IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

Crop Recommendation System using Machine Learning

Prof. Sana Pathan, Aditya Pawar, Aaditya Bandhane, Pranav Pol, Manoj Rathod
Department of Information Technology
Indala College of Engineering, Kalyan, Mumbai, India

Abstract: *India's agricultural sector remains a cornerstone of the national economy, contributing around 17% to the country's GDP and employing over 60% of the workforce. Despite advancements in agricultural technologies such as vertical farming, precision agriculture, and smart irrigation systems, many Indian farmers continue to rely on conventional farming practices and seasonal patterns. This often leads to suboptimal crop yields, especially in the face of unpredictable climate changes. Our research addresses this gap by introducing a data-driven crop recommendation system that empowers farmers to make informed decisions based on real-time environmental conditions. By analyzing key agricultural factors such as soil nutrient content (Nitrogen, Phosphorus, Potassium), soil pH, humidity, and rainfall, we predict the most suitable crops for cultivation. We explore and compare the effectiveness of multiple machine learning models, including Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB). Our approach aims to promote adaptive farming, enhance productivity, and support sustainable agriculture in India..*

Keywords: Smart agriculture, Crop recommendation, Soil nutrients, NPK, Soil pH, Humidity, Rainfall, Climate resilience, Machine Learning (ML), Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), Gaussian Naïve Bayes (GNB).

Abbreviations—DT: Decision Tree, GNB: Gaussian Naïve Bayes, SVM: Support Vector Machine, LR: Logistic Regression, NPK: Nitrogen-Phosphorus-Potassium

I. INTRODUCTION

Agriculture remains the backbone of India's economy, contributing significantly to national GDP and employing a majority of the rural population. Despite technological advancements, a substantial portion of Indian farmers still rely on traditional farming techniques, which are often vulnerable to unpredictable weather patterns, soil degradation, and inefficient crop selection. These challenges can result in poor yields, economic losses, and in extreme cases, farmer distress and suicides due to unmanageable debt.

In recent years, the integration of data-driven techniques such as precision agriculture has begun to transform farming practices. Precision agriculture leverages localized data to provide actionable insights tailored to specific plots of land. One of its critical applications is crop recommendation, which aims to help farmers select the most suitable crop for cultivation based on environmental and soil conditions. However, the accuracy of these recommendations is crucial, as poor predictions can lead to substantial financial and resource loss.

Machine Learning (ML) has emerged as a powerful tool in addressing this challenge. By learning from historical agricultural data, ML models can identify patterns and relationships among key factors such as soil nutrients (Nitrogen, Phosphorus, Potassium), pH value, humidity, temperature, and rainfall. Using this information, the system can suggest the optimal crop that would thrive under the given conditions, thereby increasing productivity and sustainability.

This research proposes a crop recommendation system powered by supervised machine learning algorithms. The system analyzes key agricultural parameters and suggests the best-suited crop from a set of 22 commonly cultivated crops in India, including rice, maize, chickpea, banana, watermelon, cotton, and coffee. We evaluate the performance of various ML algorithms—Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB)—to determine their effectiveness in providing accurate crop recommendations.

The objective of this study is to empower farmers with intelligent, data-driven decision-making tools that promote efficient land use, increase crop yield, and support sustainable agricultural development.

II. LITERATURE REVIEW

The application of machine learning in agriculture has gained momentum in recent years, with a significant focus on crop prediction and recommendation systems. Various studies have demonstrated the effectiveness of different machine learning models in enhancing crop yield, improving decision-making, and supporting sustainable agricultural practices.

Kumar et al. [1] presented a system for crop yield prediction using historical environmental data such as temperature, humidity, pH, and rainfall. Their approach leveraged Decision Tree and Random Forest algorithms, with Random Forest yielding the highest accuracy, enabling farmers to make more profitable decisions based on precise crop recommendations.

Suresh et al. [9] developed a digital farming solution that used Support Vector Machine (SVM) for predicting suitable crops based on parameters like nitrogen, phosphorus, potassium, and pH levels. Their model also estimated the required fertilizer quantities for specific crops, thus aiding in efficient resource utilization and higher productivity.

Reddy et al. [10] proposed a region-specific crop recommendation system using machine learning techniques such as Random Forest, CHAID, K-Nearest Neighbours (KNN), and Naïve Bayes. Their model considered soil characteristics, soil types, and crop yield data to suggest the most suitable crops for specific states and districts, helping farmers optimize their farming strategy based on localized weather and soil conditions.

Rajak et al. [14] extended the scope of recommendation systems by including parameters like soil texture, drainage, water retention, and erosion. Their system tested multiple algorithms—SVM, Artificial Neural Networks (ANN), Random Forest, and Naïve Bayes—to recommend crops like paddy, sugarcane, and vegetables. The research highlighted the role of machine learning in improving productivity while reducing chemical use and preventing soil degradation.

Doshi et al. [15] introduced Agro Consultant, an intelligent crop recommendation system that integrates a crop suitability predictor and a rainfall predictor. Using Decision Tree, K-NN, Random Forest, and Neural Networks, their system achieved 91% accuracy for crop prediction. The model was trained on diverse features such as aquifer thickness, topsoil characteristics, temperature, and rainfall, enabling multi-label classification across 20 different crops.

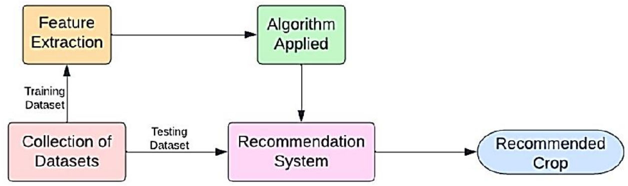
Dighe et al. [18] conducted a comprehensive survey of crop recommendation systems, reviewing a wide range of machine learning techniques including CHAID, KNN, K-means, Decision Tree, Neural Networks, and Naïve Bayes. Their study emphasized the role of scalable platforms like Hadoop for handling large-scale agricultural datasets and enhancing the computational efficiency of predictive models.

Kulkarni et al. [21] explored the use of assembling techniques to boost the accuracy of crop recommendation systems. Their model incorporated algorithms like Random Forest, Naïve Bayes, and Linear SVM, achieving a remarkable accuracy of 99.91%. The study also classified crops into Kharif and Rabi categories based on soil type, rainfall, and surface temperature, providing highly contextual recommendations for seasonal planning.

From these studies, it is evident that machine learning provides a robust foundation for developing intelligent crop recommendation systems. By leveraging diverse algorithms and environmental datasets, these systems have the potential to significantly enhance crop productivity, resource efficiency, and economic outcomes for farmers. However, achieving high accuracy and regional adaptability remains a challenge, thus motivating further research into hybrid and ensemble models tailored for local conditions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

III. METHODOLOGY



Block Diagram of Overall Methodology of Proposed System

In our framework, we have proposed a procedure that is separated into various stages as appeared in Figure .
The five phases are as per the following:
1) Collection of Datasets
2) Pre-processing (Noise Removal)
3) Feature Extraction
4) Applied Machine Learning Algorithm
5) Recommendation System
6) Recommended Crop Flow of the Proposed System As demonstrated in the figure, the methodology to extract the sentiment contains the several steps that are described below:

Data Collection:

The dataset used in this study was obtained from the Kaggle open-source platform [27], specifically from the “Crop Recommendation Dataset.” It contains 2,200 instances and includes the following features:
Soil nutrients**:** Nitrogen (N), Phosphorus (P), Potassium (K)
Environmental conditions**:** Temperature, Humidity, Rainfall
Soil pH value
Each instance is labelled with one of 22 crops commonly grown in India, such as rice, maize, banana, coconut, cotton, coffee, lentil, jute, and others. The dataset was split into training and testing subsets using an 80:20 ratio to build and evaluate the model.

Data Preprocessing:

Raw datasets often include noise, missing values, or inconsistent data. Therefore, preprocessing is essential to ensure data quality and enhance model performance. In this stage:
Noise removal was conducted using tools such as Power BI to eliminate outliers, local min-max anomalies, and junk values.
Normalizationwas applied to scale the data, ensuring all features contribute equally to model training.
Redundant and incomplete entries were removed to preserve data consistency and accuracy.

Feature Selection:

Feature selection plays a vital role in reducing dimensionality and improving model efficiency. In this step:
Only the most relevant features were retained, namely N, P, K, temperature, humidity, pH, and rainfall.
Irrelevant or redundant attributes were filtered out to optimize classification performance and reduce computational cost

Machine Learning Models:

To identify the most accurate approach for crop prediction, several supervised machine learning algorithms were implemented and compared. The models used are:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

Decision Tree Classifier (DT)

Decision Tree is a non-parametric supervised learning method used for classification. It splits the data into subsets based on the value of input features, building a tree-like structure of decisions. Each node represents a test on an attribute, and each branch represents an outcome of the test. Final decisions are made at leaf nodes. The algorithm was implemented using the `DecisionTreeClassifier` from the `sklearn.tree` module.

Steps:

Import `DecisionTreeClassifier`
Create the classifier object
Train it using the training dataset

Support Vector Machine (SVM)

SVM is a robust classification algorithm that finds the optimal hyperplane that separates data points of different classes with the maximum margin. In this project, SVM was used to handle the multi-class classification problem by transforming the data into higher-dimensional space.

Key Advantages:

Effective in high-dimensional spaces
Works well with clear margin separation

Logistic Regression (LR)

Logistic Regression is a statistical model used for binary and multiclass classification tasks. It estimates the probability of a target variable using a logistic (sigmoid) function. Although primarily used for binary outputs, it can be extended to handle multiple classes using one-vs-rest or multinomial approaches.

Application:

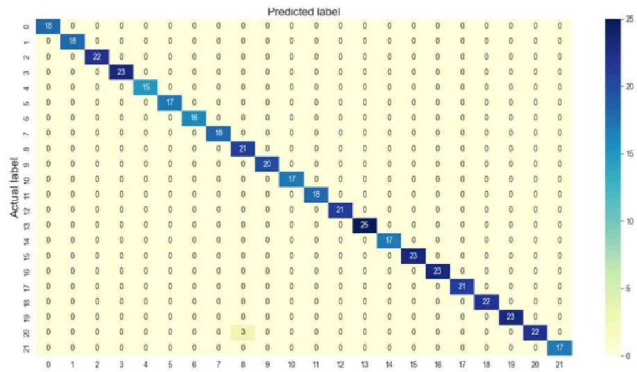
Used for its interpretability and efficiency
Handles multiclass output through softmax classification
D. Gaussian Naïve Bayes (GNB)
Gaussian Naïve Bayes is a probabilistic classifier based on Bayes' Theorem, assuming feature independence. It models the likelihood of features using a Gaussian distribution. Despite its simplicity, GNB performs remarkably well in many real-world scenarios.

Highlights:

Assumes normal distribution of input features
Highly efficient with small training datasets
Suitable for continuous input variables like temperature and pH

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

IV. RESULT AND ANALYSIS



Accuracy Comparison

The confusion matrix used to determine the performance of the classification models for a given set of test data. It can only be determined if the true values for test data are known.

The matrix itself can be easily understood, but the related terminologies may be confusing.

Training Accuracy Score	***99.5%***
Validation Accuracy Score	***99.3%***

Algorithm vice Accuracy Result in Percentage

Based on the results provided, we can see that the model performs great with 99.3% accuracy.

V. CONCLUSION

In this research, we proposed and implemented a machine learning-based crop recommendation system aimed at assisting Indian farmers in selecting the most suitable crop based on environmental and soil conditions. The system takes into account essential agricultural parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), soil pH, temperature, humidity, and rainfall to predict the most appropriate crop for a given region.

By leveraging machine learning techniques, this system has the potential to significantly improve crop productivity and enhance profitability for farmers. Through comparative analysis, we found that the **Decision Tree** and **Gaussian Naïve Bayes** algorithms provided the highest prediction accuracy among the models tested. These models effectively learned from historical data and generated reliable recommendations.

The implementation of this system not only empowers farmers with data-driven insights but also contributes to improved resource utilization, reduced crop failure risks, and a more sustainable agricultural ecosystem. Ultimately, this research supports the broader goal of increasing agricultural efficiency and boosting the overall economic development of the country

VI. FUTURE WORK

While the current crop recommendation system demonstrates promising results, there are several opportunities for further enhancement and development:

1. Dataset Expansion and Enrichment:

Future work can focus on incorporating a larger and more diverse dataset, including additional agricultural parameters such as soil texture, organic matter content, altitude, and crop rotation history. A more comprehensive dataset can improve the robustness and accuracy of predictions.

2. Crop Health and Disease Detection Integration:
Integrating a plant disease detection module using computer vision and deep learning techniques can greatly extend the system's utility. This would enable real-time classification of crop health by analyzing images of leaves and identifying specific diseases, along with appropriate treatment recommendations.

3. Development of a User-Friendly Platform:
A web-based portal and mobile application can be developed to make the system more accessible to farmers. These platforms should offer multilingual support, voice-based input, and interactive features that make it easy for users with minimal technical knowledge to obtain accurate crop suggestions.

4. Incorporation of Real-Time Weather Data:
Future versions of the system can leverage APIs to fetch real-time weather forecasts and seasonal trends, allowing for dynamic crop suggestions that adapt to current climatic conditions.

5. Fertilizer and Water Requirement Estimation:
An additional feature that estimates the quantity and type of fertilizers and irrigation needed for the recommended crop can further support sustainable farming practices and minimize resource wastage.

REFERENCES

[1] Kumar, Y. Jeevan Nagendra, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi, "Supervised Machine Learning Approach for Crop Yield Prediction in Agriculture Sector," 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 736–741, IEEE, 2020.
[2] Nigam, Aruvansh, Saksham Garg, Archit Agrawal, and Parul Agrawal, "Crop yield prediction using machine learning algorithms," 2019 Fifth International Conference on Image Information Processing (ICIIP), pp. 125–130, IEEE, 2019.
[3] Medar, Ramesh, Vijay S. Rajpurohit, and Shweta Shweta, "Crop yield prediction using machine learning techniques," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), pp. 1–5, IEEE, 2019.
[4] Jain, Sonal, and Dharavath Ramesh, "Machine Learning Convergence for Weather Based Crop Selection," 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1–6, IEEE, 2020.
[5] Gandge, Yogesh, "A study on various data mining techniques for crop yield prediction," 2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), pp. 420–423, IEEE, 2017.
[6] Suresh, A., P. Ganesh Kumar, and M. Ramalatha, "Prediction of major crop yields of Tamil Nadu using K-means and Modified KNN," 2018 3rd International Conference on Communication and Electronics Systems (ICCES), pp. 88–93, IEEE, 2018.
[7] Kamatchi, S. Bangaru, and R. Parvathi, "Improvement of Crop Production Using Recommender System by Weather Forecasts," Procedia Computer Science, vol. 165, pp. 724–732, 2019.
[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, vol. 4, no. 5, pp. 371–376, 2019.
[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, vol. 10, no. 1, pp. 906–914, 2021.
[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, vol. 6, no. 1, pp. 485–489, 2019.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

[11] Pudumalar, S., E. Ramanujam, R. Harine Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," 2016 Eighth International Conference on Advanced Computing (ICoAC), pp. 32–36, IEEE, 2017.
[12] Garanayak, Mamata, Goutam Sahu, Sachi Nandan Mohanty, and Alok Kumar Jagadev, "Agricultural Recommendation System for Crops Using Different Machine Learning Regression Methods," International Journal of Agricultural and Environmental Information Systems (IJAEIS), vol. 12, no. 1, pp. 1–20, 2021.
[13] Jejurkar, Siddhi S., Meghna S. Bhosale, and D. N. Wavhal, "Crop Prediction and Diseases Detection Using Machine Learning," 2021.
[14] Rajak, Rohit Kumar, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, and Avinash Devare, "Crop recommendation system to maximize crop yield using machine learning technique," International Research Journal of Engineering and Technology (IRJET), vol. 4, no. 12, pp. 950–953, 2017.
[15] Doshi, Zeel, Subhash Nadkarni, Rashi Agrawal, and Neepa Shah, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), IEEE, 2018.
[16] A. A. Mahule and A. J. Agrawal, "Hybrid Method for Improving Accuracy of Crop-Type Detection using Machine Learning," Int. J., vol. 9, no. 2, 2020.
[17] K. R. Akshatha and K. S. Shreedhara, "Implementation of machine learning algorithms for crop recommendation using precision agriculture," Int. J. Res. Eng. Sci. Manag. (IJRESM), vol. 1, no. 6, pp. 58–60, 2018.
[18] D. Dighe, H. Joshi, A. Katkar, S. Patil, and S. Kokate, "Survey of Crop Recommendation Systems," 2018.
[19] S. Jaiswal, T. Kharade, N. Kotambe, and S. Shinde, "Collaborative Recommendation System for Agriculture Sector," ITM Web Conf., vol. 32, EDP Sciences, 2020.
[20] R. Ghadge, J. Kulkarni, P. More, S. Nene, and R. L. Priya, "Prediction of crop yield using machine learning," Int. Res. J. Eng. Technol. (IRJET), vol. 5, 2018.
[21] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery, "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique," Proc. 3rd Int. Conf. Computational Systems and Information Technology for Sustainable Solutions (CSITSS), pp. 114–119, IEEE, 2018.
[22] A. Kumar, S. Sarkar, and C. Pradhan, "Recommendation system for crop identification and pest control technique in agriculture," Proc. Int. Conf. Communication and Signal Processing (ICCSP), pp. 185–189, IEEE, 2019.
[23] S. M. Pande, P. K. Ramesh, A. Anmol, B. R. Aishwarya, K. Rohilla, and K. Shaurya, "Crop Recommender System Using Machine Learning Approach," Proc. 5th Int. Conf. Computing Methodologies and Communication (ICCMC), pp. 1066–1071, IEEE, 2021, doi: 10.1109/ICCMC51019.2021.9418351.
[24] S. Bhuvan, A. Jain, and J. Sanjeetha, "DigiFarm – A Machine Learning Based Holistic Crop Prediction Platform for Farmers," NVEO - Natural Volatiles & Essential Oils Journal, pp. 11112–11129, 2021.
[25] M. Qiao et al., "Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks," Int. J. Appl. Earth Obs. Geoinf., vol. 102, p. 102436, 2021.
[26] V. Meshram et al., "Machine learning in agriculture domain: A state-of-art survey," Artif. Intell. Life Sci., vol. 1, p. 100010, 2021.
[27] Atharva Ingle, "Crop Recommendation Dataset," Kaggle, [Online].Available:https://www.kaggle.com/datasets/atharvaingle/croprecommendation-dataset