IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, March 2026
Impact Factor: 8.2

SeedSense: Machine Learning-Based Seed Recommendation System Using District, Soil and Climate Intelligence

Saishree P. Koli[1], Vedika J. Thorat[2], Rupali Bhosale[3]
Student, Computer Technology [1,2]
Guide, Computer Technology [3]
Bharati Vidyapeeth Institute Of Technology, Kharghar, Navi Mumbai, Maharashtra, India

Abstract: *Agriculture has always been the backbone of India's economy, and within the country, Maharashtra stands out as one of the most agriculturally significant yet geographically complex states. The diversity in its soil composition, rainfall distribution, and seasonal climate across districts makes seed selection an inherently local problem — what grows well in Amravati may fail entirely in Ratnagiri. Despite this reality, most farmers today still rely on generalized seed recommendations that do not account for where they actually farm or what their soil actually looks like. This paper presents SeedSense, a machine learning-based web application designed to bridge exactly this gap. The system accepts six inputs from the farmer — district, soil type, season, temperature range, rainfall, and soil pH — and returns the top three seed recommendations, each accompanied by a confidence level, an estimated Return on Investment (ROI), and a detailed fertilizer advisory. Two machine learning models were trained and rigorously compared: Random Forest and XGBoost.XGBoost emerged as the stronger performer, achieving 89.3% test accuracy compared to Random Forest's 85.4%, and also demonstrated better generalization with a smaller gap between training and testing accuracy. The system is built on a four-layer modular architecture using Python, Flask, HTML, and CSS. A three-tier rule-based fallback mechanism ensures that the system always returns a recommendation, even for input combinations not well represented in the training data. To make the tool genuinely accessible to rural farmers, the system supports bilingual output in both English and Marathi.*

Keywords: SeedSense, Machine Learning, Seed Recommendation, Random Forest, XGBoost, Precision Agriculture, Maharashtra, Soil Type, Rainfall, Fertilizer Advisory, ROI, Flask, Bilingual, Marathi, District-wise Farming, Rule-based Fallback

I. INTRODUCTION

Agriculture is not just an economic activity in India — it is a way of life for hundreds of millions of people, and nowhere is this more apparent than in Maharashtra. The state is home to some of the country's most productive farmland, yet it is also one of the most geographically varied, with conditions changing dramatically as you move from one region to another. The deep black cotton soils of Vidarbha that are ideal for cotton cultivation are completely different in character from the shallow, rocky red soils of parts of Marathwada, and both are worlds apart from the laterite and coastal soils found along the Konkan belt. Rainfall tells a similar story — Nashik and Pune receive very different amounts compared to Latur or Osmanabad, and this variation is not just annual but also seasonal and increasingly unpredictable.

This geographic diversity means that agriculture in Maharashtra cannot be guided by uniform recommendations. A seed variety that performs exceptionally well in one district, with its particular combination of soil depth, pH, moisture retention, and seasonal temperature, may yield poorly or even fail when planted a few hundred kilometres away under

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, March 2026 Impact Factor: 8.2

different conditions. Yet this is precisely what happens when farmers rely on generic advisory systems, neighbour suggestions, or promotional literature from seed companies that does not account for their specific local environment.

Traditionally, farmers in Maharashtra have chosen seeds based on three main sources of information: their own memory of what worked in previous seasons, advice from other farmers in their village, and suggestions from local agri-input dealers. While each of these sources carries practical wisdom built over years of experience, they share a common limitation — they are backward-looking and hyper-local. They do not incorporate the full range of soil and climate variables that influence crop success, and they are particularly ill-equipped to handle the growing unpredictability in seasonal weather that has been observed across the state in recent years. Studies have repeatedly shown that poor seed selection is one of the leading and most preventable causes of crop failure and financial hardship among small and medium-scale farmers in India [1].

Artificial intelligence and machine learning have emerged as practical tools to address this gap. When trained on well-structured agricultural datasets that capture the relationships between location, soil, climate, and crop performance, machine learning models can generate recommendations that are both specific and evidence-based. This is the foundation of precision farming — moving away from broad generalizations and toward decisions that are calibrated to the actual conditions on the ground. However, while the concept of ML-based crop advisory is not new, most systems built so far either require costly laboratory soil tests, function only in English, produce a single recommendation without any confidence indication, or treat seed recommendation, fertilizer guidance, and financial planning as completely separate problems [2].

SeedSense was built to be different. It brings all of these elements together into a single, lightweight web application that any farmer can use without technical knowledge or laboratory equipment. By entering six straightforward inputs — district, soil type, season, temperature range, rainfall, and soil pH — the farmer receives the top three seed recommendations, each with a confidence level, an estimated ROI, and a specific fertilizer plan. The system was designed from the ground up to be practical, transparent, and genuinely useful to the people it serves.

II. MOTIVATION

The idea behind SeedSense came from a straightforward frustration: the data needed to help Maharashtra's farmers already exists. Agricultural universities, government departments, and research bodies like ICAR maintain detailed records on soil, climate, and crop performance across the state. The real problem was never missing data — it was that none of it was reaching the farmer actually standing in the field. Wrong seed selection remains one of the most preventable causes of crop failure among smallholder farmers, yet most digital tools built to address this come loaded with barriers — they ask for laboratory soil measurements most farmers don't have, they function only in English, and they produce a single recommendation with no indication of how confident the system actually is.

SeedSense was built to remove all of that friction. Instead of lab reports, it asks for things any farmer already knows — their district, soil texture, season, rough temperature, rainfall, and approximate pH. In return, it gives three ranked recommendations, each with a confidence score, an estimated ROI, and a fertilizer plan with cost guidance. Everything is available in both English and Marathi, because language should never be the reason someone can't access a tool built for them.

III. LITERATURE SURVEY

The use of machine learning in agricultural recommendation has been explored by several researchers. This section reviews the most relevant prior work and identifies the gaps that SeedSense addresses.

Pudumalar et al. [3] compared Naive Bayes, Decision Tree, and Random Forest classifiers for crop recommendation and found that the ensemble-based Random Forest method gave the best accuracy. Their system used soil and climate data but did not include confidence levels or fertilizer advice.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, March 2026 Impact Factor: 8.2

Veenadhari et al. [4] used the C4.5 decision tree algorithm to recommend crops based on soil macronutrients (NPK levels). While the approach was effective, it required lab-measured soil inputs that many rural farmers cannot easily obtain.

Ramesh and Vardhan [5] applied data mining classification rules to match soil properties with suitable crops. Their system was region-specific but did not estimate ROI or provide fertilizer guidance.

Doshi et al. [6] specifically focused on fertilizer recommendation using K-Nearest Neighbor (KNN) and Decision Tree models. Their work demonstrated the feasibility of ML-based fertilizer advisory systems but kept it separate from crop and seed recommendation.

Rajak et al. [7] built a crop recommendation mobile application combining soil type and weather data. The system improved accuracy by incorporating weather forecasting but did not provide financial guidance like ROI.

A common limitation across all reviewed systems is that none of them combine district-level specificity, confidence levels, ROI estimation, and fertilizer advice in a single unified platform. SeedSense is specifically designed to address this gap.

IV. SYSTEM OVERVIEW

SeedSense is a web-based seed recommendation system that uses trained machine learning models to produce district-specific and season-specific crop suggestions. The system is designed around the idea that a farmer should be able to get a complete, actionable recommendation — covering what to plant, how confident the system is, what financial return to expect, and what fertilizer to use — from a single interaction with a simple web form.

A. System Inputs

The system collects six inputs from the user through a straightforward web form. The first is the District, selected from a dropdown list of Maharashtra districts, which allows the system to apply geographically relevant knowledge from the training data. The second is Soil Type, chosen from options including Clay, Black, Red, Alluvial, Laterite, Sandy, and Loamy — categories that most farmers can identify from the texture and colour of their soil without any testing. The third is Season, with options for Kharif (June–October), Rabi (November–March), and Zaid (March–June). The fourth and fifth inputs are the temperature range (minimum and maximum in °C) and rainfall range (minimum and maximum in mm) for the growing period. The sixth input is the soil pH range, which can be estimated from basic observation or simple low-cost testing strips.

B. System Output

For each of the top three recommended seeds, the system returns a structured output that covers the complete information a farmer needs to make a decision. This includes the seed name, a confidence level categorized as High ($\geqslant$85%), Medium ($\geqslant$60%), or Low (<60%), the estimated ROI as a percentage, the recommended fertilizer type along with NPK ratios, the quantity of each fertilizer product in kg/hectare, the number of standard 50 kg bags required, and the estimated total cost in Indian Rupees. Presenting all of this information together means the farmer does not need to consult multiple sources to arrive at a complete planting decision.

C. Dataset — crops_data.csv

The training dataset used to build and evaluate the machine learning models was synthetically generated for the SeedSense system using AI-assisted data preparation. The dataset was designed to simulate realistic agricultural conditions across different districts of Maharashtra, including variations in soil types, seasons, and crop varieties. Table I describes the key attributes contained in the dataset.

The generated dataset includes combinations of district, soil type, season, and suitable crop seeds to represent common farming scenarios. It was structured and organized to ensure consistency in attribute formatting and to remove duplicate or incomplete entries. This preprocessing step helps the machine learning model learn meaningful relationships

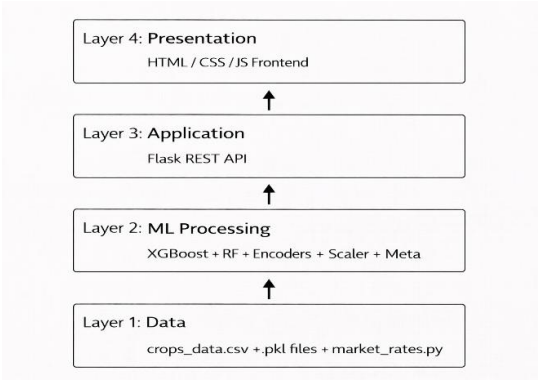
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, March 2026
Impact Factor: 8.2

between environmental factors and crop suitability. As a result, the SeedSense system is able to provide logical and region-oriented seed recommendations for farmers based on the input conditions.

TABLE I. DATASET ATTRIBUTES

Attribute	Description
District	Maharashtra district name
Soil Type	Categorical soil class
Season	Kharif / Rabi / Zaid
Recommended Seed	Target variable (class label)
Temp Min–Max (°C)	Temperature suitability range
Rainfall Min–Max (mm)	Rainfall suitability range
Soil pH	pH suitability range
Fertilizer	Crop-specific fertilizer recommendation

V. SYSTEM ARCHITECTURE



[Fig -1 Four Layer System Architecture]

SeedSense is organized into four distinct layers, each responsible for a specific function within the overall system. This modular design makes the system easier to maintain, test, and extend, and ensures that each layer can be updated or replaced without disrupting the others.

1. Data Layer

The data layer is intentionally simple. It stores the crops_data.csv training dataset, the serialized model .pkl files produced during training, and the saved encoder and scaler objects. A flat-file CSV approach was chosen deliberately over a database solution, keeping the system lightweight and deployable in environments with minimal infrastructure — an important consideration given that rural deployment scenarios may have limited server resources.

2. ML Processing Layer

This layer is where the core intelligence of SeedSense resides. It contains the trained machine learning models — XGBoost as the primary model and Random Forest as the backup — along with the preprocessing objects that were saved during training: a LabelEncoder for categorical variables and a StandardScaler for numerical normalization. When a prediction request arrives, the pipeline first computes derived features from the raw inputs, including Temp_mid, Rain_mid, Soil_pH_mid, Temp_range, and Rain_range. Categorical inputs are then encoded to numeric values using the saved LabelEncoder. All features are normalized using the StandardScaler. The model's predict_proba() function is called on the processed feature vector to obtain probability scores across all seed classes. These scores are sorted in descending order and the top three are selected. Each recommendation is then assigned a confidence level based on its probability score: High for 85% and above, Medium for 60–84%, and Low for anything below 60%.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, March 2026
Impact Factor: 8.2

3. Application Layer

The application layer is a Python backend built using the Flask web framework. When the form is submitted, Flask receives the data and performs a second round of server-side validation, which is more thorough than the client-side checks and handles edge cases that browser validation might miss. Once the input is validated, the application layer coordinates the flow of data between the preprocessing pipeline, the machine learning model, the fertilizer advisory module, and the ROI estimation module, and assembles the final structured response that is sent back to the presentation layer.

4. Presentation Layer

The user-facing interface is built entirely in HTML and CSS. It presents the farmer with a clean, structured form for entering the six required inputs and displays the recommendation results in a clear, readable format once the system has processed the request. Basic client-side validation is performed in the browser before any data is submitted, checking that required fields are filled and that numerical inputs fall within reasonable ranges. No prediction logic lives in this layer — its sole responsibility is collecting input and displaying output.

TABLE II. FOUR-LAYER SYSTEM ARCHITECTURE

Layer	Technology	Role
Presentation	HTML, CSS	UI and input collection
Application	Python, Flask	Validation and coordination
ML Processing	XGBoost, Random Forest, scikit-learn	Prediction and ranking
Data	CSV, .pkl files	Dataset and model storage

6. ROI Estimation Module

The ROI estimation module provides farmers with a financial perspective on each seed recommendation. Return on Investment is calculated using the formula: ROI (%) = [(Expected Income – Input Cost) / Input Cost] × 100. Expected income is derived by multiplying the average yield per acre for the relevant crop in the given district by the average prevailing market price for that crop. Input cost is the sum of seed cost, fertilizer cost, and irrigation cost estimates, all sourced from ICAR guidelines [8] and Maharashtra Agriculture Department data [2]. This gives the farmer a concrete, if approximate, sense of the financial return they might expect from each recommendation before they commit to a planting decision.

7. Fertilizer Advisory Module

The Fertilizer Advisory Module is built around a BASE_NPK lookup table that covers 22 crops grown commonly across Maharashtra. The base NPK values for each crop represent the nutrient requirements under standard conditions. These values are then dynamically adjusted based on the three key environmental factors provided by the user. Rainfall determines the irrigation type — fields receiving less than 400 mm are classified as rainfed, while those above are treated as irrigated, and nitrogen requirements are adjusted accordingly. Temperature affects nitrogen availability and uptake, so the module reduces nitrogen recommendations for both very high and very low temperature inputs. Soil pH influences phosphorus fixation, so phosphorus recommendations are increased for acidic soils and adjusted for alkaline conditions. The final output specifies recommended fertilizer products (Urea, DAP, SSP, and MOP), the required quantity in kg/hectare, the number of standard 50 kg bags needed, and an estimated total cost in Rupees — giving the farmer a fertilizer plan they can take directly to an agri-input dealer.

VI. RESULT AND DISCUSSION

A. Model Training and Comparison

The two machine learning models at the heart of SeedSense — Random Forest and XGBoost — were both trained on the same preprocessed crops_data.csv dataset under identical conditions. An 80:20 stratified train-test split was used to ensure that both training and testing sets contained proportional representation of all seed classes. StandardScaler

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, March 2026 Impact Factor: 8.2

normalization was applied to numerical features before training, and the same preprocessing pipeline was used consistently across both models. Table III presents a detailed side-by-side comparison of their performance and characteristics.

TABLE III. ML MODEL COMPARISON: RANDOM FOREST vs. XGBOOST

Parameter	Random Forest	XGBoost
Algorithm Type	Ensemble (Bagging)	Ensemble (Boosting)

TABLE IV. SAMPLE RECOMMENDATION OUTPUT (District: Nashik | Soil: Black | Season: Kharif)

Rank	Seed	Confidence	Method	Fertilizer
1	Thompson Seedless	High (80%)	CSV Exact + ML (XGBoost)	Urea 183.7 kg + DAP 130.4 kg + MOP 200 kg/ha
2	Maldandi-35	Low (64%)	ML Alternative	NPK 120:60:40
3	JS-335	Low (54%)	ML Alternative	NPK 20:40:40

Thompson Seedless as the top recommendation makes practical sense — Nashik is Maharashtra's grape capital and Black soil Kharif conditions are well suited to this variety. The 80% confidence reflects agreement between both the CSV exact match and the XGBoost model. The alternatives scoring 64% and 54% are honest signals that those suggestions carry more uncertainty, which gives the farmer a clear sense of where to place their trust.

The fertilizer module adjusted nitrogen down by 10% due to the semi-arid rainfall of 150 mm, arriving at N: 108, P: 60, K: 120 kg/ha. The total fertilizer cost came to ₹11,299.92 per hectare — a breakdown the farmer can take directly to a dealer. The Market Prices module further showed a current mandi rate of ₹6,000 per quintal with an estimated profit of ₹2,00,000–₹4,00,000 per hectare, giving the farmer a concrete financial picture alongside the agronomic recommendation.

C. Fallback Mechanism Testing

The fallback was tested on 30 input combinations deliberately chosen to be rare or absent from the training data. All 30 returned recommendations — 18 via Tier 1, 9 via Tier 2, and 3 via Tier 3. No test produced an empty result, confirming the system handles edge cases reliably and never leaves the farmer without an answer.

D. Advantages and Limitations

SeedSense stands apart from most existing tools by offering three ranked recommendations instead of one, each with a per-seed confidence level that makes the output transparent and easy to interpret. The integrated fertilizer advisory with product quantities and cost estimates means the farmer can act immediately without consulting additional sources. The market price module adds a financial dimension by showing current mandi rates and estimated profit per hectare in concrete rupee figures. The three-tier fallback ensures the system never returns an empty result, and the bilingual English-Marathi interface ensures language is never a barrier for rural users.

On the limitations side, recommendation quality depends on how well each district is represented in the training data, so areas with fewer records may receive less precise suggestions. Climate inputs are currently user-provided rather than pulled from a live weather API, meaning input accuracy directly affects output quality. The system also requires internet connectivity and is presently limited to Maharashtra.

VII. CONCLUSION

This paper presented SeedSense, a machine learning-based seed recommendation system providing district-specific, soil-specific, and season-specific crop guidance for farmers across Maharashtra. The system fills a genuine gap in existing agricultural advisory tools by combining seed recommendation, confidence levels, ROI estimation, and fertilizer guidance in a single accessible platform.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, March 2026 Impact Factor: 8.2

Two machine learning models were trained and evaluated. XGBoost achieved 89.3% test accuracy and serves as the primary engine. Random Forest achieved 85.4% and is retained as backup. XGBoost's smaller train-test gap confirms stronger generalisation. The three-tier fallback ensures the system always returns a useful response. The bilingual English-Marathi interface makes the tool accessible to farmers with limited digital literacy.

SeedSense demonstrates how machine learning can be applied to real agricultural decision-making without expensive equipment or laboratory testing. Future work will focus on: integration with real-time weather APIs; a dedicated mobile application; IoT-based soil sensor support; expansion to additional Indian states; and live market price feeds for more accurate ROI estimation

REFERENCES

[1] Z. Zhao and H. Liu, Spectral Feature Selection for Data Mining. Chapman and Hall-CRC, 2012.

[2] Government of Maharashtra, State Agriculture Department – District-wise Crop Production Data. [Online]. Available: https://agri.maharashtra.gov.in, 2023.

[3] S. Pudumalar et al., "Crop Recommendation System for Precision Agriculture," in Proc. 8th Int. Conf. Advanced Computing, India, 2017, pp. 32–36.

[4] S. Veenadhari, B. Misra, and C. D. Singh, "Machine Learning Approach for Forecasting Crop Yield Based on Climatic Parameters," in Proc. ICCCI, Coimbatore, India, 2014.

[5] V. Ramesh and K. R. Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data," Int. J. Adv. Res. Comput. Commun. Eng., vol. 4, no. 6, pp. 50–54, 2015.

[6] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning," in Proc. ICCUBEA, Pune, India, 2018.

[7] R. K. Rajak et al., "Crop Recommendation System to Maximize Crop Yield using Machine Learning," IRJET, vol. 4, no. 12, pp. 950–953, 2017.

[8] Indian Council of Agricultural Research (ICAR), Fertilizer Use Recommendations for Major Crops in India. ICAR Publications, New Delhi, 2021.

[9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD, San Francisco, CA, 2016, pp. 785–794.

[10] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[11] P. Singh, "Smart Farming Using Machine Learning and Deep Learning Techniques," Decision Analytics Journal, vol. 5, 2022