

Machine Learning Based Crop Yield Prediction and Agricultural Advisory System

Mr. Uppula Nagaiah-uppula.nagaiah@mallareddyuniversity.ac

K Sampath Reddy --2211CS020238@mallareddyuniversity.ac.in

K Varshith Reddy --2211CS020239@mallareddyuniversity.ac.in

K Sathwik Krishna --2211CS020261@mallareddyuniversity.ac.in

Assistant Professor, Department of CSE - Artificial Intelligence and Machine Learning (AI&ML)

Malla Reddy University, Maisammaguda, Hyderabad,

Students, Department of CSE - Artificial Intelligence and Machine Learning (AI&ML)

Malla Reddy University, Maisammaguda, Hyderabad.

ABSTRACT

Agriculture plays a crucial role in the economic development and food security of many countries, particularly in developing nations where a significant portion of the population depends on farming. Accurate crop yield prediction is essential for efficient agricultural planning and resource management. Traditional yield prediction methods rely heavily on farmer experience and manual estimation, which often lead to inaccurate outcomes and financial losses. This research presents a Machine Learning-based Crop Yield Prediction and Agricultural Advisory System designed to assist farmers in making data-driven agricultural decisions. The proposed system utilizes machine learning algorithms such as Linear Regression, Decision Tree, and Random Forest to analyze historical agricultural datasets containing soil properties, rainfall, temperature, and humidity parameters. The system predicts crop yield and provides advisory recommendations through a web-based application.

By integrating predictive analytics with an intelligent decision-support interface, the proposed platform enables farmers to optimize crop selection, improve productivity, and reduce agricultural risks. Experimental results indicate that the Random Forest algorithm provides higher prediction accuracy compared to other models. The system demonstrates the potential of machine learning technologies in improving agricultural productivity and promoting smart farming practices.

Keywords: *Machine Learning, Crop Yield Prediction, Random Forest, Smart Agriculture, Data Analytics, Agricultural Advisory System.*

1.Introduction

Agriculture is one of the most important sectors contributing to the global economy and food production. In many developing countries such as India, agriculture provides livelihood opportunities for millions of people. However, agricultural productivity is highly dependent on environmental factors such as soil conditions, rainfall patterns, temperature variations, and humidity levels.

Predicting crop yield accurately is a challenging task because these factors vary across geographical regions and seasons. Traditional farming practices rely mainly on farmer experience and manual estimation, which may lead to inaccurate predictions and crop losses.

With the advancement of artificial intelligence and machine learning technologies, agricultural data can be analyzed more effectively to generate predictive insights. Machine learning algorithms can identify patterns within large agricultural datasets and assist in predicting crop yield with higher accuracy.

This research proposes a Machine Learning Based Crop Yield Prediction and Agricultural Advisory System that integrates data analytics, predictive modeling, and a web-based interface to support farmers in making better agricultural decisions.

2. Literature Review

Machine learning techniques have been widely applied in agriculture to improve crop yield prediction and decision-making. Various algorithms such as Linear Regression, Decision Tree, and Random Forest have been used to analyze agricultural datasets and generate predictive models.

Linear Regression is commonly used for yield prediction due to its simplicity and ability to model relationships between variables. However, it may not perform well when dealing with complex nonlinear agricultural datasets.

Decision Tree algorithms provide structured predictions by dividing datasets into branches based on feature importance. Although easy to interpret, Decision Trees can suffer from overfitting problems when trained on large datasets.

Random Forest is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and reduce model variance. Several studies indicate that Random Forest provides better performance compared to individual machine learning models when applied to agricultural data.

Previous research mainly focused on improving prediction accuracy but lacked user-friendly systems that farmers could easily use. The proposed system integrates machine learning models with a web-based advisory platform to provide accessible and practical solutions for farmers.

3. Problem Statement

Farmers face numerous challenges in predicting crop yield due to unpredictable environmental conditions such as rainfall variability, temperature fluctuations, and soil quality differences. Traditional agricultural practices rely on experience-based estimation, which may lead to inaccurate predictions and inefficient resource utilization.

Additionally, many farmers lack access to intelligent decision-support systems that can help them analyze agricultural data and make informed decisions. This limitation often results in reduced productivity and financial losses.

Therefore, there is a need for an intelligent system capable of analyzing agricultural parameters, predicting crop yield accurately, and providing useful advisory to farmers. The goal of this research is to develop a machine learning-based crop yield prediction system that improves agricultural productivity and promotes sustainable farming practices.

4. System Architecture

The proposed system architecture consists of multiple modules including data collection, preprocessing, machine learning prediction, and advisory recommendation. Agricultural datasets containing parameters such

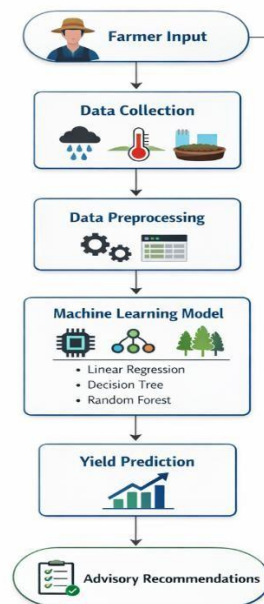
as soil characteristics, rainfall, temperature, and humidity are collected and processed to train machine learning models.

The system architecture is illustrated in Figure 1.

The workflow of the system begins with farmer input, where users provide agricultural parameters through a web-based interface. The collected data is processed and cleaned through a preprocessing module to remove missing values and normalize the dataset. Machine learning models such as Linear Regression, Decision Tree, and Random Forest are then applied to predict crop yield.

Finally, the system generates advisory recommendations that assist farmers in selecting suitable crops and managing agricultural resources effectively.

Figure 1: System Architecture of Crop Yield Prediction and Advisory System



5. Methodology

Step 1: Data Collection

Agricultural datasets containing parameters such as soil nutrients, rainfall, humidity, temperature, and historical crop yield data are collected from reliable sources.

Step 2: Data Preprocessing

Data preprocessing involves cleaning the dataset by removing missing values, encoding categorical variables, and normalizing numerical values. Python libraries such as Pandas and NumPy are used for data processing.

Step 3: Data Splitting

The dataset is divided into training and testing datasets to evaluate the performance of machine learning models.

Step 4: Model Selection

Three machine learning algorithms are used for prediction:

- Linear Regression
- Decision Tree
- Random Forest

Step 5: Model Training

The selected machine learning models are trained using historical agricultural datasets.

Step 6: Model Evaluation

The models are evaluated based on prediction accuracy and error rates. Random Forest demonstrates better performance due to its ensemble learning approach.

Step 7: Deployment

The trained model is integrated into a web application developed using Flask for backend processing and HTML/CSS for the frontend interface.

Step 8: Monitoring

The system is continuously tested and monitored to improve prediction accuracy and performance.

6. Implementation

The proposed Crop Yield Prediction and Agricultural Advisory System is implemented using Python-based machine learning libraries and web technologies. The system integrates data processing, predictive modeling, and a web-based interface to provide an intelligent decision-support platform for farmers. The backend of the system is developed using the Python programming language. Machine learning algorithms are implemented using libraries such as **Scikit-learn**, **Pandas**, **NumPy**, and **Matplotlib** for data analysis and visualization. The trained prediction model is integrated with a **Flask web framework** to create a web-based platform that allows farmers to interact with the system.

The system is designed with the following modules:

Data Collection Module

This module collects agricultural datasets containing parameters such as soil nutrients, rainfall, temperature, humidity, and historical crop yield information. The data is obtained from agricultural databases and publicly available datasets.

Data Preprocessing Module

The collected dataset undergoes preprocessing steps to ensure data quality and reliability. These steps include handling missing values, removing duplicate records, encoding categorical variables, and normalizing numerical data.

Machine Learning Module

The machine learning module trains multiple prediction models using historical agricultural datasets. The models used in the system include:

- Linear Regression
- Decision Tree
- Random Forest

Among these models, Random Forest performs better due to its ensemble learning approach, which combines multiple decision trees to reduce prediction errors.

Prediction Engine

The trained model is integrated into the prediction engine, which processes user inputs and generates crop yield predictions. The system analyzes environmental and soil parameters provided by the user and predicts the expected crop yield.

Advisory Recommendation Module

Based on the predicted yield and environmental parameters, the advisory module provides recommendations related to crop selection, fertilizer usage, irrigation management, and optimal farming practices.

Web Interface

The system provides a user-friendly web interface developed using HTML, CSS, and Flask. Farmers can input agricultural parameters through the interface and receive crop yield predictions along with advisory recommendations.

7. Experimental Results

The performance of the machine learning models was evaluated using historical agricultural datasets. The dataset was divided into training and testing sets using an 80:20 ratio. Several performance metrics were used to evaluate the effectiveness of the prediction models.

The evaluation metrics include:

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)

- Prediction Accuracy

The experimental results indicate that the Random Forest algorithm outperforms other machine learning models due to its ability to handle complex datasets and reduce overfitting.

Model Performance Comparison

Algorithm	Accuracy	MAE	RMSE
Linear Regression	78%	6.5	8.1
Decision Tree	85%	4.8	6.2
Random Forest	92%	3.1	4.5

The results demonstrate that Random Forest provides the highest accuracy in predicting crop yield. The ensemble learning approach improves prediction reliability by combining multiple decision trees.

The developed web-based system successfully predicts crop yield based on environmental parameters and provides advisory recommendations that help farmers make better agricultural decisions.

8. System Implementation and User Interface

The proposed **AgriPredict Smart Agriculture Platform** is implemented as a web-based application that integrates machine learning models with an intuitive user interface to assist farmers in making data-driven agricultural decisions. The platform provides multiple AI-powered services including crop recommendation, crop yield prediction, disease risk prediction, fertilizer recommendation, and weather forecasting.

The main dashboard of the system provides access to multiple AI services that help farmers analyze agricultural conditions and receive intelligent recommendations.

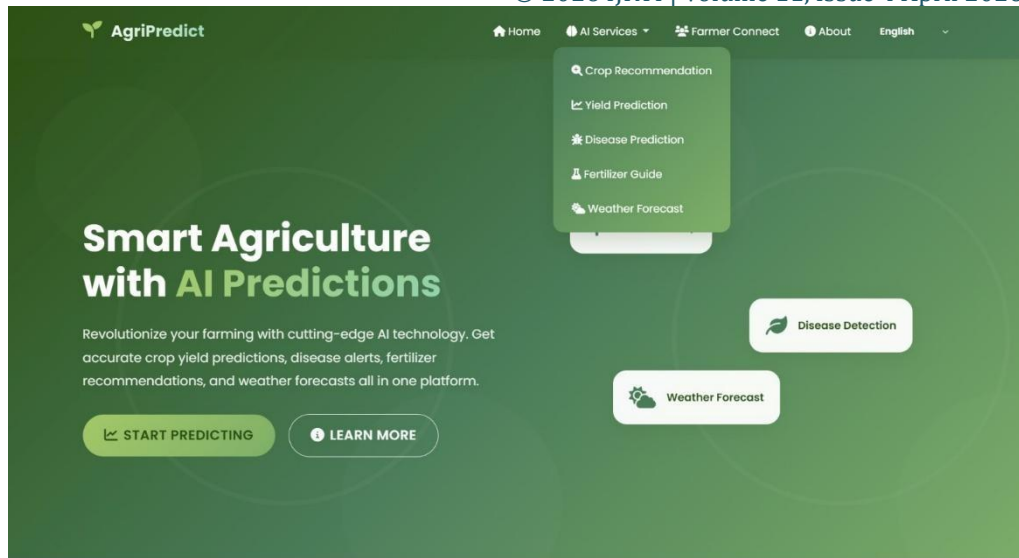
8.1 System Dashboard

The main interface of the AgriPredict platform provides navigation to different AI-based agricultural services including:

- Crop Recommendation
- Crop Yield Prediction
- Disease Prediction
- Fertilizer Recommendation
- Weather Forecast

The dashboard allows farmers to access these features through a simple and user-friendly interface.

Figure 2: AgriPredict Smart Agriculture Dashboard



8.2 Crop Recommendation System

The Crop Recommendation module analyzes environmental and soil parameters to suggest the most suitable crops for cultivation. Farmers provide input parameters such as:

- Expected Annual Rainfall
- Average Temperature
- Humidity Level
- Soil Type
- Soil pH Level
- Growing Season
- Water Availability

Based on these parameters, the machine learning model evaluates different crops and recommends suitable crops along with expected yield and investment details.

Figure 3: Smart Crop Recommendation Interface

Crop	Expected Yield (kg/ha)	Investment (₹/ha)	Duration (days)	Market Type
Rice	3,857	854,398	120-150	Food Grains
Maize	4,655	217,392	90-120	Food Grains
Wheat	3,165	222,524	-	Food Grains
Sugarcane	71,431	282,475	-	Food Grains

The system generates recommendations such as rice, maize, wheat, and sugarcane based on environmental suitability and predicted productivity.

8.3 Crop Yield Prediction

The Crop Yield Prediction module predicts the expected crop production based on environmental parameters such as rainfall, temperature, pesticide usage, and crop type.

The machine learning model analyzes the input parameters and generates predicted crop yield values. This helps farmers estimate production levels before cultivation.

Figure 4: Crop Yield Prediction Interface

The screenshot displays the 'Crop Yield Prediction' interface on the AgriPredict website. The interface includes a navigation bar with 'Home', 'AI Services', 'Farmer Connect', 'About', and 'English'. The main content area features a form titled 'Crop Yield Prediction' with the instruction 'Enter your farm conditions to get accurate yield predictions'. The form contains four input fields: 'Crop Type' (set to 'Rice'), 'Annual Rainfall (mm)' (set to '1111'), 'Pesticide Usage (kg/ha)' (set to '150'), and 'Average Temperature (°C)' (set to '28'). A green 'Submit' button is located below the form. Below the form, a green box displays the 'Yield Prediction Results': 'Crop: Rice', 'Predicted Yield: 44962.17 hg/ha', and 'Recommendation: Estimated yield for Rice: 44962.17 hg/ha'.

The system displays predicted crop yield results along with recommendation messages to help farmers plan agricultural activities efficiently.

8.4 Disease Risk Prediction

The Disease Prediction module helps farmers identify potential crop diseases based on environmental conditions. The system evaluates parameters such as rainfall, temperature, humidity, and pesticide usage to estimate disease risk levels.

Figure 5: Crop Disease Risk Prediction System

Disease Risk Prediction
Assess disease risk for your crops based on environmental conditions

Crop Type: Rice
Recent Rainfall (mm): 111
Average Temperature (°C): 28
Humidity (%): 82
Pesticide Usage (kg/ha): 0-7 days (Recent)
disease.season_label: Monsoon

Disease Risk Assessment
Risk Level: **Medium Risk**
Recommendation: Monitor crops closely, ensure proper ventilation
Based on current environmental conditions and crop type.

The system classifies disease risk levels into categories such as:

- Low Risk
- Medium Risk
- High Risk

Based on the risk level, the system provides preventive recommendations to protect crops from potential diseases.

8.5 Fertilizer Recommendation System

The Fertilizer Recommendation module suggests appropriate nutrient requirements for crops based on soil conditions and environmental parameters.

The system recommends optimal quantities of essential nutrients such as:

- Nitrogen (N)
- Phosphorus (P)
- Potassium (K)

Figure 6: Fertilizer Recommendation System

Fertilizer Recommendation
Get personalized NPK recommendations for optimal crop nutrition

Crop Type: Rice (Waddy)
Expected Annual Rainfall (mm): 111
Soil Type: Clay Soil
fert.field_size: 1.0
fert.growth_stage: Pre-Sowing/Land Preparation
Soil pH Level: Slightly Acidic (6.0-6.5)

Fertilizer Recommendation for Rice

Nitrogen (N)	Phosphorus (P)	Potassium (K)
29 kg/ha	15 kg/ha	15 kg/ha

Application Schedule
Apply all fertilizers at least 4 days before sowing.

Important Notes:

- These are general recommendations. Always conduct soil testing for precise needs.
- Apply fertilizers based on weather conditions and soil moisture.
- Consider organic alternatives and integrated nutrient management.
- Consult local agricultural experts for region-specific advice.

The module also provides application schedules and guidelines for fertilizer usage to ensure better crop growth and productivity.

8.6 Weather Forecasting Module

Weather conditions significantly affect agricultural productivity. The Weather Forecast module predicts rainfall patterns and environmental conditions based on location data.

The system allows farmers to select their location through a map-based interface and generates weather forecasts including:

- Temperature
- Humidity
- Rainfall predictions
-

Figure 7: Weather Forecast Prediction System

The screenshot displays the AgriPredict web application's weather forecasting interface. At the top, a dark green navigation bar includes the AgriPredict logo and links for Home, AI Services, Farmer Connect, About, and English. The main content area features a 'Weather Forecast' form with the following fields: 'Year' (set to 2026), 'Location' (set to South India), 'weather.season_label' (set to Zaid Season (April-June)), and 'weather.crop_cycle' (set to General Planning). A 'Submit' button is located below the form. The results section, highlighted in green, shows: 'Weather Forecast', 'Year: 2026', 'Location: South India', and 'Predicted Rainfall: 1122.5 mm'. A blue notification box below the results states: 'Normal rainfall expected. Good conditions for most crops.'

This feature helps farmers plan irrigation schedules, crop selection, and planting seasons effectively.

8.7 Location-Based Weather Analysis

The system integrates map-based location selection to provide region-specific agricultural insights. By selecting a specific location, the system retrieves historical climate data and predicts future weather conditions.

Figure 8: Location Selection and Weather Data Analysis

The screenshot displays the AgriPredict web application. At the top, there is a navigation bar with the AgriPredict logo and menu items: Home, AI Services, Farmer Connect, About, and English. Below the navigation bar is a section titled "Select Your Location". It features a search input field with the placeholder text "Search location (city, village, address)...". Below the search field is a map of Hyderabad, India, with a blue location pin placed over the city center. Below the map, the weather data is displayed: "Weather: 29.1 °C | Humidity: Humidity: 95.2 % | Rainfall: Rainfall: 1115.6 mm". A note below the weather data states "Predicted based on historical data for India".

This functionality improves the accuracy of crop recommendations and yield predictions by incorporating geographic environmental data.

9. Advantages of Proposed System

The proposed AgriPredict system provides several advantages compared to traditional agricultural decisionmaking methods.

- Provides accurate crop yield predictions using machine learning models
- Helps farmers select suitable crops based on environmental conditions
- Predicts disease risks to prevent crop damage
- Recommends appropriate fertilizer quantities
- Provides weather forecasting based on location data
- Offers a user-friendly web interface for farmers

These features help improve agricultural productivity and enable farmers to make data-driven decisions.

10. Future Work

The proposed system can be further improved by integrating additional technologies and features. •

Integration of **real-time weather APIs**

- IoT-based **soil moisture and nutrient sensors**
- Mobile application for farmers
- Deep learning models for more accurate predictions
- Integration with **satellite-based agricultural monitoring systems**

These improvements can enhance the accuracy and usability of the system.

11. Conclusion

This research presented a **Machine Learning Based Crop Yield Prediction and Agricultural Advisory System** designed to assist farmers in making informed agricultural decisions. The system integrates machine learning algorithms such as **Linear Regression, Decision Tree, and Random Forest** to analyze agricultural parameters and predict crop yield.

The developed web-based platform provides additional services including crop recommendation, disease prediction, fertilizer recommendation, and weather forecasting. Experimental results indicate that the **Random Forest algorithm provides the highest prediction accuracy.**

The proposed system demonstrates the potential of artificial intelligence and machine learning technologies in transforming traditional agriculture into a **data-driven and intelligent farming system.**

12. References

- [1] L. Breiman, "Random Forests," *Machine Learning Journal*, vol. 45, no. 1, pp. 5–32, 2001.
- [2] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
- [3] R. Sujatha and P. Isakki, "A Study on Crop Yield Prediction using Machine Learning Algorithms," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 3, pp. 1–6, 2020.
- [4] S. Jeong, J. Resop, N. Mueller, and D. Fleisher, "Random Forests for Global and Regional Crop Yield Predictions," *PLoS ONE*, vol. 11, no. 6, 2016.
- [5] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.
- [6] R. Rajeshwari and K. Arunesh, "Prediction of Crop Yield Using Machine Learning Techniques," *IEEE International Conference on Smart Technologies and Management*, pp. 1–6, 2019.
- [7] A. Kamilaris and F. Prenafeta-Boldú, "Deep Learning in Agriculture: A Survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.
- [8] S. Ramesh and N. Vardhan, "Crop Yield Prediction using Machine Learning Techniques," *International Journal of Computer Applications*, vol. 179, no. 7, pp. 10–14, 2018.