Crop Yield Prediction Using Machine Learning Based on Historical Crop Data

Dr.Dharavath Bhadru D,Sravan Kumar M.Venu
Professor Assistant Professor Assistant Professor
Department of CSE(AI&ML),Narasimha ReddyEngineeringCollege,Hyderabad,India.drdbhadru@gmail.com
Department of CSE(AI&ML),Narasimha ReddyEngineeringCollege,Hyderabad,India.shravankumar1248@gmail.com
Department of CSE(AI&ML),Narasimha ReddyEngineeringCollege,Hyderabad,India.venu.perftest11@gmail.com

Abstract

Agriculture plays a major role in the economy and food security of many countries. Accurate crop yield prediction helps farmers and governments make better decisions regarding crop planning, storage, and distribution. Traditional prediction methods often fail to handle large and complex agricultural datasets. This research proposes a Machine Learning-based crop yield prediction system using historical agricultural data such as rainfall, temperature, soil properties, fertilizer usage, and previous crop production records. Various Machine Learning algorithms including Linear Regression, Random Forest, Decision Tree, and XGBoost are used to analyze the data and predict crop yield accurately. The proposed system aims to improve prediction accuracy, optimize agricultural planning, and support smart farming practices.

Introduction

Crop yield prediction is an important task in precision agriculture. Farmers depend on environmental and historical factors to estimate production. Due to climate change and varying soil conditions, predicting crop yield manually becomes difficult.

Machine Learning techniques can analyze large agricultural datasets and discover hidden patterns that improve prediction accuracy. By using historical crop data, weather conditions, and soil information, ML models can provide reliable crop yield predictions.

This research focuses on developing an intelligent crop yield prediction model using supervised learning algorithms.

Problem Statement

Traditional crop yield estimation methods are:

- Time-consuming
- Less accurate

- Unable to process large agricultural datasets
- Highly dependent on human expertise

There is a need for an automated and accurate prediction system using Machine Learning techniques.

Objectives

1. To collect and preprocess historical agricultural data.
2. To analyze the impact of environmental and soil parameters on crop yield.
3. To develop Machine Learning models for crop yield prediction.
4. To compare the performance of different ML algorithms.
5. To improve prediction accuracy for smart farming applications.

Dataset Features

Typical input parameters:

Feature	Description
Rainfall	Annual rainfall
Temperature	Average temperature
Humidity	Environmental humidity
Soil Type	Soil category
Nitrogen	Soil nitrogen level
Phosphorus	Soil phosphorus level
Potassium	Soil potassium level
Crop Type	Type of crop
Area	Cultivation area
Historical Yield	Previous production data

Output

Predicted crop yield.

Machine Learning Algorithms

1. Linear Regression

Simple prediction based on linear relationships.

2. Decision Tree

Tree-based prediction model.

3. Random Forest

Ensemble learning method with high accuracy.

4. XGBoost

Advanced boosting algorithm for better performance.

5. Artificial Neural Network (ANN)

Deep learning-based prediction approach.

Methodology

Step 1: Data Collection

Collect agricultural datasets from:

- Government agriculture portals
- Kaggle datasets
- Historical farming records

Step 2: Data Preprocessing

- Remove missing values
- Normalize data
- Encode categorical values

Step 3: Feature Selection

Select important parameters affecting crop yield.

Step 4: Model Training

Train multiple ML models using training data.

Step 5: Prediction

Predict crop yield using test data.

Step 6: Evaluation

Evaluate models using:

- Accuracy
- MAE
- RMSE
- R^2 Score

Mathematical Model

The basic regression equation:

$$Y=\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn+\epsilon$$

Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \cdots + \beta_nX_n + \epsilon

$$Y=\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn+\epsilon$$

Where:

- YYY = Crop Yield
- X1,X2,...XnX_1, X_2, ... X_nX1,X2,...Xn = Input features
- β\betaβ = Coefficients
- ϵ\epsilonϵ = Error term

Performance Metrics

Mean Absolute Error (MAE)

MAE=1n∑i=1n∣yi−yi^∣MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y_i}|MAE=n1∑i=1n∣yi−yi^∣

Root Mean Square Error (RMSE)

RMSE=1n∑i=1n(yi−yi^)2RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y_i})^2}RMSE=n1∑i=1n(yi−yi^)2

Expected Results

- Accurate crop yield prediction
- Improved agricultural decision-making
- Better resource management
- Increased productivity
- Reduced economic loss

Advantages

- Fast prediction
- Handles large datasets
- Better accuracy
- Supports smart agriculture
- Reduces manual effort

Applications

- Smart farming
- Government agriculture planning
- Food supply management
- Fertilizer recommendation systems
- Precision agriculture

Future Enhancements

- Integration with IoT sensors
- Real-time weather analysis
- Deep learning optimization
- Mobile application development
- Federated Learning for secure agriculture data sharing

References

1. Crop yield prediction using machine learning: A systematic literature review
   Thomas Van Klompenburg, Ayalew Kassahun, Daniel Rodriguez.
   Computers and Electronics in Agriculture, 2020.
   DOI: 10.1016/j.compag.2020.105709
   Available at: ScienceDirect Paper
2. Crop yield prediction using machine learning: An extensive and systematic literature review
   Sarowar Morshed Shawon et al.
   Smart Agricultural Technology, 2024.
   DOI: 10.1016/j.atech.2024.100718
   Available at: Research Article
3. Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches
   Md Abu Jabed, Masrah Azrifah Azmi Murad.
   Heliyon, 2024.
   DOI: 10.1016/j.heliyon.2024.e40836
   Available at: PubMed Article
4. Crop yield prediction using machine learning techniques
   S. Iniyan, V. Akhil Varma, Ch. Teja Naidu.
   Advances in Engineering Software, 2023.
   DOI: 10.1016/j.advengsoft.2022.103326
   Available at: Research Paper
5. A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction
   Joshua Fan et al.
   arXiv, 2021.
   Available at: arXiv Paper
6. Predicting Crop Yield With Machine Learning: An Extensive Analysis Of Input Modalities And Models On a Field and sub-field Level
   Deepak Pathak et al.
   arXiv, 2023.
   Available at: arXiv Research
7. Recent applications of machine learning, remote sensing, and IoT approaches in yield prediction: a critical review
   Fatima Zahra Bassine et al.
   arXiv, 2023.
   Available at: Critical Review Paper

8. Winter Wheat Crop Yield Prediction on Multiple Heterogeneous Datasets using Machine Learning
Yogesh Bansal, David Lillis, Mohand Tahar Kechadi.
arXiv, 2023.
Available at: Winter Wheat Prediction Paper
9. Comparative Analysis on Crop Yield Forecasting using Machine Learning Techniques
Shubham Sharma et al.
Rural Sustainability Research, 2024.
DOI: 10.2478/plua-2024-0015
Available at: Comparative Analysis Paper
10. Food and Agriculture Organization
FAOSTAT Agricultural Database.
Available at: FAOSTAT Official Website
11. Kaggle
Crop Yield Prediction Datasets.
Available at: Kaggle Datasets
12. UCI Machine Learning Repository
Agricultural Machine Learning Datasets.
Available at: UCI Repository