Farming Assistance for Soil Fertility Improvement and Crop Prediction using XGBoost

Mangesh Deshmukh1,*, Amitkumar Jaiswar1,**, Omkar Joshi1,***, and Rajashree Shedge1,****.

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. Manhendra 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.


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 on the algorithm.

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, J48 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>
<thead>
<tr>
<th>Authors</th>
<th>Technologies Used</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kevin Tom Thomas et.al</td>
<td>KNN with cross validation, KNN, Decision Tree and Naïve Bayes algorithm are used.</td>
<td>No Environmental factors are used for recommending suitable crop.</td>
</tr>
<tr>
<td>Sonal Jain et.al [6]</td>
<td>Random Forest, Naïve Bayes and ID3 machine learning algorithm are used.</td>
<td>Only soil parameters are taken into consideration.</td>
</tr>
<tr>
<td>A. Suresh et.al [7]</td>
<td>KNN algorithm is used.</td>
<td>Soil Ingredients like Nitrogen, Phosphorous and Potassium are not taken into consideration.</td>
</tr>
</tbody>
</table>

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](image)

### 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 the 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.
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>
<thead>
<tr>
<th>Sr. No.</th>
<th>State</th>
<th>District</th>
<th>Season</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Maharashtra</td>
<td>Dhule</td>
<td>Kharif</td>
</tr>
<tr>
<td>2</td>
<td>Gujarat</td>
<td>Surat</td>
<td>All Season</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Sr. No.</th>
<th>Crop</th>
<th>Production(t/ha)</th>
<th>Area(ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cotton</td>
<td>2386223</td>
<td>1658151</td>
</tr>
<tr>
<td>2</td>
<td>Sugarcane</td>
<td>2355044</td>
<td>43800</td>
</tr>
<tr>
<td>2</td>
<td>Bajra</td>
<td>1962191</td>
<td>2193245</td>
</tr>
</tbody>
</table>

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:

a) 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.

b) 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.

c) Nitrogen fixing vegetables like peas and beans can be planted to boost nitrogen without increasing phosphorous.

d) 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

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Date (2022-02-10 21:00:00)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>14.52</td>
</tr>
<tr>
<td>Temperature Minimum</td>
<td>11.55</td>
</tr>
<tr>
<td>Temperature Maximum</td>
<td>14.52</td>
</tr>
<tr>
<td>Pressure</td>
<td>1016</td>
</tr>
<tr>
<td>Humidity</td>
<td>76</td>
</tr>
<tr>
<td>Weather</td>
<td>Clear</td>
</tr>
<tr>
<td>Description</td>
<td>Clear Sky</td>
</tr>
<tr>
<td>Wind Speed</td>
<td>3.55</td>
</tr>
<tr>
<td>Wind Degree</td>
<td>294</td>
</tr>
<tr>
<td>Wind Gust</td>
<td>5.61</td>
</tr>
<tr>
<td>Visibility</td>
<td>10000</td>
</tr>
</tbody>
</table>

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 been 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


