International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 336-340 March 2026-Sept 2026 www.csjournals.com

# Smart Agriculture: Increasing Yield and Reducing Resource Usage Using Machine Learning

Akta rani[1],prof. Vishal verma[2], Bajrang[3]
[1]MCA Student, [2]Professor, [3]Asst. Prof.
Deptt. of Comp. Sc. and Application, Ch. Ranbir Singh University, Jind, Haryana, India

## ABSTRACT

Agriculture plays a important role to ensure the food security and support economic of any country. On the other hand, increasing crop productivity while minimizing the resources that we use in our crop production remains a significant difficulty. Nowadays development in ML provide effective solutions for inspecting agricultural data and improving crop production prediction. This research based on recently build model presents a Smart cultivation system made by using machine learning techniques for crop production prediction and resource improvement. A Kaggle agricultural dataset containing 50,000 records or 16 features is use which include many features like crop type, season, soil type, irrigation type, rainfall, temperature, humidity, soil moisture, soil pH, fertilizer usage, water requirement, crop stress, and yield. The dataset was preprocessed by using exploratory data analysis methods to improve data quality. Four ML techniques, namely LR**,** DTR, RF, and XGBoost, were implemented and evaluated using $R^2$, MAE, and RMSE. The results based on the project showed that the Random Forest Regression achieved the highest prediction accuracy with lower error values compared to the other models that uses in this project. This system successfully identified the relationship between environmental conditions, resource utilization, and crop productivity, enabling reliable crop yield guesses. The study based on this project states that ML can support data-driven agricultural decision-making, improve farming efficiency, reducing unnecessary use of resources, and promote sustainable agricultural practices.

**Keywords:** Smart Agriculture, Machine Learning, Crop Yield Prediction, Random Forest, XGBoost, Resource Improvement, Sustainable Farming.

## 1. INTRODUCTION

Agriculture is one of the oldest and most important human activities in the world. It refers to the practice of cultivating crops and rearing animal to produce food, fiber and other useful products. Agriculture helps us to understand how early human shifted from hunting and gathering to settled farming life, leading to the growth of civilization.

Agriculture plays a important role in providing food security, employment and raw materials for industry. In simple words, Agriculture is the science and art of cultivating soil, growing crops and raising livestock for human use.

Classical farming methods mainly rely on farmers experience and guide decision-making, which not always provide accurate results. With the increasing demand of food and limited availability of water and other agricultural resources, there is a growing need for intelligent and data-driven farming solutions.

The development of modern technologies impacts all human activity, including the agricultural sector. Smart farming is a relatively new concept of advanced agriculture management. Smart farming in agriculture involves using various software and equipment to optimize and automate routine processes. With the spread of the Internet and mobile devices, such technologies have become available not only for large companies but also for small private farms. It helps growers optimize production and compete with larger agro companies. Since smart agriculture practices have undeniable advantages for businesses over conventional methods, it is worth exploring the features of this approach in more depth.

The term smart agriculture refers to the usage of technologies like Internet of Things, sensors, location systems, robots, machine learning and artificial intelligence on your farm. The ultimate goal is increasing the quality and quantity of the crops while optimizing the human labor used.

In this research, a Smart Agriculture system which depends on machine learning is build for crop production prediction and resource usage improvement. The system uses a Kaggle dataset containing 50,000 agricultural records with 16 important features which include crop type, season, soil type, irrigation type, rainfall, temperature, humidity, soil moisture, soil pH, fertilizer usage, water

requirement, crop stress, and crop yield. Four machine learning algorithms LR, DTR, RF, and XGBoost were implemented and compared using $R^2$ Score, MAE, and RMSE. The project results shows that the RF achieved the best prediction performance and provided reliable crop yield estimation.

The main motive of this research is to build a predictive model that helps farmers improve crop productivity while promoting logical Use of agricultural resources. The proposed system contributes to sustainable farming by enabling data-driven agricultural decision-making.

## 2. REVIEW OF LITERATURE

Jeong et al. proposed the use of the RF algorithm for global and regional crop yield prediction. Their study demonstrated that Random Forest effectively handled complex agricultural datasets and produced accurate yield predictions by analyzing environmental and climatic factors.

Khanal et al. shows detailed review of machine learning applications in agriculture. The authors highlighted that machine learning techniques are widely used for crop yield prediction, disease detection, irrigation management, and precision farming, helping farmers improve productivity and decision- making.

Chlingaryan et al. reviewed various machine learning approaches for crop yield prediction and nitrogen status estimation. Their study concluded that ensemble learning algorithms generally provide better prediction performance than conventional statistical methods because they can capture nonlinear relationships among agricultural variables.

Shahhosseini et al. integrated crop simulation models with machine learning techniques to improve crop yield prediction. Their results demonstrated that combining agricultural simulation data with machine learning significantly enhanced prediction accuracy under different environmental conditions.

Dhillon et al. investigated the integration of Random Forest with crop modeling techniques.

The study reported that ensemble learning methods improve prediction reliability and effectively analyze the influence of multiple agricultural parameters on crop productivity.

Although previous studies achieved promising results in crop yield prediction, most of them primarily focused on prediction accuracy and paid limited attention to efficient resource utilization. In this research, a machine learning- based Smart Agriculture System is proposed that not only predict the crop yield but also considers important resource-related factors such as irrigation type,

fertilizer usage, soil moisture, and water requirement. Many machine learning models were compared, and the Random Forest Regression was identified as the best-performing model, making the proposed approach suitable for supporting sustainable and resource-efficient farming practices.

## 3. PROPOSED METHODOLOGY

This study build a ML-based Smart Agriculture System for crop production prediction and resource optimization. The objective of the proposed methodology is to analyze agricultural and environmental data to accurately predict crop yield while supporting efficient utilization of farming resources such as water and fertilizers. The entire methodology contains dataset collection, EDA, feature selection, model training, performance analysis, and crop yield prediction.

The agriculture dataset used in this research was collected from Kaggle and contains 50,000 rows with 16 columns which highly relate to crop production. The dataset contains important features such as Crop Type, Season, State, Soil Type, Irrigation Type, Growth Stage, Rainfall, Air Temperature, Soil Temperature, Humidity, Soil Moisture, Soil pH, Fertilizer Usage, Water Required, Crop Stress, and Yield (q/ha). These features were selected because they directly influence crop productivity and resource utilization.

Before model development, the data is processed to improve data quality and model accuracy. Data preprocessing involved checking missing values, removing duplicate records, label encoding of categorical attributes, feature scaling, and divide the dataset into training and testing sets. These preprocessing steps ensured that the machine learning models were trained on clean and consistent data.

Four machine learning algorithms were implemented and compared for crop yield prediction LR, DTR, RF and XGBoost. LR is use as a base model, while Decision Tree captured nonlinear relationships among agricultural variables. RF and XGBoost, being ensemble learning algorithms, were implemented to improve prediction accuracy and reduce over-fitting.

The performance of all models is analysis using $R^2$, MAE, and RMSE. Based on experimental performance, the RF achieved the highest prediction accuracy and the lowest error values, making it the most suitable model for the proposed Smart Agriculture System.

## 4. RESULTS AND DISCUSSION

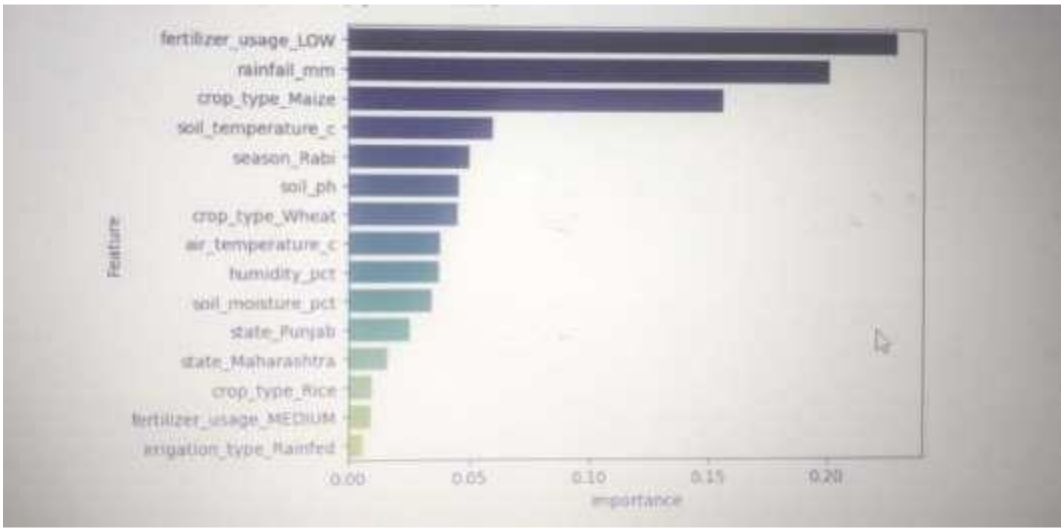
### Experiment Setup

This proposed machine learning models were implemented using Python in Kaggle Notebook. The agricultural dataset is process and divided into training and testing sets using an 80:20 ratio. Four regression algorithms, namely Linear Regression, DTR, RF, and XGBoost Gradient boosting, is trained and analysed using $R^2$ , MAE and RMSE. The Random Forest model achieved the best prediction accuracy among all implemented algorithms.

### Performance Comparison

| | Model | MAE | RMSE | R2 Score | Accuracy (%) |
|---|---|---|---|---|---|
| 2 | Random Forest | 4.782800 | 6.097185 | 0.693385 | 69.338482 |
| 1 | Decision Tree | 5.173827 | 6.621070 | 0.638431 | 63.843093 |
| 0 | Linear Regression | 5.439463 | 6.764327 | 0.622615 | 62.261541 |

### Feature Importance




### Discussion

| MODEL | STRENGTH | LIMITATION | PERFORMANCE |
|---|---|---|---|
| Linear Regression | Simple and fast | Cannot handle nonlinear data | Moderate |
| Decision Tree | **Easy to interpret** | Overfitting | Good |
| Random Forest | Robust, stable, less overfitting | Slightly higher training time | Best |
| XGBoost | High accuracy | More computationally intensive | Very good |

## 5. CONCLUSION

This research presented a ML-based Smart Agriculture System for crop yield prediction and resource optimization. The proposed approach was developed to support farmers in making accurate agricultural decisions by analyzing environmental, soil, and crop-related parameters. A Kaggle agricultural dataset containing 50,000 records and 16 features was used to train and evaluate different machine learning models.

Four regression algorithms, namely LR, DTR, RF, and XGBoost were implemented and compared using $R^2$, MAE, and Root RMSE**.** The experimental evaluation showed that the RF perform the other models by providing higher prediction accuracy and lower error values. The model effectively captured the relationship between agricultural factors such as rainfall, soil moisture, temperature, irrigation type, fertilizer usage, and crop yield.

The obtained results demonstrate that ML can importantly improve crop yield prediction and keep efficient utilization of agricultural resources. By providing reliable yield estimates, the proposed system can assist farmers in planning irrigation, fertilizer application, and crop management more effectively. This contributes to higher agriculture production, reduce resource wastage, and more renewable farming methods.

All in all, the suggested Smart Agriculture framework gives a practical and dependable solution for data-driven crop production prediction. The study confirms that random forest is the best model for the selected agriculture dataset and can serve as an effective decision-support tool for improving farming productivity and promoting renewable cultivation.

## 6. REFERENCES

[1]. J. H. Jeong *et al.*, "Random forests for global and regional crop yield prediction," *Agriculture and Forest Meterology* vol. 223, pp. 71–83, 2016.

[2]. S. Khanal *et al.*, "Machine learning in agriculture: A review," *Computers and Electronics in Agriculture*, vol. 170, p. 105–120, 2020.

[3].A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Computers and Electronics in Agriculture*, vol. 151, pp. 61–69, 2018.

[4]. A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.

[5]. K. G. Liakos *et al.*, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, p. 2674, 2018.

[6]. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[7]. T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*, Lecture Notes in Computer Science, vol. 1857, pp. 1–15, 2000.