Kiribati: A Random Forest Regression & Time-Series Analysis on Global Energy Patterns

Victoria DelaCruz, Chaminade University of Honolulu, Data Science, Fall 2024

Introduction

Kiribati is a small island state in Micronesia that faces various development challenges such as climate change (UN Knowledge Platform, n.d.). This project focuses on forecasting energy consumption for Kiribati using random forest regression and the ARIMA model. By analyzing historical data, these methods aim to identify key features influencing energy use and provide time-series predictions to help anticipate future energy demand. These insights can inform sustainable energy planning and contribute to climate change solutions.

Background

Energy production and consumption levels positively increase SDG indicators for economies and societies by contributing to a positive GDP, hence the need for energy prediction models (Santika et al., 2019). Furthermore, random forest regression allows for measures of feature importance, which indicate the strength of influence for individual variables (Gerstofer et al., 2023).

Hypothesis

Financially-centered features will display the strongest importance compared to the rest of the variables. Given the fluctuations in energy use overtime for Kiribati, the ARIMA model may not perform as well.

Methods and Programs

- Data was collected from Kaggle and originated from Our World in Data and the World Bank.
- Google Colab and Python were used for data cleaning and visualizations.
- Random forest regression and ARIMA modeling was performed in Google Colab.

Acknowledgements

- Ms. Mariah Yelenick
- Ms. Amber Camp
- Dr. Rylan Chong

Visualizations

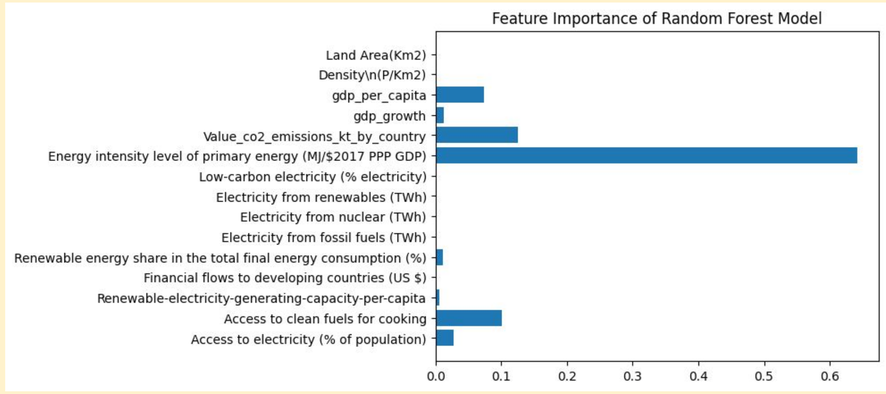

Figure 1 *Feature Importance of Random Forest Model*

p	d	q
4	1	0

Table 1 *ARIMA Model Parameters*

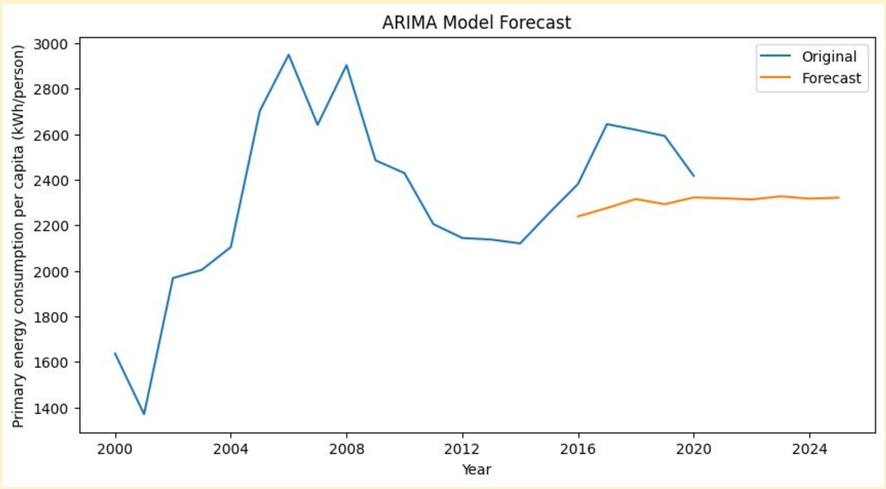

Figure 2 *ARIMA Model Forecast*

MSE	MAE	RMSE
241.79	69411.88	263.46

Table 2 ARIMA Model Evaluation Metrics

Results

- Random Forest Regression
  - Energy intensity level, access to clean fuels, and value of CO2 emissions have the highest feature importance.
  - Other important features include GDP per capita, access to electricity, and GDP growth.
- ARIMA Model
  - Comparing the testing data to forecast results (years 2016-2020), the closest points were for 2020 (training data was around 2400 and forecast around 2300).
  - Evaluation Metrics:
    - Average primary energy consumption is 2319.75. MSE and RMSE are on the lower end. Mean Absolute Error (MAE) is rather higher relative to the Mean Standard Error (MSE) and Root Mean Standard Error (RMSE).

Discussion

This analysis of Kiribati's energy consumption using Random Forest Regression and ARIMA highlights several key factors and evaluates model accuracy. These findings support the hypothesis of financially-centered variables significantly affecting energy use, but also found that access to energy is crucial. ARIMA's forecast accuracy varied with the MAE being much higher than other metrics. This suggests limitations in the model's ability to capture fluctuations in Kiribati's energy use. For future research, using SARIMA and incorporating seasonality may be a more accurate forecasting model and approach.

References

Tanwar, A. (2023). *Global data on sustainable energy (2000–2020)* [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/6327347

Gerstorfer, Y., Hahn-Klimroth, M., & Krieg, L. (2023). A notion of feature importance by decorrelation and detection of trends by random forest regression. *Data Science Journal, 22*(1), 42. https://doi.org/10.5334/dsj-2023-042

Santika, W. G., Rupf, G. V., Anisuzzaman, M., Shafiullah, G. M., Bahri, P. A., & Urmee, T. (2019). From goals to joules: A quantitative approach of interlinkages between energy and the Sustainable Development Goals. *Energy Research & Social Science, 50,* 201–214. https://doi.org/10.1016/j.erss.2018.11.016