**Project Report: Predicting Electricity Access for Sustainable Development (SDG 7)**

**SDG Problem Addressed:** This project addresses a critical challenge within **Sustainable Development Goal 7: Affordable and Clean Energy**, specifically focusing on **Target 7.1: By 2030, ensure universal access to affordable, reliable and modern energy services.** The core problem tackled is to **predict future electricity access rates** for various countries and regions based on historical trends. Accurate predictions are vital for policymakers and development organizations to identify areas needing intervention, allocate resources effectively, and track progress towards universal energy access.

**Machine Learning Approach Used:** To predict electricity access rates, a **Supervised Learning** approach was employed. Given that electricity access is a continuous numerical value (percentage), this was framed as a **regression problem**. The chosen algorithm for this task was the **Random Forest Regressor**. It builds multiple decision trees and merges their predictions to improve accuracy and control overfitting, making it a robust choice for complex datasets with diverse features. **Results:** The dataset, sourced from relevant development indicators (like World Bank Open Data), contained historical electricity access percentages for various countries from 1990 up to 2015, with the goal of predicting the 2016 access rates. After initial data loading, missing values were identified and imputed by filling numerical year columns with their respective means. Categorical features such as 'Country Name', 'Country Code', 'Indicator Name', and 'Indicator Code' were transformed using one-hot encoding to make them suitable for machine learning. The prepared data was then split into training (80%) and testing (20%) sets.

The Random Forest Regressor model was trained on the historical data and evaluated on the unseen test set. The results demonstrate the model's strong predictive capability:

* **Mean Absolute Error (MAE): 0.66** This indicates that, on average, the model's predictions for 2016 electricity access deviated by only 0.66 percentage points from the actual values. This is an excellent level of accuracy for this type of prediction.
* **R-squared (R2 Score): 0.95** An R-squared score of 0.95 means that 95% of the variance in 2016 electricity access can be explained by the features used in the model. This signifies that the model is highly effective in capturing the underlying patterns and relationships within the data.

**Ethical Considerations:**

* **Data Bias:** The historical data used for training might contain inherent biases related to data collection methodologies, reporting discrepancies, or underrepresentation of certain vulnerable or remote populations. If specific regions or demographics are historically under-documented, the model's predictions for these areas might be less accurate or even perpetuate existing inequalities if not carefully interpreted
* .**Fairness and Sustainability:** The solution aims to promote fairness by providing insights that can help identify disparities in electricity access, thereby guiding targeted interventions