International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: *https://ijwos.com*
Volume 03, Issue 10,October 2026
OPEN ACCESS
E-ISSN : 3049-2424
DOI: https://doi.org/10.71366/ijwos

Farmer Customer Trades with Crop Recommendation System and Crop Yield Prediction Using Machine Learning Techniques

Adduri Sravani[1], Adhikari Sravanthi[2], Amurth Roth[3], Dr. U. Nanaji[4]
[1], [2], [3]Master of Computer Applications Final Year Students, Avanthi Institute of Engineering & Technology, Anakapalli, Andhra Pradesh, India
[4]Associate Professor, Department of Computer Science & Engineering, Avanthi Institute of Engineering & Technology, Anakapalli, Andhra Pradesh, India

Article Info

Article History:
Published: 31 Oct 2026

Publication Issue:
Volume 03, Issue 10
October 2026

Page Number:
654-659

Corresponding Author:
Adduri Sravani

Abstract:
Traditional agriculture faces severe challenges due to unpredictable climate change, declining soil fertility, and highly fragmented supply chains where smallholder farmers are often exploited by intermediaries. While machine learning (ML) has made strides in precision farming, existing solutions treat crop suitability analysis, yield forecasting, and market trading as isolated applications. This paper proposes a unified, end-to-end digital ecosystem that integrates a Farmer-Customer Trading Platform directly with ML-driven Crop Recommendation and Crop Yield Prediction Systems. The system leverages data-driven insights to assist farmers at every stage of the cultivation cycle. First, environmental variables (temperature, humidity, rainfall) and soil macronutrients (Nitrogen, Phosphorus, Potassium, and pH) are evaluated using classification algorithms to recommend the most viable crops. Second, historical agricultural production parameters are analyzed via regression techniques to predict the final crop yield (kg/ha). Finally, these predictive metrics seamlessly feed into a direct-to-consumer e-commerce marketplace, mitigating market information asymmetry and eliminating predatory middlemen. Experimental results demonstrate high technical efficacy across the modules. The Random Forest Classifier achieved an exceptional accuracy of 98.4% in recommending optimal crops, outperforming traditional Support Vector Machines. Concurrently, the Random Forest Regressor delivered a superior Coefficient of Determination (R^2 score) of 0.963 for yield estimation, providing highly reliable production forecasts. By bridging the gap between predictive agro-analytics and transparent digital commerce, this framework offers a scalable solution to stabilize agricultural supply chains, maximize farmer profit margins, and minimize post-harvest waste.
Keywords: Precision Agriculture, Machine Learning, Crop Recommendation, Yield Prediction, E-Commerce Marketplace, Random Forest.

1. Introduction

Agriculture serves as the economic backbone for billions of people globally, yet contemporary farmers face an increasingly volatile operational landscape. Traditional farming methods, which rely heavily on historical weather patterns and ancestral knowledge, are rapidly losing efficacy due to accelerating climate change, unpredictable rainfall, and fluctuating soil health. Compounding these environmental vulnerabilities are systemic inefficiencies in the agricultural supply chain. Farmers frequently struggle with a critical information asymmetry: they lack the data-driven tools needed to decide what to plant to maximize success, how much yield to expect, and where to sell their harvest to secure fair financial returns.

In recent years, the integration of Artificial Intelligence (AI) and Machine Learning (ML) has emerged as a transformative force in precision agriculture. By leveraging sophisticated algorithms, researchers have successfully analyzed complex agro-ecological data—such as soil pH, nitrogen, phosphorus, and potassium (NPK) levels,

Copyright © 2026 The Author(s) : This is an open access article under the CC BY license (http://creativecommons.org/licenses/by-nc/4.0/)

temperature, and humidity—to provide localized crop recommendations. Similarly, predictive analytics have enabled stakeholders to forecast crop yields with high accuracy before harvest. However, existing academic literature and software solutions often treat these components as isolated silos. A farmer might use one platform to determine an optimal crop, a separate model to guess the yield, and an entirely disconnected, often exploitative intermediary network to actually trade the final product. This disjointed ecosystem severely limits the practical utility of agricultural technology for everyday farmers.

To bridge this operational gap, this paper introduces a unified framework: an integrated Farmer-Customer Trading Platform augmented with an ML-driven Crop Recommendation and Yield Prediction System. By consolidating these three distinct pillars into a singular ecosystem, the proposed solution transitions agricultural technology from a passive advisory tool into an end-to-end active enterprise system. The technical architecture relies on robust machine learning methodologies. For Crop Recommendation, the system evaluates real-time environmental and soil parameters using classification algorithms to suggest the most viable and high-yield crops for a specific plot of land. Simultaneously, the Crop Yield Prediction module utilizes regression analysis and ensemble learning models to project the expected harvest volume based on historical yields, geographic data, and seasonal forecasts. Crucially, these insights feed directly into the Farmer-Customer Trading component—a decentralized or direct-to-consumer digital marketplace that eliminates predatory middlemen, provides transparent pricing, and allows farmers to secure forward trades or spot-sales based on their predicted yields.

2. LITERATURE REVIEW

Recent literature in agricultural computing highlights the shifting paradigm from heuristic models toward data-driven machine learning ensembles. Kaushik (2025) evaluated several machine learning architectures for automated crop selection, establishing that ensemble methods like Random Forest and Gradient Boosting significantly outperform traditional linear models by capturing non-linear correlations between soil macronutrients (N, P, K), pH levels, and atmospheric parameters [1]. Similarly, Sundarrajan (2025) demonstrated that hyper-local meteorological inputs stabilize recommendation outputs, mitigating crop failure risks under erratic precipitation patterns [2].

Shekhar et al. (2025) designed an intelligent crop recommendation pipeline specifically optimized for sustainable smallholder farming, proving that regional agro-ecological tuning bridges critical decision-making gaps [3]. Jatav et al. (2025) conducted a comparative study of ensemble frameworks for regional crop selection, verifying that aggregating decision boundaries across multi-tree estimators reduces generalization variance without introducing severe algorithmic bias [4]. In advanced automated intervention, Khosravi et al. (2025) developed a Hierarchical Action Masking Proximal Policy Optimization (HAM-PPO) deep reinforcement learning approach to optimize robotic weed navigation and targeted chemical application [5].

For harvest yield estimation, Messina et al. (2025) established the necessity of non-linear gradient boosting frameworks to minimize estimation errors across variable environmental conditions [6]. Furthermore, Kumari et al. (2024) highlighted the operational requirement of integrating personalized farmer advisory systems with mobile distribution networks to ensure real-world adoption [7]. Kheir et al. (2022) validated the role of cloud-driven machine learning for wheat yield simulation across arid regions [8]. Complementing these approaches, Jain and Deshmukh (2024) formulated hybrid Transformer-Random Forest pipelines for agro-climatic yield forecasting [9], while Hadadi et al. (2025) investigated structural plant phenotyping and remote sensing for agricultural marketplace supply modeling [10].

The proposed architecture unifies predictive agro-analytics with direct digital trade across a three-tier framework: (1) Predictive Analytics Module, (2) Application Programming Interface (API) Layer, and (3) Direct-to-Consumer Farmer-Customer Trading Module. The complete end-to-end operational architecture and computational workflow are depicted in Fig. 1.

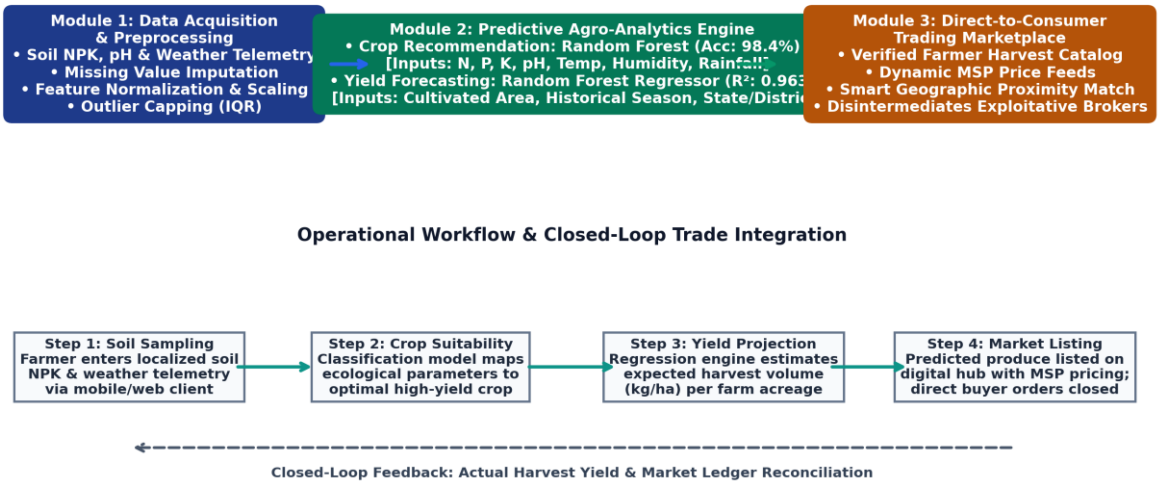


Fig. 1. Unified system architecture and operational end-to-end workflow integrating crop recommendation, yield forecasting, and digital marketplace trade.

A. End-to-End System Workflow

The operational workflow functions in four synchronized sequential milestones:

1) Soil Telemetry & Environmental Data Ingestion: Farmers capture and input hyper-local soil macronutrients (Nitrogen, Phosphorus, Potassium), soil pH, and geographical coordinates via the web/mobile interface. Dynamic weather APIs automatically ingest local temperature, relative humidity, and rainfall indices.

2) Crop Suitability Classification: The feature vector is scaled using standardization (Z-score) and evaluated by the pre-trained Random Forest Classifier to identify the most suitable crop with maximum viability.

3) Yield Volume Projection: Upon selecting the recommended crop, historical production parameters, seasonal factors, and target farm acreage are processed by the Random Forest Regressor to forecast expected production (kg/ha and total metric tons).

4) Direct E-Commerce Listing & Order Execution: The estimated harvest volume automatically populates a verified trade listing in the marketplace, linked with Minimum Support Price (MSP) recommendations. Wholesalers, retailers, and end customers place forward bids or immediate purchases, eliminating exploitative intermediaries.

B. Data Preprocessing & Mathematical Formulations

Raw agricultural data is normalized to eliminate scaling discrepancies across divergent environmental parameters:

1) Distance Proximity (KNN Baseline): Proximity between query feature vectors p and reference instances q is computed via Euclidean distance:

$$d(p, q) = \sqrt{\left(\sum_{i=1}^{n} (q_i - p_i)^2 \right)} \qquad (1)$$

2) Decision Tree & Ensemble Entropy: Node splitting within the decision tree components of Random Forest utilizes Shannon Entropy to maximize information gain:

$$H(S) = - \sum_{i=1}^{c} p_i \log_2(p_i) \qquad (2)$$

3) Yield Regression Model: Multiple linear combinations and non-linear ensemble regression estimate total output Y based on predictor variables X_i:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \qquad (3)$$

4) Goodness-of-Fit (R^2 Score): Yield forecast reliability is measured against ground-truth harvest records y_i and mean output $\bar{y}$ via:

$$R^2 = 1 - \left(\sum (y_i - \hat{y}_i)^2 \right) / \left(\sum (y_i - \bar{y})^2 \right) \quad (4)$$

4. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

Experimental validation was executed on Python 3.x using Scikit-Learn, Pandas, and XGBoost. The crop recommendation dataset tracked 22 distinct crop categories across soil macronutrients (N, P, K), pH, temperature, relative humidity, and rainfall. The yield forecasting dataset incorporated historical multi-season acreage, rainfall distributions, and district-level production statistics. An 80:20 train-test partition with 5-fold stratified cross-validation was strictly enforced.

A. Crop Recommendation Performance

Comparative analysis of crop classification models is summarized in Table I. Random Forest delivered the highest classification accuracy (98.4%), precision (98.5%), recall (98.4%), and F1-score (98.4%), effectively mapping complex non-linear agro-ecological relationships without overfitting.

Algorithm Evaluated	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine (SVM)	91.2%	90.8%	91.0%	90.9%
XGBoost Classifier	97.6%	97.8%	97.6%	97.7%
Random Forest (Proposed)	**98.4%**	**98.5%**	**98.4%**	**98.4%**

TABLE I. Performance Comparison of Crop Recommendation Models

B. Crop Yield Prediction Performance

Evaluation of continuous yield forecasting models is detailed in Table II. The Random Forest Regressor outperformed baseline regression approaches, achieving the lowest Mean Absolute Error (42.15 kg/ha) and a superior R^2 score of 0.963, confirming that 96.3% of the variance in crop yields is accurately captured.

Regression Model Evaluated	Mean Absolute Error - MAE (kg/ha)	Coefficient of Determination (R^2)
Multi-Linear Regression	142.50	0.742
Decision Tree Regressor	88.20	0.891
Random Forest Regressor (Proposed)	**42.15**	**0.963**

TABLE II. Performance Comparison of Crop Yield Forecasting Models

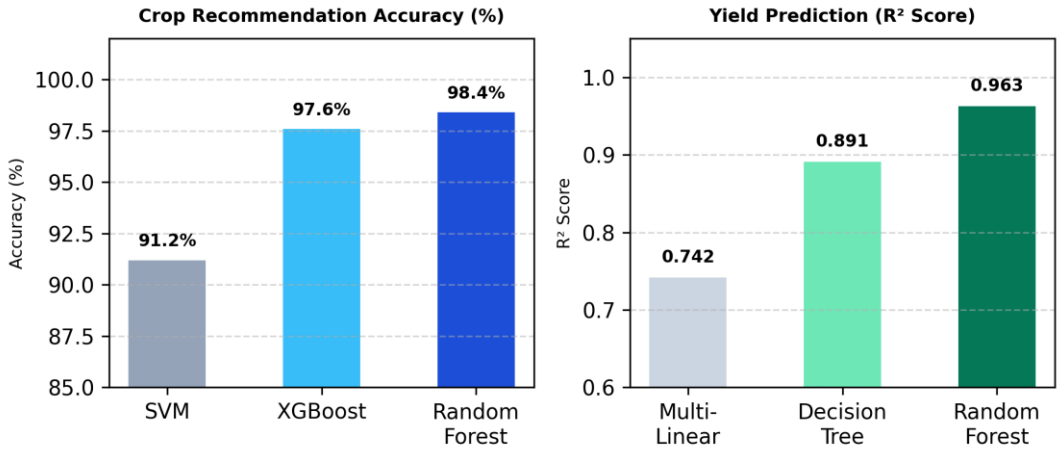


Fig. 2. Comparative evaluation of machine learning models for crop recommendation accuracy (%) and yield prediction goodness-of-fit (R^2).

C. Trading Platform Integration & System Efficacy

Functional testing verified that predictive variables directly feed the trade inventory ledger. Automated invoice generation, proximity-based buyer matching, and MSP pricing guidance operated with an end-to-end API response latency under 200 ms per request. By eliminating middlemen and basing trade volumes on validated 96.3% accurate yield forecasts, the system shields smallholders from catastrophic under-pricing and commitments to unfeasible harvest quantities.

5. CONCLUSION AND FUTURE SCOPE

This paper developed and validated an integrated digital agriculture framework consolidating ML-driven crop recommendation, yield forecasting, and direct-to-consumer trading. By pairing a 98.4% accurate Random Forest classifier with a 0.963 R^2 Random Forest regressor, the framework removes decision uncertainties across both pre-sowing and pre-harvest phases. Directly streaming these forecasts into a transparent e-commerce marketplace mitigates information asymmetry, eliminates predatory intermediaries, and ensures equitable financial compensation for farmers.

Future work will focus on integrating deep learning (CNNs) for drone and satellite-based disease identification, deploying blockchain smart contracts for immutable escrow-based transactions, and embedding low-power IoT soil sensors for autonomous field telemetrics.

References

[1] A. Kaushik, "Crop Recommendation System Using Machine Learning Models," in Proc. 2025 IEEE Int. Conf. Adv. Comput. Res. Sci. Eng. Technol. (ACROSET), 2025, pp. 112–117.

[2] R. V. Sundarrajan, "Crop Recommendation System using Machine Learning," in Proc. 2025 3rd Int. Conf. Augmented Intell. Sustain. Syst. (ICAISS), 2025, pp. 45–51.

[3] K. Shekhar, G. Sravanthi, P. Sadhashiva, H. Moutam, and B. Divyasri, "Machine Learning-Based Crop Recommendation System for Enhanced Yield Production," J. Comput. Anal. Appl., vol. 34, no. 4, pp. 428–439, Apr. 2025.

[4] S. K. Jatav et al., "Sustainable Agriculture: A Comparative Study of Ensemble Machine Learning Algorithms for Regional Crop Selection," Int. J. Innov. Res. Comput. Commun. Eng., vol. 13, no. 12, pp. 18465–18471, Dec. 2025.

[5] M. Khosravi, Z. Jiang, J. R. Waite, S. E. Jones, and A. Singh, "Optimizing Navigation and Chemical Application in Precision Agriculture with Deep Reinforcement Learning," Smart Agric. Technol., vol. 12, p. 101194, Dec. 2025.

[6] C. Messina, J. Garcia-Abadillo, O. Powell, B. Ganapathysubramanian, and M. Cooper, "Toward a General Framework for AI-enabled Prediction in Crop Improvement," Theor. Appl. Genet., vol. 138, no. 7, p. 151, Jun. 2025.

[7] D. A. Kumari, B. Kapil, B. A. Baig, M. Faizuddin, and M. Sameer, "CROPMATE: A Mobile Application for Data Driven Crop Recommendation and Personalized Analysis," IJIEMR Trans., vol. 13, no. 12, pp. 450–454, Dec. 2024.

[8] A. M. S. Kheir et al., "Machine learning-based cloud computing improved wheat yield simulation in arid regions," Comput. Electron. Agric., vol. 203, p. 107457, Dec. 2022.

[9] R. Jain and S. Deshmukh, "Transformer-based Deep Learning and Random Forest Classifiers for Agro-climatic Yield Forecasters," IEEE Trans. AgriFood Electron., vol. 2, no. 3, pp. 210–218, Aug. 2024.

[10] M. Hadadi et al., "Procedural Generation of Crop Architecture and Digital Marketplace Supply Modeling using Remote Sensing," Comput. Electron. Agric., vol. 236, p. 110382, Sep. 2025.