Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20260503133
Volume 5, Issue3 (September-December 2026), PP: 1105-1109.
www.indjcst.com
ISSN No: 2583-5300

Crop and Fertilizer Recommendation from Soil Data Using Machine Learning

Prasanna kumar Naik R[1], Chethan Kumar G S[2], Swarnalatha[3], Preethi K P[4], Meghana B R[5]

[1,2,3,4] *Assistant professor, Department of Master of Computer Application, University BDT. College Of Engineering (A Constituent College of VTU, Belagavi) Davangere, Karnataka, India.*

[5] *Student, Department of Master of Computer Application, University BDT. College Of Engineering (A Constituent College of VTU, Belagavi) Davangere, Karnataka, India.*

To Cite this Article: *Prasanna kumar Naik R[1], Chethan Kumar G S[2], Swarnalatha[3], Preethi K P[4], Meghana B R[5], "Crop and Fertilizer Recommendation from Soil Data Using Machine Learning", Indian Journal of Computer Science and Technology, Volume 05, Issue 03 (September-December 2026), PP: 1105-1109.*

Copyright: ©2026 This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution License; Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: *Agriculture is essential to maintaining rural economies and guaranteeing food security, but old farming methods frequently rely on conventional wisdom and homogenous fertilizer schedules rather than scientific soil analysis. Selecting unsuitable crops or applying imbalanced chemical fertilizers leads to reduced crop yields, nutrient depletion, soil degradation, and unnecessary financial expenditure for farmers. To enable data-driven precision agriculture, this paper presents an automated Machine learning (ML) is used by the Crop and Nourishment Recommendation System to examine soil and conservation features. The suggested framework evaluates essential soil macronutrients—Nitrogen (N), Phosphorus (P), and Potassium (K)—along with soil pH, soil type, temperature, humidity, and rainfall to provide dual decision support: foreseeing the most fit crop for agriculture and recommending the optimal fertilizer type for balanced soil nutrition. Agricultural datasets undergo systematic preprocessing, including missing-value handling, label encoding of categorical soil and crop attributes, and feature normalization. Supervised ML classification models, with Random Forest serving as the core ensemble classifier, are trained to capture non-linear relationships among soil parameters and agronomic targets, and appraised by exactness, precision, recall, and F1-score. Integrated into an accessible web-based interface, the system allows farmers and agricultural advisors to input soil test metrics and obtain immediate, reliable crop and nourishment recommendations to enhance productivity and preserve long-term soil health.*

Key Word: *Crop Recommendation, Fertilizer Prediction, Machine Learning, Precision Agriculture, Random Forest Classifier, Soil NPK Analysis.*

I. INTRODUCTION

Agriculture serves as the primary foundation of sustainable economic growth, rural employment, and worldwide food security. Reaching the highest level of agricultural productivity requires heavily on selecting crops suited to local soil characteristics and applying balanced nutrients throughout the cultivation cycle. However, rapid climate variability, soil nutrient depletion, and increasing food demand have made conventional trial-and-error farming practices insufficient for modern agricultural planning [1]. Precision agriculture addresses this challenge by applying data-driven analytical tools to guide site-specific crop cultivation and soil management decisions [1], [2].

In traditional farming workflows, crop selection and fertilizer application are often guided by inherited practices, informal regional advice, or uniform chemical schedules rather than quantitative soil testing. When crops are planted without evaluating soil macronutrient availability—specifically alongside soil pH and local weather patterns, harvest yields decline significantly. Similarly, indiscriminate or imbalanced fertilizer application increases input costs, degrades long-term soil fertility, and causes environmental contamination through nutrient leaching and runoff [2]. Providing farmers with accessible, scientifically grounded decision support directly from soil test data is therefore essential.

Advances in Machine Learning (ML) have enabled computational systems to model complex, non-linear relationships across agronomic and environmental datasets [1], [2]. Unlike static rule-based recommendations, supervised classification algorithms learn directly from historical soil and crop records to foresee results with from top to bottom accuracy [5]. Among these techniques, ensemble tree-based methods such as the Random Forest classifier [3] are particularly effective for tabular agricultural data because they combine calculations athwart many decorrelated decision trees, mitigating overfitting and maintaining robustness against noisy sensor or laboratory measurements [4], [5].

To tie the opening amongst soils laboratories metrics and actionable fields decision these papers present an integrated crops and fertilizers recommendations systems using machine learning The system process historical soil and

environmental dataset incorporating soils macronutrient (N, P, K) rainfalls through structured date cleanings categorical encodings and features preparations Supervised Machine Learning classification centered on the Random Forest algorithm is employed performs duals predictions recommending the most suitable crop for given soils and climatic condition and identifying the optimal fertilizers formulations needed to replenish nutrients deficiencies Deployed through a lightweights web interfaces the systems enable farmers and agricultural advisors to enters soils parameters and obtained immediate reliable recommendations.

The key contributions of this work are summarized as follows:

- Dual Recommendation Framework: Integrating both crops suitability predictions and nutrient specifics fertilizer commendations within a unifieded machine Learning pipeline drivens by soil and environmental parameters.
- Robust Data Preprocessing and Ensemble Modeling: Applying systematic dataset cleaning, categorical encoding and Randoms Forests classification evaluated using standard multi-class performance metrics.
- Accessible Decision-Support Interface: Delivering an intuitive web-based platform that translates raw soil test values into clear, actionable guidance for sustainable farm planning.

II. LITERATURE SURVEY

Data-driven methods have become more prevalent in precision agriculture in recent years. to optimize crop yield, reduce fertilizer wastage, and preserve soil quality [1], [2]. Existing approaches to crop selection and nutrient management reported in the literature span traditional rule-based advisory frameworks, classical single-model machine learning classifiers, ensemble tree-based methods, and deep learning architectures.

A. Traditional and Rule-Based Agronomic Advisory Methods

Historically, crop selection and fertilizer dosage decisions relied on manual soil sampling followed by generalized regional recommendations published by agricultural extension agencies [1]. While expert-defined threshold tables capture broad soil suitability rules, they rely on static boundaries and fail to account for subtle, multi-variable interactions among Nitrogen (N), Phosphorus (P), Potassium (K), soil pH, moisture, and localized climate conditions [2]. As environmental variability increases, static rule-based advisories often produce suboptimal crop matches and imbalanced chemical fertilization.

B. Classical Machine Learning Approaches for Crop and Nutrient Classification

To overcome the rigidity of manual guidelines, researchers have applied classical supervised learning algorithms—including Decision Trees [7], Support Vector Machines (SVM) [8], and probabilistic classifiers [6]—to agricultural soil datasets [1]. Single Decision Trees provide transparent if-then decision paths from soil nutrient thresholds [7], making them easy to interpret; however, they are sensitive to small dataset perturbations and prone to overfitting. Support Vector Machines [8] effectively separate non-linear feature boundaries through kernel transformations [5], [6], yet they require careful feature scaling and hyperparameter tuning when applied to multi-class crop datasets containing over twenty crop categories.

C. Ensemble and Deep Learning Approaches in Precision Agriculture

More current lessons establish that ensemble learning techniques consistently outperform individual base classifiers on tabular soil and weather datasets [1], [2]. Bootstrap aggregating (Bagging) [10] and Random Forest [3] construct multiple decorrelated decision trees and combine their predictions via majority voting, substantially improving classification stability and generalization [5]. Similarly, gradient boosting frameworks such as XGBoost [9] achieve strong predictive accuracy across structured datasets, though with higher tuning complexity.

In parallel, Deep Learning architectures [11]—such as Convolutional Neural Networks (CNNs) and Multi-Layer Perceptrons—have been explored for agricultural tasks involving satellite imagery, leaf disease detection, and continuous IoT sensor streams [1], [2]. While deep nervous web surpasses an unstructured visuals and high frequencies temporals datas they typically requires largest training corpora and substantial computationals resources offerings limited advantages over trees based ensemble on compacts tabulars soils NPK dataset.

D. Research Gaps and Motivation

Reviewing existings agriculturals decision support studies reveals three primary gaps:

1. Isolated Prediction Scope: Many existing model addresses either crops recommendations or fertilizers predictions independently forcing farmer to there consultant separates tools for completed cultivations plannings.
2. High Infrastructure Dependency: Severals smart farmings prototypes depends on costly real times IoT soils probes networks remotes sensing pipelines there remains inaccessible to small medium scales farmers.
3. Limited End-User Usability: Numerous academics models remains offlines analytical scripts without a simples web interface capables of converting routines soils health card value immediate recommendations.

To address these gap the proposed systems integrates both crop recommendation and fertilizer recommendation into a unified, lightweight Machine Learning pipeline powered by Random Forest classification [3], [4] and accessible through an intuitive web interface.

III. METHODOLOGY

The proposed crops and fertilizer commendation system follow a structure supervised Machine Learning work flows comprising datasets collections datas preprocessings features selections dataset splittings, Random Forests classifications models evaluation models serialization and web based deployments.

A. Dataset Collection and Data Preprocessing

Two structured agriculturals dataset stored Comma Separated Values (CSV) formats or used builds independents predictions models for crops selections and fertilizers recommendations:

1. Crop Recommendation Dataset: Contain soil macronutrients concentrations soil pH and climatic measurements (temperature, humidity, and rainfall) mapped to suitables crops categories.
2. Fertilizer Recommendation Dataset: Contains soils macronutrients level (N, P, K), soil moistures soil types cultivated crop types ambient temperatures humidity mapped to appropriates fertilizers class.

Before training the models, systematic preprocessing carried using Python, Pandas, NumPy:

- Data Inspection and Cleaning: Loadings there CSV datasets into structureded DataFrames, verifying column schema removings duplicate records handling missing, inconsistent values.
- Numerical Standardization: Ensuring all soils nutrient pH, moistures, and weathers value are converted into consistents numerical format.
- Categorical Encoding: Labels encoding is the processes of convertings non numerics categories variable like soil types, crop types, and targets classes label into numerics integer codes so that machines learning classifiers can process them consistently.

B. Feature Selection

Relevants agronomic and environmentals features are the cleaned datasets for each recommendation task:

- Crop Recommendation Features: (N,P,K), temperature, humidity, soil pH, and rainfall are used features to envisage the most suitable crop.
- Fertilizer Recommendation Features: Temperature, humidity, soil moisture, encoded soil type, encoded crop type, Nitrogen (N), Potassium (K), and Phosphorus (P) are used as input features to predict the optimal fertilizer.

Parameter	Applicable Module(s)	Agronomic Significance
Nitrogen (N)	Crop & Fertilizer	Governs vegetative leaf growth and chlorophyll synthesis
Phosphorus (P)	Crop & Fertilizer	Supports root expansion, flowering, and energy transfer
Potassium (K)	Crop & Fertilizer	Regulates water uptake, enzyme activation, and disease resistance
Soil pH	Crop	Determines nutrient solubility and crop soil compatibility
Temperature	Crop & Fertilizer	Controls germination rate and plant metabolic growth
Humidity	Crop & Fertilizer	Influences transpiration and atmospheric moisture balance
Rainfall	Crop	Indicates natural water availability during the growing season
Soil Moisture	Fertilizer	Affects fertilizer dissolution and root nutrient absorption
Soil Type	Fertilizer	Reflects soil texture, drainage, and nutrient-holding capacity
Crop Type	Fertilizer	Specifies the target crop's nutrient requirement profile

TABLE 1: Input Feature Specification for Crop and Fertilizer Recommendation

C. Dataset Splitting and Random Forest Classification

After preprocessings and features selection each datasets is divided into mutually exclusives training and testing subset The trainings subset is used to teach the classifications model while there testings subsets is reserved exclusively to evaluates predictions accuracies on unseen records.

The Random Forest Classifier (implemented via Scikit-learn) is select as the primar predictives algorithm for both crop recommendations and forecasting fertilizers Using bootstrapped samples of the trainings data and randoms features selections at each nodes splits the Randoms Forests ensemble learnings techniques builds severals distincts decisions tree During predictions every decisions trees evaluates there submitted soil parameters and a class vote and the final recommended crops or fertilizers is determined by majority voting acrosses the entire ensembles Combining multiples trees reduceses variances prevents overfitting on correlated soils attributes and provides reliable multi classes classifications performances.

D. Model Evaluation

The trained crops and fertilizers classifications model are evaluated on there testings subsets usings fours standard multi classes evaluations metrics:

- Accuracy: Measures the overall proportions of corrects crops or fertilizers prediction acrosses all test samples.
- Precision: Measures how reliably the model avoids false positive recommendations for each target class.
- Recall: evaluates how well each crop or fertilizer category's true cases are identified by the model.
- F1-Score: Combines precision and recall into a single balanced performance measure across uneven class distributions.

E. Model Serialization and Web Deployment

Once evaluated, the trained Random Forests crops model fertilizer model and fitted categoricals labels encoders are saved to disks as serialized pkl files using Joblib Savings the trained artifact eliminats there need to retrains there models on every predictions requests.

To makes the systems accessibles to end users a webs applications is developed usings Flask for there backend and HTML/CSS for the frontend Users enters soil NPK values pH moisture environmental readings into cleans input field selected relevants soils or crop options from dropdown men and submits the form The Flasks servers loads the saved .pkl models, processes the submitted values, decodes the predicted label, and displays the recommended crop or fertilizer directly on the webpage.

IV. TECHNOLOGY USED

The crop and nourishment recommendation system is developed using an open source Python baseded softwares stacks that unified data preprocessing machines learning classifications models serializations and web deployments technologies is selecteded for computationals reliability eases of integrations and lightweights executions.

A. Core Programming Language and Data Processing

- Python: Serves as their primaries programmings languages for datas engineerings models trainings evaluations and backend web service.
- CSV (Comma-Separated Values): Used as the structured file formats to there store historicals soils nutrient, climatic, crop suitability fertilizer datasets in tabulars form.
- Pandas: Utilized to loaded CSV files into structured DataFrames inspect column schemas handled missings duplicates entries and organizes inputs features matrices and targets labels.
- NumPy: Provides efficients multi dimensional array operations and numericals data handling during preprocessing and models inference.

B. Machine Learning and Model Persistence

- Scikit-learn: Provides the complete supervised machine learning pipeline including Label Encoder encodings categorical attributes train_test_split for datasets partitioning RandomForestClassifiers for trainings both there crops and fertilizers model and evaluations utility for computings accurac precision recall and F1-score.
- Joblib: Used to serialize and save the trained Random Forest models and fitted label encoders as
- .pkl files, enabling the web server to load pre-trained models instantaneously without retraining on each request.

C. Web Framework, Frontend, and Visualization

- Flask: A lightweight Python web micro-framework used to implement the backend server, route HTTP requests, process user-submitted soil test inputs, execute model predictions, and return results dynamically.
- HTML (HyperText Markup Language): Used to structure the web interface, including numeric input fields for N, P, K, pH, moisture, and weather parameters, dropdown selectors for soil and crop types, and prediction output sections.
- CSS (Cascading Style Sheets): Used to style the frontend layout, ensuring clean form alignment, visual clarity, and responsive presentation for end users.
- Matplotlib: Used during offline analysis to visualize soil nutrient distributions, crop frequency counts, and model evaluation charts.

V. RESULTS AND DISCUSSION

A. Model Performance and Evaluation Analysis

The crop and nourishment commendation system was implemented usings Python, Pandas, NumPy, and Scikit-learn on historicals soils nutrients and environmentals datasets stored in CSV formats After removings duplicates and incompletes record standardizing numericals value and applying Labels Encodings to categoricals attributes such soil types crop types and target classes label both dataset were divideds into trainings and testings subset.

Separates Random Forest Classifiers model were traineds for crops suitability predictions and fertilizers recommendations During evaluations on the unseens testing subset both model were assesseded usings multi-class F1-Score Evaluating all fours metrics verifies not on overall predictions correctness but also balanced classifications reliability acrosses diverses crops species and fertilizer formulations By aggregating predictions across multiple decorrelated decisions trees via majority votings there Random Forest ensemble effectively captures non-linear thresholds among Nitrogen (N), Phosphorus (P), Potassium (K), soil pH, moisture, and climatic variables while avoiding the overfitting common in single decision trees.

B. Web Application Output and Dual Recommendation Workflow

Once trained and validated, the crop classifier fertilizer classifier and fitted label encoder were serialized susing Joblib and integrated into a Flask backend paired with an HTML and CSS frontends

Through the web interface users can access both recommendation modules seamlessly:

- Crop Recommendation Output: After filling outs structured inputs area with soils N, P, and K levels soil pH temperatures humidities and rainfalls users submits there forms The pre-loaded crops models receives the value from there Flasks servers and display the best crops for cultivations rights away.

- Fertilizer Recommendation Output: Users enter ambient temperature, humidity, soil moisture, N, P, K levels and seleced the soils types and target crops from dropdowns menus The backend encodes the categorical selections, invokes the fertilizer classifier, and renders the recommended fertilizer formulation directly on the webpage.

Alongside real-time web inference, exploratory visualization using Matplotlib illustrates nutrient distributions and crop class frequencies across the training data, helping confirm balanced feature coverage prior to deployment.

C. Practical Impact and Limitations

The implemented system demonstrates supervised Machine Learning can transform routine soil test measurements practical agronomic guidance Key benefits of the system include:

- Data-Driven Crop Selection: Helps farmers match crops accurately local soil fertility and weather conditions rathe relying solely on intuition.
- Balanced Fertilizer Usage: Recommends targeted fertilizers aligned with specific soil NPK levels and crop needs, reducing chemical overuse, lowering input costs, and protecting long term soil health
- Lightweight Accessibility: Serializing the trained models via Joblib enables rapid web-based predictions on standard computing hard ware without requiring cost softwares or high-end processing resources.

VI. CONCLUSION

This paper presented the design and implementation of an integrated Crop and Fertilizer Recommendation System using Machine Learning to support data-driven precision agricultures froms soils and environmentals parameter By systematically preprocessings historicals agricultural dataset including cleaning records standardizings numericals measurement and encoding categorical soil and crop attributes—the system prepares reliable feature sets for dual agronomic prediction Using the Random Forest Classifier evaluated across Accuracy Precision, Recall and F1-Score the framework accurately captures non near relationships among soil macronutrients (Nitrogen, Phosphorus, and Potassium), soil pH, soil moisture soil type and climatic conditions to recommends boths optimals crops and suitable fertilizers.

To ensure practical usability for farmers and agricultural advisors, the trained models and label encoders were serialized using Joblib and deployed through a lightweight Flask web application with an intuitive HTML and CSS interface. Users can input standard soil test measurements and receive immediate, actionable cultivation and fertilization recommendations, reducing reliance on guesswork, minimizing unnecessary fertilizer expenditure, and promoting long-term soil sustainability.

REFERENCES

1. K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine Learning in Agriculture: A Review," Sensors, vol. 18, no. 8, p. 2674, Aug. 2018.
2. 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, Aug. 2018.
3. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001.
4. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, Oct. 2011.
5. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY, USA: Springer, 2009.
6. C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
7. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees. Belmont, CA, USA: Wadsworth, 1984.
8. C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, Sep. 1995.
9. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, Aug. 2016, pp. 785–794.
10. L. Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp. 123–140, Aug. 1996.
11. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.