

# Crop Selection and it's Yield Prediction



Aksheya Suresh, K. Monisha, R. Pavithra, B. Marish Hariswamy

**Abstract:** *The field of Agriculture plays a major role in the Indian economy. This sector helps to meet the basic needs of human and their civilization. Hence agriculture would be the enterprise in the globe. Considering the parameters of the agriculture, selection of crops plays a very vital role in farming. The proposed model for Crop selection and it's yield prediction mainly focusses on the season and location to display the desired crop for cultivation . This requirement is implemented with Machine Learning algorithms like Decision tree for classification and Linear regression for yield prediction to maximize the crop yield. This model helps the farmers to know about the correct crop to be cultivated in a particular location . And also gives a approximate percentage of yield based on the data available in the dataset. Thus selecting crop for cultivation becomes a easier task for farmers because selection of correct crop for their location is precisely implemented using this project.*

**Keyword :** *Decision tree, Linear Regression, Crop selection , Yield Prediction.*

## I. INTRODUCTION

Around 2/3rd of the Indian population is engaged in agriculture activities. Hence agriculture plays a important role in the economic development of India. In recent times, modern people don't have the complete awareness about the cultivation of the crops in a right season and at a right location. Due to which modern farmers lack the knowledge of proper selection of crops. Selecting a wrong crop for cultivation may lead to loss in achieving high yield rate and also simultaneously leads to shortage of food. These difficulties implies the need of smart farming which can be achieved with various machine learning algorithm. The model uses machine learning algorithms to make predictions of the desired crop and it's approximate yield production and Python as the programming language that lets you work quickly and integrate systems more efficiently in machine learning area.

Manuscript received on March 15, 2020.

Revised Manuscript received on March 24, 2020.

Manuscript published on March 30, 2020.

\* Correspondence Author

**Aksheya Suresh\***, B.Tech., M.E., Assistant Professor in Rajalakshmi Engineering College.

**K. Monisha**, Student, Rajalakshmi Engineering College, Computer Science and Engineering, Chennai, India Email: [monishaamalu@gmail.com](mailto:monishaamalu@gmail.com)

**R. Pavithra** Student, Rajalakshmi Engineering College, Computer Science and Engineering, Chennai, India Email: [madhuma6778@gmail.com](mailto:madhuma6778@gmail.com)

**Marish Hariswamy.B.**, Student, Rajalakshmi Engineering College, Computer Science and Engineering, Chennai, India Email: [marishhariswamy@gmail.com](mailto:marishhariswamy@gmail.com)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Machine learning algorithms uses historical data and information to gain experiences and generate a trained classifier by training. This classifier makes output predictions.

The accuracy of the model will be better if dataset is accurate .It has been observed that machine learning methods such as regression and classification perform better than various statistical models.

Crop production using machine learning has showed a greater improvement in the field of agriculture. It predicts the data from historical data which is provided by authorized organisation or sector. Improving the farming with latest technologies not only improves the production of the crop but also the quality of the crop produced . Hence smart farming is the means to overcome the farmers economical poverty and improve the country's growth.

## II. RELATED WORKS

### A. Crop Prediction System Using Machine Learning

[1]This is a android based application that prove to be beneficial to agriculture which will increase crop productivity resulting in better yields to the farmer. The system aims at smart agriculture by monitoring the agriculture field which will increase the productivity. In this proposed system data analytics is being used for processing the datasets to draw conclusion to the information collected and Multiple linear regression is being used for prediction analysis.

### B. Smart Farming Prediction Using Machine Learning

[2]This system aims at smart farming which is practical, cheap and easy to develop the task that results in higher productivity of the crop to the farmers. In this system ,classification is implemented with Random Forest Algorithm , Bayesian Network for statistical analysis and Artificial Neural Network for comparing the patterns.This proposed smart system provides real time suggestion and also make long term forecasts based on the user choices.

### C .Crop Selection Method Based Various Environmental Factors Using Machine Learning

[3]This system suggest the method that provide details regarding the maximum yield of the crop by considering the affecting parameters. This system is implemented through two phases namely crop selecting method and crop sequencing method . Crop selection uses classification algorithm and Crop sequencing algorithm for sequencing of crop.

### D. Prediction Of Crop Yield Using Machine Learning

[4]This system focuses on checking the soil quality for predicting the crop for cultivation according to the soil type to maximize the yield and also recommending the appropriate fertilizer.



This project uses both supervised and unsupervised algorithms such as Kohonen's SOM and BPM. BPM is used for testing the data set and the other algorithm is being used for checking the effectiveness of the prediction.

**E . A Survey On Crop Prediction Using Machine Approach**

[5]This system uses various techniques to compare which provide the best result. They use ANN, fuzzy network and also some of the Data mining techniques but finally ANN produces the best result to their project. This paper describes more about the technique and its usage rather than the implementation.

**III. PROPOSED SYSTEM**

The proposed system determines the suitable crop to be cultivated and its appropriate yield production . The system gets the input such as state name, district name, season for classifying the crop and inputs such as year and available yield area for predicting the crop production. This system analyzes the crop yield production based on available data in the dataset. In order to classify the crops based on location and season we use Decision tree algorithm and for crop yield production we use Linear regression method. The suitable crop to be cultivated is obtained from decision tree classifiers such as state name, district name , season and fed as an input to the linear regression algorithm to find the yield of that selected crop. This proposed system reduces the need for costlier hardwares and the time to acquire the knowledge about suitable crops. Figure 1 overall shows the architecture of the proposed model.

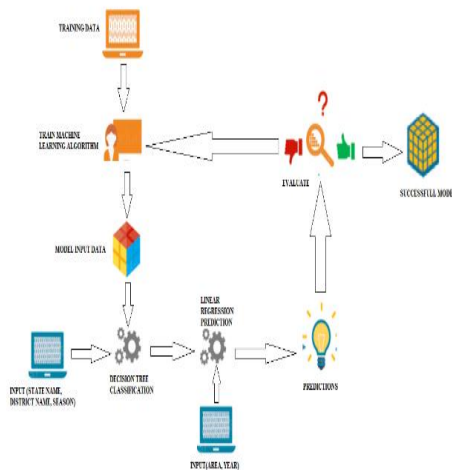


Figure 1 : Architecture

**IV. IMPLEMENTATION**

**A. Overview of Dataset**

To start with any machine learning technique, it is first necessary to bring all the data together. The preliminary data collection is carried out for the states in India. Each area in this collection is identified by the respective state name and district name. The data are taken in six input variables. The variables are state name, district name,crop\_year,

season,crop,area, production. Table 1 shows the first five records of the dataset.

**Table 1 : First five records of dataset**

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0

**B. Preprocessing**

Machine learning algorithm does not work well with processing raw data. Therefore before the data fed into the system it has to be processed. Preprocessing converts raw data into clean data. In preprocessing the dataset is imported into the model and preprocessed using preprocessing algorithms.

**Import of dataset :** The dataset obtained from the authorized organization is given as an input to the model. The dataset contains raw data which has to be processed before the implementation of the model.

**Deleting rows :** The one way of handling missing values is deleting the rows with null values. This method is a quick solution and it is typically is preferred in cases where the percentage of missing values is relatively low.

**Label encoding:** Label encoding converts word labels into numbers to let algorithms work on them. This is an important pre-processing step for the structured dataset in supervised learning. This ensures that the data in the dataset are in the specified format for usage in the algorithm.

**C. Split the data for Training and testing**

We need to split the dataset into the training and testing dataset before training the model for prediction. Initially we have to separate the target variable from the attributes in the dataset.

```
features = data.iloc[:,0:3].columns
target = 'Crop'
X = data[features]
y = data[target]
```

Above are the lines from the code that separates the dataset. The variable X contains the features considered for prediction while the variable Y contains the target variable of the dataset. Then split the dataset for training and testing purpose.

```
X_train, y_train, X_test, y_test = train_test_split(X, Y, test_size = 0.3, random_state = 100)
```

Above code split the dataset for training and testing. Since we are splitting the dataset in a ratio of 70:30 between training and testing so we are pass *test\_size* parameter's value as 0.3. The function *random\_state* variable is a pseudo-random number generator state used for random sampling.

**D. Make predictions**

In our proposed model we have two predictions made.

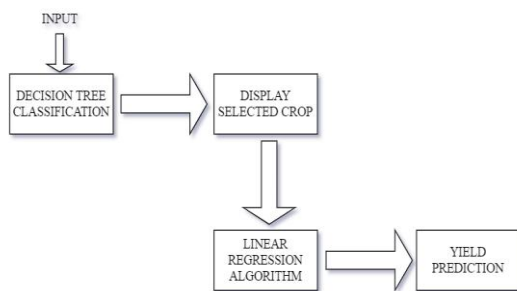
*1.Prediction of crop*

The prediction of suitable crop is dependent on various factors such as location,season, year and past crop production in order to predict the crop accurately. These factors are given as input to the model and state name, district name and season are passed as the classifiers. Based on the classification made by the algorithm it provides a suitable crop to be cultivated.

*2.Prediction of yield*

The output from the previous phase is given as an input along with the area available for cultivation to the model. The model predicts the approximate crop production by linear regression algorithm.

Figure 2 shows the flow of the work of the proposed system.



**Figure 2 : Make Predictions**

**V. CONCLUSION**

Crop selection and it’s yield prediction is still remaining as a challenging issue for farmers. The aim of this research is to propose and implement a model that helps the user in a wider way and cost efficiently use the system. This model classifies the crops that can be cultivated based on the location and season that the user gives and predict the crop yield production from the collection of past data. This has been achieved by using machine learning techniques such as decision tree and linear regression algorithms on dataset.

**VI. RESULTS AND DISCUSSION**

The proposed idea is successfully developed with the machine learning algorithms. The model gives an accuracy of 88.7% when tested. This model can be tested with different dataset for improved accuracy. This model can be enhanced by using it with mobile application. With the scope of mobile application it can be more flexible to use for the farmers. Since agriculture determines the life of people, innovation in this field is mandatory . Therefore this proposed model provides a wider use. This project has a vast scope in future.

**REFERENCES**

1. Prof. D.S. Zingade, Omkar Buchade , Nilesh Mehta, Shubham Ghodekar, Chandan Mehta , “Crop Prediction System Using Machine Learning”, International Journal of Advance Engineering and Research Development Special Issue on Recent Trends in Data Engineering Volume 4, Special Issue 5, Dec.-2017.

2. S.R Rajeswari, Parth Khunteta, Subham Kumar, Amrit Raj Singh, Vaibhav Pandey , “ Smart Farming Prediction Using Machine Learning”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-7 May, 2019

3. Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, Prajakta Kulkarni, “Crop Selection Method Based Various Environmental Factors Using Machine Learning” , International Research Journal of Engineering and Technology (IRJET) Volume: 04 ,Issue: 02, Feb -2017.

4. Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya R L, “ Prediction Of Crop Yield Using Machine Learning” , International Research Journal of Engineering and Technology (IRJET), Volume: 05,Issue: 02, Feb-2018.

5. Sriram Rakshith.K, Dr. Deepak.G, Rajesh M, Sudharshan K S, Vasanth S, Harish Kumar, “A Survey On Crop Prediction Using Machine Approach”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 7 Issue IV, Apr 2019.

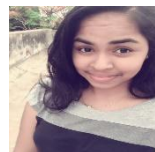
6. <https://technostacks.com/blog/machinelearning-in-agriculture/>

7. <https://en.wikipedia.org/wiki/Agriculture>

**AUTHORS PROFILE**



**Aksheya Suresh**  
Aksheya Suresh., B.Tech., M.E.,  
Assistant Professor in Rajalakshmi Engineering College.Areas of Interest are Data Mining, Machine Learning and Internet of Things. Have published paper titled Automatic Generation of Multi Modal Learning Objects from Electronic Textbooks in Springer Series (AISC) ISSN 2194-5357 Volume 1, pp 427-437. Have also published paper in various International and National Conferences.  
Email:[aksheya.suresh@rajalakshmi.edu.in](mailto:aksheya.suresh@rajalakshmi.edu.in)



**K. Monisha** Student, Rajalakshmi Engineering College, Computer Science and Engineering, Chennai, India Email: [monishaamalu@gmail.com](mailto:monishaamalu@gmail.com)



**R. Pavithra** Student, Rajalakshmi Engineering College, Computer Science and Engineering, Chennai, India Email: [madhuma6778@gmail.com](mailto:madhuma6778@gmail.com)



**Marish Hariswamy.B** Student, Rajalakshmi Engineering College, Computer Science and Engineering, Chennai, India Email: [marishhariswamy@gmail.com](mailto:marishhariswamy@gmail.com)

