International Journal of Advanced Multidisciplinary Research and Educational Development
Volume 2, Issue 2 | March – April 2026 | www.ijamred.com
ISSN: 3107-6513

CROP YEILD PREDICTION

Akash A

ABSTRACT

Agriculture plays a vital role in the economy, and accurate crop yield prediction is essential for improving productivity and ensuring food security. This project focuses on developing a crop yield prediction system using machine learning techniques. The system analyzes various factors such as temperature, rainfall, humidity, soil type, soil nutrients, crop type, and farming practices to estimate the expected yield of crops.

Traditional methods of predicting crop yield are often based on past experiences and limited data, which may lead to inaccurate results. To overcome these limitations, this project utilizes machine learning algorithms such as Linear Regression, Decision Trees, Random Forest, and Support Vector Machine (SVM) to build predictive models. These models are trained using historical agricultural data to identify patterns and relationships between different factors affecting crop production.

The proposed system improves prediction accuracy, supports better decision-making, and helps farmers optimize the use of resources like water, fertilizers, and pesticides. It also reduces risks caused by changing climatic conditions and enhances overall agricultural efficiency. The system can be further extended with real-time data integration and deployed as a web or mobile application for easy access.

INTRODUCTION

Crop yield prediction is an essential part of modern agriculture that helps farmers and agricultural organizations estimate the amount of crop production before harvesting. Accurate prediction plays a crucial role in planning storage, transportation, pricing, and ensuring food security. In earlier times, crop yield was predicted using traditional methods based on past experiences and limited data, which often resulted in less accurate outcomes. With the advancement of technology, machine learning has emerged as a powerful tool to improve prediction accuracy.

Machine learning algorithms analyze large datasets, including weather conditions such as rainfall, temperature, and humidity, soil characteristics, crop type, and other environmental factors, to identify patterns and make reliable predictions. This data-driven approach enables farmers to make better decisions, increase productivity, reduce risks caused by climate change, and efficiently use resources like water and fertilizers. Thus, crop yield prediction using machine learning provides a smart and effective solution for improving agricultural practices.

In addition to improving accuracy, machine learning-based crop yield prediction systems can handle large and complex datasets collected from various sources such as satellite imagery, sensors, and historical agricultural records. These systems use advanced algorithms like regression models, decision trees, and neural networks to learn relationships between different factors affecting crop growth. By continuously learning from new data, these models become more reliable over time and can adapt to changing environmental conditions. This makes them highly useful for predicting yields in different regions and for various types of crops.

EXISTING SYSTEM

The existing system for crop yield prediction mainly relies on traditional methods and manual analysis. Farmers generally depend on their past experience, knowledge, and local environmental conditions to estimate crop production. In some cases, basic statistical methods and historical data are used by agricultural experts to make predictions. However, these methods often lack accuracy as they do not consider real-time data or complex relationships between multiple factors such as weather changes, soil conditions, and crop health.

Moreover, the traditional system is time-consuming and less efficient, as it involves manual data collection and analysis. It also fails to handle large volumes of data and cannot adapt quickly to sudden climate changes or unpredictable environmental conditions. As a result, farmers may face issues like low productivity, improper resource utilization, and financial losses. Therefore, the existing system highlights the need for an advanced, automated, and data- driven approach like machine learning to improve the accuracy and efficiency of crop yield prediction.

DRAWBACKS FOR EXISTING SYSTEM

The existing system for crop yield prediction has several limitations that affect its accuracy and reliability. One of the major drawbacks is its heavy dependence on manual methods and farmers' experience, which may not always provide accurate results. Since these predictions are often based on past data and assumptions, they fail to consider real-time changes in weather, soil conditions, and other environmental factors.

Another significant drawback is the inability to handle large and complex datasets. Traditional methods cannot efficiently process data from multiple sources such as satellite images, sensors, and climate records. This leads to incomplete analysis and less effective decision-making. Additionally, the system is time-consuming and lacks automation, making it difficult to provide quick predictions when needed.

Furthermore, the existing system is not adaptable to sudden climate changes or unexpected situations like droughts or floods. This can result in poor crop planning, low productivity, and financial losses for farmers. It also leads to inefficient use of resources such as water, fertilizers, and pesticides. Overall, these drawbacks highlight the need for a more advanced and intelligent system using machine learning techniques to improve crop yield prediction.

PROPOSED SYSTEM

The proposed system introduces a machine learning-based approach for accurate and efficient crop yield prediction. Unlike the traditional methods, this system uses advanced algorithms to analyze large amounts of data collected from various sources such as weather conditions, soil properties, crop type, and historical yield records. By processing this data, the system can identify patterns and relationships that help in predicting crop yield with higher accuracy.

The proposed system is automated and capable of handling real-time data, making it faster and more reliable. Machine learning models such as regression algorithms, decision trees, and random forests are used to build predictive models that continuously improve as more data is provided. This allows the system to adapt to changing environmental conditions and provide better predictions over time.

Additionally, the system helps farmers make informed decisions regarding crop selection, irrigation, fertilizer usage, and harvesting time. It reduces risks caused by unpredictable weather and improves overall productivity. The proposed system also supports efficient resource management and promotes sustainable agriculture. Overall, this approach provides a smart, data-driven solution to overcome the limitations of the existing system and enhance agricultural outcomes.

ADVANTAGES OF PROPOSED SYSTEM

Improved Accuracy:

Machine learning models such as SVM, Linear Regression, and Random Forest provide higher prediction accuracy compared to traditional methods used in agriculture. Among these, SVM often achieves better performance in analyzing complex relationships between factors like soil, weather, and crop conditions, making it a reliable algorithm for predicting crop yield.

Better Resource Management:

Accurate crop yield predictions help farmers optimize the use of resources such as water, fertilizers, and pesticides, reducing wastage and improving overall farming efficiency.

Scalability and Flexibility:

The proposed system can be applied at different levels, from small farms to large agricultural regions, making it suitable for various farming practices and crop types.

Integration of Environmental Data:

The prediction models effectively utilize environmental data such as rainfall, temperature, humidity, and soil nutrients, which are essential for accurate crop yield forecasting.

SYSTEM SPECIFICATION

HARDWARE REQUIREMENTS:

- **Processor**: Intel Core i3 or above (Preferably multi-core processors for faster model training and data processing)
- **RAM**: Minimum 8 GB (Recommended 16 GB or higher for handling large agricultural datasets efficiently)
- **Storage**: 500 GB HDD or SSD (SSD is recommended for faster data access and improved system performance)

- **Network Connectivity**: High-speed internet connection for cloud-based deployment, real-time data access, and model updates

SOFTWARE REQUIREMENTS:

- **Operating System**: Windows 10/11, Linux (Ubuntu preferred), or macOS
- **Programming Language**: Python (Version 3.7 or above)
- **Development Environment/IDE**: Jupyter Notebook, PyCharm, or Visual Studio Code
- **Libraries & Frameworks**:
- **NumPy** (for numerical computations)
- **Pandas** (for data manipulation and analysis)
- **Matplotlib Seaborn** (for data visualization)
- **Scikit-Learn** (for machine learning models - Linear Regression, Random Forest, SVM)

LANGUAGE SPECIFICATION

OVERVIEW OF WINDOWS 11:

Windows 11 is the latest operating system developed by Microsoft, designed to provide a modern, user-friendly, and efficient computing experience. It features a redesigned interface with a centered Start menu, improved taskbar, and enhanced visual elements that make navigation easier and more intuitive. The operating system focuses on performance, security, and productivity, making it suitable for both personal and professional use.

Windows 11 offers improved support for multitasking through features like Snap Layouts and virtual desktops, allowing users to organize multiple applications efficiently. It also includes better integration with cloud services and supports a wide range of software and development tools, which are essential for projects like machine learning and data analysis. The operating system is optimized for speed and provides better memory management, ensuring smooth performance even when handling large datasets.

OVERVIEW OF BACK-END TOOL:

The back-end tool in crop yield prediction using machine learning algorithms serves as the foundation for data processing, model training, and predicting crop yield . Key features of the back-end tool may include:

Data Preprocessing: The back-end system uses libraries such as Pandas and NumPy for cleaning, transforming, and feature engineering of raw agricultural data. This includes data from sources like weather stations, soil sensors, historical crop records, and rainfall datasets. Proper preprocessing ensures that the data is in the right format for efficient analysis and modeling.

Model Development and Training: Frameworks like Scikit-learn, TensorFlow, or PyTorch are used to develop and train predictive models. This phase involves building models using algorithms such as Linear Regression, Decision Trees, Random Forest, and SVM, fine-tuning hyperparameters, and performing cross-validation to ensure accurate and robust crop yield predictions.

DATA PREPROCESSING:

The collected data is preprocessed to prepare it for training the crop yield prediction . This includes steps such as splitting the dataset into training and testing sets, and formatting the data into appropriate sequences suitable for input into random forest regressor algorithms.

DATA SPLITTING:

The dataset was split into training and testing sets using a random sampling approach,
with 80% of the data allocated for training and 20% for testing. This ensured that the model was trained on a sufficient portion of the data while retaining a separate subset for evaluation.

ALGORITHM SELECTION

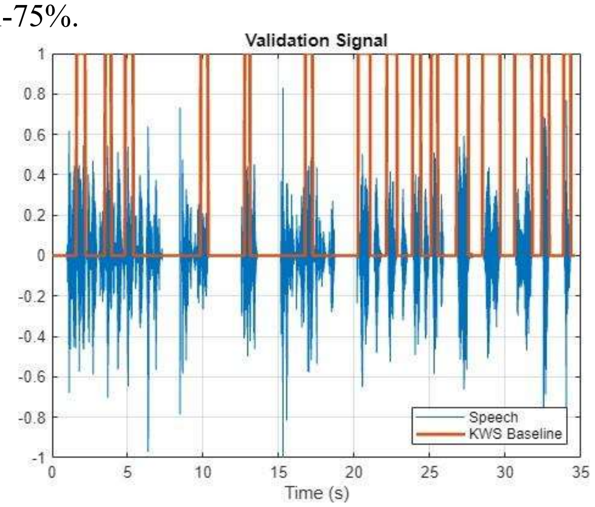
- Random Forest Regressor
- Long Short-Term Memory

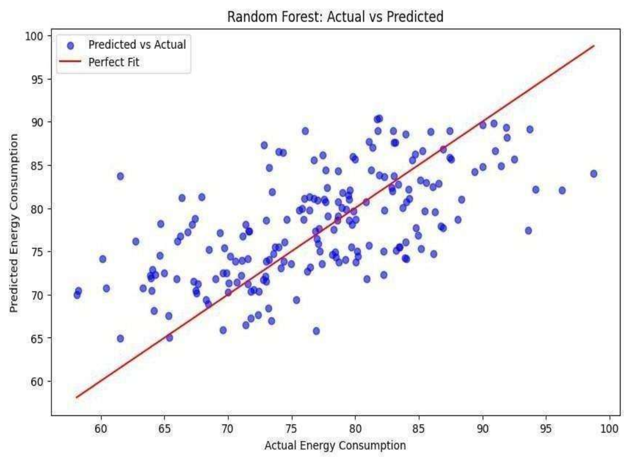
Long Short-Term Memory:

Long Short-Term Memory (LSTM) is an advanced supervised deep learning algorithm used for predictive modelling, especially with sequential or time-series data. It is a type of Recurrent Neural Network (RNN) designed to capture long-term dependencies by maintaining a memory of past information. The algorithm works through a series of memory cells and gates—namely the forget gate, input gate, and output gate—which regulate the flow of information and decide what to retain or discard over time. Unlike traditional models, LSTM does not assume a simple linear relationship but instead learns complex temporal patterns in the data. This makes it highly suitable for tasks where historical context plays a crucial role, such as crop yield prediction, weather forecasting, and stock analysis. The model processes input sequences and produces predictions based on learned patterns across time steps. In the research, the LSTM algorithm demonstrated an accuracy of 79%, precision of 79%, F1 score of 76%, and recall of 74%.

RANDOM FOREST REGRESSION:

Random Forest Regression is a powerful ensemble learning method used for energy consumption prediction. It builds multiple decision trees on different subsets of the data and averages their predictions to improve accuracy and reduce overfitting. Unlike linear regression, it can capture nonlinear relationships between energy consumption and factors like temperature, humidity, and appliance usage. In the research the Random Forest algorithm demonstrated an accuracy -80%, precision-74%, F1 Score -74%, Recall-75%.





REFERENCES

[1] Ramesh Medar, Vijay S Rajpurohit and Shweta S, "Crop Yield Prediction using Machine Learning Techniques," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 6, pp. 1425–1428, 2019.

[2] K. R. Patil and S. S. Thorat, "Crop Yield Prediction using Data Mining Techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, 2016.

[3] T. M. Mitchell, Machine Learning, McGraw-Hill, 1997.