© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1712190

ML-CropAdvisor: A Data-Driven Approach to Crop Selection for Sustainable Farming

AMIT KUMAR[1], ANISH KUMAR[2], DR. ANUJ CHANDILA[3], PROF. (DR.) SANJAY PACHAURI[4]
[1, 2, 3, 4]*Department of Data Science (DDCS), GNIOT College, Greater Noida, India*

Abstract- ***This paper presents a machine learning–based crop recommendation system designed to assist farmers in making data-driven agricultural decisions. The proposed system analyzes multiple environmental factors including soil nutrients (NPK), temperature, humidity, rainfall, and pH to recommend the most suitable crop for a given region. By applying machine learning algorithms such as Random Forest, Decision Trees, and Support Vector Machines (SVM), the system achieves high accuracy and reliability. The study demonstrates how data science and ML can contribute to smart farming, resource optimization, and improved agricultural productivity.***

Keywords — 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. 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

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

Multiple machine learning algorithms were tested including:

Decision Tree Classifier

Support Vector Machine

Logistic Regression

Random Forest Classifier

The Random Forest model performed best with high accuracy and stability.

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.

III. RESULTS AND DISCUSSION

The Random Forest model showed strong performance with high prediction accuracy across all crop categories.

Accuracy: 95% on test data

Precision & Recall: Balanced across all labels

Feature Importance: Rainfall, Nitrogen, and pH were identified as the most influential factors in predicting suitable crops.

The confusion matrix analysis confirmed low misclassification rates. This highlights the model's reliability for real-world applications.

IV. CONCLUSION

In this study, a machine learning–based crop recommendation system was developed and evaluated. The model effectively analyzes soil and weather parameters to suggest the most suitable crop with high accuracy. This system can assist farmers in making informed decisions, optimizing resource usage, and improving agricultural productivity.

Future work may include integrating real-time weather APIs, IoT sensor data, and deploying the model using mobile applications for wider accessibility.

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.