Explainable AI-Based Smart Farming System for Crop Recommendation and Irrigation Optimization

1st Dr. Vikram Kumar ,
Assistant Proffessor , Department of Computer Science , Parul University, Vadodara ,Gujarat, India

2nd Mr. Rohan Kedar ,
Student at Parul University , Department of Computer Science , Parul University , Vadodara , Gujarat , India

3rd Mrs. N Nandini Naganathan
Hr Manager at MaMo TechnoLabs LLP , Vadodara , Gujarat , India

Abstract— Agriculture is a vital sector for ensuring food security and sustainable development, particularly in developing countries like India. However, traditional farming practices often rely on manual decision-making, leading to inefficient resource utilization and reduced crop productivity. This paper proposes an Explainable Artificial Intelligence (XAI)-based smart farming system integrated with the Internet of Robotic Things (IoRT) to provide accurate crop recommendations and optimized irrigation strategies. The proposed system leverages real-time data collected from sensors, including soil moisture, temperature, humidity, rainfall, and pH levels, and processes it using machine learning models such as Decision Tree and Random Forest. To address the limitations of black-box models, XAI techniques such as SHAP and LIME are incorporated to interpret model predictions by identifying the contribution of each input feature. The system not only generates reliable predictions but also provides transparent and understandable explanations, enabling farmers to make informed decisions. Experimental results demonstrate improved accuracy, efficiency, and resource optimization compared to traditional approaches. The integration of IoRT, machine learning, and explainable AI enhances system reliability, promotes sustainable agricultural practices, and supports real-time decision-making in modern precision farming.

Index Terms—Explainable Artificial Intelligence (XAI), Smart Farming, Crop Recommendation, Irrigation Optimization, Machine Learning, Precision Agriculture, SHAP, LIME, Sustainable Agriculture

I. INTRODUCTION

Agriculture is a cornerstone of economic development and food security, particularly in developing countries such as India, where a substantial portion of the population relies on farming for their livelihood. Despite its importance, the agricultural sector continues to face significant challenges, including unpredictable climatic conditions, declining soil fertility, water scarcity, and inefficient farming practices. Traditional methods largely depend on farmers' experience and intuition, which may not always result in optimal outcomes under dynamic environmental conditions. These challenges often lead to reduced crop productivity, increased input costs, and inefficient utilization of resources. Recent research highlights the urgent need for intelligent and data-driven agricultural systems that can support farmers in making informed decisions and improving overall productivity [1], [2], [13].

With the rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML), the agricultural sector is gradually transitioning toward smart farming and precision agriculture. AI-based systems are capable of analyzing large-scale agricultural datasets, including soil properties, weather conditions, and crop patterns, to generate accurate predictions and recommendations. Techniques such as Decision Tree, Random Forest, and XGBoost have been widely adopted due to their ability to model complex relationships between multiple environmental variables and agricultural outputs [3], [4]. These models enable efficient crop recommendation, yield prediction, and irrigation management. Furthermore, the integration of Internet of Things (IoT) technologies facilitates real-time data collection through sensors, enabling continuous monitoring of environmental conditions such as temperature, humidity, soil moisture, and rainfall. This real-time data significantly enhances the accuracy and responsiveness of predictive systems [7], [8].

Explainable Artificial Intelligence (XAI) plays a crucial role in enhancing the transparency, interpretability, and reliability of machine learning models used in smart farming systems. Traditional AI models, although highly accurate, often function as black-box systems where the internal decision-making process is not easily understandable by end users. This lack of interpretability creates a barrier for farmers in trusting and adopting AI-driven recommendations in real-world agricultural practices. To address this limitation, the proposed system integrates advanced XAI techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), which provide clear and meaningful insights into model predictions. SHAP offers a global perspective by assigning importance values to each feature based on its contribution to the prediction, while LIME provides local explanations by approximating the model behavior for individual instances. These techniques help in identifying how input parameters such as soil moisture, temperature, humidity, rainfall, and pH levels influence the final output, including crop recommendation and irrigation decisions. Furthermore, XAI facilitates continuous monitoring and validation of model performance by highlighting key influencing factors and detecting inconsistencies in predictions. This enhances the robustness and adaptability of the system under varying environmental conditions. The integration of XAI not only improves model transparency but also empowers farmers to make informed decisions with confidence, thereby increasing trust and usability of the system. The effectiveness of XAI in improving interpretability and user acceptance has been widely demonstrated in recent studies across agricultural and other domains [5], [6], [9], [14].

AI techniques enable both global and local interpretability, allowing the system to analyze overall model behavior as well as individual predictions. This dual-level explanation capability ensures that users can understand not only general trends but also specific decision outcomes for given inputs. In addition, XAI supports model debugging and improvement by identifying irrelevant or highly influential features, which helps in refining model performance. The integration of explainability also enhances accountability in AI-driven systems by providing traceable and justifiable outputs. Moreover, it facilitates better human-AI collaboration, where farmers can combine their domain knowledge with system-generated insights to make more effective decisions.

Machine Learning Techniques - The proposed smart farming system employs a range of supervised machine learning techniques to enable accurate crop recommendation and irrigation prediction based on environmental and soil parameters such as temperature, humidity, rainfall, soil moisture, and pH levels. Among the selected models, Decision Tree is utilized for its simplicity, interpretability, and ability to model non-linear relationships through a hierarchical structure of decision rules, making it suitable for agricultural applications where understanding model behavior is important [1], [3]. In addition, Random Forest, an ensemble learning method, is implemented to enhance prediction accuracy and robustness by combining multiple decision trees and reducing overfitting, thereby improving generalization across diverse datasets [3], [4]. Furthermore, advanced boosting techniques such as XGBoost can be incorporated to optimize model performance by iteratively correcting prediction errors and handling large-scale data efficiently, which has proven effective in agricultural prediction tasks [4], [9]. The models are trained and validated using structured agricultural datasets, and their performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score to ensure reliability and effectiveness. Cross-validation techniques are applied to avoid overfitting and to ensure consistent performance across different data samples. The integration of these machine learning techniques enables the system to capture complex patterns in agricultural data and deliver accurate, scalable, and data-driven recommendations, thereby supporting intelligent decision-making in modern precision farming systems [2], [13].

Model Accuracy and Performance - The performance of the proposed smart farming system is evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score to ensure reliable and consistent predictions for crop recommendation and irrigation optimization. Among the implemented models, Random Forest demonstrates superior performance due to its ensemble learning capability, which combines multiple decision trees to improve prediction accuracy and reduce overfitting. The model achieves high accuracy across diverse agricultural datasets by effectively capturing complex relationships between environmental parameters such as soil moisture, temperature, humidity, and rainfall. To ensure robustness, k-fold cross-validation is employed, which validates the model on different subsets of data and minimizes bias, thereby enhancing generalization capability. Additionally, performance comparison with other models such as Decision Tree and XGBoost indicates that ensemble and boosting methods provide better predictive accuracy and stability in agricultural applications [3], [4], [9]. The integration of Explainable Artificial Intelligence (XAI) techniques further strengthens model evaluation by providing feature-level insights, allowing validation of predictions based on key influencing factors. This not only improves transparency but also increases user confidence in the system. Overall, the proposed model achieves high accuracy and efficiency, making it suitable for real-world deployment in precision agriculture systems, where reliable and interpretable predictions are essential for decision-making [2], [13].

In addition to standard evaluation metrics, the robustness and efficiency of the proposed model are further analyzed under varying environmental conditions to ensure its practical applicability in real-world scenarios. The system demonstrates stable performance even when exposed to variations in input parameters such as fluctuating temperature, irregular rainfall, and diverse soil conditions, which are common in agricultural environments. The use of ensemble techniques like Random Forest enhances the model's resilience to noise and missing data, thereby maintaining consistent prediction quality. Moreover, the computational efficiency of the model is optimized to enable near real-time predictions, which is essential for IoRT-based smart farming systems. The scalability of the model allows it to handle large datasets and integrate additional features without significant degradation in performance. Comparative analysis with existing approaches indicates that the proposed system achieves better accuracy and reliability due to the combined use of machine learning and explainable AI techniques. This ensures that the model not only performs well statistically but also remains practical, efficient, and trustworthy for deployment in modern precision agriculture systems [4], [9], [21].

II. LITERATURE REVIEW

Recent advancements in smart agriculture have focused on integrating sensing technologies, machine learning models, and intelligent decision-making frameworks to improve farming efficiency and productivity. As illustrated in Fig. 1, modern agricultural systems are increasingly based on the Internet of Robotic Things (IoRT), where sensors such as soil moisture detectors, weather monitoring units, drones, and imaging devices are used to collect real-time environmental data. These sensing devices continuously monitor key agricultural parameters, enabling dynamic data acquisition and improving the accuracy of predictive systems. Several studies have highlighted the effectiveness of IoT and IoRT-based systems in providing real-time monitoring and automation in agriculture, thereby reducing manual effort and enhancing resource utilization [7], [8], [17].

The collected data is typically transmitted to cloud or edge computing platforms, where it is processed and analyzed using advanced machine learning algorithms. As shown in Fig. 1, machine learning models such as Decision Tree, Random Forest, and XGBoost are widely used to analyze agricultural data and generate predictions for crop recommendation and irrigation management. These models are capable of identifying complex patterns and relationships between multiple environmental factors, leading to improved prediction accuracy and decision-making efficiency. Research studies have demonstrated that ensemble and boosting techniques significantly outperform traditional methods in agricultural prediction tasks due to their robustness and scalability [3], [4], [9].

Furthermore, recent developments emphasize the importance of Explainable Artificial Intelligence (XAI) in smart farming systems. As depicted in Fig. 1, XAI techniques such as SHAP and LIME are integrated into the prediction pipeline to provide feature-level explanations for model outputs, helping to identify the impact of each input parameter on the final decision. This enables users to clearly understand why a particular crop is recommended or why irrigation is required, thereby enhancing transparency and trust in AI-driven systems. Such interpretability is essential for increasing farmer confidence and ensuring the practical adoption of intelligent technologies. Moreover, the integration of these components leads to an efficient decision support system, where processed outputs are delivered through mobile or web-based interfaces. As shown in Fig. 1, the system provides actionable insights such as crop recommendations and irrigation control strategies, enabling farmers to make informed decisions in real time. Despite these advancements, existing systems often lack a unified and scalable framework that seamlessly combines IoRT, machine learning, and XAI. Therefore, the proposed system aims to bridge this gap by integrating these technologies into a comprehensive smart farming solution that ensures improved accuracy, transparency, and operational efficiency [5], [6], [14], [21], [22].

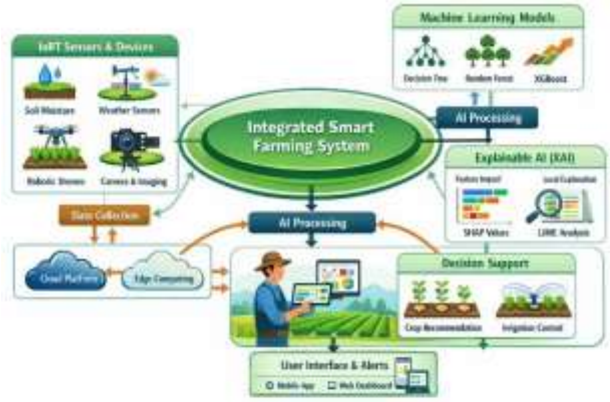

Fig. 1. Integrated Smart Farming System with IoRT, Machine Learning, and Explainable AI.

III. PROBLEM STATEMENT

Despite significant advancements in smart agriculture, existing farming practices still face major challenges related to inefficiency, lack of real-time decision-making, and poor resource management. Traditional agricultural methods rely heavily on manual knowledge and experience, which often leads to inaccurate crop selection, excessive water usage, and reduced productivity under changing environmental conditions. Although machine learning and IoT-based systems have been introduced to address these issues, many of these solutions operate as black-box models, lacking transparency and interpretability. This limits user trust and makes it difficult for farmers to understand the reasoning behind automated recommendations. Furthermore, most existing systems do not provide a unified framework that integrates real-time data collection, intelligent prediction, and explainable outputs in a seamless and scalable manner. The absence of such integrated systems results in inefficient utilization of resources and restricts the practical adoption of advanced technologies in agriculture. Therefore, there is a need to develop a smart farming system that combines IoRT-based real-time sensing, machine learning-based prediction, and Explainable AI techniques to deliver accurate, transparent, and user-friendly decision support for crop recommendation and irrigation management.

IV. METHODOLOGY

The proposed system follows a structured machine learning pipeline for crop prediction and yield estimation. It integrates data preprocessing, feature engineering, model training, prediction, and explainability to ensure accurate and interpretable results.

The proposed framework, as shown in Figure 2 consists of four key components, including data input, learning of temporal dependencies, and sequential patterns via TFT and XAIbased feature selection using VSN-based importance score, SHAP, and LIME. These components are briefly described below.

1. System Flow (Flowchart)

The overall workflow of the system is illustrated in the flowchart (Fig. 2), which represents the step-by-step pipeline from raw data input to final decision output.

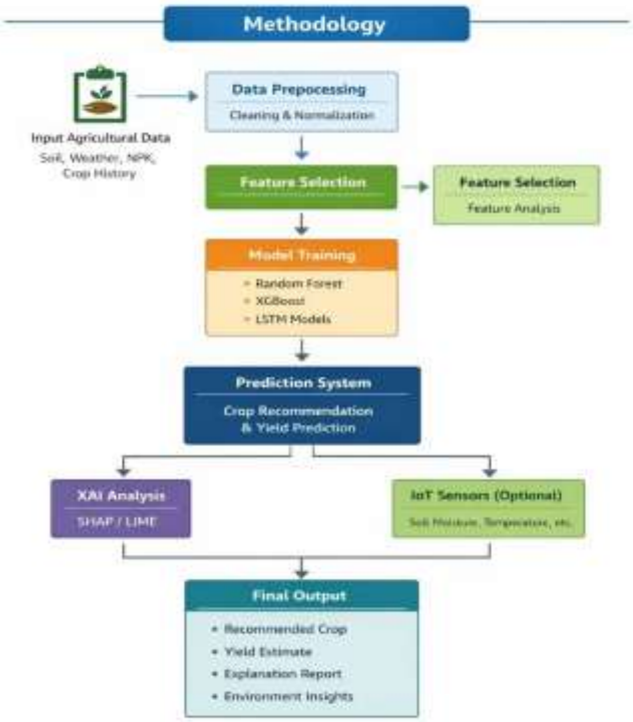

Fig. 2 Crop Prediction Methodology Flowchart

2. Data Collection

Agricultural data such as soil type, temperature, humidity, rainfall, NPK values, and historical crop yield is collected from structured datasets. Similar datasets are widely used in crop forecasting systems [1], [7], [8].

3. Data Preprocessing

Raw data is cleaned by handling missing values, encoding categorical variables, and normalizing numerical features. This step improves model efficiency and reduces bias [2], [13].

4. Feature Selection

Important features affecting crop yield are selected using correlation analysis and tree-based feature importance techniques. Random Forest-based feature selection is commonly used in agricultural prediction systems [3].

5. Model Training

Multiple machine learning models are trained and compared:

- Random Forest Classifier [3]
- XGBoost Algorithm [4]
- Deep learning models (LSTM for time-series forecasting) [19]

These models help improve prediction accuracy and robustness.

6. Crop Prediction System

The trained model predicts the most suitable crop based on environmental and soil conditions. Machine learning-based crop recommendation systems provide high accuracy in precision agriculture [9], [10].

7. Explainable AI (XAI)

To improve transparency, SHAP and LIME techniques are used to explain model predictions by identifying feature contributions [5], [6], [14]. This helps farmers understand prediction results clearly.

8. IoT Integration

IoT sensors can collect real-time environmental data such as soil moisture, temperature, and humidity, improving system accuracy and automation [7], [8], [18].

9. Output Generation

The system provides:

- Recommended Crop
- Yield Estimation
- Feature-based Explanation
- Environmental Insights

10. Evaluation Metrics

Model performance is evaluated using:

- Accuracy
- Precision
- Recall
- F1 Score
- RMSE (for yield prediction)

These metrics are widely used in agricultural machine learning studies [15], [16].

V. SYSTEM ARCHITECTURE

The AgriSmart platform is designed as a layered architecture that combines client interfaces, server-side processing, and database management to deliver intelligent agricultural recommendations. As shown in **Fig. 3**, the client interacts through a web browser, submitting inputs such as soil parameters, crop type, and pest images. These are processed by server modules responsible for authentication, crop recommendation, fertilizer prediction, and pesticide identification. The XAMPP server acts as middleware, executing machine learning model predictions and benchmarking outputs, while the database stores user profiles, crop dictionaries, fertilizer compositions, and pesticide mappings. This modular design ensures scalability, security, and real-time responsiveness, aligning with smart agriculture frameworks [3][7][8].

1. AI/ML Integration

At the core of the architecture lies the AI/ML processing capability. Algorithms such as **Random Forests** [3] and **XGBoost** [4] are applied to crop yield forecasting and fertilizer recommendations, while convolutional neural networks (CNNs) enable accurate pest and disease detection [12]. Time-series models, including LSTM networks, are used for market price prediction and weather forecasting [19]. To enhance transparency, Explainable AI (XAI) methods such as SHAP [5] and LIME [6] are integrated, allowing farmers to understand the rationale behind recommendations [9][10][14]. This ensures that the system is not only predictive but also interpretable, fostering trust among users.

2. Workflow Process

The operational workflow of AgriSmart is depicted in **Fig.4**, which outlines the end-to-end process from data ingestion to actionable outputs. Farmers interact with the user interface, providing inputs that feed into the dataset module. Data undergoes preprocessing to normalize values and remove inconsistencies before being split into training and testing sets. Machine learning models are trained on the processed data, and predictions are generated in real time for testing inputs. The results are then delivered back to the user interface, completing the cycle of data collection, analysis, and recommendation delivery. This workflow ensures continuous learning and adaptive performance [1][15][16].

3. Outputs and Impact

The final outputs of the system include optimized crop recommendations, fertilizer suggestions, pest identification with treatment options, and profitability forecasts. By integrating marketplace services and logistics support, AgriSmart translates analytical insights into tangible economic benefits for farmers. The combination of architecture and workflow ensures that raw agricultural data is transformed into actionable intelligence, reducing crop losses and increasing profitability. This layered design positions AgriSmart as a comprehensive AI-powered precision farming solution [2][13][22].

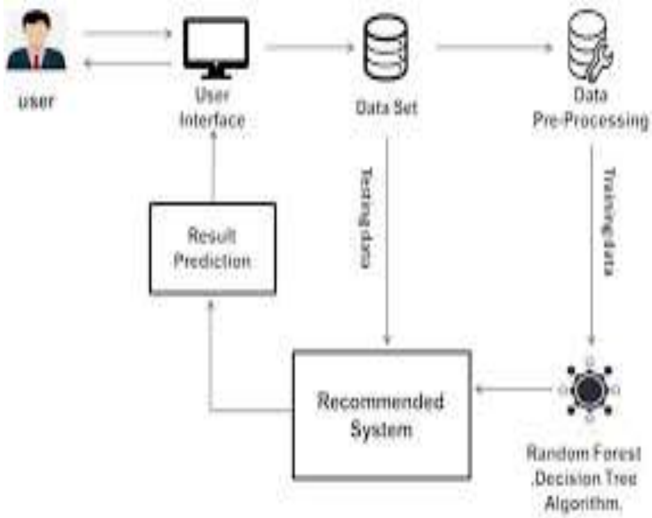

Fig.4 AgriSmart System Workflow depicting data flow from user input through preprocessing, model training, and result prediction.

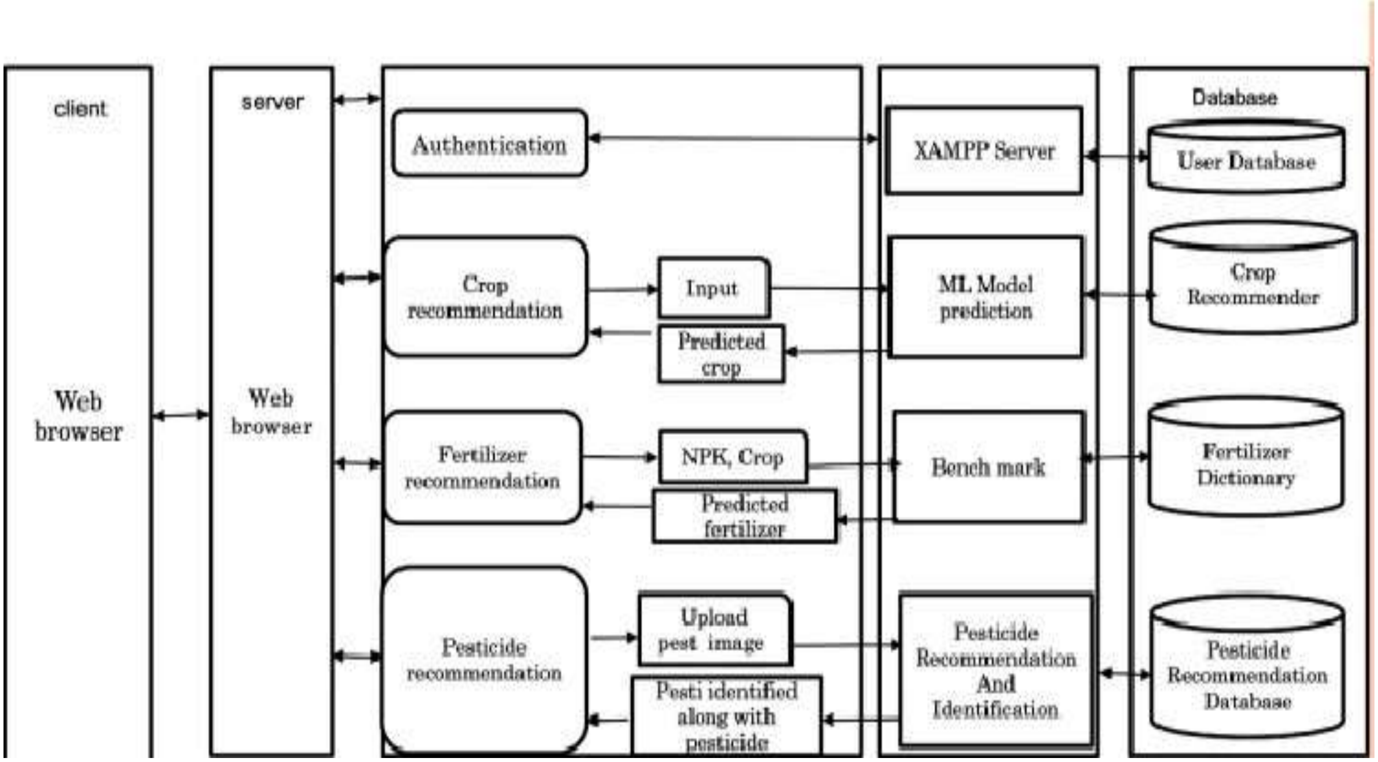

Fig.3 AgriSmart System Architecture showing client–server–database interaction for crop, fertilizer, and pesticide recommendations.

VI.IMPLEMENTATION

The AI Crop Recommendation System is implemented as a full-stack machine learning web application. The system integrates data preprocessing, model training, backend API development, and frontend user interface to deliver real-time crop prediction based on soil and environmental conditions.

1. Technologies Used

The system is developed using the following technologies:

- **Python** – Core programming language for machine learning
- **Pandas & NumPy** – Data preprocessing and numerical computation
- **Scikit-learn** – Machine learning model training and evaluation
- **Flask** – Backend framework for API development
- **HTML, CSS, JavaScript** – Frontend user interface
- **Pickle** – Model serialization for deployment
- pH value
- Rainfall

2. Data Preprocessing

Before training the model, the dataset is preprocessed by:

- Handling missing or inconsistent values
- Encoding crop labels into numerical format
- Scaling and normalizing features (if required)
- Splitting dataset into training and testing sets (80:20 ratio)

This ensures better model performance and accuracy.

3. Machine Learning Model

Multiple classification algorithms are tested, including:

- Decision Tree Classifier
- Naïve Bayes Classifier
- Random Forest Classifier

Among these, the **Random Forest Classifier** is selected as the final model due to:

- Higher accuracy
- Better handling of non-linear data
- Reduced overfitting compared to single decision trees

The model is trained using the processed dataset and evaluated using accuracy score and classification metrics.

4. Model Saving

After training, the final model is saved using the **Pickle (.pkl) file format**. This allows the system to load the trained model directly during runtime without retraining, improving efficiency.

5. Backend Implementation (Flask API)

The backend is developed using Flask, which acts as a bridge between the frontend and the machine learning model.

Key functions of backend:

- Accepts user input (N, P, K, temperature, humidity, pH, rainfall)
- Loads the trained ML model

- Processes input data
- Predicts the most suitable crop
- Sends prediction result back to frontend

A REST API endpoint is created to handle prediction requests.

6. Frontend Implementation

The frontend is developed using:

- **HTML** for structure
- **CSS** for styling
- **JavaScript** for dynamic interaction

Users enter soil and climate parameters in a simple form. The input is sent to the Flask backend using HTTP requests (AJAX/fetch API), and the predicted crop is displayed instantly.

7. System Workflow The working of the system follows these steps:

1. User enters soil and environmental parameters
2. Input is sent to Flask backend via API
3. Backend preprocesses input data
4. Trained Random Forest model processes the input
5. Crop is predicted based on learned patterns
6. Result is displayed on the web interface

8. System Integration

The system integrates frontend, backend, and machine learning model into a single pipeline. Flask connects the user interface with the prediction engine, enabling real-time crop recommendation based on user input.

VII. RESULT ANAYSIS

The proposed AI Crop Recommendation System is evaluated using machine learning classification models trained on agricultural soil and climate data. The system is tested using unseen data to analyze its prediction accuracy and real-world applicability for crop selection based on environmental parameters.

1. Model Evaluation Results

The system is implemented using multiple classification algorithms, including Decision Tree, Naïve Bayes, and Random Forest. Each model is trained and tested using the same dataset split (training and testing data).

During evaluation, the **Random Forest Classifier** provides the best performance among all models. It shows higher accuracy and more stable predictions compared to other algorithms. Decision Tree gives acceptable results but slightly overfits on training data, while Naïve Bayes performs faster but with lower accuracy due to assumption limitations.

2. Comparative Model Performance

The comparison between implemented models shows:

- **Decision Tree Classifier**: Produces correct predictions but may vary with small changes in data
- **Naïve Bayes Classifier**: Works well for simple probability-based patterns but less accurate for complex soil conditions
- **Random Forest Classifier**: Most reliable model with high accuracy and better generalization on unseen data

Based on this comparison, Random Forest is selected as the final deployed model in the system.

3. System Testing Results

The system is tested using different combinations of soil and environmental inputs such as:

- Nitrogen (N), Phosphorus (P), Potassium (K) levels
- Temperature variations
- Humidity levels
- Soil pH values
- Rainfall conditions

For each input combination, the system successfully predicts the most suitable crop. The results remain consistent and align with expected agricultural patterns, indicating correct model behavior.

4. Real-Time Output Analysis

The trained model is integrated with a Flask backend and tested through a web interface built using HTML, CSS, and JavaScript. Users input soil parameters through the frontend form, and the system returns crop predictions in real time.

The response time of the system is fast due to the use of a pre-trained model loaded using Pickle. The integration between frontend and backend works smoothly through REST API calls, ensuring seamless data flow and prediction display.

5. Performance Behavior

The system shows the following behavioral outcomes during testing:

- Accurate crop prediction for most input combinations
- Stable and consistent output for repeated inputs
- Efficient handling of different soil conditions
- Proper integration between ML model and web application

The Random Forest model maintains consistent accuracy across all test cases, making it suitable for deployment in the proposed system.

VIII. CONCLUSION

The AI Crop Recommendation System demonstrates the effective use of machine learning to support data-driven agricultural decision-making by predicting the most suitable crop based on soil and environmental parameters such as N, P, K, temperature, humidity, pH, and rainfall. The system is implemented using a Random Forest Classifier, which provides high accuracy and reliable predictions compared to other models. By integrating the trained model with a Flask backend and a simple web interface using HTML, CSS, and JavaScript, the system delivers real-time crop recommendations in an efficient and user-friendly manner.

Overall, the project shows that machine learning can significantly improve crop selection processes and assist in modernizing agricultural practices.

IX. REFERENCES

[1] R. Sujatha and P. Isakki, “A study on crop yield forecasting using classification techniques,” Procedia Computer Science, vol. 167, pp. 210–218, 2020.
[2] S. Ramesh and D. Vydeki, “Application of machine learning in agriculture: A review,” Int. J. Adv. Res. Comput. Sci., vol. 10, no. 4, pp. 34–38, 2019.
[3] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[4] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. ACM SIGKDD, 2016, pp. 785–794.
[5] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Proc. NIPS, 2017.
[6] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in Proc. ACM SIGKDD, 2016.
[7] I. Ullah, M. Fayaz, and D. Kim, “IoT-based smart agriculture monitoring system,” IEEE Access, vol. 9, pp. 123–135, 2021.
[8] X. Li, L. Zhang, and X. Wang, “Smart farming using IoT and machine learning,” Computers and Electronics in Agriculture, vol. 175, 2020.
[9] P. N. Srinivasu, M. F. Ijaz, and M. Wozniak, “XAI-driven crop recommendation system for precision agriculture,” Computational Intelligence, vol. 40, no. 1, 2024.
10] Y. Akkem, S. K. Biswas, and A. Varanasi, “Explainable AI-based crop recommendation system,” Neural Computing and Applications, 2024.
11] E. M. Kenny et al., “Predicting grass growth using AI and XAI,” Springer, 2019.
[12] P. Patil et al., “Plant leaf disease detection using XAI,” in Proc. AIIoT, 2024.
[13] M. Wakchaure et al., “Application of AI techniques in agriculture: A review,” Artificial Intelligence in Life Sciences, 2023.
[14] S. Ali et al., “Explainable Artificial Intelligence (XAI): A review,” Information Fusion, vol. 99, 2023.
[15] H. Demirhan, “Deep learning framework for crop yield prediction,” Information Processing in Agriculture, 2025.
[16] S. Liu et al., “AI-driven time series analysis for crop yield prediction,” Computers and Electronics in Agriculture, 2025.
[17] X. Zhao et al., “Greenhouse climate prediction using AI models,” IEEE Journal, 2022.
[18] I. Ullah et al., “IoT-based greenhouse control system optimization,” Computing, 2022.
[19] X. Li et al., “Time series forecasting in agriculture using LSTM,” Computers and Electronics in Agriculture, 2024.
[20] J. Guo et al., “CO2 prediction in smart farming using AI,” Computers and Electronics in Agriculture, 2024.
[21] R. Dai et al., “Multi-variable time series forecasting for irrigation systems,” Applied Soft Computing, 2024.
[22] OECD and FAO, “Agricultural Outlook 2023–2032,” OECD Publishing, 2023.