Project REGIS: Renewable Energy Growth and Investment Strategies

Sameen Shaik, Aareb Chowdhury, Maycol Meza Padilla, Ayman Alabbasi
Department of Computer Science
University of Massachusetts Amherst
Amherst, MA, USA

***Abstract*—Project REGIS applies data-driven methods to explore how economic indicators and financial flows affect renewable energy adoption. Using Kaggle's Global Sustainable Energy Dataset (2000–2020), we implement statistical analysis, clustering (K-Means), and predictive modeling (Linear Regression and Decision Tree Regressor) to identify countries where renewable energy investment would have the greatest impact. Our results show that GDP per capita is a far stronger predictor of renewable share than financial investment alone, pointing to the need for strategies that consider broader economic development.**

***Index Terms*—Renewable energy, clustering, regression, GDP, sustainability, investment strategy, machine learning**

I. Introduction

Climate change demands a global transition to renewable energy sources. Although global investments in clean energy are increasing, outcomes vary widely across countries—some experience rapid growth in renewable adoption, while others struggle to gain traction. This disparity highlights the need for a smarter, more targeted approach to clean energy investment.

Project REGIS, which stands for Renewable Energy Growth and Investment Strategies, aims to address this challenge. Our objective is to identify where renewable investments will have the highest impact. We use Kaggle's Global Sustainable Energy Dataset (2000–2020), which contains energy and economic indicators from over 200 countries, to analyze these trends. By applying a combination of statistical testing, clustering algorithms, and predictive modeling, we provide actionable insights to optimize investment strategies for sustainable energy development.

II. Motivation

Existing funding strategies for renewable energy often follow general trends or political priorities rather than data-driven insights. Prior research from the IEA and UN suggests that investment generally improves renewable adoption. However, these studies rarely identify which countries are most in need—or most likely to benefit—from such investments.

Our project goes beyond surface-level analysis. Using t-tests, ANOVA, and ANCOVA, we examined the influence of economic indicators on renewable adoption. We also applied clustering (K-Means) to group countries with similar characteristics and trained predictive models (Linear Regression and Decision Tree Regressor) to assess key drivers. One of our most important findings is that GDP per capita is a stronger predictor of renewable adoption than external financial flows. This insight challenges the assumption that money alone drives clean energy growth and instead emphasizes the role of local economic readiness.

III. Related Work

Reports from the International Energy Agency (IEA) and the United Nations Department of Economic and Social Affairs (UNDESA) emphasize the need for financial support in transitioning to renewable energy. These sources confirm that investments play a key role but often lack specificity in identifying which countries should be prioritized. Morgan Stanley also notes an increase in private-sector investments, though their reports don't outline strategic allocation based on data. Our project fills this gap by combining rigorous analysis and machine learning to provide granular, country-level recommendations.

IV. Dataset and Methods

We used the Kaggle Global Sustainable Energy Dataset (2000–2020), which includes over 200 countries and the following features:

- GDP per capita
- Financial flows (USD)
- Renewable electricity generation (TWh)
- Population density
- Electricity access (%)
- Renewable energy share (%)

We cleaned and standardized the dataset: null rows were dropped, feature names were normalized (lowercase, underscores), and values were scaled between 0 and 1. We used:

- **Statistical Testing:** t-test, ANOVA, and ANCOVA
- **Clustering:** K-Means with 3 clusters
- **Regression:** Linear Regression and Decision Tree Regressor
- **Train-test split:** 70/30 for all models

V. Key Findings

A. Statistical Tests

T-Test: Low-income countries (GDP ¡ $10,000) had a significantly higher renewable share than high-income countries (45.63% vs. 18.88%). **ANOVA:** Financial flows varied significantly between income tiers (p = 0.00117). **ANCOVA:** GDP per capita was a stronger predictor than financial flows (p ¡ 0.001 vs p = 0.12).

B. Trends

- Electricity access steadily improved in countries like India, Brazil, and Nigeria.
- Heatmaps show strong positive correlation between GDP and electricity access.
- Economic readiness—not funding—drives renewable transition.

VI. Modeling and Analysis

A. Model 1: Multiple Linear Regression

- Inputs: financial flows, GDP per capita, renewable electricity, population density
- R^2 Score: 0.12, RMSE: 28
- MLR underestimated high renewable users and overestimated low ones.
- Baseline model showed that linearity fails to capture complexity.

B. Model 2: Decision Tree Regression

- R^2 Score: 0.78, RMSE: 14.06
- Key predictors: electricity access and renewable generation
- Financial flows had low importance in feature ranking
- Economic infrastructure proved more predictive than monetary aid

C. Final Model: K-Means + DTR Hybrid

We grouped countries into 3 clusters using K-Means on economic and energy indicators, then trained a separate DTR for each group:

- **Cluster 0 (Low-income, high renewable):** R^2 = 0.64, RMSE = 18.06
- **Cluster 1 (Mid-income, low renewable):** R^2 = 0.76, RMSE = 9.46
- **Cluster 2 (Upper/mid-income, mixed renewables):** R^2 = 0.68, RMSE = 11.60

Key insights from each cluster provided a more tailored view of what factors matter most per region.

VII. Conclusion and Future Work

Conclusion: Our results show that economic development—measured through GDP per capita and electricity access—is a stronger driver of renewable energy growth than financial flows. Tree-based models outperformed linear ones in identifying these relationships, and clustering helped highlight actionable country groupings.

Future Work:

- Incorporate policy incentives, infrastructure indices, and environmental risks
- Try ensemble methods like Random Forest or forecasting with LSTMs
- Develop an interactive investment dashboard
- Investigate causality, not just correlation, between funding and adoption

Acknowledgment

We would like to thank our course instructors for their continued guidance and support, and Kaggle for providing access to open renewable energy datasets.

References

[1] International Energy Agency. https://www.iea.org/energy-system/renewables
[2] UN Department of Economic and Social Affairs. https://sdgs.un.org/goals/goal7
[3] Kaggle Global Sustainable Energy Dataset. https://www.kaggle.com/datasets
[4] Morgan Stanley Reports on Renewable Investment. https://www.morganstanley.com/articles/decarbonization-renewable-energy-investment-ideas