A Machine Learning-Powered Framework for Predictive Soil Analysis and Smart Fertilizer Recommendation

S. Venkatesh
Department of Computer Science and Engineering
Dr. M.G.R. Educational and Research Institute Chennai, India
venkatmcc007@gmail.com

V. Muthu Krishnan
Department of Computer Science and Engineering
Dr. M.G.R. Educational and Research Institute Chennai, India
vmkrishnan08@gmail.com

Dr. M. Sujitha
Department of Computer Science and Engineering
Dr. M.G.R. Educational and Research Institute Chennai, India
sujitha.cse@drmgrdu.ac.in

Dr. Saravanan Elumalai
Department of Computer Science and Engineering
Dr. M.G.R. Educational and Research Institute Chennai, India
saravanan.e@drmgrdu.ac.in

Dr. V. Sai Shanmuga Raja
Department of Computer Science and Engineering
Dr. M.G.R. Educational and Research Institute Chennai, India
saishanmugaraja.cse@drmgrdu.ac.in

Abstract**—**In contemporary agriculture, optimizing soil health and delivering precise fertilizer guidance are indispensable to enhancing crop yield, sustainability, and resource efficiency. This paper presents AGROMIND, an intelligent Machine Learning-Powered Framework for Predictive Soil Analysis and Smart Fertilizer Recommendation, designed to transform conventional soil management practices. The framework employs machine learning algorithms to evaluate key soil parameters including pH, nitrogen (N), phosphorus (P), potassium (K), moisture, and temperature. Based on these parameters, the system carries out two primary functions: it evaluates soil quality to suggest necessary corrective measures through targeted fertilizer usage, and it identifies the most appropriate fertilizer type for specific regional soil conditions. By training predictive models on varied soil datasets, the system delivers accurate, data-driven fertilizer guidance that encourages sustainable farming. The solution is deployed via the Django web framework, providing farmers and agricultural specialists an accessible and intuitive interface for enhanced soil management and crop production decision-making.

Keywords**—***soil health analysis; machine learning; fertilizer recommendation; soil salinity prediction; soil parameter classification; sustainable agriculture; Django web deployment; supervised learning algorithms*

I. INTRODUCTION

Soil health is a foundational determinant of agricultural success and long-term sustainability. As global population growth intensifies demand for food and land resources become increasingly strained, modern agriculture must integrate data- driven technologies to enhance productivity while protecting environmental integrity. Conventional soil testing and fertilizer recommendation approaches are frequently time-consuming, inconsistent, and imprecise in their outputs.

To overcome these challenges, this paper introduces AGROMIND, a Machine Learning-Powered Framework for Predictive Soil Analysis and Smart Fertilizer Recommendation. The framework leverages advanced machine learning algorithms to analyze critical soil characteristics including pH, nitrogen (N), phosphorus (P), potassium (K), moisture, and temperature, thereby assessing soil fertility and recommending fertilizers tailored to specific regional conditions.

By training the system on comprehensive soil datasets, AGROMIND delivers contextually aware recommendations that optimize fertilizer application, reduce environmental impact, and maximize crop output. The solution is integrated within a Django-based web application, providing an intuitive interface that enables farmers and agricultural practitioners to make well-informed soil management decisions. This work represents a meaningful contribution toward precision agriculture and sustainable farming through the application of artificial intelligence.

II. RELATED WORK

Considerable research has explored the application of machine learning in agriculture to improve crop productivity and support decision-making. Chappidi [1] proposed a crop yield prediction system employing machine learning algorithms to analyze agricultural datasets and identify suitable crops based on regional, seasonal, and temporal factors. Similarly, Champaneri and Chandvidkar [2] developed a web-based crop prediction model utilizing the Random Forest algorithm to estimate yield by analyzing environmental variables such as rainfall, temperature, and humidity.

Varma and Singh [3] investigated the effectiveness of machine learning algorithms including Random Forest Regression and Support Vector Regression for crop yield forecasting using meteorological variables, demonstrating their superiority over conventional statistical approaches. Ishwarya and Nagapooja [4] focused on evaluating soil parameters such as pH, temperature, and moisture to determine soil quality and generate crop and fertilizer recommendations. More recently, Cyril, Archana, and Vignesh [5] introduced a crop yield prediction system integrating historical and real-time agricultural data to support precision farming and resource optimization.

Collectively, these studies underscore the growing importance

ISSN: 2583-6129
DOI: 10.55041/ISJEM05791

of machine learning in agricultural applications and establish the theoretical and methodological foundation upon which AGROMIND is developed.

III. PROPOSED METHODOLOGY

The proposed methodology develops an intelligent machine learning-based system capable of performing predictive soil analysis and delivering smart fertilizer recommendations to enhance agricultural productivity. The workflow proceeds through five principal stages: data collection, preprocessing, model development, evaluation, and web deployment.

A. Data Collection

Relevant soil and environmental data are sourced from reliable agricultural repositories. The dataset encompasses key soil indicators including pH level, nitrogen (N), phosphorus (P), potassium (K), soil moisture, and temperature, each of which is critical to determining soil fertility and crop growth potential.

B. Data Preprocessing

The raw dataset undergoes preprocessing to ensure quality and consistency. This involves handling missing values, removing duplicate or noisy records, normalizing numerical features, and converting variables to appropriate formats for machine learning analysis.

C. Exploratory Data Analysis and Feature Selection Exploratory analysis and visualization techniques are applied to identify patterns, correlations, and distributions within the soil attributes. Feature selection methods isolate the most influential predictors, reducing computational overhead and improving model accuracy.

D. Model Training and Evaluation

The prepared dataset is partitioned into training and testing subsets. Multiple supervised machine learning algorithms— Decision Tree, Random Forest, Support Vector Machine (SVM), Linear Regression, MLP Classifier, and Bagging Classifier—are trained and evaluated using accuracy, precision, recall, and F1-score metrics. Cross-validation is employed to mitigate overfitting and ensure generalizability. The algorithm demonstrating the highest predictive accuracy and consistency is selected as the deployment model.

E. Web-Based Deployment

The best-performing model is integrated into a Django web application, providing a user-friendly interface for soil data input and fertilizer recommendation retrieval. Users enter soil parameter values obtained from field testing, and the system responds with tailored fertilizer guidance, enabling data-driven agricultural decisions without requiring specialized technical expertise.

IV. SYSTEM ARCHITECTURE

The architecture of the proposed system is structured as an interconnected multi-layer framework, encompassing data acquisition, preprocessing, feature analysis, model training, prediction, and application deployment. Each layer contributes a distinct function to the overall operation.

A. Data Acquisition Layer

This layer gathers soil parameter data from agricultural research databases, soil testing laboratories, and publicly available datasets. The collected attributes—pH, N, P, K, moisture, and temperature—serve as the primary inputs for subsequent analysis.

B. Data Preprocessing Layer

Raw data is cleaned by handling missing entries, eliminating duplicates, correcting inconsistencies, normalizing features, and encoding categorical values. This stage is critical as dataset quality directly influences model performance.

C. Feature Analysis Layer

Exploratory data analysis is performed using graphical tools including histograms, scatter plots, and box plots. Feature selection methods are applied to retain only those attributes with significant predictive value, enhancing model efficiency.

D. Model Training and Evaluation Layer

The preprocessed dataset is divided into training and testing subsets. Algorithms including Decision Tree, Random Forest, SVM, Linear Regression, MLP Classifier, and Bagging Classifier are trained to learn soil-fertilizer relationships. Model performance is benchmarked using accuracy, precision, recall, and F1-score, and the highest-performing model is retained for deployment.

E. Prediction and Recommendation Module

This module accepts user-supplied soil parameters and processes them through the trained model to produce fertilizer recommendations. It communicates the predicted fertilizer type and any suggested nutrient corrections to the web interface.

F. Application and Deployment Layer

The Django web framework serves as the interface between the end user and the prediction engine. Farmers or agricultural specialists input soil values and receive instant fertilizer recommendations presented in an interpretable format. The architecture supports future enhancements including cloud integration, mobile accessibility, and IoT-based real-time soil monitoring.

V. ALGORITHM AND MODULE DESCRIPTION

The system incorporates four primary machine learning algorithms, each contributing uniquely to predictive performance.

A. MLP Classifier (Multi-Layer Perceptron)

The MLP Classifier is a feedforward artificial neural network comprising an input layer, multiple hidden layers, and an output layer. Soil parameters serve as input features, and the network learns complex nonlinear soil-fertilizer mappings through backpropagation. Weight adjustments are iteratively applied to minimize prediction error, enabling the network to classify soil conditions and recommend appropriate fertilizers.

B. Bagging Classifier

The Bagging (Bootstrap Aggregating) Classifier is an ensemble method that constructs multiple models trained on independently sampled subsets of the training data. Final predictions are determined through majority voting, reducing variance and improving stability. This approach is particularly advantageous when dealing with noisy or variable agricultural datasets.

C. Linear Regression

Linear Regression establishes a linear mapping between soil parameters and fertilizer targets using least squares optimization. While relatively simple, this model provides a valuable baseline and aids in understanding individual nutrient contributions to fertilizer requirements.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI: 10.55041/ISJEM05791
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

D. Random Forest Regression

Random Forest Regression constructs an ensemble of decision trees, each trained on random data subsets, and aggregates their outputs through averaging. This method effectively handles nonlinear feature interactions, mitigates overfitting, and delivers robust predictions across diverse agricultural scenarios.

The system is organized into six functional modules:

- **Data Collection Module:** Gathers soil parameters (N, P, K, pH, moisture, temperature, humidity) from agricultural sources.
- **Data Preprocessing Module:** Cleans, normalizes, and structures the raw dataset for analysis.
- **Model Training Module:** Applies machine learning algorithms to learn patterns from the prepared soil data.
- **Model Evaluation Module:** Measures predictive performance using accuracy, precision, recall, and F1- score.
- **Prediction Module:** Processes user-supplied soil inputs through the trained model to generate fertilizer recommendations.
- **Web Application Module:** Provides an interactive Django-based interface for user interaction and recommendation display.

VI. DATASET DESCRIPTION

The dataset employed in this study encapsulates soil parameter data and associated environmental conditions that collectively determine fertilizer requirements and influence crop productivity. It is organized in a tabular format where each record corresponds to a distinct soil sample and each column represents a specific attribute.

Primary nutrient attributes include nitrogen (N), phosphorus (P), and potassium (K), which govern essential plant physiological processes. Nitrogen drives foliar growth and photosynthesis; phosphorus supports root development and reproductive processes; potassium enhances plant resilience and disease resistance. The soil pH value, indicating acidity or alkalinity, critically governs nutrient bioavailability and is therefore an indispensable feature. Environmental variables including temperature, humidity, and rainfall capture climatic influences on soil moisture dynamics and nutrient absorption.

Prior to model training, the dataset undergoes thorough preprocessing involving removal of missing values, correction of inconsistencies, and feature normalization. The processed dataset is then partitioned into training and testing subsets. A representative sample of dataset records is presented in Table I.

TABLE I. SAMPLE DATASET RECORDS

N (mg/kg)	P (mg/kg)	K (mg/kg)	Temp (°C)	Humidity (%)	pH	Rainfall (mm)	Recommendation
90	42	43	20.8	82	6.5	202	Urea
85	58	41	22.3	80	6.8	210	DAP
60	55	44	24.0	75	7.0	180	NPK
74	35	40	26.5	70	6.4	190	Potash
78	40	38	25.2	72	6.7	195	Urea

VII. EXPERIMENTAL ENVIRONMENT AND PERFORMANCE METRICS

The experimental environment was established on a standard computing platform configured to support machine learning model training and data analysis. Development was conducted using the Python programming language, widely adopted in data science and machine learning for its flexibility and extensive ecosystem.

Key libraries utilized include: Pandas for dataset management and manipulation; NumPy for numerical computation; Scikit- learn for implementing MLP Classifier, Bagging Classifier, Linear Regression, and Random Forest Regression; and Matplotlib for performance visualization. The dataset was stored in CSV format, enabling seamless integration with Python- based analysis tools.

Following preprocessing, the dataset was partitioned into a training subset for model learning and a testing subset for performance evaluation. The system's web interface was constructed using the Django framework, through which users supply soil parameters and receive fertilizer recommendations from the trained prediction engine.

A. Performance Evaluation Metrics

Four standard metrics were employed to quantify model performance:

- **Accuracy:** The proportion of correct predictions over total predictions, reflecting overall model reliability.
- **Precision:** The ratio of true positive predictions to total positive predictions, measuring the model's specificity in fertilizer classification.
- **Recall:** The ratio of true positives to the sum of true positives and false negatives, capturing the model's sensitivity to relevant outcomes.
- **F1-Score:** The harmonic mean of precision and recall, offering a balanced performance indicator particularly suitable for multi-class classification tasks.

B. Algorithm Performance Comparison

Performance comparison was conducted across the four implemented algorithms. Table II summarizes the evaluation results on the test dataset.

TABLE II. ALGORITHM PERFORMANCE COMPARISON

Algorithm	Accuracy (%)	Precision	Recall	F1-Score
MLP Classifier	91.4	0.91	0.91	0.91
Bagging Classifier	93.2	0.93	0.93	0.93
Linear Regression	78.6	0.79	0.78	0.78
Random Forest Regression	96.8	0.97	0.97	0.97

VIII. RESULTS AND DISCUSSION

The experimental outcomes confirm the viability of machine learning approaches for predictive fertilizer recommendation based on soil nutrient and environmental indicators. Following preprocessing and dataset partitioning, four algorithms—MLP Classifier, Bagging Classifier, Linear Regression, and Random

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI: 10.55041/ISJEM05791
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Forest Regression—were independently trained and comparatively evaluated.

All implemented algorithms successfully identified soil- fertilizer patterns to varying degrees. Random Forest Regression achieved the highest overall accuracy of 96.8%, attributable to its ensemble structure that aggregates outputs across multiple decision trees, effectively mitigating overfitting while accommodating nonlinear feature interactions. The Bagging Classifier also demonstrated strong performance at 93.2% accuracy, reflecting the reliability inherent in bootstrap aggregation methodologies.

The MLP Classifier attained 91.4% accuracy, demonstrating the network's capacity to capture complex nonlinear soil-fertilizer relationships, albeit at the cost of greater computational resource consumption during training. Linear Regression recorded the lowest accuracy at 78.6%, constrained by its assumption of linear variable relationships, which inadequately represents the multidimensional complexity of soil nutrient interactions.

Graphical analysis through accuracy comparison charts and training-versus-testing performance plots confirmed that Random Forest maintained consistent performance across both subsets, indicating strong generalization to previously unseen soil data. Linear Regression exhibited greater divergence between training and testing accuracy, indicative of limited modeling capacity for this agricultural prediction task.

These findings highlight the practical value of ensemble learning approaches, particularly Random Forest, in agricultural decision support applications. The integration of the best- performing model within a Django web application enables farmers to input soil test values and obtain reliable, instant fertilizer guidance, facilitating more sustainable resource utilization, reduced chemical over-application, and improved crop productivity.

IX. CONCLUSION

This paper has presented AGROMIND, an intelligent machine learning-based framework for soil parameter analysis and fertilizer recommendation aimed at enhancing agricultural productivity. By evaluating soil nutrient values including nitrogen, phosphorus, potassium, pH, temperature, humidity, and rainfall, the system identifies appropriate fertilizer interventions to support optimal crop growth.

Comparative evaluation of four supervised machine learning algorithms demonstrated that Random Forest Regression achieved the highest predictive accuracy and the most stable generalization performance. Integration of the trained model within a Django web application delivers an accessible, user- friendly interface enabling real-time fertilizer recommendations for farmers and agricultural practitioners without specialized technical knowledge.

The proposed system contributes meaningfully to precision agriculture by providing data-driven insights that minimize excessive fertilizer application, improve soil health, and support sustainable crop production. Future extensions may incorporate IoT-based real-time soil sensing, mobile application interfaces, cloud-based model scaling, and deep learning architectures to further strengthen predictive capabilities and operational reach.

REFERENCES

[1] S. Chappidi, "Crop yield prediction system using machine learning algorithms," International Journal of Agricultural Informatics, vol. 10, no. 4, pp. 12–19, 2022.
[2] M. Champaneri and C. Chandvidkar, "Web-based crop yield prediction using random forest algorithm," International Journal of Computer Applications, vol. 169, no. 7, pp. 1–6, 2017.
[3] A. Varma and R. Singh, "Comparative analysis of machine learning algorithms for crop yield forecasting using weather variables," Computers and Electronics in Agriculture, vol. 195, p. 106782, 2022.
[4] N. Ishwarya and B. N. Nagapooja, "Soil quality analysis and crop-fertilizer recommendation using machine learning," International Journal of Advanced Research in Computer Science, vol. 13, no. 2, pp. 45–52, 2022.
[5] J. Cyril, S. Archana, and R. Vignesh, "Machine learning- based crop yield prediction integrating real-time and historical data for precision agriculture," Journal of Agricultural Technology, vol. 20, no. 1, pp. 89–99, 2024.
[6] R. Sujatha and P. Isakki, "A study on crop yield forecasting using classification techniques," International Journal of Computer Science and Engineering, vol. 5, no. 8, pp. 1–5, 2017.
[7] S. Ramesh and B. Vishnu Vardhan, "Data mining techniques and applications to agricultural yield data," International Journal of Advanced Research in Computer Science, vol. 4, no. 11, pp. 84–88, 2013.
[8] S. Jeong, J. Kim, and H. Lee, "Crop yield prediction using machine learning techniques," Journal of Agricultural Informatics, vol. 7, no. 3, pp. 45–53, 2016.
[9] A. K. Tripathy, J. Adinarayana, and D. Sudharsan, "Machine learning approach for crop yield prediction," Computers and Electronics in Agriculture, vol. 142, pp. 321–330, 2017.
[10] L. Breiman, "Random forests," Machine Learning Journal, vol. 45, no. 1, pp. 5–32, 2001.
[11] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," Informatica, vol. 31, no. 3, pp. 249–268, 2007.
[12] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.
[13] S. A. K. Ahamed, "Smart agriculture system using machine learning," International Journal of Engineering Research and Technology, vol. 8, no. 5, 2019.
[14] R. Rajalakshmi and S. D. Mahalakshmi, "IOT based crop field monitoring and irrigation automation," International Journal of Advanced Research in Computer Science, vol. 8, no. 9, pp. 107–111, 2017.
[15] M. S. Kiran and S. Kumar, "Machine learning techniques for soil fertility prediction," International Journal of Computer Applications, vol. 179, no. 18, pp. 20–24, 2018.
[16] P. B. Nithya and V. Ilango, "Predictive analytics in agriculture using machine learning techniques," International Journal of Engineering Research, vol. 5, no. 3, pp. 1–5, 2016.
[17] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016.
[18] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.
[19] P. C. Senthilkumar and K. S. Ramasamy, "Soil nutrient analysis and crop recommendation using machine learning," International Journal of Advanced Research in Computer Science, vol. 9, no. 2, 2018.
[20] Food and Agriculture Organization (FAO), Soil Fertility Management in Agriculture, FAO Report, Rome, Italy, 2015.
[21] Indian Council of Agricultural Research (ICAR), Soil Health Management and Fertilizer Recommendation, ICAR Publications, New Delhi, India, 2018.
[22] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann Publishers, 2011.

[23] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed., O'Reilly Media, 2019.
[24] T. Mitchell, Machine Learning, McGraw-Hill Education, 1997.
[25] K. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.