.in

***email: omk.jos.rt18@rait.ac.in

****email: rajashree.shedge@rait.ac.in

Among the weather parameter rainfall is calculated by using previous year data with Support Vector Machine (SVM) algorithm.

Mansi Shinde et.al [3] Crop Recommendation is done using Random Forest algorithm and only soil parameters has been taken into consideration i.e. NPK (Nitrogen, Phosphorous and Potassium) values of the soil.

Kiran Shinde et.al [4] Recommendation system for farmers using Random Forest Algorithm and gives recommendation about fertilizer. Swapneel Chakraborty et.al [5] has implemented Crop Recommendation using Neural Network and have taken soil parameters and Environmental factors into consideration.

Sonal Jain et.al [6] Crop Selection is done based on soil parameters and weather but some important soil parameters such NPK values are not taken into consideration. A. Suresh. et.al [7] Crop yield prediction of major crops of Tamilnadu using modified KNN and K-Mean is done in this study and other crops are not taken into consideration. S.Pudumalar et.al [8] Recommendation of crop is implemented using ensemble approach by only taking only soil type into consideration. Naive Bayes, K-Nearest neighbour and random tree are combined as an ensemble approach.

This research paper [9] enumerates the significance of selecting the correct crop and factors such as government policies, market price and production rate which affects the crop selection. The paper presents a Method for selecting crop, also known as Crop Selection Method (CSM). It also helps in improving the rate of net yield of the crop. Shilpa Pande et.al [10] has implemented Crop Recommendation system using Random Forest Algorithm and taken location and soil properties as input and predict the crop yield for the desired crop. Angu Raj et.al [11] Crop Recommendation using Internet of things (IOT). The soil parameters such as temperature, humidity, soil moisture and PH are collected from the sensor using IOT and then applied Random Forest and Naïve Bayes algorithm on the data.

T. Ragunthar et.al[12] has implemented Crop Selection problem by making use of the Apriori algorithm and Decision tree induction by taking attributes as duration, water needed, budget, soil type, season sowed, profit per hectare and market price into consideration. Sridhar Mhasikar et.al [13] Survey for Predicting Agricultural Cultivation using IOT. The sensor used to get data for the values of moisture, PH and temperature of the soil and then Random Forest and Naïve Bayes algorithm was used but only soil factors are taken into consideration.

Lakshmi N. et.al [14] proposed a crop recommendation system considering factors such as drainage, texture, color, depth, soil erosion, soil pH, permeability and water holding using big data. M.V.R Vivek et.al [15] Based on wind, precipitation, and temperature, I did a complete survey on the use of several machine learning algorithms for crop recommendation, including Naive Bayes, multi-layer perceptron, JRIP, Jf48 and SVM. Hao Zhang et.al [16] Meteorological data, crop data and soil data are used to discuss the implementation and design of crop recommendation and fertilizer recommendation systems.

Following Table 2.1 shows the limitations of reviewed work.

Table 2.1. Limitations of reviewed work

Authors	Technologies Used	Limitations
Kevin Tom Thomas et.al [1]	<ul style="list-style-type: none"> •KNN with cross validation, KNN, Decision Tree and Naïve Bayes algorithm are used. •Parameters: Nitrogen, Phosphorous, Potassium and soil pH value. 	No Environmental factors are used for recommending suitable crop.
Mansi Shinde et.al [3]	<ul style="list-style-type: none"> •Random Forest, Naïve Bayes and ID3 machine learning algorithm are used. •Parameters: Nitrogen, Phosphorous and Potassium. 	Only soil parameters are taken into consideration. Weather parameters are not taken into consideration.
Sonal Jain et.al [6]	<ul style="list-style-type: none"> •Random Forest algorithm is used. •Parameters: Temperature, Rainfall, soil type, soil pH, water capacity and soil fertility. 	Soil Ingredients like Nitrogen, Phosphorous and Potassium are not taken into consideration.
A. Suresh et.al [7]	<ul style="list-style-type: none"> •KNN algorithm is used. •Parameters: Rainfall, groundwater, soil type and cultivated area. 	Crop recommendation is done only for major crops i.e. Rice, Sugarcane, Maize, Raggi and Tapioca. Other crops are not taken into consideration.

3 Proposed System

The backbone of country's economy is agriculture. It is the important source of income. Soil type, PH value, rainfall, environment factors are major factors on which agriculture is dependent. Selection of wrong crop may lead to reduction of crop production and farmer may have to bear huge loss. If it is possible to get the right crop before sowing, it can be very helpful for farmers to increase crop production and to make the right decisions about storage and the business side. Farmer is not able

to grow the crop that he desires as the soil may lack fertility or some ingredients required for the desired crop. With proper recommendation of improving soil fertility, the farmer can grow the crop that he desires. Also, bad weather condition may affect the crop production on large scale so weather forecast of next five days will help farmer to take necessary precautions.

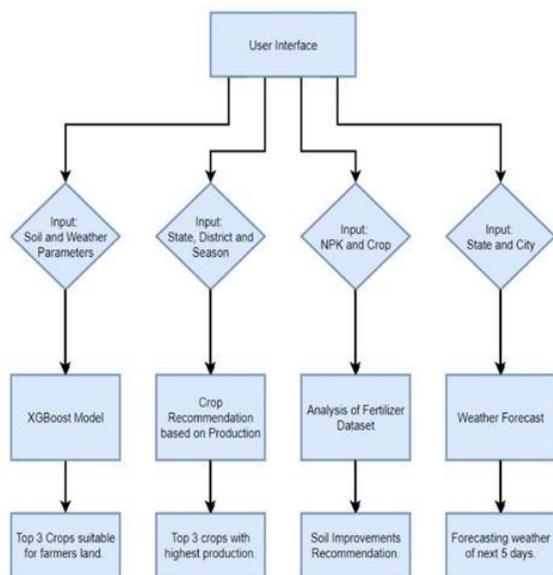


Figure 3.1 Crop Recommendation Model

3.1 Crop Prediction

After the dataset is processed and split into training and testing dataset. Different Machine Learning algorithms were used such as Decision Tree [17], Naïve Bayes[18], SVM [19], Linear Regression [20], Random Forest[21] and XGBoost [22]. From these algorithm XGBoost gives us the best accuracy.

XGBoost:

It is a gradient boosting algorithm which enhances performance and speed of a tree based (decision trees) machine learning model. In the first iteration, XGBoost will have less accuracy as initially the model will be basic. But with increase in iterations the model will optimise the loss function using Gradient Descent technique. This process is repeated until model reaches a threshold i.e. loss can't be optimised further. Hence the model's accuracy will improve with increase in number of iterations.

In the proposed system the train size is kept to 0.8 i.e. 80% of the data in dataset will be used for training purpose. Later the trained model will be tested on 20% of remaining test data. Then the XGBoost model is trained using input parameters like Nitrogen, Phosphorous, Potassium, pH, rainfall, temperature and humidity. This values will be given as input by the farmer. Then using predict_proba() function we find the probability of each of the output class. Then top three values with maximum probability is selected and then we get the corresponding crop for this given probabilities. Finally these three crops are recommended as they are more suitable to grow for

given input features.

3.2 Crop Recommendation based on Production

Some Farmers may not have the data of Ingredients present in their soil so for them the system will take input as State, District and Season. After analyzing the input and dataset the system will recommend top three crops which have the highest production in the given State, District and Season.

3.3 Soil Improvements Recommendation

Fertility of the soil depends upon its NPK value present in the soil. The proposed system will take the input as soil properties i.e. Nitrogen, Potassium, Phosphorus and the crop that farmer desire to grow.

The System then analyze the input and dataset and gives recommendation about what can be added in the soil to improve its fertility to grow the desired crop.

3.4 Weather Forecast

Climate forecasting and prediction include gathering and communicating data about future weather conditions based on weather predictions.

The farmer has to give its state and city as input then the system will fetch the weather by using Weather API [23] this will help farmer to take necessary precaution and can avoid huge loss.

4 RESULTS AND DISCUSSION

This section contains implementation and result of all the features used in the proposed system.

To train and test any Machine Learning model, you'll need a dataset. The dataset for the proposed crop recommendation system was obtained from Kaggle [24], a well-known website. The data is collected from soil testing lab, it contains the soil specific attributes. Different crops like maize, rice, coffee, etc are considered in this system. The attributes considered different nutrients like nitrogen, potassium, phosphorus and pH value of the soil in the dataset. It also contains weather factors such as temperature, humidity and rainfall as its parameters. The other dataset that is used in this system is also extracted from Kaggle [25] and has attributes such as Season, State name, Area, District name and Production.

Database collected from open source should be cleaned. Noisy, incomplete and incompatible data cannot be processed by Machine Learning Algorithm so they must be cleaned.

4.1 Crop Prediction

After preprocessing of dataset then the data set is split into 80% training data set and 20% testing data set. Different Machine Learning Algorithms such as Decision Tree, Naïve Bayes, SVM, Logistic Regression, Random Forest and XGBoost were used on the dataset. After calculating the accuracies of the model, the accuracy of Decision Tree is 90%, Naïve Bayes is 99%, SVM is 10%, Logistic Regression is 95.22%, Random Forest is 99% and XGBoost is 99.3%. The same is shown in the following Figure 4.1.

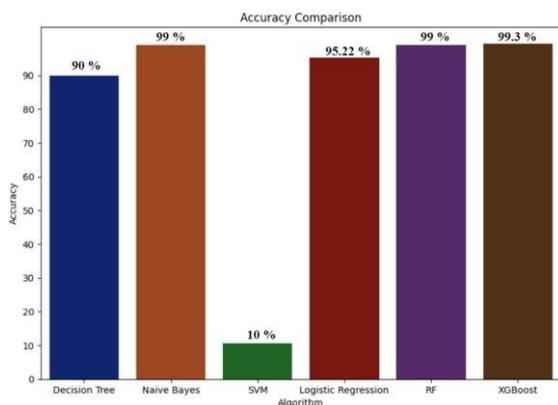


Figure 4.1 Comparing Accuracy of different Models

The proposed system uses the XGBoost algorithm to predict the best top three crops for farmers based on soil and weather parameters as it has the highest accuracy.

Table 4.1 Tabular Comparison of the Classifiers (in %)

Algorithm	Accuracy (%)	Precision	Recall	F1-Score
Decision Tree	90	84.29	88.16	85.36
Naïve Bayes	99	99.43	98.86	99.04
SVM	10	66.06	12.52	13.93
Logistic Regression	95.22	95.40	94.86	95.04
Random Forest	99	99.29	98.90	99.03
XGBoost	99.3	99.34	99.44	99.37

The Columns of Table 4.1 are defined as follows.

- **Classifiers:** Classifiers are models that are used to train and test data.
- **Accuracy:** The percentage of times the algorithm successfully classifies a data point.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad \text{eq. 4.1}$$

Where,

- TP = True Positive
- FP = False Positive
- TN = True Negative
- FN = False Negative

- **Precision:** The number of accurately predicted positive observations divided by the total number of predicted positive observations.

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{eq. 4.2}$$

- **Recall:** The percentage of accurately anticipated positive observations to the total number of observations in a class is known as recall.

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{eq. 4.3}$$

- **F1 Score:** The F1 Score is the weighted average of Precision and Recall.

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{eq. 4.4}$$

4.2 Crop Recommendation based on Production

The Proposed system takes inputs as State, District and Season as shown in the following Table 4.2.

Table 4.2 Input for Crop Recommendation based on Production

Sr. No.	State	District	Season
1	Maharashtra	Dhule	Kharif
2	Gujarat	Surat	All Season

After taking the input the system then recommends the best three crops which has the highest crop production in the given season in that region as shown in the Table 4.3.

Table 4.3 Output for Crop Recommendation based on Production

Sr. No.	Crop	Production(t/ha)	Area(ha)
1	Cotton	2386223	1658151
	Sugarcane	2355044	43800
	Bajra	1962191	2193245
2	Sugarcane	126331300	1778100
	Banana	2448600	36300
	Rice	149100	81900

4.3 Soil Fertility Improvement

The Farmer gives the input such as the desired crop he wants to grow and NPK of the soil. After analyzing the dataset the system will tell which ingredient is low or high in the soil and recommend improvements in soil to make it suitable for the desired crop.

For example, the desired NPK value for Rice are 80-40-40 respectively. If the soil with composition of NPK values is 80-80-40 i.e. the phosphorous value is high, then the system will tell that the phosphorous value of the soil is high and will recommend improvements as follows:

- Avoid using manure since it includes a lot of important nutrients that your soil needs, but it also contains a lot of phosphorus. Reducing the amount of manure applied will assist to limit the amount of phosphorus added.
- Usage of phosphorus free fertilizer can help you to control the quantity of phosphorous that is been added to your soil. This will help the soil as plants will absorb the phosphorous that is already present in the soil while still supplying other essential nutrients like Potassium and Nitrogen
- Nitrogen fixing vegetables like peas and beans can be planted to boost nitrogen without increasing phosphorous.
- Water the soil - letting your soil soak the water will help in the removal of phosphorous from the soil. This is a last ditch effort.

4.4 Weather Forecast

Farmer give input as State and City then the system will forecast weather for next five days which includes parameters as temperature, minimum temperature, maximum temperature, pressure, sea level, ground level, humidity, weather description, wind speed, wind gust, visibility, date and time. The following is an example of the same.

Table 4.4 Sample result of weather forecast

Parameters	Date (2022-02-10 21:00:00)
Temperature	14.52
Temperature Minimum	11.55
Temperature Maximum	14.52
Pressure	1016
Humidity	76
Weather	Clear
Description	Clear Sky
Wind Speed	3.55
Wind Degree	294
Wind Gust	5.61
Visibility	10000

In the above Table 4.4 we have shown the sample result of weather forecast for one day. Similarly weather forecast of next five days will be done.

5 Conclusion

Farmers can choose the most profitable crop with the help of this system. It helps the farmer to make better decision on the storage and business side if they know the best crop before sowing. In this proposed system most widely used techniques, Random Forest, SVM, Decision Tree, Logistic Regression, Naive Bayes and XGBoost are compared to choose the optimal classifier for recommending best three crops. After the comparison, it can be seen that XGBoost algorithm has the highest accuracy of 99.3%. Also soil improvement recommendation was done by the system so that the soil is suitable to grow the crop that farmer desires. The proposed system provides weather forecast of next five days so that farmer can take necessary precaution. Thus the result shows that there is significant potential to develop the agricultural sector using machine learning based techniques in the most of the situations.

References

1. Kevin Tom Thomas, Varsha S, Merin Mary Saji, Lisha Varghese and Er. Jinu Thoma, "Crop Prediction Using Machine Learning", International Journal of Future Generation Communication and Networking, Vol. 13, Issue: 3, pp. 1896-1901, 2020
2. Mahendra N , Dhanush Vishwakarma , Nischitha K , Ashwini, and Manjuraju M. R, "Crop Prediction using Machine Learning Approaches", INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 08 (August 2020).
3. Mansi Shinde, Kimaya Ekbote, Sonali Ghorpade, Sanket Pawar and Shubhada Mone, "Crop Recommendation and Fertilizer Purchase System", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (2), 665-667, 2016.
4. Kiran Shinde, Jerrin Andrei and Amey Oke, "Web-Based Recommendation System for Farmers", International Journal on Recent and Innovation

Trends in Computing and Communication ISSN: 2321-8169, Volume: 3 Issue: 3 1444 – 1448.

5. P. A, S. Chakraborty, A. Kumar and O. R. Pooniwala, "Intelligent Crop Recommendation System using Machine Learning," 5th International Conference on Computing Methodologies and Communication (ICCMC), , pp. 843-848, doi: 10.1109/ICCMC51019.2021.9418375, 2021.
6. S. Jain and D. Ramesh, "Machine Learning convergence for weather based crop selection," IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-6, doi: 10.1109/SCEECS48394.2020.75, 2020.
7. Suresh, P. Ganesh Kumar and M. Ramalatha, "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN," 3rd International Conference on Communication and Electronics Systems (ICES), pp. 88-93, doi: 10.1109/CESYS.2018.8723956, 2018.
8. S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, "Crop recommendation system for precision agriculture," Eighth International conference on Advanced Computing (ICoAC), pp. 32-36, doi: 10.1109/ICoAC.2017.7951740, 2017.
9. R. Kumar, M. P. Singh, P. Kumar and J. P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique," International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), pp. 138-145, doi: 10.1109/ICSTM.2015.7225403, 2015.
10. S. M. PANDE, P. K. RAMESH, A. ANMOL, B. R. AISHWARYA, K. ROHILLA and K. SHAURYA, "Crop Recommender System Using Machine Learning Approach," 5th International Conference on Computing Methodologies and Communication (ICCMC), , pp. 1066-1071, doi: 10.1109/ICCMC51019.2021.9418351, 2021
11. Raj, Anju & Balashanmugam, Dr.Thiyaneswaran & Jayanthi, J & Yoganathan, N & Srinivasan, PCrop. . Recommendation on Analyzing Soil Using Machine Learning. Turkish Journal of Computer and Mathematics Education (TURCOMAT). 12. 1784-1791, 2021.
12. T. Rangunthar, S. Selvakumar and G. Ilamurugan, " Counsel System for Effective Farming using Data Mining Algorithm", International Journal of Pure And Applied Mathematics, Volume 117 No. 21, 921-924 ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (online version), 2017.
13. Shridhar Mhaiskar, Chinmay Patil, Piyush Wadhai, Aniket Patil and Vaishali Deshmukh, "A Survey on Predicting Suitable Crops for Cultivation Using IoT ", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization), Website: www.ijirce.com, Vol. 5, Issue 1, January 2017.
14. Lakshmi N, Priya M, Mrs. Sahana Shetty and Mr. Manjunath C.R, "Crop Recommendation System for Precision Agriculture ", International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653; IC Value: 45.98; SJImpact Factor: 6.887, Volume 6

Issue V, May 2018.

15. Lakshmi. N, Priya. M, Mrs. Sahana Shetty and Mr. Manjunath C.R , "Crop Recommendation System for Precision Agriculture ", International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653; IC Value: 45.98; SJImpact Factor: 6.887, Volume 6 Issue V, May 2018.
16. M.V.R. Vivek, D.V.V.S.S. Sri Harsha and P. Sardar Maran, "A Survey on Crop Recommendation Using Machine Learning", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-7, Issue-5C, February 2019.
17. Zhang H. et al. Design and Implementation of Crop Recommendation Fertilization Decision System Based on WEBGIS at Village Scale. In: Li D., Liu Y., Chen Y. (eds) Computer and Computing Technologies in Agriculture IV. CCTA 2010. IFIP Advances in Information and Communication Technology, vol 345. Springer, Berlin, Heidelberg, 2011.
18. Patel, Harsh & Prajapati, Purvi. "Study and Analysis of Decision Tree Based Classification Algorithms". International Journal of Computer Sciences and Engineering. 6. 74-78. 10.26438/ijcse/v6i10.7478, 2018.
19. Kaviani, Pouria & Dhotre, Sunita. "Short Survey on Naive Bayes Algorithm". International Journal of Advance Research in Computer Science and Management. 04, 2017.
20. Evgeniou, Theodoros & Pontil, Massimiliano. Support Vector Machines: Theory and Applications. 2049. 249-257. 10.1007/3-540-44673-7_12, 2001.
21. Rong, Shen & Bao-wen, Zhang. "The research of regression model in machine learning field". MATEC Web of Conferences. 176. 01033. 10.1051/mateconf/201817601033, 2018.
22. Schonlau, Matthias, and Rosie Yuyan Zou. "The Random Forest Algorithm for Statistical Learning." The Stata Journal, vol. 20, no. 1, pp. 3–29, doi:10.1177/1536867X20909688, 2018.
23. Santhanam, Ramraj & Uzir, Nishant & Raman, Sunil & Banerjee, Shatadeep. Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets, 2017.
24. Weather Forecast API- <https://openweathermap.org/api>
25. <https://www.kaggle.com/atharvaingle/crop-recommendation-dataset>
26. <https://www.kaggle.com/abhinand05/crop-production-in-india>