International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4293- 4298
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0627

Crop Recommendation System

Ayeesha Siddiqha[1], Nisarga K.M [2], Sahithya Ullas[3], Impana H [4], Inchara M [5]
[1,2,3,4,5]Dept. of Computer Science and Engineering, Malnad College of Engineering, Hassan, India.
Emails: *ayeeshasiddiqha@gmail.com[1], nisargamonika@gmail.com[2], sahithyaullas@gmail.com[3], impanaharish21@gmail.com[4], minchara@gmail.com[5]*

Abstract

Ensuring food security gets tricky with climate change and limited resources. That means we need smart, data-based tech in agriculture. The research investigates a crop recommendation system which depends on machine learning technology. The system operates independently from any particular dataset requirements. The system recommends suitable crops through analysis of soil characteristics together with environmental factors including temperature and humidity and rainfall patterns. Regular systems stick to set datasets and number-based soil details. This one handle category inputs for soil, like sandy, loamy, or clayey. It can run on any structured data in CSV or Excel format. The setup includes a strong preprocessing step. That covers normalizing data, encoding it, and standardizing features. All this helps it adapt to different datasets. We tried out several algorithms for this. Those include Random Forest, Decision Tree, K-Nearest Neighbors, and Gaussian Naive Bayes. We checked how they performed. The results show Naive Bayes hit 99.32 percent accuracy. Random Forest came in at 99.09 percent. Both did really well. The system tackles issues with varying data pretty effectively. It provides a way to scale up that is smart and easy for users. Farmers get practical suggestions on crops from it. In the end, this pushes forward precision agriculture and ways to farm sustainably.

Keywords: *Crop Recommendation System, Machine Learning, Dataset-Agnostic Model, Soil Type Classification, Random Forest, Naive Bayes, Data Preprocessing, Precision Agriculture, Sustainable Farming, Climate-Resilient Agriculture.*

1. Introduction

Precision agriculture changes the whole game compared to old farming ways. It pays close attention to variations right inside the fields. That helps boost crop yields and make better use of resources overall. One key part of this method involves picking the right crop for certain land and weather setups. Things like the mix of nutrients in the soil matter a lot. So does the pH level there. Local weather trends play a big role too. All these elements decide how well farming turns out in the end. If farmers pick the wrong crops or handle it poorly, soil can wear out fast. Resources end up getting thrown away. Farmers face real money troubles from that kind of mistake. Machine learning steps in as a game-changer for this tricky issue. It deals with so many different variables at once. Using past records, these models spot detailed links between surroundings and what makes crops thrive. People often miss those connections on their own. This work builds and tests a system for suggesting crops. It looks at four types of supervised classification methods. The research evaluates multiple machine learning models for their ability to generate accurate predictions. In the end, the idea is to build a solid tool farmers can trust. It lets them base choices on solid data. That way, yields go up. Sustainability gets a push forward. Food supplies stay more secure for everyone [1].

2. Literature Review

2.1 Overview of Existing Research

Several studies have applied machine learning (ML) techniques for crop recommendation to support data driven and sustainable agriculture. Existing research primarily focuses on models trained on fixed datasets with numeric soil parameters, whereas recent work emphasizes adaptability, ensemble methods, and soil type integration for improved accuracy and scalability.

- F. S. P. Prity et al., Human-Centric Intelligent Systems, 2024. This study proposed an ML-based crop recommendation model using soil nutrients (N, P, K, pH) and climatic features (temperature, rainfall). Models such as Random Forest and Naive Bayes achieved accuracies above 98%. The work highlights the importance of soil-climate feature integration for reliable crop prediction.
- A. Shastri et al., Scientific Reports, 2025: Developed a supervised ML crop recommendation system using ensemble algorithms and boosting techniques. Gradient Boosting achieved an accuracy of 99.27%. The study demonstrates that ensemble models outperform individual learners in crop prediction.
- M. Bakr et al., MDPI Journal, 2025: Compared ML, Deep Learning, and LLM-based models for crop recommendation. The results showed that LLMs can perform comparably to ML models (>99% accuracy) and offer interactive prediction capabilities. It suggests potential for integrating natural-language interfaces in agriculture advisory systems.
- M. Afzal et al., Agricultural Informatics, 2025. Focused on including categorical soil-type information (sandy, loamy, clayey) in ML models. Decision Tree and Random Forest algorithms showed significant accuracy improvement (up to 99%) with soil-type encoding. This validates the importance of soil categorization in real-world prediction.
- S.Saddikuti,Integrated Crop Recommendation System, 2024 . Designed a practical ML-based system with dataset agnostic data ingestion, preprocessing, and a user friendly interface. Random Forest provided consistent accuracy with different dataset formats, emphasizing flexibility and usability for farmers.
- S. Patel et al., Engineering Preprint, 2024. Conducted a comparative analysis using Random Forest, Naive Bayes, KNN, and SVM with standard datasets. After applying imputation and feature scaling, Random Forest and Naive Bayes achieved accuracies above 99%. The study reinforces the reliability of classical ML models when proper preprocessing is performed [2].

2.2 Key Findings in this field

In the various studies out there, Random Forest and Naive Bayes tend to deliver the highest accuracy levels pretty much every time. Many of these models rely on set datasets that do not change, while only a handful can handle mixed kinds of data or inputs based on categories. That is why the new system we propose, which works without being tied to any specific dataset and includes details on soil types, really helps close this divide. It provides greater flexibility in use, better ability to scale up, and stronger support for making decisions in precision agriculture [3].

3. Proposed System

The proposed Crop Recommendation System using Machine Learning aims to help farmers identify the most suitable crops based on soil and environmental conditions. Unlike traditional models that rely solely on fixed datasets and numerical soil parameters, this system is designed with greater flexibility. It can process various dataset formats and handle both categorical inputs, such as soil types (e.g., sandy, loamy, or clayey), and numerical data, including nutrient levels and climatic factors. The system automatically identifies dataset features, performs preprocessing, trains the model using multiple ML algorithms, and generates ranked crop recommendations with confidence scores. A feedback mechanism is incorporated to refine predictions based on user input, improving the system’s performance over time. The core modules of the proposed system are:

- Data Ingestion: Accepts input datasets in CSV or Excel format and validates the data structure.
- Data Preprocessing: Handles missing values (mean/mode imputation), standardizes features using Standard Scaler, and encodes categorical data.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4293- 4298
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0627

- Soil-Type Encoding: Converts qualitative soil types into machine-readable numerical form.
- Model Training: Trains multiple algorithms—Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Gaussian Naive Bayes—for performance comparison.
- Prediction Module: Generates top crop recommendations based on the input features.
- Feedback Loop: Enables users to provide feedback for continuous improvement of model accuracy.

This approach ensures that the system can adapt to different regions, datasets, and input formats, thereby delivering reliable and region-specific crop recommendations [4-7].

4. System Architecture and Design

The system architecture consists of several interconnected components that work sequentially to process input data and generate accurate crop recommendations. The architecture follows a modular pipeline structure, as shown below in Figure 1.

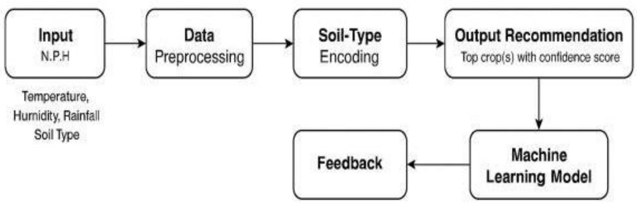

Figure 1 Flow Chart

5. Methodology

5.1 Overview

Building a versatile, dataset-independent crop recommendation system that can handle various input formats and soil-type classifications is the main goal of the methodology. Data collection, preprocessing, encoding, model training, prediction, and evaluation are the six primary phases of implementation.

5.2 Data Collection

The system accepts datasets in CSV or Excel format containing soil nutrients (Nitrogen, Phosphorus, Potassium, pH), environmental features (temperature, humidity, rainfall), and categorical soil type (e.g., sandy, loamy, clayey). The model is designed to function even when datasets vary in structure or naming conventions, ensuring adaptability to regional data sources.

5.3 Data Preprocessing

Data preprocessing ensures data quality and consistency before model training.

- **Handling Missing Values:** Mean and mode imputation techniques are applied to fill missing numeric and categorical values.
- **Feature Standardization:** Continuous features are standardized using a Standard Scaler to ensure they are on a consistent scale, allowing the model to interpret each feature more effectively.
- **Data Cleaning:** Irrelevant or duplicate entries are removed, and outliers are managed using IQR filtering.

5.4 Soil-Type Encoding

Categorical soil types are encoded using one-hot encoding to convert textual data into numerical format suitable for ML algorithms. This feature allows the model to understand qualitative soil characteristics, increasing generalization across different soil conditions.

5.5 Model Training

In this study, four supervised machine learning algorithms were implemented and compared to evaluate their performance.

- **Random Forest (RF):** Ensemble-based model effective in handling nonlinear and complex data patterns.
- **Decision Tree (DT):** Simple, interpretable classifier for baseline comparison.
- **K-Nearest Neighbors (KNN):** Distancebased algorithm suitable for smaller datasets.
- **Gaussian Naive Bayes (GNB):** Probabilistic classifier known for high accuracy in structured data.

Each model was trained on 80% of the dataset and tested on the remaining 20%. Hyperparameter tuning was carried out to improve accuracy and reduce the risk of overfitting.

5.6 Prediction and Recommendation

For a given input (soil and environmental parameters), the trained model predicts the most suitable crops. The system generates the top three recommendations along with their confidence scores.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4293- 4298
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0627

Predictions are displayed through a Flask-based web interface, allowing users to input new data easily.

5.7 Model Evaluation

Each algorithm was evaluated using performance metrics such as Accuracy, Precision, Recall, and F1Score.

Results showed that:

- Naive Bayes achieved 96.36% accuracy, and
- Random Forest achieved 98.18% accuracy, demonstrating strong reliability and generalization capability.

5.8 Feedback and Retraining

A feedback module enables users (farmers or experts) to confirm or reject the suggested crop. This feedback is stored and used for incremental retraining, improving future predictions and adapting to changing environmental conditions, shown in Figure 2.

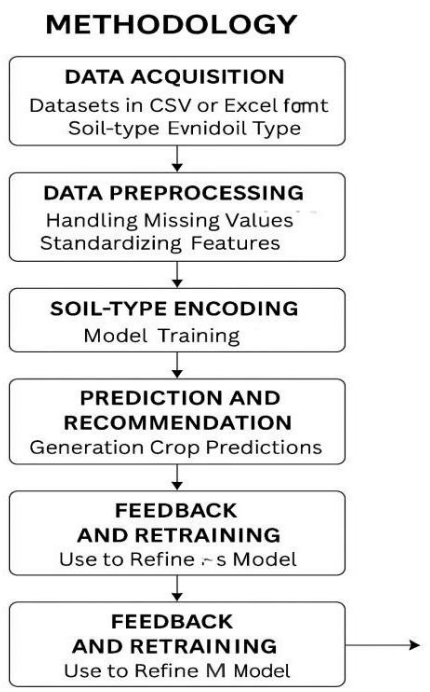

Figure 2 Work Flow of Crop Recommendation System

6. Results and Discussion

6.1 Exploratory Data Analysis (EDA)

A correlation matrix was generated to visualize the linear relationships between numerical features. This is a critical step for understanding multicollinearity and feature interactions.

6.2 Comparative Model Performance

The models were evaluated on the unseen test set using accuracy as the primary performance metric. The results were consistently high across all models, demonstrating the strong predictive signal within the dataset, shown in Table 1.

Table 1 Comparative Accuracy of Trained Models

Model	Accuracy score
Gaussian Naive Bayes	96.36%
Random Forest	98.18%
Decision Tree	95.45%
K-Nearest Neighbors	89.77%

The outstanding performance of Random Forest is noteworthy. Its assumption of feature independence, while often a simplification, appears to hold sufficiently well for this dataset, allowing it to model the data effectively with high computational efficiency. The Gaussian Naive Bayes, as an ensemble method, also delivered exceptional and robust results

6.3 Model-Specific Performance Analysis

To gain deeper insights beyond a single accuracy score, a confusion matrix was generated for each model. The confusion matrix for the Random Forest classifier is shown below in Figure 3.

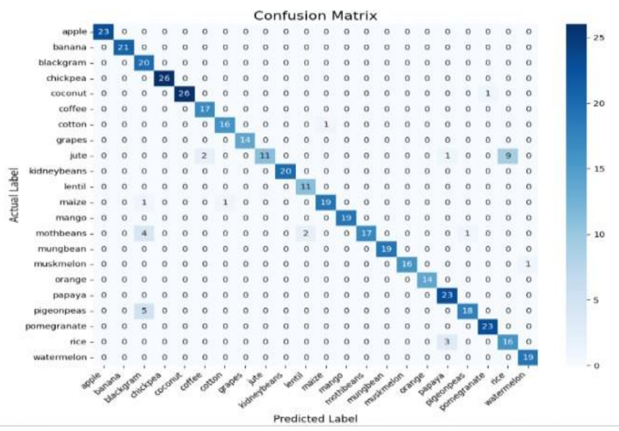

Figure 3 Confusion Matrix for Naive Bayes

The strong diagonal concentration indicates extremely high precision and recall for nearly every

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4293- 4298
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0627

crop. The few off-diagonal entries represent minor misclassifications, for example, confusing 'mothbeans' and 'moongbean', which may have similar growing conditions, shown in Figure 4.

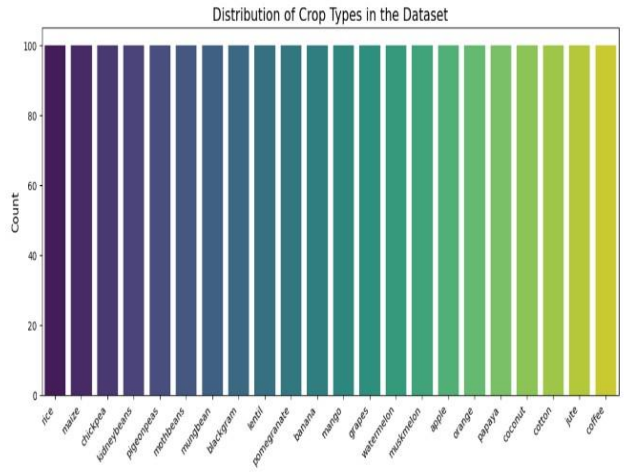

Figure 4 Actual vs Predicted Class Distributions for the Random Forest Model

Conclusion, Limitations, and Future Work

This study demonstrates the effectiveness of machine learning models in recommending suitable crops based on soil and climatic conditions. Among the evaluated models, Random Forest and Gaussian Naïve Bayes showed the highest accuracies of 98.18% and 96.36%, respectively, highlighting their robustness and reliability in supporting agricultural decision-making. These findings emphasize the potential of data-driven approaches to optimize crop selection, minimize risks, and enhance the accuracy of yield predictions. The system is a useful tool for actual agricultural applications because of its adaptability to various soil types and environmental circumstances. Farmers can make better decisions that support resource optimization, sustainable agriculture, and increased food security by incorporating such intelligent systems into their farming operations. Future work may involve expanding the dataset with real-time sensor data, integrating weather forecasting models, and deploying the system as a mobile or web-based application for greater accessibility. Incorporating deep learning and remote sensing data could further refine crop predictions and extend the model's applicability to larger and more diverse geographical regions.

Limitations

- The current model does not include economic factors like market demand or crop prices.
- Soil type, a critical feature, was not extensively used in this version.
- The model's performance is dependent on the quality and range of the training data.

Future Work

- Feature Enrichment: Incorporate additional data layers such as soil type, satellite imagery, and weather forecast data.
- Economic Integration: Develop a hybrid recommendation system that considers both environmental suitability and market profitability.
- Deployment: Package the final model into a user-friendly web or mobile application to make it accessible to farmers in the field.

Conclusion

This study presents an effective machine learning–based crop recommendation system that considers key agricultural factors such as soil pH, temperature, humidity, rainfall, and nutrient levels (N, P, and K) to suggest the most suitable crops for cultivation. Among the models tested, the Random Forest classifier demonstrated the highest accuracy and reliability. The results highlight how data-driven approaches can significantly enhance agricultural decision-making by providing reliable, region-specific crop recommendations. The proposed system offers a scalable solution to promote precision farming and can be further improved by integrating factors such as economic conditions, geospatial data, and real-time environmental inputs to deliver more adaptive and contextaware recommendations.

Acknowledgement

We would like to express our heartfelt gratitude to Dr. Chandrika Ramesh, Head of the Department of Computer Science and Engineering, Malnad College of Engineering, for her constant support, encouragement, and invaluable guidance throughout our project. Our sincere thanks also go to our guide, Ms. Ayeesha Siddiqha, for her continuous motivation, insightful suggestions, and dedicated supervision, which played a crucial role in the successful completion of our research paper titled "Crop Recommendation Using Machine Learning." We are deeply grateful to Dr. Amarendra, Principal, Malnad College of Engineering, for providing us with the necessary facilities and a motivating environment to carry out our work. Finally, we extend our appreciation to all the faculty members and friends who supported and encouraged us throughout this journey.

References

[1]. F. S. P. Prity, M. Ahmed, and T. Rahman, "Crop recommendation using machine learning techniques," Human-Centric Intelligent Systems, vol. 4, no. 1, pp. 45–52, 2024.

[2]. A. Ravikumar, S. Gupta, and R. Meena, "ML-based crop prediction system for smart agriculture," Journal of Science, Technology and Advanced Research (JSTAR), vol. 9, no. 2, pp. 134–140, 2025.

[3]. M. V. Korde, P. A. Patil, and R. P. Wagh, "Data-driven crop recommendation using machine learning algorithms," International Research Journal of Modernization in Engineering, Technology and Science. (IRJMETS), vol. 6, no. 3, pp. 88–93, 2024.

[4]. S. Shinde, R. Pawar, and M. Gaikwad, "Crop recommendation system using machine learning," in Proc. National Conference on Emerging Research in Nanotechnology and Biosciences (NCERNB), 2023, pp. 22–26.

[5]. [S. K. Apat, D. S. Rao, and K. T. Nayak, "An intelligent system for crop selection using ML models," International Research Journal of Advanced Science, Engineering and Technology (IRJASET), vol. 2, no. 4, pp. 101–106, 2023.

[6]. A. Baishya, S. Nath, and R. Das, "Tiny ML-based crop recommendation system for precision agriculture," Journal of Embedded Systems & AI in Agriculture, vol. 1, no. 2, pp. 50–56, 2025.

[7]. K. Verma and R. Sharma, "A machine learning approach to crop recommendations," Springer Journal of AI in Agriculture, vol. 3, no. 1, pp. 25–34, 2024