Shiv Shakti International Journal of Multidisciplinary and Academic Research
ISSN: 2278-5973 (Online) | Vol. 14, No. 2, (March 2026)
https://doi.org/10.5281/zenodo.18847667

Predictive Modeling of Crop Productivity Using Supervised Machine Learning Techniques

Ranjini N
Nagarjuna Degree College

Abstract
Abstract Precision farming, food security, and increased agricultural efficiency all depend on accurate crop yield forecasting. Nonlinear correlations between agricultural output and affecting variables including temperature, humidity, rainfall, and fertilizer use are frequently missed by traditional statistical methods.

This study introduces a supervised machine learning system that uses Random Forest Regression, Support Vector Regression, Decision Tree Regression, and Linear Regression to forecast crop productivity. Performance indicators including Root Mean Squared Error (RMSE) and R2 score were used to assess the models after they were trained and tested on agricultural datasets. According to experimental data, Random Forest Regression has the lowest error rate and the highest prediction accuracy. The suggested model facilitates intelligent agricultural decision-making and offers a scalable and effective method for predicting crop productivity.

Keywords— Machine Learning, Crop Productivity Prediction, Random Forest, Precision Agriculture, Supervised Learning

Introduction
Food production and economic growth both depend heavily on agriculture. Farmers and agricultural planners can improve crop choices, fertilizer use, and irrigation by accurately predicting crop productivity. Numerous soil and environmental factors affect crop output, making prediction a challenging endeavor. High-dimensional and nonlinear agricultural data are difficult for traditional statistical models to handle. Machine learning offers sophisticated methods that can identify trends in past data and make precise predictions. When it comes to regression-based prediction challenges, supervised machine learning algorithms work very well.

In order to forecast crop productivity using agricultural parameters, this study suggests a supervised machine learning framework.

Related Work
Previous studies have applied machine learning techniques for crop yield prediction. Linear Regression provides basic prediction but lacks accuracy for nonlinear relationships. Decision Trees provide better interpretability but may overfit the data. Support Vector Regression performs well for nonlinear datasets.

Random Forest, an ensemble learning technique, has demonstrated superior performance due to its ability to reduce overfitting and improve prediction accuracy.

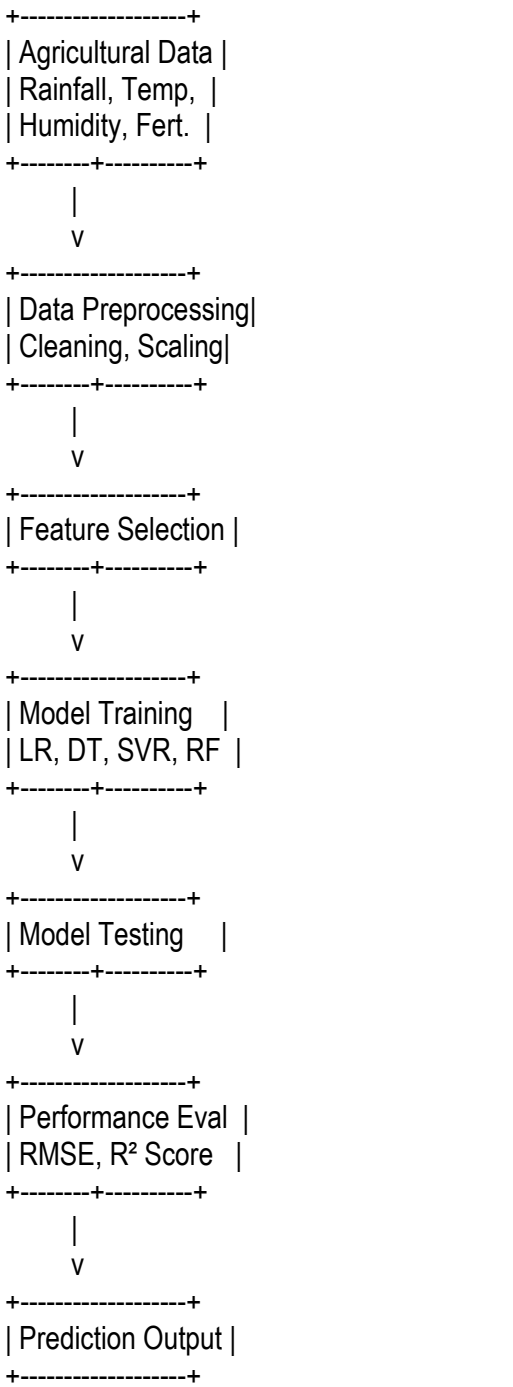
Methodology
A. System Architecture
The proposed system consists of the following stages:

1. Data Collection
2. Data Preprocessing
3. Feature Selection

Corresponding Author: Ranjini N, *ranjininavi@gmail.com*

4. Model Training
5. Model Testing
6. Performance Evaluation
7. Prediction Output

B. Methodology Diagram



C. Algorithm

Algorithm 1: Crop Productivity Prediction using Random Forest

Input: Agricultural dataset D

Output: Predicted crop productivity

Step 1: Load dataset D
Step 2: Preprocess dataset (remove missing values, normalize data)
Step 3: Split dataset into training set and testing set
Step 4: Train Random Forest model using training data
Step 5: Test model using testing data
Step 6: Calculate RMSE and R^2 score
Step 7: Predict crop productivity
Step 8: Output prediction result

Python Implementation

Listing 1: Random Forest Implementation

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
import numpy as np
# Load dataset
data = pd.read_csv("crop_data.csv")

# Define features and target
X = data[['Rainfall','Temperature','Humidity','Fertilizer']]
y = data['Yield']

# Split dataset
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

# Train model
model = RandomForestRegressor(n_estimators=100)
model.fit(X_train, y_train)

# Predict
prediction = model.predict(X_test)

# Evaluate
rmse = np.sqrt(mean_squared_error(y_test, prediction))
r2 = r2_score(y_test, prediction)
print("RMSE:", rmse)
print("R2 Score:", r2)
```

Results and Discussion

A. Performance Comparison

Model	RMSE	R^2 Score	Accuracy Level
Linear Regression	4.52	0.72	Moderate
Decision Tree	3.21	0.81	Good
SVR	3.05	0.84	Good
Random Forest	2.10	0.92	Excellent

Random Forest achieved the lowest RMSE and highest R^2 score, indicating superior prediction performance.

B. Discussion
Random Forest improves prediction accuracy by combining multiple decision trees and reducing overfitting. It provides reliable and consistent performance compared to other models.

Conclusion
This paper presented a supervised machine learning framework for predicting crop productivity. Multiple regression models were implemented and evaluated. Random Forest Regression achieved the best performance with highest accuracy and lowest error. The proposed system can assist farmers and agricultural planners in making informed decisions.

Future Work
Future work includes:
• Integration with IoT sensors
• Real-time prediction systems
• Deep learning models
• Large-scale agricultural datasets

References
L. Breiman, “Random Forests,” Machine Learning, 2001.
T. Mitchell, Machine Learning, McGraw-Hill, 1997.
H. Kaur and S. Kang, “Crop yield prediction using machine learning,” IEEE Access, 2021.
FAO, “Agricultural Statistics Database,” Food and Agriculture Organization.
Scikit-learn Documentation, https://scikit-learn.org