

RESEARCH ARTICLE

AI-based Smart Crop Recommendation System for Sustainable Agricultural Production: A Data-driven Approach to Minimize Resource Use and Maximize Yield.

Aarthi S^{*1}, Manimegalai S¹, Sakthivel R²

¹Dept. of Physical Sciences & Information Technology, AEC&RI, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India.

²Dept. of Civil Engineering, Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India.

ABSTRACT

This paper presents an innovative crop recommendation system powered by artificial intelligence to support sustainable agricultural practices. The proposed solution uses multiple machine learning algorithms to predict the most suitable crops based on key environmental and soil parameters, including nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall. Several models were evaluated, including Logistic Regression, Decision Trees, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes, Random Forests, and XGBoost. Among these, the Random Forest classifier achieved the highest accuracy at 98.2 percent. A web-based application was developed using Flask, providing an interactive and accessible platform for farmers to receive categorized crop suggestions as Recommended, Slightly Recommended, and Not Recommended. The system is designed for scalability, ease of use, and real-time responsiveness, offering a promising tool for data-driven, resource-efficient, and yield-optimized farming.

Received: 14 May 2025

Revised: 27 May 2025

Accepted: 16 Jun 2025

Keywords: Artificial intelligence, Crop recommendation system, Machine learning, Precision agriculture, Sustainable farming, Soil nutrient analysis.

INTRODUCTION

Agriculture forms the cornerstone of food security, rural livelihoods, and economic development worldwide. However, traditional crop selection methods in many regions still rely heavily on farmer intuition and outdated practices, which often lead to inefficient resource use, poor yields, and environmental strain. With the advent of artificial intelligence (AI) and machine learning (ML), agriculture is undergoing a data-driven transformation, enabling precision farming techniques that improve productivity while preserving natural resources (Chlingaryan *et al.*, 2018; Kamilaris and Prenafeta-Boldú, 2018).

Machine learning algorithms can analyze large volumes of soil and climate data to extract patterns and generate predictive insights that support more informed agricultural decisions. For instance, Random Forest and other ensemble methods have been proven highly effective in predicting regional and global crop yields under diverse conditions (Jeong *et al.*, 2016). These models outperform traditional statistical methods in terms of accuracy, scalability, and noise resilience in real-world data.

In recent years, researchers and developers have focused on creating intelligent crop recommendation

*Corresponding author mail: ssaarthi11@gmail.com



Copyright: © The Author(s), 2025. Published by Madras Agricultural Students' Union in Madras Agricultural Journal (MAJ). This is an Open Access article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited by the user.

systems that consider key agronomic factors, including soil nutrient levels, pH, rainfall, temperature, and humidity. These parameters directly influence crop suitability and yield outcomes. By processing this data through ML models, such systems offer farmers objective, site-specific recommendations that can guide cultivation decisions and reduce the risk of crop failure (Chlingaryan *et al.*, 2018).

This study introduces an AI-based crop recommendation system that incorporates seven critical input variables: nitrogen (N), phosphorus (P), potassium (K), pH level, temperature ($^{\circ}\text{C}$), humidity (%), and rainfall (mm). Multiple machine learning models were trained and tested, including Logistic Regression, Decision Trees, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes, XGBoost, and Random Forests. Among these, the Random Forest classifier achieved the highest prediction accuracy, reaching 98.2 percent.

To make this system accessible and practical for end-users, a web application was developed using Python and Flask, with a responsive and straightforward front-end interface. Users can input local environmental and soil data and receive categorized crop suggestions: Recommended, Slightly Recommended, and Not Recommended. This tool not only supports sustainable agricultural practices but also demonstrates how AI technologies can be effectively translated into field-level solutions for smallholder and commercial farmers alike.

MATERIALS AND METHODS

Data Collection and Preprocessing

The dataset used for this study was sourced from publicly available agricultural data repositories that include region-specific soil and climatic conditions. The dataset consists of soil nutrients (Nitrogen, Phosphorus, Potassium), environmental factors (temperature in $^{\circ}\text{C}$, humidity in %, rainfall in mm), and soil pH values. Each instance in the dataset is labeled with a crop suitable for the given conditions. A total of 22 different crops were included, covering cereals, pulses, fruits, and vegetables.

The dataset was cleaned by removing null values and duplicate entries. Numerical features were normalized to improve the performance of algorithms sensitive to scale, such as K-Nearest Neighbors and Support Vector Machine.

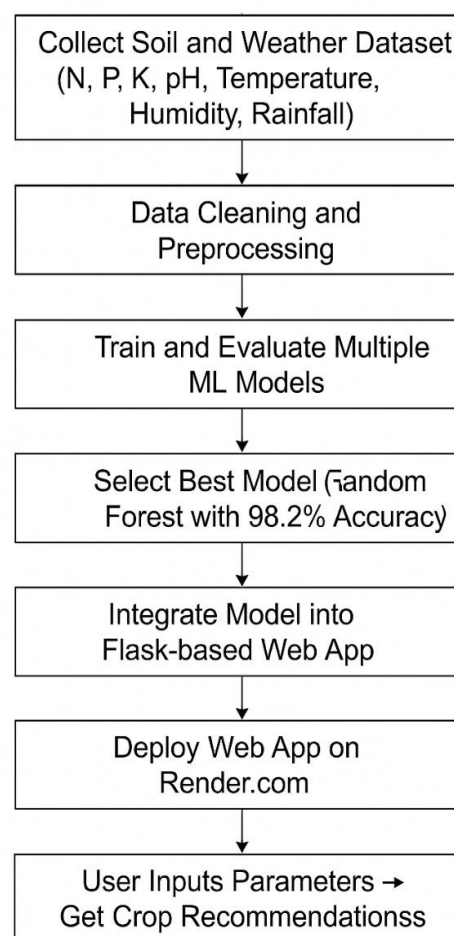


Fig 1. Methodology

Tools and Technologies

The following tools and libraries (Table 1) were used in this project:

Tool	Purpose
Python 3.10	Backend logic and ML model
scikit-learn	ML algorithm implementation
Pandas, NumPy	Data processing and analysis
Flask	Backend web framework
HTML/CSS, Bootstrap	Frontend interface
Render.com	Web deployment platform
Matplotlib/Seaborn	Visualization
Jupyter Notebook	Model training and testing

2.3 Model Selection and Evaluation

Seven machine learning algorithms were trained and evaluated on the dataset listed in Table 2.

Table 2. Models and their accuracy

Model	Accuracy (%)
Logistic Regression	85.6
Decision Tree	91.5
KNN	94.3
SVM	93.4
Naive Bayes	80.1
Random Forest	98.2
XGBoost	97.5

Each model was evaluated using accuracy as the primary metric, supported by a confusion matrix and a classification report. The dataset was split into a training set (80%) and a testing set (20%) using stratified sampling to maintain class balance.

The Random Forest Classifier achieved the highest accuracy of 98.2%, outperforming other models due to its ensemble nature and robustness against overfitting (Fig. 2).

2.4 Web Application Development

A web application was developed to make the model accessible to end-users. It allows farmers to input seven parameters (N, P, K, pH, temperature, humidity, and rainfall) through a simple form interface. Based on the input, the model classifies crops into three categories:

- Recommended Crops
- Slightly Recommended Crops
- Not Recommended Crops

- Not Recommended Crops

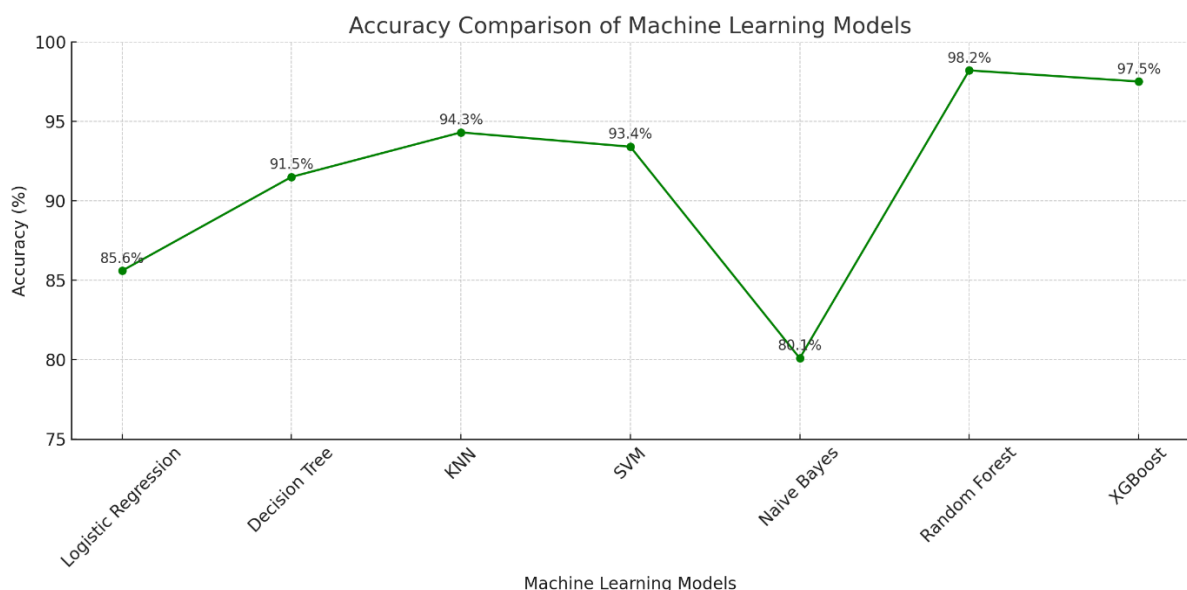
Flask handled the back-end logic, and the model was integrated using a pickle-serialized .pkl file. The application was deployed on Render.com, a free cloud hosting platform (Fig. 1).

3. RESULTS AND DISCUSSION:

The performance of multiple machine learning models was evaluated for crop recommendation using soil and environmental parameters, including nitrogen (N), phosphorus (P), potassium (K), pH, temperature, humidity, and rainfall. Each algorithm was assessed based on classification accuracy, supported by confusion matrix analysis and precision-recall metrics.

The Random Forest classifier demonstrated the highest prediction accuracy at 98.2%, outperforming other models such as Logistic Regression (85.6%), Decision Tree (91.5%), K-Nearest Neighbors (94.3%), Support Vector Machine (93.4%), Naive Bayes (80.1%), and XGBoost (97.5%).

These findings align with previous research that highlights the superior performance of Random Forest models in agricultural classification tasks due to their ensemble nature and robustness to overfitting (Biau and Scornet, 2016; Jeong *et al.*, 2016). Other studies have also found that Random Forest often outperforms traditional classifiers for crop suitability prediction and soil-based recommendations (Prity *et al.*, 2024; Shingade *et al.*, 2025).



The developed model was integrated into a web-based application using Flask and Bootstrap technologies. The app interface allows users to input soil and climate parameters and receive categorized recommendations, such as:

- Recommended Crops – ideal for given inputs
- Slightly Recommended Crops – moderately suitable
- Not Recommended Crops – unsuitable for the given conditions

A screenshot of the deployed web interface is shown below, demonstrating both the input form and prediction output:

The web application is hosted and publicly accessible at: <https://crop-recommendation-w330.onrender.com/>

The model's high accuracy and deployment readiness suggest strong potential for real-world adoption. By integrating AI predictions with user-friendly interfaces, the system facilitates data-informed farming decisions, particularly in regions where expert agronomic guidance is scarce.

While promising, the system is currently limited to static data inputs and does not account for spatial or temporal variability. Future iterations may integrate satellite data, geolocation tagging, or seasonal

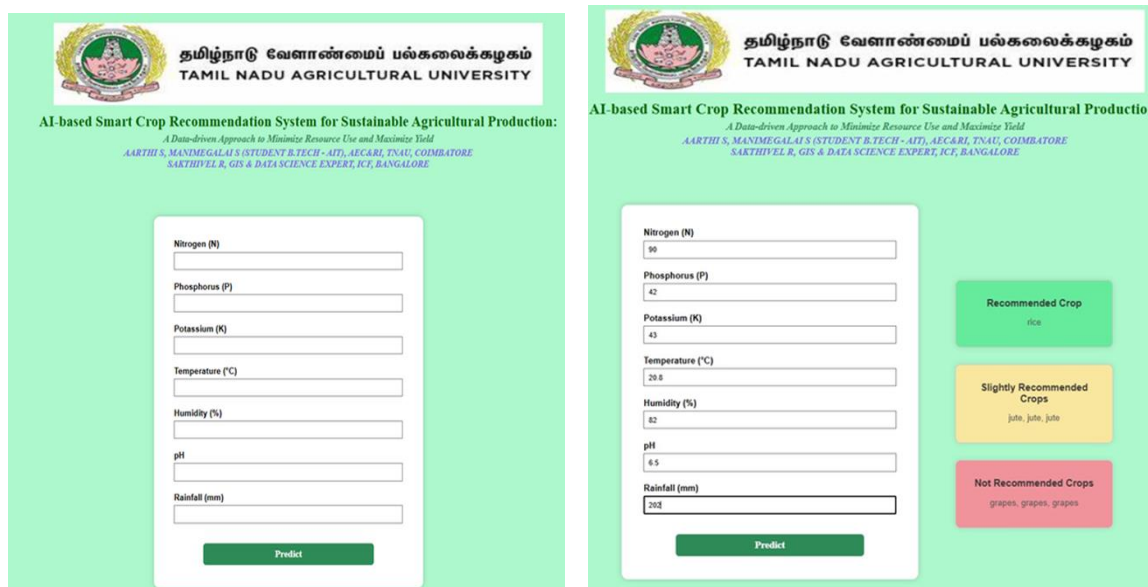


Figure 2. Web Application Interface for Crop Recommendation (Input and Output)

trends to enhance context-aware recommendations. Additionally, expanding to multilingual interfaces could improve accessibility for farmers across different regions.

4. Conclusion and Future Work

This study proposed an AI-based crop recommendation system that leverages machine learning techniques to support data-driven agricultural decision-making. By analyzing key parameters, including soil nutrient levels (N, P, K), pH, temperature, humidity, and rainfall, the system effectively predicts suitable crops for cultivation. Among the seven machine learning models evaluated, the Random Forest classifier achieved the highest accuracy (98.2%), confirming its robustness and generalizability for classification tasks in agricultural datasets.

The deployment of the model into a Flask-based web application offers an accessible and interactive platform for farmers and agricultural advisors. Users can input localized data and receive categorized crop suggestions, including Recommended, Slightly Recommended, and Not Recommended, thus enabling more informed and resource-efficient farming practices. The system represents a practical implementation of precision agriculture by translating complex model predictions into simple, actionable insights for end-users.

While the system demonstrates high performance and usability, there are still avenues for further enhancement. Currently, the model relies on static inputs and does not incorporate spatial or temporal

dynamics. In real-world agricultural settings, factors such as seasonal variability, geographical location, and real-time weather updates can significantly impact crop performance.

Future Work:

To further improve the system, the following enhancements are proposed:

- Integration of Geospatial and Remote Sensing Data: Incorporating satellite imagery and GIS layers can enable spatially explicit crop recommendations tailored to specific locations.
- Dynamic and Real-time Inputs: Linking the model to live weather APIs and soil sensors can allow for time-sensitive recommendations.
- Multilingual and Voice-enabled Interface: To improve accessibility for farmers in rural areas, the application can be extended to support regional languages and voice input/output functionalities.
- Mobile Application Development: A lightweight Android application can broaden the reach of the tool, especially in areas with limited access to computers.
- Adaptive Learning Models: Incorporating user feedback into the system can help refine model predictions over time, making the system more personalized and locally adapted.

In conclusion, this AI-powered crop recommendation system offers a scalable and impactful solution for promoting sustainable and precision-based agriculture. With continued development, it has the potential to assist millions of farmers in optimizing crop selection, conserving resources, and increasing overall productivity.

REFERENCES:

- Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151, 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>
- Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., & Kim, S. H. (2016). Random Forests for global and regional crop yield predictions. *PLOS ONE*, 11(6), e0156571. <https://doi.org/10.1371/journal.pone.0156571>
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., & Kim, S. H. (2016). Random Forests for global and regional crop yield predictions. *PLOS ONE*, 11(6), e0156571. <https://doi.org/10.1371/journal.pone.0156571>
- Prity, N., Meena, D. P., Suman, A., & Pant, D. (2024). Comparative analysis of ensemble learning models for intelligent crop prediction. *Discover Artificial Intelligence*, 4(1), 22. <https://doi.org/10.1007/s44230-024-00081-3>