Optimizing Crop Recommendations using Machine Learning: A Comparative Study for Enhanced Yield Prediction

Submitted: 21st July 2024; accepted: 17th September 2024

Sanket Gupta, Trishna Panse, Kailash Chandra Bandhu, Ratnesh Litoriya, Shivani Patnaha, Divya Kumawat, Lishika Pargi, Tisha Modi

DOI: 10.14313/jamris-2026-021

Abstract:
For an important segment of the Indian people, agriculture serves as a primary source of income. Most Indian farmers choose to produce crops in a field using traditional farming methods; hence, one of their biggest issues is that they frequently choose to cultivate the incorrect crop for their soil type. The crop recommendation system proposed in this research would assist farmers and educate them on decision-making regarding which crops to plant on their property. Using soil parameters like potassium, nitrogen, and phosphorus as well as environmental variables like humidity, rainfall, and pH levels, to build this recommendation system, we used ML methods such as Random Forest, KNN, Naïve Bayes, SVM, and Logistic Regression. As a result, we also present comparative performance on the model for the dataset. Therefore, finally, these technologies will be helpful for farming and agriculture. Today's smart agricultural solutions, can address the growing concern about the world population's food consumption and environmental impact. The accuracy of this crop recommendation system will depend on the following: The quality and quantity of our dataset, the relevance and effectiveness of our features, the choice and tuning of our machine learning models, the balance of our dataset and the complexity of the crop prediction task, performing thorough training, validation, and testing will give the accuracy metric we need.

Keywords: *Machine Learning, SVM, KNN, Crop Recommendations*

1. Introduction

Since independence, the primary contributor to the country's GDP has been agriculture. Agronomy and their associated industries constituted 59% of the nation's overall GDP during the fiscal year 1950-1951 [1]. Despite the relatively low agricultural productivity, agriculture remains one of the most prominent economic sectors in India. Precision agriculture is one method we may utilize to boost productivity. Applying precise and suitable amounts of soil, fertilizers, and other elements is what precision farming entails, as the term implies. Globalization has, however, caused a significant shift in the agricultural trend in recent years. Numerous factors have adversely affected India's agricultural sector. In order to restore crop vitality numerous innovative technologies have been developed. Precision farming is one such method, applied at the right time to the crop to boost yields and production. Precision agriculture refers to farming technology that is site-specific. But in agriculture, it's important that the guidance provided be precise and correct.

Machine learning, as defined by Arthur Samuel in 1959, is the study of how computers learn without being explicitly programmed. ML algorithms are trained on vast volumes of facts to generate expectations or findings. Recent years have seen a surge in crop prediction research. For example, using IoT and machine learning (ML) technology to improve agricultural decision-making [2]. Another research proposes employing neural networks to create a robust, precise, and clear recommendation system [3].

To anticipate the most productive crop, this work proposes a crop suggestion method that makes source of machine learning algorithms to evaluate soil (pH, phosphorus, nitrogen, and potassium) and meteorology (temperature, moisture, and rainfall) data. The F1 score, Recall, and Precision have been utilized to assess the performance of the approach suggested for every class and method. Crop recommendation systems can assist farmers in selecting which crops to plant, increasing yields and reducing resource consumption. Crop suggestion systems can also increase agriculture's ability to adapt to climate change. The remaining part of this work includes: The literature review and details on the model are provided in Sections 2 and 3. Sections 4 and 5 cover the Experimental Setup.

2. Literature Review

Including important environmental variables improves the dataset which is used in [4]. The dataset contains Temperature, Humidity, pH, rainfall and label which includes sugarcane, coconut, jute, cotton, papaya, groundnut, maize, graphs, rice, mango, rubber etc. It uses an SVM decision tree (Hybrid approach) which maintains an accuracy rate of 91.8% and Random Forest shows an accuracy of 95%. The Table 1 itself serves as the literature review summary.

The methodology, which integrates machine learning with the IoT, is not well recognized. The authors of [2] proposed it as the Crop Monitoring and Recommendation System utilizing sensors to record certain

Table 1. Crop Recommendation Techniques ML

Authors	Methodology	Features	Accuracy	Dataset	Advantages	Limitations
Srilakshmi A.,Madhumitha K., Geetha K[4]	SVM decision tree (Hybrid approach), Random Forest	Temperature, Humidity, pH, rainfall, label	SVM decision tree (Hybrid approach) - 91.8%. Random Forest - 95%	sugarcane, coconut, jute, cotton, papaya, groundnut, maize, graphs, rice, mango, rubber etc	Predict crop for any type of field	Small dataset
R. Pallavi Reddy, B. Vinitha, K. Rishita, K. Pranavi [2020] [2]	Linear Regression Model	N, P, K, and moisture values			generate recommendations to improve crop production and estimates the price of the yield	Limited in capturing non-linear patterns, Assumes homoscedasticity and independence of errors
S. Mamatha Jajur, Soumya N. G. [2019] [5]	KNN, Decision trees, SVM, CNN and LSTM, ANNs, K-means clustering	Soil Type, pH value, NPK content of the soil, Water holding, Temperature, Average rainfall, Previously Harvested crop	-	wheat, rice, bajra, maize, jawar,	select the optimum crop while keeping a number of variables in mind to boost the output of agriculture, minimise the deterioration of the soil in fields that are under cultivation and use less fertiliser when growing crops.	Many algorithms are used
Mr. Santosh Mahagaonkar, Devdatta A. Bondre [2019] [6]	Random Forest, Support Vector Machine algorithm	crop, crop yield dataset, Location, soil and crop nutrients, fertilizer datasets	soil classification, RF -86.35% crop yield prediction SVM -99.47%	Soybean, Rice, Jowar, Wheat, Sunflower, Cotton, Sugarcane, Tobacco, Onion, Dry Chili, etc.	future prediction of crop yield	Low accuracy in soil classification performance heavily depends on parameter tuning and it is memory intensive, particularly for large datasets
D. Anantha Reddy, Bhagyashri Dadore, Aarti Watekar [2019] [7]	Naïve Bayes, K-NEAREST NEIGHBOUR, RANDOM FOREST, CHAID	Depth, Texture, pH, Soil Colour, Permeability, Drainage, Water holding and Erosion	-	groundnut, pulses, cotton, vegetables, paddy, sugarcane, coriander.	Assist farmers in planting the appropriate seed according to the needs of the soil in order to boost output.	The Naïve Bayes algorithm pretends feature independence, which might not be true when dealing with real-world data., CHAID - Limited to categorical target variables and predictors, making it less versatile for handling continuous data
Nidhi H. Kulkarni [8] 2018	Linear SVM algorithms, Random Forest, Naïve Bayes	Soil type, pH soil, NPK, average rainfall, porosity of soil, sowing season temperature	99.9 1%	Cotton, Sugarcane, Rice, Wheat	Crop productivity has improved exponentially for rice, wheat, cotton, and sugarcane.	restricted to a fairly small number of crops

Table 1. Continued

Authors	Methodology	Features	Accuracy	Dataset	Advantages	Limitations
Zeel Doshi [3] 2018	Neural Network Random Forest, Decision Tree, KNN	Temperature rainfall, Location, soil condition	91%	Jute, sesame, soybean, sugarcane, tobacco, sunflower seeds, ragi, potato, tur, grapeseed, and mustard, bajra, maize wheat, rice gram, barley, cotton, groundnut, and pulses	Neural Networks have the highest accuracy percentage.	predict the crop using the harvest from the previous cycle. Crop supply and demand are not considered
Rohit Kumar Rajak [9] 2017	Random Tree, NB-classifier, ANN, SVM	depth, pH, texture, permeability to store water, color of the soil, and drainage from erosion	-	vegetables, rice, sugarcane, sorghum, coriander, bananas, legumes, and groundnuts	boosts agricultural productivity	larger dataset for model training
S. Pudu-malar [10] 2016	Random Tree, Naïve Bayes, KNN, CHAID,	Depth, pH, texture, water-holding permeability, Soil color, erosion drainage,	88%	millet, pulses, groundnut, cotton, banana, vegetables, paddy, sugarcane, sorghum, coriander	Boost productivity	larger dataset for model training
Rakesh Kumar [11] 2015	CSM, Gradient Boosted Decision Tree, and Greedy Forest	soil type, weather, crop type, water density,		ratoi, toria, wheat, potato, sarso, linseed, masoor, khesari, onion, sugarcane, Kanda, mung, til, pumpkin, nenua, ladies' finger, rice, soybean, sweet potato, toor, vegetable seed, and so on	offers a method to select crops while taking into account the yield forecast rate influenced by various factors.	Adopting a prediction technique that performs well and has greater accuracy is necessary.

characteristics of the soil, such as its moisture content and nutrients, and uploading the data to a cloud platform. An Android app receives this data and gives recommendations for crop selection based on soil type among other factors. Furthermore, a price prediction module has been integrated using linear regression. This combined approach is expected to help farmers make good choices and increase farm productivity and profitability aimed at taking into consideration soil health worries.

Authors in [5] have used data including soil type, acidity level, NPK content, permeability, water holding capacity, average rainfall, temperature, as well as previously grown crops. For classification tasks they have supervised learning methods KNN, ensemble learning (EL), and SVM and also unsupervised learning methods (K-means clustering for data analysis).

A technique that farmers throughout India can simply employ is the intelligent crop recommendation system. Three processes are involved in this research: the classification of the soils, crop yield prediction, and fertilizer suggestion utilizing Random Forest and Support Vector Machine, which provide more robust

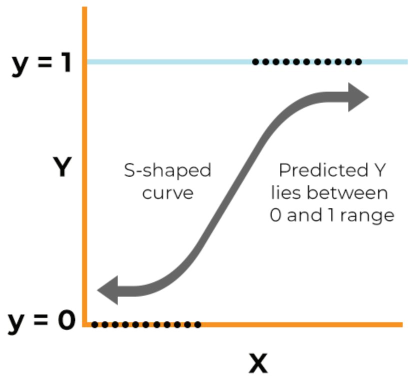

Figure 1. Logistic function [12]

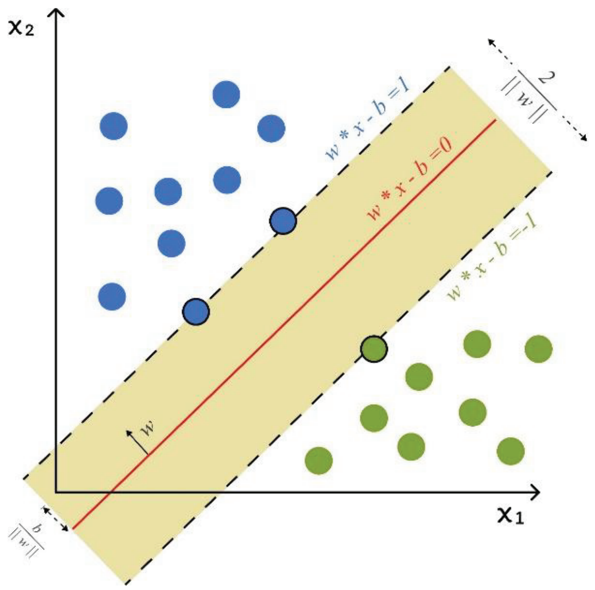

Figure 2. SVM [5]

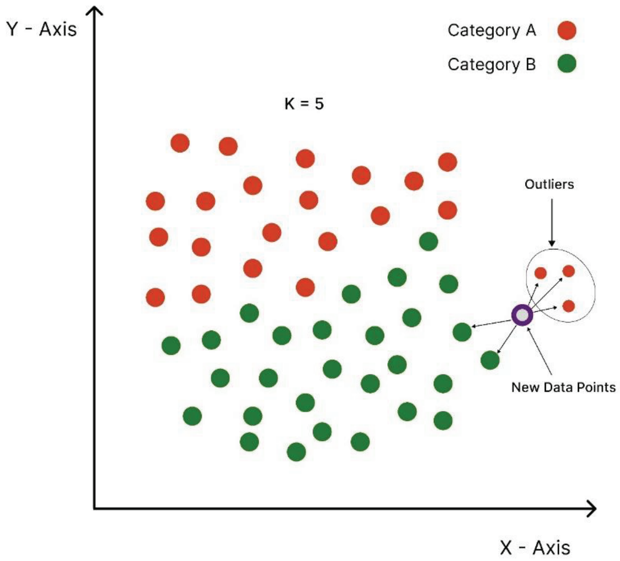

Figure 3. KNN [3, 5]

models than traditional logistic regression model as shown in Figure 1. Additionally, the system has apps from third parties that provide weather data. The result of this experiment reveals that soil classification using Random Forests and crop yield prediction with Support Vector Machines are effective. The future is to improve it by including development of a mobile application for farmer as well as implementing crop disease detection through image processing [6].

The authors in [7] suggested a method so that they can help farmers to make decisions about crops depending on their soil types. It applies soil-specific characteristics such as depth, texture, pH, and water-holding capacity to recommend appropriate crops. It makes use of a group method that incorporates Random Tree, CHAID, Naïve Bayes and KNN machine learning algorithms as demostrated in Figure 3 and 5. This study demonstrates the use of Random Forests for soil categorization and the use of Vector Machines for crop production prediction.

Four crops have been taken into consideration in a crop recommendation system: wheat, cotton, sugarcane, and rice in [8]. Accurate crop selection is provided by crop recommendation systems, which take soil, surface temperature, and rainfall into account. The proposed model uses a high degree of efficiency and accuracy of ensembling approach to predict the crop that will boost yield.

The authors of this work [3] have introduced Agro-Consultant, an intelligent system that Indian farmers may use to make well-informed decisions on which crops to produce by utilizing data on soil properties, geographic location, and climatic parameters like rainfall and temperature. This is accomplished by the utilization of various ML methods, like neural networks, KNNs, Random Forests, and decision trees.

The authors [9] have concluded that crop yield prediction is essential for the country's planned guiding principles made in the field of agriculture development, to provide greater agricultural output and effective use of water resources while assisting farmers in minimising the use of pesticides in crop production and preventing soil deterioration.

Based on the needs of the soil, data mining tools would assist farmers in choosing the best seeds to plant, guaranteeing higher yield to make a profit. To precisely and effectively recommend a crop based on site-specific data, a collaborative recommendation model is built using techniques including the algorithms Random Tree, CHAID, KNN, and Naïve Bayes [10].

To grow the crop's net yield rate, the authors of [11] recommend the Crop Selection Method (CSM), which recommends the series of crops that will be sown throughout the season depending on the crop yield forecast. A solution for crop selection based on factors such as crop type, weather, soil type, and water density is offered by the proposed method. This approach suggests a crop sequence's daily production is maximal for a given season, considering the crop, sowing time, number of days in planting, and season-wise yield rate as inputs.

Machine learning is one of several methods to predict crop production in agriculture. Machine learning techniques such as Random Forests and Support Vector Regression are used, which provide more robust models than traditional linear regression models [12].

3. Model Used

3.1. Logistic Regression

One technique used for binary classification is called Logistic Regression. In this technique we predict the probability of an input example falling into one of two classes. Therefore, the output should be in discrete value i.e., either Yes or No, true or false, etc. It is classified into three basic categories: Binomial, Multinomial and Ordinal.

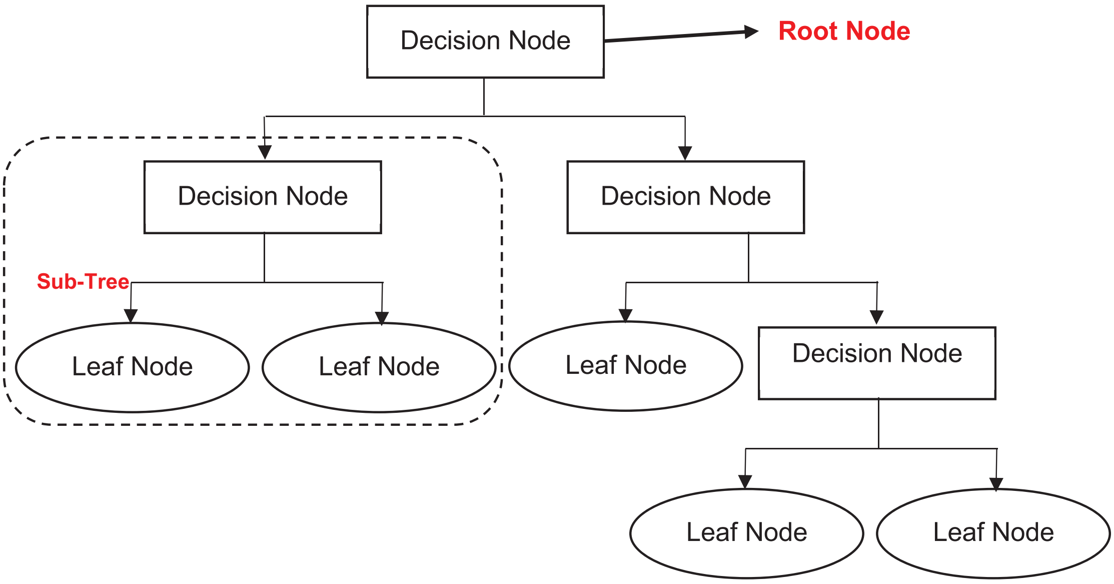

Figure 4. Decision Tree [3]

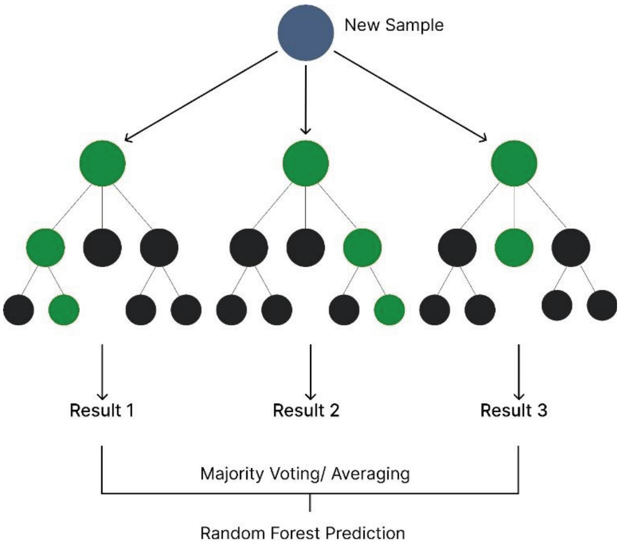

Figure 5. Random Forest [3]

$$y = \frac{1}{1 + e^{-(a_0 + a_1 x)}} \tag{1}$$

- x = input value
- y = predicted output
- a_0 = bias or intercept term
- a_1 = coefficient for input (x)

3.2. Naïve Bayes

One of the most fundamental and successful probabilistic classification methods, Naïve Bayes, is based on the Bayes' theorem. It generates a likelihood table by finding probabilities of the features. Gaussian Naïve Bayes is specifically applied when continuous features follow a Gaussian distribution. It's efficient, simple, and performs well in limited training data.

$$P\,(A \mid B) = \frac{P\,(B \mid A) \cdot P\,(A)}{P(B)} \tag{2}$$

Eq.(2) is Bayes' theorem in which:
$P(A)$= The probability of A occurring
$P(B)$= The probability of B occurring
$P(A|B)$=The probability of A given B
$P(B|A)$= The probability of B given A

3.3. SVM

Support Vector Machine (SVM) is a supervised machine learning technique that may be applied to both regression and classification applications. The objective of this method is to determine the hyperplane that effectively divides the two classes with the widest possible margin. The SVM algorithm utilizes two crucial elements to choose the most suitable hyperplane for classifying labels: two measurements: the support vectors, or the data points closest to the hyperplane, and the margin, or the distance between the hyperplane and the closest data points. Figure 2 highlights the SVM plan and support vectors.

$$w \cdot x + b = 0 \tag{3}$$

Eq.(3) equation of hyperplane in which:
w = a vector normal to hyperplane
b = an offset

3.4. KNN (K-Nearest Neighbours)

KNN, K-Nearest Neighbours is a supervised ML method, is usually used for division but also for regression. This method starts with selecting the nearest Neighbors. Then, on can try different values for K to find the optimal one. The Euclidean distance between K Neighbors is also calculated. We count the number of data points in each category between these K Neighbors, and we assign a new data point to the category with the highest number of Neighbors. Expanding the training data collection could enhance this technique.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{4}$$

Eq. (4) is Euclidean distance formula in which:

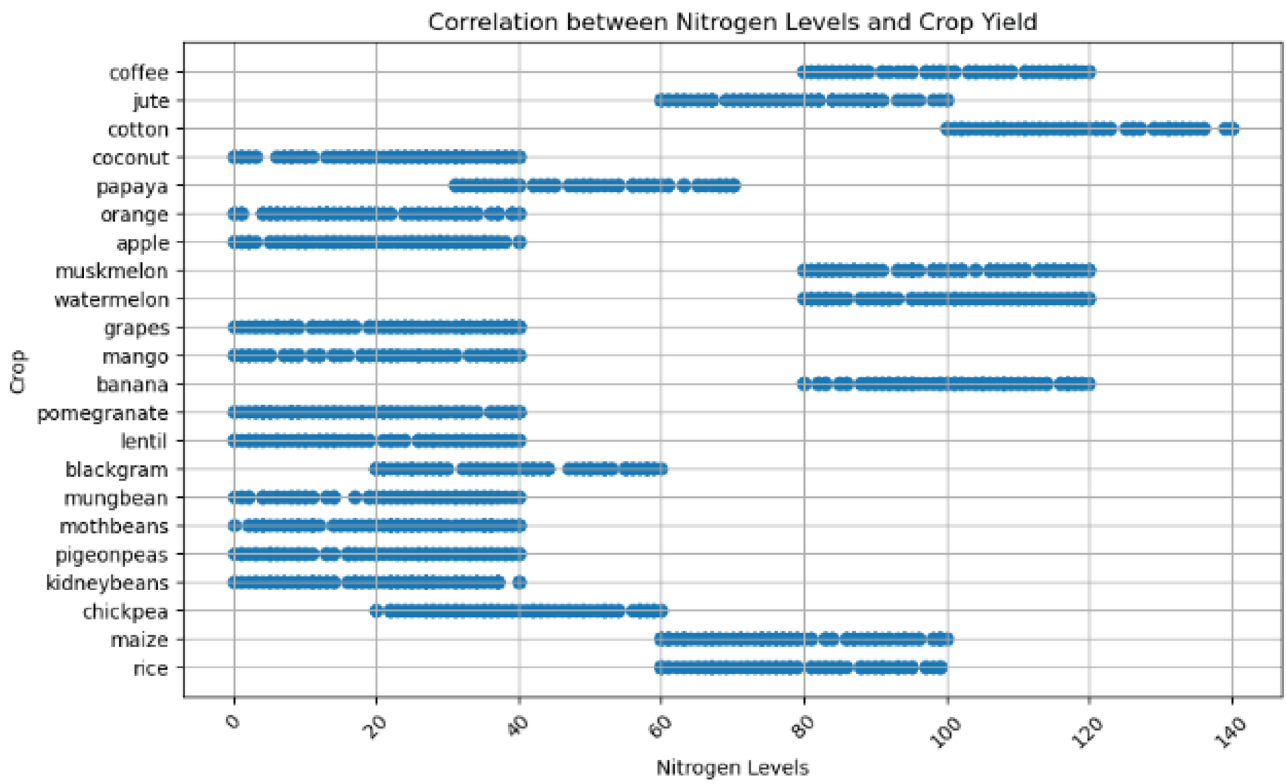

Figure 6. (a). Relationship between Nitrogen Levels and Crop Yield

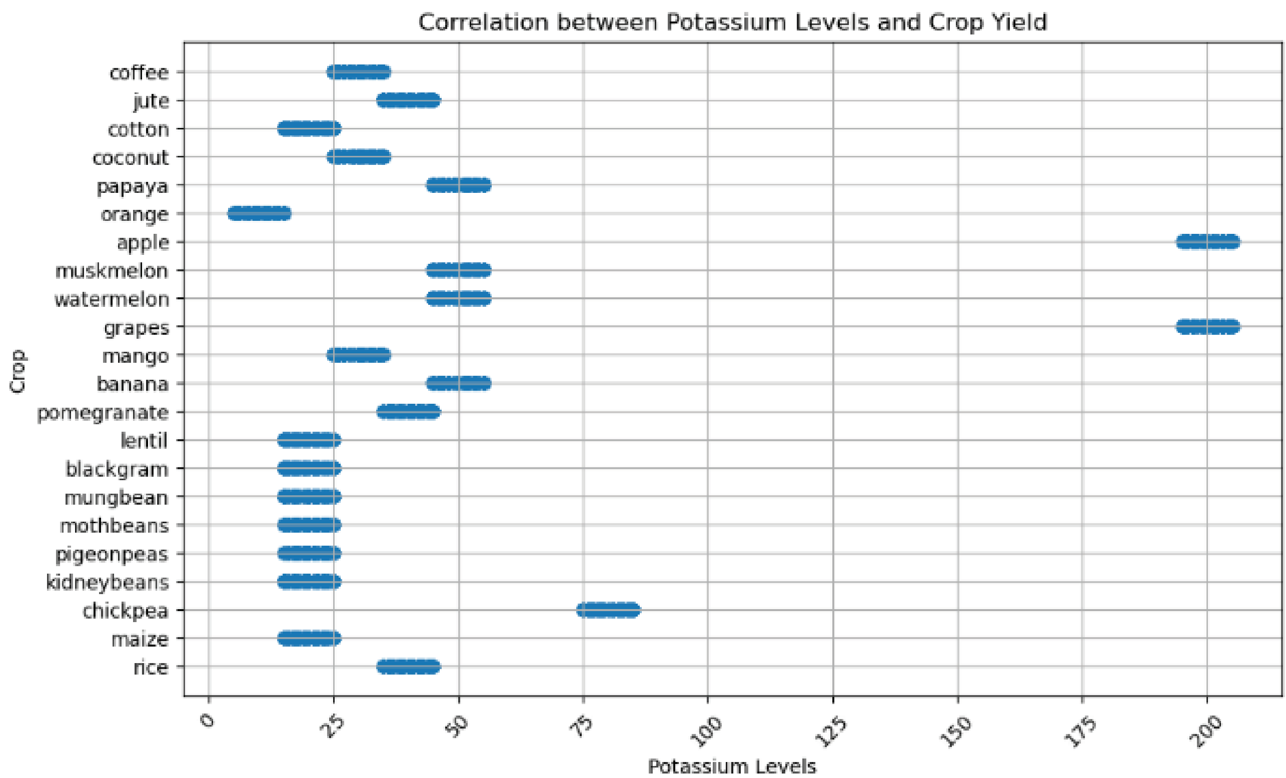

Figure 6. (b). Relationship between Potassium Levels and Crop Yield

The coordinates of one point are (x_1, y_1)
The coordinates of the other point are (x_2, y_2)
Distance between (x_1, y_1) and (x_2, y_2) is d.

3.5. Decision Tree

Although decision trees are among the most powerful tools available, they are typically employed for classification jobs. They can also be utilized for regression assignments. As shown in Figure 4, the Decision

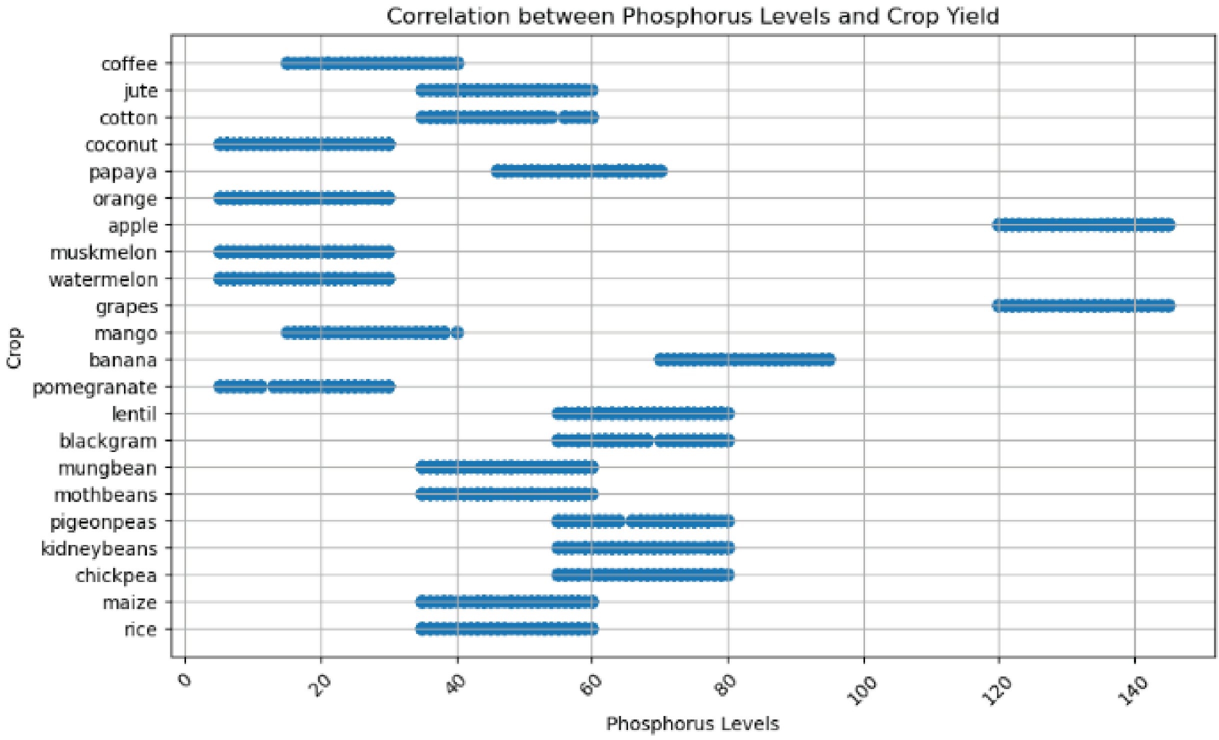

Figure 6. (c). Relationship between Phosphorus Levels and Crop Yield

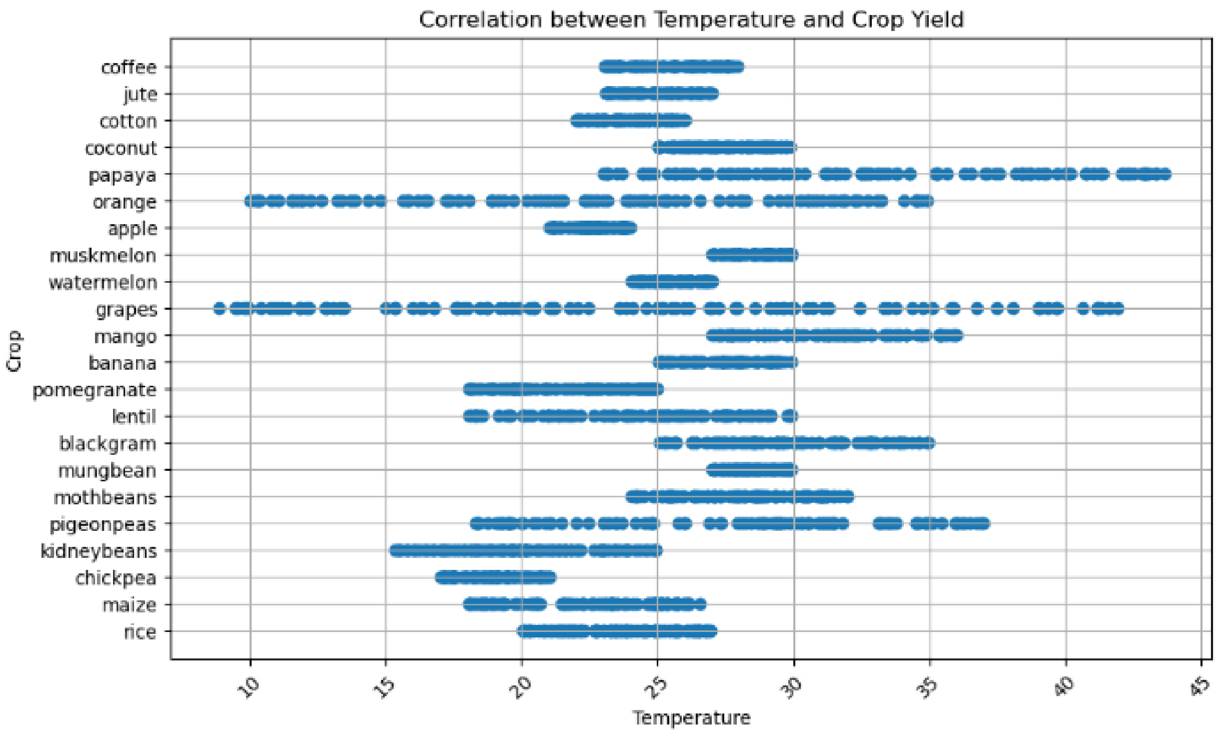

Figure 6. (d). Relationship between Temperature and Crop Yield

Node and Leaf Node are the two nodes that make up this system. Leaf Nodes represent the decision's output, and Decision Nodes are used to make decisions. Decision trees are easy to understand, treatment of both numerical and categorical data though their failure to prune could lead them into over-fit situations especially when noisy datasets are considered.

3.6. Random Forest

Among the widely used ensemble models based on decision trees is Random Forest. Every observation is fed into one of the many decision trees that are generated in Random Forest. Random Forest can deal with high-dimensional data, maintain some level of interpretability, and has less chance of being affected

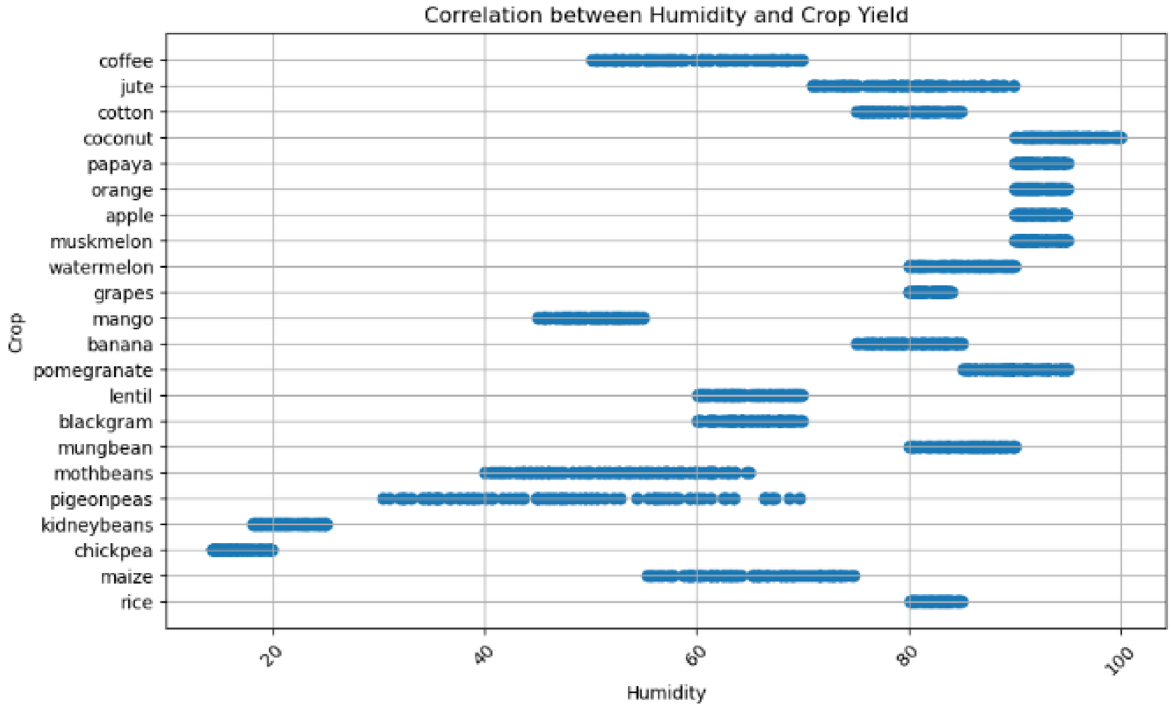

Figure 6. (e). Relationship between Humidity and Crop Yield

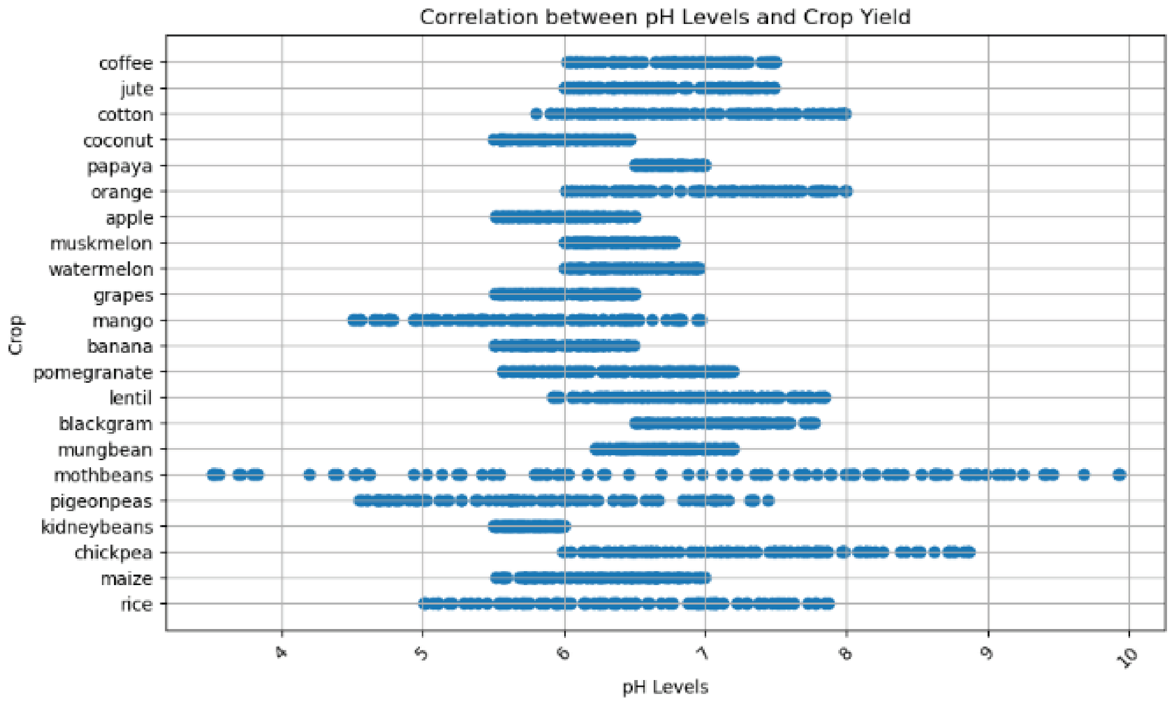

Figure 6. (f). Relationship between pH and Crop Yield

by overfitting than individual decision trees themselves [13].

4. Proposed Method

We aim to develop a CR System by using ML techniques to help farmers select which crop should yield based on soil and climate parameters. The dataset used contains nitrogen (N), phosphorus (P), potassium (K) levels, temperature, humidity, pH, and amount of rainfall, along with the label of crops suitable for the environment and to get a better understanding of the dataset we have visualized the correlation between label and other features.

To ensure data integrity, exploratory data analysis identifies the dataset's dimensions, contents, missing values, and duplicates. Furthermore for smooth flow of data, crop labels were mapped into numerical identifiers. Each crop is associated with a unique numerical value, ranging from 1 to 22. A new column

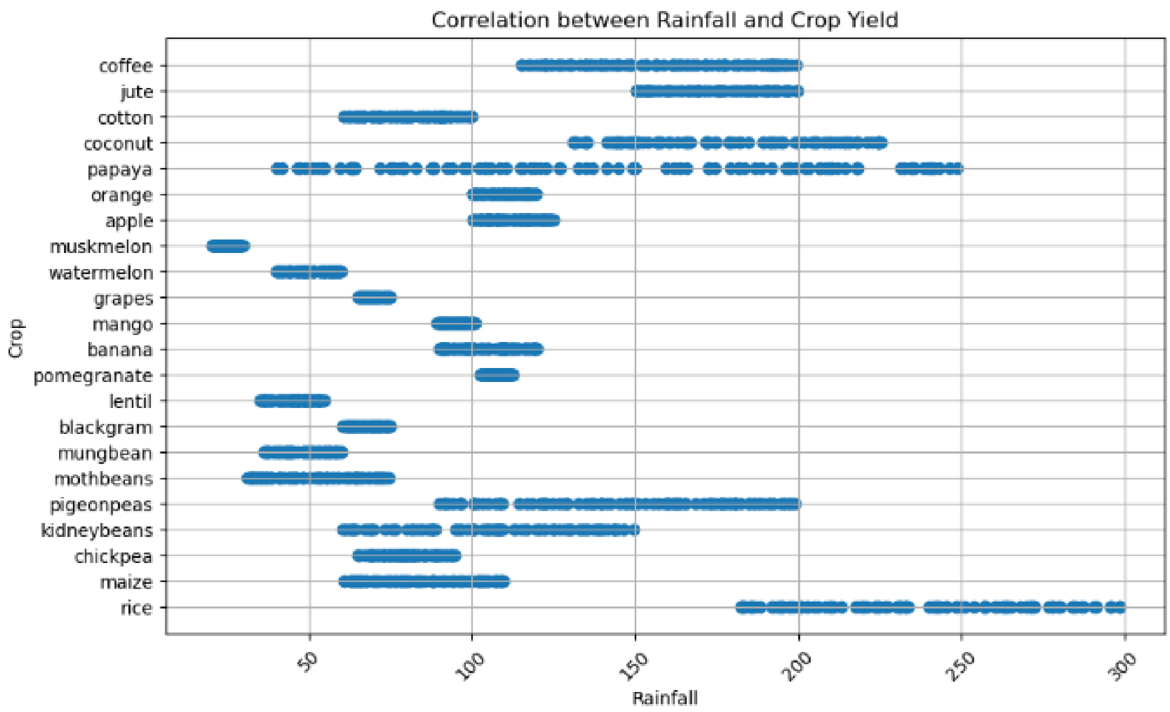

Figure 6. (g). Relationship between Rainfall and Crop Yield

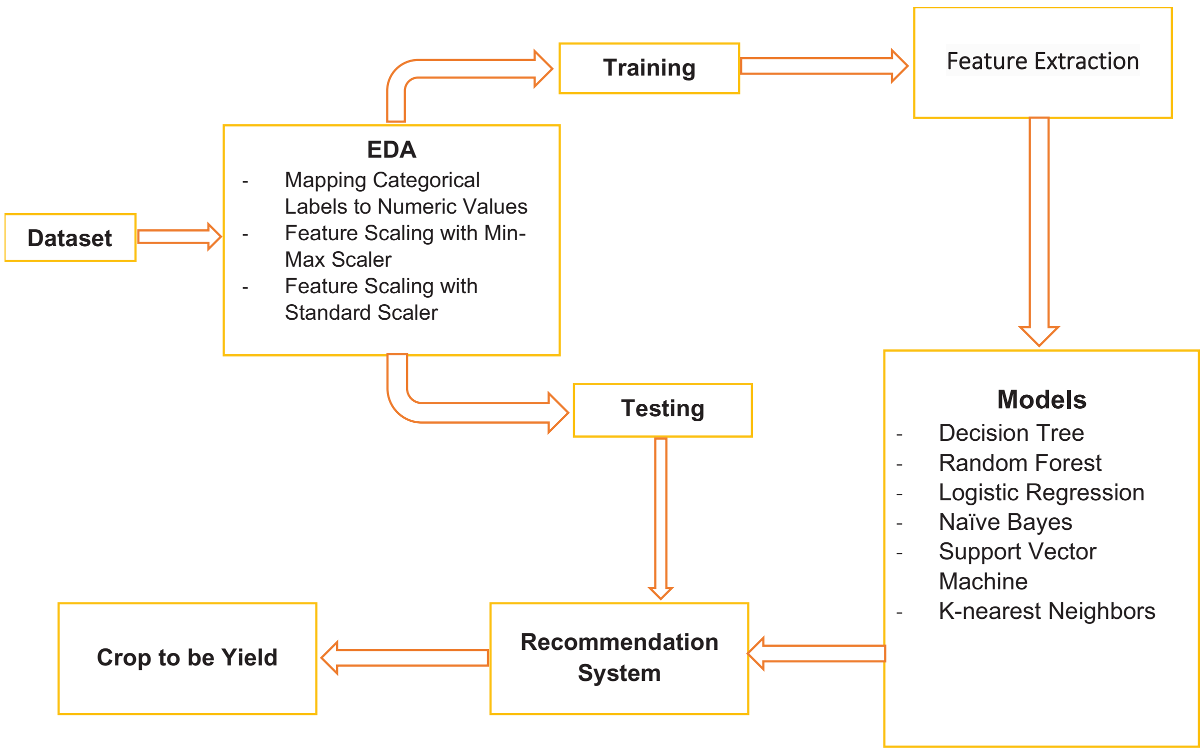

Figure 7. Block Diagram of Crop Recommendation System

'crop_num' is created corresponding to existing column 'label' containing crop names.

The dataset is divided into features (X) containing N, P, K, temperature, humidity, pH and rainfall and labels (y) containing newly created column 'crop_num'. To standardize and normalize the data for model compatibility, we have used Min-Max Scaler. So that it scales and translates each feature individually using the function MinMaxScaler().

The Figure 7 above demonstrates Various Machine learning algorithms were used for the development of the model. Each model's accuracy in crop prediction was evaluate using training dataset (X_train, y_train) after that, the suitable model was trained using a training dataset (X_train, y_train). Naïve Bayes has shown to have the highest accuracy among these models.

For the most accurate crop recommendation as shown in the Figure 8, a Gaussian Naïve Bayes model

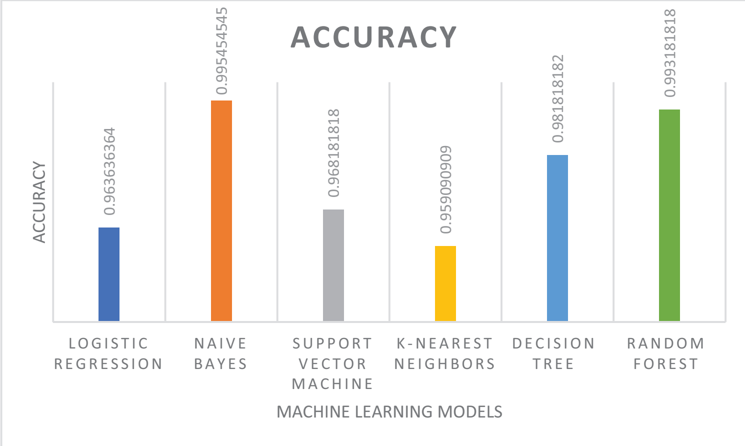

Figure 8. Accuracy Comparison

Table 2. Crop Label and Corresponding Numerical Representation

label	Crop_num
Rice	1
Maize	2
Jute	3
Cotton	4
Coconut	5
Papaya	6
Orange	7
Apple	8
Muskmelon	9
Watermelon	10
Grapes	11
Mango	12
Banana	13
Pomegranate	14
Lentil	15
Blackgram	16
Mungbean	17
Mothbeans	18
Pigeonpeas	19
Kidneybeans	20
Chickpea	21
Coffee	22

is trained on the entire dataset. A recommendation function is developed, which will allow users to input soil and climatic parameters such as nitrogen, phosphorus, potassium levels, temperature, humidity, pH, and rainfall to predict the most suitable crop for cultivation then a trained model is used, and the function simply predicts while mapping the predicted crop number against a predefined dictionary storing its corresponding crop name [14] as displayed in Table 2.

For better understanding of the performance of all models used and to make right decisions we have also generated Confusion Matrix demonstrated in Figures 9 & 9 and Classification report containing Precision, Recall and F1 score of all models (Logistic Regression, Naïve Bayes, SVM, KNN, Decision Tree, and Random Forest) [15].

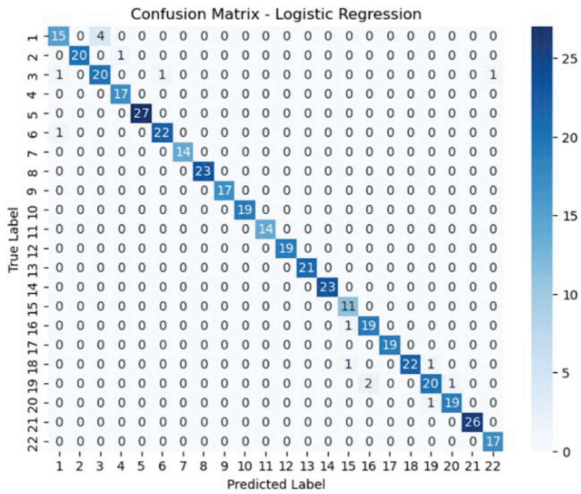

Figure 9. (a). Confusion Matrix - Logistic Regression

Classification Report:

	precision	recall	f1-score	support
1	0.88	0.79	0.83	19
2	1.00	0.95	0.98	21
3	0.83	0.87	0.85	23
4	0.94	1.00	0.97	17
5	1.00	1.00	1.00	27
6	0.96	0.96	0.96	23
7	1.00	1.00	1.00	14
8	1.00	1.00	1.00	23
9	1.00	1.00	1.00	17
10	1.00	1.00	1.00	19
11	1.00	1.00	1.00	14
12	1.00	1.00	1.00	19
13	1.00	1.00	1.00	21
14	1.00	1.00	1.00	23
15	0.85	1.00	0.92	11
16	0.90	0.95	0.93	20
17	1.00	1.00	1.00	19
18	1.00	0.92	0.96	24
19	0.91	0.87	0.89	23
20	0.95	0.95	0.95	20
21	1.00	1.00	1.00	26
22	0.94	1.00	0.97	17
accuracy			0.96	440
macro avg	0.96	0.97	0.96	440
weighted avg	0.96	0.96	0.96	44

Figure 9. (b). Classification Report - Logistic Regression

5. Result

The research presented here compares machine learning approaches used in crop recommendation

Table 3. Before MinMax Scaling

	N	P	K	temperature	humidity	pH	rainfall
1656	17	16	14	16.396243	92.181519	6.625539	102.944161
752	37	79	19	27.543848	69.347863	7.143943	69.408782
892	7	73	25	27.521856	63.132153	7.288057	45.208411
1041	101	70	48	25.360592	75.031933	6.012697	116.553145
1179	0	17	30	35.474783	47.972305	6.279134	97.790725

Table 4. After MinMax Scaling

0.12142857	0.07857143	0.045	0.21723408	0.9089898	0.48532225	0.29685161
0.26428571	0.52857143	0.07	0.53710965	0.64257946	0.56594073	0.17630752
0.05	0.48571429	0.1	0.53647858	0.57005802	0.58835229	0.08931844
0.72142857	0.46428571	0.215	0.47446209	0.708898	0.39001747	0.34576958
0.	0.08571429	0.125	0.76468429	0.39318139	0.43145185	0.2783274

Table 5. Accuracy of ML Models

Models	Accuracy
Logistic Regression	0.9636
Naive Bayes	0.9954
Support Vector Machine	0.9681
K-Nearest Neighbors	0.9590
Decision Tree	0.9818
Random Forest	0.9931

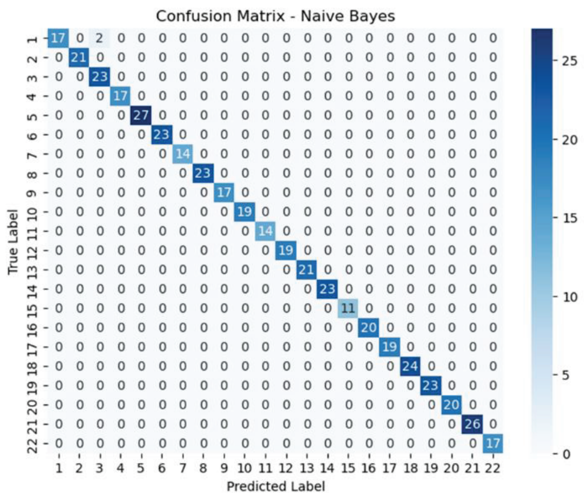

Figure 10. (a). Confusion Matrix - Naïve Bayes

```
Classification Report:
              precision    recall  f1-score   support

           1       1.00      0.89      0.94        19
           2       1.00      1.00      1.00        21
           3       0.92      1.00      0.96        23
           4       1.00      1.00      1.00        17
           5       1.00      1.00      1.00        27
           6       1.00      1.00      1.00        23
           7       1.00      1.00      1.00        14
           8       1.00      1.00      1.00        23
           9       1.00      1.00      1.00        17
          10       1.00      1.00      1.00        19
          11       1.00      1.00      1.00        14
          12       1.00      1.00      1.00        19
          13       1.00      1.00      1.00        21
          14       1.00      1.00      1.00        23
          15       1.00      1.00      1.00        11
          16       1.00      1.00      1.00        20
          17       1.00      1.00      1.00        19
          18       1.00      1.00      1.00        24
          19       1.00      1.00      1.00        23
          20       1.00      1.00      1.00        20
          21       1.00      1.00      1.00        26
          22       1.00      1.00      1.00        17

    accuracy                           1.00       440
   macro avg       1.00      1.00      1.00       440
weighted avg       1.00      1.00      1.00       440
```

Figure 10. (b). Classification Report - Naïve Bayes

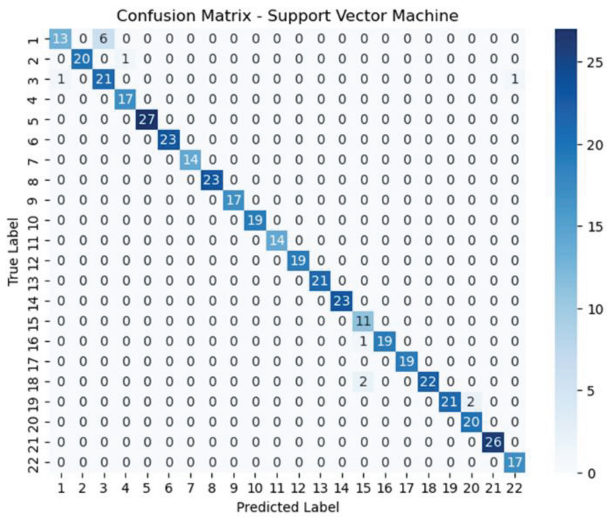

Figure 11. (a). Confusion Matrix - Support Vector Machine

systems to recommend high-yielding crops. To get the better insight of the dataset we have plotted graphs of correlation between each feature (Nitrogen, Phosphorus, Potassium, Temperature, pH and Humidity) and the label. Furthermore, each model is evaluated for accuracy using testing data. Table 3 and 4 displays the standardize and normalize the data for model compatibility, using Min-Max Scaling.

The Figure 8 shown above demonstrates that Random Forest and Naïve Bayes algorithms exhibit the utmost level of accuracy, while Logistic Regression and K-nearest Neighbors methods display the lowest level of accuracy as seen in the Table 5. Metrics such as the confusion matrix, precision, recall, and F1 score offer a more insightful analysis of the prediction results than accuracy alone. Confusion matrices display the true positive, true negative, false positive, and false negative predictions given by each model shown

Classification Report:

	precision	recall	f1-score	support
1	0.93	0.68	0.79	19
2	1.00	0.95	0.98	21
3	0.78	0.91	0.84	23
4	0.94	1.00	0.97	17
5	1.00	1.00	1.00	27
6	1.00	1.00	1.00	23
7	1.00	1.00	1.00	14
8	1.00	1.00	1.00	23
9	1.00	1.00	1.00	17
10	1.00	1.00	1.00	19
11	1.00	1.00	1.00	14
12	1.00	1.00	1.00	19
13	1.00	1.00	1.00	21
14	1.00	1.00	1.00	23
15	0.79	1.00	0.88	11
16	1.00	0.95	0.97	20
17	1.00	1.00	1.00	19
18	1.00	0.92	0.96	24
19	1.00	0.91	0.95	23
20	0.91	1.00	0.95	20
21	1.00	1.00	1.00	26
22	0.94	1.00	0.97	17
accuracy			0.97	440
macro avg	0.97	0.97	0.97	440
weighted avg	0.97	0.97	0.97	440

Figure 11. (b). Classification Report - Support Vector Machine

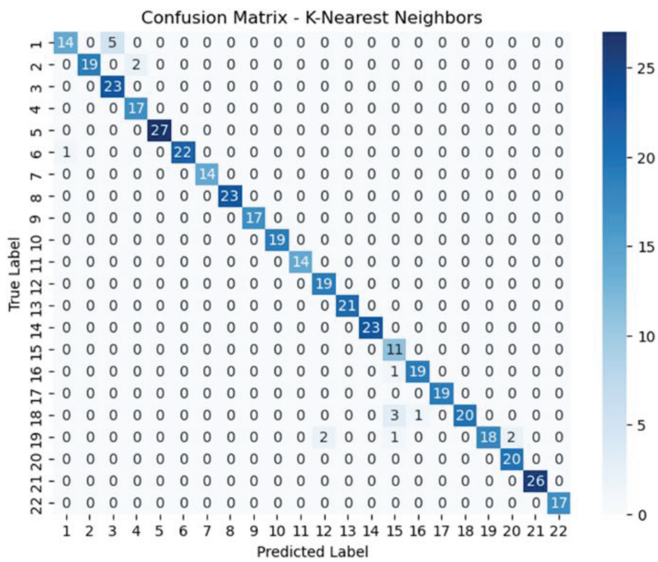

Figure 12. (a). Confusion Matrix - K-Nearest Neighbors

in the Figures 14(a) & 14(b), enabling us to evaluate the precision and efficacy of crop categorization. Additionally, the categorization reports provide a thorough evaluation of performance parameters such as precision, recall, and F1 score for each crop category. Precision assesses the model's accuracy in identifying specific crops by determining the ratio of correctly identified instances to all instances predicted for that crop, for example, correctly identifying wheat crops out of all predicted wheat instances. Recall provides crucial insights into the model's capability to capture and include all relevant events, such as ensuring all instances of a specific crop like rice are correctly identified and included in the predictions. The term refers to the proportion of true positive predictions relative to the combined number of false negative predictions

Classification Report:

	precision	recall	f1-score	support
1	0.93	0.74	0.82	19
2	1.00	0.90	0.95	21
3	0.82	1.00	0.90	23
4	0.89	1.00	0.94	17
5	1.00	1.00	1.00	27
6	1.00	0.96	0.98	23
7	1.00	1.00	1.00	14
8	1.00	1.00	1.00	23
9	1.00	1.00	1.00	17
10	1.00	1.00	1.00	19
11	1.00	1.00	1.00	14
12	0.90	1.00	0.95	19
13	1.00	1.00	1.00	21
14	1.00	1.00	1.00	23
15	0.69	1.00	0.81	11
16	0.95	0.95	0.95	20
17	1.00	1.00	1.00	19
18	1.00	0.83	0.91	24
19	1.00	0.78	0.88	23
20	0.91	1.00	0.95	20
21	1.00	1.00	1.00	26
22	1.00	1.00	1.00	17
accuracy			0.96	440
macro avg	0.96	0.96	0.96	440
weighted avg	0.97	0.96	0.96	440

Figure 12. (b). Classification Report - K-Nearest Neighbors

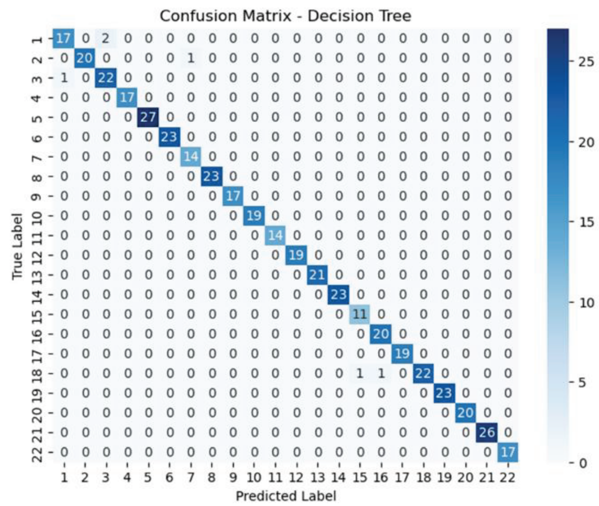

Figure 13. (a). Confusion Matrix - Decision Tree

Classification Report:

	precision	recall	f1-score	support
1	0.94	0.89	0.92	19
2	1.00	0.95	0.98	21
3	0.92	0.96	0.94	23
4	1.00	1.00	1.00	17
5	1.00	1.00	1.00	27
6	1.00	1.00	1.00	23
7	0.93	1.00	0.97	14
8	1.00	1.00	1.00	23
9	1.00	1.00	1.00	17
10	1.00	1.00	1.00	19
11	1.00	1.00	1.00	14
12	1.00	1.00	1.00	19
13	1.00	1.00	1.00	21
14	1.00	1.00	1.00	23
15	0.92	1.00	0.96	11
16	0.95	1.00	0.98	20
17	1.00	1.00	1.00	19
18	1.00	0.92	0.96	24
19	1.00	1.00	1.00	23
20	1.00	1.00	1.00	20
21	1.00	1.00	1.00	26
22	1.00	1.00	1.00	17
accuracy			0.99	440
macro avg	0.98	0.99	0.99	440
weighted avg	0.99	0.99	0.99	440

Figure 13. (b). Classification Report - Decision Tree

and true positive predictions for each category. Taking the weighted harmonic mean of precision and accuracy, one can calculate the F1 score. The number of actual occurrences of the class in the provided dataset is referred to as support. Consequently, we have generated a classification report and visualization for each model. To obtain the most accurate forecast, it is essential to utilize above mentioned metrics provided in the analysis [16].

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1\,Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (7)$$

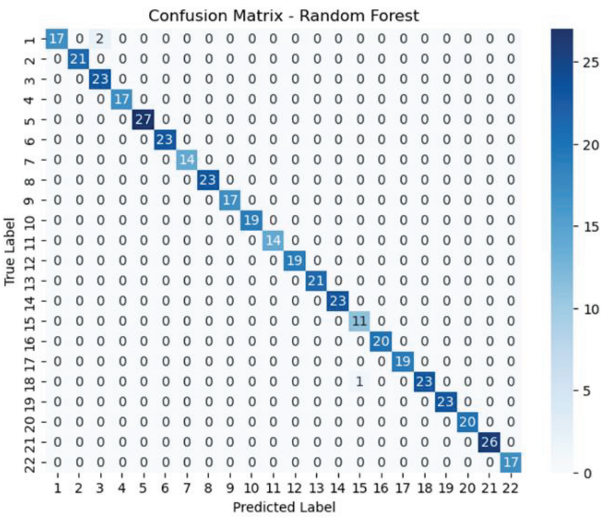

Figure 14. (a). Confusion Matrix - Random Forest

Classification Report:

	precision	recall	f1-score	support
1	1.00	0.89	0.94	19
2	1.00	1.00	1.00	21
3	0.92	1.00	0.96	23
4	1.00	1.00	1.00	17
5	1.00	1.00	1.00	27
6	1.00	1.00	1.00	23
7	1.00	1.00	1.00	14
8	1.00	1.00	1.00	23
9	1.00	1.00	1.00	17
10	1.00	1.00	1.00	19
11	1.00	1.00	1.00	14
12	1.00	1.00	1.00	19
13	1.00	1.00	1.00	21
14	1.00	1.00	1.00	23
15	0.92	1.00	0.96	11
16	1.00	1.00	1.00	20
17	1.00	1.00	1.00	19
18	1.00	0.96	0.98	24
19	1.00	1.00	1.00	23
20	1.00	1.00	1.00	20
21	1.00	1.00	1.00	26
22	1.00	1.00	1.00	17
accuracy			0.99	440
macro avg	0.99	0.99	0.99	440
weighted avg	0.99	0.99	0.99	440

Figure 14. (b). Classification Report - Random Forest

6. Conclusion and Prospective Intent

The proposed study will assist farmers in increasing agricultural output, reducing soil degradation in cultivated fields, and using less fertiliser during crop production by suggesting the optimal crop amount to plant based on a wide variety of criteria. The suggested activity helps farmers sustainability by helping them choose the right crops to cultivate. We have identified the advantages and disadvantages of each model through comparative analysis, providing important information for decision-making in the agricultural sector. We can improve the system later by adding more features to the dataset. Furthermore, with the support of remote sensing technologies and IoT devices, real-time monitoring can be made possible which would allow the system to recommend crops along with climate change. To make the project beneficial to farmers in every corner of our nation, we may also incorporate all regional languages.

Declarations

Funding None

Conflicts of interest The authors confirm that they have no competing interests or conflicts of interest.

Code availability The code for implementation is available upon request, subject to privacy and other restrictions.

Technology used Python programming, Machine learning libraries

AUTHORS

Sanket Gupta – Department of Computer Science and Engineering, Medi-Caps University, Indore, India, e-mail: sanket.jec@gmail.com.

Trishna Panse – Department of Computer Science and Engineering, Medi-Caps University, Indore, India, e-mail: trishnapanse@gmail.com.

Kailash Chandra Bandhu – Department of Computer Science and Engineering, Medi-Caps University, Indore, India, e-mail: kailashchandra.bandhu@gmail.com.

Ratnesh Litoriya – Department of Computer Science and Engineering, Medi-Caps University, Indore, India, e-mail: litoriya.ratnesh@gmail.com.

Shivani Patnaha – Department of Computer Science and Engineering, Medi-Caps University, Indore, India, e-mail: patnahashivani45@gmail.com.

Divya Kumawat – Department of Computer Science and Engineering, Medi-Caps University, Indore, India, e-mail: ddoraya@gmail.com.

Lishika Pargi – Department of Computer Science and Engineering, Medi-Caps University, Indore, India, e-mail: lishi1146@gmail.com.

Tisha Modi* – Department of Computer Science and Engineering, Medi-Caps University, Indore, India, e-mail: tishamodi0212@gmail.com.

*Corresponding author

References

[1] D. Pal, "Changes that have Taken Place in the Indian Economy after 1951", Dec. 2015.

[2] R. P. Reddy, B. Vinitha, K. Rishita, and K. Pranavi, "Crop Monitoring and Recommendation System using Machine Learning and IOT", *International Journal of Innovative Technology and Exploring Engineering*, 2020, DOI: 10.35940/ijitee.I7008.079920.

[3] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms", 2018, DOI: 10.1109/ICCUBEA.2018.8697349.

[4] Srilakshmi A., Madhumitha K., and Geetha K, "Machine learning approach: Recommendation of suitable crop for land using meteorological factors", *SSRN Electronic Journal*, 2020, Elsevier BV. DOI: https://doi.org/10.2139/ssrn.3736550

[5] S. Mamatha jajur, Soumya N.G., G.T. Raju, "Crop Recommendation using Machine Learning Techniques", International *Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 2S, 2019, 658–661.DOI: 10.35940/ijitee.b1106.1292s19.

[6] D. A. Bondre and M. Santosh Mahagaonkar, "PREDICTION OF CROP YIELD AND FERTILIZER RECOMMENDATION USING MACHINE LEARNING ALGORITHMS," 2019, DOI: 10.33564/IJEAST.2019.v04i05.055

[7] D. A. Reddy, B. Dadore, and A. Watekar, "Crop Recommendation System to Maximize Crop Yield in Ramtek region using Machine Learning," *Int J Sci Res Sci Technol*, 485–489, Feb. 2019, DOI: 10.32628/ijsrst196172.

[8] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery, "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique". in *2018* 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)., 2018, 114–119. DOI: 10.1109/CSITSS.2018.8768790.

[9] Jabed, Md & Murad, Masrah, "Crop Yield Prediction in Agriculture: A Comprehensive Review of Machine Learning and Deep Learning Approaches, with Insights for Future Research and Sustainability", 2024 DOI: 10.1016/j.heliyon.2024.e40836

[10] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," in 2016 Eighth International Conference on Advanced Computing (ICoAC)., 2017, 32–36. DOI: 10.1109/ICoAC.2017.7951740.

[11] R. Kumar, M. Singh, P. Kumar, and J. Singh, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", 2015, DOI: 10.1109/ICSTM.2015.7225403.

[12] Ramu, K Priyadarsini, K, "A Review on Crop Yield prediction Using Machine Learning Methods", 2021, DOI: 10.1109/ICOSEC51865.2021.9591764

[13] F. Shahrin, L. Zahin, R. Rahman, A. J. Hossain, A. H. Kaf and A. K. M. Abdul Malek Azad, "Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning", 2020 11th International Conference on Electrical and Computer Engineering (ICECE), 21-24, 2020, DOI: 10.1109/ICECE51571.2020.9393066

[14] M. Kavita and P. Mathur, "Crop Yield Estimation in India Using Machine Learning", 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 220-224, 2020, DOI: 10.1109/ICCCA49541.2020.9250915

[15] P. S. Nishant, P. Sai Venkat, B. L. Avinash and B. Jabber, "Crop Yield Prediction based on Indian Agriculture using Machine Learning", 2020 International Conference for Emerging Technology (INCET)., 1–4, 2020, DOI: 10.1109/INCET49848.2020.9154036

[16] P.S. Maya Gopal and R. Bhargavi, "Optimum Feature Subset for Optimizing Crop Yield Prediction Using Filter and Wrapper Approaches", Applied Engineering in Agriculture., vol. 35, no. 1, 2019, 9-14. DOI: https://doi.org/10.1016/j.compag.2019.104968