Sowing Success: How Machine Learning Helps Farmers Select the Best Crops

Objective: This project aims to help farmers select the most suitable crops based on soil conditions using machine learning. The dataset provided contains essential soil metrics such as nitrogen, phosphorus, potassium, and pH values, with the goal of predicting the best crop for optimal yield.

Dataset Overview: The dataset `soil_measures.csv` contains the following columns:

- N: Nitrogen content ratio
- P: Phosphorous content ratio
- K: Potassium content ratio
- pH: pH value of the soil
- `crop`: Target variable (crop type)

Each row represents soil conditions in a field, with the target variable indicating the ideal crop.

Model Approach:

- Features: Nitrogen (N), Phosphorous (P), Potassium (K), and pH (ph) were used to predict the best crop.
- Model: Logistic Regression was applied for multi-class classification, trained for each feature individually.
- Evaluation Metric: F1 score was used to evaluate performance, balancing precision and recall.

Training & Results:

- Data Split: 80% of the data was used for training and 20% for testing.
- Best Feature: Potassium (K) produced the highest F1 score, indicating it was the most predictive feature for selecting the crop.
- F1 Score: The overall average F1 score was 0.13, showing limited predictive power.

Feature Performance Visualization: A bar chart was used to visualize F1 scores for each feature, with potassium (K) emerging as the most influential feature.

Conclusion & Recommendations:

- Best Feature: Potassium (K) proved most important for predicting the best crop.
- Model Improvement: The overall F1 score of 0.13 suggests the model needs improvement. Future work could focus on using more complex models like random forests or neural networks and incorporating additional features such as weather data or crop history to boost accuracy.
- Cross-validation: To prevent overfitting and enhance generalization, cross-validation should be implemented.

This project demonstrates the potential of using machine learning to assist farmers in making informed crop decisions, but improvements are needed to enhance the model's predictive performance.