Title: Predictive model for Energy Consumption and Electricity Generation

1. Executive Summary:

This code analyzes a dataset on world energy consumption and electricity generation trends. It explores various aspects such as energy sources, greenhouse gas emissions, and energy self-sufficiency across countries. The analysis includes data cleaning, exploratory data analysis (EDA), data visualization, and classification models to predict a country's energy self-sufficiency level.

2. Problem Statement:

The primary goal is to gain insights into global energy consumption patterns, identify trends in renewable and non-renewable energy sources, and classify countries based on their energy self-sufficiency levels. This analysis aims to provide a comprehensive understanding of the energy sector and its impact on greenhouse gas emissions, as well as identify countries that are heavily reliant on energy imports and potential areas for increasing domestic energy production or diversification. Develop a predictive model to forecast global energy consumption based on historical data.

3. Data Sources:

Using this World Energy Data collected and collated by https://ourworldindata.org/energy

4. Methodology:

1. **Data Collection:**

    - Use historical global energy consumption data from a publicly available dataset.

2. **Data Preparation:**

    - Load and inspect the dataset for missing values and anomalies.
    - Perform data cleaning, such as handling missing values and removing duplicates.
    - Standardize and normalize the data to prepare it for modeling.

3. **Exploratory Data Analysis (EDA):**

    - Visualize the data using libraries like Matplotlib, Seaborn, and Plotly.
    - Identify trends, patterns, and correlations in the data.

4. **Modeling:**

    - Apply various machine learning algorithms such as Linear Regression, Decision Trees, and Random Forest.
    - Train and test the models using appropriate metrics to evaluate their performance.
    - Select the best-performing model based on accuracy and other relevant metrics

5. Expected Outcomes:

- A predictive model capable of accurately forecasting global energy consumption.
- Identification of key factors influencing energy consumption.
- Insights to support energy management and policy decisions.
- Recommendations for further improvements and research in energy forecasting.

6. Tools and Technologies:

- Python programming language.
- Libraries including Pandas, NumPy, Matplotlib, Seaborn, Plotly, Scikit-learn.

7. Risks and Challenges:

- Data Quality: The presence of missing or inconsistent data can impact model accuracy, necessitating thorough data cleaning.
- Model Accuracy: Ensuring the model's predictions are reliable for real-world applications requires rigorous testing and validation.
- External Factors: Unpredictable events or changes in global energy policies may affect consumption patterns, posing challenges to the model's predictive capabilities.

8. Conclusion:

This project aims to develop a robust model for predicting global energy consumption, which can significantly aid in energy planning and management. By leveraging machine learning techniques, the project seeks to provide accurate forecasts and actionable insights, contributing to more efficient and sustainable energy use.