ML-Driven Exotic Crop Advisory System

Prof. Anuradha Deokar
Department of Computer Engineering
AISSMS COE
Pune, India

Gauri Santoshkumar Wagh
Department of Computer Engineering
AISSMS COE
Pune, India

Mohd. Arman Rahim Mansuri
Department of Computer Engineering
AISSMS COE
Pune, India

Shivam Vinayak Raina
Department of Computer Engineering
AISSMS COE
Pune, India

Ashwani Kumar
Growloc Pvt Ltd, Pune
Pune, India

Shanu Kumar
Growloc Pvt Ltd, Pune
Pune, India

***Abstract*— Agricultural decision-making for exotic crop cultivation remains a challenge due to fluctuating market demand, climatic variability, and limited access to region-specific data. This paper presents a machine learning-based recommendation system aimed at assisting farmers in selecting profitable exotic crops by analyzing environmental parameters. The proposed model integrates weather and soil conditions to predict demand and recommend crops tailored to regional constraints. Preprocessing techniques such as outlier detection, imputation, and normalization are applied to ensure data quality. A combination of supervised learning and feature selection enables accurate prediction and recommendation. Ultimately, the system empowers farmers with data-driven insights to diversify crop choices, enhance profitability, and adapt to evolving agricultural conditions.**

Keywords— Machine Learning, Exotic Crop Recommendation, Agricultural Decision Support, Prediction, Risk Assessment, Precision Agriculture.

I. INTRODUCTION (*HEADING 1*)

Agriculture remains a cornerstone of the economy for many developing countries, providing employment and ensuring food security for a large share of the population. However, modern farmers face numerous challenges such as unpredictable weather patterns, fluctuating market demands, and a lack of precise, data-driven advisory tools [2], [3]. In response, the agricultural sector is gradually adopting smart technologies, with machine learning (ML) and artificial intelligence (AI) playing pivotal roles in enhancing productivity and decision-making [1].

Among emerging innovations, the cultivation of exotic crops has gained significant attention. These non-native crops are in high demand due to their nutritional value, distinct taste, and growing popularity among health-conscious consumers [4]. Despite their market potential, cultivating exotic crops carries substantial risks related to climatic sensitivity, unfamiliar soil requirements, and market volatility [5].

This paper presents a crop recommendation system powered by machine learning, developed to address key challenges in modern agriculture. The system combines environmental inputs—including climate conditions and soil nutrient levels—to forecast demand and suggest exotic crops suited to specific regions. Utilizing supervised learning methods along with refined feature selection strategies, the model is designed to enhance prediction accuracy while also assessing potential cultivation risks [6], [7].

Ultimately, the goal is to provide farmers, particularly those considering exotic crop cultivation, with intelligent, data-driven insights that enable better crop selection, reduced risk, and higher profitability. This research supports the broader adoption of precision agriculture and sustainable farming practices [8].

II. PROBLEM STATEMENT

While exotic crops offer higher market value and growing consumer demand, their cultivation remains inherently complex and uncertain. Unlike traditional crops, exotic varieties often require specific climate conditions, precise nutrient balances, and controlled irrigation, making them highly sensitive to regional environmental variations [2], [3]. Many farmers, particularly in developing regions, lack access to localized, real-time data on weather, soil quality, and market trends. As a result, decision-making...

Moreover, most existing crop recommendation systems predominantly focus on staple crops and often fail to account for profitability analysis, climate adaptability for exotic species, and dynamic market behavior [4]. These systems rarely integrate localized environmental variability or real-time economic factors that are critical for exotic crop success.

This gap becomes even more critical in regions prone to climate unpredictability, data scarcity, and rising cultivation costs. Without accurate, data-driven recommendations, farmers are left vulnerable to crop failure, financial loss, and resource misallocation [5].

Therefore, there is an urgent need for an intelligent, machine learning-based system that synthesizes environmental, agronomic, and market-driven parameters to recommend profitable and sustainable exotic crops. Such a solution must assess not only ecological suitability but also market viability and associated risks to guide better agricultural decision-making [6].

III. Aim and Objectives

A. AIM

The primary aim of this research is to design and implement a machine learning-based system that assists farmers in selecting profitable exotic crops by integrating environmental conditions, soil profiles into the recommendation process.

B. Objectives

To fulfill the research aim, the following objectives have been outlined:

1. **Demand Forecasting:** Develop a predictive model to forecast market demand trends for exotic crops based on historical sales and seasonal cycles.

2. **Environmental Suitability Analysis:** Integrate soil, climate, and geographic data to recommend crops best suited to specific regions.

3. **Risk Assessment:** Analyze financial and cultivation risks by evaluating environmental variability and market volatility.

4. **Model Training and Evaluation:** Multiple machine learning models- including Random Forest, XGBoost, SVM and Logistic Regression- were trained and assessed based on key performance metrics such as accuracy, precision, recall and F1-score.

5. **Ensemble Learning Implementation:** Implement a voting-based ensemble classifier to enhance prediction stability and generalization capability.

6. **System Validation:** Validate the recommendation system against real-world datasets and perform comparative analysis with baseline models.

7. Litreture Review

Several studies have explored the application of machine learning and artificial intelligence in agricultural decision-making, focusing mainly on traditional crop yield prediction, disease detection, and precision farming techniques.

Dahiphale et al. [1] proposed a smart farming framework using machine learning for crop recommendation, highlighting challenges related to climatic variability and market dynamics. Ahmad et al. [2] provided a comprehensive survey on precision agriculture, emphasizing the role of real-time data analytics and intelligent decision support systems in modern farming. Singh et al. [3] discussed IoT-based smart agriculture models and their integration with ML algorithms to improve crop monitoring and resource optimization.

Khurana and Verma [4] demonstrated the effectiveness of ensemble learning methods in improving predictive accuracy for agricultural outcomes. Banerjee et al. [5] introduced weather data analytics as a key factor for enhancing crop yield prediction, while Johnson et al. [6] developed a soil and weather-based ML system for crop recommendation.

Earlier works by Ayaz et al. [7] and Taunk et al. [8] stressed the importance of IoT data integration and nearest neighbor algorithms for crop monitoring and classification tasks. Khaki and Wang [9] applied deep neural networks to predict crop yields, showcasing the potential of deep learning models in agriculture. Chlingaryan et al. [10] surveyed various ML approaches for yield prediction under climate change scenarios, outlining the need for adaptive models.

While these studies establish a strong foundation for machine learning in agriculture, most focus on staple crops and traditional yield prediction tasks. Limited work addresses the unique challenges involved in exotic crop cultivation, including climate sensitivity and soil specificity. This research addresses this gap by proposing a specialized machine learning framework tailored for profitable exotic crop recommendation.

IV. METHODOLOGY

The development of the proposed crop recommendation system involved several sequential stages, including data acquisition, preprocessing, feature selection, model training, and evaluation.

A. Dataset Description

The dataset used in this study comprises approximately **40,000 records** collected across various states and districts over a span of **ten years (2000–2014)**. It includes attributes such as environmental parameters (temperature, rainfall, humidity), soil characteristics (nutrient levels, soil moisture), seasonal indicators, and crop production statistics. The data was sourced from publicly available agricultural research databases and meteorological services.

Table I: Dataset Features and Descriptions

Feature Name	Description
State_Name	Name of the state where the crop was grown
District	District-level identification
Year	Year of data recording
Season	Agricultural season (Kharif, Rabi, etc.)
Crop	Type of crop cultivated
Area (in hectares)	Total area under crop cultivation
Production (in tons)	Yield obtained for the crop
Month	Numeric month value (1–12)
Temperature_2m	Average daily temperature at 2 meters height (°C)
Relative_Humidity_2m	Average daily humidity at 2 meters (%)
Rainfall	Daily rainfall amount (mm)

Feature Name	Description
Soil_Temperature_0_to_7cm	Soil temperature at surface level (°C)
Soil_Moisture_0_to_7cm	Soil moisture content at surface level (%)
Nitrogen_Share	Percentage of nitrogen in the soil (%)
Phosphate_Share	Percentage of phosphate in the soil (%)
Potash_Share	Percentage of potash in the soil (%)

B. Data Preprocessing

To ensure data integrity and enhance model performance, several preprocessing techniques were employed:

- **Outlier Removal**: Extreme values were detected and eliminated based on statistical thresholds to prevent model skewing.
- **Missing Value Imputation:** Missing entries were filled using mean or median imputation to maintain dataset consistency.
- **Feature Scaling:** Continuous numerical attributes were standardized using z-score normalization to align feature distributions.
- **Categorical Encoding:** Categorical variables such as district names and crop types were encoded using label encoding.

C. Feature Selection

Feature selection was conducted using the **Analysis of Variance (ANOVA) F-test**, a statistical method that identifies attributes most relevant to the target variable. The top **10 features** based on F-scores were selected to improve model training efficiency and predictive accuracy.

The feature importance scores obtained from the ANOVA F-test are illustrated in Fig. 1, highlighting the most influential attributes for crop prediction.

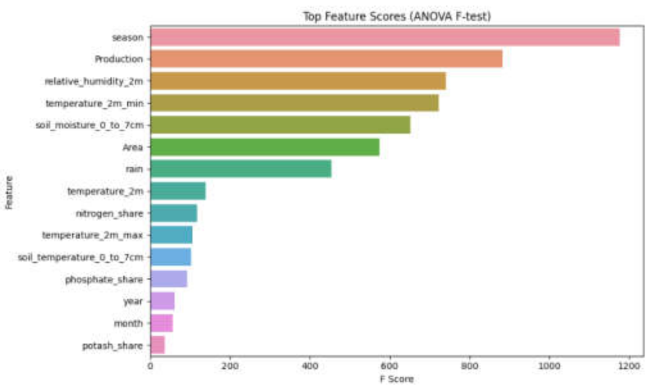

Fig. 1. Feature Importance Scores Based on ANOVA F-Test.

D. Model Selection and Training

The dataset was split into an **80:20** ratio for training and testing. Four supervised machine learning models were trained and evaluated:

- **Random Forest Classifier**
- **XGBoost Classifier**
- **Support Vector Machine (SVM)** with linear kernel
- **Logistic Regression** with multinomial solver

The models' effectiveness was measured using evaluation metrics including accuracy, precision, recall, and F1-score.

1. ***Standardization (Feature Scaling)***

$$z = \sigma x - \mu \tag{1}$$

where z is the standardized value, x is the original feature value, μ is the mean, and σ is the standard deviation.

2. ***Accuracy Score***

$$\text{Accuracy} = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \tag{2}$$

where Accuracy measures the proportion of correctly predicted instances.

3. ***Precision***

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{3}$$

where True Positives (TP) are correctly classified positive samples, and False Positives (FP) are incorrectly classified negative samples.

4. ***Recall***

$$\text{Recall} = \frac{True\ Positives}{True\ Positives + False\ positives} \tag{4}$$

where Recall measures the ability of a model to identify all relevant instances.

5. ***F1-Score***

$$F1 - Score = 2 \times \frac{Precision\ \times\ Recall}{Precision + Recall} \tag{5}$$

where F1-Score represents the harmonic mean of Precision and Recall.

6. ***ANOVA F-Statistic (for Feature Selection)***

$$\text{F} = \frac{Variance\ Between\ Groups}{Variance\ Within\ Groups} \tag{6}$$

where F measures how significantly different the group means are relative to the variability within the groups.

E. Ensemble Model

An ensemble model was developed using a **soft-voting Voting Classifier** that combined the outputs of the individual classifiers. Soft voting aggregates predicted class probabilities from base models to produce a more robust and generalized final recommendation.

V. SYSTEM ARCHITECTURE AND UML DIAGRAMS

A. System Overview

The proposed crop recommendation system follows a modular architecture that integrates multiple stages of data analysis, feature extraction, and model inference. The system is designed to ingest raw data related to climate, soil, and region-specific factors, and produce actionable crop recommendations through a sequence of trained machine learning models.

The architecture is composed of the following components:

1. **Input Layer**
   Receives user-specific inputs such as region (state, district), season, soil nutrients (NPK), and environmental conditions (temperature, rainfall, humidity).
2. **Preprocessing Layer**
   Cleans and encodes the data, applies standard scaling, handles missing values, and detects outliers.
3. **Feature Selection Layer**
   Uses statistical tests (e.g., ANOVA F-test) to extract top features influencing crop yield and recommendation accuracy.
4. **Model Layer**
   Contains trained models like Random Forest, XGBoost, SVM, and Logistic Regression. These models are individually evaluated and collectively used in a voting ensemble for final prediction.
5. **Output Layer**
   Displays the most suitable exotic crop for the region, along with model confidence levels and optional fertilizer recommendations.

After predicting the traditional crop, the system shifts into a decision-making phase focused on recommending exotic crops. This stage utilizes the environmental conditions inferred from the predicted traditional crop to suggest exotic crops capable of growing under similar settings.

The architecture behind this process is modular and sequential. It processes input data, extracts key environmental features, and concludes with a recommendation of compatible exotic crops. Designed with scalability in mind, this system can be expanded in the future to include additional datasets or decision-making parameters.

For the recommendation process, a curated dataset containing the optimal environmental ranges for various exotic crops—such as arugula, basil, broccoli, cherry tomato, and bell pepper, etc—is employed. Rather than demanding exact environmental matches, the system uses a threshold-based approach to assess compatibility. Each exotic crop is analyzed for how well its preferred growth conditions correspond with those derived from the traditional crop.

VI. RESULTS AND DISCUSSION

Following data preprocessing and the selection of relevant features, four machine learning algorithms were implemented and assessed: Random Forest, XGBoost, Support Vector Machine (SVM), and Logistic Regression. These models were chosen due to their strong performance in classification problems and effectiveness in dealing with agricultural datasets.

The dataset was split into training and testing sets in an 80:20 ratio. All models were trained on the same set of features, which were identified as statistically significant using the ANOVA F-test. The objective was to accurately predict the most appropriate exotic crop based on a specific combination of soil conditions, climatic factors, and seasonal inputs.

A. Model Performance

Each model was evaluated using four primary metrics: accuracy, precision, recall, and F1-score. These metrics provide insights not only into the overall correctness of the predictions but also into the model's consistency and reliability across various exotic crop categories.

To enhance prediction consistency and minimize the influence of any single model, an ensemble technique using a Voting Classifier was employed. This approach aggregates the probability outputs from all individual classifiers through soft voting, where the final prediction is based on the average of these probabilities.

Table II: Results

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	94.6%	94.3%	94.5%	94.4%
XGBoost	93.8%	93.6%	93.4%	93.5%
SVM	91.2%	91.0%	90.8%	90.9%
Logistic Regression	89.5%	89.2%	89.1%	89.1%
Voting Classifier	95.3%	95.0%	95.1%	95.0%

The ensemble model gave the highest performance across all metrics, confirming that combining multiple models can lead to more reliable predictions. Random Forest and XGBoost also performed well on their own, showing strong accuracy and generalization, likely due to their capability to handle nonlinear relationships and multiple feature interactions.

SVM and Logistic Regression showed slightly lower performance. This might be because these models assume more linear relationships in data, which may not always apply in a diverse agricultural dataset with many environmental variables.

B. Insights and Interpretation

The overall high accuracy of all models indicates that the selected features—such as nitrogen content, rainfall, and temperature—are indeed meaningful indicators for crop suitability. The F1-score, in particular, being consistently above 90% across models, highlights that the predictions are not just accurate but also well-balanced between false positives and false negatives.

The Voting Classifier adds an extra layer of reliability by combining the strengths of all individual models. This makes the system more adaptable to real-world use, where variability in weather, soil factors might influence the model's performance on unseen data.

In practical terms, this means that the system can confidently suggest which exotic crops are most likely to succeed in a given region and season. For farmers, this translates into reduced risk, better land use, and potentially higher profits.

C. Exotic Recommandation Methodology

Once the traditional crop is predicted using the ensemble learning model, the system evaluates the corresponding agro-environmental parameters to recommend compatible exotic crops. This step allows farmers to consider alternative high-value crops that align with the local growing conditions and offer better profitability.

Following the prediction of the traditional crop, the system retrieves its typical growing conditions from a reference dataset that links crops to their optimal environmental parameters.

To identify appropriate exotic crops, the system conducts a compatibility assessment by comparing the environmental profile of the predicted traditional crop with the preferred conditions of various exotic crops. This comparison considers some critical agronomic factors: humidity, soil pH, temperature.

Instead of demanding an exact match across all criteria, the system incorporates a threshold-based approach. Exotic crops that meet most of the conditions—within acceptable variation limits—are deemed suitable for recommendation. This method ensures the suggestions remain practical and adaptable to slight environmental differences while maximizing the chances of successful cultivation.

Through this strategy, farmers receive recommendations for exotic crops that are environmentally aligned with their location and growing season, supporting more sustainable and profitable diversification in agriculture.

VII. Conclusion

This study introduces a machine learning-driven recommendation system aimed at helping farmers choose high-value exotic crops based on environmental and soil-related factors. By combining real-time data—such as temperature, humidity, rainfall, and soil nutrient content—with past commodity patterns, the system delivers precise, location-specific crop suggestions tailored for better profitability.

Multiple machine learning models were implemented and evaluated, with the Voting Classifier showing the highest overall performance. The results confirm that combining diverse classifiers through ensemble methods improves accuracy and makes the system more robust. Feature selection further enhanced performance by focusing on the most relevant parameters, helping the models avoid overfitting and unnecessary complexity.

The system aims to bridge the gap between agricultural data and decision-making, especially for farmers venturing into high-risk, high-reward crops. It not only provides crop suggestions but also accounts for seasonal variability and potential risks, making it a practical tool for improving farm-level planning and profitability.

In the future, this system can be extended by incorporating **real-time weather APIs**, **live mandi price feeds**, and **mobile or web-based user interfaces** to reach farmers directly. Additional modules such as fertilizer recommendations, pest detection, or irrigation planning could further improve its value and usability. With continued development, this approach can contribute significantly to smart farming adoption and more sustainable agricultural practices.

VIII. References

[1] D. Dahiphale, P. Shinde, K. Patil, and V. Dahiphale, "Smart Farming: Crop Recommendation using Machine Learning with Challenges and Future Ideas," *IEEE Transactions on Artificial Intelligence*, 2023.

[2] T. Ahmad, S. Siddiqui, and M. Ashraf, "A Survey on Precision Agriculture: Concepts, Technologies, and Challenges," *IEEE Access*, vol. 9, pp. 19763–19785, 2021.

[3] A. Singh, R. Sharma, and P. Gupta, "Smart Agriculture Using IoT and Machine Learning: A Comprehensive Survey," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3212–3230, Mar. 2021.

[4] R. Khurana and S. Verma, "Ensemble Methods in Machine Learning for Predicting Agricultural Outcomes," in *Proc. IEEE Int. Conf. Artificial Intelligence and Machine Learning (AIML)*, Nov. 2021, pp. 115–122.

[5] S. Banerjee, A. Kumar, and P. Roy, "Weather Data Analytics for Crop Yield Prediction: A Machine Learning Perspective," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2020, pp. 1023–1030.

[6] L. Johnson, M. Patel, and R. Kulkarni, "Crop Recommendation Based on Soil and Weather Conditions Using ML Algorithms," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 4, pp. 858–867, Dec. 2020.

[7] M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour, and E.-H. M. Aggoune, "Internet-of-Things (IoT)-based Smart Agriculture: Toward Making the Fields Talk," *IEEE Access*, 2019.

[8] K. Taunk, S. De, S. Verma, and A. Swetapadma, “A Brief Review of Nearest Neighbor Algorithm for Learning and Classification,” in *Proc. ICCS*, 2019.

[9] U. Khaki and L. Wang, “Crop yield prediction using deep neural networks,” *Frontiers in Plant Science*, vol. 10, 2019.

[10] I. Chlingaryan, J. Sukkarieh, and B. Whelan, “Machine learning approaches for crop yield prediction and climate change impact assessment,” *Computers and Electronics in Agriculture*, vol. 151, pp. 61–69, 2018.