Crop Yield Classification Report

Project Title

Crop Yield Classification

Objective: To classify agricultural crop yield into categories - low, medium, or high based on input features: soil type, rainfall, and seed type.

Dataset Description

The dataset used in this project includes the following columns:

- soil_type: Numeric value representing the type of soil (e.g., 0, 1, 2)
- rainfall: Numeric value indicating rainfall in mm
- seed_type: Categorical variable (e.g., A, B, C) indicating seed variety
- yield: Target variable representing crop yield category (low, medium, high)

The dataset was uploaded in CSV format and automatically preprocessed for model training.

Data Preprocessing

1. Loaded the CSV dataset using pandas.
2. Encoded categorical features using Label Encoding:
   - seed_type: Converted into numeric labels
   - yield: Converted into numeric target labels
3. Split the dataset into training and testing sets using an 80/20 split.
4. Visualized the class distribution of 'yield' using a bar chart.

Crop Yield Classification Report

Model Training

A Random Forest Classifier was used to train the model. This model is effective for classification tasks and handles both numeric and categorical data well.

Key parameters:

- n_estimators = 100 (number of trees)
- test_size = 0.2 (20% data used for testing)
- random_state = 42 (to ensure reproducibility)

Model Evaluation

The model was evaluated using Accuracy Score and a Classification Report (Precision, Recall, F1-score).

Results showed that the model is effective at correctly classifying yield categories based on the given features.

The trained model was saved using joblib for future use without retraining.

Prediction Process

After training, the model is used to predict the crop yield based on user input:

- User provides values for soil_type, rainfall, and seed_type.
- seed_type is encoded using the same LabelEncoder.
- The model predicts one of the classes: low, medium, or high.

Crop Yield Classification Report

Conclusion

This project demonstrates the application of machine learning for predicting crop yields.

It showcases how data preprocessing, model training, evaluation, and prediction can be integrated into a robust pipeline.

This can help in agricultural planning and decision-making to improve crop production.

Report Generated On

2025-04-22 09:44:39