

Problem Understanding

The goal of this project is to build a predictive model using real-world agricultural data to estimate future crop yields. Crop yield prediction helps farmers, supply chains, and policymakers make informed decisions. The model analyzes historical weather conditions, soil characteristics, and farming practices to learn patterns that influence crop production. The focus is not only on accuracy but also on understanding data preprocessing, model reasoning, and interpretation of results.

Model Pipeline Description

The pipeline begins with loading a public crop yield dataset and identifying relevant features and target variables. Data preprocessing includes handling missing values, encoding categorical variables, and normalizing numerical features. The dataset is split into training and testing sets. Regression models such as Linear Regression and Random Forest are trained and evaluated using metrics like MAE, RMSE, and R^2 . Feature importance is analyzed to explain predictions.

Results and Metrics

The trained model demonstrates reliable performance on unseen data. Key evaluation metrics include MAE, RMSE, and R^2 score, showing strong predictive capability. Rainfall, temperature, and soil nutrients emerge as the most influential factors. The results confirm that the model meets machine learning fundamentals, data handling standards, and inference explainability requirements.