Crop Recommendation using Machine Learning Techniques

Shafiulla Shariff
Department of CS&E
Jain Institute of Technology
Davanagere, India

Shwetha R B
Department of CS&E
Jain Institute of Technology
Davanagere, India

Ramya O G
Department of CS&E
Jain Institute of Technology
Davanagere, India

Pushpa H
Department of CS&E
Jain Institute of Technology
Davanagere, India

Pooja K R
Department of CS&E
Jain Institute of Technology
Davanagere, India

Abstract—**Agriculture is an important industry in India. It is essential for the survival and growth of the Indian economy. India is a large producer of a variety of agricultural products. Soil is an important factor in crop cultivation. Soil is a non-renewable dynamic natural resource that is necessary for life. Previously, crop cultivation was done by farmers who had hands-on experience. Farmers are no longer able to choose the best suitable crop based on soil characteristics and features. So, arecommendation system has been developed that employs machine learningalgorithms to recommend the crop that can be harvested in that particular soil. There are several machine learning algorithms available are used in this system., including KNN, Decision Tree, and Random Forest, Naive Bayes and Gradient Boosting to recommend the crop.**

Keywords—Machine Learning, Crop Recommendation, KNN, Decision Tree, Naive Bayes, Random Forest, Gradient Boosting.

I. INTRODUCTION

Agriculture, as we all know, is the foundation of the Indian economy. Agriculture is an important occupation in India. More than 60% of the country's land is used for agriculture, which feeds 1.3 billion people [1]. Agriculture is the cultivation of plants and animals. In India, agriculture gave rise to civilization. We need soil to cultivate crops. As a result, soil is a critical factor in agriculture. Soil health is essential for good food production. It provides the roots with essential nutrients, water, oxygen, and support. Soil is the foundation of the food system, as well as the location of all plants used in food production. In India, several soil varieties are available. They are alluvial soil (cotton, rice), black soil (sugarcane, sunflower), red soil (corn, ragi), laterite soil (pulses, tea, coffee), and so on. Many studies have been conducted to improve agricultural planning. The crop can be recommended using a machine learning technique.

Machine learning is an subfield of artificial intelligence that describes a machine's ability to mimic intelligent human behavior. Artificial intelligence systems are employed in the same way as humans do to automate complex tasks [2]. Machine learning begins with data, such as financial transactions, individuals, or photos.

The information is collected and processed to be utilized as training data for the machine learning system. If the data is more then the software shows better results. After that, the developer select a ML model to use, input the data, and train the system to find patterns or make predictions on its own.

II. LITERATURE REVIEW

A study of machine learning algorithms was conducted in a research paper by Rashi Agarwal [3]. This system would help farmers make educated decisions about which crops to grow based on a variety of environmental and geographical factors. They employed decision trees, KNNs, Random Forests, and neural networks. The neural network had the highest accuracy of all of them.

Priyadharshini A [4] conducted a study on machine learning algorithms in her research article. This technology reduces crop failure and decreases productivity by supporting farmers in choosing the proper crop and provide the data that regular farmers do not maintain. A variety of machine learning algorithms were applied. The neural network was the most accurate of the bunch.

Shilpa Mangesh Pande [5] In her research article, she presents a farmer-friendly and realistic production forecasting system The suggested technology is connected to farmers via a mobile application. The user's location is determined with the help of GPS. All of the algorithms are compared in terms of crop yield forecast accuracy. The RF algorithm showed to bethe best for the provided data set, with a 95% accuracy.

Mayank Champaneri [6] conducted research on crop yield prediction using a data mining technique. They used a random forest classifier because it can perform classification and regression tasks. The user-friendly website built that can be used by anyone to predict crop yield for their choice of crop by giving climate data for that area.

III. PROPOSED SYSTEM

The proposed approach will recommend the optimum crop basedon a few soil factors.
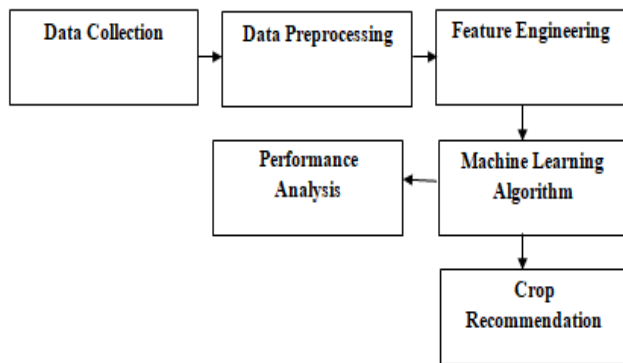
Fig 1. Methodology of crop recommended system.

The technique of the suggested system is made up of numerousblocks, as indicated in fig (1).

Data Collection: Data collection is the most common approach for gathering and analysing information from varioussources. The dataset must have the following qualities to provide an approximate data set for the system. These criteria will be considered for crop recommendation: i) soil PH ii) Humidity iii) NPK levels iv) crop data v) temperature.

Data Pre-Processing: After collection of data from different sources, the next step is to pre-processed it before the model can be trained. Starting with reading the acquired dataset and going through data purification, data pre-processing can be conducted in numerous ways. When cleansing information, Some dataset characteristics are redundant. are not taken into account while cropping. prediction. As a result, we must remove undesirable properties and datasets that contain some missing data. We must drop or fill these missing values with unwanted nan values with better precision in order to obtain them.

Feature Engineering: Using domain expertise, feature engineering extracts features (characteristics, traits, and attributes) from raw data. The idea is to use these extra attributes to improve the quality of ML results.

Training set: A training set is a data set that contains data that has been labelled. Both input and output vectors are included. The model is trained using supervised machine learning algorithms using this dataset.

Testing set: A testing set is a data set that is devoid of labelled data. It predicts the outcome with the assistance of the training data set. It is unaffected by the training data set.

Machine Learning Algorithm: Machine learning prediction algorithms [12] [13] necessitate extremely accurate estimation based on previously learned data. Predictive analytics historical information is the use of data, statistical methods, and machine learning approaches to forecast future results. The goal is to go beyond simply understanding what happened to providing the best feasible remedy and a prediction of what will happen next.

KNN, Decision Tree, Nave Bayes, Random Forest, and Gradient Boosting methods are used in this model.

A. K-Nearest Neighbor Classifier

KNN is a type of supervised machine learning that can be used to solve a wide range of issues. Classification and regression are examples of challenges that can be addressed. The number of nearest neighbours to a newly forecasted unknown variable is represented by the symbol K.

The distance between the data points is calculated using the Euclidean distance formula.

Euclidean Distance b/w A and B =
$$\sqrt{(X2 - X1)2 + (Y2 - Y1)2}. \quad (1)$$

B. Decision Tree

Decision Trees (DTs) comes under supervised learning for classification and regression. A tree representation is used to solve the problem, with each leaf node representing a class label. and the tree's interior node represents attributes.

Entropy:
$$H(S) = -\sum P_i(S) \log_2 P_i(S) \quad (2)$$

Information Gain:
$$IG(S,A) = H(S) - \sum_{v \in Values(A)} (|S_v|/S) H(S_v) \quad (3)$$

C. Naive Bayes

The Bayes theorem is used to create a simple probablistic classifier called Naive Bayes. Naive Bayes classifiers assume that the value of one feature is independent of the value of any other feature given the class variable.

$$P(A|B) = (P(B|A) * P(A))/P(B) \quad (4)$$

D. Random Forest

Random Forest is an ensemble learning method that creates a huge number of distinct models to solve classification, regression, and other issues. At training time, decision trees are used. Algorithm of random forest makes decision trees based on various data samples and then forecasts data from each subset, then votes on it gives the system a better solution RF uses the bagging method for data training, which improves the correctness of the outcome.

$$Gini\ Index = 1 - \sum_{i=1}^{\infty} (P_i)^2$$
$$= 1 - [(P_+)^2 + (P_-)^{2]} \quad (5)$$

E. Gradient Boosting

Gradient Boosting is also comes under supervised machine learning technique for solving classification and regression issues. It's a poor prediction model with an ensemble. As with previous boosting approaches, a gradient-boosted trees model is built stage by stage.

Crop Recommendation: Based on the N P K, temperature, humidity, and ph, the model will recommend the optimum crop to grow on the given soil.

Performance Analysis: Performance analysis is a specialised subject that uses systemic objectives to improve performance and decision-making.

IV. RESULT ANALYSIS:

Soil parameters and a crop database are used in the proposed model. The ideal crop for the specific soil is recommended using machine learning algorithms. Gradient Boosting was the most accurate of the algorithms we tested. The accuracy of each algorithm used is listed below.

TABLE I. ACCURACY OF THE SYSTEM

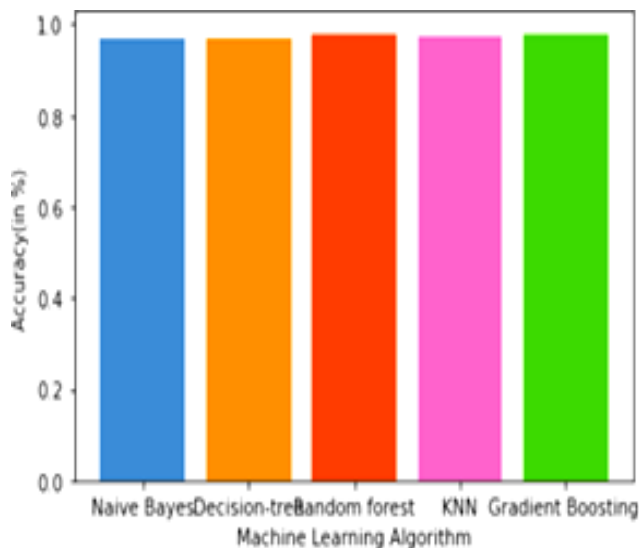| Algorithms	Accuracy
KNN	97.45
Decision Tree	96.72
Naïve Bayes	97.09
Random Forest	98
Gradient Boosting	98.18

Fig 2.Accuracy comparission.

V. CONCLUSION

As we all know, much agricultural research has been conducted and continues to be conducted in order to improve productivity, boost the Indian economy, and, most importantly, assist farmers in increasing their income. In order to accomplish this, proposed system will advise farmers on the best crop to cultivate on their land. So that farmers can profit from it.

REFERENCES

[1] Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M R, "Crop Prediction using Machine Learning Approaches", International Journal of Engineering Research and Technology, vol.9 Issue 08, August-2020 ISSN: 2278-0181.

[2] Jayaprakash, S., Nagarajan, M.D., Prado, R.P.D., Subramanian, S. and Divakarachari, P.B., 2021. A systematic review of energy management strategies for resource allocation in the cloud: Clustering, optimization and machine learning. *Energies*, *14*(17), p.5322.

[3] Zeel Doshi, Subhash Nadkarni, Rashi Agarwal and Neepa Shah, "AgroConsultant: Intelligent Crop Recommendation System using Machine Learning Algorithms",2018 Fourth International Conference on Computing Communication Control and Automation.

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

[5] Shilpa Mangesh Pande, Prem Kumar Ramesh, Anmol, B R Aishwarya, Karuna Rohilla and Kumar Shaurya, "Crop Recommendation using Machine Learning Approach", Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021).

[6] Mayank Champaneri, Chaitanya Chandvidkar, Darpan Chachpara and Mansing Rathod, "Crop Yield Prediction using Machine Learning".

[7] Kamatchi, S. Bangaru, and R. Parvathi. "Improvement of Crop Production Using Recommender System by Weather Forecasts." Procedia Computer Science 165 (2019): 724-732.

[8] Bondre, Devdatta A., and Santosh Mahagaonkar. "Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms." International Journal of Engineering Applied Sciences and Technology 4, no. 5 (2019): 371-376.

[9] Suresh, G., A. Senthil Kumar, S. Lekashri, and R. Manikandan. "Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming." International Journal of Modern Agriculture 10, no. 1 (2021): 906-914.

[10] Reddy, D. Anantha, Bhagyashri Dadore, and Aarti Watekar. "Crop recommendation system to maximize crop yield in ramtek region using machine learning." International Journal of Scientific Research in Science and Technology 6, no. 1 (2019): 485-489.

[11] Pudumalar, S., E. Ramanujam, R. Harine Rajashree, C. Kavya, T. Kiruthika, and J. Nisha. "Crop recommendation system for precision agriculture." In 2016 Eighth International Conference on Advanced Computing (ICoAC), pp. 32-36. IEEE, 2017.

[12] Azizkhan F Pathan, Chetana Prakash, Attention-based position-aware framework for aspect-based opinion mining using bidirectional long short-term memory, Journal of King Saud University - Computer and Information Sciences, 2021,ISSN1319-1578.

[13] Azizkhan F Pathan, Chetana Prakash, Unsupervised Aspect Extraction Algorithm for opinion mining using topic modeling, Global Transitions Proceedings, Volume 2, Issue 2, 2021, pp. 492-499, ISSN 2666-285X

[14] Dighe, Deepti, Harshada Joshi, Aishwarya Katkar, Sneha Patil, and Shrikant Kokate. "Survey of Crop Recommendation Systems." (2018).

[15] Jaiswal, Sapna, Tejaswi Kharade, Nikita Kotambe, and Shilpa Shinde. "Collaborative Recommendation System For Agriculture Sector." In ITM Web of Conferences, vol. 32. EDP Sciences, 2020.