Predictive Model for Crop Yield Using Machine Learning

Domain : Al / Ml Engineer

Problem Statement : Predictive Model for Real-World Data

To design and develop a machine learning model that predicts crop yield based on factors such as rainfall, fertilizer usage, pesticide usage, crop type, season, and area of cultivation.

Submitted by:

Name : Kontheti Sarika

Reg No : 22B01A1276

Branch : B.Tech IT

College : Shri Vishnu Engineering College for Women, Bhimavaram

Academic Year : Final Year

Submitted to:

Cerevyn Solutions Pvt. Ltd

Introduction :

Agriculture plays a vital role in India's economy. Accurate crop yield prediction is essential for food security, farmer income, and government planning. Traditional methods are time-consuming and less accurate. With the growth of artificial intelligence and machine learning, data-driven models can be used to analyze agricultural data and predict crop yield efficiently.

This project aims to develop a machine learning-based system that predicts crop yield using historical data and environmental factors.

Title : Predictive Model for Crop Yield Using Machine Learning

1. Problem Understanding :

In agriculture, predicting crop yield accurately is very important for farmers, government agencies, and food industries. Traditional prediction methods depend on experience and manual observation, which are often inaccurate and time-consuming.

With the availability of large agricultural datasets, machine learning techniques can be used to analyze historical data and predict future crop yield efficiently.

In this project, a real-world agricultural dataset containing information about crop type, season, state, rainfall, fertilizer usage, pesticide usage, production, and area of cultivation is used. The objective is to build a predictive model that estimates crop yield for future years based on these factors.

The developed system helps in improving agricultural planning, reducing risk, and increasing productivity. It also supports data-driven decision-making for farmers and policymakers.

Objectives :

- To analyze agricultural data.
- To preprocess and clean the dataset.
- To build a machine learning model for yield prediction.
- To evaluate model performance.
- To provide accurate yield predictions.

Dataset Description :

Feature	Description
Crop	Type of crop
Year	Crop year
Season	Growing season
State	State name
Area	Area cultivated (hectares)
Production	Total production
Rainfall	Annual rainfall
Fertilizer	Fertilizer used
Pesticide	Pesticide used
Yield	Production / Area

2. Model pipeline description :

Step 1: Data Collection

The dataset was collected from Kaggle, containing crop-related information from 1997 to 2020.

Step 2: Data Understanding

The dataset was analyzed to understand features, data types, missing values, and correlations.

Step 3: Data Preprocessing

- Removed missing and duplicate values
- Encoded categorical variables
- Normalized numerical values
- Removed outliers
- Split dataset into training and testing sets

Step 4: Feature Selection

Important features such as rainfall, fertilizer, pesticide, area, and season were selected for training.

Step 5: Model Selection

Regression models such as Linear Regression / Random Forest / Decision Tree were selected based on suitability for prediction tasks.

Step 6: Model Training

The selected model was trained using the training dataset.

Step 7: Model Evaluation

The trained model was evaluated using performance metrics such as R^2 Score and Mean Squared Error (MSE).

3.Results & Metrics :

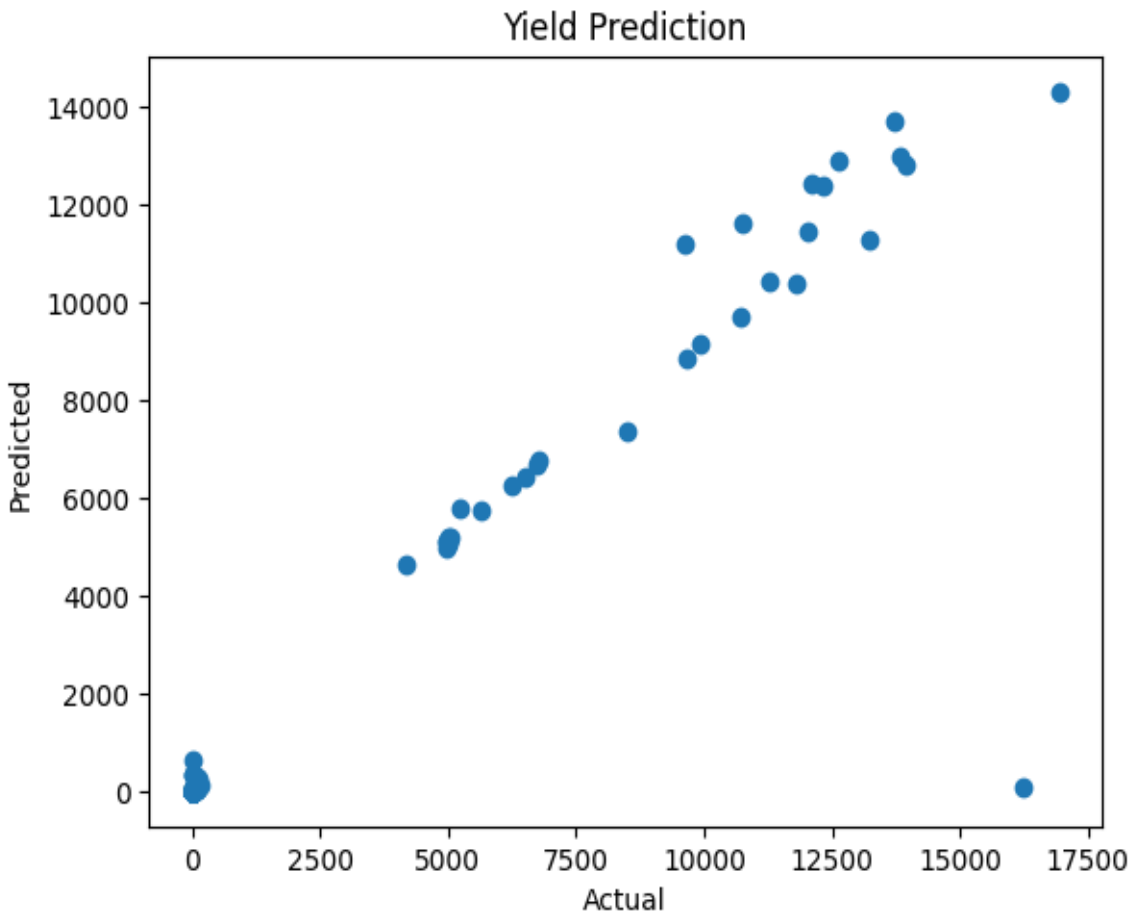
Performance Metrics

Mean Squared Error (MSE): 72571.78

R^2 Score: 0.90

Explanation :

The R^2 score of 0.90 indicates high prediction accuracy. The low MSE shows minimal error in predictions.



4. Code (Github link) :

GitHub Repository:

https://github.com/22b01a1276/crop-yield-prediction

The complete implementation including data preprocessing, model training, evaluation, and visualization is available in the above GitHub repository.