*International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)*

**ISSN: 2456-236X**

Vol. 10 Issue 01 | 2025

# Predictive Modeling for crop Recommendation using Machine Learning Algorithm

Dr. J.M.Patil[1], Vaishnavi. S. Kanherkar [2]

[1] *Associate Professor, Dept of Computer Science &Engineering, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, Maharastra, India*

[2] *M.E Student, Dept of Computer Science &Engineering, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, Maharastra, India*

**DOI: 10.5281/zenodo.15751505**

## ABSTRACT

***India's economy and employment are heavily reliant on agriculture. The most prevalent issue among Indian farmers is that they fail to select the appropriate crop for their soil conditions. They will experience a significant decline in production as a result. Precision agriculture has been used to alleviate this issue for farmers. Precision agriculture is a contemporary agricultural method that offers the best crop for farmers based on site-specific features by using research data on soil kinds, characteristics, weather, and crop output. Several districts in the Indian state of Maharashtra are the subject of the analysis. This project makes use of a dataset with roughly 2200 rows of data and machine learning algorithms like Random Forest, logistic regression, decision trees. In order to meet the demands of the current socioeconomic crises that many farmers are facing, the crop prediction system aims to provide the best crop options for a farmer. This system's implementation can assist farmers in optimizing their crop production, reducing, costs, and boosting efficiency in agricultural processes. Introducing a machine learning-based Agricultural Crop Suggestion System could radically change agriculture. By assisting farmers in selecting the best crops for their fields, this crop suggestion system will boost output.***

***Keyword: - Machine Learning, Random Forest, Logistic regression, Decision tree, crop recommendation system.***

## 1 .INTRODUCTION

One of the oldest nations currently engaged in agriculture is India. However, as a result of globalization, agricultural practices have changed significantly in recent years. India's agricultural health has been impacted by a number of issues. To restore health, numerous innovative technologies have been developed. Precision farming is one such method. In India, precision agriculture is growing. Precision agriculture refers to "site-specific" farming technologies. We have benefited from efficient inputs, production, and improved farming decision-making. Precision agriculture has made more strides, but there are still certain problems. Numerous techniques are available that suggest inputs for a specific agricultural area. Higher agricultural crop production is the main goal of crop yield estimation, and numerous well-established models are used to boost crop production yield. Due to its effectiveness in a number of fields, including pattern identification, defect detection, forecasting, etc., machine learning is being employed globally these days. When there is a loss under adverse circumstances, the ML algorithms also aid in increasing the crop yield production rate. The crop is subjected to machine learning techniques. selecting technique to lower agricultural yield production losses regardless of distracting surroundings [4].

A single crop failure brought on by insufficient soil fertility, climate change, flooding, insufficient groundwater, and other similar issues devastates the crops, which then impacts the farmers. In some countries, farmers are advised by society to boost the production of particular crops based on local conditions and environmental variables. Since the population has been growing at a much faster rate, crop production estimation and monitoring are required. Therefore, a suitable approach that takes into account the influencing factors must be created in order to improve crop selection with regard to seasonal variance. The best crop for a given plot of land will be suggested by the system. based on soil composition and meteorological factors such pH, humidity, temperature, and rainfall. They are gathered from the meteorological service and government website. The system receives the necessary inputs, including soil pH, temperature, and humidity. The purpose of the suggested approach is to advise farmers on how to increase crop production and recommend the most lucrative crop for the area [1].

## 2 .LITERATURE SURVEY

Agricultural machine learning is a relatively new technology, and numerous studies have been conducted utilizing machine learning to apply the technology to the agricultural sector. By identifying and characterizing driving data consistency and pattern, machine learning enhances machine efficiency by enabling the system to learn on its own without explicit programming.

M. Bajgai et al. [1] The research underscores the superior performance of Support Vector Machines (SVM) over Back-propagation Neural Networks (BPNN) and Random Forests (RF) when it comes to predicting palm oil yield. The ability of SVM to regularize contributes to its robustness in the face of noisy agricultural data, thereby improving its generalization performance. Remote sensing tools such as drones and satellites, when combined with machine learning algorithms, have demonstrated effectiveness in tasks like counting trees, detecting diseases, and monitoring growth stages in oil palm cultivations. S. K. Roy, S. C. Paul et al.[2] The study offers a thorough analysis of the different machine learning methods for predicting agricultural yield. The authors address the benefits and drawbacks of several machine learning techniques, including decision trees, artificial neural networks, support, vector machines, and regression models, while highlighting the significance of crop production prediction in agriculture. The significance of each data source for precise crop yield prediction is also covered in the paper. These data sources include weather, soil, and crop phenology data.

P. Patil, A. Shinde et al.[4] This survey paper examines the application of data mining methods for predicting crop yields, as investigated by the authors. The article emphasizes the importance of forecasting crop yield in farming and the difficulties that farmers encounter as a result of unpredictable weather and pest invasions. N. K. Gupta and V. G. S. Kumar [5] The authors of this conference article examine how different machine learning methods are used to estimate agricultural productivity. The significance of crop yield prediction in agriculture and the possible advantages of applying machine learning algorithms to this task are introduced at the beginning of the paper. The authors then give a summary of the many machine learning methods such as Bayesian networks, decision trees, random forests, artificial neural networks, and support vector machines that are used to forecast agricultural production. The study also covers the different elements that influence crop output, including soil composition, climate, and pest infestations, as well as how machine learning algorithms can be utilized to model and forecast crop yield.

N. Murthy [6] They provide a thorough examination of different machine learning algorithms, such as decision trees Algorithm, neural networks, SVM, and k-nearest neighbor algorithms. In addition, the authors address the pros and cons of every method and offer a comparison of how they perform on various datasets. S. G. S. Saini, S. S. Rajput [7] The article outlines different data mining methods like decision tree, logistic regression, clustering, and neural networks that are employed for predicting crop yield. In addition, it addresses the pros and cons of every approach and underscores how crucial precise crop yield forecasting is to farming.

## 3. PROPOSED SYSTEM

The proposed system will use environmental characteristics including temperature, humidity, soil pH, and rainfall to determine which crop would be best suited for a given plot of land. By creating a user-friendly application that takes into account factors that directly affect cultivation, such as rainfall, temperature, soil type, etc., the proposed solution seeks to address these limitations and provide a greater variety of crops that can be grown throughout the season. The system would also help farmers choose crops more easily and maximize yield, which would lower the rate of crop failures.

### 3.1 Data Collection

The most effective technique for gathering and analyzing data from many sources, such as official websites, is data collection.to obtain a system's estimated dataset. The following characteristics must be included in this dataset as shown in figure1. For crop forecast, factors including soil pH, temperature, humidity, rainfall, and NPK levels will be taken into account.

.

| | N | P | K | temperature | humidity | ph | rainfall |
|---|---|---|---|---|---|---|---|
| 1 | 90 | 42 | 43 | 20.87974371 | 82.00274423 | 6.502985292000001 | 202.9355362 |
| 2 | 85 | 58 | 41 | 21.77046169 | 80.31964408 | 7.038096361 | 226.6555374 |
| 3 | 60 | 55 | 44 | 23.00445915 | 82.3207629 | 7.840207144 | 263.9642476 |
| 4 | 74 | 35 | 40 | 26.49109635 | 80.15836264 | 6.980401905 | 242.8640342 |
| 5 | 78 | 42 | 42 | 20.13017482 | 81.60487287 | 7.628472891 | 262.7173405 |
| 6 | 69 | 37 | 42 | 23.05804872 | 83.37011772 | 7.073453503 | 251.0549998 |
| 7 | 69 | 55 | 38 | 22.70883798 | 82.63941394 | 5.70080568 | 271.3248604 |
| 8 | 94 | 53 | 40 | 20.27774362 | 82.89408619 | 5.718627177999999 | 241.9741949 |
| 9 | 89 | 54 | 38 | 24.51588066 | 83.53521629999999 | 6.685346424 | 230.4462359 |
| 10 | 68 | 58 | 38 | 23.22397386 | 83.03322691 | 6.336253525 | 221.2091958 |
| 11 | 91 | 53 | 40 | 26.52723513 | 81.41753846 | 5.386168788 | 264.6148697 |
| 12 | 90 | 46 | 42 | 23.97898217 | 81.45061596 | 7.50283396 | 250.0832336 |
| 13 | 78 | 58 | 44 | 26.80079604 | 80.88684822 | 5.108681786 | 284.4364567 |
| 14 | 93 | 56 | 36 | 24.01497622 | 82.05687182 | 6.98435366 | 185.2773389 |
| 15 | 94 | 50 | 37 | 25.66585205 | 80.66385045 | 6.94801983 | 209.5869708 |
| 16 | 60 | 48 | 39 | 24.28209415 | 80.30025587 | 7.042299068999985 | 231.0863347 |
| 17 | 85 | 38 | 41 | 21.58711777 | 82.7883708 | 6.249050656000005 | 276.6552458999995 |
| 18 | 91 | 35 | 39 | 23.79391957 | 80.41817957 | 6.970859754 | 206.2611855 |
| 19 | 77 | 38 | 36 | 21.8652524 | 80.1923008 | 5.953933276 | 224.5550169000003 |

**Fig -1** Dataset

### 3.2 Data Preprocessing

Following the collection of datasets from multiple sources. Prior to training the model, the dataset needs to be preprocessed. Reading the gathered dataset is the first step in the data preprocessing process, which then moves on to data cleaning. During the process of data cleaning, any redundant attributes present in the datasets are disregarded for the purpose of crop prediction. To achieve better accuracy, we must eliminate unwanted characteristics and datasets with missing values. This means we need to either discard the missing values or replace them with appropriate values instead of using unwanted values.

### 3.3 Data Analysis

The correlation matrix was applied to visualize the dataset attributes' correlation coefficient. This highlights that there are significant relationships among P and K, humidity and temperature, humidity and nitrogen, humidity and rainfall as shown in figure2.

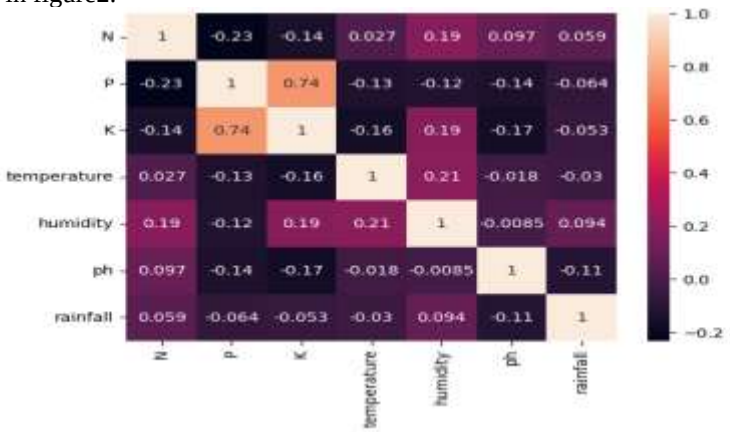


Fig 2 correlation matrix

### 3.4 Feature Extraction

Characterizing a large collection of data would require less data if the features were extracted. The final training data collection is determined by the soil, crop, and meteorological variables gathered throughout the pre-treatment phase.

### 3.5 Data Prediction

The data must be divided into train and test datasets prior to this stage. The data is trained using the provided input and output data by using the Naïve Bayes Gaussian classifier. During the test phase, the data are examined to see if the model's accuracy is satisfactory. The machine learning module then makes predictions based on the new data. Thus, based on the input values, it becomes perceptible which crop will yield better, as shown in fig 3.


Fig 3 Predicted crop (Output)

## 4. MACHINE LEARNING ALGORITHMS

Applications of artificial intelligence (AI), like machine learning, allow systems to independently learn from experience and improve without requiring explicit programming. The primary aim of machine learning is to develop software that can retrieve data and use it for self-directed learning.

### 4.1 Supervised Learning

The machines can forecast the output by using well labelled training data, which consists of input that has been assigned to the correct output. The supervised learning approach fundamentally provides the machine learning algorithm with the correct input and output data. Supervised learning algorithms aim to connect the input variable (x) with the output variable by identifying a mapping characteristic or other comparable mechanism (y).

### 4.2 Unsupervised Learning

Unsupervised learning, as its name suggests, is a category of machine learning where models operate without the direction of training datasets. Models, conversely, investigate the available data to clarify hidden patterns and insights. The cognitive process that occurs in the human brain while acquiring a new skill is remarkably similar to this.

### 4.3 Reinforcement Learning

People always endeavor to enhance their interaction with the environment based on past experiences. In the RL setup, an artificial agent attempts to exhibit the same behavior. Given our understanding of the concepts of state, action, environment, reward, and policy, we can define the goal of an RL agent as searching for an optimal policy among a specified set of states in order to maximize long-term reward. The RL, which is the most target-oriented branch of ML, also integrates delayed rewards within highly intricate delayed response configurations where pinpointing the beneficial action across numerous time slots poses a challenge.

## 5. PROPOSED ALGORITHM

### 5.1 Decision Tree

A supervised machine learning approach for regression and classification is called a decision tree. It builds a tree model of a choice and all of its potential outcomes, including utilities, resource costs, and contingencies. Recursively dividing the data into subgroups according to the eigenvalues is how the decision tree algorithm operates. Each subset's members share output values or are members of the same classes in this fashion. A class label or expected value is represented by each leaf node in the tree, while a decision based on one of the features is represented by each interior node. The root node is where the decision-making process begins.

### 5.2 Random Forest

One machine learning algorithm is called Random Forest. It falls within the category of supervised learning methods. It can be applied to regression as well as classification. It is predicated on the idea. Summary. Combining several classifiers to solve a challenging problem and enhance model performance is known as ensemble learning. Compared to other methods, the random forest classifier requires less training time and produces highly accurate output predictions, even for big datasets.

**5.3 Logistic Regression:**

A mathematical technique for forecasting the likelihood of binary responses based on one or more independent factors is called logistic regression. This indicates that logistic regression is used to predict an outcome with two values, such as zero or one, pass or fail, or yes or no, given certain factors. etc. Logistic regression is a prognostic analysis, just like the other regression models.

**5.4 Support Vector Machine (SVM)**

Support Vector Machine is a method used in supervised machine learning. This is used in both regression and classification analyses. We employ the linear Support Vector Machine classifier in our research as the data is linearly separable and can be graphed in n-employ.

**5.5 K-Nearest Neighbor**

The K nearest neighbor algorithm is the most straightforward one. It can be applied to tackle issues related to classification and regression. Widely utilized in image recognition technology. Simple systems for recommendations and decisions. KNN rely on widely accepted mathematical concepts. The first step in implementing KNN is to convert the data items into their precise values. It operates by indicating the gap between the numeric rate and the numeric rate of points. The way to determine the distance is through the Euclidean distance.

## 6. SYSTEM ARCHITECTURE DESIGN

Figure 4 shows the data flow and system architecture in analysis. The initial stage of the analytical process is data collection. The purpose of Kaggle is to gather huge datasets. Data pre-processing is thus necessary since, for instance, soil pH, P, and K cannot be zero, null, or any other value. After preprocessing, data sets are prepared, and models are constructed for different crops utilizing the datasets that are available. After that, the data is divided into training and testing datasets; training datasets comprise 80% and 20% of the total data, respectively. The feature selection is now complete. Feature selection is a method for reducing the input variable in your model by utilizing only relevant data and eliminating extraneous data. Next, we employed machine learning methods including decision trees, logistic regression, and random forests. After then, each algorithm's performance is examined, and the most accurate algorithm is selected.

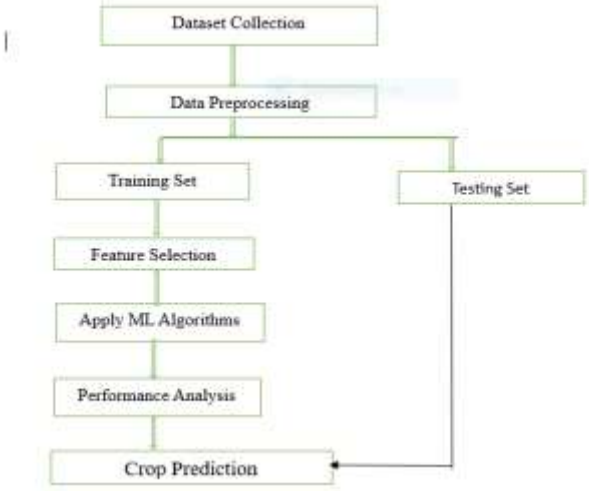


Fig 4 Flow chart of proposed System

## 7. CONCLUSION

The ineffective use of technology and analysis by our farmers at the moment could result in the incorrect crop selection being grown, which would lower their income. A crop recommendation system that can identify the optimal crop for a given plot of land is introduced to lessen those kinds of losses. Several Machine learning methods, including Random Forest, Decision Tree, and Logistic Regression, were applied and evaluated on the provided datasets in order to determine yield to accuracy. The accuracy of the various algorithms is compared. With a 95% accuracy rate, the results show that Random Forest Regression outperforms the other common techniques when applied to the provided datasets. Because of this, farmers are able to choose crops wisely, allowing for the development of the agricultural sector through creative thinking.

ISSN: 2456-236X

Vol. 10 Issue 01 | 2025

**8. REFERENCES**

[1] M. Bajgai, D. Grgicak-Mannion, and A. Mukherjee, "A comprehensive review of crop yield prediction using machine learning methods," Journal of Agricultural Informatics, vol. 12, no. 2, pp. 1-25, 2021

[2] S. K. Roy, S. C. Paul, M. S. Islam, and M. A. Hasan, "Crop yield prediction using machine learning algorithms: A comprehensive review," IEEE Access, vol. 7, pp. 78907- 78926, 2019.

[3] Japneet Kaur, "Impact of Climate Change on Agricultural Productivity and Food Security Resulting in Poverty in India", Università Ca' Foscari Venezia, 2017.

[4] P. Patil, A. Shinde, and M. Kadam, "A survey on crop yield prediction using data mining techniques," International Journal of Advanced Research in Computer Science, vol. 8, no. 1, pp. 421-425, 2017

[5] N. K. Gupta, V. G. S. Kumar, "Crop yield prediction using machine learning algorithms: A review," in 2018 International Conference on Smart Computing and Informatics (SCI), pp. 647-652.

[6] N. Murthy, "Crop yield prediction using machine learning: A review," in 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)

[7] S. G. S. Saini, S. S. Rajput, and S. S. Khillare, "Crop yield prediction using data mining techniques: A review," in 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 1953-1957.

[8] T. H. Shroff and D. R. Patel, "A review on crop yield prediction using machine learning techniques," in 2019 International Conference on Intelligent Sustainable Systems (ICISS)

[9] D. Liu, J. Zhang, and J. Sun, "Crop yield prediction using machine learning: A review," Journal of Physics: Conference Series, vol. 1634, no. 1, p. 012050, 2020.

[10] D. K. Pal and B. Datta, "Crop yield prediction: A review," Journal of Agrometeorology, vol. 16, no. 2, pp. 121-130, 2014.

[11] N. Mishra, and N. Singh, "A survey on crop yield prediction using machine learning techniques," in 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)