IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 7.047 ⁜ Vol. 10, Issue 2, February 2022

DOI: 10.17148/IJIREEICE.2022.10239

Crop Recommendation using Machine Learning Techniques

Syed Suhaila S[1]

Assistant Professor, Department of Computer Science Engineering, Alagappa Chettiar Government College of Engineering and Technology, Karaikudi, India[1]

Abstract: Agriculture is one of the most significant industries in India, forming the backbone of the nation's economy and contributing substantially to its growth and development. It provides employment to millions and ensures food security for the population. India is renowned for its diverse production of agricultural crops, making it a global leader in the sector. Among the many factors that influence agricultural productivity, soil plays a pivotal role. Soil, being a non-renewable, dynamic natural resource, is essential for the cultivation of crops and sustains life on Earth. In earlier times, farmers relied heavily on their experience and traditional knowledge to decide which crops to cultivate. This experience-based approach enabled them to assess the suitability of crops for their land. However, with changing times, rapid urbanization, and technological advancements, farmers are increasingly unable to make precise decisions about crop selection based solely on soil characteristics and environmental factors. This gap has necessitated the development of a robust system to assist farmers in choosing the most appropriate crop for their land. To address this challenge, a crop recommendation system has been introduced, leveraging advanced machine learning algorithms. These algorithms analyze soil features, climatic conditions, and other attributes to provide tailored crop recommendations. Key algorithms used in the system include K-Nearest Neighbors (KNN), Decision Tree, Random Forest, Naive Bayes, and Gradient Boosting. Each of these algorithms plays a specific role in enhancing the accuracy and reliability of the recommendations. For instance, KNN identifies similarities with historical data to suggest the best crop, while Decision Trees and Random Forests provide logical classification and ensemble predictions. Naive Bayes offers probabilistic insights, and Gradient Boosting improves the system's performance by minimizing errors iteratively. By incorporating such advanced machine learning techniques, the system empowers farmers with data-driven insights, ensuring better resource utilization, higher yields, and improved sustainability in agriculture. This approach not only enhances productivity but also contributes to the long-term health of the environment and the agricultural sector.

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

I. INTRODUCTION

Agriculture forms the backbone of the Indian economy and plays a pivotal role in sustaining the livelihoods of a large segment of the population. It serves as the primary source of income for millions of people and contributes significantly to the nation's GDP. Over 60% of India's land is utilized for agriculture, which supports the nutritional needs of approximately 1.3 billion people. Despite the vast agricultural landscape and diverse climatic conditions, crop yields per hectare in India remain relatively low when compared to global standards. Several factors influence agricultural productivity, including soil properties, topography, irrigation, weather patterns, and fertilizer management. Soil, as a non-renewable and dynamic natural resource, is fundamental to agriculture. It serves as the foundation for crop cultivation, providing essential nutrients, water, oxygen, and physical support to plant roots. Healthy soil is integral to achieving good food production and ensuring sustainable agricultural practices. India boasts a variety of soil types, such as alluvial, black, red, and laterite soils, each suited to specific crops. For instance, black soil is ideal for sugarcane and sunflower, while laterite soil supports pulses, tea, and coffee cultivation. However, despite the abundance of natural resources, Indian farmers often rely on traditional methods and experience-based knowledge to make decisions about crop selection. This approach is no longer adequate to address the challenges posed by modern agriculture, including resource scarcity, climate change, and the increasing demand for food. [1-2].

To overcome these challenges, Precision Agriculture (PA) has emerged as a transformative approach. PA involves the precise utilization of agricultural inputs, such as seeds, water, fertilizers, and pesticides, to maximize crop yield and

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 7.047 ⁜ Vol. 10, Issue 2, February 2022
DOI: 10.17148/IJIREEICE.2022.10239

quality. By integrating advanced technologies such as sensors, data analytics, and mapping tools, PA enables farmers to optimize resource usage, conserve natural resources, and reduce environmental impact. PA technologies not only increase productivity but also contribute to sustainable agriculture by addressing ecological and economic constraints. A key component of PA is the application of artificial intelligence (AI) and machine learning (ML) techniques. ML, a subfield of AI, empowers machines to mimic human intelligence and solve complex problems autonomously. In the context of agriculture, ML provides practical solutions for crop recommendation, yield prediction, and resource management. Machine learning systems begin with data collection, which may include information on soil characteristics, climatic conditions, rainfall, temperature, and agro-ecological factors. This data is processed and used to train ML models, which then identify patterns and make predictions based on the input data. The more comprehensive the dataset, the more accurate the predictions become.

Various machine learning algorithms, such as K-Nearest Neighbors (KNN), Decision Tree, Random Forest, Naive Bayes, and Gradient Boosting, have been successfully employed in crop recommendation systems. These algorithms analyze complex datasets and provide tailored recommendations to farmers, helping them select the most suitable crops for their land. For example, Decision Trees and Random Forests classify soil and climatic attributes, while Gradient Boosting enhances predictive accuracy by iteratively minimizing errors. By leveraging these techniques, farmers can make informed decisions about crop selection, ultimately improving productivity, profitability, and sustainability. Crop recommendation systems consider a range of factors, including soil properties, applied rainfall, crop rotation, land preparation, and uncontrollable variables such as weather conditions. These systems serve as a valuable tool for Indian farmers, enabling them to transition from traditional farming practices to data-driven agricultural planning. By integrating ML-based solutions, farmers can optimize the use of resources, reduce ecological harm, and ensure food security for the growing population.

II. LITERATURE SURVEY

A research paper by Rashi Agarwal highlighted the importance of machine learning in supporting farmers' decisions on crop selection by factoring in environmental and geographical elements. This research emphasizes the significance of accurate predictions, suggesting that neural networks provide the best results among the tested algorithms. Our system adopts similar machine learning techniques to offer precise crop recommendations based on regional data and weather patterns, ensuring the best possible crop selection for Indian farmers. The study utilized decision trees, K-Nearest Neighbors (KNN), Random Forests, and neural networks. Among these methods, the neural network demonstrated the highest accuracy. [3]

Mayank Champaneri conducted an in-depth study on crop yield prediction utilizing data mining techniques, focusing specifically on the random forest classifier. This machine learning algorithm was chosen due to its strong capabilities in handling both classification and regression tasks, making it well-suited for predicting crop yields based on various input features. One of the key contributions of his research was the development of an easy-to-use web-based platform that allows farmers or users to predict the potential yield of their selected crops. By inputting climate-related data specific to their region, the platform can offer predictions tailored to the environmental conditions of that area. The use of such predictive models aligns closely with the goals of our project, which also aims to provide farmers with an accessible, user-friendly platform. Just as Champaneri's platform empowers users to make informed decisions based on climate data, our system leverages similar machine learning algorithms to generate reliable crop yield predictions. The real-time integration of climate data allows our system to offer precise, context-aware crop recommendations to farmers, helping them maximize productivity and make more informed decisions on what crops to plant based on current and future environmental conditions. [4]

Priyadharshini A explored the use of machine learning algorithms in her research study. This technology helps minimize crop failure and boosts productivity by assisting farmers in selecting the right crops and providing data that is typically not maintained by conventional farmers. Various machine learning algorithms were tested, with the neural network achieving the highest accuracy among them. [5]

In her research article, Shilpa Mangesh Pande introduces a practical and farmer-friendly crop production forecasting system. The proposed technology is accessible to farmers through a mobile application, with the user's location determined via GPS. A comparison of various algorithms was conducted to assess crop yield prediction accuracy. The Random Forest (RF) algorithm proved to be the most effective for the given dataset, achieving an accuracy rate of 95%.[6]

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 7.047 ⁜ Vol. 10, Issue 2, February 2022
DOI: 10.17148/IJIREEICE.2022.10239

Farmers across India can easily utilize this intelligent crop recommendation system Zeel Doshi. This system is designed to assist farmers in making well-informed decisions regarding which crops to cultivate by considering a variety of geographical and environmental factors. These factors include soil characteristics, climate conditions, and regional weather patterns, all of which play a critical role in determining the suitability of specific crops for a given area. By leveraging this data, the system helps farmers optimize their crop selection, ensuring that they choose the best crops based on local environmental conditions and, consequently, maximize crop yield and quality. Additionally, a secondary system, the Rainfall Predictor, can be integrated to forecast rainfall for the upcoming 12 months. These models are highly effective for real-time and practical applications, thanks to their impressive accuracy rates.[7]

III. PROPOSED METHOD

The proposed system architecture is shown in Figure 1. The proposed approach will recommend the optimum crop based on a few soil factors.

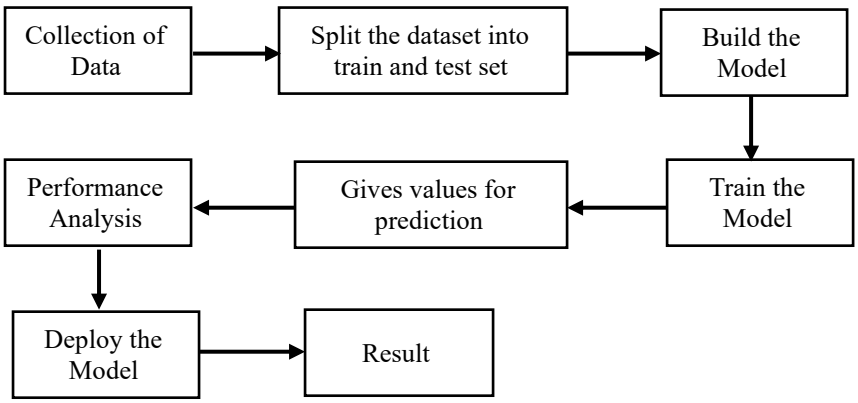

Fig 1: Proposed System Architecture

The technique of the suggested system is made up of numerous blocks, as indicated in Fig 1.

Data Collection:
Data collection is a fundamental and widely used method for gathering and analyzing information from various sources. In the context of crop recommendation systems, collecting accurate and relevant data is crucial for the system's effectiveness. The data used must meet specific criteria to ensure its quality and relevance for generating reliable crop recommendations. 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 collecting the data, the next step is to pre-process it to ensure it is ready for model training. This involves cleaning the dataset by addressing issues such as redundancy and missing values. Redundant features are removed to simplify the dataset and avoid unnecessary complexity. Missing data is handled by either removing rows or columns with excessive missing values or imputing them with appropriate substitutes, such as the mean or median. Any 'NaN' (Not a Number) values are also addressed, either by dropping or replacing them with meaningful data to maintain consistency. This process ensures that the dataset is clean, complete, and suitable for training the model.

Training Set:
A training set is a labelled dataset that includes both input and output vectors. It is used to train machine learning models by providing examples with known outcomes. The model learns to make predictions by identifying patterns between the input and corresponding output. Supervised machine learning algorithms are employed during this phase to adjust the model's parameters based on the input-output relationships in the training set. This process allows the model to generalize and make predictions on new, unseen data.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 7.047 ⁜ Vol. 10, Issue 2, February 2022
DOI: 10.17148/IJIREEICE.2022.10239

Machine Learning Algorithm:

Machine learning prediction algorithms rely on highly accurate estimations derived from previously learned data. Predictive analytics involves using historical data, statistical methods, and machine learning techniques to forecast future outcomes. The goal is not just to understand past events but to provide the best possible solutions and predict what is likely to happen in the future. By learning from historical patterns, these algorithms help anticipate trends, enabling better decision-making and more informed predictions. [8-9].

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

K-Nearest Neighbour Classifier:

The K nearest neighbor algorithm is the simplest algorithm. It can be used to solve problems based on classification and regression. Commonly used in image recognition technology. Simple recommendation and decision systems. Online platforms such as Amazon or Netflix use KNN to recommend various books for users to buy products or watch movies. KNN are based on well-established mathematical concepts. The first thing to do when implementing KNN is to convert the data items to their exact values. This is how it works by pointing the space between the numeric rate numeric rate of points. The method to find the distance is the Euclidean distance. 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.

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

Decision Tree:

A decision tree is a supervised machine learning algorithm used for both classification and regression tasks. It builds a tree-like model where each decision and its possible outcomes, including contingencies, costs, and utilities, are represented. The algorithm works by recursively dividing the dataset into subsets based on the feature values. These subsets are formed in a way that members within each group are either in the same class or have similar output values. Each internal node in the tree represents a decision based on a feature, while each leaf node represents either a class label (for classification) or a predicted value (for regression). The process begins at the root node, and the algorithm recursively splits the data using the features that yield the highest information gain or best split. Information gain is a metric that measures the difference between the impurity of the parent node and the sum of the impurities of its child nodes, helping to identify the most informative features for partitioning the data.

$$\text{Entropy: } H(S) = -\sum Pi(S) \log 2\ Pi(S) \quad (2)$$

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

Naive Bayes:

Bayes' theorem is used to construct a probabilistic classifier known as Naive Bayes. This classifier makes a simplifying assumption that each feature is independent of the others, given the class variable. In other words, Naive Bayes assumes that the presence of a particular feature in a data point is unrelated to the presence of any other feature, once the class label is known. Despite this "naive" assumption of independence, Naive Bayes classifiers perform surprisingly well, especially in tasks like text classification, where features (such as words) can often be treated as independent. The model calculates the probability of each class given the features and chooses the class with the highest probability.

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

Random Forest:

Random Forest is a machine learning algorithm that falls under supervised learning techniques, capable of handling both classification and regression tasks. It is built on the concept of ensemble learning, which involves combining multiple classifiers to address complex problems and enhance model performance. The Random Forest algorithm works by constructing a collection of decision trees, where each tree is trained on a random subset of the data, and the final output is determined by aggregating the results of all trees. One of the key advantages of Random Forest is that it requires less training time compared to other algorithms, while maintaining high prediction accuracy, even with large datasets.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 7.047 ⁜ Vol. 10, Issue 2, February 2022
DOI: 10.17148/IJIREEICE.2022.10239

Additionally, Random Forest is resilient to missing data, as it can still produce accurate predictions when a significant portion of the data is unavailable.

Gradient Boosting:

A gradient-boosting classifier is a machine learning algorithm that belongs to the ensemble methods category, primarily used for classification tasks. It works by combining multiple weak classifiers to form a single strong classifier. The algorithm iteratively adds new decision trees, where each tree is designed to correct the errors made by the previous ones, improving the overall model's performance with each step. Gradient boosting is a supervised learning technique that can be applied to both classification and regression problems. This approach gradually builds the model by focusing on the misclassified data points in each iteration. In the context of crop recommendations, a gradient-boosting model can be used to predict optimal cultivation practices by analyzing factors such as NPK (Nitrogen, Phosphorus, Potassium) levels, temperature, humidity, and soil pH. The model would recommend the best crop cultivation strategies based on these soil parameters, ensuring optimal growth conditions for the crops.

Performance Analysis:

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

IV. RESULTS AND DISCUSSION

In the proposed model, soil parameters such as pH, temperature, humidity, and NPK levels, along with a comprehensive crop database, are leveraged to recommend the most appropriate crop for a particular soil type. These parameters are crucial in determining the fertility and suitability of the soil for different crops. By applying machine learning algorithms, the model processes the data to identify patterns and make predictions on which crop would yield the best results under the given conditions. The algorithms used in this model include decision trees, Random Forest, Naive Bayes, and Gradient Boosting, each bringing its own strengths to the prediction process. Among all the algorithms tested, Gradient Boosting proved to be the most accurate in providing crop recommendations. This algorithm's ability to iteratively improve predictions by correcting the errors of previous iterations made it particularly effective in handling complex, non-linear relationships between soil conditions and crop suitability. The high accuracy of Gradient Boosting was evident when compared to the other models, as it outperformed them in terms of prediction reliability and precision. The accuracy of each algorithm tested is detailed below, offering a comparative analysis of how each one performed in terms of making the correct crop recommendations based on the given soil parameters. This comparison highlights the effectiveness of Gradient Boosting as the most reliable choice for this application.

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

Table 1: Accuracy the System

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 7.047 ⁜ Vol. 10, Issue 2, February 2022
DOI: 10.17148/IJIREEICE.2022.10239

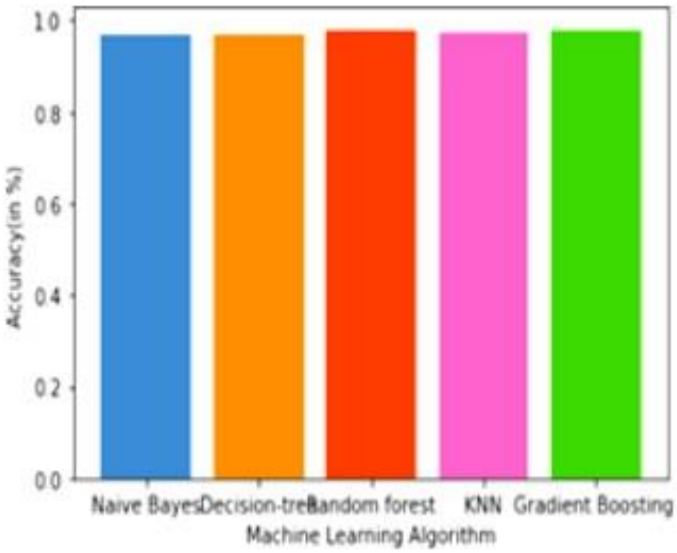

Fig 2: Accuracy Comparison

V. CONCLUSION

In this article, we propose and successfully implement a crop recommendation system designed to be user-friendly for farmers worldwide. This system assists farmers in making informed decisions about which crops to plant based on various parameters, including nitrogen, phosphorus, potassium, pH, humidity, temperature, and rainfall. By utilizing this research, we aim to boost agricultural productivity, which, in turn, will generate higher profits for farmers. With the right crop selection, farmers can optimize their harvests, increasing both their personal profit and contributing to national growth.

The proposed system is expected to have a significant positive impact on the Indian economy by helping farmers make better decisions, leading to increased income and sustainable agricultural practices. We recommend that farmers adopt this system to ensure they plant the most suitable crops for their land, thereby maximizing their yield and financial benefit. This approach will not only help individual farmers but also support the broader agricultural landscape, benefiting the economy.

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

[5]. 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).

[6]. 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)

[7]. Zeel Doshi, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms" in 2018 IEEE.

DOI: 10.17148/IJIREEICE.2022.10239

[8]. 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.

[9]. 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.

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

[11]. 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.

[12]. 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