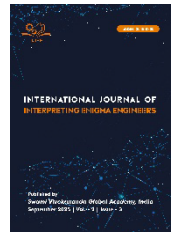




CROP YIELD PREDICTION AND FERTILIZATION OPTIMIZATION USING MACHINE LEARNING



Purvi Lakhota¹, Keshav Kant², K.Uday³, M.Yoshitha⁴, V.Varshita⁵, S. Mothi Sree⁶

Original Article

¹UG Scholar, CSE-AIML, Stanley College of Engineering and Technology for Women(A), Hyderabad, India.

²⁻⁶Computer Science and Engineering, Sree Vidyanikethan Engineering College, Tirupati, India.

School of Computing, Mohan Babu University, Tirupati, India.

*Corresponding Author's Email: lakhotiapurvi@gmail.com¹, keshavkant7859@gmail.com², k.udayuday777@gmail.com³, yoshithamoolinti@gmail.com⁴, varshitaachowdary21@gmail.com⁵, mothisreesannapaneni@gmail.com⁶

Abstract

Crop selection presents a number of complex issues for the agricultural industry, requiring an informed and flexible strategy for the best possible production and resource management. Farmers struggle with the complex variety of soil properties, the ever-changing dynamics of weather patterns, and the never-ending quest to maximize crop yield while consuming the fewest resources possible. Without data-driven solutions, conventional approaches or trial-and-error techniques frequently result in less-than-ideal crop selections, lower yields, and financial losses for farmers. To develop a platform that can read and analyze a variety of environmental parameters, the suggested solution promotes the fusion of state-of-the-art technology and data analytics. This approach aims to produce customized suggestions for farmers by combining soil analysis data, weather forecasting algorithms, and historical crop performance. With the help of these suggestions, they should be able to confidently make strategic planning decisions that are in line with the complex needs of their unique agricultural environments. In conclusion, the suggested solution promotes a creative strategy that uses data analytics and technology breakthroughs to provide farmers with precise, tailored, and flexible crop suggestions. This solution seeks to transform agricultural decision-making processes by providing farmers with practical insights, promoting increased productivity, resource efficiency, and long-term farming profitability.

Keywords – *Crop selection, soil analysis, weather forecasting, yield optimization, resource management, smart farming, agri-tech, historical crop performance.*

Introduction

Since agriculture produces a large amount of food, it is one of the primary areas of societal concern. Many countries are still experiencing hunger today as a result of a lack of food and a growing population. The combined effects of soil loss, climate change, natural weather unpredictability, and population increase necessitate methods to guarantee agricultural development and output in a timely and dependable manner. It must also help increase the sustainability of agricultural food production. According to these criteria, predicting crop yields, protecting crops, and assessing land are more crucial to the global food forecast.

In order to improve national food security, policymakers must be able to make well-informed judgments about export-import assessments, which requires precise crop yield forecasts. However, forecasting agricultural output is difficult because of many intricate elements. In essence, a variety of elements, such as landscapes, soil quality, pest infestations, genotype, water quality and accessibility, climate, harvest planning, and so on, affect crop yield. Crop yield methods and processes are essentially nonlinear and time-specific. These tactics are also complicated due to the inclusion of numerous interconnected elements that are characterized and impacted by external and non-arbitration factors. In the

past, farmers made important cultivation decisions based on their predictions of crop output, which they derived from their own experiences and reliable historical data. Mostly farmers face a multitude of intricate challenges when it comes to deciding which crops to cultivate on their agricultural land. These difficulties include the always shifting and unexpected weather patterns, the various and varied characteristics of soils, and the ongoing requirement to optimize crop yield while consuming the fewest resources possible. Farmers sometimes rely on trial-and-error techniques or conventional wisdom in the absence of data-driven advice, which can result in less-than-ideal crop selections, lower yields, and monetary losses. A complete system that gives farmers precise and tailored crop suggestions based on their unique environmental conditions is therefore desperately needed. This system can enable farmers to make well-informed decisions that maximize output, preserve resources, and improve overall profitability by utilizing cutting-edge technologies and data analytics.

The main focus of the problem statement is the use of machine learning techniques to predict crop productivity. The project's objective is to assist users in selecting an appropriate crop to cultivate in order to optimize yield and, consequently, profit. The suggested system analyzes structured data to create predictions and attempts to address the shortcomings of current systems. In order to develop a crop and obtain a better selection of crops that can be grown throughout the season, we are suggesting that a system be designed that takes into account the most important factors.

This would lessen the challenges farmers experience in choosing crops that will yield a lot of money and optimize profitability, which will lower the suicide rate. There are two primary modules in the system: i. Module for Yield Prediction ii. Crop Suggestion System iii. The weather report.

The goal of this research is to forecast crop production under certain weather conditions in order to suggest appropriate crops for that field. The following steps are involved:-

- i. Gather rainfall, crop yield, soil type, and weather data, combine them in a structured format, and prepare the data. Data cleaning improves data quality and, thus, overall productivity by eliminating erroneous, lacking, and irrational data.
- ii. Select a suitable machine learning model to recommend crops. The Random Forest Classifier was chosen in this instance because of its capacity to manage intricate data relationships and generate precise predictions.
- iii. To estimate the crop yield for specified inputs, separate the analyzed crop data into training and testing sets. Then, train the model using the training data.
- iv. Evaluate several algorithms by running the analyzed dataset through them and determining the accuracy and error rate of each. Select the algorithm that has the lowest error rate and the maximum accuracy.
- v. Put the system into practice as a web application and incorporate the algorithm to get the desired outcomes.

By combining farming and machine learning, we can optimize the use of resources and maximize yield, which will further progress agriculture. Predicting the present yield requires knowledge of the previous year's production data. This project aims to assist farmers by fusing technology and agriculture. An application that is accessible online is the final product. The following functionalities are available in the application: i. Login: By entering their username and password, users can log in on their own. Following registration, users can log in to continue using the application and see all of the available options. One of the modules of the application that allows the user to view crop yield estimates is ii. Yield Prediction. In addition to providing production value and yield per acre, it lets users specify the name of the state, location, crop, and soil type. iii. Crop Suggestion: This is one of the application's modules that lets the user see the crop recommendation.

By entering the state name, soil type, and region name, it provides the crop projection. iv. Sign out: The user can return to the login page by signing out at the conclusion. One of India's primary sources of income is agriculture, and with the rising number of farmer suicides, there is a pressing need to preserve agricultural sustainability. As a result, it

significantly improves the agricultural and economic well-being of nations worldwide.

Literature Survey

A literature survey is a methodical and comprehensive study of all published literature and additional sources, such as dissertations, to find as many articles as possible that are pertinent to a certain topic. In agriculture, forecasting agricultural goods is crucial. It facilitates improved planning, higher net produce, and increased profitability. We read a few research papers on the subject of our project in order to get better outcomes.

According to this research, current agricultural systems primarily rely on hardware, which can be expensive and difficult to maintain while frequently producing inaccurate results. Crop sequencing is recommended by certain systems according to market prices and yield rates. In order to overcome these limitations, this research suggests a software-based method that uses structured data analysis to forecast crops. It emphasizes accuracy by taking into account variables like soil composition, soil type, pH value, and meteorological conditions throughout the prediction process, minimizing maintenance problems as a fully software approach. Both supervised and unsupervised learning methods, including Kohonen's Self Organizing Map (SOM) and Back Propagation Network (BPN), are used in the implementation. Learning networks are used to train the dataset, enabling a comparison of the accuracy of various network learning strategies. After that, end consumers receive the most accurate result. The suggested system evaluates soil quality, forecasts crop production, and makes recommendations for fertilizer based on soil quality. Two controllers process user inputs, including location and pH value. A predefined "nutrients" data store is used to compare the results of Controllers 1 and 2, and the predefined crop dataset is integrated with the compared results in Controller 3 for additional analysis. The accuracy percentage is displayed with the final findings in bar graph form.⁷ The system, which uses both supervised and unsupervised machine learning techniques, produces the best results based on accuracy, the report concludes. The algorithm that produces the most accurate results is determined by comparing the two. By providing effective information for good yields and greater income, this method seeks to lessen farmer difficulties and possibly prevent suicides. Although the report discusses soil and fertilizer experiments, it notes that climate conditions were not taken into account, offering a possible direction for future research.

In order to highlight R's importance as a flexible tool for statistics, data analysis, and machine learning, this study combines R programming with machine learning approaches. R is more than just a statistical package; it is also a programming language that makes it easier to create unique objects, functions, and packages. It is available on a variety of operating systems due to its cost-free and platform-independent nature. Every dataset used in this study comes from publicly available Indian government records from 1997 to 2013, with an emphasis on various rice-producing seasons such as Kharif and Rabi. A small number of critical elements that have the greatest influence on agricultural yield were selected for the study from the large initial dataset. These elements include of temperature, season, rainfall, and crop production. Two machine learning algorithms—Random Forest and Decision Trees—are compared in this research. Decision Tree: Employs a greedy strategy, making a characteristic selected in the first phase unusable in subsequent steps.⁸ Has a propensity to overfit training data, which could lead to subpar outcomes for unseen data. The study uses ensemble models to reduce constraints, aggregating findings from various models to improve overall accuracy. Random Forest: An ensemble classifier that predicts results by using several decision tree models. uses a distinct, randomly chosen sample of training data with replacement to train each tree. averages results for regression problems and uses the total votes of all trees for class assignment in classification tasks. The paper's methodology consists of dividing loaded datasets into training (67%) and test (33%). summarizing the datasets and determining the mean and standard deviation for pertinent tuples. computing probability and comparing the condensed data list with the original datasets. The highest probability obtained is used to make the prediction, and the accuracy is evaluated by comparing the resulting class value (which ranges from 0% to 100%) with the test dataset. According to the study's findings, the Random Forest algorithm shows promise in accurately predicting crop yields and has benefits in terms of model count and applicability for large-scale crop yield forecasts in agricultural planning. The work admits the restriction of a relatively limited dataset, which may affect forecast accuracy, even when it takes into account meteorological elements like temperature and rainfall. The comparison of Random Forests with decision trees offers important information on algorithm performance. [3] Augusta Sophy and S. Pavani Beulet P. With a focus on current climatic indicators including temperature, humidity, rainfall, and soil moisture, this study explores the application of machine learning, specifically the K-Nearest Neighbors (KNN) algorithm, to forecast Telangana's agricultural output. The goal is to address the

challenges farmers face in increasing production while optimizing end-user quality and affordability. The importance of increasing farmer incomes and ensuring Telangana's sustained growth is emphasized in the report. Temperature, humidity, soil moisture, rainfall, and other factors that affect crop yield were all investigated in this study using data collected in May 2019 from different districts. Predictions were made using the KNN algorithm, which took into consideration the closeness of nearby features. The KNN algorithm is implemented by loading the dataset, setting the initial value for 'k,' calculating the distances between test and training data, classifying distances, choosing the top 'k' rows, and identifying the most common class for yield prediction. According to the study's findings, the KNN algorithm is appropriate and accurate at forecasting agricultural yield for particular Telangana districts. In order to support the suggested model, the study emphasizes the significance of a carefully created dataset and machine learning methods. In order to increase accuracy, future considerations should include adding more variables that affect crop yield and growing the dataset. Significantly, the study compares KNN, Support Vector Machine (SVM), and Linear Regression and finds that KNN is the most accurate and suitable. Although the report analyzes soil and climate characteristics, it makes no suggestions for crops or fertilizer based on these aspects, nor does it identify possible areas for further research.

Proposed Model

Implementing the proposed system would assist farmers in choosing crops that yield more and in improving our nation's agricultural practices. It can also be used to decrease losses incurred by farmers and increase crop yields to obtain better capital for agriculture. Therefore, the proposed system will assist in reducing farmers' difficulties and preventing them from attempting suicide. It will also serve as a medium for providing farmers with the effective information they need to encourage high yield, thereby maximizing profits, which will ultimately reduce suicide rates and lessen farmers' difficulties. In order to maximize crop yield, it is helpful to compare the productivity of various crops. This helps farmers choose the best crop for their chosen area and season, which helps them overcome their challenges in the agricultural industry. As a result, the suggested approach suggests a method for forecasting crop yield. Before planting in order to increase production, the farmer will measure the crop's yield per acre.

Methodology

Data Collection and Preprocessing:

Data source: The process begins with collection data that includes historical data on crop growth, earth properties, weather conditions and crop yield. This data will cover a wide range of environmental conditions and crop types.

Data processing: Clean and preffromous data to handle missing values, Outs and deviations. This step ensures that the computer is suitable for training the learning model.

Machine Learning Model Development:

Functional choice: Identify the relevant features of important data sets to create crop recommendations. Facilities may include nutritional levels for soil, temperature, humidity, rainfall, soil type, weather conditions, weather conditions and geographical rooms.

Model choice: Select a suitable machine learning model for the recommendation of the crop. In this case, the Random One Classifier model is created and trained using scaled training data, with the ability to handle complex conditions in data and provide accurate predictions.

Model training: Train random forest classifies using prepared data sets. This includes dividing data into training and verification sets, assessing model performance through fine-tuning model hyperparameters and cross-satypa.

Algorithms Used:

1. Random forest: Random forest is a versatile and widely used machine learning algorithm known for its simplicity and efficiency, often provides excellent results without a larger hyperparameter setting. This applies to both classification and regression tasks, and creates many decision-making trees and incorporates production for more accurate and stable predictions. In particular, it easily measures the importance of each function in the prediction process. The most important hyperparameters in the random forest work either to increase the power of the future or to increase the speed of the model. Notable parameters include:

2. 'N_estimators': determines the number of trees for the construction of the algorithm.
3. 'Max_features': Node indicates the maximum features considered for the division.
4. 'Min_sample_leaf': indicates the minimum number of leaf knots required to divide the internal node.
5. For speed optimization:
6. 'N_JOBS': Checking the number of processors used for parallel processing.
7. 'Random_state': ensures replication by providing frequent results with similar hyperparameters and training data.
8. 'OOB_SCORE': A cross -validation method When using selected samples, the model assessment increases.
9. Random forestry process includes:
 1. Select samples randomly from the dataset.
 2. Build a decision tree for each sample.
 3. Votes for approximate results.
 4. Choose the most voted prediction as the final production.

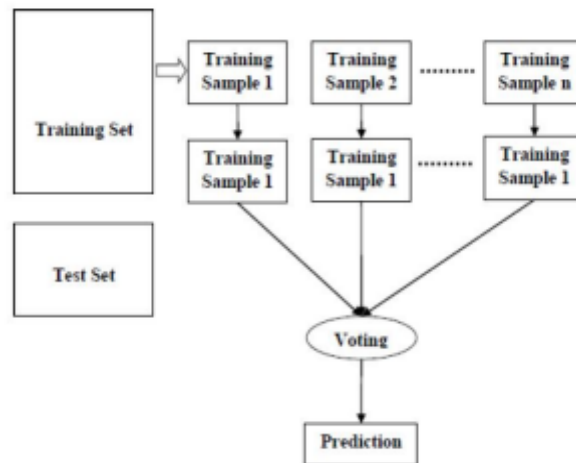


Fig.1. Flowchart of Training and Testing Model

Decision Tree Regression:

Decision Tree Regression is a gadget studying set of rules more often than not used for regression obligations. It's a predictive modelling device that works by recursively partitioning the information into subsets primarily based on the most extensive attribute at each node. Unlike Decision Tree Classification, which predicts express effects, Decision Tree Regression predicts continuous numeric values.

Hyper-parameters in Decision Tree Regression:

Max Depth: Determines the maximum intensity of the tree, proscribing the quantity of nodes and splits.

Min Samples Split: Sets the minimum quantity of samples required to break up an inner node.

Min Samples Leaf: Specifies the minimal quantity of samples required to be in a leaf node.

Max Features: Limits the range of functions taken into consideration for splitting at every node.

Working Process:

Initialization: Start with the whole dataset as the foundation node. **Attribute Selection:** Choose the characteristic that provides the high-quality split based totally on a criterion which includes variance reduction or suggest squared error.

Node Splitting: Divide the dataset into subsets based totally on the chosen attribute. **Recursive Process:** Repeat the method for every subset till achieving a preventing criterion, creating a tree structure.

Prediction: For a brand new information factor, traverse the tree following the found out rules to a leaf node and output the continuous value related to that leaf.

Architecture

The architecture of a Crop Recommendation System typically involves several components working together to provide accurate and relevant crop suggestions based on various factors. Below is a simplified representation of the architecture:

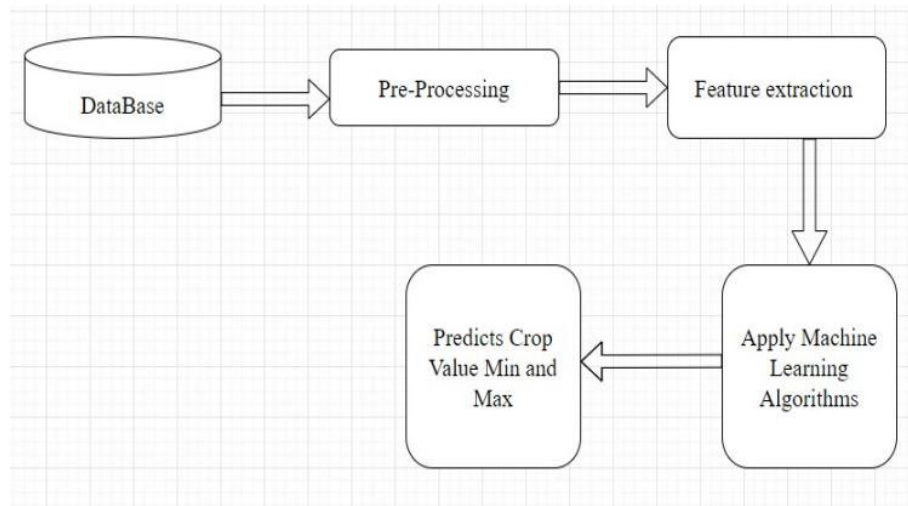


Fig. 3. Graph of Architecture of Model

MODULES:

Data Preparation:

1. This is the first real step towards the real development of a machine learning model, collects data. This is an important step that the model will be, the more and the better the data we get, the better.
2. There are many techniques for collecting data, such as scraping the network, manual intervention and etc.
3. The dataset used in this crop is taken from another source in India.
4. In an established systematic way, the data collection is the process of collecting and measuring information about the variable of interest, which allows someone to answer research questions, test the hypothesis and evaluate the results.
5. The data collection component in research is common for all fields of study, including physical and social science, humanities, business, etc., while the methods are different by discipline, the emphasis is to ensure accurately and honest collection.

Calculate Yield of Production:

In this project, vegetation fee is calculated through the first-rate of the crop identified the use of ranking method. By this manner the min and max fee of crop manufacturing is likewise notified. The significance of crop manufacturing is associated with harvested regions, returns in keeping with hectare (yields) and quantities produced. Crop yields are the harvested production in step with unit of harvested place for crop merchandise.

Predict Crop Value:

In this module the crop price is predicted via applying system getting to know algorithms to the accumulated and train information. So that we will know the crop min max price of the crop at any precise place i.E. Primarily based on the enter.

Accuracy on Test Set:

We got an accuracy of 90.7% on test set.

System Development:

Net app: Develop a user -friendly web application using bottles, HTML, CSS and Bootstrap. The application will act as a user interface to enter data and get crop recommendations.

Integration: Integrates trained machine learning models into web applications. Make sure the application can accept the user entrance for different environmental factors.

User Interface and Input Features:

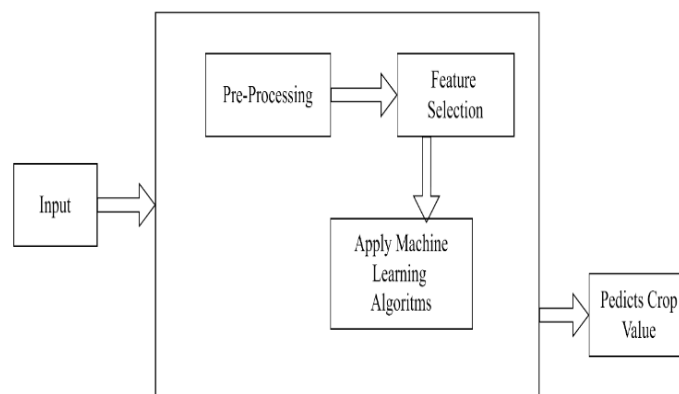
User entrance form: Create an intuitive form in the web application where farmers can provide information on their agricultural conditions, including down -to -earth levels (nitrogen, potassium and phosphorus), temperature, moisture, rain, rain, rain, Earth's pH, soil type, weather conditions, weather conditions and room for crop room cultivation.

Data verification: To use computer information mechanisms, to ensure that users are within valid limits and meet the necessary criteria.

Crop Recommendations:

Prediction: When a user submits the input, you must send them in a random classification model to predict. The model will analyze the data and predict the 5 5 crops that are most likely to flourish under the stated conditions.

Data Flow Diagram:



Results and Discussions

In the final implementation of the application the first screen the user can view is the login page. Here, the user can login using his/her credentials into the application

i) Yield Prediction: The system takes the required inputs to predict the yield of the given crop. The inputs to be given are state name, crop, area, soil type.

ii) Crop Prediction: For this module the system takes the required inputs i.e., soil type and area as seen in the Fig 6.5. The system returns a screen with the crop name.

iii) Weather report: For this module the system gives the weather report of the region for today and tomorrow. Many algorithms were explored and the error rates and accuracy were checked for each. From the above table, we can conclude that the Random Forest Algorithm gives the best accuracy for our dataset.

S.No	Algorithm	Accuracy
1	Multivariate linear regression	73.3%

2	Support Vector Machine (SVM)	75.9%
3	Artificial Neural Network (ANN)	86.5%
4	K Nearest Neighbours (KNN)	90.2%
5	Random Forest	9..65%

The prediction and recommendation system for crop is a bright future ahead of them. It is expected that deep learning models and other advanced machine learning approaches will improve the accuracy of prognosis. It will be necessary to integrate IoT units for monitoring real -time and optimization of environmental status. Because geophysical technology takes into account site -specific data, it will help improve regional predictions. It is estimated that a number of data sources and safety use of blockchain will increase the data set and guarantee openness. Integration of automated machines and expansion of these systems to estimate and reduce crop diseases will maximize the use of use. The extensive adoption facility will be made by the manufacture of user -friendly mobile applications and collaboration with agricultural technical companies. Use a chatbot powered by AI to comply with accurate agricultural methods and provide immediate help.

The direction that these systems take will be affected by government programs that promote digital changes in agriculture. Farmers must participate in educational programs to use technology properly. It is expected that these systems will spread all over the world as they develop further, support different types of agricultural methods and persecuting durable agricultural techniques. The future of these systems will be shaped by ongoing research, technological development and collaborative activities that will solve global food security problems and increase sustainable agricultural practices.

Conclusion

In short, using state -art -art technologies, this effort is a big step towards changing agriculture. Inclusion of machine learning methods, such as random forest, provides an opportunity to improve the prediction's accuracy of the crop yield. This helps farmers solve problems around food security, as well as helping them adapt the agricultural methods. Including environmental factors such as temperature, rainfall and soil moisture suggests that agriculture is widespread. Farmers can enter significant data, get estimates and make educated decisions on crop production with the project's user -friendly web application. Applied Machine Learning models have proved reliable and effective, which appears from a high accuracy of 90.7% on the test set. The architecture of the system is distributed for future improvement. Opportunities include deeper learning and use of other sophisticated machine learning techniques, as well as to incorporate real -time data from the Internet of Things devices to produce more dynamic forecasts.

The prediction and recommendation system for crop dividends becomes an important tool for farmers as climate change and increasing demand continues difficulties for agriculture. The initiative optimizes resource use, takes into account environmental issues and provides adapted insight to support permanent agricultural methods. The achievement of this project paves the way for further successes, teamwork and extensive implementation of technically oriented solutions in the agricultural sector..

References

- [1] Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya R L , International Research Journal of Engineering and Technology (IRJET) Volume 05, Issue 02, Feb-2018
- [2] P.Priya, U.Muthaiah & M.Balamurugan, International Journal of Engineering Sciences & Research Technology (IJESRT), April, 2018.

- [3] S. Pavani, Augusta Sophy Beulet P, International Journal of Engineering and Advanced Technology (IJEAT) Volume-9, December 2019.
- [4] Ramesh Medar, Vijay S.Rajpurohit, Shweta “Crop yield Prediction using Machine Learning Techniques” 2019 IEEE 5th International Conference for Convergence in Technology (I2CT).
- [5] Niketa Gandhi, OwaizPetkar, Leisa J Armstrong “Rice crop yield prediction using Support Vector Machines” 2019 IEEE Technological Innovations in ICT for Agriculture and Rural Development.
- [6] J.P. Singh, M.P. Singh, Rakesh Kumar and Prabhat Kumar “Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique”, International Journal on Engineering Technology, May 2015.
- [7] Prof. D. S. Zingade, OmkarBuchade, NileshMehra, ShubhamGhodekar, ChandanMehta “Crop Prediction system using machine Learning”2020.
- [8] S. S. Kale and P. S. Patil, "A Machine Learning Approach to Predict Crop Yield and Success Rate," 2019 IEEE Pune Section International Conference (PuneCon), Pune, India, 2019.
- [9] Gour Hari Santra, Debahuti Mishra and Subhadra Mishra, Applications of Machine Learning Techniques in Agricultural Crop Production, Indian Journal of Science and Technology, October 2016. 20
- [10]Karan deep Kauri, Machine Learning: Applications in Indian Agriculture, International Journal of Advanced Research in Computer and Communication Engineering, April 2016