Machine Learning for Sustainable Development Goal 13: Climate Action

1. Introduction

Project Objective: The primary objective of this project is to develop a machine learning model that analyzes energy consumption patterns to recommend optimal energy usage that minimizes carbon emissions. By identifying trends in energy consumption and emissions, this project aims to contribute towards actionable strategies that support carbon footprint reduction, aligning with global sustainability goals.

Motivation: This project is inspired by the United Nations Sustainable Development Goals (SDGs), especially the goal centered on climate action. As global energy demands rise, there is an urgent need to adopt efficient energy practices that reduce carbon footprints and foster sustainable resource management.

2. Data Collection

Data Source: Kaggle Dataset

Dataset Description:
The dataset combines monthly data on energy consumption, renewable energy production, and carbon emissions, helping to identify patterns that reduce emissions:

1. Date: Month and year of each record.
2. Energy Consumption :Residential, Commercial, Industrial: Energy use in different sectors.
3. Renewable Production (MWh): Solar, Wind: Renewable energy generation metrics.
4. Carbon Emissions (metric tons):Coal, Natural Gas, Oil: Emissions from each energy source.

3. Exploratory Data Analysis (EDA)

- Distribution and Outliers: Histograms revealed general distribution trends, while box plots identified outliers, especially in industrial energy consumption and emissions.

- Correlation Analysis: Scatter plots showed strong correlations between commercial/industrial energy use and coal emissions, indicating these sectors as major contributors.

4. Data Preprocessing

- Splitting Data: The data was divided into training and testing sets to ensure that the model could generalize to new, unseen data. Typically, 80% of the data was used for training, and 20% for testing.

- Features and Target Variable: Key energy consumption features (e.g., Residential, Commercial, and Industrial energy usage) and renewable energy production data (e.g., Solar and Wind production) were used as input features (X), while Coal Emissions was the target variable (y).

5. Machine Learning Model Selection

-Objective: Choose a simple model to predict coal emissions based on energy consumption data for baseline insights.

- Model: Linear Regression: Selected for its simplicity, interpretability, and ability to identify linear relationships between energy consumption and emissions.

- Evaluation Metrics: Mean Squared Error (MSE) and R^2 Score were used to assess model accuracy.

- Scikit-Learn: Easy implementation, variety of algorithms, and effective performance metrics.

6. Model Implementation

A simple Linear Regression model was used to predict coal emissions. The model was trained and tested on energy consumption and renewable production data, aiming for transparency and interpretability.

Code Example:

```
# Prepare data for modeling

X = data[['Residential_Consumption', 'Commercial_Consumption', 'Industrial_Consumption', 'Solar_Production', 'Wind_Production']]

y = data['Coal_Emissions']

# Split the dataset into training and testing sets

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)

from sklearn.linear_model import LinearRegression# Create a linear regression model

model = LinearRegression()

model.fit(X_train, y_train)
```

```
# Make predictions on the test set

y_pred = model.predict(X_test)

# Evaluate the model

from sklearn.metrics import mean_squared_error, r2_score

mse = mean_squared_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)

# Display the evaluation results

print(f'Mean Squared Error: {mse:.2f}')

print(f'R² Score: {r2:.2f}')

# Plotting the predicted vs actual emissions

plt.figure(figsize=(8, 6))

plt.scatter(y_test, y_pred, color='blue')

plt.plot(y_test, y_test, color='red', linestyle='--')  # Line for perfect predictions

plt.xlabel('Actual Coal Emissions')

plt.ylabel('Predicted Coal Emissions')

plt.title('Actual vs Predicted Coal Emissions')

plt.show()
```

7. Results and Evaluation

Performance Metrics: The model achieved an R^2 score of 0.78, explaining 78% of the variance in emissions.

Prediction Accuracy: A scatter plot showed that predictions closely matched actual emissions, proving the model's baseline effectiveness.

8. Conclusion and Future Work

The project demonstrates a practical approach to predicting coal emissions based on energy consumption patterns. The model provides insights into how residential, commercial, and industrial consumption impacts emissions, with a focus on identifying high-emission sectors that could benefit from targeted optimization strategies.

Future Work: Experimenting with time series or more complex machine learning models like Random Forest or Gradient Boosting and Adding regional, seasonal, and policy-related data to capture broader factors affecting emissions.

9. References

- Kaggle Dataset
- Scikit-Learn Documentation and Google Colaboratory