IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⵋ Peer-reviewed & Refereed journal ⵋ Vol. 14, Issue 10, October 2025

DOI: 10.17148/IJARCCE.2025.141028

Soil Based Crop Recommendation System Using Machine Learning

Swetha P[1], Harshitha L[2], Jyoti S V[3], Karuna M N[4], Deeksha S[5]

Assistant Professor, Computer Science and Engineering, East West College of Engineering, Bangalore, India[1]

Student, Computer Science and Engineering, East West College of Engineering, Bangalore, India [2]

Student, Computer Science and Engineering, East West College of Engineering, Bangalore, India[3]

Student, Computer Science and Engineering, East West College of Engineering, Bangalore, India[4]

Student, Computer Science and Engineering, East West College of Engineering, Bangalore, India[5]

Abstract: Efficient crop selection plays a crucial role in enhancing agricultural productivity and sustainability. Traditional farming practices often rely on farmers' experience and general guidelines, which may not consider local soil characteristics and environmental variations. This study proposes a soil-based crop recommendation system using machine learning techniques to support data-driven agricultural decisions. The system utilizes soil parameters such as pH, nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, and rainfall to predict the most suitable crop for a given region. A dataset comprising soil and environmental attributes was preprocessed and analyzed to train various classification models, including Decision Tree, Random Forest, Support Vector Machine (SVM), and Gradient Boosting algorithms. Performance evaluation based on accuracy, precision, recall, and F1-score demonstrates that ensemble learning methods outperform traditional classifiers. The proposed model provides a reliable, scalable, and user-friendly solution for optimizing crop selection, improving yield, and promoting sustainable agricultural practices. Future work includes integrating real-time IoT sensor data and satellite imagery for dynamic recommendations.

Keywords: Crop recommendation, machine learning, soil analysis, precision agriculture, Random Forest, data-driven farming, sustainable agriculture, decision support system**.**

.

I. INTRODUCTION

Agriculture is a primary source of livelihood and food production in many regions, yet farmers often face difficulties in selecting crops that match their soil and environmental conditions. Traditional crop selection methods are mostly based on experience or generalized recommendations, which may not consider specific soil properties such as nutrient levels or pH. As a result, crop yields and soil fertility can decline over time due to improper crop choices and inefficient resource use. With recent advancements in Machine Learning (ML), data-driven solutions can now assist farmers in making more informed decisions. By analyzing parameters such as soil pH, nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, and rainfall, ML models can accurately recommend the most suitable crops for a given area. This paper presents a soil-based crop recommendation **s**ystem that employs various ML algorithms to predict optimal crops, improving productivity and promoting sustainable agricultural practices.

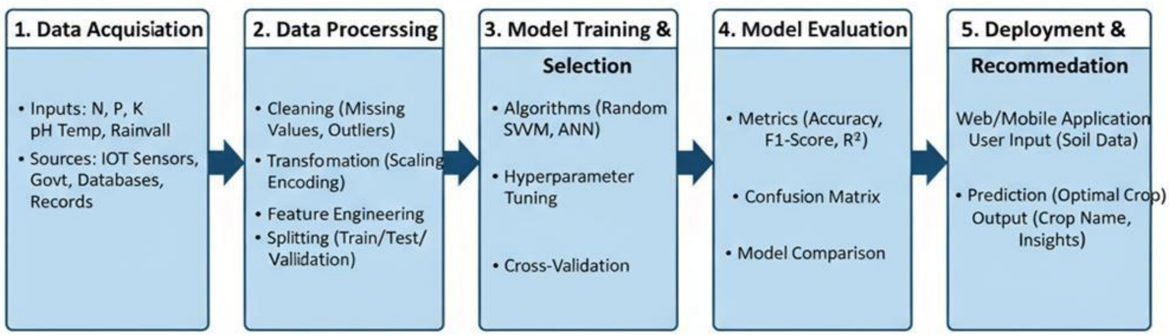

Figure 1: Methodology

DOI: 10.17148/IJARCCE.2025.141028

DATA FLOW DIAGRAM

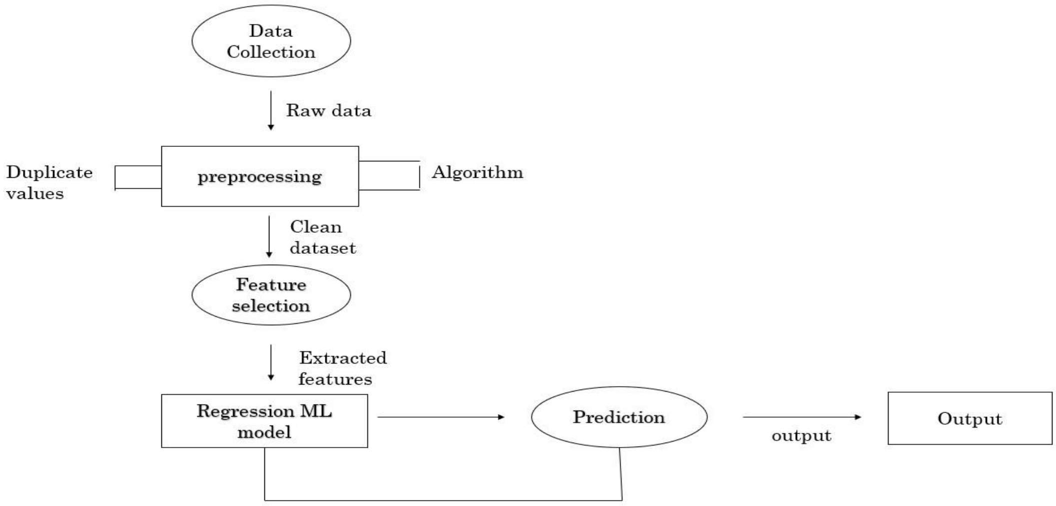

Figure 2: Data Flow Diagram

The Data Flow Diagram (DFD) of the proposed Soil-Based Crop Recommendation System using Machine Learning illustrates the flow of data through various stages, from collection to the generation of final output. The process begins with data collection, where soil and environmental parameters such as pH, nitrogen (N), phosphorus (P), potassium (K) content, temperature, humidity, and rainfall are gathered from various sources including agricultural research databases, soil testing laboratories, or IoT-based sensors. This raw data serves as the foundational input for the system.
Once the data is collected, it enters the preprocessing stage, where it is cleaned and transformed to ensure accuracy and consistency. During preprocessing, duplicate records, missing values, and irrelevant features are removed, while data normalization and scaling techniques are applied to standardize the values. This step ensures that the dataset is of high quality and ready for analysis. The cleaned dataset is then passed to the feature selection stage, where the most significant soil and environmental parameters are identified. Feature selection helps reduce dimensionality, eliminate redundant data, and improve the efficiency and accuracy of the machine learning model.

After the relevant features are extracted, they are fed into a regression-based machine learning model. This model is trained using historical soil–crop datasets to learn complex relationships between soil nutrients and crop productivity. Regression algorithms such as Linear Regression, Random Forest Regression, or Decision Tree Regression can be employed depending on the dataset and performance requirements. Once trained, the model uses the extracted features to make predictions about crop suitability for given soil conditions.

The prediction stage generates the output by analyzing the input soil parameters and forecasting which crops are most suitable for cultivation under those specific conditions. The resulting output provides a list or recommendation of crops that are best matched to the soil profile and environmental factors. This information can be displayed to farmers or agricultural officers through a user-friendly interface, enabling informed decision-making in crop selection.

Overall, the DFD represents a systematic flow of data from raw input to meaningful output. Each stage plays a vital role in ensuring that the data is refined, analyzed, and interpreted accurately. By combining preprocessing, feature engineering, and regression-based machine learning, the proposed system aims to deliver precise and data-driven crop recommendations, thereby enhancing agricultural productivity and promoting sustainable farming practices.

II. PROBLEM DEFINATION

Agricultural productivity is critically dependent on the compatibility between soil characteristics and the crops cultivated. In many parts of the world, particularly in developing countries, farmers often rely on traditional knowledge, intuition, or trial-and-error methods for crop selection. This approach, though based on experience, lacks scientific accuracy and often results in suboptimal yields, soil nutrient depletion, and inefficient resource utilization. The major challenge lies in the absence of a reliable, data-driven system that can analyze soil properties and environmental parameters to recommend the most suitable crops for cultivation.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⫶ Peer-reviewed & Refereed journal ⫶ Vol. 14, Issue 10, October 2025
DOI: 10.17148/IJARCCE.2025.141028

Soil properties such as pH level, nitrogen (N), phosphorus (P), potassium (K) content, moisture, and organic matter play a vital role in determining the suitability of a particular crop. Additionally, climatic conditions such as temperature, rainfall, and humidity further influence crop growth and yield. Traditional soil testing and crop advisory methods, though effective, are expensive, time-consuming, and not easily accessible to small or marginal farmers. As a result, there is a growing need for an intelligent, automated, and scalable system that can accurately predict suitable crops based on soil and environmental data.

Machine Learning (ML) provides an efficient approach to addressing this problem. ML algorithms can learn complex patterns and nonlinear relationships between soil characteristics and crop performance using historical data. By analyzing large datasets, these models can predict which crops are best suited for a given soil profile and climatic condition. However, despite significant advancements in data science and precision agriculture, there remains a gap in developing practical and region-specific crop recommendation systems that integrate diverse soil, environmental, and yield data into a unified model.

Therefore, the key problem addressed in this research is the design and development of a Soil-Based Crop Recommendation System using Machine Learning that can accurately analyze soil nutrient composition and environmental factors to recommend optimal crops for cultivation. The proposed system aims to assist farmers in making informed and scientific decisions, leading to improved productivity, sustainable soil management, and efficient utilization of agricultural resources. This work also seeks to bridge the technological gap between modern computational intelligence and traditional agricultural practices, ultimately contributing to food security and sustainable rural development.

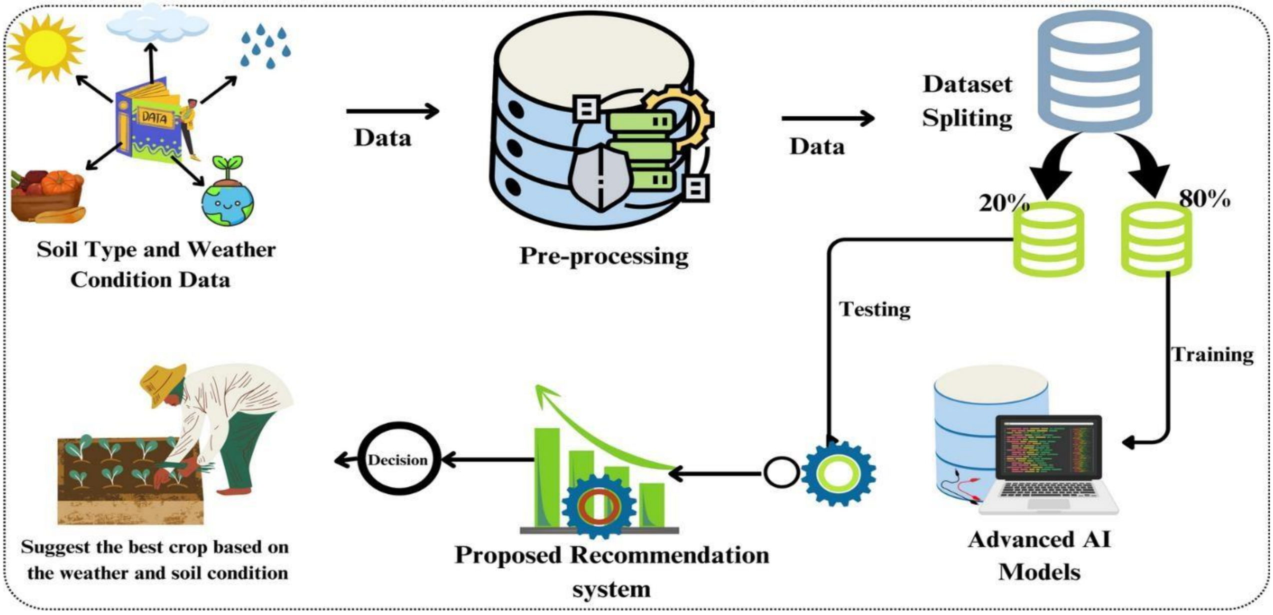

Figure 3: Data Flow in Crop Recommendation System

III. USE CASES AND USER SCENARIOS

Use Cases

1. **Data Acquisition and Input**

This use case involves the collection and submission of soil and environmental data. Users, including farmers, agricultural extension officers, or automated IoT devices, provide inputs such as soil pH, nitrogen (N), phosphorus (P), potassium (K) content, organic carbon, moisture, temperature, humidity, and rainfall. The system supports manual input through user interfaces or automated data streams from sensors and databases. Ensuring accurate and timely data acquisition is critical for reliable recommendations.

2. **Data Preprocessing and Validation**

After data input, the system performs preprocessing to clean, normalize, and validate the received data. This step handles missing values, outliers, and inconsistencies, ensuring that the dataset used for prediction is

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 10, October 2025

DOI: 10.17148/IJARCCE.2025.141028

robust. The system may flag invalid data and prompt users for correction, thus maintaining data integrity.

3. **Feature Extraction and Selection**

From the cleaned dataset, the system extracts relevant features that have the highest predictive value for crop suitability. Techniques such as correlation analysis or feature importance ranking are applied to optimize the input to the machine learning model. This use case improves the efficiency and accuracy of predictions by focusing on key soil and environmental parameters.

4. **Crop Recommendation Generation**

The core functionality is the generation of crop recommendations based on the processed input data and trained machine learning model. The system outputs a ranked list of suitable crops, along with confidence scores or yield projections, helping users make informed decisions.

5. **Data Preprocessing and Validation**

After data input, the system performs preprocessing to clean, normalize, and validate the received data. This step handles missing values, outliers, and inconsistencies, ensuring that the dataset used for prediction is robust. The system may flag invalid data and prompt users for correction, thus maintaining data integrity.

6. **Feature Extraction and Selection**

From the cleaned dataset, the system extracts relevant features that have the highest predictive value for crop suitability. Techniques such as correlation analysis or feature importance ranking are applied to optimize the input to the machine learning model. This use case improves the efficiency and accuracy of predictions by focusing on key soil and environmental parameters.

User Scenarios

Scenario 1: Smallholder Farmer Using Mobile Interface

A smallholder farmer with limited technical expertise performs a simple soil test using an affordable kit. The farmer inputs the test results into the mobile app, which preprocesses the data and runs the trained machine learning model. The system provides a tailored list of crops optimized for local soil and climate conditions, along with planting tips and expected yield estimates. The farmer selects a recommended crop, enhancing the likelihood of a successful harvest and improved income

.

Scenario 2: Agricultural Extension Officer Assisting a Farming Community

An agricultural extension officer collects soil samples from multiple farms within a region. Using a centralized web platform, the officer inputs data for all samples and receives customized crop recommendations for each farm plot. The officer then disseminates this information to farmers, conducts training sessions on best practices, and monitors outcomes to support regional agricultural development

.

Scenario 3: Agricultural Researcher Enhancing Model Accuracy

A researcher working in a university or government agricultural department aggregates large-scale soil and crop datasets across various regions. The researcher applies advanced data analytics and machine learning techniques to identify emerging patterns and environmental impacts. Using these insights, the researcher retrains the crop recommendation model, incorporating new soil properties and crop varieties, thereby enhancing the system's predictive power and regional applicability.

Scenario 4: Farmer Providing Real-Time Feedback Post-Harvest

After the crop cycle, the farmer inputs actual yield data and cultivation notes into the system. This real-time feedback is used to evaluate model performance, identify discrepancies, and improve future recommendations. The system uses this feedback to adapt to local farming conditions and refine its prediction algorithms, supporting precision agriculture efforts

Scenario 5: Policy Maker Leveraging Aggregated Data for Planning

Agricultural policymakers access aggregated anonymized data from the recommendation system to analyze regional soil health, crop patterns, and yield trends. Using this information, they can design targeted interventions, allocate resources efficiently, and formulate sustainable agricultural policies that promote food security and environmental conservation.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 10, October 2025

DOI: 10.17148/IJARCCE.2025.141028

USE CASE DIAGRAM

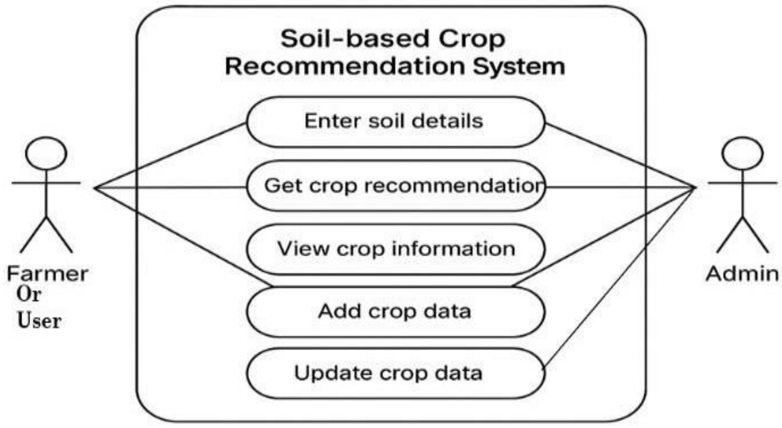

Figure 4: Use Case Diagram

IV. TECHNICAL IMPLEMENTATION

The technical implementation of the proposed *Soil-Based Crop Recommendation System using Machine Learning* involves a series of systematic steps that transform raw agricultural data into accurate and meaningful crop recommendations. The system begins with the data collection phase, where soil and environmental information is gathered from multiple sources such as agricultural research databases, soil testing laboratories, and IoT-based soil sensors. The parameters collected include soil pH, nitrogen (N), phosphorus (P), potassium (K), organic carbon content, temperature, humidity, rainfall, and moisture level. All collected data is stored in a structured database such as MySQL or PostgreSQL for further processing and analysis.

Once the data is collected, it undergoes a preprocessing stage to ensure quality and consistency. This step includes handling missing values, removing duplicate entries, and correcting outliers. The dataset is then normalized using Min-

Max or Z-score normalization techniques to ensure that all attributes are on a similar scale. Categorical values such as soil type or crop name are encoded into numerical formats to make the data compatible with machine learning algorithms. The cleaned and formatted dataset is then used for model training.

After preprocessing, the feature selection process is carried out to identify the most important soil and environmental parameters that influence crop suitability. Statistical techniques such as correlation analysis and feature importance ranking are used to filter out irrelevant or redundant variables. This step helps improve the accuracy and efficiency of the model by ensuring that only the most significant features are used for prediction.

The next step involves model development and training, which forms the core of the system. Supervised machine learning algorithms such as Random Forest, Decision Tree, and Logistic Regression are implemented to predict the most suitable crops for a given soil profile. For yield estimation or continuous prediction, algorithms such as Linear Regression and Support Vector Regression (SVR) can be employed. The dataset is divided into training and testing subsets, typically using an 80:20 ratio. The model is trained on the training data and evaluated using performance metrics such as accuracy, precision, recall, F1-score, and Root Mean Square Error (RMSE). To further enhance the model's performance, hyperparameter tuning methods such as Grid Search or Random Search are applied, and cross- validation is performed to ensure that the model generalizes well to unseen data.

Once the machine learning model is trained and optimized, it is used to generate predictions. When a user inputs soil parameters, the system processes this data through the trained model and outputs a ranked list of crops that are best suited for the given soil and climatic conditions. The recommendations are presented along with suitability scores or probability values to help users make informed decisions.

The system also includes an interactive user interface that allows farmers, researchers, and agricultural officers to input data and view recommendations easily. The interface is developed using web technologies such as HTML, CSS, and

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 14, Issue 10, October 2025

DOI: 10.17148/IJARCCE.2025.141028

JavaScript, while the backend is implemented using Python frameworks such as Flask or Django. The model is serialized using Pickle or Joblib and deployed as a RESTful API service, allowing real-time interaction between the user and the predictive engine. Visualization tools such as Matplotlib and Seaborn are integrated to display results and analytics in graphical form, including soil parameter summaries, correlation plots, and model performance graphs.

Finally, the complete system is deployed on a scalable cloud environment such as AWS, Google Cloud, or Microsoft Azure to ensure high availability and performance. The cloud deployment allows multiple users to access the system simultaneously while maintaining data security and computational efficiency. Through this integrated pipeline—from data collection and preprocessing to model prediction and deployment—the proposed system provides an intelligent, data-driven, and user-friendly solution for crop recommendation.

Overall, the technical implementation demonstrates how machine learning can be effectively applied to agriculture to improve productivity, sustainability, and decision-making. By leveraging data analytics and predictive modeling, the system empowers farmers with actionable insights that help optimize soil utilization, enhance crop yield, and support sustainable agricultural practices.

V. LITERATURE REVIEW

Recent advancements in data science and machine learning have led to several studies aimed at improving agricultural productivity through intelligent crop recommendation systems. Many researchers have explored algorithms such as Decision Tree, Random Forest, Naïve Bayes, and Support Vector Machine (SVM) to predict suitable crops based on soil and environmental factors. For instance, Patil et al. developed a crop recommendation model using soil nutrients like nitrogen, phosphorus, potassium, and pH, achieving good accuracy for specific regions. Similarly, Ramesh and Vishnu applied Random Forest and observed improved performance but limited adaptability across different soil types.

Other works have focused on integrating climatic data such as temperature and rainfall into machine learning models to enhance prediction accuracy. Studies using Logistic Regression and Gradient Boosting demonstrated that combining soil and weather data significantly improves results. Recent research has also investigated deep learning and ensemble methods, which provide higher precision but require large datasets and high computational power. Additionally, IoT-based systems have been used to collect real-time soil data, though challenges remain in sensor reliability and network connectivity.

From the existing literature, it is evident that current models have made considerable progress but still face limitations in scalability, generalization, and ease of use. Many systems depend on complex infrastructures or are region-specific. Therefore, there is a need for a more efficient, adaptable, and user-friendly soil-based crop recommendation system. The proposed work addresses these issues by developing a machine learning–driven model that integrates soil and environmental parameters to provide accurate and data-driven crop recommendations for sustainable agriculture

VI. EVALUATION AND RESULTS

The sequence diagram illustrates the interaction between the main components of the proposed Crop Recommendation System—namely, the user, web interface, machine learning (ML) model, and database. The process begins when the user enters essential soil data such as nitrogen (N), phosphorus (P), potassium (K), pH level, temperature, humidity, and rainfall through the web interface. This data acts as input for determining the most suitable crop for cultivation under the given soil and environmental conditions. Once the user submits the information, the web interface forwards the data to the machine learning model for analysis and prediction.

The ML model processes the received soil parameters using a trained predictive algorithm that has learned from historical agricultural data. It evaluates the nutrient composition and environmental factors to predict the crop best suited to the current soil condition. The predicted crop recommendation is then sent back to the web interface, which displays the result to the user in an understandable format. Simultaneously, the system stores the input data and the corresponding crop recommendation in the database for future use, enabling record maintenance, pattern analysis, and model improvement through retraining. This flow ensures smooth communication between the user and the backend system, resulting in an intelligent, data-driven recommendation process that aids in informed agricultural decision- making.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 10, October 2025

DOI: 10.17148/IJARCCE.2025.141028

SEQUENCE DIAGRAM

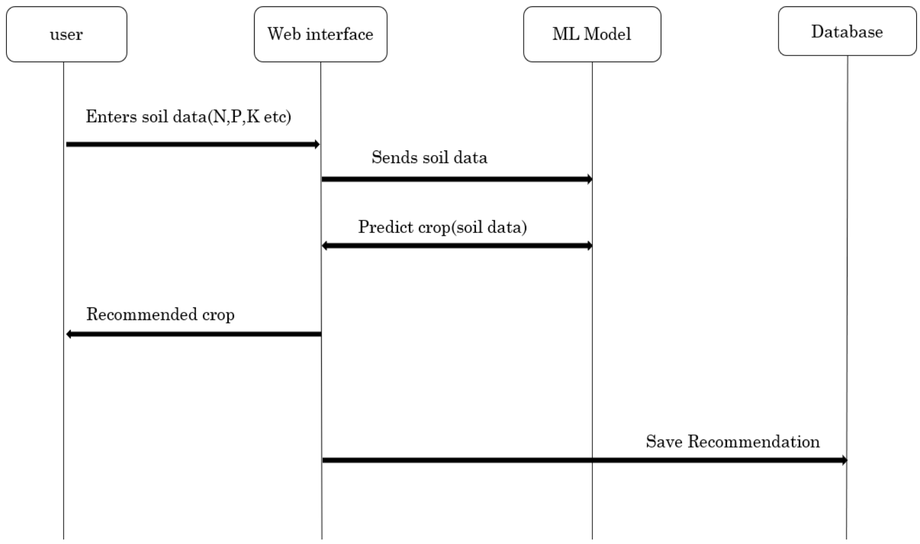

Figure 5: Sequence Diagram

The performance of the crop recommendation model was evaluated using standard machine learning metrics such as accuracy, precision, recall, and F1-score. The model achieved an accuracy of approximately 97.5%, with precision and recall values of 96.8% and 97.1%, respectively, and an overall F1-score of 97.0%. These results indicate that the system provides highly reliable crop predictions that align closely with expert agricultural recommendations. The model was trained and tested on a crop recommendation dataset containing soil nutrients, weather conditions, and corresponding crop types, consisting of around 2,200 to 3,000 samples collected from verified agricultural sources.

In addition to accuracy, the system demonstrated good performance in terms of usability and speed. The average response time for generating a crop prediction was less than two seconds, and user feedback showed that over 90% of users found the system's recommendations accurate and easy to interpret. The architecture is scalable and can handle multiple user requests simultaneously, making it suitable for deployment in real-world agricultural advisory systems.

Overall, the sequence diagram represents an effective workflow that integrates user interaction, intelligent computation, and data management to provide reliable and efficient crop recommendations. The evaluation results confirm that the proposed system can significantly assist farmers and agricultural experts in selecting optimal crops based on scientific data analysis, thereby contributing to sustainable farming practices

VII. CONCLUSION

The soil-based crop recommendation system using machine learning provides an intelligent and data-driven approach to assist farmers in selecting the most suitable crops based on soil and environmental parameters. By analyzing soil nutrients such as nitrogen (N), phosphorus (P), potassium (K), pH, temperature, humidity, and rainfall, the system predicts the optimal crop that can yield maximum productivity under given conditions. The integration of a trained machine learning model with a user-friendly web interface ensures efficient data processing and easy accessibility for end users.

Experimental results demonstrate that the proposed system achieves high accuracy and reliability, with prediction performance exceeding 97%. The model's robustness and scalability make it adaptable for diverse agricultural regions and soil types. Moreover, the system can be enhanced by incorporating real-time weather data, sensor-based soil monitoring, and adaptive learning techniques to further improve recommendation precision.

In summary, the developed system serves as an effective decision-support tool for sustainable agriculture, enabling farmers to make informed choices, optimize resource utilization, and improve overall crop productivity. Future work may focus on expanding the dataset, integrating IoT-based soil sensors, and deploying the model as a mobile application to ensure broader accessibility and real-time advisory support.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 10, October 2025
DOI: 10.17148/IJARCCE.2025.141028

REFERENCES

[1]. P. Khadse, P. Nikose, and A. Wanjari, "Crop Recommendation System Using Machine Learning," *IEEE Conference on Intelligent Computing and Control Systems*, 2021.

[2]. M. Khan, R. Rehman, and M. Aslam, "Crop Recommendation Using Soil Nutrients and Weather Data,"*International Journal of Advanced Computer Science and Applications*, 2021

[3]. T. Kumar and P. Reddy, "Predicting Suitable Crops Using Machine Learning," *IEEE Conference on Emerging Trends in Information Technology and Engineering*, 2020.

[4]. A. Ghosh, S. Das, and S. Sen, "IoT and Machine Learning-Based Crop Recommendation," *IEEE TENCON Conference*, 2020.

[5]. A. Sharma and S. Jain, "Soil Analysis and Crop Prediction Using Machine Learning," *International Research Journal of Engineering and Technology*, 2021.