Machine Learning for Sustainable Development Goal 2: Zero Hunger

1. Introduction

Project Objective: To use machine learning to tackle challenges in food security and hunger, supporting Sustainable Development Goal 2 (SDG 2: Zero Hunger). This project aims to predict food insecurity trends, identify regions at high risk of undernourishment, and analyze crop yield patterns, with the goal of aiding in resource allocation and policy-making for food distribution and agricultural development.

Motivation: Access to adequate food and nutrition is essential for health, growth, and socio-economic development. Hunger and food insecurity continue to affect millions globally, leading to malnutrition and stunted economic progress. By leveraging machine learning, this project seeks to develop predictive tools that can support efforts to reduce hunger, optimize food production, and improve food distribution, thereby contributing to global efforts toward Zero Hunger.

2. Data Collection

Data Source: Kaggle Dataset (e.g., "Zero_Hunger")

Dataset Description:
- Features: Entity, Code, Year, Prevalence of undernourishment (% of population)
- Size: X rows by Y columns
- Target Variable: Prevalence of undernourishment (% of population)

3. Exploratory Data Analysis (EDA)

Summary Statistics: Mean, median, and distribution of Year , Undernourishment.
Visualizations:
- linear regression for Training set
- linear regression for Testing set.

4. Data Preprocessing

Handling Missing Values: No Nan values present in this dataset

5. Machine Learning Model Selection

Model Choices:
- Simple Linear Regression

Why Scikit-Learn: Easy implementation, variety of algorithms, and effective performance metrics.
Evaluation Metric: MSE,MAE,R-square

6. Model Implementation

Data Splitting: Split dataset into 80% training and 20% testing sets using `train_test_split` from Scikit-Learn.
Splitting the dataset into the Training set and Test set

Training the Simple Linear Regression model on the Training set

Predicting the Test set results

Code Example:

```
#Splitting the dataset into the Training set and Test set

from sklearn.model_selection import train_test_split

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=42)


# Training the Simple Linear Regression model on the Training set

from sklearn.linear_model import LinearRegression

regressor=LinearRegression()

regressor.fit(X_train,y_train)


# Predicting the Test set results

y_pred=regressor.predict(X_test)
```

7. Results and Evaluation

Model Performance:
- Linear Regression achieved Mean Squared Error (MSE) ,Mean Absolute Error (MAE) ,R-squared (R^2) ,values indicating the model's strength in predicting Undernourishment.

<u>CODE:</u>

```
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
```

```
# Assuming y_test contains actual values and y_pred contains predictions

mse = mean_squared_error(y_test, y_pred)

mae = mean_absolute_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)


print("Mean Squared Error (MSE):", mse)

print("Mean Absolute Error (MAE):", mae)

print("R-squared (R²):", r2)
```

Target Importance:
- Here we estimated the Undernourishment based on the given Year

8. Conclusion and Future Work

Key Takeaways: Machine learning models effectively predict Undernourishment based on Year(of past dataset values).The project demonstrates potential for real-time monitoring and resource allocation.
Future Improvements:
- Incorporating real-time data for continuous learning.
- Expanding to a broader dataset covering multiple regions.
- Implementing models on edge devices for on-site analysis in remote areas.

9. References

- Kaggle Dataset
- Scikit-Learn Documentation