Review Article

Crop Prediction Methods-A Comparative Study

Abstract

In India 58% of the population depend upon agriculture and agriculture contributes about 18% of the gross domestic product. But the farmers face numerous challenges in agriculture due to various reasons. The farmers' crop selection criteria generally depended upon their traditional base and testimonials of the success stories. The failure of selecting appropriate crop for a particular area makes them vulnerable to other vagaries also. Scientifically crop selection depends upon various factors like soil parameter (Ex: Soil moisture), weather parameters (Ex: Rainfall, temperature, Humidity), Geographical parameters (Ex: Slope) etc. Farmers and those who are venturing into farming requires a crop prediction model or a technique for the best crop selection decision in order to cope up and tide over exigencies like climate change, flood, drought etc which occur globally. The model will be accurate to tackle the issues arised by the climate change and also gives reasonable results to the farmers. This paper review on the existing crop prediction model and also their implication on the farming community.

Keywords : crop prediction,soil parameter,topography,machine learning,big data analytics

1. Introduction

The suitability of a crop for a given land and its resultant yield depends upon various parameters like soil, climate and geography. This signifies that crop selection is very relevant in this era of climatic changes. It will be beneficial for the farmers to tide over the climate change exigencies and to those who wish to start farming as their profession and don't have much traditional knowledge on farming practices. Some major features influencing crop selection for a specific area are:

1. Soil factors
    a. Soil Type
    b. Soil fertility level
    c. Soil reactions like acidity and alkalinity
    d. Problem soils like salinity and acidity presence in the area

2. Climatic factors
    a. Rainfall pattern
    b. Average Temperature
    c. Relative Humidity
    d. Availability of solar radiation
    e. Wind
3. Land topography, elevation and altitude
4. Water availability and source of water
5. Specific pest/diseases prevailing in the area
6. Marketing and transportation facilities
7. Labour availability in the area

All these factors together decide the performance and yield of a crop and income of the farmers. Hence while predicting the suitability of a particular crop/crop variety, all these factors should be considered directly or indirectly. This paper reviews various crop selection models in agriculture.

Section 2 describes the various methods available for crop prediction. Most of the work is done using machine learning, where crop data is collected and identified the factors/attributes influencing the yield. This data is split into training data and testing data set. Using a classification algorithm, a model is created with training and this model is tested using testing data. The accuracy of data set and feature selection decides the accuracy of the system. Section 2 also contains the methodologies which help to identify the influencing features/attributes. Section 3 analyses various methods discussed in section 2.

2. Literature Survey

Existing crop recommendation methods are discussed in this section. A crop sequencing technique also considered, as it returns sequence of crop which produces maximum yield over a given season.

Suruliandia et. al. (2021) proposed a system which predicts crops based on environmental and soil characteristics[1]. The dataset has 1000 instances and 16 attributes, in which 4 attributes are environmental characteristics and 12 attributes are soil characteristics. Target class has multiclass representation with 9 classes. The collected data set is preprocessed and essential attributes were extracted from the preprocessed data by using the wrapper feature selection techniques. Then they applied classification algorithm on the mined attributes to find the suitable crop for cultivation in a specific area. The algorithms used for feature selection are: Boruta, Recursive Feature Elimination (RFE) and Sequential Forward Feature Selection (SFFS). Algorithms used for Classification are: k-Nearest Neighbor, Naive Bayes, Decision Tree, Random Forest, Bagging and SVM. From the analysis, they found that RFE (for feature Selection) with bagging (for classification) has the best crop prediction accuracy based on the considered environment and soil characteristics.

RISHI GUPTA et. al. (2021) proposed a method using big data analytics to predict the top three crops which provides the best yield for a particular season and top 3 crops with the best yield throughout the year[9]. The required data was collected from a university website (The University of Dayton) and Kaggle. This Data contains the daily average temperature (1995 to 2020), monthly rainfall (1901 to 2015), seed, soil, wind speed and humidity data. It also contains the data of 125 crops along with the area it was sown and their production(2000 to 2014) for six seasons and the whole year. Then they used MapReduce and K-means clustering for crop recommendation. Incase weather related data set, year and region is taken as the key, and the corresponding parameter for all months has taken as the value for Map Function. In Reduce function, these parameters were calculated and assigned to a season. In case of Map function of the crop dataset, the year, crop, season and region were taken as the key and the area and production has taken as the value. After that produce per area of each row of data is calculated as part of reduce function. After that they combined all reduced datasets to form a super dataset and done classification with this. The inputs of the proposed system were month, state and region. System will assign the respective season/s, based on the month selected by the user. Then system returns the crops that give the best yield for a particular season and crops which give the best yield throughout the entire year by passing through the dataset with state and region selected by the user.

Nischitha K et. al. (2020) has suggested a system which recommends the suitable crop for a given land[3]. It will also recommend the required seed for cultivation and the amount of nutrients required for the crop. In addition to this, system will display the approximated yield as well. The main attribute considered for crop prediction are: Soil pH, Temperature, Humidity, Rainfall and Crop data, where the rainfall value was predicted based on the previous year rainfall data. For rainfall prediction, they used SVM classifier and Decision tree algorithm is used to predict suitable crop. The data required for crop prediction and rainfall prediction was collected from Government website, V C Farm Mandya and weather department.

G. Mariammal et. al. (2021) proposed a feature selection technique called as modified recursive feature elimination (MRFE), which helps to select relevant parameters from the collected data set for predicting the suitable crops [4]. MRFE technique selects and ranks salient features using ranking method. The analysis results showed that the MRFE technique is best for feature selection and bagging is the accurate method for predicting suitable crops. The environmental dataset is collected from Tamilnadu Agricultural University website and soil characteristics are collected from Sankarankovil Taluk, Tenkasi, India. The data set includes 1000 instances with 9 classes and 16 attributes, where 12 attribute are soil characteristics and the 4 are environmental characteristics. In this paper they have done analysis using different existing wrapper feature selection techniques such as Boruta, SFFS, RFE and proposed MRFE technique.

M.Kalimuthu et. al.(2020) developed a system which will predict the suitable crop for the given input parameter using machine learning[5]. In this method they considered temperature, relative humidity, moisture content, pH & rainfall as the

features, where preceding data and present data of humidity of air, temperature, rainfall and moisture of the soil are collected. Preceding data is collected from a weather station and present data is collected using designed Arduino setup. Then they performed the consolidation of preceding and present data to check accuracy between them. The Seed data is collected seed data contains the crops and its corresponding parameter values to have a healthy growth. Then seed data and feature set fed into Naive Bayes classifier, to predict seeds based on various data (Temperature, Humidity, pH, Rainfall and Moisture). The accuracy of Naïve Bayes method was 97%, and boosting method is used to improve the accuracy.

Aruvansh Nigam et. al.(2019) proposed a method for crop yield and crop name prediction on the basis of Temperature, Rainfall, Area and Season. 100 years of Temperature and Rainfall dataset is arranged according to the month and the districts in India [6]. 17 years of production dataset is arranged according to the crops, districts, area and season. All the dataset were collected from official website of the Indian government. Then temperature and rainfall dataset recategorized on the basis of seasons and then appended with the production dataset. They observed that Sequential model (Simple Recurrent Neural Network) performs better for rainfall prediction and LSTM (long short-term memory) for temperature prediction. Predicted Rainfall and Temperature are then used in crop name prediction and crop yield prediction. Crop yield data are grouped season wise and it is used as an input for predicting suitable crop. During analysis they observed that Random Forest Regression has the highest yield prediction accuracy with the above data set.

Pavan Patil et. al.(2020) has suggested an efficient crop recommendation system using classifier models[7]. They considered Soil Parameters(Soil pH value, Soil Type), Climatic Parameters (Temperature, Humidity, Rainfall, Wind) and production parameter(Cost of cultivation, Previous year yield details for that region) as the features. Environmental data is collected from Indian Meteorological Department. Classification is done using Decision tree classifier. In addition to this yield is predicted based on attributes like crop name, cost of irrigation, cost of cultivation, cost of production. Prediction is done using Decision Tree Classifier, KNN and naïve bias. 76.8% accuracy is obtained when tested with decision tree and 89.4% accuracy is obtained for KNN.

Sayed Mazhar Ali et. al.(2021) proposed system which predicts the crop based on temperature[8]. Dataset consists of mostly cultivated crop such as wheat, sugarcane, rice and cotton. They used data collected from Nawabshah area and Polynomial Regression is used for prediction. The dependence of crop production on temperature is also examined in this paper and concluded that maximum wheat crop production will be obtained in 12°C to 22°C, Maximum cotton crop yield will be at temperature 25-35°C, maximum Sugarcane production is achieved at 20°C to 32°C and maximum rice yield is at 30°C to 45°C.

FAN-HSUN TSENG et. al. (2019) suggested a method to check whether a particular crop is suitable or not[2]. For this, they established IoT sensors in farms (which can help detect temperature, soil EC, humidity, soil moisture content, atmospheric pressure, salinity and soil illumination) to observe the farm environment. The data generated by these IoT sensors send out to the server (XML format). Then data from all sensors exported to perform analysis. Before analysis they did data cleaning and normalization of the exported data. 3D correlation analysis also has been done to examine the cultivation techniques. Then they classified various features as necessary and other condition. Necessary conditions are soil electrical conductivity, temperature, and soil salinity features. Other conditions include soil moisture content, air humidity, atmospheric pressure, and illumination (these values can be improved using manual intervention). After this, system determines whether a crop is suitable or not by using the below equation:

$$\mathbb{CRE}_i = \begin{Bmatrix} true, & \mathbb{G}'_{i,\max} < \mathcal{CR}_{i,\max} \text{ and } \mathbb{G}'_{i,\min} > \mathcal{CR}_{i,\min} \\ flase, & otherwise \end{Bmatrix}$$

$G_{i,max}$ is maximum sensor data value, $G_{i;min}$ is minimum sensor data value, $CR_{i;min}$ is minimum crop environment value, and $CR_{i;max}$ is maximum crop environment value. In the result if 'other conditions' is found unsuitable, it provides the assessment result to the farmer. Based on these result, farmers can decide whether to proceed with the cultivation or not. Normal distribution is used to perform risk analysis. They tested this system in celery, green beans, water spinach and daikon.

Rakesh Kumar et. al.(2015) has suggested an approach of crop sequencing technique for improving net yield rate of crop over season[10]. The Crops can be Seasonal crops, Whole year crops, Short time plantation crops (crops which takes short time for growing), Long time plantation crops(Crops takes long time for growing). So they created a crop sowing table with the below attributes, data are collected from farmers of Patna, Bihar (India).

a. Crop name
b. Sowing period
c. Harvesting period
d. Growing day
e. Predicted yield

Then used below rules to identify the sequence of crops which will maximize the planting in net yield of crops over a season

$$\text{cropSelector(presentTime)} = \begin{cases} 0 \quad ; \text{if presentTime} \geq \text{End of season} \\ \text{cropSelector(presentTime+1)} \ ; \text{if presentTime} \neq \text{sowingTime} \\ \text{crop} \leftarrow \max_{\text{crop} \in \text{cropSowingTable}} \{\text{crop} \rightarrow \text{productionRate}/\text{crop} \rightarrow \text{plantingDay}\} \\ \quad + \text{cropSelector(presentTime+ crop} \rightarrow \text{plantingDay)}; \\ \qquad\qquad\qquad\qquad \text{if presentTime} = \text{sowingTime} \end{cases}$$

3. Performance Analysis

Since the accuracy of crop machine learning based crop recommendation depends on the selected features and classification algorithm, different method discussed in section 2 can be summarized as shown in the table 1

Table 1 . Performance Analysis

Author	Features Considered	Classification algorithm
Suruliandia et. al.[1]	Soil Characteristics (Electrical Conductivity, Ph, Organic Carbon, Phosphorus, Nitrogen, Potassium, Sulphur, Boron, , Zinc, Iron, Manganese, Copper) Environmental Factors(Season, Texture, Average Temperature ,Rainfall)	RFE(For feature selection) Bagging classifier(For Classification)
Nischitha K et. al.[2]	Soil pH, Temperature, Humidity, Rainfall	Decision tree algorithm
M.Kalimuthu et. al.[5]	Temp, Humidity, Moisture, pH & rainfall	Naive Bayes classifier
Aruvansh et. al.[6]	Temperature, Rainfall , Area , Season	Random Forest Regression
Pavan Patil et. al.[7]	Climatic Parameters (Temperature, Humidity, Rainfall, Wind), Soil Parameters(Soil pH value, Soil Type)	Combination of classification algorithms like naïve bayes and decision tree classifier are better performing
Sayed Mazhar Ali et. al.[8]	Temperature	Polynomial Regression
Rishi Gupta et. al.[9]	rainfall, temperature, crop production, wind speed, humidity, soil type	MapReduce framework
G. Mariammal et. al.[4]	Soil Characteristics (Electrical Conductivity, Ph, Organic Carbon, Phosphorus, Nitrogen, Potassium, Sulphur, Boron, , Zinc, Iron, Manganese, Copper) Environmental Factors(Season, Texture, Average Temperature ,Rainfall)	MRFE(For feature selection) Bagging classifier(For Classification)

Also, **CSM Technique** suggested a crop sequencing technique for improving the net yield of crop over season. They considered Sowing time, Harvesting time, Growing period , Predicted yield (kg/hectare) for the suitable crop sequence selection. FAN-HSUN TSENG et. al considered temperature, soil moisture content, air humidity, soil EC, atmospheric pressure, soil salinity and illumination, to check whether a particular crop is suitable for the land or not. Crop analysis is done in green beans, celery, daikon and water spinach,

From the above comparison, it is inferred that:

- Crop Prediction mainly depends on various Soil characteristics (Soil Type, Soil reaction etc.), Climatic Parameters (Humidity, Wind, Temperature, Rainfall etc.) and geographical parameters.
- It will be more effective if we include season wise prediction and crop sequencing as well so that farmers can have better farm planning. Also crop varieties suitability can be considered as part of data set.
- More than one crop can be return as the result of prediction, this will help farmer to choose a better one for him based on his knowledge and labour availability.

Some of the factors which are not considered in the above research work are:

- Inter cropping is not considered, which is a practice of cultivation two or more crops simultaneously in the same field. So that farmers can have greater productivity of the farm land available.
- Yield based prediction will only results the crop which are part of production data set. This will restrict the crop name returned by the prediction algorithm.
- Specific pest and disease prevailing in the area, topology and elevation, availability of sunlight and water availability are not considered in the existing works

4. Conclusion

Different research works are available which predict suitable crop for an area, yield per area, crop name etc. along with sequencing technique. The major features used for crop prediction are soil and climatic parameters. Of this climatic factors can be predicted using past data. Recursive Feature Elimination (RFE) and MRFE can be used to select the major influencing features from data set. As a next step classification algorithm such as bagging classifier, Decision tree algorithm, and Random Forest Regression, Naive Bayes etc. or MapReduce framework can be used for predicting the crop name. Including crop sequencing technique along with crop prediction will help farmers to have better accuracy in decision making.

References

[1] A. Suruliandi, G. Mariammal & S.P. Raja, "Crop prediction based on soil and environmental characteristics using feature selection techniques" .Mathematical and Computer Modelling of Dynamical Systems, 27:1, 117-140, DOI: 10.1080/13873954.2021.1882505 (2021)

[2] FAN-HSUN TSENG , HSIN-HUNG CHO, HSIN-TE WU, "Applying Big Data for Intelligent Agriculture-Based Crop Selection Analysis" IEEE ACCESS, Special Section On Data Mining For Internet Of Things, August 15, 2019,

[3] Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini Manjuraju M.R, "Crop Prediction using Machine Learning Approaches", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 9 Issue 08, August-2020

[4] G. Mariammal, A. Suruliandi , S. P. Raja , and E. Poongothai, "Prediction of Land Suitability for Crop Cultivation Based on Soil and Environmental Characteristics Using Modified Recursive Feature Elimination Technique With Various Classifiers", IEEE Transactions On Computational Social Systems, VOL. 8, NO. 5, October 2021

[5] M.Kalimuthu, P.Vaishnavi, M.Kishore, "Crop Prediction using Machine Learning", Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020), IEEE Xplore Part Number: CFP20P17-ART; ISBN: 978-1-7281-5821-1

[6] Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal, "Crop Yield Prediction Using Machine Learning Algorithms", 2019 Fifth International Conference on Image Information Processing (ICIIP), IEEE Xplore

[7] Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate, "Crop Prediction System using Machine Learning Algorithms", International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056, p-ISSN: 2395-0072,Volume: 07 Issue: 02, Feb 2020,

[8] Sayed Mazhar Ali; Bhagwan Das, Dileep Kumar, "Machine Learning based Crop Recommendation System for Local Farmers of Pakistan" ., Innovation,Technologies Management Journal, ISSN: 2237-0722 Vol. 11 No. 4 (2021)

[9] Rishi Gupta, Akhilesh Kumar Sharma, Oorja Garg, Krishna Modi, Shahreen Kasim, Zirawani Baharum, Hairulnizam Mahdin And Salama A. Mostafa , "WB-CPI: Weather Based Crop Prediction in India Using Big Data Analytics",IEEE Access, October 4- 2021

[10] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 6 - 8 May 2015. pp.138-145.