SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V11I11P126

Volume 11 Issue 11, 275-283, November 2024

Original Article

Analysis of a Crop Recommendation System for Farmers Based on Machine Learning

Ujwala Ghodeswar[1], Minal Keote[2]

[1]*Department of Electronics, YCCE, Maharashtra, India.*
[2]*Department of Electronics & Telecommunication, YCCE, Maharashtra, India.*

[1]*Corresponding Author : ghodeswarujwala83@gmail.com*

Received: 14 September 2024 Revised: 15 October 2024 Accepted: 13 November 2024 Published: 30 November 2024

Abstract - *Farmers must identify a certain soil type crop before sowing seeds. A farmer's ability to determine which crop to plant will boost the production of the land. Farmers risk losing their crops, wasting their time, and losing the money they invested in cultivating them. A system that recommends crops based on machine learning is suggested to make this process easier for farmers. Eight machine learning algorithms are applied to determine the crop for a given plot of land. Some techniques include decision Tree, Naïve Bayes, K Nearest Neighbour, Random Forest, Adaboost, Logistic Regression, Gradient Boosting, and Support Vector Machine. The dataset utilized for the system was obtained from Kaggle. This data collection consists of 2200 rows of varying values of seven features: N, P, K, temperature, humidity, rainfall, soil pH, and one output label. The above algorithms are trained using a 20% test and 80% training data set. Accuracy is calculated from the machine learning algorithm compared to other algorithms, and the Naïve Bayes algorithm gives good accuracy.*

Keywords - *Accuracy, Decision tree, Logistic regression, Machine Learning, XGBoost.*

1. Introduction

Farming is the most important occupation in the world for the existence of humans. It generates the most basic requirements for humans to exist in this world. The agriculture industry strongly relies on novel ideas because of the constant demand for crops and vegetables. The solution to today's farmers' challenges is the necessity for crop analysis for specific soils. This can speed up production, prevent time loss, and preserve crops from being wasted and damaged due to weather conditions. This analysis can be performed using machine learning algorithms. These machine learning algorithms provide a solution for which crop suits specific soil and climatic circumstances. Several studies utilised various machine learning techniques, focusing on their accuracy. Some of the studies conducted are listed below.

Machine learning techniques categorize, detect, and predict disease in tomato crops. Smart farming ideas aim to improve traditional agricultural processes [1]. The study compares Explainable Artificial Intelligence (XAI) concepts to machine learning models like Gradient Boosting, Decision Tree, Random Forest, Gaussian Naive Bayes, and Multimodal Naive Bayes. Three performance evaluation metrics are considered, i.e. R-squared, Mean Squared and Absolute Error [2]. Cloud computing and the Internet of Things are better approaches to agriculture that allow farmers to produce superior crops. Ensemble methods are also used to improve the precision of recommendation systems [3]. Random Forest predicts crops with 95% accuracy compared to Support Vector Machines, Artificial Neural Networks, Multivariate Linear Regression, and K-Nearest Neighbor algorithms [4]. A convolutional neural network determines whether a plant is at risk of illness [5].

Different machine learning algorithms recommend crops for specific soil types [6]. The accuracy of the crop recommendation system is calculated using various algorithms [7]. Internet of things-based machine learning model is used to calculate the efficiency of algorithms [8]. The importance and implementation of the crop recommendation system are discussed [9]. Naïve Bayes and XGBoost algorithms give 99% accuracy compared to other algorithms [10]. The Arduino Uno board implements the appropriate crop for a certain land [11]. Data analysis and neural networks provide crop fertilization suggestions with 97% accuracy [12]. The dataset used in this paper was taken from Kaggle [13].

Machine learning algorithms are used for recommendation systems, achieving an accuracy of 98.48% for decision trees and 99.31% for random forest classifiers [14]. The Synthetic Minority Oversampling Technique (SMOTE) is used to find the accuracy of different classifiers. The author also calculated precision, recall, and F1 score parameters without using SMOTE [15].

A convolutional neural network is used for recommendation systems with 99.98% accuracy [16]. The smart web-based application is developed using machine learning algorithms, and the accuracy of the algorithms is calculated [17]. Graph convolution neural network model is used to recommend crops [18]. Farmers can use a blockchain-based method to recommend the best crop for a certain soil [19]. IoT based methods are also employed to determine the accuracy of crop recommendation systems [20]. According to the literature mentioned above, several researchers have developed various machine learning algorithms by considering different features given in the dataset. According to the literature survey, the most significant research gap in recommendation systems is the lack of feature analysis. If the features are evaluated, agricultural productivity will improve. Hence, in this study, data is analyzed in depth. This work's primary goal is to give crop suggestions to the farmers based on the feature values available in the dataset. Section 2 gives the information about the dataset. Section 3 gives the methodology used, and a discussion of the results is given in Section 4.

2. Dataset Information

The machine learning-based crop recommendation dataset comprises 17600 elements with 2200 distinct values for each of the seven features and one output. A crop recommendation system produces labels as its output. The machine learning algorithms decide which crop should be planted in the soil based on the values of seven features. The seven features are phosphorus, nitrogen, potassium, temperature, humidity, pH, and rainfall, and they are labeled as the output-dependent variable.

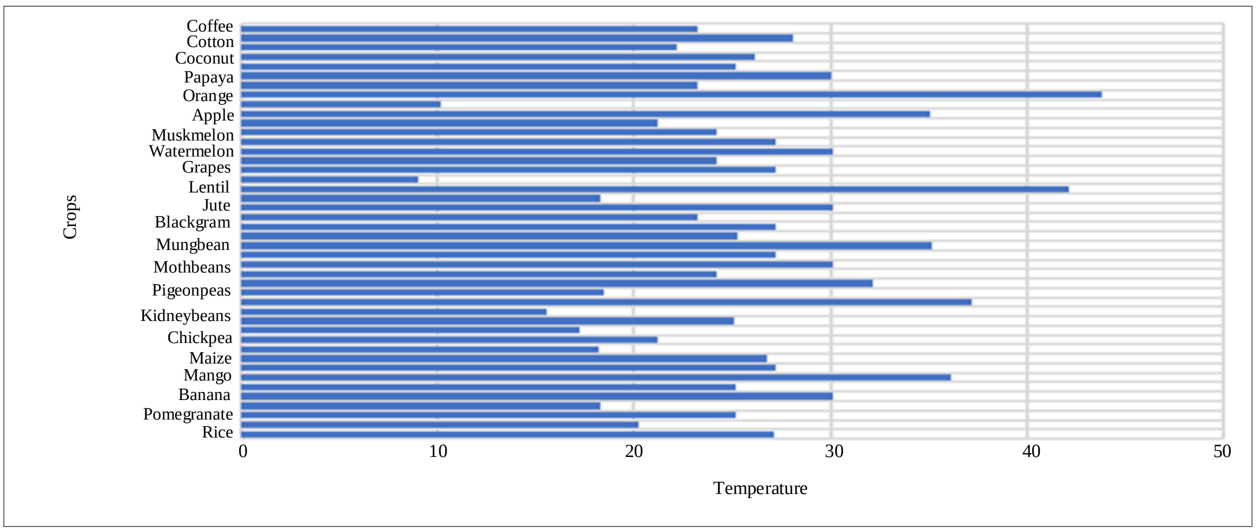

Fig. 1 Temperature extremes for crops

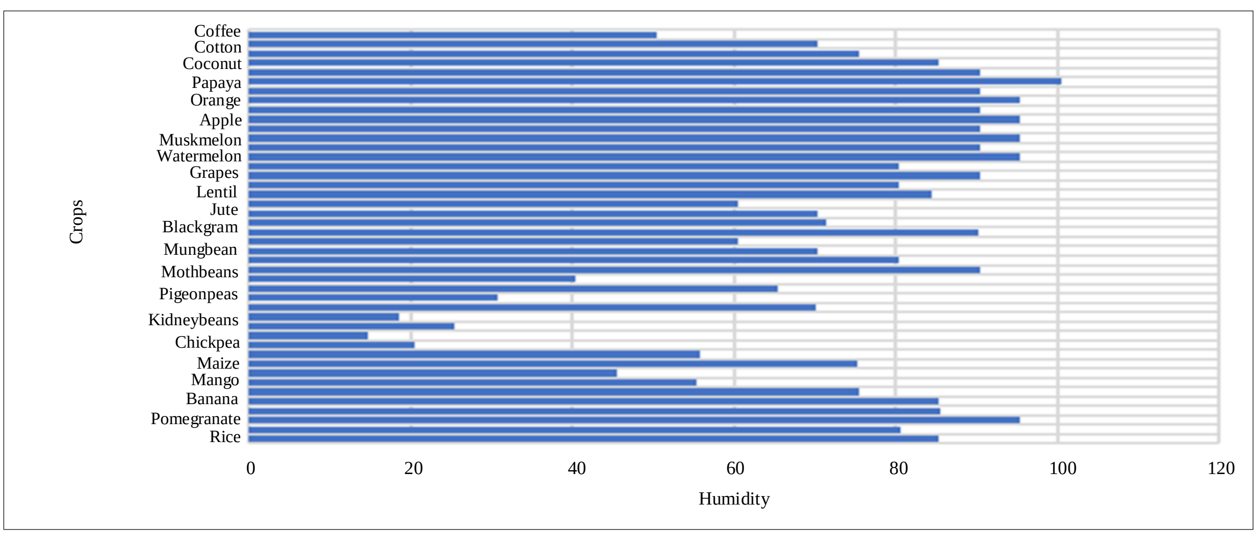

Fig. 2 Humidity extremes for crops

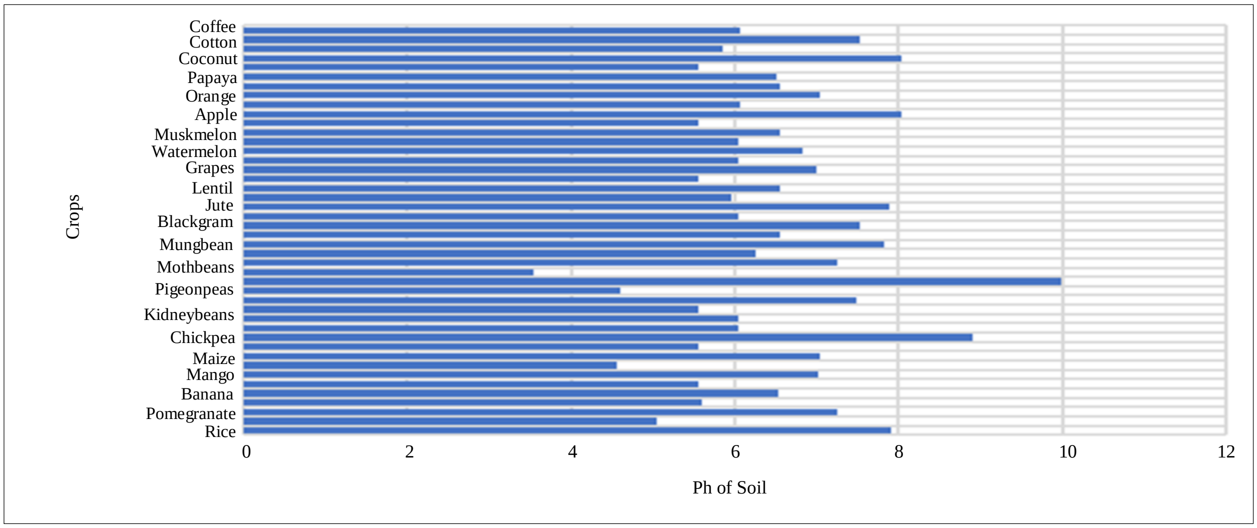

Fig. 3 pH of soil extremes for crops

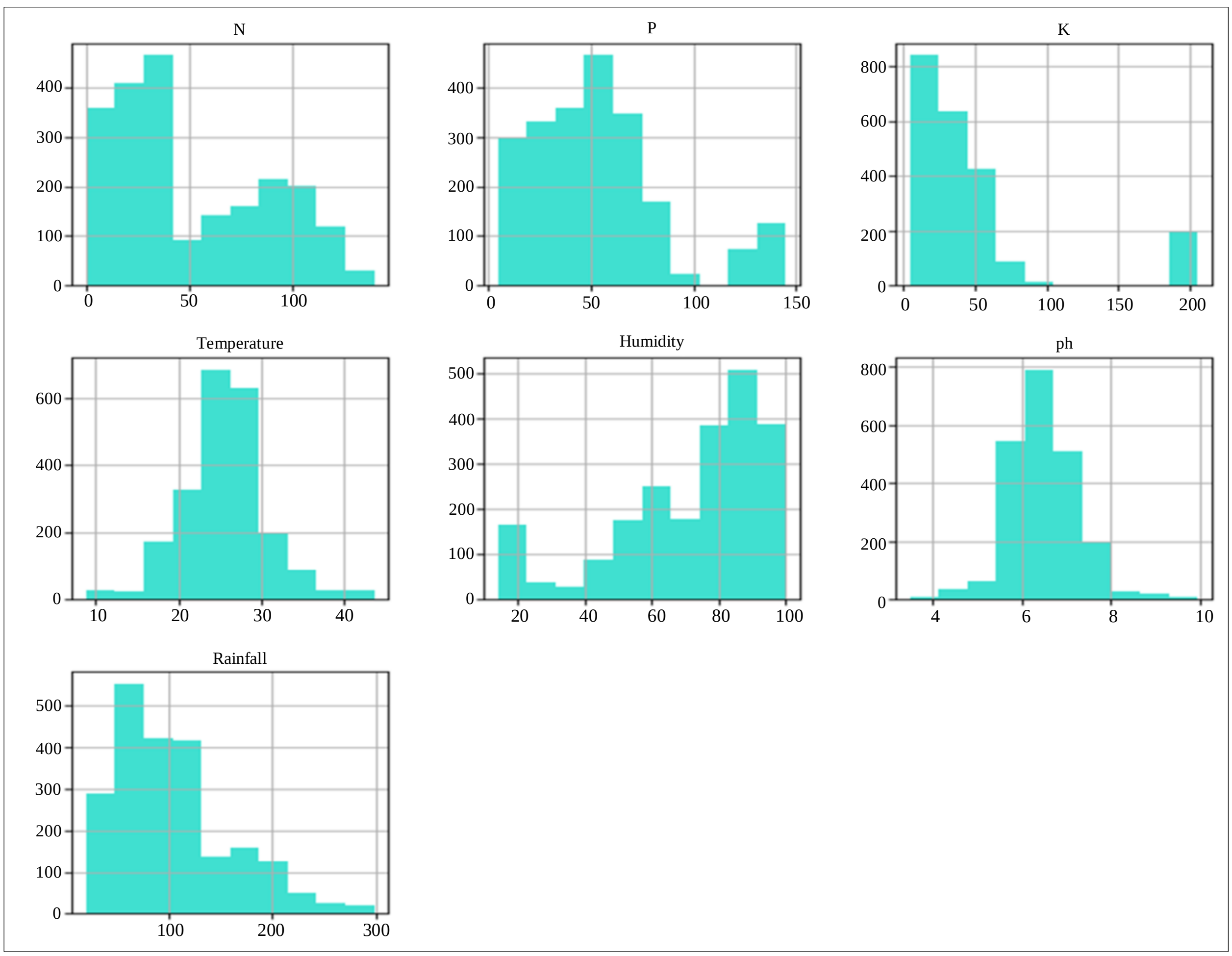

Fig. 4 Histogram plot of features

Nitrogen and phosphorus are key nutrients required for plant growth. Plants rely on phosphorus for cell division and development. Potassium is required for water circulation and is essential for the development of crops. The acidity or alkalinity of soil is indicated by pH value, which indicates the concentration of hydrogen ions in the soil. If the pH value is less than 7, the soil is considered acidic; if it is greater than 7, it is considered alkaline; and at pH=7, the soil is neutral. The pH value of plants should be maintained for general development; else, crop damage would occur.

Muskmelon, watermelon, banana, and cotton crops demand more nitrogen. In contrast, moth beans, lentils, mung beans, kidney beans, grapes, mango, pigeon peas, apples, oranges, pomegranate, and coconut require less amount of nitrogen. Crops like oranges, pomegranate, coconut, muskmelon, and watermelon require less phosphorus, whereas grapes and apples demand more. The potassium need for oranges is low. However, for grapes and apples, it is considerable, at 205 kg/ha. The temperature needed for papaya crops is higher, at 43.675 Celsius, but grapes require a lower temperature of 8.825 Celsius, as shown in Figure 1. Chickpeas demand less humidity (14.258%) than coconut, as shown in Figure 2. The soil pH requirement for mothbeans ranges from 3.504 to 9.935. The graphical representation of Figure 3 illustrates the maximum and minimum soil pH requirements for crops. Muskmelon requires the least rainfall, whereas rice crops need 298.560mm.

Table 1 gives the lower and higher range of feature values. The types of crops given as output labels are rice, mango, watermelon, maize, apple, banana, black gram, chickpea, coffee, cotton, grapes, jute, kidney beans, lentil, muskmelon, orange, papaya, pomegranate, mung beans, moth beans, orange and muskmelon. There are a total of 100 different values for each type of crop. A total of 22 types of crops can be recommended depending on the characteristics of features in the soil. The total values of the features are 2200.

Table 1. Range of feature values

Features	**Range of values**
Nitrogen	0 -140 kg/ha
Phosphorus	5 - 145 kg/ha
Potassium	5 -205 kg/ha
Temperature	8.825 - 43.675 Celsius
Humidity	14.258 - 99.981%
Ph of soil	3.504 - 9.935
rainfall	20.211 mm - 298.560 mm

Figure 4 gives the histogram plot of seven features. This histogram provides a graphical representation of the feature values. It displays the maximum and minimum ranges for each feature value. The heatmap shows the relation between the required soil parameters and each other. Some values are positively correlated with others, and some are negatively correlated.

Fig. 5 Heatmap of dependent and independent variables

Figure 5 shows the relationship between the feature variables. Rainfall is positively correlated with nitrogen and humidity and negatively correlated with remaining features. The pH value of soil is positively correlated with nitrogen, but it has a negative relation with other features. Humidity positively correlates with rainfall, temperature, Nitrogen, and potassium. The shape of the crop database is 2200 by 8. There are 7 feature variables and one label as the dependent variable.

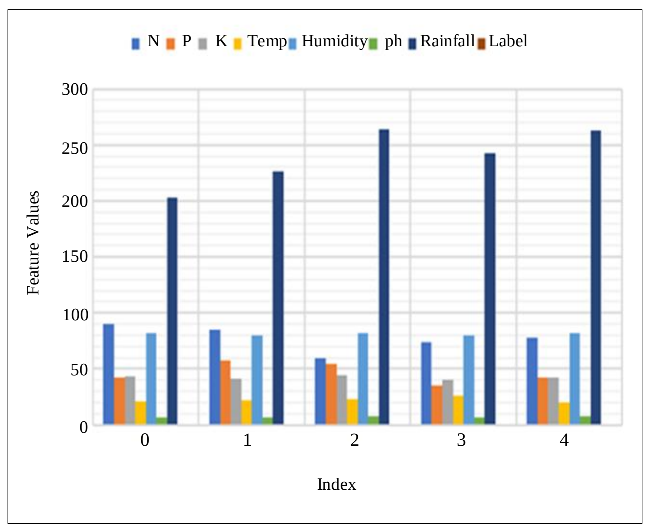

Fig. 6 Five rows of rice crop data

Figure 6 indicates the variation of feature values with respect to the index for the crop dataset using a bar chart. From this, it can be observed that rainfall is the most important feature for rice crops as the rainfall has greater value than other features. Similarly, Figure 7 bar chart indicates coffee plant requirements. The last five rows of coffee crop data are indicated in Figure 7. In this, the least requirement is for the pH feature value. It is approximately 6.

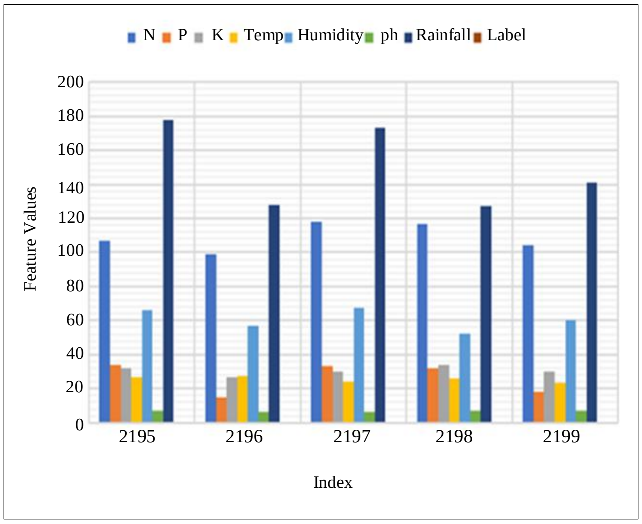

Fig. 7 Five rows of coffee crop data

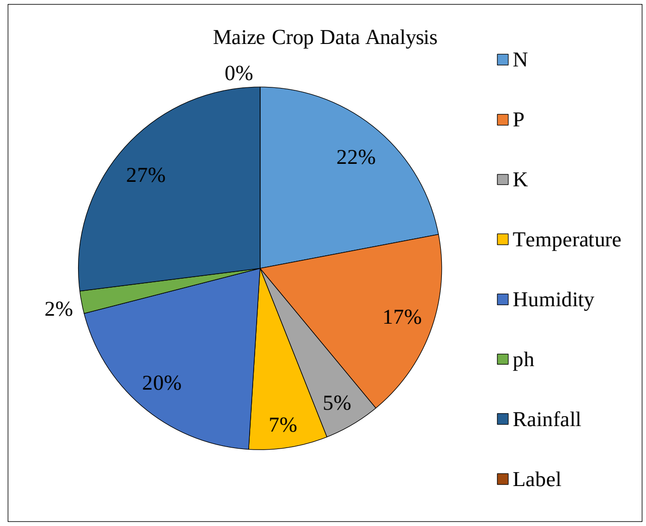

Fig. 8(c) Chickpea crop

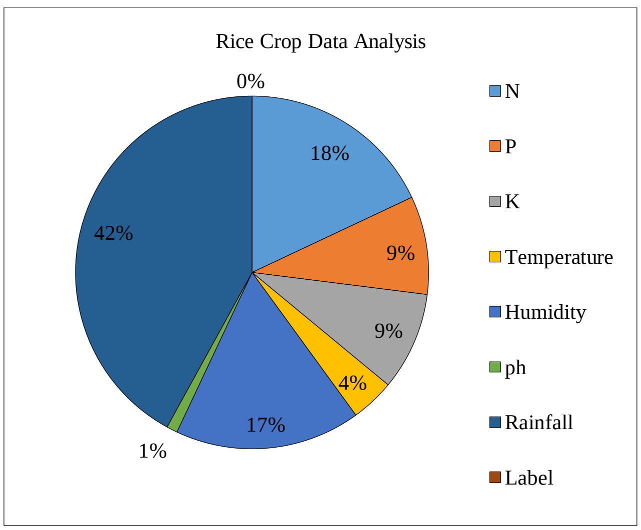

Fig. 8(a) Rice crop

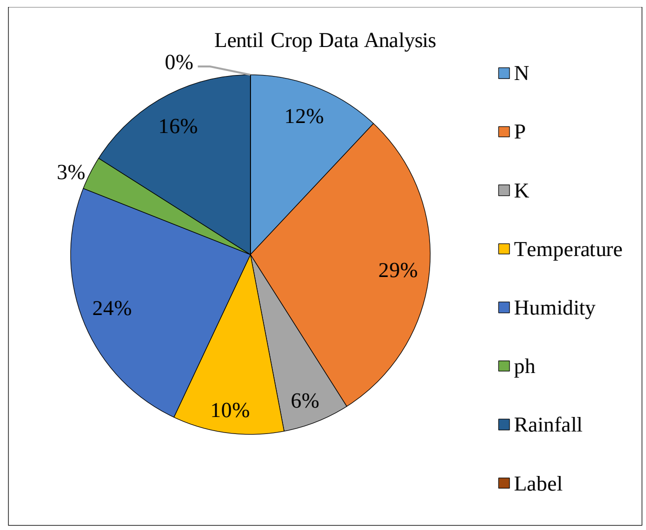

Fig. 8(d) Lentil crop

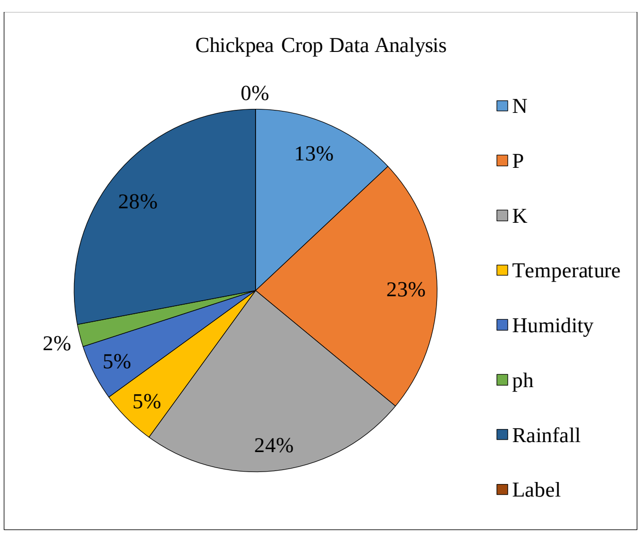

Fig. 8(b) Maize crop

Figure 8(a) gives a pie chart of rice feature values for the 100 index parameters. Similarly, Figures 8(b),(c), and (d) give Maize, chickpea, and lentil crop feature values. This pie chart indicates the percentage requirement of feature values for the corresponding crop. This percentage requirement is also given in Table 2. Table 2 gives the percentage requirement of the feature values for the four crops whose pie chart is in Figure 8. It can be observed that the requirement for nitrogen is higher in maize than in the other four features. Similarly, phosphorus, temperature, and humidity are higher in Lentil crops.

3. Materials and Methods

A content-based recommendation system is used for this. This is based on the contents of feature vectors for the crop identified and recommended to the farmers. Collaborative filtering is based on the approaches of different farmers for recommending crops in their fields.

Table 2. Percentage requirement of feature values

Sr. No.	N	P	K	Temp	Humidity	ph	Rainfall	Label
1	18	9	9	4	17	1	42	Rice
2	22	17	5	7	20	2	27	Maize
3	13	23	24	5	5	2	28	Chickpea
4	12	29	6	10	24	3	16	Lentil

Algorithms are implemented in python. All necessary libraries are imported into python. The input to this system is a csv file containing the crop recommendation dataset. This dataset consists of different values for seven features. After this, empty fields from the data are analyzed. Then, the data is divided into train and test datasets.

The accuracy of eight machine learning algorithms is calculated. Based on accuracy, a relevant crop is recommended for farmers. The train size is 80%, and the test size is 20% applied to the algorithms. The machine learning model is trained using this training size to forecast and identify which crop should be planted based on new data values. Utilizing varying feature values improves the accuracy of this model.

This is applied to the machine learning algorithms. Train and test Accuracy is calculated. This is done to recommend rice crops. The same method is applied to recommend the remaining crops.

Performance analysis of the system is determined by eight algorithms: Naïve Bayes, Decision Tree, Logistic Regression, Gradient boosting, Adaboost classifier, support vector machine, k nearest neighbor, and random forest.

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \quad (1)$$

Where,

TN : True Negative
TP : True Positive
FP : False Positive
FN : False Negative

The algorithm's accuracy depends on TN, TP, FP, and FN values. True positive indicates that both the expected and actual values are positive. True negative indicates that both the expected and actual values are negative. False positive means the predicted value is positive, and the actual value is negative. False negative indicates that the actual value is positive while the predicted value is negative.

$$\mathrm{Re}\,call = \frac{TP}{FN + TP} \quad (2)$$

$$\mathrm{Pr}\,ecision = \frac{TP}{FP + TP} \quad (3)$$

Precision and recall parameters are calculated from Equations (2) and (3). Precision is related to a false positive rate of prediction, and Recall is related to a false negative rate. Precision is related to the correct predictions given by the model. These two terms are crucial in machine learning since they indicate the model's accuracy.

4. Results and Discussion

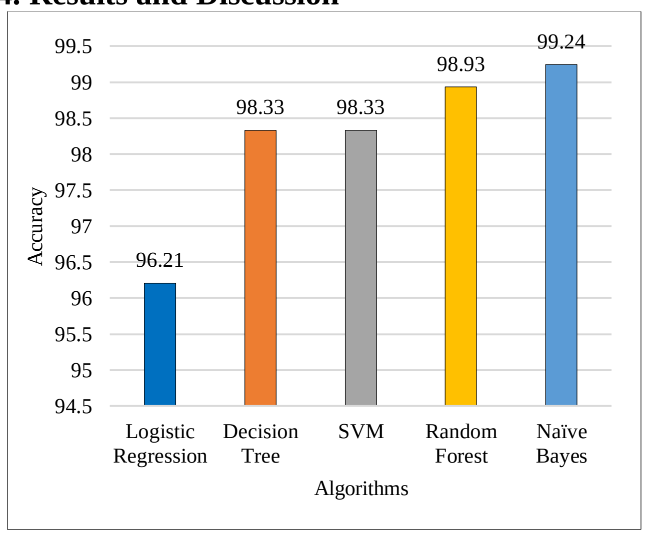

Fig. 9 Histogram of accuracy

The algorithms' accuracy is plotted using a histogram in Figure 9. This histogram shows that the accuracy of naïve Bayes, bagging classifier and gradient boosting algorithm is near 100 percent.

Table 3 shows the TN, TP, FN, and FP parameter values. Table 3 indicates the train test accuracy parameter. Figure 10 gives the bar chart indicating a graphical representation of the accuracy obtained from the algorithms. Figure 11 gives the Precision, Recall and F1 score obtained. These numerical values are also indicated in Table 4. As shown in Figure 10, the decision tree's train test accuracy is higher than the rest of the algorithms. The Adaboost algorithm has the highest precision, recall and F1 score values, as shown in Figure 11.

Table 3. Accuracy parameters

Algorithms Applied	TN	TP	FN	FP
Naïve Bayes	418	15	0	7
Decision Tree	424	11	3	2
Logistic Regression	422	12	3	3
Gradient Boosting Classifier	417	18	5	0
Adaboost Classifier	420	21	0	0
Support Vector Machine	415	16	8	1
K Nearest Neighbor	416	19	3	2
Random Forest Classifier	420	14	6	0

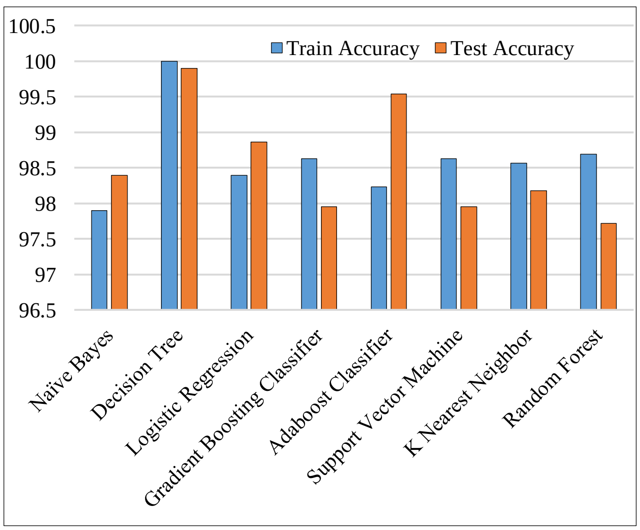

Fig. 10 Train, test accuracy

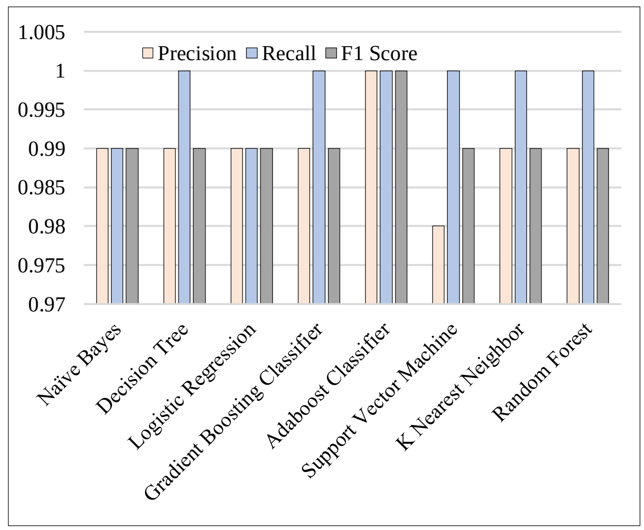

Fig. 11 Calculation of precision, recall, f1 score

Table 4. Train and test accuracy

Algorithms Applied	Train Accuracy	Test Accuracy	Precision	Recall	F1 Score
Naïve Bayes	97.9	98.40	0.99	0.99	0.99
Decision Tree	100	99.9	0.99	1	0.99
Logistic Regression	98.4	98.86	0.99	0.99	0.99
Gradient Boosting Classifier	98.63	97.95	0.99	1	0.99
Adaboost Classifier	98.23	99.54	1	1	1
Support Vector Machine	98.63	97.95	0.98	1	0.99
K Nearest Neighbor	98.57	98.18	0.99	1	0.99
Random Forest	98.69	97.72	0.99	1	0.99

Table 5 gives a comparative analysis of the earlier published research paper. The accuracy of reference [10] is compared with the accuracy obtained in this work. This significantly improves the accuracy of decision trees, SVM, and random forest machine learning algorithms.

Table 5. Comparative analysis

Algorithms	Ref. [10]	This Work (2)
Logistic Regression	95.22%	96.21%
Decision Tree	90%	98.33%
SVM	10.86%	98.33%
Random Forest	95.22%	98.93%
Naïve Bayes	99%	99.24%

5. Conclusion

Farmers must know about recommended crops before planting them in the ground. This study discusses the importance of a crop decision system for farmers. This paper uses machine learning techniques to recommend which crop is best suited to a certain plot of land. Eight algorithms based on machine learning are applied to determine the best harvest. The percentage accuracy for training and testing has been established. After analysis, the decision tree algorithm has higher accuracy than other algorithms, while the Adaboost approach has higher precision, recall, and F1 score. When high precision is required, the Adaboost algorithm is utilised, while in accuracy, the decision tree approach is used. Most rural farmers are unaware of technology and its applications, which significantly challenges the recommendation system. As a result, delivering this technology to farmers in remote areas is challenging. For this challenge, further development could include developing a mobile application for the recommendation system so that all farmers can benefit from it and boost crop growth.

References

[1] Tiago Domingues, Tomás Brandão, and João C. Ferreira, "Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey," *Agriculture*, vol. 12, no. 9, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Mehmoud Y. Shams, Samah A. Gamez, and Fatma M. Talaat, "Enhancing Crop Recommendation Systems with Explainable Artificial Intelligence: A Study on Agricultural Decision-Making," *Neural Computing and Applications*, vol. 36, pp. 5695-5714, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Sasmita Subhadarsinee Choudhury et al., "An Acquisition Based Optimised Crop Recommendation System with Machine Learning Algorithm," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 11, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Shilpa Mangesh Pande et al., "Crop Recommender System Using Machine Learning Approach," *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, pp. 1066-1071, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Dhruvi Gosai et al., "Crop Recommendation System Using Machine Learning," *International Journal of Scientific Research in Computer Science, Engineering and Information*, vol. 7, no. 3, pp. 554-557, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Shafiulla Shariff et al., "Crop Recommendation Using Machine Learning Techniques," *International Journal of Engineering Research & Technology*, vol. 10, no. 11, pp. 199-201, 2022. [Publisher Link]

[7] Devendra Dahiphale et al., "Smart Farming: Crop Recommendation Using Machine Learning with Challenges and Future Ideas," *TechRxiv*, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Navod Neranjan Thilakarathne et al., "A Cloud Enabled Crop Recommendation Platform for Machine Learning-Driven Precision Farming," *Sensors*, vol. 22, no. 16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Sumanth Dubbudu et al., "Crop Recommendation Using Machine Learning," *International Journal of Creative Research Thoughts*, vol. 11, no. 6, pp. 609-612, 2023. [Publisher Link]

[10] Immanni Venkata, and Narayana Chowdary, "Crop Recommendation Using Machine Learning," *International Journal of Research Publication and Reviews*, vol. 4, no. 12, pp. 4007-4009, 2023. [Publisher Link]

[11] Pradeepa Bandara et al., "Crop Recommendation System," *International Journal of Computer Applications*, vol. 175, no. 22, pp. 22-25, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Christine Musanase et al., "Data-Driven Analysis and Machine Learning-Based Crop and Fertilizer Recommendation System for Revolutionizing Farming Practices," *Agriculture*, vol. 13, no. 11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Atharva Ingle, Crop Recommendation Dataset, Maximize Agricultural Yield by Recommending Appropriate Crops. [Online]. Available: https://Www.Kaggle.Com/Datasets/Atharvaingle/Crop-Recommendation-Dataset

[14] Mohamed Bouni et al., "Integrated IoT Approaches for Crop Recommendation and Yield-Prediction Using Machine-Learning," *IoT*, vol. 5, no. 4, pp. 634-649, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Murali Krishna Senapaty, Abhishek Ray, and Neelamadhab Padhy, "A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms," *Agriculture*, vol. 14, no. 8, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[16] P. Ankit Krishna, Neelamadhab Padhy, and Archana Patnaik, "Applying Machine Learning for Sustainable Farm Management: Integrating Crop Recommendations and Disease Identification," *Engineering Proceedings*, vol. 67, no. 1, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[17] K. Sutha, N. Indumathi, and S. Uma Shankari, "Recommending and Predicting Crop Yield Using Smart Machine Learning Algorithm (SMLA)," *Current Agriculture Research Journal*, vol. 11, no. 2, pp. 686-694, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[18] P. Ayesha Barvin, and T. Sampradeepraj, "Crop Recommendation Systems Based on Soil and Environmental Factors Using Graph Convolution Neural Network: A Systematic Literature Review," *Engineering Proceedings*, vol. 58, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[19] Devangi Hitenkumar Patel et al., "Blockchain-Based Crop Recommendation System for Precision Farming in IoT Environment," *Agronomy*, vol. 13, no. 10, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[20] Ersin Elbasi et al., "Optimizing Agricultural Data Analysis Techniques through AI-Powered Decision-Making Processes," *Applied Sciences*, vol. 14, no. 17, 2024. [CrossRef] [Google Scholar] [Publisher Link]