AI Based Agriculture Optimization

Dr. Rachana Dhannawat[1]; Smruti Sawant[2]; Ritika Sawant[3]; Shivanshi[4]

[1,2,3,4]Department of Computer Engineering, Usha Mittal Institute of Technology, Mumbai, India

Publication Date: 2026/05/18

Abstract: Agriculture remains central to food security and rural livelihoods, yet farmers continue to face persistent challenges due to environmental variability, fluctuating market conditions, and limited access to data-driven decision-support tools. This paper presents AgriSmart, an integrated AI-driven framework designed to support modern farming through three key functional modules: crop harvest optimization, intelligent crop recommendation, and flood risk early warning. The harvest optimization module utilizes hybrid machine learning models to forecast optimal harvest timing using agricultural market data from Pune spanning 2020–2025. The dataset includes price information for 22 crops: onion, potato, rice, maize, wheat, black gram (whole), cabbage, grapes, cauliflower, watermelon, orange, papaya, beetroot, garlic, carrot, spinach, green peas, brinjal, bottle gourd, ridge gourd, and tomato.The crop recommendation module evaluates multiple machine learning classifiers, including Random Forest, XGBoost, CatBoost, Support Vector Machine (SVM), and Gaussian Naive Bayes, to recommend suitable crops based on soil and environmental attributes. The Early Flood Alert Module operates as a Multi-Class Classification System using an XGBoost algorithm able to process multiple variables rainfall, humidity, soil type, terrain slope and crop growth stage to predict the risk of waterlogging for a farmer and generate advance notice of this risk to the farmer.

Keywords: *Artificial Intelligence, Crop Harvest Optimization, Crop Recommendation, Flood Risk Prediction, Machine Learning, Sustainable Agriculture, Precision Agriculture, XGBoost, Random Forest, Land Suitability Rating.*

How to Cite: Dr. Rachana Dhannawat; Smruti Sawant; Ritika Sawant; Shivanshi (2026) AI Based Agriculture Optimization. *International Journal of Innovative Science and Research Technology*, 11(4), 4817-4822.
https://doi.org/10.38124/ijisrt/26apr2478

I. INTRODUCTION

Agriculture is much more than merely being a business; it has become a part of everyone's culture and an important support structure for food security and also acts as the invisible engine of the rural economy. This is particularly true in a developing country like India that has either directly or indirectly affected almost every single household through farming, and therefore the strength of the agricultural sector greatly affects the strength of the economy as a country. Farming is about feeding people; it also provides farmers with the means to make a living and provides manufacturers with materials to produce goods. Farming is a significant contributor to gross domestic product (GDP); however, currently the people growing food face pressures that previous generations of farmers never had to contend with - such as unpredictable rainfall, decreased groundwater supply, over-exploitation of the land and/or natural resources, increasing production costs and markets that are highly variable from season-to-season.

Farmers have relied on legacy knowledge, practical experience and their own community for generations to deal with such uncertainties. There is value in all aspects of this knowledge as they are all based on real life and earned through many years of experience. This approach works in relatively consistent conditions; therefore, as climate variability increases and market volatility increases, there are times when even the best, most experienced farmer will find that many of their instinctive decisions have not been enough to make their farming practice successful. Choosing the wrong crop, miscalculating soil quality or waiting too long to harvest creates both an opportunity for profit and will cause you to lose all the money you've invested into your farm this year. A great number of farmers face this challenge on a daily basis and all of them could lose everything.

There are a large number of components that are involved in making good decisions in the agricultural industry, having said this the complexities of the agricultural industries are great because of all the factors that are considered at any one time. A farmer wanting to make a decision about what crop to plant must factor in soil nitrogen, phosphorus, potassium and acidity levels, local humidity and temperature, forecasted precipitation, irrigation access, crop growth timelines, risk from pests, and anticipated prices by the harvest date [1] . All of these factors do not exist in a vacuum, rather they are interrelated and impact each other in many ways that basic spreadsheets and simple rules cannot capture. All of this is further complicated by the constant threat of natural disasters. A single flood can quickly destroy the entire existing crop, and with it the nutrients within the topsoil, cause a delay in

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr2478

planting during the next cycle, and make the land unsuitable for growing crops for several subsequent growing seasons. Therefore, to develop a system that truly supports farmers, all of this must be addressed as an integrated entity.

We now see that technological advances, such as artificial intelligence and machine learning, can quickly create a positive impact through their ability to filter through large amounts of past data within their respective fields and generate accurate forecast models for activities which may not have been possible for any human analyst to accomplish manually before now.

Technologies such as random forest and gradient boosting have shown to be very effective at finding those relationships between data sets in agricultural settings that do not have the linear characteristics associated with "normal" data. Most of the existing agricultural solutions are very narrow in their application capabilities and provide little to no guidance to farmers in making decisions relative to growing crops. For example, there are several models used in agriculture to predict when crops should be planted. However, these models do not consider whether or not the crop will produce enough profit for farmers to plant and expand. However, this model does not take into account if this crop will be profitable for the farmer to grow and harvest. Alternatively, there are models whose yield predictions are accurate but do not indicate if current growing conditions will allow for the production of a crop with that expected yield. Another model may provide accurate forecasts of crop yields, but does not include an assessment of whether current growing conditions are conducive to producing a viable crop from that yield. The same can be said for flood warning systems which often rely on fixed thresholds for rainfall measurements and do not adjust the thresholds that trigger a flood warning as weather patterns and amounts of rainfall change seasonally.

This research grew out of a desire to connect these gaps. We're proposing an Integrated AI-Driven Agricultural Decision-Support System that combines 3 interdependent pillars of smart farming: (1) Harvest Optimization, (2) Crop Recommendation, and (3) Flood Risk Early Warning, which we are viewing as separate problems; in fact they are different aspects of the same problem every farmer faces every day as they wake up and try to figure out if this season will be successful.

The intent is not to merely develop an improved tool, but to provide very intelligent tools to those people that need them. Our aim is to assist farmers in developing sustainable methods of growing their crops, managing their risk with confidence, and creating greater resilience for themselves, their families and the communities that rely upon them.

II. RELATED WORK

Optimizing Agricultural Yield via Market Dynamics (Sable et al., 2025) [2] study focuses on forecasting crop prices and identifying the most profitable harvest months to assist farmers in optimizing selling time and maximizing revenue. Several single machine learning models, including Decision Tree, Random Forest, XGBoost, CatBoost, and Support Vector Machine (SVM) were implemented. The system also incorporated a Streamlit-based visualization interface to present predictions and insights in an accessible manner. Among the evaluated models, the Decision Tree model achieved the highest accuracy of 96.76%. . Dey et al. [3] proposed a crop recommendation system using the Ingle dataset (22 crops, NPK/pH/rainfall), achieving 98.51% accuracy with XGBoost but limited by lack of profitability analysis or hybrid rules. Costache et al. [4] and Wang et al. [5] demonstrated the effectiveness of ML-based flood susceptibility mapping. A key gap across existing literature is the absence of a unified system integrating yield optimization, agronomically validated crop recommendation, and real-time flood risk assessment into a single framework.

III. PROPOSED SYSTEM

➢ *Research Design*

AgriSmart follows a quantitative and experimental design employing a data-driven predictive modeling approach across three interconnected modules. The unified workflow encompasses data acquisition, preprocessing, feature engineering, model selection, hybrid model development.

➢ *Data Sources*

Each module draws from distinct datasets tailored to its objective. The Harvest Optimization Module uses agricultural market data from the Government of India opendata portal (data.gov.in), covering the Pune region (2020– 2025) for 22 crops. The Crop Recommendation Module uses the publicly available Crop Recommendation Dataset [6], comprising 2,200 samples across 22 crop classes described by seven features: Nitrogen (N), Phosphorus (P), Potassium (K), pH, temperature, humidity, and rainfall. The Flood Early Warning Module uses a curated dataset from environment and forest data from the Government of India open-data portal (data.gov.in)representing flood risk scenarios across five severity levels: no risk, low risk, moderate risk, high risk and severe risk, with features including short-term rainfall, cumulative rainfall, humidity, soil type, land slope, crop type, and crop growth stage.

➢ *Data Preprocessing*

Data preprocessing in the Harvest Optimization Module focused on preparing the historical market dataset for accurate regression modeling while preserving seasonal trends and market behavior patterns.

The original date attribute was decomposed into separate *month* and *year* components. Agricultural markets are strongly influenced by seasonal cycles driven by sowing periods, harvesting windows, storage patterns, and supply-demand dynamics. Extracting the month feature enabled the model to capture recurring seasonal price fluctuations, while the year component helped in identifying long-term trends

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr2478

and interannual variations across the 2020–2025 dataset. This temporal transformation allowed the hybrid models to learn periodic patterns more effectively than using raw date values.

Missing values and duplicate records were also examined. Any incomplete entries were handled using appropriate numerical computation techniques to maintain dataset continuity. Duplicate observations were removed to avoid biased learning toward specific temporal instances.

These preprocessing steps ensured cleaner, seasonally structured, and statistically stable inputs for the hybrid ensemble models, ultimately contributing to improved predictive accuracy and reliable harvest timing recommendations.

The Crop Recommendation dataset was partitioned into features (environmental variables: nitrogen, phosphorus, potassium, temperature, humidity, pH, rainfall) and target (crop labels). Categorical crop labels were encoded into numerical values to enable processing by machine learning algorithms.

An 80/20 stratified split was applied, with 80% allocated for training and 20% reserved as a hold-out set for unbiased evaluation on unseen data.

➢ *Machine Learning Models*

- *Harvest Optimization:*
  Three tree-based base models were selected: Decision Tree (DT), Random Forest (RF), and Gradient Boosting (GB). To improve prediction accuracy, generalization capability, and overall stability, hybrid models were proposed using an ensemble averaging [7] strategy, where predictions from two base models were combined to produce a unified output.

Three hybrid combinations were evaluated: DT+RF, RF+GB, and DT+GB. Datasets were split using an 80/20 ratio for training and testing to ensure unbiased performance evaluation.

- *Crop Recommendation:*
  Five classifiers were evaluated: Gaussian Naive Bayes (GNB), Random Forest, XGBoost, CatBoost, and Support Vector Machine (SVM) [8]. This enables users to compare predictions and identify consensus recommendations (such as crops endorsed by the majority), facilitating informed decision-making tailored to their preferences. [9]

- *Flood Risk Classification:*
  Five algorithms were evaluated: Logistic Regression (LR), K-Nearest Neighbors (KNN), SVM, Random Forest, and XGBoost [10]. XGBoost was selected for its high predictive accuracy and robustness to overfitting in nonlinear environmental data. Rule-based agronomic filters suppress flood warnings when short-term and cumulative rainfall values fall below predefined negligible thresholds, preventing false alarms during dry periods [11].

IV. RESULTS AND DISCUSSION

This section presents a comprehensive evaluation of the three modules comprising the proposed AgriSmart system. Each module was independently validated using appropriate performance metrics.

➢ *Harvest Optimization Module*
The performance of the harvest optimization module was evaluated to assess the predictive capability and stability of different machine learning models. Three statistical metrics were used for evaluation:

- *R^2 Score (Coefficient of Determination):*
  This metric represents the proportion of variance in crop prices explained by the model. Values closer to 1 indicate better predictive capability.

- *RMSE (Root Mean Square Error):*
  This metric measures the average magnitude of prediction error in price units. Lower RMSE values indicate predictions closer to actual market prices.

- *Standard Deviation (Variance):*
  This measures the variation of model performance. Lower values indicate greater stability and consistency of the model.

Table 1 Overall System Performance Of Harvest Prediction Models

Model Strategy	**R2**	**RMSE (Bias)**	**Std (Variance)**
Decision Tree	0.9741	86.60	0.0789
Random Forest	0.9861	81.85	0.0276
Gradient Boosting	0.9869	77.51	0.0260
Hybrid (DT + RF)	0.9859	76.72	0.0320
Hybrid (RF + GB)	0.9881	74.08	0.0246
Hybrid (DT + GB)	0.9866	71.92	0.0341

As shown in Table 1, the Hybrid (Random Forest + Gradient Boosting) model achieved the highest overall predictive performance with a R^2 score of 0.9881 and the lowest variance of 0.0246, indicating both high accuracy and strong stability.

Although the Hybrid (Decision Tree + Gradient Boosting) model produced the lowest RMSE value of 71.92, its higher variance indicates slightly less stable performance. In contrast, the Random Forest and Gradient Boosting hybrid combination provides a more balanced

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr2478

trade-off between bias and variance. The superior performance of the hybrid models can be attributed to the complementary strengths of the constituent algorithms. Gradient Boosting sequentially corrects residual errors to reduce bias, while Random Forest uses bagging and ensemble averaging to reduce variance. Combining these techniques produces a robust predictive model capable of handling fluctuations in agricultural market prices.

Overall, the results demonstrate that hybrid ensemble models significantly improve prediction reliability for harvest optimization, enabling the AgriSmart system to generate more accurate and stable recommendations for farmers.

➢ *Crop Recommendation Module*

A multi-model decision support framework is provided through the crop recommendation module with various validated classifiers selected from an integrated dropdown menu for users to choose between them. The comparative performance of the algorithms was evaluated using a dataset consisting of 2,200 samples from 22 crop classes: rice, maize, cotton, jute, chick- peas, lentils, pigeonpeas, mungbean, blackgram, kidneybeans, mothbeans, coconut, coffee, papaya, mango, banana, pomegranate, orange, grapes, watermelon, muskmelon, and apple.

All of the models in the system examined demonstrated exceptionally accurate predictions, all were greater than 96%.

Table 2 Crop Recommendation: Model Performance Comparison

Algorithm	**Accuracy (%)**	**Precision (%)**	**F1-Score (%)**
Gaussian Naive Bayes	99.55	99.58	99.54
Random Forest	99.32	99.37	99.32
XGBoost	98.64	98.69	98.63
CatBoost	98.18	98.27	98.18
SVM	96.14	96.73	96.12

The highest performance was reached by Gaussian Naive Bayes (GNB) with an F1-Score of 99.54%, denoting a very close to perfect accuracy,precision and F1 score balance among these three measures.

Probabilistic methods such as GNB closely model the average nutrient allocation for each crop, while both tree-based and boundary-optimised techniques present a good alternative to determine specific environmental thresholds.

In order to confirm that the recommendations are reliable, the module was tested using the following performance metrics:

- *Accuracy:*

It measures the overall proportion of correct predictions made by the model out of all predictions. It reflects how effectively the model correctly classifies instances across all categories.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Where *TP* represents True Positives, *TN* represents True Negatives, *FP* represents False Positives, and *FN* represents False Negatives.

- *Precision:*

It evaluates the proportion of correctly predicted positive instances among all instances predicted as positive. It indicates the model’s ability to avoid false positive predictions.A higher precision value indicates that the model produces fewer incorrect positive predictions.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

- *F1-Score:*

It is the harmonic mean of Precision and Recall and provides a balanced measure of a model’s performance, particularly when dealing with imbalanced datasets.It combines both precision and recall into a single metric to evaluate the overall effectiveness of the classification model.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

Overall, these evaluation metrics provide a comprehensive assessment of the model’s classification performance, enabling effective analysis of its predictive capability and reliability across different categories.

➢ *Flood Early Warning Module*

Table 3 compares the five classifiers evaluated for multiclass flood risk prediction across five severity levels(No risk, Low, Moderate, High, Severe). These levels represent increasing degrees of flood and waterlogging threat based on rainfall intensity, humidity, soil characteristics, land slope, crop type, and crop growth stage. No risk indicates negligible flooding probability, while Low and Moderate reflect gradually increasing water accumulation that may begin affecting crops. High denotes serious waterlogging conditions with likely agricultural impact, and Severe represents extreme flood scenarios with significant risk to crops and infrastructure.

Table 3 Flood Early Warning: Model Performance Comparison

Model	Accuracy	Precision	F1-Score
Logistic Regression	0.72	0.70	0.69
KNN	0.75	0.73	0.72
SVM	0.81	0.80	0.79
Random Forest	0.87	0.86	0.85
XGBoost	0.91	0.90	0.89

XGBoost achieved the highest performance across all evaluation metrics, with 91% accuracy and an F1-Score of 0.89. The improvement over linear classifiers such as Logistic Regression demonstrates the presence of strong nonlinear interactions among environmental variables.

➢ *Comparative Analysis*

The AgriSmart system has shown consistently strong predictive ability. The Harvest Optimization Module achieved an R^2 score of 0.9881%, for its hybrid (Random Forest and Gradient Boosting) model indicating that it is highly capable of accurately predicting future market windows, based on historical trends. The Crop Recommendation Module achieved 99% classification accuracy. The Flood Early Warning Module achieved 91% multi-class classification accuracy; hence, it will help in predicting and mitigating environmental risk using adaptive and context-aware methods.

Hybrid regression models reduce variance and improve harvest prediction stability. Hybrid Regression Models effectively balanced the bias–variance trade-off, with the Random Forest + Gradient Boosting combination reducing variance to a stable 0.0246.Probabilistic classifiers effectively model nutrient distributions in crop recommendation. Gradient boosting demonstrates superior capability in capturing nonlinear hydrological relationships.

As a whole, the AgriSmart framework is shown to effectively combine predictive analytics, domain-specific validation and environmental risk assessment for a singular decisionsupport system. This approach enables the simultaneous pursuit of economic optimization, agronomic compatibility and environmental resilience in order to resolve the issue of fragmentation seen within the existing AI solutions currently developed for use in agriculture. The empirical evidence demonstrates that the AgriSmart framework provides a basis for supporting informed and data-driven decisions within agricultural operations that vary greatly by their operational characteristics.

V. CONCLUSION AND FUTURE SCOPE

This paper outlines AgriSmart, an intelligent integrated decision support system that aims to solve some of today's most important issues within agriculture, using a data-driven approach. The AgriSmart system incorporates artificial intelligence (AI) and machine learning along with relevant agriculture-specific expert knowledge to help facilitate informed decision-making. The analytical framework presented here illustrates how using long-term data about historical market sales of crops, soil nutrient parameters, weather and climate conditions, and other significant environmental attributes can be leveraged to create accurate predictive models of future agricultural success. The results of the experimental analyses confirm that ensembling techniques and the use of probabilistic models are effective at modeling the complex interactions that occur within agronomics, including nonlinear relationships between environmental factors, market and economic forces, and various events that lead to variations in yield across crops. AgriSmart also demonstrates that AI technologies can be systematically applied to increase productivity and profitability of the agricultural sector as well as enhance resiliency to climate change through improved operational decision-making by farmers.

➢ *Future Scope*

AgriSmart's innovative solutions enable real-time, accurate data to support sustainable (and thus resilient) agricultural development through a cloud-based platform with best practices integrated. This implementation has been demonstrated through the research to provide farmers with better decision making capabilities than they previously possessed by using AgriSmart's technology.As demonstrated in the research, the implementation of AgriSmart supports sustainable and resilient agricultural development. The research outcomes produced multiple opportunities for AgriSmart to expand its options for enhancing capabilities and scalability. Potential enhancements to AgriSmart's capabilities are the development of multiple interfaces with other real-time data acquisition and automation solutions (e.g., Integrated through IoT enabled soil sensors), improved temporal accuracy, and reduced reliance on manual input using continuous and autonomous data streaming.

Incorporating advanced deep learning architectures into AgriSmart (i.e., long-term temporal modeling) can improve forecasting accuracy (i.e., by improving forecasts of seasonal pricing trends and rainfall variability). Additionally, deploying AgriSmart utilizing both a cloud-native and edge-computing architecture can create a scalable, regionally focused advisory solution.

Integrating large-scale governmental agricultural databases (e.g., United States Department of Agriculture), and other meteorological services can enable AgriSmart to create a validated nationwide advisory solution and produce data that is more accurate. Further, using explainable AI can provide additional transparency in AgriSmart and create user confidence in the decision support resources provided to create greater adoption by nontechnical stakeholders.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr2478

As AgriSmart continues to evolve, it will serve as an opportunity for building an intelligent, resilient, and sustainable future agricultural ecosystem.

ACKNOWLEDGMENT

We express our appreciation to A. Ingle for making the Crop Recommendation Dataset publicly available [6], and to data.gov.in for providing access to relevant public datasets that supported this research.

REFERENCES

[1]. S. C. Liu *et al.*, "A crop harvest time prediction model for better sustainability integrating feature selection and artificial intelligence methods," *Sustainability*, vol. 14, no. 21, Art. no. 14101, 2022.
[2]. N. P. Sable, V. K. Shukla, P. N. Mahalle, and V. Khedkar, "Optimizing agricultural yield: A predictive model for profitable crop harvesting based on market dynamics," *Frontiers in Computer Science*, vol. 7, Art. no. 1567333, 2025, doi: 10.3389/fcomp.2025.1567333.
[3]. B. Dey, J. Ferdous, and R. Ahmed, "Machine learning based recommendation of agricultural and horticultural crop cultivation," *Heliyon*, vol. 10, no. 3, Art. no. e25112, 2024, doi: 10.1016/j.heliyon.2024.e25112.
[4]. R. Costache *et al.*, "Flash-flood susceptibility assessment using multicriteria decision making and machine learning supported by remote sensing and GIS techniques," *Environ. Model. Softw.*, vol. 160, Art. no. 105583, 2023, doi: 10.1016/j.envsoft.2022.105583.
[5]. Z. Wang *et al.*, "Flood susceptibility mapping using ensemble machine learning models based on a novel PCA-weighting method," *Sci. Total Environ.*, vol. 886, Art. no. 163925, 2023, doi: 10.1016/j.scitotenv.2023.163925.
[6]. A. Ingle, "Crop recommendation dataset,"Kaggle, Dec. 2020.
[7]. [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/croprecommendation-dataset
[8]. A.F. Jadama, B. Jobarteh, M. M. Islam, and M. K. Toray, "Ensemble learning: Methods, techniques, application," University of New Brunswick, Fredericton, NB, Canada, 2024. doi:
[9]. 10.13140/RG.2.2.28017.08802.
[10]. M. K. Senapaty, A. Ray, and N. Padhy, "A decision support system for crop recommendation using machine learning classification algorithms," *Agriculture*, vol. 14, no. 8, Art. no. 1256, 2024, doi: 10.3390/agriculture14081256.
[11]. S. Pariselvam, C. Loganath, M. Kiran Kumar, and M. Aravind, "AI based crop recommendation and finding the profitable region to sell the crop," *African Journal of Biomedical Research*, vol. 27, no. 4S, pp. 15109– 15114, 2024, doi: 10.53555/AJBR.v27i4S.7565.
[12]. A.Khan, S. Shahid, H. Ahmed, T. Ismail, N. Nawaz, and M. Wang, "Machine learning-based flood prediction techniques: A comprehensive review," *Environmental Modelling & Software*, vol. 143, Art. no. 105284, 2021, doi: 10.1016/j.envsoft.2021.105284.
[13]. Y. Zhang, X. Liu, and H. Chen, "A comprehensive review of machine learning approaches for flood depth estimation," *International Journal of Disaster Risk Science*, vol. 16, pp. 433–445, 2025, doi: 10.1007/s13753025-00639-0.
[14]. O. Turgut, I. Kok, and S. Ozdemir, "AgroXAI: Explainable AI-driven crop recommendation system for agriculture 4.0," arXiv:2412.16196, 2024.