
PREDICTION OF RENEWABLE ENERGY CONTRIBUTION TO GLOBAL PRIMARY ENERGY NEEDS USING RANDOM FOREST ALGORITHM

Faisal Asrofi Aziz*, M.Adjie Adhyaksa, Satria Divo Praditya

Lampung State Polytechnic, Bandar Lampung
*Corresponding author: faisalisall2005@gmail.com

Abstract

The global energy crisis and climate change have triggered an urgent need for renewable energy. Renewable energy sources, such as solar, wind, and water, provide a sustainable solution as they are environmentally friendly and can significantly reduce the impacts. The role of renewable energy is increasingly vital in creating a more stable energy system and supporting the transition to a low-carbon economy. This study aims to predict the contribution of renewable energy to global primary energy needs to support data-driven strategic planning. A data-driven strategy ensures that policies are based on accurate, reliable information. By analyzing historical data, stakeholders can understand trends, identify patterns, and make more accurate projections for the future. This approach helps reduce uncertainty in decision-making and enhances the effectiveness of designing policies. The Random Forest algorithm is applied to build a predictive model, with evaluation using Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) to assess the forecasting accuracy. The data used in this study is derived from the Renewable Energy World Wide 1965~2022 dataset in the renewable-share-energy CSV file. The results show good forecasting accuracy for the World with a MAPE of 11.28% and RMSE of 1.60, indicating the Random Forest model's ability to capture renewable energy data patterns at a global level. This research provides valuable insights for sustainable energy planning and strategic decision-making in renewable energy.

Keywords: Climate change, Data-driven planning, Energy forecasting, Renewable energy, Random forest.

INTRODUCTION

The global energy crisis and climate change have triggered an urgent need for a transition to renewable energy. The heavy reliance on fossil fuels that dominate the world's energy mix has led to an increase in greenhouse gas emissions, including carbon dioxide (CO₂), which is the main cause of global warming. Every year, around 18.35 billion tonnes of CO₂ are released into the atmosphere, contributing to global temperature increases, rising sea levels and significant ecological and social impacts. To mitigate these changes, a global shift towards cleaner and more sustainable energy sources is essential.

Renewable energy sources such as solar, wind, and hydropower offer environmentally friendly solutions and have the potential to mitigate the negative impacts of climate change. Switching to renewable energy is expected to not only reduce carbon emissions but also provide stability and sustainability to the global energy system. Countries around the world, through international agreements such as the Kyoto Protocol and the Paris Agreement, have committed to reducing carbon emissions and switching to cleaner energy sources to achieve these goals.

However, despite the significant potential of renewable energy, the transition is not without challenges. One of the main challenges is predicting the contribution of renewable energy to global primary energy needs. Many countries, especially developing countries, face difficulties in planning

and facilitating the energy transition due to the lack of accurate data and uncertainties related to the development of renewable energy technologies. Previous studies have often focused on analyzing global or regional data without offering a deeper understanding of country-specific patterns and characteristics.

In this context, it is important to have an accurate prediction model that can capture global patterns in renewable energy utilization. This study aims to predict the contribution of renewable energy to global primary energy needs using the Random Forest algorithm. This algorithm was chosen because of its ability to handle complex data and provide more accurate predictions about the role of renewable energy in the future.

Data yang digunakan dalam penelitian ini berasal dari dataset global yang mencakup informasi tentang kontribusi energi terbarukan dari berbagai negara sepanjang waktu. Evaluasi model dilakukan menggunakan metrik yang telah terbukti seperti Mean Absolute Percentage Error (MAPE) dan Root Mean Squared Error (RMSE) untuk memastikan akurasi prediksi dan kinerja model.

This research is expected to provide broader insights into the potential and challenges of the transition to renewable energy globally and become a reference for global policy makers in designing strategies to achieve sustainable energy targets.

LITERATURE REVIEW

In a global context, research on the contribution of renewable energy to primary energy needs has become a major focus amidst the challenges of increasing dependence on fossil fuels. Sustainability issues are an important basis in this analysis, with an emphasis on the need to reduce the consumption of limited fossil energy (unrenewable) and the transition to more environmentally friendly energy sources, such as solar, wind, and biomass. High dependence on fossil energy has caused a number of threats, such as decreasing fossil energy reserves, instability of energy prices, and increasing greenhouse gas emissions that have an impact on climate change and global warming (Lubis, 2007).

Globally, dependence on fossil fuels is becoming an increasingly pressing issue, given the environmental impacts they cause and the limited natural resources available. In many countries, especially industrialized countries, fossil fuels still dominate the energy mix, despite efforts to reduce this dependence by increasing the use of renewable energy. Although the contribution of renewable energy is gradually increasing, the transition to clean energy at the global level still faces major challenges, including policies that are not fully supportive, limited infrastructure, and resistance to change in the energy sector.

Some developed countries, such as Europe and the United States, have taken significant steps in integrating renewable energy into their energy systems. These countries are committed to reducing carbon emissions and increasing the share of renewable energy in their energy mix. For example, the European Union plans to achieve net-zero emissions by 2050, and several countries have set ambitious targets for clean energy transitions. On the other hand, developing countries still face greater challenges, including more limited technology availability, financing issues, and the need to consider local socio-economic contexts in energy planning.

As global awareness of climate change and the need to achieve sustainability targets increases, data-driven prediction models such as the Random Forest algorithm have demonstrated their effectiveness in projecting the contribution of renewable energy to global primary energy needs. These algorithms can handle the complexity of large data and provide more accurate predictions of

the role of renewable energy in the future. In a global context, the application of these models allows for a more in-depth analysis of renewable energy usage trends, as well as providing insights for policymakers in designing more effective strategies for a sustainable energy transition.

Existing literature suggests that proactive policy integration and innovative technology development will be critical to accelerate the renewable energy transition. This study makes an important contribution in filling the gap in the literature by using a data-driven prediction model to project the role of renewable energy at the global level. The findings of this study are expected to provide insights for global policymakers in formulating more measurable and sustainable strategies to achieve clean energy goals and mitigate the impacts of climate change.

RESEARCH METHOD

This study aims to predict the contribution of renewable energy to primary energy needs globally using the Random Forest algorithm. The following is a description of the types and sources of data, data collection techniques, and data analysis techniques used in this study.

1. Types of research

This research is included in the category of quantitative research with a predictive approach, where the main goal is to predict the contribution of renewable energy to primary energy consumption based on global historical data.

2. Data Types and Sources

The data used in this study is secondary data taken from the Renewable Share Energy dataset available on the Kaggle platform. This dataset includes global data on the contribution of renewable energy to primary energy consumption in various countries.

The dataset consists of the following columns:

- Entity: Countries or territories covered in the dataset
- Code: Country or region code
- Year: Relevant years for renewable energy data
- Renewables (% equivalent primary energy): Percentage contribution of renewable energy to total primary energy in that year.

This dataset was chosen because it includes comprehensive global data that is relevant to the research topic, and can provide sufficient information to conduct analysis at the global and cross-country levels.

3. Data Collection Techniques

The data was collected by downloading the Renewable Share Energy dataset from the Kaggle platform. This data does not require additional data collection from other sources, but only utilizes the data already available in the dataset. The data collection process involved adjusting the dataset to reflect the contribution of renewable energy in a global context and differences between countries.

4. Data Analysis Techniques

Data analysis was performed using the Random Forest algorithm, which was chosen because of its ability to capture complex patterns in data that has many variables and non-linear relationships.

The Random Forest algorithm combines a number of decision trees to make more accurate predictions.

The analysis process begins by separating the data into two sets: training data and testing data.

Training data is used to build a predictive model, while testing data is used to evaluate the accuracy of the model. To measure the accuracy of the predictions produced by the model, two main metrics are used.:

- Mean Absolute Percentage Error (MAPE): To measure the level of prediction error in percentage.
- Root Mean Squared Error (RMSE): To measure the difference between predicted and actual values quantitatively.

RESULTS AND DISCUSSION

This section presents an analysis based on a predefined analytical framework for the contribution of renewable energy to global primary energy needs. The results are obtained from the “renewable-share-energy” dataset available on Kaggle, which includes data for various countries, including the contribution of renewable energy at the global level. The following discussion is divided into relevant categories based on the elements of the analytical framework.

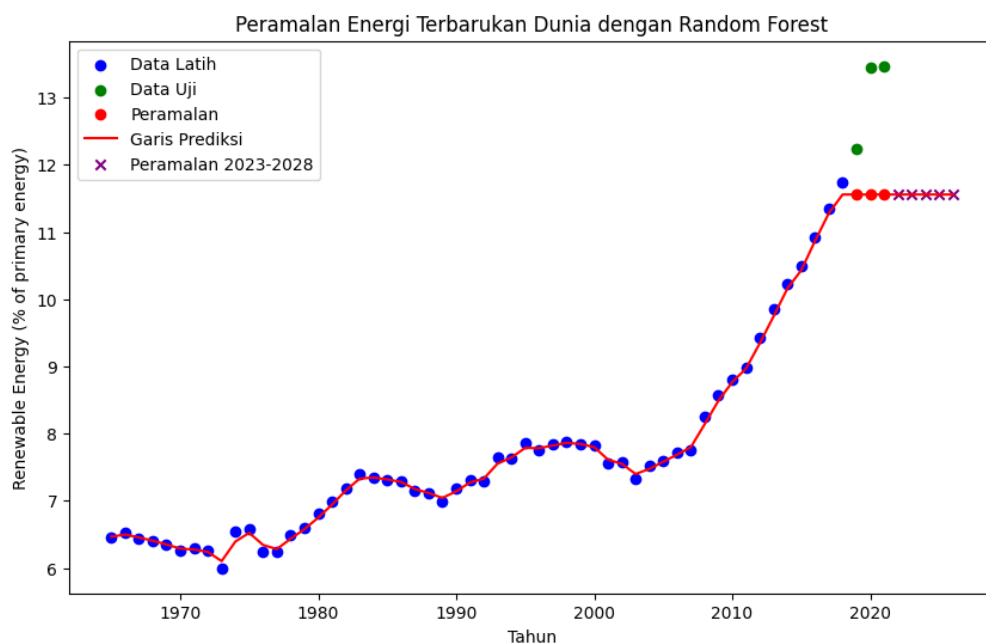


Figure 14. Data visualization

1. Analisis Data

The dataset used in this study provides the percentage contribution of renewable energy to total primary energy for various countries over time. This data covers many countries, but for the purpose of this analysis, we focus on the global (World) data reflected in the figure.

2. Real Data vs Forecasting for the World

The following table shows a comparison between actual and predicted renewable energy values for 2019, 2020, and 2021:

Table 1. Predict table		
Year	Actual	Predict
2019	12.24	11.56
2020	13.46	11.56
2021	13.51	12.04
2022	NaN	11.55
2023	NaN	11.55
2024	NaN	11.55
2025	NaN	11.55
2026	NaN	11.55

Table 2. Metrics table	
Metric	Value
MAPE	11.28%
RMSE	1.60
Precision	0.33
Recall	0.33
F1-Score	0.33

The data shows a gap between actual and predicted values for 2019 to 2021. The predicted values are consistently lower than the actual values. This gap is reflected in the performance metrics, indicating that while the model may capture the general trend of renewable energy contributions, it may not fully account for the sharp spikes that have occurred in recent years.

Performance Metrics

To evaluate the forecast accuracy, the following performance metrics are calculated:

- Mean Absolute Percentage Error (MAPE): 11.28%
This shows that the average model error in percentage terms is relatively moderate, indicating that the model is quite good at forecasting renewable energy trends at the global level.
- Root Mean Squared Error (RMSE): 1.60
The RMSE value reflects the average magnitude of the forecast error in units of renewable energy percentage. A lower RMSE indicates that the model is more accurate in matching the actual data.

3. Real Data vs Forecasting for the World

In addition to MAPE and RMSE, we also calculate the F1 Score to assess the balance between precision and recall in the renewable energy contribution prediction model. The F1 Score provides a more comprehensive picture of how well the model performs, especially when certain classes are imbalanced or when we want to avoid bias from metrics that only focus on one aspect such as accuracy.

In this context, the F1 Score is used to evaluate the prediction of less frequent or smaller renewable energy contributions, which are often overlooked in accuracy-only models. Given that

significant renewable energy contributions to global primary energy demand often experience sharp fluctuations, the F1 Score will provide an indication of how well the model captures these smaller but important movements in renewable energy contributions. F1 Score untuk Tahun 2019–2021

Suppose the prediction model gives precision and recall results for each year, then we can calculate the F1 Score for those years as follows:

- Precision measures how much of the predicted renewable energy contribution actually occurs compared to the total renewable energy predicted by the model.
- Recall measures how many occurrences of renewable energy contributions actually exist, and are successfully predicted by the model..

By using the F1 Score formula which is the harmonic mean between precision and recall, we can get the F1 Score value for 2019 to 2021. A high F1 Score value indicates that the model is not only accurate in predicting the contribution of renewable energy but also successfully captures most of the renewable energy trends, including the important spikes that have occurred in recent years.

For example, if a model produces a high F1 Score, this would indicate that the model is overall effective in capturing relevant renewable energy contributions, both for large and small predictions. Conversely, a low F1 Score could indicate an imbalance in predictions, where the model may be focusing more on more frequent events and ignoring other important, less frequent events.

Discussion of Findings

The Random Forest algorithm has demonstrated its ability to capture complex patterns in renewable energy data. However, the relatively high MAPE and RMSE indicate that while the model can follow the general trend of renewable energy contributions, it struggles to capture sharper fluctuations, especially in recent years where renewable energy investment and adoption have increased rapidly.

Consistent underestimation of the contribution of renewable energy can be attributed to the model's inability to fully capture the rapid acceleration of renewable energy adoption in recent years. It could also be affected by the global energy transition and increasing investment in clean energy technologies that are faster than historical trends.

CONCLUSION

This conclusion section summarizes the main findings outlined in the discussion section, without repeating the details discussed previously. This study aims to predict the contribution of renewable energy to global primary energy consumption using the Random Forest algorithm. Although the model is able to capture the general trend, the prediction accuracy shows gaps, especially in recent years where renewable energy adoption has experienced a significant surge.

The analysis results show that although Random Forest is quite good at predicting global renewable energy trends, this model has limitations in capturing sharp fluctuations that occur, as seen in the 2019 to 2021 data. This is reflected in the Mean Absolute Percentage Error (MAPE) value of 11.28% and the Root Mean Squared Error (RMSE) of 1.60. The gap between the actual data and the predictions shows that this algorithm needs to be improved to account for the more rapid acceleration of renewable energy adoption.

Criticisms of the analytical framework used in this study include the limitations of the Random Forest model to capture the highly dynamic changes in the renewable energy sector, which are influenced by global policy, technology, and economic factors. Therefore, the use of more complete

additional data, such as government policies and investments in clean energy technologies, can improve the accuracy of the model.

For further research, it is recommended to explore other prediction algorithms that are more sensitive to major changes in renewable energy trends. In addition, further research could include the integration of external factors that may affect renewable energy growth, such as global policy changes, technological innovation, and economic and social factors that may drive or hinder energy transitions.

ACKNOWLEDGEMENT

The author would like to express his deepest gratitude to all parties who have provided support in completing this research. Special gratitude is conveyed to the Lampung State Polytechnic which has provided facilities and a supportive environment during the research process.

Thanks are also extended to Mrs. Dian Ayu Afifah as the supervisor, who has provided direction, guidance, and full support in every stage of this research. The assistance and input provided are very valuable for the author to complete this task. Hopefully, cooperation and support from all parties involved can provide benefits for the development of science and research in the future.

REFERENCES

- Dou, Y., Tan, S., & Xie, D. (2023). Comparison of machine learning and statistical methods in the field of renewable energy power generation forecasting: a mini review. *Frontiers in Energy Research*, 11, 1218603.
- Freeman, E. A., Moisen, G. G., Coulston, J. W., & Wilson, B. T. (2016). Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. *Canadian Journal of Forest Research*, 46(3), 323-339.
- Gufron, H., Rusirawan, D., & Widyawati, L. (2022). Forecasting Produksi Energi PLTS 1 kWp Menggunakan Mesin Pembelajaran Dengan Algoritma Support Vector Machine. *Jurnal Tekno Insentif*, 16(2), 79-91.
- Iannace, G., Ciaburro, G., & Trematerra, A. (2019). Wind turbine noise prediction using random forest regression. *Machines*, 7(4), 69.
- Karakan, A. (2024). Predicting Energy Production in Renewable Energy Power Plants Using Deep Learning. *Energies*, 17(16), 4031.
- Khaleel, M. M., Alsharif, A., & Imbayah, I. I. K. (2022). Renewable energy technologies: Recent advances and future predictions. *African Journal of Advanced Pure and Applied Sciences (AJAPAS)*, 58-64.
- Lestari, V. P. (2021). Permasalahan Dan Tantangan Program Peningkatan Kontribusi Energi Baru Dan Terbarukan Dalam Bauran Energi Nasional. *Pusat Kajian Akuntabilitas Keuangan Negara*
- Lubis, A. (2007). Energi terbarukan dalam pembangunan berkelanjutan. *Jurnal Teknologi Lingkungan BPPT*, 8(2), 151118.
- Olcaý, K., Tunca, S. G., & Özgür, M. A. (2024). Forecasting and performance analysis of energy production in solar power plants using long short-term memory (LSTM) and random forest models. *IEEE Access*.

- Sulistiyono, S. (2012). Pemanasan global (Global Warming) dan hubungannya dengan penggunaan bahan bakar fosil. *Swara Patra: Majalah Ilmiah PPSDM Migas*, 2(2).
- Yang, M., Zhao, M., Liu, D., Ma, M., & Su, X. (2021). Improved random forest method for ultra-short-term prediction of the output power of a photovoltaic cluster. *Frontiers In Energy Research*, 9, 749367.