© APR 2026 | IRE Journals | Volume 9 Issue 10 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I10-1716864

ML-Cropadvisor: An Enhanced Data-Driven Approach for Crop Selection in Sustainable Agriculture

AMIT KUMAR[1], DR. MOHD DANISH[2], PROF. (DR.) SANJAY PACHAURI[3]
[1,2,3] *Department of Data Science (DDCS), GNIOT College, Greater Noida, India*

Abstract- ***This paper presents an advanced machine learning–based crop recommendation system aimed at enhancing decision-making in sustainable agriculture. The proposed system utilizes critical environmental and soil parameters, including Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, rainfall, and pH, to predict the most suitable crop for cultivation. Unlike traditional approaches, this study incorporates multiple machine learning models, including Random Forest, Support Vector Machine (SVM), and Light Gradient Boosting Machine (LGBM), to improve prediction performance. Experimental results demonstrate that the LGBM classifier outperforms other models, achieving an accuracy of 97.58%, indicating superior predictive capability and robustness. Feature importance analysis reveals that rainfall, soil nutrients, and pH play a crucial role in crop selection. The system provides a scalable and data-driven solution for precision agriculture, contributing to improved yield, efficient resource utilization, and sustainable farming practices.***

Index Terms- ***Machine Learning, Crop Recommendation, Smart Farming, Random Forest, Agriculture Technology***

I. INTRODUCTION

Agriculture is a primary sector in many developing countries, and farmers often rely on traditional or experience-based methods to select crops. These decisions are highly affected by variations in soil health, weather conditions, and environmental changes. With the rise of data science and machine learning, it is now possible to make accurate predictions and provide personalized recommendations to farmers.

This study aims to build a crop recommendation system using machine learning models trained on soil and climatic datasets. The system suggests the ideal crop based on input parameters, enabling smart agricultural practices and improving crop yield.

II. LITERATURE REVIEW

Recent advancements in machine learning have significantly influenced precision agriculture. Various studies have explored crop prediction using supervised learning techniques.

Patel and Prajapati (2020) analyzed crop prediction systems using traditional machine learning models such as Decision Trees and Naïve Bayes, highlighting the importance of soil nutrients and climatic conditions. Breiman (2001) introduced Random Forest, which has been widely adopted due to its robustness and ability to handle high-dimensional data.

However, most existing approaches suffer from limitations such as:

- Limited dataset diversity
- Lack of model comparison
- Absence of advanced boosting algorithms
- Lower prediction accuracy

To address these gaps, this study integrates ensemble learning techniques and boosting algorithms (LGBM) to enhance prediction accuracy and model efficiency.

III. METHODOLOGY

The proposed crop recommendation system uses a supervised machine learning approach. The following steps were employed:

A. Dataset

The dataset includes essential agricultural parameters:

Nitrogen (N)
Phosphorus (P)
Potassium (K)
Temperature
Humidity

© APR 2026 | IRE Journals | Volume 9 Issue 10 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I10-1716864

pH value
Rainfall
The dataset contains labeled entries where each row corresponds to an environment condition and its suitable crop.

B. Data Preprocessing
Handling missing values
Normalization of continuous features
Encoding categorical labels
Train-test split (80:20)

C. Model Selection
In addition to traditional classifiers, this study incorporates the Light Gradient Boosting Machine (LGBM), a gradient boosting framework that uses tree-based learning algorithms. LGBM is optimized for speed and performance by utilizing histogram-based learning and leaf-wise tree growth, which significantly improves accuracy compared to level-wise algorithms.
Hyperparameter tuning was performed using grid search to optimize model performance. Key parameters such as number of estimators, learning rate, and maximum depth were adjusted to achieve optimal results.

D. Model Training
The Random Forest model was trained using 100 decision trees. Cross-validation was performed to prevent overfitting.
E. Deployment Format
The final trained model was saved using the pickle format (model.pkl) and integrated with a simple user interface for easy input and prediction.

IV. RESULTS AND DISCUSSION

The performance of multiple machine learning models was evaluated using accuracy, precision, recall, and F1-score.

Model	Accuracy
Logistic Regression	89.2%
Decision Tree	92.4%
SVM	94.1%
Random Forest	95.0%
LGBM	97.58%

The results clearly indicate that the LGBM classifier outperforms all other models, achieving the highest accuracy of 97.58%.
Key observations:

- LGBM provides better generalization due to boosting
- Random Forest performs well but slightly less efficient
- SVM struggles with scalability on larger datasets

Feature importance analysis shows:

- Rainfall → highest impact
- Nitrogen → second most important
- pH → critical for crop suitability

The confusion matrix confirms minimal misclassification, demonstrating the robustness of the proposed model.

V. CONCLUSION

This study presents an enhanced crop recommendation system using advanced machine learning techniques. The integration of LGBM significantly improves prediction accuracy compared to traditional models.

The proposed system demonstrates strong potential for real-world agricultural applications by enabling data-driven crop selection, improving productivity, and supporting sustainable farming practices.
Future work will focus on:

- Integration with real-time weather APIs
- IoT-based soil data collection
- Mobile/web deployment for farmer accessibility

- Incorporation of deep learning models for further improvement

REFERENCES

[1] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, 2009.

[2] J. Patel and P. Prajapati, "Crop Prediction Using Machine Learning: A Review," International Journal of Computer Applications, 2020.

[3] Breiman, L., "Random Forests," Machine Learning Journal, 2001.

[4] Ke, G. et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," NeurIPS, 2017.

[5] FAO, "Climate-Smart Agriculture Sourcebook," 2019.

[6] K. Gopal et al., "Machine Learning in Agriculture: A Review," IEEE Access, 2021.