

# CROP YIELD OPTIMIZATION AND PREDICTION BY MACHINE LEARNING ALGORITHMS

Sakthivel.E M.E.,/Assistant professor

Jeppiaar Engineering college

Haritha.B

Lakshmi Priya.M

Monika.D

Jeppiaar Engineering college

Jeppiaar Engineering college

Jeppiaar Engineering college

**ABSTRACT**--Agriculture is one of the major and lowest paid occupations in India. Agriculture is that sector which plays an important role in improving the economy of our country. Machine learning can drive a boom in the agriculture sector by changing the income landscape through optimum crop growing. Farmers are growing the same crop over and over again without trying new variety of crops and they are applying fertilizer in random amount without knowing the small quantity and quantity. So, it is directly affecting the crop yield and causes acidification of soil and damages the top layer. Hence, we have designed the system using machine learning algorithms for the betterment of the farmers. This project will help the farmers to know the yield of their crop before taking up cultivation in the agriculture field and thus help them to take proper decision. Data mining is useful for predicting crop yield production. Generally, data mining is the process of analyzing data from different perspectives and summarizing it into important information. This research proposes and implements a system to predict crop yield from past data. This paper present a brief analysis of crop yield prediction using machine learning techniques. The experimental results show that the proposed functions efficiently predict crop yield output.

**KEYWORDS**--Crop yield prediction, Machine Learning, Data Mining, Random forest, Linear Regression and Logistics Regression algorithm.

## INTRODUCTION

Since ancient times, agriculture is considered as the main source of supply to meet the daily needs of human life. It is also considered a primary occupation, and is also one of the major industrial areas of India. Farmers must follow a traditional naked eye observation and produce healthy crops without the involvement of animals and chemicals for their cultivated land in order to maintain healthy diversity. But nowadays, weather conditions are increasingly turning against food scarcity and fundamental assets to increase security. Meanwhile, GDP in the agriculture sector continues to decline, where it was around 17.2% in 2005, 11.1 in 2012, 5% in 2018 and declined to 2109 in Q1 of 2020-2%. About 80 percent of farmers come from rural areas, and if the revenue from crop production is less, their lifestyle will be affected by farms at the industry level. It is understandable for farmers in India to show some special concern towards effective and precision farming. There are many ways to grow crops in India, learn the benefits and improve the level of crops so as to maintain economic growth in agriculture. Therefore, the deployment of one of the most recent advances in technology, such as machine learning, is one of the answers to crop prediction in relation to atmospheric and soil parameters of agricultural land. Since, today's climatic conditions are not as predictable as decades ago.

Due to globalization it is changing day by day. Therefore, farmers are facing difficulties in predicting the weather and crops based on the climate data. The advancement of machine learning in recent years plays an important role in every sector including agriculture, where the crop prediction process is carried out along with integrating the past data and the current data of a particular month to prove the accuracy of climate data. Machine learning can be a method of analyzing information to automate a given model and can be a branch of AI that relies on the concept that the system will study the data to make selections with minimal human intervention. . One might be the logical classifier, where a naive mathematician who predicts membership opportunities for each group, such as the probability that the knowledge belongs to a specific class. Our work proposes to help farmers determine soil quality by analyzing its various parameters and suggest crops based on the results obtained using a data mining approach. The system uses the classification algorithm of the random forest to improve the efficiency of the crop recommendation system. The system maps soil and crop data to predict the list of suitable crops for the soil and also provides information about the nutrients that are deficient in the soil for the particular crop. Hence it is left to the user to decide about the crop to be sown. Thus, the system helps in imparting knowledge to the hardworking farmers. There are many ways to increase and improve crop yield and crop quality. the main objectives are

- A. Using machine learning techniques to estimate crop yield.
- b. To provide an easy to use user interface.
- C. To increase the accuracy of crop yield prediction.
- D. To analyze various climatic parameters (cloud cover, precipitation, temperature)

## LITERATURE SURVERY

In [1] Crop Selection Method to maximize crop yield rate using machine learning algorithm[May 2015]International Conference on Smart Technologies and Management for Computing,Communication,Contolrs,Energy and Materials(ICSMT).This Paper presents a technique named CSM to select sequence of crops to be planted over season crop method may improve net yield rate of crops to be planted over season .It helps to resolve selection of crop based on prediction yield rate influenced by parameters(e.g.weather,soil type, water density, crop type)

In [2] Yield Forecasting for Indian crops with Ensemble Model[July 2019].International Journal of Recent Technology and Engineering(IJRTE).Ensemble based crop yield forecasting is proposed in this work. Each of machine learning models of ARIMA-LR, Kmeans- Neural Network and Baiyes Net are trained using different parameter are their result are ensemble using weighted average ensembling.

In [3] Crop Yield Prediction on Indian Agriculture using Machine Learning[June 2020].International Conference for Emerging Technology(INCET).It uses advanced regression techniques like Kernel Ridge ,Lasso and ENet algorithms to predict the yield and uses the concept of stacking Regression for enhancing the algorithms to give a better prediction.

In [4] Machine learning approach for forecasting crop yield based on climatic parameters[Jan2014].International Conference on Computer Communication and Informatics(ICCCI).In this paper,the study was aimed to develop a website for finding out the influence of climatic parameter on the crop prediction in selected districts.The selection of districts in which the selected crop area under the particular crop.

In [5] Crop Yield Prediction and Efficient use of Fertilizers [October 2018].International Journal of Creative Research Thoughts(IJCRT).It uses classification techniques like ANN, Naïve Byes and support vector Machines.They have additionally covered each climate parameters and soil parameter as functions of their modelling.

In [6] Crop Prediction using Machine Learning Approaches[August 2020].International Journal of Engineering Research and Technology(IJERT).It describes the different machine learning methods used for accelerating crop yield.In this paper they gone through different artificial intelligence techniques such as machine learning algorithm ,big data analysis for precision agriculture.

In [7] Crop Yield Prediction Based on Weather using Machine Learning[9 Dec 2019].International Journal of Engineering and Advanced Technology(IJEAT).The paper says ,vast research has been done and several attempts are made for application of machine learning in agricultural fields.

In [8] Recognition of Yield in Crop Images using Deep Learning models for Smart Agriculture [19 Oct 2021].International Journal of Engineering Research&Technology (IJERT).This paper elucidates the diverse automation approaches for crop yield detection techniques with virtual analysis and classifier approaches and highlights the machine vision and deep learning models which need to be explored for improving automated precision farming.

In [9] Design and Development of Machine Learning Model for crop yield prediction [6 Sep 2021].International Conference on Emerging Technologies. It concluded the method which is helpful in finding the suitable crop according to the land. Agro Algorithm is used in this paper.

In [10] Crop Yield Prediction using Machine Learning algorithm [July 2021].Fifth International Conference on Intelligent Computing and Control

System(ICICCS).This research paper focused on process of crop yield analysis, the analysis of several ML methods used to examine Crop yield prediction ,the objectives and problem statement of crop yield prediction and comparative analysis of several types of research.

## METHODOLOGY

1.Dataset Collection: In this step, we collect data from various sources and prepare datasets. And the dataset provided is in use for analytics (descriptive and diagnostic). There are many online abstract sources such as Data.gov.in and indiastat.org. The annual essence of a crop shall be used for at least ten years. These datasets generally accept the behavior of chaotic time series. Combination of primary and essential essence. Random forests are global and regional crop yield prediction.

|    | A           | B        | C        | D        | E         | F           | G               |
|----|-------------|----------|----------|----------|-----------|-------------|-----------------|
| 1  | Temperature | Humidity | Moisture | Nitrogen | Potassium | Phosphorous | Fertilizer Name |
| 2  | 26          | 52       | 38       | 37       | 0         | 0           | Urea            |
| 3  | 29          | 52       | 45       | 12       | 0         | 36          | DAP             |
| 4  | 34          | 65       | 62       | 7        | 9         | 30          | 14-35-14        |
| 5  | 32          | 62       | 34       | 22       | 0         | 20          | 28-28           |
| 6  | 28          | 54       | 46       | 35       | 0         | 0           | Urea            |
| 7  | 26          | 52       | 35       | 12       | 10        | 13          | 17-17-17        |
| 8  | 25          | 50       | 64       | 9        | 0         | 10          | 20-20           |
| 9  | 33          | 64       | 50       | 41       | 0         | 0           | Urea            |
| 10 | 30          | 60       | 42       | 21       | 0         | 18          | 28-28           |
| 11 | 29          | 58       | 33       | 9        | 7         | 30          | 14-35-14        |
| 12 | 27          | 54       | 28       | 13       | 0         | 40          | DAP             |
| 13 | 31          | 62       | 48       | 14       | 15        | 12          | 17-17-17        |
| 14 | 25          | 50       | 65       | 36       | 0         | 0           | Urea            |
| 15 | 32          | 62       | 41       | 24       | 0         | 22          | 28-28           |
| 16 | 26          | 52       | 31       | 14       | 0         | 41          | DAP             |
| 17 | 31          | 62       | 49       | 10       | 13        | 14          | 17-17-17        |
| 18 | 33          | 64       | 34       | 38       | 0         | 0           | Urea            |
| 19 | 25          | 50       | 39       | 21       | 0         | 19          | 28-28           |
| 20 | 28          | 54       | 65       | 39       | 0         | 0           | Urea            |
| 21 | 29          | 58       | 52       | 13       | 0         | 36          | DAP             |
| 22 | 30          | 60       | 44       | 10       | 0         | 9           | 20-20           |
| 23 | 34          | 65       | 53       | 12       | 14        | 12          | 17-17-17        |
| 24 | 35          | 68       | 33       | 11       | 0         | 37          | DAP             |
| 25 | 28          | 54       | 37       | 36       | 0         | 0           | Urea            |

2. Data Partitioning: The entire dataset is divided into 2 parts: for example, 75% of the dataset is used for training the model and 25% of the data is set aside for testing the model.

3. Pre-processing Step: An important step in pre-processing the data is considered as the machine learning phase. Pre-processing involves adding missing values, extracting the correct set of data, and functionality. Data set form is critical to the process of analysis. The data collected in

this step will be driven into Google Colab platform as python Programming to get the desired output.

4. Feature Extraction: The extraction of features will reduce the included data size to characterize a wide collection of data. Soil, crop characteristics are collected from the pretreatment process, setting up the final training data collection. This approach selects features based on the correlation matrix i.e. the features which have higher correlation value are selected as a significant predictive function to yield.

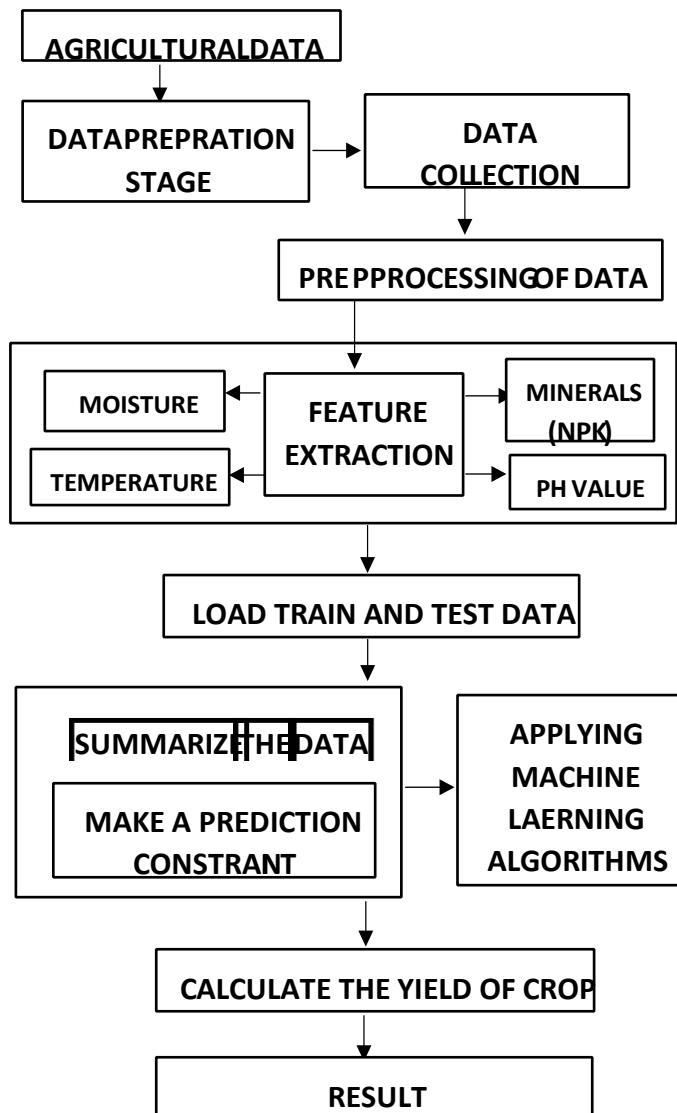
5. Comparison and Selection of Machine Learning Algorithm

- Random Forest Classifier: Random Forest is the most popular and Powerful Supervised Machine Learning Algorithms capable of doing both classification and regression functions, that work by producing a multitude of Decision tree at training and generating the output of the class that's the way of classes (classification) or mean prediction (regression) Of individual trees.
- Linear regression : Linear regression is one of the easiest and the most popular machine learning algorithms. It is a statistical method used for predictive analysis. Linear regression makes predictions for continuous/real or numerical variables like sales, salary, age, product price, etc. It shows a linear relationship between a dependent (y) and a. more independent (y) variables, so it is called linear regression. It provides a sloping straight line representing the relationship between the variables.
- Logistic regression : Logistic regression is a supervised learning classification algorithm used to predict the probability of an target variable. The nature of the target or dependent variable is

dichotomous, meaning that there will be only two possible classes. When the logistic regression algorithm provides an accuracy of 87.8%.

|   | Model | Accuracy | Train_acc |
|---|-------|----------|-----------|
| 0 | KNN   | 1.00     | 0.949367  |
| 2 | RFC   | 1.00     | 1.000000  |
| 4 | XGB   | 0.95     | 1.000000  |
| 1 | DT    | 0.90     | 1.000000  |
| 3 | GBC   | 0.90     | 1.000000  |

### BLOCK DIAGRAM



**Arduino**-Arduino is the open-source platform used for building electronics project.

Arduino consists of the physical programmable circuit boards and a piece of software, or IDE, that runs on your computer, is used to write and upload computer code to the physical board.

**Moisture Sensor** - Soil moisture sensors measure the amount of water in the soil. Since and the weight of a sample, soil moisture sensors measure the volumetric water content indirectly as a proxy for moisture content, such as some other property of soil, such as electrical resistance,

dielectric constant, or interaction with neutrons. .

**NPK Sensor**: Soil NPK sensor are suitable for detecting the content of nitrogen, Phosphorus and potassium in the soil. It helps as to determine soil fertility thereby facilitating systematic assessment of soil condition. The sensors can be buried in the soil for a long time.

**PH Sensor**: A pH sensor is one of the most essential instruments commonly used for water measurement. This type of sensor is capable of measuring the amount of alkalinity and acidity in water and other solutions.

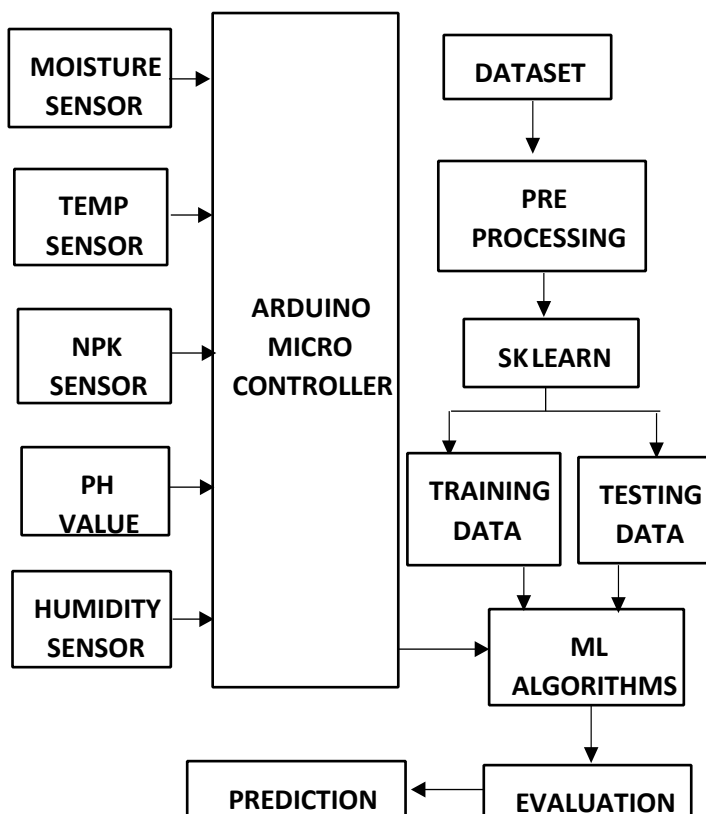
**Temperature Sensor** - A temperature sensor is an device, typically, a thermocouple or a RTD that provides the temperature measurement by means an electrical signal. The Temperature sensor that uses a transistor connected to an external diode as a sensing element to measure the outside temperature of the sensor.

**Humidity Sensor**-This module is based on HR202 resistive humidity sensor, which exhibits excellent linearity, has wide measurement range and low power consumption. The module has both the features a Power output indicator LED and a digital output indicator.

**SK Learn** -scikit-learn is a free software machine learning library for the Python programming language. scikit-learn is probably the most useful library for machine learning Python.



### ARCHITECTURAL DIAGRAM

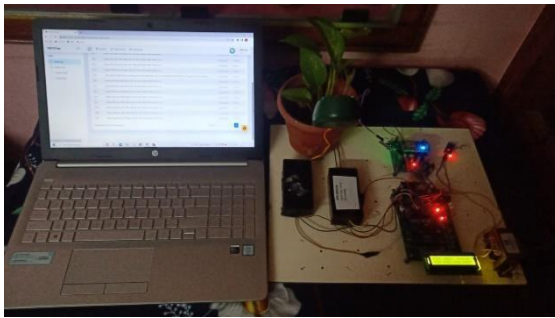
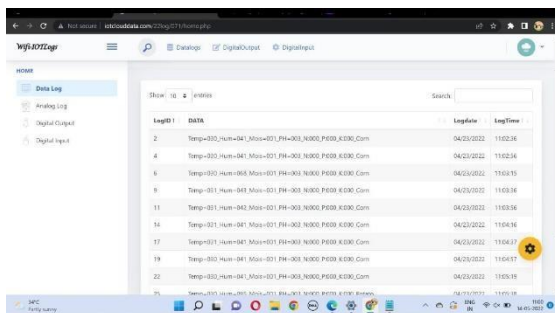


The components used in the proposed work are Arduino UNO with Atmega processor quoted as 1, Dht11 temperature and moisture sensor quoted as 2 and soil moisture quoted as 3 has

gone. The collected data is sent to Google Colab, which is an open source platform completely written in Python and designed for data analysts. Google Colab offers built-in integration with many popular scientists Packages including Numpy, Pandas. Data collection Pre-processing data, selection of features, and prediction are considered as different steps involved in crop yield prediction and process flow. The preceding data collection includes temperature, air humidity, soil moisture data from the stage at high accuracy. And the collected data is indicated in the sheet. In this set-up the component currently performs special operations to collect the data. Once the controller triggers the sensor The data will be stored as a .csv file with some time delay associated with it and transferred to the controller board and then transferred to the cloud. The value is displayed on the screen as the output value sensed by the respective sensor. Datasets are collected, preprocessing techniques Used to remove unwanted data from dataset and perform training and testing process. The collected sensor values are Dataset comparison using machine learning algorithms such as random forest, linear regression and logistics regression. Finally, by comparing those three algorithms which shows a higher accuracy is taken for crop prediction.

### CONCLUSION

Proposed Supervised Machine Learning using Random Forest Algorithm, Linear Regression, Logistic Regression is used to predict Crop at high accuracy. Thus, the seed is estimated as the output of the given input parameter. This work can be of great help to the needy farmers who have little knowledge of predicting crops for a sustainable future development. In future it may also expand to suggest suitable guidelines for fertilizers, crop lands and crops for the given inputs. In addition, the sources of sunlight and the health of the crop are monitored at regular intervals and also taken into account to get better crop yield.

| LogID | DATA                                                          | LogTime             |
|-------|---------------------------------------------------------------|---------------------|
| 2     | Temp=32.0 Hum=44.1 Moist=0.1 PH=6.8 NO2=0.000 P009=0.000 Corn | 04/23/2022 11:02:36 |
| 4     | Temp=32.0 Hum=44.1 Moist=0.1 PH=6.8 NO2=0.000 P009=0.000 Corn | 04/23/2022 11:02:36 |
| 6     | Temp=32.0 Hum=44.1 Moist=0.1 PH=6.8 NO2=0.000 P009=0.000 Corn | 04/23/2022 11:02:36 |
| 8     | Temp=32.1 Hum=44.2 Moist=0.1 PH=6.8 NO2=0.000 P009=0.000 Corn | 04/23/2022 11:03:36 |
| 10    | Temp=32.1 Hum=44.2 Moist=0.1 PH=6.8 NO2=0.000 P009=0.000 Corn | 04/23/2022 11:03:36 |
| 12    | Temp=32.1 Hum=44.2 Moist=0.1 PH=6.8 NO2=0.000 P009=0.000 Corn | 04/23/2022 11:03:36 |
| 14    | Temp=32.1 Hum=44.2 Moist=0.1 PH=6.8 NO2=0.000 P009=0.000 Corn | 04/23/2022 11:04:36 |
| 16    | Temp=32.1 Hum=44.2 Moist=0.1 PH=6.8 NO2=0.000 P009=0.000 Corn | 04/23/2022 11:04:36 |
| 18    | Temp=32.1 Hum=44.2 Moist=0.1 PH=6.8 NO2=0.000 P009=0.000 Corn | 04/23/2022 11:04:36 |
| 20    | Temp=32.1 Hum=44.2 Moist=0.1 PH=6.8 NO2=0.000 P009=0.000 Corn | 04/23/2022 11:05:36 |
| 22    | Temp=32.1 Hum=44.2 Moist=0.1 PH=6.8 NO2=0.000 P009=0.000 Corn | 04/23/2022 11:05:36 |

## REFERENCES

- [1] R. Ghadge, J. Kulkarni, P. More, S. Nene, and R. L. Priya, "Prediction of crop yield using machine learning," *Int. Res. J. Eng. Technolgy*, vol. 5, 2018.
- [2] F. H. Tseng, H. H. Cho, and H. T. Wu, "Applying big data for intelligent agriculturebased crop selection analysis," *IEEE Access*, vol. 7, pp. 116965-116974, 2019.
- [3] A. Suresh, N. Manjunathan, P. Rajesh, and E. Thangadurai, "Crop Yield Prediction Using Linear Support Vector Machine," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 6, pp. 2189-2195, 2020.
- [4] M. Alagurajan, and C. Vijayakumaran, "ML Methods for Crop Yield Prediction and Estimation: An Exploration," *International Journal of Engineering and Advanced Technology*, vol. 9 no. 3, 2020
- [5] P. Kumari, S. Rathore, A. Kalamkar, and T. Kambale, "Prediction of Crop Yeild Using SVM Approch with the Facility of E-MART System" *Easychair* 2020.
- [6] S. D. Kumar, S. Esakkirajan, S. Bama, and B. Keerthiveena, "A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier," *Microprocessors and Microsystems*, vol. 76, pp.103090, 2020
- [7] P. Tiwari, and P. Shukla, "Crop yield prediction by modified convolutional neural network and geographical indexes," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 8, pp. 503-513, 2018.
- [8] P. Sivanandhini, and J. Prakash, "Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 5, pp. 1092-1096, 2020.
- [9] N. Nandhini, and J. G. Shankar, "Prediction of crop growth using machine learning based on seed," *Ictact journal on soft computing*, vol. 11, no. 01, 2020
- [10] A. A. Alif, I. F. Shukanya, and T. N. Afee, "Crop prediction based on geographical and climatic data using machine learning and deep learning", *Doctoral dissertation, BRAC University* 2018.