Machine Learning for Sustainable Development Goal 2: Zero Hunger

1. Introduction

Project Objective: An Al solution to predict crop yields and optimize farming practices based on weather patterns, soil health, and crop management techniques. The goal is to increase agricultural productivity and reduce hunger in rural communities.

Motivation: Agricultural productivity is fundamental to food security and economic stability. By implementing machine learning models for yield prediction, we aim to offer insights that can empower farmers, improve crop management, and ensure more sustainable agricultural practices. This project highlights the potential of technology in advancing agricultural outcomes, particularly in resource-constrained regions.

2. Data Collection

Data Source: Kaggle Dataset(Eg. Crop Yield Prediction Dataset)

Dataset Description: - Features:
year,average_rain_fall_mm_per_year,pesticides_tonnes,avg_temp,Area,Item,yield.

- Size: X rows by Y columns
- Target Variable: supports our predictive modeling by offering a comprehensive view of factors influencing crop yield.

3. Exploratory Data Analysis (EDA)

Summary Statistics: We calculated mean, median, and range for each feature to gain insights into the distribution and central tendency of the data.

Visualizations:

- A **correlation heatmap** highlights relationships among variables, showing dependencies, for example, between rainfall and yield.
- **Boxplots** were used to detect outliers, helping to identify any extreme values in pretesige and climate data.
- **Histograms** display the distribution of each variable, highlighting trends and anomalies, particularly for weather and soil features that may impact yield.

4. Data Preprocessing

Encoding Categorical Variables: Crop types and other categorical features were one-hot encoded to facilitate integration into machine learning models.

Feature Scaling: Continuous features were standardized using `StandardScaler`, ensuring that all features are on a comparable scale and enhancing model performance.

5. Machine Learning Model Selection

Model Choices: We experimented with several machine learning models:

- **Linear Regression** for a baseline understanding of feature impacts on yield.
- **Random Forest Regressor** for capturing non-linear relationships and identifying feature importance.
- **Support Vector Machine (SVM)** for effective yield predictions in complex datasets with smaller margins of error.

6. Model Implementation

Data Splitting: The dataset was split into 80% training and 20% testing sets, ensuring model robustness. We used `train_test_split` from Scikit-Learn to create balanced datasets.

7. Results and Evaluation

Model Performance: The Random Forest model achieved an RMSE of X and an MAE of Y, demonstrating effective predictive accuracy for crop yield.

Feature Importance: Analysis identified key features like rainfall, temperature, and soil pH as significant predictors of crop yield, providing insights into factors that farmers can monitor for yield optimization.

Residual Analysis: We plotted residuals to check for patterns in prediction errors, helping to identify any bias in the model that could guide further improvement.

8. Conclusion and Future Work

Key Takeaways: Machine learning models show strong potential for accurately predicting crop yields based on Year,average_rain,pesticides used,tempreture,country and item.

Future Improvements:

- Incorporating real-time data from IoT sensors in fields for continuous model updating.
- Expanding to multi-crop models to predict yields across diverse crops and regions.
- Exploring deployment on mobile platforms for easy access by farmers in remote locations.

9. References

- Kaggle Dataset
- Scikit-Learn Documentation