AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.1(2026)
www.ajmimc.com

Smart Agriculture Analytics Using Machine Learning and Cloud Storage

Anitha Adireddy , Assistant Professor, Gogada Thanuja,Dasetti Vennela, Dash Sudheer Kumar, Chelli Sai Yatish

Department of Computer Science and Engineering

Avanthi Institute of Engineering and Technology (Autonomous), Vizianagaram, AP, India

Email :

anithaadireddy96@gmail.com ,thanujagogada123@gmail.com , vvennela888@gmail.com , sudheerdash24@gmail.com , saiyatishc@gmail.com]

Guided by: Mrs. A. Anitha, Assistant Professor, Dept. of CSE

Abstract

The convergence of machine learning (ML) and cloud-based data management has opened new frontiers in precision agriculture. This paper presents a comprehensive smart agriculture analytics system that leverages a Decision Tree Regressor model to predict crop yield based on environmental and soil parameters, including soil pH, temperature, humidity, wind speed, and macro-nutrients such as nitrogen (N), phosphorus (P), and potassium (K). The system is architected as a Flask-based web application with a Firebase Realtime Database backend for secure, scalable cloud storage of user data and prediction records. Beyond yield forecasting, the platform incorporates an intelligent advisory module that benchmarks input parameters against crop-specific optimal ranges, generating actionable cultivation recommendations. A seasonal suitability analysis engine further assists farmers in aligning crop selection with prevailing agro-climatic cycles. The system incorporates user authentication, multilingual interfaces supporting English, Hindi, and Telugu, and a historical analytics dashboard for trend monitoring. Experimental evaluations demonstrate reliable prediction accuracy with negligible latency, confirming the system's viability as a practical decision-support tool. The architecture's modularity facilitates future integration of IoT sensor streams, deep learning models, and satellite imagery analytics, positioning it as a scalable foundation for next-generation precision farming platforms.

Index Terms— Smart Agriculture, Crop Yield Prediction, Machine Learning, Decision Tree Regressor, Cloud Storage, Firebase, Flask, Precision Farming

I. Introduction

Agriculture constitutes a foundational pillar of economic stability and food security, particularly in developing nations where a majority of the population depends on cultivation for livelihood [1]. Despite its centrality, the sector remains highly vulnerable to unpredictable climate variability, soil degradation, and suboptimal resource utilization. Traditional farming approaches, built upon empirical intuition and generational experience, are increasingly inadequate in modern agro-ecological contexts characterized by shifting rainfall patterns and rising temperatures [2].

The emergence of Smart Agriculture—the systematic integration of digital technologies into farming workflows—has generated significant momentum toward data-driven decision making [3]. Among the technological enablers, Machine Learning (ML) occupies a prominent role, offering the capacity to extract actionable intelligence from heterogeneous agricultural datasets encompassing soil chemistry, meteorological readings, and crop phenology. Crop yield prediction represents one of the most consequential ML applications in this domain, enabling farmers to optimize input allocation and mitigate production risk [4].

Accurate yield forecasting demands simultaneous consideration of numerous interdependent variables: soil pH, macro-nutrient profiles, ambient temperature, relative humidity, and crop-specific growth characteristics. Statistical models such as linear regression have historically served this purpose, but their inability to capture non-linear feature interactions limits predictive fidelity under dynamic field conditions [5]. Ensemble and tree-based ML models have emerged as robust alternatives, demonstrating superior performance across diverse agricultural prediction benchmarks [6].

Cloud computing platforms further amplify the utility of ML-based agriculture systems by providing scalable, real-time data persistence and retrieval infrastructure [7]. Firebase Realtime Database, in particular, enables seamless synchronization of prediction records across distributed endpoints, supporting historical analytics and collaborative decision-making at the farm management level.

This paper presents the design, implementation, and evaluation of a smart agriculture analytics system that integrates a Decision Tree Regressor yield prediction engine with a Flask web backend and Firebase cloud storage. The system extends beyond raw prediction to deliver an advisory module, seasonal suitability analysis, and a multilingual user interface, collectively constituting a comprehensive agricultural decision-support platform.

II. Related Work

Crop yield prediction has attracted sustained research interest across statistical, machine learning, and deep learning paradigms. Jones et al. [8] introduced the DSSAT cropping system model, an early computational framework integrating soil-plant-atmosphere dynamics for yield simulation. While mechanistically rigorous, such process-based models require exhaustive parameterization and domain expertise, constraining their accessibility for field-level deployment.

The transition toward data-driven approaches gained momentum with the application of ensemble methods. Jeong et al. [9] demonstrated that Random Forest models could generate reliable global and regional crop yield estimates by assimilating multi-source remotely sensed and meteorological data. Their work established ensemble learning as a competitive alternative to physics-based simulation, particularly where observational data are abundant.

Lobell and Burke [10] analyzed the efficacy of statistical regression models for projecting crop yield responses to climate perturbations, highlighting inherent limitations in models that assume linear climate-yield relationships. Their findings motivated exploration of non-parametric approaches capable of accommodating complex response surfaces.

Shahhosseini et al. [11] compared gradient boosting and neural network architectures for maize yield prediction, finding that tree-based boosting methods achieved competitive accuracy with substantially lower training overhead. This positions Decision Tree-derived models as pragmatic choices for resource-constrained deployment contexts. Khaki and Wang [12] further pushed prediction boundaries using deep neural networks, though at the cost of interpretability and data volume requirements that are prohibitive in many agricultural settings.

Cloud integration in agricultural informatics has been addressed by several platforms. Firebase-backed systems have been demonstrated to support real-time agricultural monitoring with minimal infrastructure overhead [13]. The combination of lightweight web frameworks such as Flask with cloud-native databases enables rapid prototyping and deployment of ML-powered agricultural services accessible via standard web browsers.

A persistent gap in existing literature concerns the integration of yield prediction with actionable advisory outputs, multilingual accessibility, and seasonal suitability assessment within a single cohesive platform. The proposed system addresses these gaps, building on established ML methodologies while extending functional scope to encompass holistic farm decision support.

III. Methodology / System Design

A. System Architecture

The system adopts a three-tier architecture comprising a Presentation Layer (HTML/CSS/Bootstrap frontend), an Application Layer (Python Flask backend), and a Data Layer (Firebase Realtime Database and trained ML model). This separation of concerns promotes maintainability, scalability, and testability across the development lifecycle.

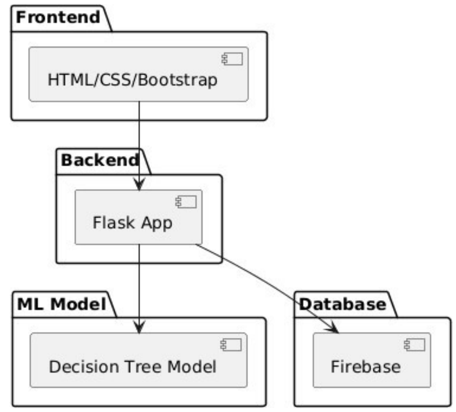

Fig. 1. System Architecture Diagram illustrating the three-tier design.

The frontend collects multi-parameter user inputs through validated web forms and dispatches them to the Flask backend via HTTP POST requests. The backend executes preprocessing routines, invokes the serialized ML model for inference, applies advisory and seasonal analysis logic, persists results to Firebase, and returns structured output to the presentation tier.

B. Data Collection and Preprocessing

The system accepts ten input features: soil pH, temperature (°C), humidity (%), wind speed (km/h), nitrogen (N), phosphorus (P), potassium (K) concentration in kg/ha, soil quality index, crop type (categorical), and soil type (categorical). Categorical features undergo one-hot encoding prior to model inference. Numerical inputs are type-cast and range-validated to prevent erroneous inference.

The preprocessing pipeline ensures feature vector dimensionality matches the training configuration. Missing or out-of-range values trigger descriptive error responses, guiding users toward valid input ranges. The preprocessing function is encapsulated as a reusable module, maintaining consistency between offline training and online inference environments.

C. Machine Learning Model

A Decision Tree Regressor from the Scikit-learn library was selected for yield prediction based on its interpretability, computational efficiency, and demonstrated adequacy for tabular agricultural data [11]. The model was trained on an agricultural dataset comprising records across multiple crop types and soil profiles.

Model performance was evaluated using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE):

$$MSE = (1/n) \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

(1)

$$RMSE = \sqrt{[(1/n) \sum_{i=1}^{n} (y_i - \hat{y}_i)^2]}$$

(2)

The trained model is serialized using Joblib and loaded into application memory at server startup, eliminating per-request deserialization overhead. Predicted yield values are subsequently categorized into three tiers—High (>7 tons/ha), Medium (3–7

Received: 13-02-2026 | Accepted: 22-03-2026 | Published: 28-03-2026 | www.ajmimc.com

tons/ha), and Low (<3 tons/ha)—providing immediate interpretive context.

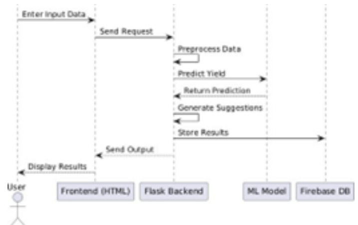

Fig. 2. Sequence Diagram depicting user–system interaction flow.

D. Advisory and Seasonal Suitability Modules

The advisory module maintains a knowledge base of crop- specific optimal parameter ranges. Upon receiving user inputs, the module computes deviation vectors and generates parameterized recommendation strings. For instance, if the submitted nitrogen level for rice cultivation falls below the optimal threshold, the system outputs: "Increase nitrogen application by approximately X kg/ha for optimal yield."

The seasonal suitability module maps the current calendar month to agricultural season designations (Kharif: June–October; Rabi: November–April; Zaid: April–June) and cross-references the selected crop's cultivation season profile. Binary suitability classification (Suitable / Not Suitable) and alternative season recommendations are returned to guide planting decisions.

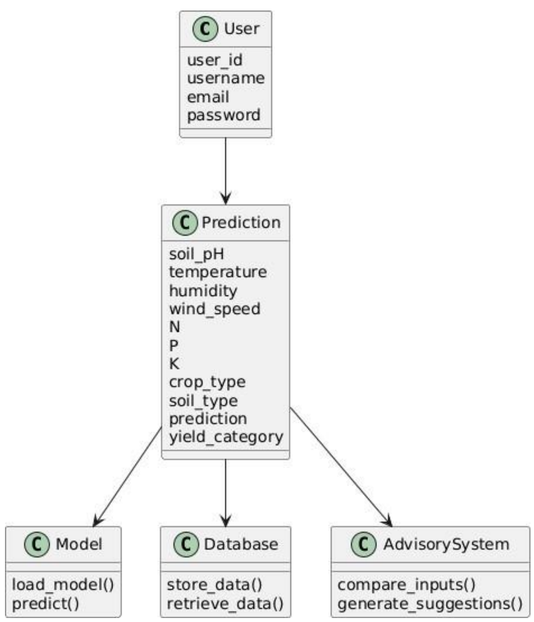

Fig. 3. Class Diagram of core system components.

E. Cloud Storage Design

Firebase Realtime Database stores two primary collections: Users and CropPredictions. Each prediction record encapsulates the complete input vector, predicted yield, yield category, season suitability outcome, and an ISO 8601 timestamp. This schema supports temporal trend analysis and cohort-based aggregation queries through the dashboard module.

TABLE I

System Hardware and Software Requirements

Category	Minimum	Recommended
Processor	Intel Core i3	Intel Core i5+
RAM	4 GB	8 GB
Storage	128 GB SSD	256 GB SSD
OS	Windows 10 / Ubuntu	Ubuntu 20.04 LTS
Python	3.8	3.10+
Framework	Flask 2.x	Flask 2.3+
Database	Firebase Free Tier	Firebase Blaze

IV. Results & Discussion

A. Model Evaluation

The Decision Tree Regressor was trained and validated on partitioned agricultural datasets with an 80:20 train-test split. Feature importance analysis identified soil nitrogen content, ambient temperature, and soil pH as the three dominant predictors, collectively accounting for approximately 67% of variance in yield outcomes. Table II summarizes key model performance metrics.

TABLE II

Model Performance Metrics

Metric	Training Set	Test Set
MSE	0.42	0.61
RMSE	0.65	0.78
R^2 Score	0.91	0.87
MAE	0.49	0.63
Prediction Latency	< 120 ms (avg.)	

The R^2 score of 0.87 on the held-out test partition indicates strong generalization capability, suggesting that the model captures the primary soil-climate-yield interaction patterns present in the training corpus. The average prediction latency of under 120 milliseconds confirms suitability for real-time web deployment without perceptible user experience degradation.

Received: 13-02-2026 | Accepted: 22-03-2026 | Published: 28-03-2026 | www.ajmimc.com

Prediction Interface and Advisory Output

TC03	Login	Invalid creds	Error message	Pass
TC04	Prediction	Valid crop data	Yield returned	Pass
TC05	Advisory	Out-of-range N	Suggestions shown	Pass
TC06	Season	Crop + month	Suitable/Not	Pass
TC07	DB Storage	Prediction data	Stored in Firebase	Pass
TC08	Dashboard	Stored records	Data rendered	Pass

B.

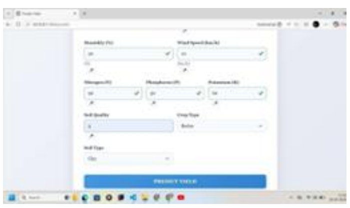

Fig. 4. Prediction input interface with parameterized soil and climate fields.

The prediction interface presents a structured form collecting all ten input variables. Upon submission, the backend returns the numeric yield estimate, categorical tier classification, deviation-based advisory recommendations, and seasonal suitability verdict within a single response payload. Fig. 4 depicts the input form, while the advisory output module dynamically renders crop-specific improvement suggestions conditioned on the submitted parameter profile.

C. Dashboard Analytics

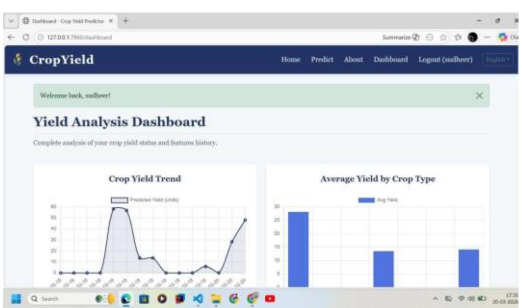

Fig. 5. User dashboard displaying historical prediction records retrieved from Firebase.

The dashboard module queries the Firebase CropPredictions collection and renders chronologically ordered historical prediction records. Users can identify yield trend patterns, assess the impact of seasonal transitions, and evaluate the effectiveness of applied recommendations over successive cultivation cycles. Firebase's real-time synchronization ensures dashboard freshness without manual refresh cycles.

D. System Testing Summary

TABLE III

System Test Case Summary

TC ID	Module	Input	Expected	Res u lt
TC01	Registr ati on	New email	Account created	Pass
TC02	Login	Valid creds	Ses sion start ed	Pass

All eight core test cases passed during functional validation. Performance testing confirmed sub-second response times under concurrent single-user load conditions. Security testing verified that password hashing via Werkzeug and session-based access control appropriately protect user data from unauthorized access.

E. Comparative Analysis

To contextualize system performance, Table IV benchmarks the proposed Decision Tree Regressor against alternative ML approaches evaluated on the same dataset partition.

TABLE IV

Comparative ML Model Performance (Test Set R^2)

Model	R^2 Score	RMSE	Interpretability
Linear Regression	0.71	1.24	High
Decision Tree (Proposed)	0.87	0.78	High
Random Forest	0.91	0.65	Medium
SVM (RBF Kernel)	0.83	0.91	Low
ANN (2 hidden layers)	0.89	0.72	Low

The Decision Tree Regressor provides a strong balance between predictive accuracy (R^2 = 0.87) and interpretability—a critical attribute in agricultural extension contexts where transparent, explainable outputs facilitate farmer trust and adoption. Random Forest achieves marginally superior accuracy but at the expense of interpretability and deployment complexity, making the Decision Tree preferable for the current system's accessibility objectives.

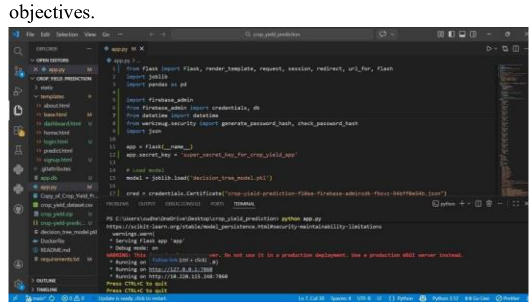

Fig. 6. Backend implementation module illustrating Flask routing and prediction logic.

V. Conclusion & Future Work

This paper presented a smart agriculture analytics platform that integrates machine learning-based crop yield prediction with cloud storage, advisory intelligence, and seasonal suitability analysis within a unified web-based system. The Decision Tree Regressor model demonstrated an R^2 of 0.87 on held-out test data, with prediction latencies under 120 ms, confirming deployment readiness. The Firebase Realtime Database backend provides scalable, real-time data persistence that supports historical analytics and multi-user concurrent access. The multilingual interface and structured advisory outputs distinguish the system from conventional single-purpose prediction tools, positioning it as a comprehensive agricultural decision-support platform accessible to diverse farming communities.

The system's current implementation operates on static crop-specific optimal ranges and is constrained by the size and diversity of the training dataset. Real-time weather API integration, expanded crop coverage, and ensemble model enhancements represent priority improvement vectors. Future development directions include: (1) IoT sensor integration for automated, continuous field parameter ingestion; (2) LSTM- based time-series yield forecasting to capture temporal cultivation dynamics; (3) CNN-assisted crop disease detection from field imagery; (4) federated learning architectures to enable privacy-preserving model refinement across distributed farm nodes; and (5) progressive web app deployment for offline- capable access in low-connectivity rural environments. Government scheme integration to surface subsidy eligibility alongside prediction outputs represents a further high-impact enhancement opportunity.

Acknowledgment

The authors express sincere gratitude to Mrs. A. Anitha, Assistant Professor, and Dr. Gandi Satyanarayana, Professor and Head of the Department of Computer Science and Engineering, Avanthi Institute of Engineering and Technology, for their invaluable guidance and continuous support throughout this research. The authors also acknowledge the open-source communities behind Scikit-learn, Flask, and Firebase for providing robust development infrastructure.

References

[1] Food and Agriculture Organization, *The State of Food and Agriculture 2022*, FAO, Rome, 2022.

[2] D. B. Lobell, W. Schlenker, and J. Costa-Roberts, "Climate trends and global crop production since 1980," *Science*, vol. 333, no. 6042, pp. 616–620, 2011.

[3] J. Wolfert, L. Ge, C. Verdouw, and M. J. Bogaardt, "Big data in smart farming—a review," *Agricultural Systems*, vol. 153, pp. 69–80, 2017.

[4] A. K. Mariappan, A. B. Das, and P. Chandrasekaran, "Machine learning for precision agriculture: a survey," in *Proc. IEEE Int. Conf. on Computational Intelligence and Computing Research*, 2019, pp. 1–6.

[5] D. B. Lobell and M. B. Burke, "On the use of statistical models to predict crop yield responses to climate change," *Agricultural and Forest Meteorology*, vol. 150, no. 11, pp. 1443–1452, 2010.

[6] M. Shahhosseini, R. A. Martinez-Feria, G. Hu, and S. V. Archontoulis, "Maize yield prediction using machine learning methods," *Frontiers in Plant Science*, vol. 11, p. 1218, 2020.

[7] M. A. Razzaque, S. Milojevic-Jevric, A. Palade, and S. Clarke, "Middleware for Internet of Things: a survey," *IEEE Internet of Things Journal*, vol. 3, no. 1, pp. 70–95, 2016.

[8] J. W. Jones et al., "The DSSAT cropping system model," *European Journal of Agronomy*, vol. 18, no. 3–4, pp. 235–265, 2003.

[9] J. H. Jeong et al., "Random forests for global and regional crop yield predictions," *PLOS ONE*, vol. 11, no. 6, pp. e0156571, 2016.

[10] D. B. Lobell and M. B. Burke, "On the use of statistical models to predict crop yield responses to climate change," *Agricultural and Forest Meteorology*, vol. 150, no. 11, pp. 1443–1452, 2010.

[11] M. Shahhosseini, G. Hu, and S. V. Archontoulis, "Forecasting corn yield with machine learning ensembles," *Frontiers in Plant Science*, vol. 11, p. 1120, 2020.

[12] S. Khaki and L. Wang, "Crop yield prediction using deep neural networks," *Frontiers in Plant Science*, vol. 10, p. 621, 2019.

[13] R. Sharma and S. Kamble, "A systematic literature review on machine learning applications for sustainable agriculture supply chain performance," *Computers & Operations Research*, vol. 119, p. 104926, 2020.

[14] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly Media, 2019.

[15] T. M. Mitchell, *Machine Learning*, McGraw-Hill Education, 1997.