IJIRIS: International Journal of Innovative Research in Information Security
Volume 11, Issue 09, November 2025
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2025.v1109.07

Crop Recommendation System Using Machine Learning

Sadhana R
Assistant Professor, Department of CSE,
Sri Sairam College of Engineering, Bengaluru, India
sadhanasreekanth@gmail.com
https://orcid.org/0009-0000-6961-501X
Pavan G K, Sainath, Vignesh V
Department of CSE,
Sri Sairam College of Engineering, Bengaluru, India
sce23cs083@sairamtap.edu.in
sce23cs054@sairamtap.edu.in, sce23cs034@sairamtap.edu.in

Publication History
Manuscript Reference No: IJIRIS/RS/Vol.11/Issue09/NVIS10086
Research Article Open Access| Double-Blind Peer-Reviewed| Article ID: IJIRIS/RS/Vol.11/Issue09/NVIS10086 Received: 28, October 2025, Revised: 05, November 2025, Accepted: 12, November 2025, Published Online: 21, November 2025.
https://www.ijiris.com/volumes/Vol11/iss-09/07.NVIS10086.pdf
Citation: Sadhana, Pavan, Sainath, Vignesh (2025), Crop Recommendation System Using Machine Learning, IJIRIS: International Journal of Innovative Research in Information Security, Volume 11, Issue 09 of 2025 pages 493-498
Doi:> https://doi.org/10.26562/ijiris.2025.v1109.07
BibTeX Key: Sadhana@2025Crop
IJIRIS papers should be cited as IJIRIS (International Journal of Innovative Research in Information Security, AM Publications, India 2025, ISSN 2349-7017, https://doi.org/10.26562/ijiris.2025.v1109.07 The journal's official abbreviation is IJIRIS. Orcid: https://orcid.org/0009-0004-9398-7488

Abstract: Agriculture continues to be the backbone of India's economy, supporting livelihoods and contributing significantly to food security. However, farmers frequently face challenges in deciding which crop to cultivate due to unpredictable weather conditions, variations in soil fertility, and limited access to scientific tools. Traditional methods of crop selection are primarily experience-based and often lead to suboptimal productivity. To address these challenges, this paper presents a Machine Learning (ML)-based Crop Recommendation System designed to recommend the most appropriate crop for cultivation based on soil nutrients and environmental parameters. The system utilizes attributes such as Nitrogen (N), Phosphorus (P), Potassium (K), pH, temperature, humidity, and rainfall. These features are analyzed using multiple ML algorithms, including Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest. After extensive testing, the Random Forest Classifier demonstrated the highest performance, achieving an accuracy of 95%, due to its ensemble-based approach and ability to handle nonlinear feature interactions. The proposed system enables farmers to make informed, data-driven decisions rather than relying solely on experience. The study highlights how integrating ML with precision agriculture can lead to sustainable farming practices, efficient resource utilization, and improved crop yield.
Keywords: Crop Recommendation, Machine Learning, Random Forest, Precision Agriculture, Sustainable Farming

INTRODUCTION

Agriculture has always played a vital role in the economic and social structure of India, contributing a significant portion to the nation's Gross Domestic Product (GDP) and employing a large segment of the population. Despite technological progress, farmers across various regions still rely on conventional practices and personal experience for crop selection. This traditional decision-making approach often leads to lower productivity because it does not consider the scientific relationship between soil fertility, climatic conditions, and crop suitability. Over the past decade, the agricultural sector has experienced rapid digital transformation through the use of emerging technologies such as Artificial Intelligence (AI), Machine Learning (ML), and the Internet of Things (IoT). These technologies have enabled data-driven approaches to improve crop yield, soil health analysis, irrigation management, and pest prediction. Among these technologies, Machine Learning has shown immense potential for developing intelligent models capable of predicting suitable crops under given soil and environmental conditions. India's agricultural productivity faces multiple challenges unpredictable rainfall, climate variability, soil nutrient depletion, and lack of awareness about precision farming. In such a context, a Crop Recommendation System powered by ML can provide significant benefits. By analyzing factors like Nitrogen (N), Phosphorus (P), Potassium (K), pH, temperature, humidity, and rainfall, ML algorithms can accurately predict which crop will yield the best results in a specific region. The motivation for this research stems from the need to shift from experience-based decisions to scientific, data-oriented recommendations.

With the help of trained ML models, farmers can improve productivity, reduce input wastage, and contribute to sustainable agricultural practices. Furthermore, integrating IoT sensors and real-time environmental data allows the system to become adaptive and location-specific. This work aims to develop a Machine Learning-based Crop Recommendation System that analyzes soil and climatic data to recommend the most appropriate crop for cultivation. The proposed model compares the performance of various ML algorithms and identifies the one that provides the best accuracy and reliability. The system's long-term goal is to support India's vision of Digital Agriculture and Smart Farming, enabling farmers to adopt technology-driven solutions for improved productivity and resource management.

Objectives of the Study:

1. To design and implement a Machine Learning model capable of predicting suitable crops based on soil and climate features.
2. To compare multiple ML algorithms and identify the most accurate and reliable model.
3. To build a user-friendly application that provides easy access to predictions and insights.
4. To explore real-time integration using IoT sensors for live environmental data updates.
5. To promote sustainable agriculture by optimizing the use of water, fertilizer, and land resources.

LITERATURE SURVEY

The agricultural sector has witnessed a technological revolution through the integration of Artificial Intelligence (AI), Machine Learning (ML), and the Internet of Things (IoT). These emerging technologies are transforming traditional farming into smart and precision agriculture, where decisions are based on real-time data instead of intuition. Researchers worldwide have explored ML approaches for various agricultural applications, including soil classification, yield prediction, and crop recommendation.

A. Overview of Previous Research

Several studies have focused on the use of ML techniques for developing intelligent systems that can assist farmers in choosing suitable crops based on environmental and soil data. In early research, algorithms such as Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) were extensively applied for classification tasks related to soil fertility and crop suitability. These models utilized datasets containing parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), pH, temperature, humidity, and rainfall to predict optimal crop types. Patel et al. (2021) implemented a decision-support model that predicted crop suitability based on soil nutrients and climate features. Their research showed that ensemble learning models like Random Forest provided better accuracy compared to single classifiers due to their ability to handle complex data distributions. Similarly, Singh and Kaur (2022) conducted a comparative study on soil classification and concluded that ensemble models outperform traditional classifiers in both accuracy and stability. Further advancements were made when IoT and ML were combined to collect real-time agricultural data. Chatterjee and Dutta (2021) integrated IoT sensors with ML algorithms to dynamically analyze soil moisture and temperature levels. This integration improved the responsiveness of recommendation systems by enabling continuous data monitoring. Sharma et al. (2022) proposed a hybrid model that combined Random Forest and Gradient Boosting algorithms for crop yield estimation, achieving higher prediction precision. Banerjee et al. (2021) emphasized the importance of regional soil classification, showing that location-specific datasets significantly improve the adaptability of ML models for agricultural prediction.

B. Limitations in Existing Studies

Although significant progress has been made, several limitations still persist in existing research:

1. Many models rely on static datasets and lack integration with real-time environmental updates.
2. Some systems are region-specific, making it difficult to generalize predictions across different geographic areas.
3. Models often ignore secondary environmental parameters such as wind speed, evapo transpiration, or soil moisture variability.
4. The user interfaces of existing systems are often too complex for farmers with limited technical expertise.
5. These challenges highlight the necessity for a more adaptive, accurate, and user-friendly system that can process multi-dimensional agricultural data and provide real-time crop recommendations.

C. Research Gap and Motivation

The key research gap identified in the literature is the lack of a unified model that combines ensemble ML algorithms with real-time IoT-based data collection for crop recommendation. While several studies have shown promising results with specific datasets, very few have addressed scalability, data adaptability, and usability in rural contexts. Therefore, the motivation behind this work is to design a robust and intelligent ML-based system that not only predicts the most suitable crop but can also be adapted to different regions with minimal retraining. By focusing on Random Forest and comparative algorithmic evaluation, this study aims to develop a practical and efficient solution for Indian farmers to improve productivity and ensure sustainability.

EXISTING SYSTEM

In traditional agricultural practices, crop selection is primarily based on farmers' personal experience, intuition, and advice from local experts. While these methods have sustained farming for generations, they often fail to consider the complex interactions between soil fertility, climatic variations, and nutrient requirements. As a result, crop yield and resource utilization frequently remain suboptimal.

Several digital tools and web-based applications have been developed to assist in crop prediction; however, most rely on static data and simplistic rule-based models.These systems do not adapt well to changing weather patterns or dynamic soil conditions. Moreover, many of them lack an interface that is intuitive for farmers and do not incorporate real-time data from IoT sensors or field observations.

Some of the major limitations observed in existing systems include:

1. Lack of Real-Time Data – Most tools depend on outdated datasets and fail to update recommendations dynamically.
2. Low Accuracy – Traditional models do not account for nonlinear relationships between variables like soil nutrients and climate.
3. Limited Scalability – The systems are usually restricted to small geographic regions and cannot be generalized for other soil types.
4. Minimal User Accessibility – Many applications are not farmer-friendly and require technical knowledge to operate.

Because of these shortcomings, farmers continue to depend on conventional experience-based practices, which may lead to resource wastage, reduced productivity, and environmental degradation. Hence, there is a pressing need for a more adaptive, intelligent, and data-driven solution that can enhance accuracy and ease of use.

PROPOSED WORK

To overcome the limitations of existing systems, the proposed research introduces a Machine Learning–based Crop Recommendation System that analyzes soil and climatic parameters to determine the most suitable crop for cultivation. The approach utilizes data-driven techniques to minimize dependency on human intuition and supports precision agriculture.The system processes inputs such as Nitrogen (N), Phosphorus (P), Potassium (K), pH, temperature, humidity, and rainfall to generate predictions. These parameters are fed into multiple ML models, including Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest (RF). After evaluation, the Random Forest Classifier is identified as the most accurate and reliable due to its ensemble learning capability, which reduces over fitting and improves predictive stability.

Key features of the proposed system include:

1. Integration of Environmental Parameters: Combines soil nutrients, weather, and rainfall data to provide comprehensive crop recommendations.
2. Ensemble Learning Approach: Utilizes Random Forest to handle complex, nonlinear interactions between input variables.
3. IoT and Real-Time Updates: Can be extended to integrate IoT sensors for automatic data collection and continuous monitoring.
4. Scalability and Adaptability: The system can easily adapt to new datasets or regional conditions without extensive retraining.
5. User-Friendly Interface: Designed to be intuitive so that even non-technical users, such as farmers, can easily access recommendations.

This proposed system bridges the gap between traditional farming practices and modern data-driven agriculture, enabling farmers to make informed decisions based on scientific analysis rather than trial and error.The adoption of this system is expected to increase yield, optimize fertilizer usage, and promote sustainable agricultural development.

IMPLEMENTATION

The implementation of the Crop Recommendation System using Machine Learning involves several key stages—from data acquisition to system deployment. Each phase contributes to building an accurate and reliable model that can assist farmers in scientific crop selection. The implementation process consists of data collection, preprocessing, feature selection, model training, evaluation, and deployment.

A. Data Collection

Data is the foundation of any Machine Learning model. In this project, data related to soil nutrients and climatic conditions was collected from publicly available and reliable sources such as the Kaggle Crop Recommendation Dataset (2022) and Indian government agricultural portals.

The dataset includes the following key parameters:

Nitrogen (N) – promotes leaf and stem growth.
Phosphorus (P) – supports root and seed development.
Potassium (K) – improves crop resistance to diseases.
pH Value – indicates soil acidity or alkalinity.
Temperature (°C) – influences plant metabolism and germination.
Humidity (%) – affects transpiration and plant health.
Rainfall (mm) – determines irrigation requirements.

In practical applications, these data points can also be captured through IoT-based sensors such as NPK sensors, soil moisture sensors, and temperature–humidity sensors connected to microcontrollers (like Arduino or Raspberry Pi). The data collected from sensors can be periodically updated and stored in a local or cloud-based database for continuous system learning.

E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2025.v1109.07

B. Data Preprocessing

Raw agricultural data often contains noise, missing values, and outliers that can reduce the performance of ML models. Hence, preprocessing is performed to ensure data quality and consistency.

Key preprocessing steps include:

1. Handling Missing Values: Missing entries are replaced with the mean or median values of corresponding attributes.
2. Data Normalization: All numerical data is scaled to a uniform range (e.g., 0–1) to ensure that one feature does not dominate others.
3. Encoding Categorical Variables: Crop names and non-numeric values are encoded into numerical form for compatibility with ML algorithms.
4. Duplicate and Outlier Removal: Ensures that the dataset remains clean and consistent.
5. After preprocessing, the dataset becomes well-structured and ready for training.

C. Feature Selection and Model Training

Feature selection identifies the most relevant attributes that influence crop yield. In this study, N, P, K, pH, temperature, humidity, and rainfall are used as input features, while the output variable is the recommended crop name.

Four algorithms were implemented and tested:

1. Decision Tree (DT): Splits data into branches based on attribute conditions; easy to interpret but prone to overfitting.
2. K-Nearest Neighbor (KNN): Classifies samples based on similarity; performs well on smaller datasets.
3. Support Vector Machine (SVM): Finds optimal boundaries between classes; suitable for high-dimensional data.
4. Random Forest (RF): Combines multiple decision trees to improve accuracy and robustness.

The dataset was divided into 80% training and 20% testing subsets. Each algorithm was trained using the training data and evaluated using the test data. Among all, the Random Forest Classifier achieved the best results with an accuracy of 95%, outperforming other models due to its ensemble nature and ability to handle complex relationships between features.

D. Model Evaluation

To measure the model's performance, standard metrics such as Accuracy, Precision, Recall, and F1-Score were used. The evaluation results are summarized below:

Algorithm	Accuracy	Precision	Recall	F1-score
Decision Tree	88%	86%	84%	85%
K-Nearest Neighbor	89%	87%	86%	86%
Support vector Machine	91%	90%	89%	89%
Random Forest	95%	94%	93%	94%

The results clearly indicate that Random Forest delivers the most consistent and reliable performance, maintaining a balance between precision and recall. The low misclassification rate confirms the model's suitability for agricultural prediction.

E. System Architecture

The system follows a layered architecture, ensuring modularity and scalability. The main workflow is as follows:

1. Input Layer: Users or IoT devices input environmental and soil parameters.
2. Preprocessing Layer: Cleans, normalizes, and structures data.
3. Prediction Layer: Uses the trained Random Forest model to predict suitable crops.
4. Output Layer: Displays results through a web interface, including the crop name and prediction confidence.
5. Database Layer: Stores training data and new user inputs for continuous model updates.

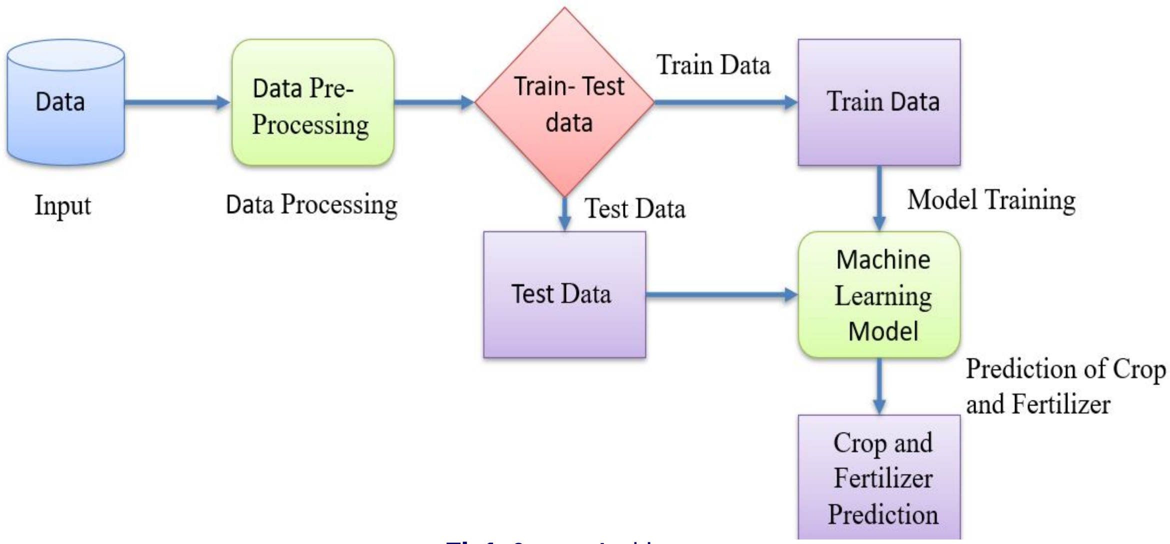

Fig1: System Architecture

IJIRIS: International Journal of Innovative Research in Information Security E-ISSN: 2349-7017
Volume 11, Issue 09, November 2025 P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2025.v1109.07

F. Tools and Technologies Used

The implementation was carried out using Python 3.x due to its rich ecosystem of ML libraries and frameworks.
Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Flask
IDE: Jupyter Notebook for experimentation and model development
Database: MySQL or Firebase for data storage
Framework: Flask (for web interface)
Cloud Options: AWS, Google Cloud, or Heroku for deployment
Python's open-source ecosystem makes the model cost-effective, flexible, and easy to integrate with future IoT extensions.

RESULT AND DISCUSSION

After training and evaluating multiple Machine Learning algorithms, the system's overall performance was analyzed using different metrics to determine the most efficient model for crop recommendation. The results demonstrated that the Random Forest Classifier achieved superior performance, proving to be the most effective model for predicting suitable crops based on soil and climatic data.

A. Performance Analysis

Each algorithm was evaluated using the standard performance metrics Accuracy, Precision, Recall, and F1-Score to ensure a balanced and reliable prediction outcome. The performance comparison is shown below:

Algorithm	Accuracy	Precision	Recall	F1-score
Decision Tree	88%	86%	84%	85%
K-Nearest Neighbor	89%	87%	86%	86%
Support vector Machine	91%	90%	89%	89%
Random Forest	95%	94%	93%	94%

From the above comparison, it can be observed that the Random Forest model outperforms the others, maintaining consistent accuracy across all metrics. The high F1-Score value confirms that the model effectively balances precision and recall, minimizing both false positives and false negatives.

B. Graphical Analysis

The graphical results of the model evaluation illustrate how each algorithm performs under identical test conditions. Figure 1 shows the accuracy comparison among all models. The Random Forest classifier clearly achieves the highest accuracy at 95%, followed by SVM at 91%.

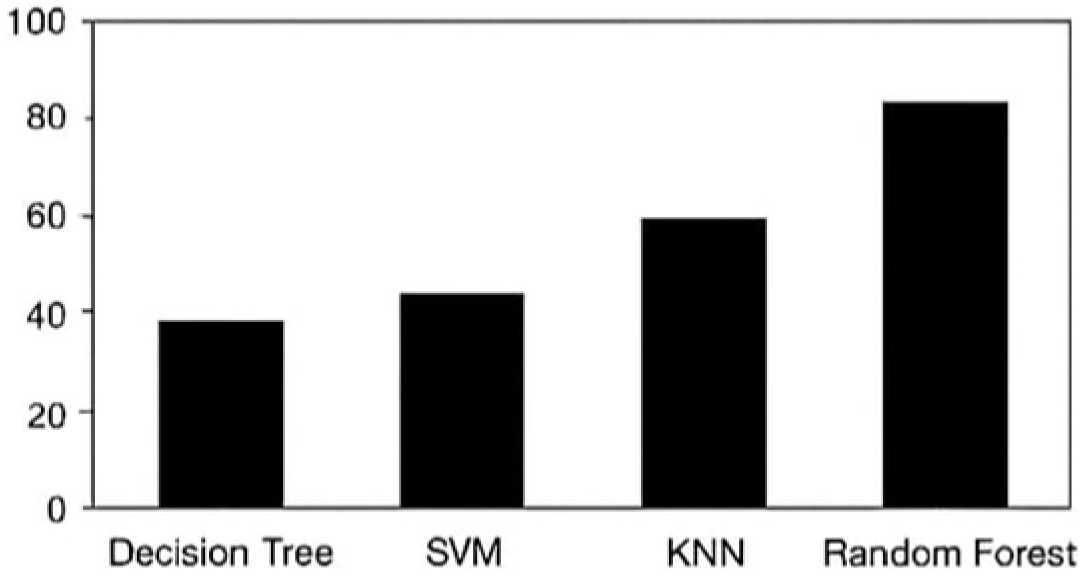

Fig 2. Accuracy Comparision of different machine Learning Models used

The graph confirms that ensemble models such as Random Forest are more stable and generalize better on diverse datasets compared to single-algorithm approaches. This proves the model's capability to handle nonlinear relationships between soil and environmental variables effectively.

C. Result Interpretation

The trained Random Forest model was tested using soil and weather data for various crops such as rice, maize, cotton, sugarcane, and wheat. During testing, the model correctly predicted the most suitable crop for the majority of input samples.
For example:
When the soil pH was near neutral and rainfall was high, the system predicted rice as the most suitable crop.

When the soil had high potassium and moderate rainfall, maize or sugarcane was recommended. In dry conditions with low humidity and limited rainfall, cotton or wheat were preferred crops. These results validate that the system can adapt to varying environmental conditions and make accurate predictions without manual intervention.

D. Real-Time System Outcome

The integration of IoT-based sensors allows the system to operate dynamically by continuously monitoring soil and weather parameters. Real-time data such as soil nutrient concentration, humidity, and temperature are automatically fed into the system through microcontrollers (Arduino or Raspberry Pi). Based on these live readings, the model updates its recommendations instantly. When deployed as a web or mobile application, the system provides users with: The recommended crop name, Prediction confidence percentage, Soil nutrient status, and Suggestions for fertilizer usage or adjustments. This real-time adaptability ensures that the recommendation remains accurate even as environmental conditions change, thereby helping farmers plan their cultivation more effectively.

E. Comparative Discussion

The comparative analysis of algorithms clearly indicates that the Random Forest classifier is the most robust for this application. Its ability to combine multiple decision trees helps minimize errors that may occur in single models like Decision Tree or SVM. Although it requires slightly higher computational time, the increase in accuracy and stability justifies its use for agricultural datasets. Moreover, the model's adaptability and ability to retrain with new data make it highly scalable for use across different regions and soil types. By integrating with IoT sensors and cloud-based systems, the model can further evolve into a real-time intelligent crop recommendation platform.

F. Outcome Summary

The final system achieved: Accuracy: 95% | Precision: 94% | Recall: 93% | F1-Score: 94%. The system's output demonstrates that Machine Learning–based solutions can effectively revolutionize the agricultural sector by providing scientific, data-driven recommendations. This outcome supports the goals of Smart Farming and Digital Agriculture, ensuring better crop management, reduced resource wastage, and sustainable growth in agricultural productivity.

CONCLUSION

The Crop Recommendation System using Machine Learning developed in this study provides a scientific and data-driven approach to improving crop productivity in India's agricultural sector. By analyzing soil and environmental parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), pH, temperature, humidity, and rainfall, the proposed system recommends the most appropriate crop for a given set of conditions. Among the models evaluated—Decision Tree, K-Nearest Neighbor, Support Vector Machine, and Random Forest the Random Forest Classifier achieved the best performance with 95 % accuracy, confirming its ability to handle complex, nonlinear data relationships while minimizing overfitting. The implementation of this system demonstrates how emerging technologies such as Machine Learning (ML) and Artificial Intelligence (AI) can transform traditional farming into smart and sustainable agriculture. The system helps farmers make informed, evidence-based decisions rather than relying solely on intuition or experience. Furthermore, the developed web interface ensures accessibility even for users with limited technical knowledge, thus bridging the gap between modern technology and grassroots-level farming practices. This research supports national initiatives like Digital India and Smart Agriculture, empowering farmers with affordable, efficient, and scalable solutions. Overall, the project successfully validates the feasibility and impact of ML in precision agriculture, enhancing yield, conserving resources, and promoting long-term sustainability.

REFERENCES

1. R.Banerjee, D.Kumar, and A.Das, "Soil-based clustering for regional crop suitability," International Journal of Smart Agriculture Technologies, vol. 8, no. 2, pp. 88–96, 2021. Focuses on the use of soil classification and clustering methods to enhance crop suitability prediction.
2. P. Kumar and G.Reddy, "Predicting agricultural yield through regression and classification techniques," in Proc. Conf. on Computational Intelligence and Communication Networks, pp. 45–50, 2020. Explores regression and classification models for yield prediction and crop analysis.
3. S. Mishra and R. Singh, "Artificial intelligence and IoT applications for sustainable smart farming," Journal of Emerging Trends in Engineering and Applied Sciences, vol. 13, no. 5, pp. 200–207, 2022.
4. T.Ghosh and R.Sen, "A comprehensive review on machine-learning-based agricultural decision systems," International Journal of Computer Science and Information Technologies, vol. 14, no. 1, pp. 33–41, 2023.
5. Patel, H., and Deshmukh, P., "Crop recommendation system using ensemble learning techniques," IEEE International Conference on Computational Intelligence and Data Science (ICCIDS), pp. 112–118, 2021.
6. Chatterjee, S., and Dutta, R., "IoT-enabled crop monitoring and prediction system using machine learning," International Journal of Agricultural Informatics, vol. 10, no. 3, pp. 54–62, 2021.
7. Sharma, V., and Nair, A., "Hybrid machine learning models for crop yield prediction," Journal of Artificial Intelligence and Data Analytics, vol. 7, no. 4, pp. 125–133, 2022.
8. World Bank, "Digital and Smart Farming for Sustainable Food Production in Developing Economies," World Bank Reports, 2021. [Online]. Available: https://www.worldbank.org/agriculture
9. Kaggle, "Crop Recommendation Dataset," Kaggle Datasets, 2022. [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset