

# Crop Yield Prediction Using Ensemble Algorithm

G. Abirami<sup>1</sup>, R. Reni Hena Helan<sup>2</sup>, K. Anandhan<sup>3</sup>, G. Vishnuvardhan Reddy<sup>4</sup>, S. Naveen Kumar<sup>5\*</sup>,  
V. Ruban Karthick<sup>6</sup>

<sup>1,2</sup>Assistant Professor, Department of Computer Science and Engineering, Dhanalakshmi College of Engineering, Chennai, India

<sup>3,4,5,6</sup>Student, Department of Computer Science and Engineering, Dhanalakshmi College of Engineering, Chennai, India

**Abstract:** Machine studying is a pivotal standpoint for greedy real-global and useful use instances for yield prediction of crops. Machine studying is a supportive device for the rural area which allows us to determine which plant to develop and whilst to develop the preferred plant. This study suggests the utilization and implementation of predicting the crop kind primarily based totally on ensemble techniques. From a hard and fast of given parameters, gadget studying can forecast the final results via unsupervised or supervised studying techniques. To get the desired output parameter, we have to produce an appropriate and pleasant characteristic via the means of a few sets of variables if you want to depict the output using the given entered variables or parameters. This consists of the ensemble (combination) of or extra gadget studying algorithms which improves the crop yield prediction accuracy.

**Keywords:** Accuracy, AdaBoost regressor, Classification, Crop yield, Decision tree regressor, Ensemble, Metrics, Prediction, Churn.

## 1. Introduction

India is one of the most important manufacturers of agricultural merchandise and nevertheless has little or no farm productivity. Productivity desires to be accelerated in order that farmers can get greater pay from the identical piece of land with much less labor. Precision agriculture presents a manner to do it. Precision farming, because the call implies, refers back to the making use of unique and right general remarks like fertilizers, soil etc. however nowadays the fashion in agriculture has significantly advanced because of globalization. Various elements have affected the fitness of agriculture in India. Many new technologies have been advanced to regain fitness. One such method is precision agriculture, at the right time to grow its productiveness and its yields. Not all precision agriculture structures provide nice results. Precision agriculture is a generation of site-unique farming. But in agriculture it's essential that the suggestions made are correct and unique due to the fact that in case of mistakes it can result in heavy fabric and capital loss. Recommendation of vegetation is one most important area in precision agriculture.

## 2. Literature Survey

This proposed machine has long passed through diverse

references and studies over exceptional journals and convention papers which have advanced for the past 10 years. The studies and referrals of the survey is carried out on exceptional domain names like AI, NLP. Home environments that helped to broaden the proposed machine of AI primarily based totally Control AAS. Those references and the particular file at the references are indexed as follows

*Priyadharshini A, Swapneel Chakraborty, Aayush Kumar, Omen Rajendra Pooniwala (2021) Proceedings of the Fifth International Conference on Computing Methodologies and Communication Intelligent Crop Recommendation System using Machine Learning*

Agriculture performs an essential function within the socioeconomic cloth of India. Failure of farmers to determine at the best-proper crop for the land the usage of conventional and non-medical techniques is a severe difficulty for a rustic wherein about fifty eight percent of the populace is worried in farming. Sometimes farmers didn't pick out the proper vegetation primarily based totally on the soil conditions, sowing season, and geographical location. This consequences in suicide, quitting the agriculture field, shifting closer to city regions for livelihood. To triumph over this difficulty, this study of paintings has proposed a gadget to help the farmers in crop choice with the aid of thinking about all of the elements like sowing season, soil, and geographical location. Furthermore, precision agriculture is being applied with cutting-edge agricultural generation and its miles evolving in growing international locations that focus on site-unique crop management.

*Fariha Shahrin, Labiba Zahin, Ramisa Rahman, ASM Jahir Hossain 2020 Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning International Conference on Electrical and Computer Engineering*

Bangladesh is predominantly an agriculture primarily based totally in the USA, where in an intensive a part of its populace is basically hired in its agriculture sector. However, unsure crop yields and inefficient farming infrastructure cause destructive outcomes on meal security. Habiganj is chosen to examine the vicinity due to its vulnerability to floods and drought because of its specific terrain. This paper offers a mixture of agricultural

\*Corresponding author: naveennaveensiva@gmail.com

mapping and tracking of Habiganj with crop increase and yield prediction. Multi-spectral band pictures of Habiganj from Landsat-eight are processed and far-flung sensing indices, correlating to crop increase and yield are extracted. With alternatives of K-approach and Mask R-CNN methods, the alternative over time is evaluated for crop increase estimation in each Python and Matlab. Then the usage of a kind system gaining knowledge of algorithms crop yield of Habiganj is expected from its current parameters and for higher accuracy destiny values of its datasets are expected via means of the usage of sorts of time collection evaluation models. Such analytical reviews and prediction lets in tracking crop increase dynamic and figuring out early symptoms and symptoms of discount in crop productivity. Furthermore, comparative research is concluded among platforms, algorithms and time collection evaluation to decide the maximum proper surroundings and version for this study's purpose.

*S. Pudumalar, E. Ramanujam, R. Harine Rajashreeñ, C. Kavyañ, T. Kiruthikañ, J. Nishañ, 2016 Crop Recommendation System for Precision Agriculture IEEE Eighth International Conference on Advanced Computing*

Data mining is the exercise of inspecting and deriving functional records from the facts. Data mining reveals its utility in diverse fields like finance, retail, medicine, agriculture etc. Data mining in agriculture is used for studying the diverse biotic and abiotic factors. Agriculture in India performs a most important position in the financial system and employment. The not unusual place trouble current a number of the Indian farmers is they don't select the proper crop primarily based totally on their soil requirements. Due to this they face a critical setback in productivity. This trouble of the farmers has been addressed through precision agriculture. Precision agriculture is a present-day farming method that makes use of studies facts of soil characteristics, soil types, crop yield facts series and shows the farmers the proper crop primarily based totally on their web website online unique parameters. This reduces the incorrect preference on a crop and will increase productivity. In this paper, this trouble is solved via way of means of presenting a advice device thru an ensemble version with majority balloting method the use of Random tree, CHAID, K-Nearest Neighbor and Naive Bayes as freshmen to propose a crop for the web website online unique parameters with excessive accuracy and efficiency.

*B Mahesh, N. Ramakrishna, Dudekula Nabi Rasool, S. Aiman Ajmal, K. Rajasekhar. 2021 Crop Yield Prediction using Machine Learning Algorithms Journal of Resource Management and Technology.*

Agriculture is one of the predominant and the least paid occupations in India. Machine getting to know can carry an increase within the agriculture subject through converting the profits situation thru developing the most advantageous crop. This paper specializes in predicting the yield of the crop through making use of diverse gadget getting to know strategies. The final results of those strategies are compared on the premise of suggesting absolute error. The prediction made through gadget getting to know algorithms will assist the farmers to determine which crop to develop to get the most

yield through thinking about elements like temperature, rainfall, area, etc.

*Nidhi H. Kulkarni, G N Srinivasan, B. M. Sagar, N. K. Cauvery 2018 Improving Crop Productivity Through a Crop Recommendation System Using Ensembling Technique IEEE International Conference on Computational Systems and Information Technology*

Agriculture performs a most important position withinside the monetary boom and improvement of the country. The main and severe setback in crop productivity is that the farmers no longer pick the proper crop for cultivation. In order to enhance the crop productivity, a crop advice gadget is to be evolved that makes use of the ensemble approach of system mastering. The ensemble approach is used to construct a version that mixes the predictions of a couple of system mastering fashions collectively to propose the proper crop primarily based totally at the soil particular kind and traits with excessive accuracy. The unbiased base newbies used within the ensemble version are Random Forest, Naive Bayes, and Linear SVM. Each classifier offers its personal set of sophistication labels with an appropriate accuracy. The elegance labels of character base newbies are mixed with the use of the bulk balloting approach. The crop advice gadget classifies the entered soil dataset into the recommendable crop kind, Kharif and Rabi. The dataset incorporates the soil particular bodily and chemical traits similarly to the climatic situations which include common rainfall and the floor temperature samples.

*Kavita, Pratistha Mathur 2020 Crop Yield Estimation in India Using Machine Learning International Conference on Computing Communication*

Agriculture in India has massive financial support. Population increase is the predominant task for food security. The population increase makes an upward thrust in name for which requires farmers to offer extra from the same agricultural land a good way to boom the supply. Technology can help farmers to offer extra with the help of crop yield prediction. The vital cause of this paper is to anticipate crop yield, the use of location, yield, production, and location under irrigation. Four systems gaining knowledge of techniques Decision Tree, Linear Regression, Lasso regression, and Ridge Regression have been completed to estimate the crop yield. Cross validations methods, for validation, propose absolute error, propose squared error, and root propose squared error, had been used to validate. The Decision tree outperforms distinctive systems gaining knowledge of techniques.

*Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber, 2020 International Conference Crop Yield Prediction based on Indian Agriculture using Machine Learning*

In India, all of us realize that Agriculture is the spine of the country. This script is novel via means of the use of easy parameters like State, district, season, region and the consumer can be expecting the yield of the crop where she or he wishes to. The paper makes use of superior regression strategies like Kernel Ridge, Lasso and ENet algorithms to be expecting the yield and makes use of the idea of Stacking Regression for boosting the algorithms to provide a higher prediction.

Thomas van Klompenburg, Ayalew Kassahun, Cagatay Catal, 2020 Crop yield prediction using machine learning: A systematic literature review

Machine mastering is a vital choice help device for crop yield prediction, inclusive of helping selections on what plants to develop and what to do at some point of the developing season of the plants. Several system mastering algorithms had been carried out to help crop yield prediction research. In this study, we completed a Systematic Literature Review (SLR) to extract and synthesize the algorithms and functions which have been utilized in crop yield prediction research. Based on our seek criteria, we retrieved 567 applicable research from six digital databases, of which we've decided on 50 research for in addition evaluation the usage of inclusion and exclusion criteria. We investigated those decided on research carefully, analyzed the strategies and functions used, and furnished guidelines for in addition research.

3. Methodology

Crop yield prediction is a system which is used to predict the suitable crop based on the parameters considered in the dataset. Based on the predicted crop we can also suggest fertilizer, pesticides and water level that need to be used for better crop growth. The process consists of the following modules.

A. Dataset Collection

Collecting information lets you to seize a file of beyond activities in order that we are able to use information evaluation to discover ordinary patterns. From the one pattern, you construct predictive fashions: the usage of gadgets gaining knowledge of algorithms that search for tendencies and are expecting destiny changes. Predictive fashions are simplest as proper because the information from which they may be built, so proper information series practices are critical to growing high-acting fashions. The information wants to be error-free (rubbish in, rubbish out) and include applicable facts for the mission at hand. For example, a mortgage default version might now no longer be advantageous from tiger populace sizes however may want to take advantage of fuel line costs over time. In this module, we gather the crop advice information from kaggle dataset archives. This dataset incorporates the facts of divorce in preceding years. The datasets used for crop prediction is shown in Fig. 1.

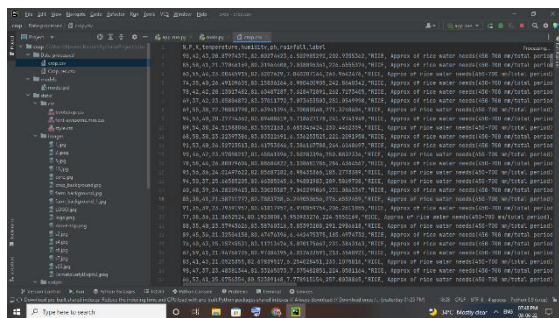


Fig. 1. Datasets which include attributes like N, P, K, Temperature, Humidity, PH, Rainfall

B. Data Cleaning

Data cleansing is a severely critical step in any system studying project. In this module records cleansing is executed to put together the records for evaluation through eliminating or editing the records that can be incorrect, incomplete, duplicated or improperly formatted. In tabular records, there are numerous distinctive statistical evaluation and records visualization strategies you could use to discover your records for you to pick out records cleansing operations you could need to perform.

C. Feature Extraction

This is performed to lessen the quantity of attributes within the dataset subsequently offering benefits like dashing up the education and accuracy improvements. In device mastering, sample recognition, and photo processing, function extraction begins off evolved from a preliminary set of measured records and builds derived values (functions) supposed to be informative and non-redundant, facilitating the following mastering and generalization steps, and in a few instances main to higher human interpretations. Feature extraction is associated with dimensionality reduction When the records entered to a set of rules is just too big to be processed and its miles suspected to be redundant (e.g., the identical size in each toes and meters, or the repetitiveness of pix provided as pixels), then it may be converted into a discounted set of functions (additionally named a function vector). Determining a subset of the preliminary functions is referred to as function selection. The decided-on functions are predicted to incorporate the applicable facts from the enter records, in order that the preferred undertaking may be finished through the usage of this decreased illustration as opposed to the whole preliminary records.

D. Model Training

An education version is a dataset that is used to teach an ML set of rules. It includes the pattern output facts and the corresponding units of entered facts which have a power at the output. The education version is used to run the entered facts via the set of rules to correlate the processed output towards the pattern output. The end result from this correlation is used to alter the version. This iterative method is called "version fitting". The accuracy of the education dataset or the validation dataset is vital for the precision of the version. Model education in device language is the method of feeding an ML set of rules with facts to assist pick out and examine accurate values for all attributes involved. There are numerous varieties of device getting to know models, of which the maximum not unusual place ones are supervised and unsupervised getting to know. In this module we use supervised class algorithms like linear regression to teach the version at the wiped clean dataset after dimensionality reduction.

E. Testing Model

In this module we check the skilled gadget studying version of the usage of the check dataset. In gadget studying, version checking out is called the system wherein the overall performance of a totally skilled version is evaluated on a checking out set. In gadget studying, a programmer generally

inputs the information and the favored behavior, and the common sense is elaborated through the gadget. Therefore, the reason of gadget studying checking out is, first of all, to make certain that this found out common sense will stay consistent, regardless of how typically we name this system with inside the checking out.

#### F. Performance Evaluation

In this module, we examine the overall performance of an educated device, getting to know fashions, the use of overall performance assessment standards together with F1 score, accuracy and type error. In case the version plays poorly, we optimize the device by getting to know algorithms to enhance the overall performance. Performance assessment is a critical component of the device getting to know the process. However, it's a complicated task. It, therefore, desires to be performed cautiously so as for the software of device getting to know radiation oncology or different domain names to be reliable, used to assess a type version for accuracy, precision, and recollect is used for crop recommendation.

#### G. Prediction

Prediction" refers back to the output of a set of rules after it's been skilled on an ancient dataset and carried out to new information whilst forecasting the probability of a specific outcome, along with whether or not or now no longer a consumer will churn in 30 days. The set of rules will generate likely values for an unknown variable for every report with inside the new information, permitting the version builder to discover what that fee will maximum in all likelihood be. The word "prediction" may be misleading. In a few cases, it simply does suggest which you are predicting a destiny outcome, along with whilst you're using the device to decide the following fine motion in an advertising campaign. Other times, though, the "prediction" has to do with, for example, whether or not or now no longer a transaction that already took place changed into fraudulent. In that case, the transaction already happened, however you're making a knowledgeable wager approximately whether or not or now no longer it changed into legitimate, permitting you to take the correct motion. In this module we use a skilled and optimized device getting to know the version to expect whether or not the crop recommendation.

### 4. System Architecture

This machine is a utility which predicts the call of the crop in addition to calculating its corresponding yield. Name of the crop is decided with the aid of using numerous functions like temperature, humidity, wind-speed, rainfall etc. and yield is decided with the aid of using the region and production. In this challenge CNN and a random tree is used for prediction. It will acquire the crop prediction with high-quality correct values. This utility used for authorities' approach is going to analyze the use of preceding facts and supply the float chart. This manner is used because of insecticide advice and fertilizer advice. The Block Diagram of System Architecture is shown in Fig. 2.

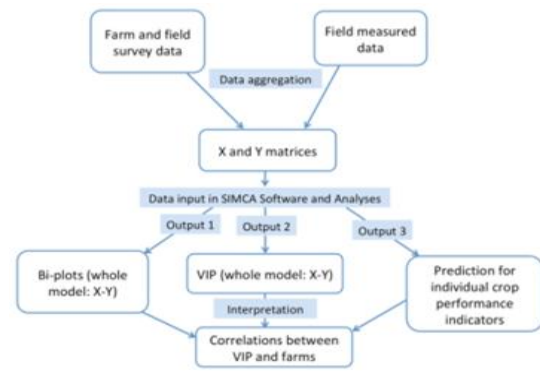


Fig. 2. Block diagram

### 5. Workflow

The working of Crop Yield Prediction involves collecting data from the kaggle website. Collected datasets are trained and tested for crop prediction based on the natural parameters like Humidity, Rainfall, Temperature and soil content like Nitrogen (N), Potassium (K), Phosphorous (P). Summarize the data based on the algorithms and predict the constraints like Crop Type, Amount of Water, Fertilizer and Pesticides needed for the predicted crop. The work flow of the system is shown in Fig. 3.

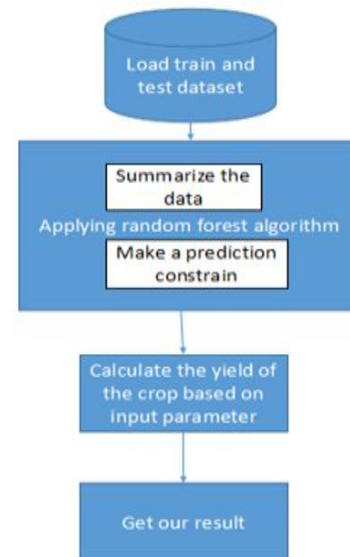


Fig. 3. System flow

### 6. Implementation

We have used PyCharm compiler, Seaborn and Matplotlib for statistics visualization, Pyplot as open source, pandas, numpy, warnings, pickle, requests programs in PyCharm. These are the principal programs that produce the preferred results.

#### A. PyCharm

PyCharm is a committed Python Integrated Development Environment (IDE) presenting a extensive variety of critical equipment for Python developers, tightly incorporated to create a handy environment for effective Python, web, and information technological know-how development.

**B. Seaborn vs. Matplotlib**

Seaborn makes use of charming themes, whilst Matplotlib is used for making simple graphs. Seaborn includes some plots and styles for records visualization, whilst in matplotlib, datasets are visualized with the help of lines, scatter plots, pie charts, histograms, bar-graphs, etc.

**C. Pyplot**

Pyplot is an API (Application Programming Interface) for Python's matplotlib that correctly makes matplotlib a feasible open supply opportunity to MATLAB. Matplotlib is a library for records visualization, normally within the shape of plots, graphs and charts.

**7. Results**

A histogram graph is a bar graph representation of data. It is a representation of a range of outcomes into columns formation along the x-axis. in the same histogram, the number count or multiple occurrences in the data for each column is represented by the y-axis. It is the easiest manner that can be used to visualize data distributions.

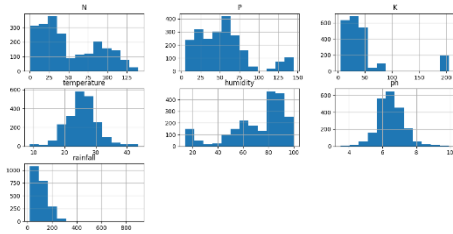


Fig. 4. Histogram graphical representation of attributes range

A correlation matrix is simply a table which displays the correlation coefficients for different variables. The matrix depicts the correlation between all the possible pairs of values in a table. It is a powerful tool to summarize a large dataset and to identify and visualize patterns in the given data. A correlation matrix consists of rows and columns that show the variables. Each cell in a table contains the correlation coefficient.

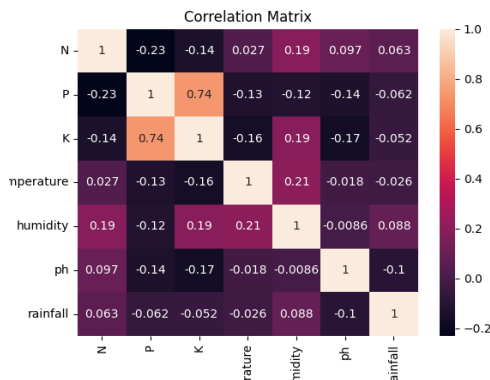


Fig. 5. Correlation matrix of input data

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. The confusion matrix itself is relatively simple to understand, but

the related terminology can be confusing. The most basic terms, which are whole numbers (not rates):

1. True positives (TP): These are cases in which we predicted yes (they have the disease), and they do have the disease.
2. True negatives (TN): We predicted no, and they don't have the disease.
3. False positives (FP): We predicted yes, but they don't actually have the disease. (Also known as a "Type I error.")
4. False negatives (FN): We predicted no, but they actually do have the disease.

Using the Confusion Matrix, we can conclude the following metrics and compare which algorithm is best for predicting the suitable Crop based on the input datasets given by the user. The Metrics are as follows,

1. Accuracy: The best accuracy is 1.0, whereas the worst is 0.0. It can also be calculated by  $1 - \text{ERR}$ . Accuracy is calculated as the total number of two correct predictions (TP + TN) divided by the total number of a dataset (P + N).
2. Precision: The precision value lies between 0 and 1. Recall. Out of the total positive, what percentage are predicted positive. It is the same as TPR (true positive rate).
3. F Score: Metrics derived from the confusion matrix. TN = true negative. FP = false positive. FN = false negative. TP = true positive.
4. Recall: The ratio of correct positive predictions to the total predicted positives. Recall also called Sensitivity, Probability of Detection, True Positive Rate. The ratio of correct positive predictions to the total positive's examples.

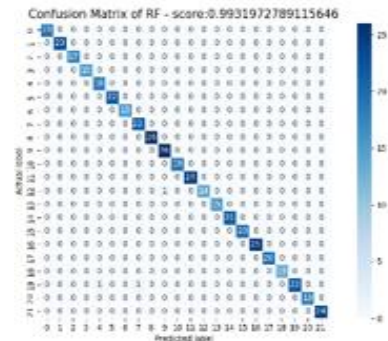


Fig. 6. Confusion matrix



Fig. 7. Home page 1

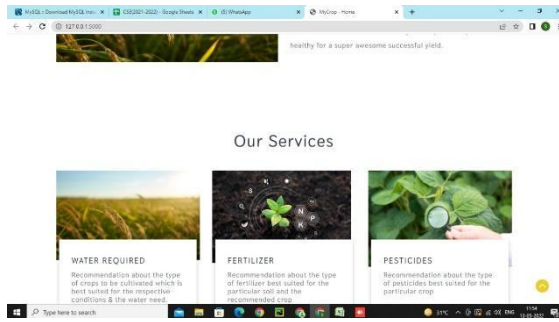


Fig. 8. Home page 2

Agricultural Crop Recommendation

Enter ratio of Nitrogen content in soil:

Enter ratio of Phosphorus content in soil:

Enter ratio of Potassium content in soil:

Enter temperature in degree Celsius:

Fig. 9. Gathering input data from the user

Enter relative humidity in % :

Enter ph value of the soil:

Enter rainfall in mm :

Click here to Recommend

Fig. 10. Gathering input data from the user

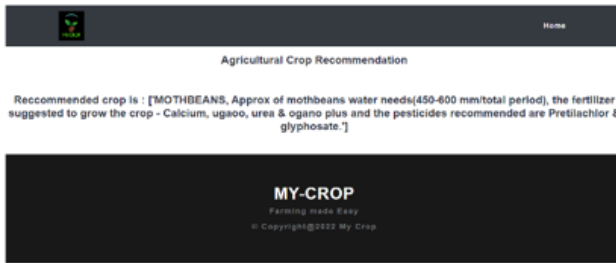


Fig. 11. Recommendation of crop based on the given input data

### 8. Conclusion

India is a state where agriculture performs a high role. In the prosperity of the farmers, the state prospers. Thus, our paintings could assist farmers in sowing the proper seed primarily based totally on soil necessities to boom productiveness and collect income out of any such technique. Thus, the farmer can plant the proper crop growing his yield and additionally growing the general productiveness of the state. Our destiny paintings are aimed toward a stepped forward fact set with a massive quantity of attributes and additionally implements yield prediction.

### 9. Future Enhancement

Crop advice prediction is a vital venture for the decision-makers at countrywide and local levels (e.g., the EU level) for speedy decision-making. A correct crop advice prediction version can assist farmers to determine on what to develop and whilst to develop. There are one-of-a-kind tactics to crop advice prediction.

### References

- [1] Aruvansh Nigam, Saksham Garg, Archit Agrawal, "Crop Yield Prediction using ML Algorithms," 2019.
- [2] Leo Brieman, "Random Forests," 2019.
- [3] Priya, P., Muthaiah, U., Balamurugan, M., "Predicting Yield of the Crop Using Machine Learning Algorithm," 2020
- [4] Mishra S, Mishra D, Santra G. H, "Applications of machine learning techniques in agricultural crop production," 2020
- [5] Y Jeevan Kumar, "Supervised Learning Approach for Crop Production," 2020.
- [6] Ramesh Medar, Vijay S, Shweta, "Crop Yield Prediction using Machine Learning Techniques," 2018.
- [7] Ranjini B. Guruprasad, Kumar Saurav, Sukanya Randhawa, "Machine Learning Methodologies for Paddy Yield Estimation in India: A Case Study," 2019.