A

Project Report On

Crop Recommendation System using Machine Learning

By

Mr. Ankit Jeevankumar Jadhav

ABSTRACT

The Crop Recommendation System is a machine learning-based solution designed to assist farmers in selecting the most suitable crop based on environmental and soil conditions. The system analyzes essential agricultural parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, soil pH, and rainfall. Using a trained predictive model—supported by data preprocessing, feature scaling, and label encoding—the system predicts the optimal crop that can yield the highest productivity under the given conditions. By leveraging algorithms such as Random Forest, the model learns patterns from historical datasets and provides accurate and data-driven crop recommendations. This project aims to support modern precision farming, reduce crop failure risks, and promote efficient agricultural decision-making. The tool can be integrated into web applications to provide real-time crop suggestions, ultimately empowering farmers with intelligent and user-friendly technology.

- ## Introduction

Agriculture plays a vital role in the global economy, and selecting the right crop for cultivation is a crucial decision for farmers. Traditional crop selection relies heavily on experience and intuition, which may lead to inaccurate decisions, especially under rapidly changing climatic conditions. With advancements in data science and machine learning, it is now possible to make more informed, data-driven,decisions.

This project, Crop Recommendation System using Machine Learning, aims to recommend the most suitable crop based on real-time environmental and soil parameters. The system analyzes key features such as soil nutrients (N, P, K), temperature, humidity, pH value, and rainfall to predict the best crop that can thrive in those conditions. By integrating predictive analytics into agriculture, the project supports precision farming and assists farmers in maximizing productivity.

- ## Problem Statement

Farmers often face challenges in selecting the optimal crop for cultivation due to limited knowledge about soil nutrient levels, climatic conditions, and the compatibility of different crops with environmental factors. Incorrect crop selection can result in poor yield, financial loss and wastage of resources.

Therefore, there is a need for a data-driven recommendation system that can accurately suggest the best crop based on measurable soil and weather parameters. This system should be simple to use, reliable, and accessible to both farmers and agricultural decision-makers.

- ## Objective

The primary objectives of this project are:

1. To analyse agricultural datasets containing soil nutrient values and climate features.
2. To build a machine learning model capable of predicting the best crop based on soil and environmental inputs.
3. To implement data preprocessing techniques such as scaling, encoding, and cleaning for improved model accuracy.
4. To compare different machine learning algorithms and select the best-performing model.
5. To develop a user-friendly interface (Flask web app) that allows users to input environmental parameters and receive real-time crop recommendations.

6. To support precision agriculture by promoting resource-efficient and high-yield farming practices.

- ## Dataset Description

Dataset Name: Crop_recommendation.csv

Total Records: 2200

Total Features (Columns): 8

Column Name	Description	Type
N	Nitrogen content in soil	Numeric
P	Phosphorus content in soil	Numeric
K	Potassium content in soil	Numeric
temperature	Temperature in °C	Numeric
humidity	Humidity in %	Numeric
ph	Soil pH value	Numeric
rainfall	Rainfall in mm	Numeric
label	Crop name (Target Variable)	Categorical

The dataset covers 22 different crops including rice, maize, chickpea, kidney beans, pigeon peas, moth beans, mung bean, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, and more.

- EDA

The dataset contains soil nutrients (N, P, K), climate features (temperature, humidity, rainfall), soil pH, and the corresponding crop label. The dataset is clean with **no missing values** and has a **balanced distribution of crops**, ensuring fair model training.

- Key Insights:

1. Crops like rice and maize prefer higher rainfall and humidity.
2. Lentil and chickpea thrive in dry conditions with moderate nitrogen content.
3. Soil pH is a major influencing factor for citrus crops.
4. The dataset is balanced, meaning all crops have nearly equal representation.

- Key Visual Insights :
    1. Histogram: Nutrient & climate distribution (helps detect skewness)
    2. Boxplot: Outlier detection for rainfall and temperature

3. Correlation Heatmap: Identify related features
4. Crop Frequency Bar Chart: Show class balance
5. Pairplot (Seaborn): Visualize multi-dimensional separation between crops

- Statistical Findings :

1. pH range between 5.5–7.5 supports majority of crops — ideal soil.
2. Rainfall and Humidity are the strongest environmental determinants.
3. Nutrient combinations vary significantly by crop; e.g., Rice requires high N & rainfall, Lentil prefers low P and dry weather.
4. Dataset covers 22 crops with balanced representation → suitable for classification modeling.

- ## Feature Engineering

1. **Handling Missing Values**
- Checked for null values using df.isnull().sum()
- Dataset contained *no missing values*, so no imputation was required.

2. **Label Encoding**
- The crop names in the label column were converted into numeric values using Label Encoder to make them suitable for machine learning algorithms.

3. **Feature Scaling**
- Numerical features (N, P, K, temperature, humidity, pH, rainfall) were standardized using Standard Scaler to ensure all features are on similar scales and to improve model performance.

4. **Outlier Detection (Optional)**
- Boxplots were used to inspect extreme values, but the dataset showed minimal outliers, so no removal was needed.

5. **Train–Test Split**
- Dataset was divided into training (80%) and testing (20%) sets to evaluate model performance on unseen data.

- Model Training

  Multiple classification algorithms were trained and compared:

Model	Accuracy (%)
Decision Tree	95.1
Random Forest	98.3
SVM	96.7
KNN	93.4

Best Model: Random Forest Classifier

- High accuracy
- Low variance
- Interpretable via feature importance

Trained model saved as: crop_model.pkl

- Model Evaluation

Evaluation metrics included:

- Accuracy Score
- Confusion Matrix
- Classification Report (Precision, Recall, F1-score)

Random Forest Results:

- Accuracy: 98.3%
- Precision: 98%
- Recall: 98%
- F1-Score: 98%

Feature Importance:

Rainfall, humidity, and nitrogen had the highest impact on predictions.

- **Hyperparameter Tuning**

Used GridSearchCV for parameter optimization.

Best parameters for Random Forest:

{'n_estimators': 200, 'max_depth': 15, 'min_samples_split': 5}

Post-tuning accuracy improved slightly to 98.7%.

- ## Deployment (Short Summary)

The Crop Recommendation model was deployed using **Streamlit** and **Flask**.

Streamlit Deployment

- Built an interactive UI for entering soil and climate values.
- Loaded the trained model and displayed predictions instantly.
- Simple, fast, and ideal for demos and local testing.

Flask Deployment

- Created a backend server with HTML forms for user input.
- Loaded the model and preprocessor to generate predictions.
- Suitable for web integration and scalable production use.

- ## Conclusion

The Crop Recommendation System successfully uses machine learning to suggest the most suitable crop based on soil nutrients and environmental conditions. By analyzing inputs such as NPK values, temperature, humidity, pH, and rainfall, the model delivers accurate and data-driven crop predictions that help farmers make informed decisions. The system's deployment through Streamlit and Flask makes it easily accessible, user-friendly, and practical for real-world use. Overall, the project demonstrates how technology can significantly support agricultural productivity and sustainable farming.

- ## Key Outcomes:

• Achieved 98%+ accuracy with Random Forest

• Created an interactive web interface for real-world usability

• Demonstrated potential integration with IoT sensors

- ## Future Scope

• Real-time deployment with Raspberry Pi & soil sensors

• Integration with satellite weather data APIs

• Mobile application for field-level accessibility

• Use of Deep Learning (e.g., LSTM for seasonal prediction)

- ## References

1. Kaggle – Crop Recommendation Dataset
2. Scikit-learn Documentation (https://scikit-learn.org/)
3. Streamlit Documentation (https://docs.streamlit.io/)
4. OpenWeatherMap API (https://openweathermap.org/api)