# Crop Recommendation System Based On High Yield Using Machine Learning Techniques

**Maria Sobana S[1], Shyla Shree M M[2], Susmitha M[3], Swetha M[4]**

[1] *Assistant Professor and Faculty Mentor, Dept of Computer Science and Engineering, K L N College of Engineering, Pottapalayam, Sivagangai, Tamil Nadu, India*
[2,3,4] *Student, Dept of Computer Science and Engineering, K L N College of Engineering, Pottapalayam, Sivagangai, Tamil Nadu, India*

---

**Abstract -** *Agriculture is the backhand of the Indian economy, providing employment to an excellent extent and generating an outstanding amount in the country's income. It plays a vital role in providing food and it supports rural livelihoods. Using machine learning for improving agricultural productivity through the Crop Recommendation System based on High Yield crop offers exact advice to farmers. It analyzes crucial factors like soil type, climate, location, and real-time weather data - temperature, rainfall, humidity, and altitude to determine the best crops that can be planted for specific regions and seasons. Machine learning algorithms give it an ability to recognize some specific patterns within the data to predict which crops will most likely bring better results in a given condition. It ranks the top five crops according to predicted yields so that farmers can quickly identify the best crop for their farm. Such a system is scalable and adaptable, allowing it to be applied across an vast spectrum of agricultural practices and regional diversities. Such a system, aligning crop recommendations with local environmental factors, thus promotes sustainable farming, boosts food production, and optimizes resource use, thereby aiding the farmers to decide whether to use whatever means to improve both crop selection and the general farm productivity.*

***Key Words***: **Crop recommendation, Crop Yield Prediction, Machine Learning, Random Forest, Support Vector Machine(SVM), XGBoost.**

## 1.INTRODUCTION

Many countries have become highly advanced in agricultural practices: they have used scientific discoveries to improve their efficiency and yield in order to benefit the agribusiness in their country. India, on the other hand, has remained the same in farming techniques used and remains the biggest sector contributing to its GDP in agriculture. Obsolete methods are not helping in increasing productivity in farming, but there's a deep need for a statistic-based technique that could guide a farmer into appropriate crops selection for his farmland. It generates estimates on crop choice based on environmental indicators such as rainfall, weather, and composition of soils, which can more directly lead to a decrease in losses due to nutrient deficiencies. We use models like Random Forest, XGBoost, and SVM to analyze environmental traits and improve crop selection through yield forecasting. Such algorithms provide suitable insights for higher yields and productivity. They detect nutrient deficiencies and recommend soil amendments.

Crop recommendation and environmental analysis is meant to link historical systems with modern scientific techniques in Indian agriculture by equipping the farmers with the ability to decide for higher output and sustainability.

## 2.LITERATURE SURVEY

[1]Training sample selection for robust multi-year within-season crop classification using machine learning - Zitian Gao*, Danlu Guo*, Dongryeol Ryu*, Andrew W. Western. The system decides on optimizing training sample selection by adjusting parameters like size and class balance. The use of Random Forest and SVM algorithms makes accurate predictions. The model addresses the problem of using past season data that can decrease accuracy with unseen years due to weather variations. These techniques enhance crop recommendations that are climate and environment responsive while performing well in various seasons.

[2]Crop recommendation and forecasting system for Maharashtra using machine learning with LSTM: a novel expectation-maximization technique - Yashashree Mahale1, Nida Khan1, Kunal Kulkarni, Shivali Amit Wagle, Preksha Pareek, Ketan Kotecha, Tanupriya Choudhury, Ashutosh Sharma. This system is set based on the optimized dataset through the EM technique. The accuracy level is better. The Random Forest algorithm with a 92% accuracy level states that this one is well effective for prediction regarding suitable crops. The focus is also limited to Maharashtra. A highly region-specific model will be arrived at, but the applicability of the results will be limited to other regions due to varied climatic conditions and agricultural practices.

[3]Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables - Biplob Dey , Jannatul Ferdous , Romel Ahmed. The system will try to give practical choices of crops where to actually cultivate or other nutrients needed for optimal yields. However, the very tendency of the system to combine datasets in this case from both agricultural and horticultural crops has been seen to be associated with reduced accuracy in its predictions.

[4]Crop yield prediction using multi-attribute weighted tree-based support vector machine - M. Rajakumaran, G. Arulselvan, S. Subashree , R. Sindhuja. This model optimizes the choice of training samples based on several parameters, including sample size, balance of crop classes, and selective inclusion of non-crop classes. Such techniques will improve the performance of the model but may, unbeknownst to the operators, delete some useful information or distort the data, which can affect the suboptimality of real-world predictions.

## 3.METHODOLOGY

### 3.1 DATA AGGREGATION

The process of data aggregation as done here will require accrual and integration of varied types of data in making an overall dataset that further enhances and adds credence to crop recommendations. Here, it begins with the entry of user-provided geographic data for more specific specification of the location for crop recommendations. As such, in this regard, Loamy, Sandy, and Clay soil type information will be collated so that the local soil condition is included in this process. Aggregating current real-time weather information on temperature, rainfall, humidity, wind speed, and altitude refines the recommendations to present as close a variation of prevailing environmental conditions as possible. Historical yield information is also sourced for a variety of crops such that it is determined whether the generated crops would be appropriate for those specific local environmental conditions. Seasonal data is also taken into consideration so that the crops suggested would fall within the appropriate growing season. This integrated approach allows for very precise crop recommendations based on exact conditions.

### 3.2 MACHINE LEARNING MODELS

#### i)SUPPORT VECTOR MACHINE

Support vector machine algorithm used for model training and evaluation. In its training process, the parameters get optimized based on historical crop data learned over temperature, rainfall, humidity, and type of soil. SVM generates a hyperplane that maximally separates varied classes of crops: it iteratively improves the ability to recognize high-yielding crops. Its performance is calculated by comparing the actual yields from a validation set and the predictions it made, through metrics such as accuracy, precision, recall, and F1-score. These metrics would measure how well it correctly classifies crops of high yields. Confusion matrices further develop the visualization of performance over crop classes by ensuring the SVM model accurately predicts suitable crops over its data input.

#### ii)RANDOM FOREST

The Random Forest algorithm has been used for the training and assessment of the model in the crop recommendation system. An ensemble of decision trees is used to improve the predictions. It takes the historical crop data as input and focuses on characteristics like temperature, rainfall, humidity, or type of soil. The decision trees will all be trained on the random subset of the data, which makes the model learn diverse patterns and less prone to over fitting. This makes predictions on the output of all decision trees in aggregate. It also identifies the top high-yielding crops with majority voting. The performance of the model is assessed by evaluating the predictions against actual yields based on metrics such as accuracy, precision, recall, and F1-score. Therefore, it ensures that the Random Forest model is credible in giving advice on crops from input data.

#### iii)XGBOOST REGRESSOR

The XGBoost algorithm is the popular model that uses efficiency and accuracy in data sets with large data, used for crop recommendation systems to train and evaluate a model. Historical crop data trained on the model based on key features like temperature, rainfall, humidity, and soil type. XGBoost uses the gradient boosting framework to train an ensemble of decision trees by iteratively optimizing the model by minimizing the loss function with new trees added on to correct mistakes made by previously added trees. Model accuracy is evaluated against actual yields in validation data using some metrics, which include accuracy, precision, recall, and F1-score. This analysis can be done too to extract feature importance to determine their influence on the predictions. The evaluation process ensures that the XGBoost model based on the given input data provides crop recommendations that are reliable and of quality.

### 3.3 METRICS EVALUATION

Common metrics used to evaluate a model in machine learning are precision, recall, F1 score, MAE, MSE and R-squared ($R^2$). Precision is the percentage of correctly identified positive instances, whereas recall is the percentage of actually relevant instances that fall within a relevant class or are selected, in all cases. The F1 score is a balancing of precision and recall, making the score useful especially if there is a problem of class imbalance. The MAE calculates the average absolute difference between the prediction and actual values, while MSE penalizes large differences in errors, reflecting large deviations. $R^2$ is the measure of how much the model is able to predict the variability in the data; the higher the value of $R^2$ gets close to 1, the better the performance of the model would be. Thus, all these metrics give a balanced view of error as well as accuracy of the model.

```
print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1 Score: {f1:.2f}")

Regression Metrics:

Random Forest:
Mean Absolute Error (MAE): 58.53
Mean Squared Error (MSE): 5651.86
R-squared (R²): 0.70

XGBoost:
Mean Absolute Error (MAE): 61.26
Mean Squared Error (MSE): 6168.98
R-squared (R²): 0.67

Classification Metrics:

Random Forest:
Precision: 0.72
Recall: 0.81
F1 Score: 0.76

XGBoost:
Precision: 0.78
Recall: 0.88
F1 Score: 0.82

SVM:
Precision: 0.72
Recall: 0.81
F1 Score: 0.76
```

**Fig -1**: Metrics Evaluation

## 3.4 TOP 5 HIGH YIELD PREDICTION

The system predicts the top five high-yielding crops given the input data by using Random Forest, SVM, and XGBoost machine learning models. It evaluates the features such as the soil type, temperature, rainfall, and humidity and ranks the crops according to the potential yields predicted. An aggregation of the ranking is established by model voting to select the five crops with the highest yields. The results dashboard displays the crops that have been ranked on top, with necessary information, like predicted yields and confidence scores, to make efficient decisions. All available crops for any given scenario can be predicted, as the models are trained with historical crop data and environmental factors. Crop rankings and details evolve with changing input data and iteratively update recommendations over time. What the final dashboard does is it visually presents the top crops and their suitability in such a way that a person could use it to point out ideal choices for the region.

## 3.5 CROP RECOMMENDATION

During training, the system optimizes its parameters on the basis of historical crop data, and on the basis of various key features like temperature, rainfall, and soil type, improves the accuracy in predicting those crops. The model goes through these data iterations to enhance its ability to recommend high-yield crops. Using machine learning algorithms like Random Forest, SVM, or XGBoost, the system predicts the top 5 crops based on new input data. The performance of the model is then compared to actual yields by using the metrics such as accuracy, precision, and recall. The system is adaptable to environmental factors because it adjusts the predictions according to the regional climate and soil conditions. Users can input data in real-time that generates personalized recommendations for crops.

```
Enter Temperature (°C): 22
Enter Humidity (%): 80
Enter Rainfall (mm): 250
Enter Soil Type: Silt
Enter Location: Jaipur
Enter Altitude (m): 600
Enter Wind Speed (m/s): 2

Top 5 crops recommended by Random Forest:
['Onion' 'Radish' 'Peanuts' 'Pineapple' 'Banana']

Top 5 crops recommended by SVM:
['Banana' 'Peanuts' 'Starfruit' 'Finger Millet' 'Coconut']
```

**Fig -2**: Top 5 High Yield Prediction

## 4.SYSTEM ARCHITECTURE

A software architecture diagram is another term for the high-level structure of the system and its interactions. They show key components, their relationships, and how they collaborate to achieve functionality. The major parts of such diagrams include modules or services that represent its components; methods or protocols of their interaction, which are interfaces; and the relationships themselves, indicating where dependencies lie. Data flows are used to illustrate how information travels and is processed between the components. At times, the diagram also depicts how components are dispersed within hardware or software environments. As a blueprint, these diagrams allow developers, architects, and other stakeholders to understand and communicate system design, thereby aiding in collaboration and decision-making.

The crop recommendation system architecture diagram is as shown in figure 5.3.

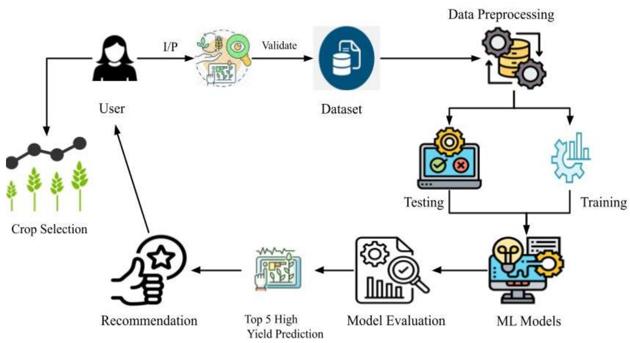


**Fig-3** Architecture of the system

The Crop Recommendation System inputs location, type of soil, and weather data, validates the information for accuracy. In the module, the pre-processed dataset cleanses the data by dealing with missing values, adjusts formats to standard ones, and makes sure that data quality is ensured. This preprocessed data is split into training and testing datasets, on which models are built and evaluated.

The system uses a few models like Random Forest, SVM, and XGBoost, which get trained and have evaluated performance on bases such as accuracy, precision, recall, and F1-score for predicting the Top 5 High-Yielding Crops. Then the recommendation of the crops is done by using the Recommendation Engine, which provides proposals based upon crop suitability. Models get retrained with new data along with feedback from the users so that the recommendations are refined in the system.

## 5.USE CASE DIAGRAM

A use case diagram in software engineering is a graphical presentation of how users or actors interact with a system to achieve specific goals or use cases. It describes the functional requirements of the system in addition to its relationship with actors and use cases. Actors could be users, other systems, or devices whereas, use cases would mean tasks or activities that the system performs. The use cases diagram provides how actors initiate or engage with different types of use cases in most cases with the interaction of several actors.

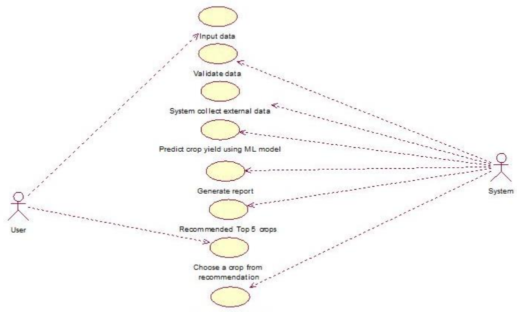


**Fig -4**: Use Case Diagram

The two main users of the system represented by figure 5.4 are the user and the system.

Here, the user feeds into the system data relating to historical crop yields, soil quality, as well as weather conditions while making recommendations to the system for further analysis. The system checks the input fed by the user in terms of accuracy and completeness before use. In addition, the system autonomously collects external data, including weather forecasts and market trends. This brings about an increased accuracy in predicting crop yields. After machine learning models based on inputting and gathering data predict crop yields, it produces a general report detailing the probable yield and other insights. Using these predictions and factors such as suitability to the soil and market prices, the system recommends the top five crops for planting. The suggestions are then reviewed by the user, who then selects a crop based on their choice. The system then comes up with elaborate data about the expected yield of the chosen crop under conditions of climate and soil. This process clearly defines user interactions and functionalities of the system.

## 6.EXISTING SYSTEM

This is challenging for generation in a timely manner, due to the fact that there are no satellite images and ground truth data available, especially on management during the season for both water and crops. The accuracy of the model increases with more imagery, which by December has reached 86%, with user accuracies over 80%. Having the requirement to include non-crop classes such as fallow land reduces accuracy, mainly in the type of crops, while their being minimal in number helps. Accurate classification requires a large number of training samples, which may not be feasible in areas whose history is not documented or well-documented. Model accuracy depends intensely on ground truth data. When the information is sparse or outdated, classification quality will drop appreciably.

## 7.PROPOSED SYSTEM

This system applies advanced algorithms from the field of machine learning, including Random Forest, XGBoost, and SVM, to evaluate critical factors affecting the crop environment, such as soil type, climate, and location. With these features, it offers precise crop recommendations tailored to the conditions the user has. In addition, the model integrates weather data in real-time to update crop recommendations dynamically, as they will be relevant during the season at hand.

The model ranks the top 5 most suitable crops and gives users a clear, data-driven process of optimizing their yield. This ranking helps farmers and agriculture professionals at large make well-informed decisions by selecting crops that are suitable for their environment. It is more scalable and can be used for different regions and climates. It does away with the complexity of satellite images and works with soil conditions and, above all, the weather data, which makes it very user-friendly. Designed for ease of use, the system simplifies the decision-making process for improved crop yield across varying agricultural settings.

## 8. RESULT

After refining and analyzing the dataset, we implemented several machine learning algorithms to enhance our crop recommendation system. The algorithms used include

Random Forest, Support Vector Machine (SVM), and XGBoost. The selected features of the dataset include crucial environmental factors such as temperature, humidity, rainfall, altitude, wind speed and soil type.

We have tested three algorithms; we found that for Random Forest accuracy, it was at 92.09% accuracy, with highest prediction. XG Boost follows at 89.67% accuracy level while the SVM model recorded approximately 87.72%. These have shown how each model accurately portrays a top 5 high yielding crop suitable for a specified environment.

## 9. FUTURE ENHANCEMENT

The Crop Recommendation System is thereby being upgraded with mechanisms for continuous improvement of the model such that it should learn from new data and feedback by users; thus, crop recommendations will be improved and updated over time. Transition to the system in a web-based platform and accessible from any device by users. The system personalizes recommendations according to the individual inputs of users by coming up with specific crop suggestions via user accounts and thereby ensures relevance and optimization of crop yields for specific user conditions. Other characteristics would also include interactiveness, such as visual representations of recommended crops, implemented suggestion tracking progress, and feedback mechanisms aimed at further improvement in future predictions and recommendations.

## 10. CONCLUSIONS

The Crop Recommendation System will provide that crop which seems the fittest according to its threshold of real-time data like temperature, humidity, and type of soil. This system is adaptive to all changes in the climate conditions by using this real-time data and user feedback. The leverage of lessons learned from machine learning and analytics allows the system to let farmers make better decisions regarding crop yield for more precise and efficient crop management. This approach improves efficiency and promotes agricultural sustainability in the long run by only suggesting factors that are conducive to particular environmental factors.

## REFERENCES

[1] Liyakathunisa Syed (2024) ‘Smart Agriculture using Ensemble Machine Learning Techniques in IoT Environment’, International Conference on Machine Learning and Data Engineering, ScienceDirect, Volume 235, Pages 2269-2278

[2] Zitian Gao, Danlu Guo, Dongryeol Ryu, Andrew W. Western [2023] ‘Training sample selection for robust multi-year within-season crop classification using machine learning’, Computers and Electronics in Agriculture, ScienceDirect, Volume 210

[3] Senthil Kumar, Swami Durai, Mary, Divya Shamili (2022) ‘Smart farming using Machine Learning and Deep Learning techniques’, ScienceDirect, Decision Analytics Journal, Vol no.2

[4] Reddy, D. J., & Kumar, M. R. (2021, May) ’Crop yield prediction using machine learning algorithm’, 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1466-1470). IEEE

[5] P. S. Nishant, P. Sai Venkat, B. L. Avinash and B. Jabber, ‘Crop Yield Prediction based on Indian Agriculture using Machine Learning,’ 2020 International Conference for Emerging Technology (INCET), 2020

[6] J. N. Kumar, V. Spandana, and V. S. Vaishnavi, ‘Supervised Machine Learning Approach for Crop Yield Prediction in Agricultural Sector,’ 2020 5th International Conference on Communication and Electronics Systems (ICCES), June 2020

[7] Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, Vinu Williams, 2021, ‘Crop Yield Prediction using Machine Learning Algorithms,’ International Journal of Engineering Research & Technology (IJERT) NCREIS – 2021 (Volume 09 – Issue 13)

[8] D. A. Bondre, and S. Mahagaonkar,’Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms’, International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 5, pp. 371-376, 2019

[9] Akash Manish Lad, K. Mani Bharathi, B. Akash Saravanan, R. Karthik,’Factors affecting agriculture and estimation of crop yield using supervised learning algorithms’,on ScienceDirect, Volume 62, Part 7, 2022, Pages 4629-4634

[10] Kavita Jhajharia, Pratistha Mathur,’Prediction of crop yield using satellite vegetation indices combined with machine learning approaches’,on ScienceDirect, Volume 72, Issue 9, 1 November 2023, Pages 3998-4007