IJE | INDONESIAN JOURNAL OF ENERGY
P-ISSN: 2549-1016 | E-ISSN: 2549-760X
available online at www.ije-pyc.org

Published by: The Purnomo Yusgiantoro Center

Big Data and Satellite Imagery for Energy Efficiency Mapping in Indonesia: A Future Shaped by Advanced Analytics

Firman Afrianto [1, *], Andini P. Salsabillah [2], Annisa D. Hariyanto [2]

[1] Department of Urban and Regional Planning, Faculty of Engineering, Gadjah Mada University, Bulaksumur, Sleman, Daerah Istimewa Yogyakarta 55281, Indonesia
[2] PT Sagamartha Ultima Indonesia, Griya Shanta J 295, Kelurahan Mojolangu, Kota Malang, Jawa Timur 65141, Indonesia

*Corresponding Author: firmanafrianto@mail.ugm.ac.id

Article History

Received 25 June 2024
Accepted 17 January 2025
Available 27 February 2025

Abstract

In the sophisticated realm of big data, analyzing energy efficiency in Indonesia has become crucial for identifying savings opportunities. This study analyzes and forecasts energy efficiency across various Indonesian provinces by using advanced regression techniques in machine learning—Support Vector Regression, Artificial Neural Network, and Random Forest. We utilize large-scale raster data, including carbon dioxide (CO_2) emissions from the OCO-2 (Orbiting Carbon Observatory-2) GEOS satellite, nocturnal satellite images from the Visible Independent Imaging Radiometer Suite (VIIRS), and demographic and infrastructural data from WorldPOP and EsriWorld Cover. The analysis results highlight a notable increase in CO_2 emissions from 2019 to 2023, with a significant reduction in night-time light emissions in 2020 due to the pandemic, which temporarily decreased human activities. Despite these fluctuations, the continuous increase in population density and built-up areas underscores the persistent influence of urbanization on emissions. The Random Forest model, which provided the most accurate predictions, showed a 65% increase in total CO_2 emissions by 2030, driven by urbanization and economic growth, followed by a decline by 2045 due to targeted government policies. These insights contribute significantly to understanding the distribution of energy efficiency and support the development of sustainable energy policies in Indonesia. The study not only enriches scientific literature but also guides policy-making, offering a framework for tailored energy efficiency improvements. This research marks a pivotal advancement in utilizing big data and satellite technology to optimize energy use in a context that was previously underexplored.

Keywords:
big data, energy efficiency, energy management, machine learning, predictive analysis

1. Introduction

Energy is crucial for economic and social development and is closely linked to improving quality of life (Bologna, 2013; Ibrahim et al., 2023; Lloyd, 2017). It is also a key focus of the Sustainable Development Goals (SDGs), as it supports poverty alleviation, education, health, industrialization, and water supply (Santika et al., 2019; United Nations, 2016). However, most global energy production and consumption remains unsustainable, causing increases in energy use and carbon dioxide (CO_2) emissions (Ibrahim et al., 2023; Miškinis et al., 2014; Zakari et al., 2022).

Urbanization has transformed cities worldwide, with 55% of the global population now residing in urban areas. This shift drives social and economic changes and heavily impacts energy consumption, as urban areas account for about 80% of global energy use (Bilgili et al., 2017; Lee et al., 2013; Min et al., 2022; Nejat et al., 2015). Previous research shows a close link between urbanization and high energy consumption, influenced by factors like human activity intensity, population density, and land expansion (Elliott et al., 2017; Guan & Zhou, 2015; Jones, 1991; Papa et al., 2014; Zhao & Zhang, 2018). Factors such as the intensity of human activity (Fehrer & Krarti, 2018; Hipskind et al., 2011), population density (Mazur, 1994; Zarco-Periñán et al., 2021), and the expansion of built-up land (Hu & Fan, 2020; Polydoros & Cartalis, 2015) have been identified as primary drivers of increased energy consumption in urban areas. High energy consumption also significantly contributes to the rise in CO_2 emissions (Jian et al., 2021; Peters et al., 2007; Vieira & Ceretta, 2021). CO_2 is one of the gases most implicated in global pollution and is a key factor in causing the greenhouse effect, which contributes to global climate change (Romero-García et al., 2022). However, there is potential to reduce urban energy consumption by up to 50% by 2050 by improving energy efficiency and promoting energy-saving behaviors among the public (Min et al., 2022; Svenfelt et al., 2011; Ürge-Vorsatz et al., 2012).

The concept of energy efficiency is increasingly emphasized as a crucial aspect of achieving sustainable development (Di Foggia, 2018; Soltangazinov et al., 2020; Zakari et al., 2022). Energy efficiency is a measure of how effectively and efficiently energy is used to produce the desired output. (Lin & Zhai, 2023). The higher the energy efficiency, the less energy is required to produce the same amount of output (Milovanovic et al., 2012; Wang et al., 2023), thereby reducing energy use, costs, greenhouse gas emissions, and CO_2 air pollution. Various studies have consistently shown that higher energy efficiency generally results in lower CO_2 emissions (Longa et al., 2022; Tajudeen et al., 2018; Tu et al., 2022). Many studies have examined the link between energy efficiency and CO_2 emissions, focusing on urbanization's role. Ma et al. (2014) and Sun & Huang (2020) noted the need for more detailed data to estimate CO_2 emissions from transport accurately and understand urbanization’s impact on carbon efficiency. Nuţă et al. (2021) and Zhang & Lin (2012) further emphasized the importance of considering the impacts of urbanization and related energy factors on CO_2 emissions. However, using CO_2 emission estimates to assess energy efficiency, as suggested by Guo et al. (2022), requires further research. Accurately mapping CO_2 patterns in space and time is essential for informed carbon reduction policies, but data limitations often restrict studies to large-scale models (Guo et al., 2022; Wang & Liu, 2017). Most research relies on administrative data, which lacks internal spatial patterns (Cao et al., 2014).

Several studies have attempted to investigate energy consumption and efficiency in Indonesia. For instance, Jafari et al. (2012) explored the relationship between economic growth, CO_2 emissions, and energy consumption from 1971 to 2007 using official statistical data from Indonesian government institutions on a linear logarithmic specification within a causality framework. Another study by Cahyo et al. (2023) aimed to identify the environmental, population, and economic impacts on CO_2 emissions in Indonesia from 1990 to 2021, employing multiple linear regression analysis with numeric data from various international organizations and the Central Bureau of Statistics. Although both studies provide valuable insights, they are limited by their reliance on conventional numerical data and statistical techniques, which have inherent weaknesses compared to more modern methods like Machine Learning (ML) algorithms. A recent 2023 study in Indonesia took a novel approach by using spatial data, such as satellite imagery. Swardika and Santiary (2023) attempted to model past energy consumption patterns spatially (location) and temporally (year) using nighttime satellite dataset. The modelling technique used was curve-fitting, evaluating indicators through R-squared (R2) and Root Mean Square Error (RMSE) values. However, this study's limitation lies in its inability to predict future energy consumption, only identifying and modelling patterns or relationships in data plotted as curves or graphs. Additionally, Farida et al. (2023) analyzed factors influencing CO_2 emissions in Indonesia and projected an increase up to 2030 using the econometric method Vector Error Correction Model (VECM). The findings suggest that energy consumption is a primary factor contributing to the rise in CO_2 emissions, with an estimated increase of 65% from 2020 to 2030 in Indonesia. These results underscore the potential for developing more sophisticated models to understand, predict, and manage energy consumption in Indonesia, especially through innovative approaches such as ML.

With the rapid development of networks, data storage, and data collection capabilities, Big Data has swiftly expanded across all fields of science and engineering (Su et al., 2020). Numerous studies have explored the use of Big Data, including satellite imagery, to estimate CO_2 emissions and energy efficiency (Baldwin et al., 2017; Falchetta & Noussan, 2019; Townsend & Bruce, 2010). These data sources offer solutions for gathering electricity and energy-related data at small scales and low costs. Recent advancements in the energy sector are propelled by the utilization of Big Data and advanced analytical methods such as ML. Various ML algorithms have been applied to predict energy consumption with varying degrees of success. Wu & Chu (2021) found that the random forest algorithm was most effective for constructing energy consumption predictions, while Islam et al. (2023) evaluated energy consumption from various ML classifier algorithms, recommending Gaussian Naive Bayes as the most effective. Mohapatra et al. (2021) provided a comprehensive review of computational intelligence approaches for energy consumption prediction, highlighting the potential of ML and deep-learning models in predicting energy consumption.

To the best of our knowledge, the utilization of big data technology and ML algorithms for estimating and predicting energy efficiency in lighting within Indonesia remains relatively limited. Previous studies have fallen short in evaluating and forecasting energy efficiency, particularly in the context of urbanization, as well as the use of Big Data and ML approaches in future modelling. Typically, assessing energy efficiency in lighting levels necessitates expensive energy audits. This underscores the need for in-depth research in this area to comprehensively understand the dynamics of energy efficiency in Indonesia.

Therefore, our study aims to fill these gaps. Our novel study refines the literature insights on energy efficiency modelling in Indonesia. We focus on evaluating past dynamics of energy efficiency and forecasting future trends by leveraging Big Data and ML approaches. This approach will form an integral part of the data collection and analysis process in this research, and this is expected to provide deep and detailed insights into energy efficiency from the past to the future across various provinces in Indonesia. The uniqueness of this research lies in its further contribution to the literature on energy efficiency modelling. The outcomes of this study are expected to enhance understanding of regional disparities in energy efficiency in both temporal and spatial terms and provide a robust scientific basis for formulating effective CO_2 emission mitigation policies aligned with the energy efficiency characteristics of each province, prioritizing the most critical efforts.

2. Methods and Materials

This study is a quantitative research that focuses on the application of several algorithms and statistical models in the domain of ML. The ultimate goal of this research is to predict future energy efficiency levels with factors from urbanization (human activity intensity, population density, and built-up land development), thereby formulating policy recommendations related to development and efforts to mitigate CO_2 emissions in various provinces in Indonesia. The prediction process is carried out based on the best model generated from several ML algorithms, including SVM, ANN, and RF. These algorithms allow computers to learn from data without needing explicit instructions. The integration of these algorithms is done using the Google Colab platform. Analysis and predictions are conducted for all provinces in Indonesia. As one of the largest countries in the world, human activity patterns, population, and energy consumption in Indonesia vary greatly among provinces and regions. Therefore, a comprehensive analysis of CO_2 emissions is needed to formulate appropriate policies.

2.1 Data Collection

Data plays a crucial role in research as it forms the foundation of structured and essential information. However, conventional data tends to have several limitations, such as scope and detail limitations, delays in availability, as well as costs and difficulties in collection. Additionally, conventional data often has rigid formats, making it difficult to integrate with other data or analyze using modern analysis methods such as ML or spatial analysis. Conversely, with technological advancements and the widespread use of Big Data, researchers, as well as urban and regional planners, are beginning to understand the importance of Big Data in uncovering temporal and spatial patterns related to urban and regional contexts, including energy consumption patterns and carbon emissions. Based on this justification, an attempt has been made to focus on the utilization of Big Data by collecting various digital data sources. The details of the dataset used will be elaborated in detail in Table 1 and visualized in Figure 1.

The data extracted from WorldPOP for the years 2021–2023 is the result of extrapolation from previous years. Subsequently, due to varying resolutions of the extracted data, an upscaling process was performed to standardize the resolution to 1000 meters. Thus, the data is prepared for analysis at the raster level.

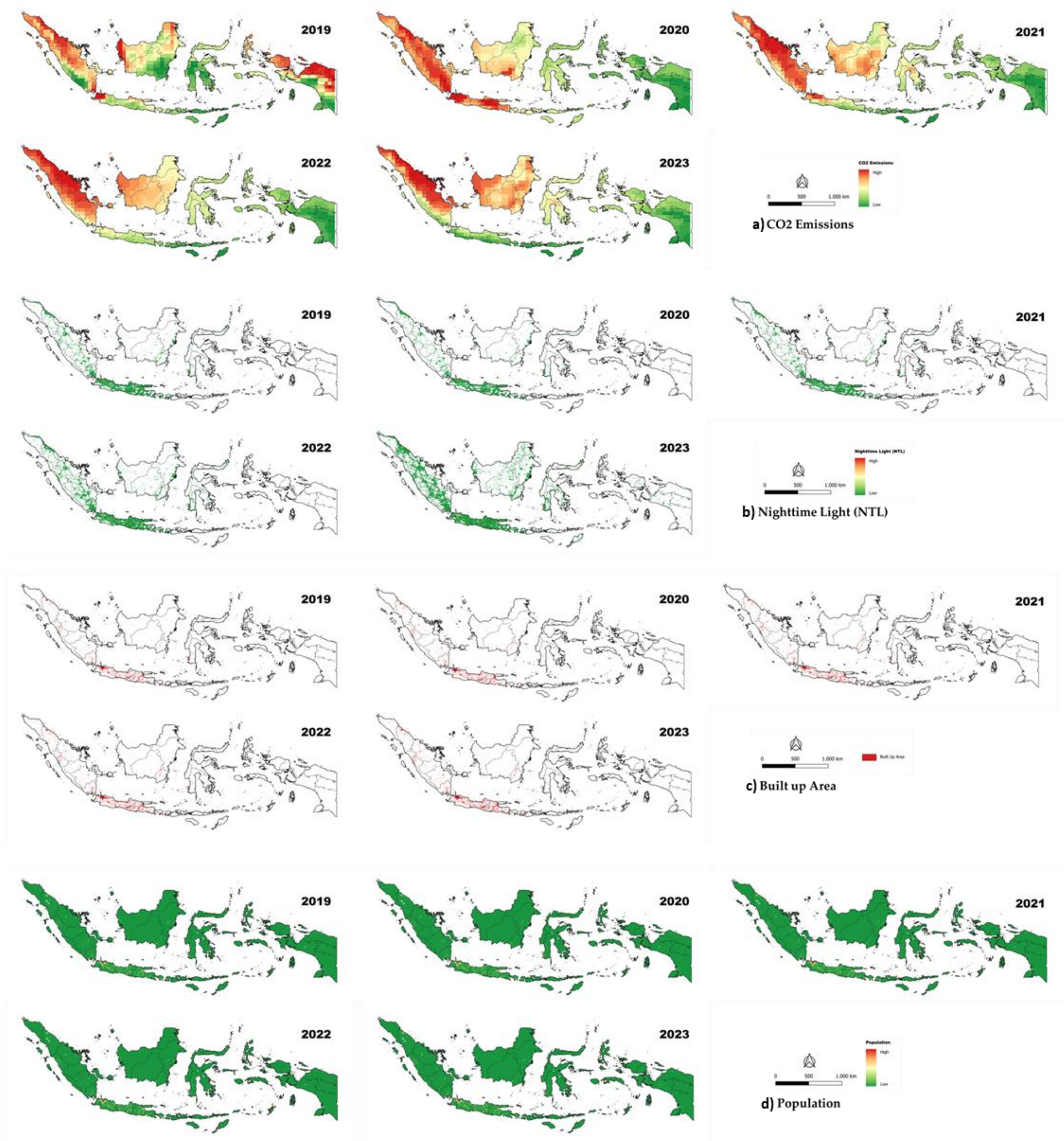

Figure 1. Research spatial data (from 2019–2023), a) CO_2 emissions; b) NTL; c) Built up area; d) Population.

Table 1. Research dataset.

Data	Data source	Remarks
Administrative boundaries	Indonesian Geospatial Information Agency	▪ Data accessed on November 2023
Human activities	Night-time Light Imagery (NTL)	▪ VIIRS, Annual NTL (2019–2023) ▪ Data accessed on 15 April 2024 and downloaded at https://eogdata.mines.edu/products/vnl/#v1
Population density	WorldPOP	▪ Population density, resolution 1km (2019–2020, 2021–2024 extrapolated) ▪ Data accessed on 15 April 2024 and downloaded at https://www.worldpop.org/datacatalog/
Built-up land	Esri Land Cover	▪ Sentinel-2, 10-Meter Land Cover ▪ Data accessed on 15 April 2024 and downloaded at https://livingatlas.arcgis.com/landcover/
CO_2 emissions	The OCO-2 Satellite by NASA's Goddard Space Flight Center	▪ Spatial resolution: 0.5 ° x 0.625 ° ▪ Data units: Parts per million (ppm) ▪ Data accessed on 15 April 2024 and downloaded at https://earth.gov/ghgcenter/data-catalog/oco2geos-co2-daygrid-v10r

2.2 Analysis Process

The data generated from various digital data sources is raw data that is not yet ready for further analysis, thus requiring a preprocessing or pre-analysis stage. The preprocessing stage is a process performed on data before it is processed by an algorithm or model. The goal of the preprocessing stage is to clean, normalize, and transform raw data into a format that is easier to understand and can be processed by a model. One of the preprocessing stages that will be undertaken is Zonal Statistics, which is a spatial analysis technique for calculating statistics within a specific area (zone) by considering the pixel values or spatial grid data within that area (Singla & Eldawy, 2020; Erdem et al., 2021). This technique is applied to various dependent and independent factors in this study, including human activity intensity, population density, built-up land development, and CO_2 emissions, to generate statistical information such as averages in each zone (Figure 2).

After the preprocessing stage, the next step involves ML using three different algorithms or models, namely SVM, ANN, and RF. After modelling using these three algorithms, the model evaluation process is conducted by considering the RMSE and Standard Deviation values. RMSE is a commonly used metric in statistics and ML to evaluate the accuracy of a model (Chai & Draxler, 2014). Models with the smallest RMSE are generally considered the best models (Chai & Draxler, 2014; Mentaschi et al., 2013; Moriasi et al., 2007; Singh et al., 2005). Meanwhile, Standard Deviation is used to measure how far data is spread out from its mean value in a data distribution. The result provides information about the variability or diversity of the data. In other words, standard deviation provides information on how much dispersion or difference there is between each data point and the mean value. The larger the standard deviation, the greater the variability in the data, and vice versa (Pfeiffer, 1990).

The next step is to predict energy efficiency until 2045 using historical growth data from the three independent variables and using the best-performing algorithm model based on RMSE and standard deviation values. The entire modelling process in this research utilizes the Google Colab platform. Google Colab is a free cloud platform that allows users to write and run Python code in a web browser. This platform eliminates the need for complex local installation and configuration, thus saving time and resources. The main advantage of Google Colab is the utilization of Google's advanced cloud infrastructure to accelerate intensive computing, such as ML model training and large-scale data analysis. The analysis process is detailed through the research framework in Figure 3.

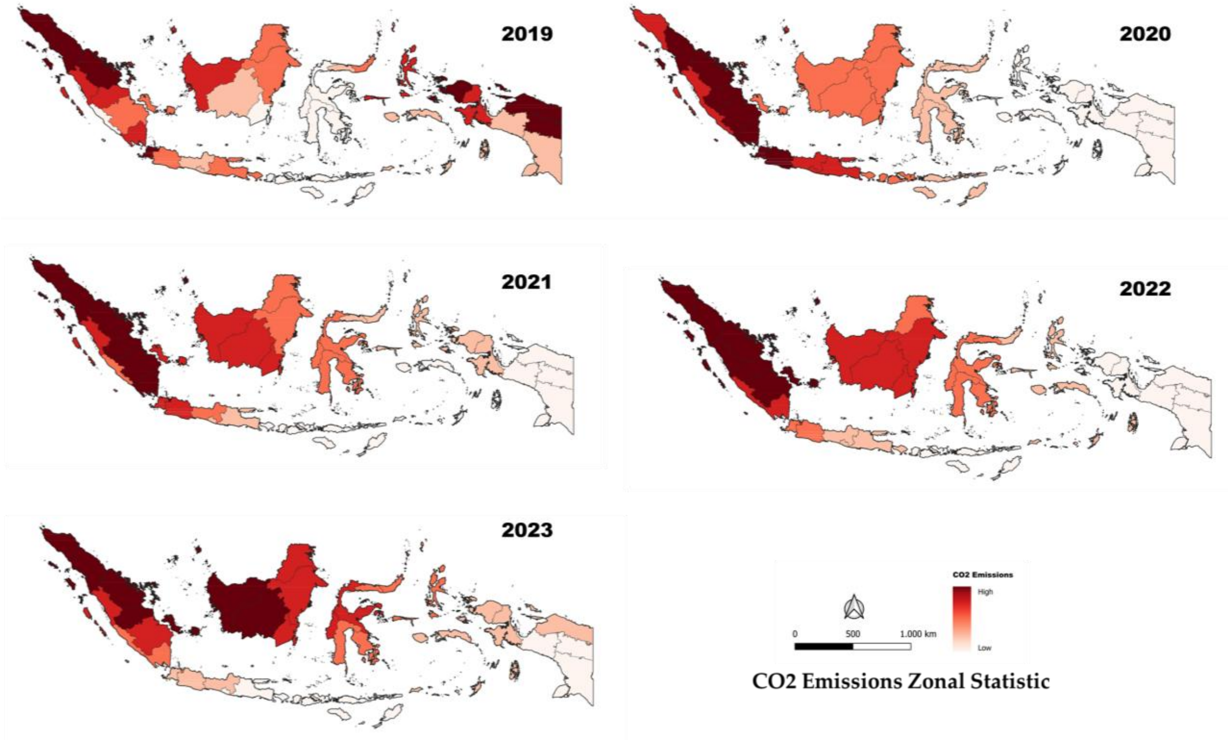

Figure 2. CO_2 Emissions zonal statistics.

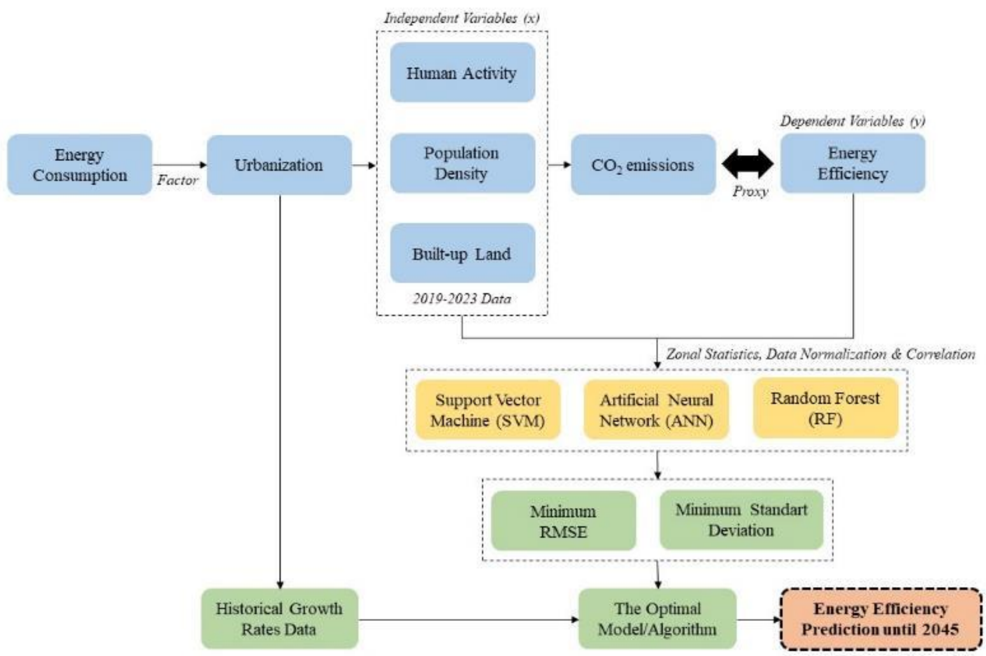

Figure 3. Conceptual research framework.

3. Results and Discussions

3.1 Dynamics of Energy Consumption Factors Development

In the last five years, the development of energy consumption factors has shown diverse dynamics. Analyzing the trends in energy consumption during this period is important for understanding patterns and their implications for the sustainability of energy resources and the environment. Figure 4 is a visualization in the form of a graph depicting the dynamics of the development of energy consumption factors.

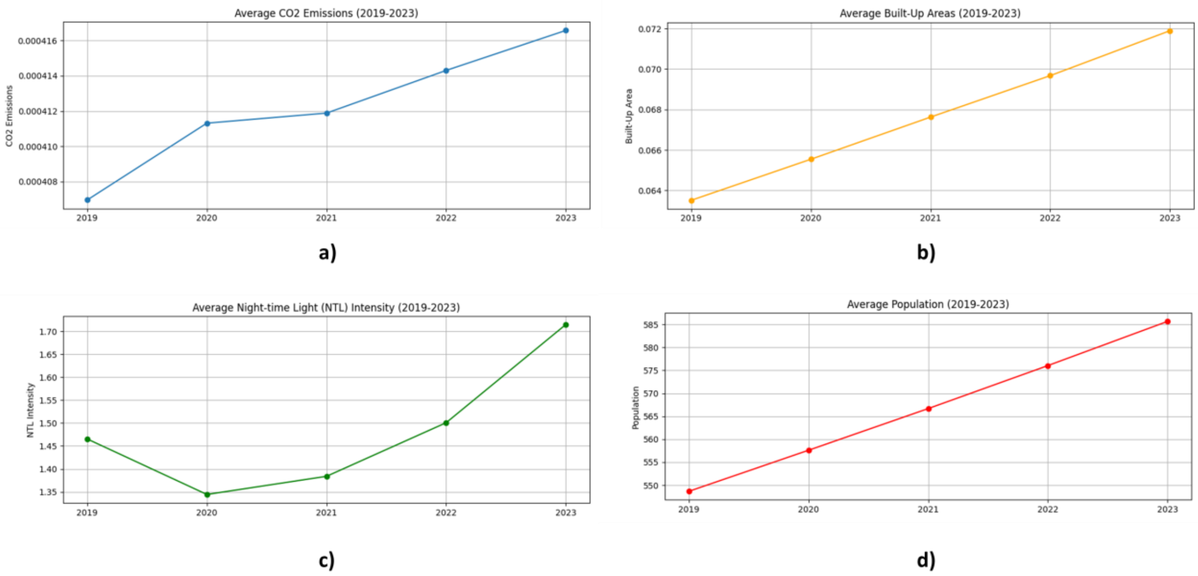

Figure 4. The average development of each variable from 2019 to 2023, a) Average CO_2 emissions; b) Average built-up areas; c) Average NTL intensity; d) Average population.

Figure 4 depicts the visualization of the development of CO_2 emissions and several associated factors, including built-up area, nighttime light intensity, and population. It is known that all the aforementioned variables exhibit diverse trends. CO_2 emissions show an increase each year. In contrast, the other two factors, namely built-up area and population, tend to experience a stable and sustainable increase from year to year. Meanwhile, the human activity factor tends to fluctuate.

3.2 The Correlation Among Variables

The following section presents a correlation analysis between CO_2 emissions and three key factors: built-up area, human activity, and population. By examining these relationships, we aim to gain insights into the extent and nature of the correlation between CO_2 emissions and each factor. This analysis enables us to assess whether increases in built-up area, nighttime light intensity, or population size have a positive, negative, or negligible association with CO_2 emissions.

Figure 5 illustrates a heatmap visualization depicting the correlation between CO_2 emissions and the three factors (built-up area, human activity, and population) from 2019 to 2023. Each cell in the heatmap represents the correlation coefficient between two variables, with a color scale indicating the strength of the correlation, where blue indicates low values and red indicates high values. From the heatmap above, it can be observed that in 2019, the correlation of variables with CO_2 emissions ranged from 0.598 to 0.628. This indicates a relatively strong relationship, signifying that CO_2 emissions have a positive and significant correlation with built-up areas, human activity, and population. This relationship suggests that an increase in one variable tends to be followed by an increase in the other variables. The same relationship is identified in the following year, 2020.

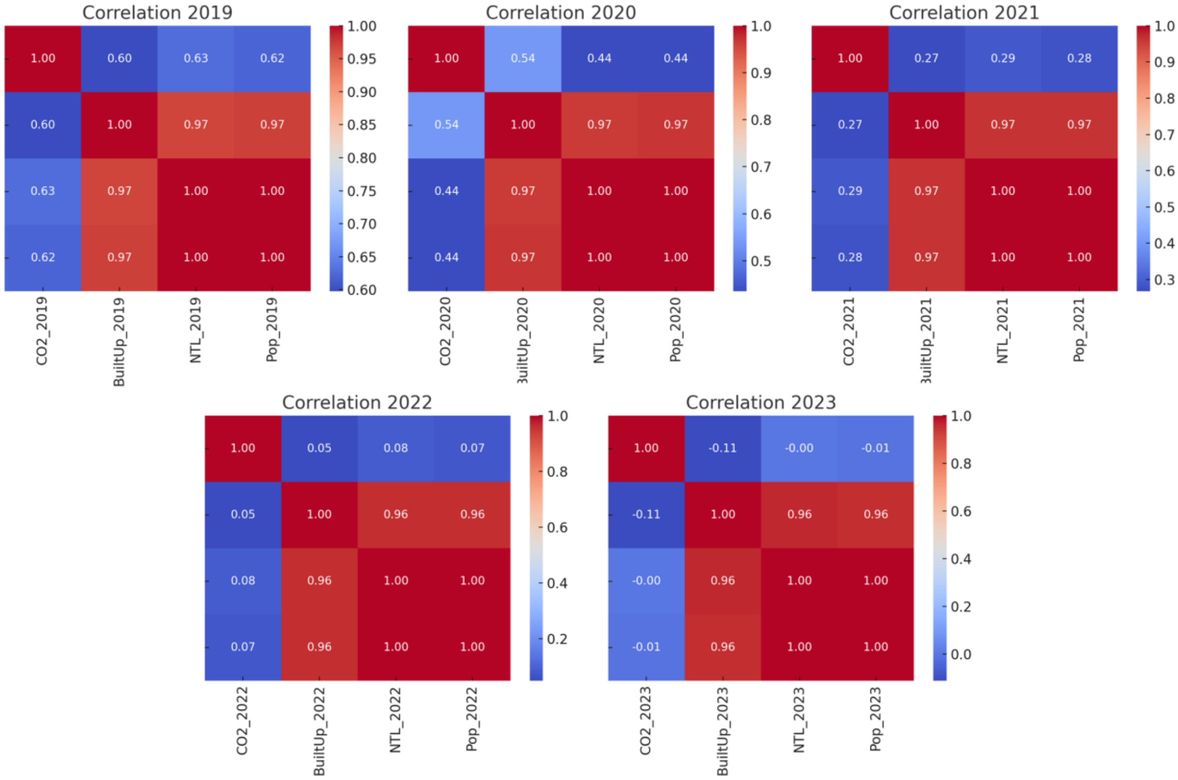

Figure 5. Visualization of the average development of each variable from 2019–2023.

However, there is a shift in the trend in the years 2021–2023. In 2021, the correlation coefficient of variables decreases to approximately 0.267–0.289. In 2022, the correlation continues to decrease to very low values, only around 0.048–0.084. In 2023, the correlation tends to reverse, with a negative correlation value for built-up area with CO_2 emissions at around -0.0112 and almost no correlation for human activity and population variables.

These shifts in correlation could be attributed to several other factors such as changes in economic conditions or policies, for example, significant changes in industrial or energy sector policies affecting CO_2 emissions. Additionally, there may be influences from extraordinary events such as pandemics, natural disasters, land, forest, and peat fires, or other events that alter normal human activity patterns. This underscores the importance of considering external context and data quality when drawing conclusions from correlations.

3.3 Energy Efficiency Modelling

Before predictions can be made, ML builds a model to predict CO_2 emissions using three factors including building area, human activity and population. The data used to train this model is from 2019 to 2022, and the validity of the model was tested using 2023 data. ML carried out the steps taken to train the model, including:

a. Data preparation for training, including separating feature data and target data.
b. Training on three types of models, namely RF, SVM, and ANN.
c. Validate the model using cross-validation techniques to ensure the model performs well on unseen data.
d. Produce an evaluation and comparison of model performance to determine which is most suitable.

The prepared data is then divided into training and testing sets, with a ratio of 30 samples for training and 8 samples for testing. Cross-validation was then performed to evaluate the performance of each model and produce the mean RMSE and standard deviation for each model, as in Table 2.

Table 2. RMSE and Standard Deviation of RF, SVM, and ANN Models.

Model	**RMSE**	**Standard Deviation**
Random Forest	$4{,}78\times10{-}7$	$1{,}26\times10{-}7$
Support Vector Machine (SVM)	$5{,}69\times10{-}7$	$1{,}10\times10{-}7$
Artificial Neural Network (ANN)	13,76	20,26

Table 2 shows that the RF model consistently performs well on each sample of the research dataset due to having the lowest RMSE compared to the other models. The SVM model is also good but still performs worse than the RF model. Meanwhile, the ANN model tends to be less stable and not suitable for the data in this study compared to the other two models. Based on this modelling, it is concluded that the RF model is the best choice for predicting CO_2 emissions based on built-up area, human activity, and population factors. This is because it has better stability and accuracy compared to the other models.

3.4 Energy Efficiency Prediction

Based on the modelling conducted, the RF model appears to be the best choice for predicting CO_2 emissions based on the variables of human activity, population density, and built-up area. The RF model offers a good combination of accuracy and stability. To predict CO_2 emissions per province for future years, including 2025, 2030, 2035, 2040, and 2045 using the RF model, it is necessary to consider how to protect the values of independent variables (human activity, population density, and built-up area) for those years. An approach that can be used is data extrapolation, which involves estimating variable values based on existing trends.

The next step is to estimate the values of the independent variables (human activity, population density, and built-up area) for future years based on the annual average growth from the available data. First, the annual average growth for each feature from 2019 to 2023 will be calculated and then used to project these values for the years 2025, 2030, 2035, 2040, and 2045. After that, the trained RF model will be used to predict CO_2 emissions based on the projected values.

Figure 6 displays the projection of total CO_2 emissions from all provinces from 2025 to 2045. As observed, there is a tendency to increasing total CO_2 emissions from 2025 to 2030, followed by a stable decline from 2030 to 2045, according to the RF model predictions. Meanwhile, the analysis of the trend of predicted CO_2 emissions per province, as shown in Figure 7, reveals that only two provinces in Indonesia experience an increase in CO_2 emissions. Four other provinces are projected to experience a significant decrease in CO_2 emissions, namely North Sumatra, West Sumatra, South Kalimantan, and Riau Islands. Meanwhile, other provinces show fluctuations or stagnation in CO_2 emission rates.

Figure 8 displays the average growth of CO_2 emissions per province. From the analysis results, it is revealed that provinces showing positive average growth or the highest increase in CO_2 emissions are generally located in the eastern regions of Indonesia, such as South Papua, Highlands Papua, East Nusa Tenggara, and Maluku. On the other hand, provinces showing negative average growth or a decrease in CO_2 emissions include North Sumatra, South Kalimantan, and Riau Islands.

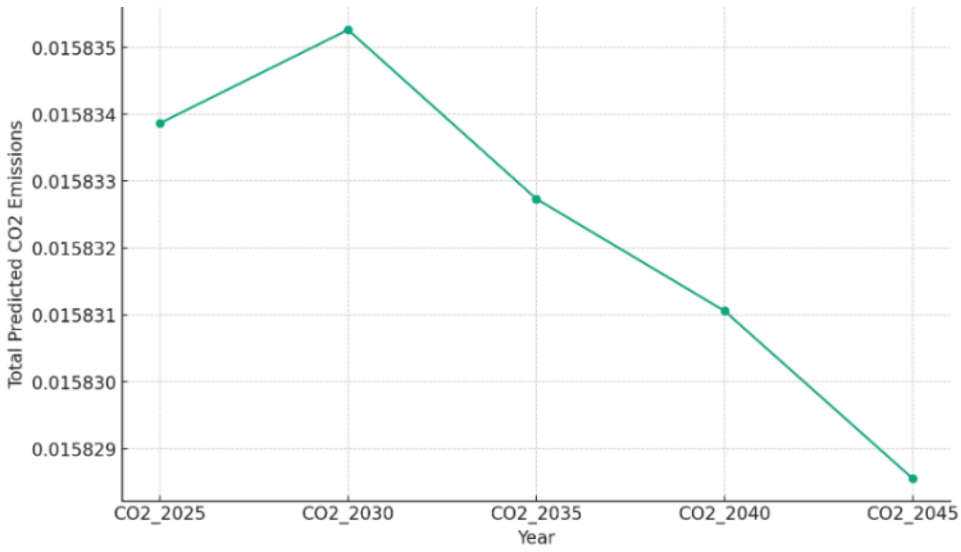

Figure 6. Total Predicted CO_2 emissions from 2025–2045.

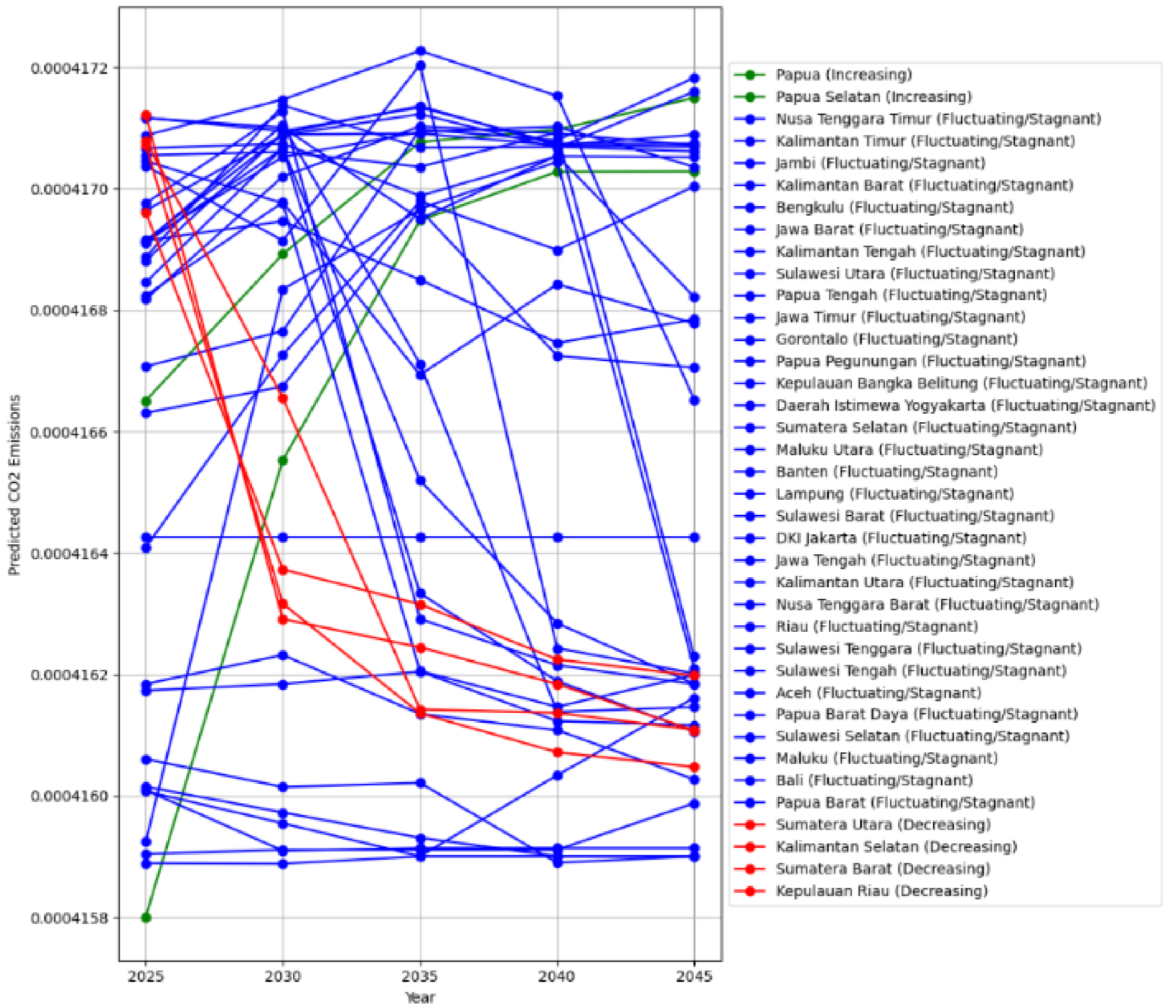

Figure 7. Predicted CO_2 emissions per province categorized by trends.

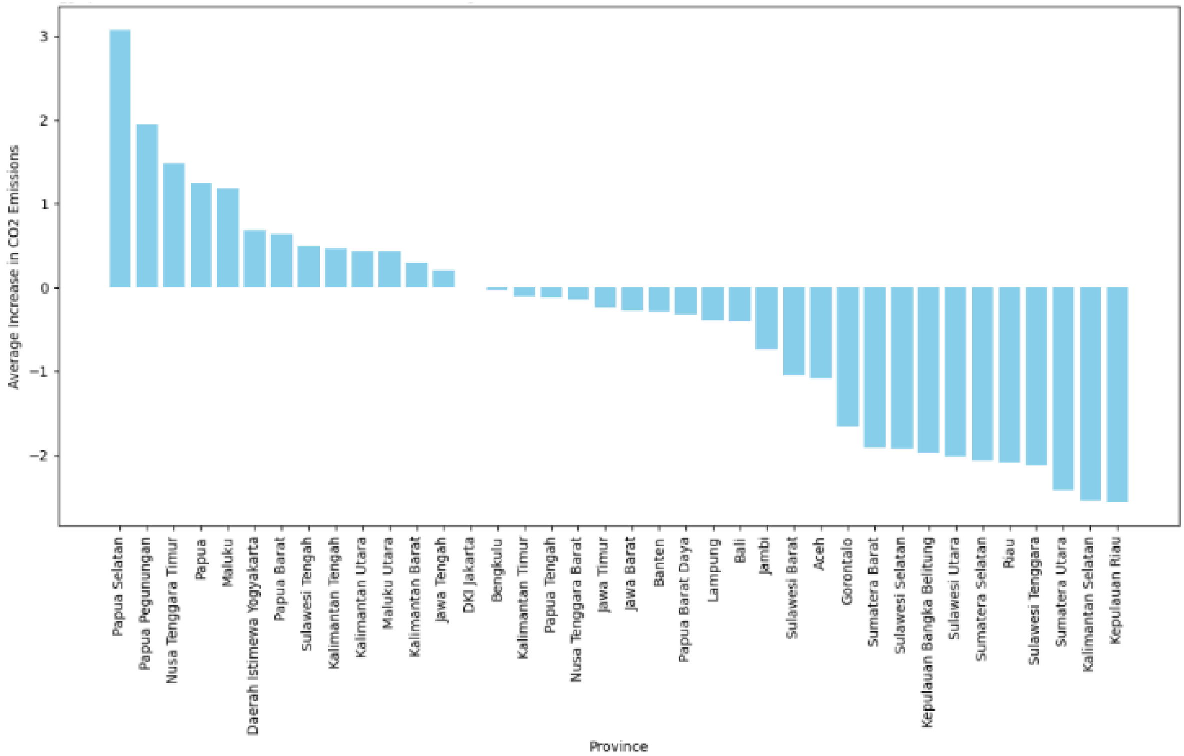

Figure 8. Average growth in CO_2 emissions per province**.**

3.5 Discussion

Based on the data on CO_2 emissions development in various provinces in Indonesia from 2019 to 2023, there is a noticeable upward trend in emissions. However, upon further analysis, only the human activity factor shows a decline, particularly evident from the decrease in nighttime light values in 2020, which can be attributed to the COVID-19 pandemic drastically reducing human activity. Meanwhile, factors such as population density and built-up area show stability in their increase during this period. This indicates a positive relationship between population growth and built-up area expansion, affirming the ongoing urbanization dynamics in Indonesia, where population growth often impacts built-up land expansion.

We have attempted to model and predict CO_2 emissions as a proxy for energy efficiency in Indonesia. Our results indicate that the RF algorithm is the best model capable of capturing the dynamics of CO_2 emission-producing factors in Indonesia from various aspects of urbanization. The prediction results show that total CO_2 emissions tend to increase from 2025 to 2030. This is indicative of the direct impact of urbanization and rapid economic growth, such as the growth of human activity, population density, and land development, which also increase over time, directly contributing to CO_2 emissions. These findings are also consistent with research conducted by Farida et al. (2023), where econometric models predicted a 65% increase in CO_2 emissions from 2020 to 2030 in Indonesia. However, interestingly, after peaking in 2030, our research using the RF model predicts that total CO_2 emissions will gradually decline until 2045.

The decline is indicated by various government policies in Indonesia aimed at reducing CO_2 emissions, particularly in the context of urbanization. At the Conference of the Parties (COP) 26 or COP26 Glasgow, Indonesia reaffirmed its commitment to achieve Net Zero Emissions by 2060 or even earlier, and in response, the government has implemented various policies. The Ministry of Public Works and Housing, for example, has committed to reducing carbon emissions through sustainable construction practices and the development of green infrastructure. Furthermore, Ministerial Regulation Number 14 of 2021 has been enacted to implement Minimum Energy Performance Standards (MEPS) for energy-using equipment, with the aim of improving energy efficiency and providing information to users about energy-efficient equipment. This regulation, along with Ministerial Decree No. 135.K.EK.07/DJE/2022, establishing minimum energy performance standards and energy-saving labels for LED lamps, is intended to prevent inefficient household appliance products from entering the Indonesian market, with a primary focus on energy savings.

Additionally, there is an energy transition that will gradually reduce the use of coal and promote the development of power plants from renewable energy sources. The Indonesia Green Growth Program, initiated by the government together with the Global Green Growth Institute (GGGI) and the Ministry of National Development Planning (BAPPENAS), supported by several ministries and local governments, aims to promote green growth and increase investment in renewable energy projects and energy efficiency, with the hope of creating innovative and creative financing schemes. However, investments related to renewable energy still tend to be concentrated in the Kalimantan and Nusa Tenggara regions.

Our predictive results align with various statements in global studies stating that energy demand or consumption can be reduced by up to 50% before 2050 by improving energy efficiency and promoting energy-saving behaviors (Min et al., 2022; Svenfelt et al., 2011; Ürge-Vorsatz et al., 2012). The concept of energy efficiency has been widely promoted as one of the indicators for achieving sustainable development (Di Foggia, 2018; Soltangazinov et al., 2020; Zakari et al., 2022). The International Renewable Energy Agency has also emphasized that, in response to climate change, many countries have developed ambitious renewable energy plans to achieve zero net emissions goals by 2050 (International Renewable Energy Agency, 2018), including Indonesia. Kusumadewi & Limmeechokchai (2015) showed that by implementing various energy efficiency scenarios in Indonesia, energy can be saved by 27.6% of total energy demand in 2050, while cumulative CO_2 emissions can be reduced by 16% of total CO_2 emissions in 2050.

Figure 9 illustrates the pattern of CO_2 emissions development across various provinces in Indonesia. In 2045, the highest CO_2 emissions are found in North Kalimantan, East Kalimantan, Central Kalimantan, Central Sulawesi, Maluku, North Maluku, West Papua, and Papua. This indicates that the lowest energy efficiency is found in those 9 provinces. It can be observed that the highest emission levels in the final year of the prediction are generally located in Kalimantan Island and the eastern regions of Indonesia. However, upon examining the analysis of the CO_2 emission prediction trends per province, it is noted that only two provinces experience an increase in CO_2 emissions, namely Papua and South Papua. This highlights the differences in emission dynamics between provinces. More interestingly, it was found that four other provinces are expected to undergo a significant reduction in CO_2 emissions, namely North Sumatra, West Sumatra, South Kalimantan, and Riau Islands. One justification for this phenomenon is that the decrease in CO_2 emissions in these four provinces may be due to local policies or initiatives to reduce deforestation or convert degraded land, along with more aggressive mitigation efforts in the energy or plantation sectors (especially palm oil). Provinces showing fluctuations or stagnation in CO_2 emission levels require higher policy support to achieve CO_2 reduction. A thorough evaluation of the implemented policies is needed, involving an assessment of the effectiveness of current policies and adjustment of strategies to better align with national CO_2 reduction goals. Inter-provincial cooperation can also be enhanced through the exchange of knowledge and experiences in addressing CO_2 emission-related challenges, allowing these provinces to learn from each other and improve their performance in reducing CO_2 emissions.

When considering their average growth rates, provinces in eastern Indonesia, such as South Papua, Papua Highland, East Nusa Tenggara, and Maluku, tend to exhibit positive growth in CO_2 emissions. This is indicated by the challenges these regions may face in controlling CO_2 emissions, such as rapid infrastructure development, land conversion leading to deforestation, or intensive industrial growth that heavily relies on energy consumption. This aligns with previous studies, emphasizing the crucial role of energy in accelerating economic and social development, particularly in areas with limited infrastructure and economic development (Bologna, 2013; Ibrahim et al., 2023; Lloyd, 2017). On the other hand, there are provinces showing negative average growth or decline in CO_2 emissions, including North Sumatra, South Kalimantan, and Riau Islands. This decrease in CO_2 emissions may be related to the implementation of pro-environment policies or the adoption of green economy practices in the energy sector.

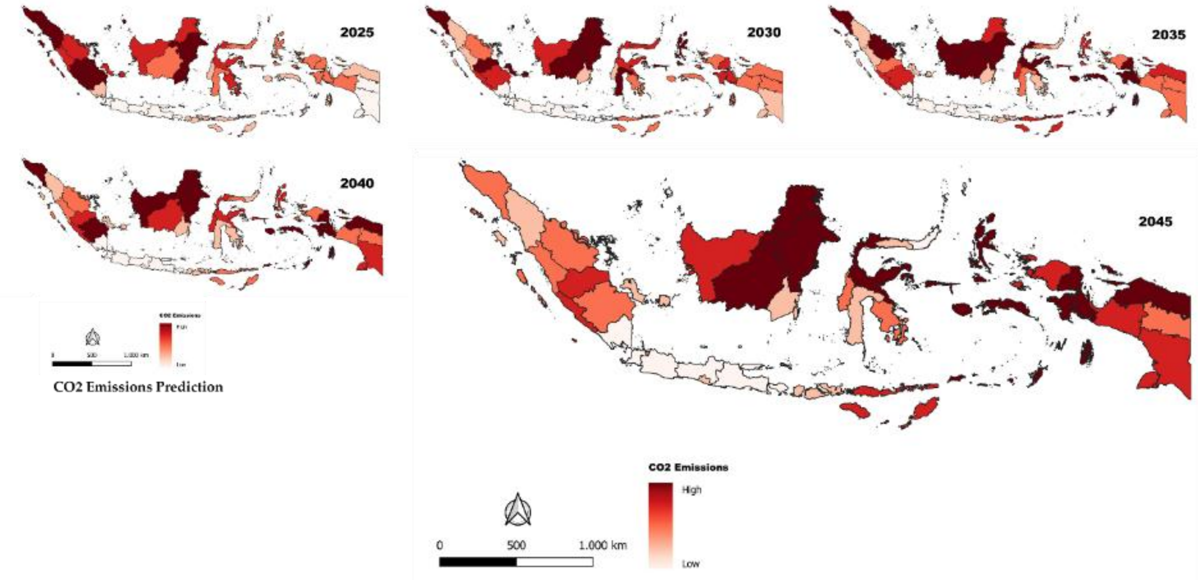

Figure 9. CO_2 Emission prediction based on RF algorithm.

This study also underscores the importance of considering urbanization factors such as human activities, built-up areas, and population density in designing CO_2 emission mitigation policies. With rapid urbanization, human activities in urban areas become the primary contributors to CO_2 emissions. Therefore, policies aimed at more efficient built-up land management and sustainable urban development can help reduce their negative impact on the environment and carbon emissions. Additionally, increased population density in urban areas demands careful planning to provide efficient public transportation, green infrastructure, and environmentally friendly public services to reduce reliance on private vehicles and fossil fuels, ultimately aiding in significant CO_2 emission reduction.

4. Conclusions

Our study represents a pioneering effort, showcasing the effectiveness of big data technology and machine learning algorithms in predicting and modelling energy efficiency across different provinces in Indonesia. The research findings indicate that the RF algorithm performs the best based on its low RMSE compared to the SVM and ANN algorithms. The predictions show a trend of increasing CO_2 emissions until 2030, followed by a stable decrease until 2045, reflecting the effects of urbanization and varied energy policy implementations across provinces.

As a contribution, our research reinforces the importance of adaptive and regionally-based policies in managing energy efficiency in Indonesia. The study suggests that with the right approach to data collection and predictive modelling, a deeper understanding of how to significantly improve energy efficiency, thereby reducing CO_2 emissions, can be achieved. Implementing policies based on the results of these models will help Indonesia not only meet its global emission reduction commitments but also promote broader sustainable development.

Although providing valuable insights, this study has several limitations that may affect the accuracy and applicability of its findings. One major limitation is the use of limited datasets, which may not fully represent real-world conditions due to constraints in socio-economic variables and ongoing government policy influences. These limitations potentially affect the accuracy of the developed model predictions. Given these shortcomings, future research should aim to expand the availability and diversity of data and incorporate other variables that may influence energy efficiency. More complex modelling approaches are also recommended to accommodate more dynamic interactions between different variables. Implementing these models would significantly support the formulation of more adaptive and effective energy policies, particularly in adjusting to the specific conditions and needs of each province.

Acknowledgments

We would like to express our sincere gratitude to PT Sagamartha Ultima Indonesia for funding this research. Their generous support has been instrumental in enabling us to conduct this study and advance our understanding of energy efficiency modelling in Indonesia. We also extend our appreciation to all individuals and organizations who have contributed their expertise and assistance throughout the course of this research. Their valuable insights and support have greatly enriched our work.

References

Baldwin, H. B., Klug, M., Tapracharoen, K., & Visudchindaporn, C. (2017). Utilizing Suomi NPP's day-night band to assess energy consumption in rural and urban areas as an input for poverty analysis. *AGU Fall Meeting Abstracts*, *2017*, IN52B-09. https://ui.adsabs.harvard.edu/abs/2017AGUFMIN52B..09B

Bilgili, F., Koçak, E., Bulut, Ü., & Kuloğlu, A. (2017). The impact of urbanization on energy intensity: Panel data evidence considering cross-sectional dependence and heterogeneity. *Energy*, *133*, 242–256. https://doi.org/10.1016/j.energy.2017.05.121

Bologna, S. (2013). Energy and sustainable economic development. In *Renewable energy for unleashing sustainable development* (pp. 181–193). Springer International Publishing. https://doi.org/10.1007/978-3-319-00284-2_8

Cahyo, H., Purnomo, S. D., Octisari, S. K., Surveyandini, M., Sundari, S., & Purwendah, E. K. (2023). Environment, population, and economy on CO2 emission in Indonesia. *International Journal of Energy Economics and Policy*, *13*(6), 295–303. https://doi.org/10.32479/ijeep.14938

Cao, X., Wang, J., Chen, J., & Shi, F. (2014). Spatialization of electricity consumption of China using saturation-corrected DMSP-OLS data. *International Journal of Applied Earth Observation and Geoinformation*, *28*(1), 193–200. https://doi.org/10.1016/j.jag.2013.12.004

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Di Foggia, G. (2018). Energy efficiency measures in buildings for achieving sustainable development goals. *Heliyon*, *4*(11). https://doi.org/10.1016/j.heliyon.2018.e00953

Elliott, R. J. R., Sun, P., & Zhu, T. (2017). The direct and indirect effect of urbanization on energy intensity: A province-level study for China. *Energy*, *123*, 677–692. https://doi.org/10.1016/j.energy.2017.01.143

Erdem, F., Atun, R., Yigit Avdan, Z., Atila, I., & Avdan, U. (2021). Drought analysis of Van Lake Basin with remote sensing and GIS technologies. Egyptian Journal of Remote Sensing and Space Science, 24(3), 1093–1102. https://doi.org/10.1016/j.ejrs.2021.10.006

Falchetta, G., & Noussan, M. (2019). Interannual variation in night-time light radiance predicts changes in national electricity consumption conditional on income-level and region. *Energies*, *12*(3). https://doi.org/10.3390/en12030456

Farida, Y., Siswanto, N., & Vanany, I. (2023). Forecasting CO2 emission in Indonesia from the economic and environment impact using vector error correction model. 2*023 5th International Conference on Management Science and Industrial Engineering*, 42–49. https://doi.org/10.1145/3603955.3603963

Fehrer, D., & Krarti, M. (2018). Spatial distribution of building energy use in the United States through satellite imagery of the earth at night. *Building and Environment*, *142*, 252–264. https://doi.org/10.1016/j.buildenv.2018.06.033

Guan, X., & Zhou, M. (2015). Decomposing the effecting factors of urbanization on energy consumption of China. *The Open Cybernetics & Systemics Journal*, *9*(1), 2035–2043. https://doi.org/10.2174/1874110X01509012035

Guo, W., Li, Y., Li, P., Zhao, X., & Zhang, J. (2022). Using a combination of nighttime light and MODIS data to estimate spatiotemporal patterns of CO2 emissions at multiple scales. *Science of the Total Environment*, *848*. https://doi.org/10.1016/j.scitotenv.2022.157630

Hipskind, S., Elvidge, C., Gurney, K., Imhoff, M., Bounoua, L., Sheffner, E., Nemani, R., Pettit, D., & Fischer, M. (2011). Global night-time lights for observing human activity. *34th International Symposium on Remote Sensing of Environment - The GEOSS Era: Towards Operational Environmental Monitoring*. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84879760185&partnerID=40&md5=6fad1ed654df2b3f6488d4ada0bd3ee7

Hu, W., & Fan, Y. (2020). City size and energy conservation: Do large cities in China consume more energy? *Energy Economics*, *92*, 104943. https://doi.org/10.1016/j.eneco.2020.104943

Ibrahim, M. D., Alola, A. A., & Ferreira, D. C. (2023). Assessing sustainable development goals attainment through energy-environmental efficiency: The case of Latin American and Caribbean countries. *Sustainable Energy Technologies and Assessments*, *57*(December 2021), 103219. https://doi.org/10.1016/j.seta.2023.103219

International Renewable Energy Agency. (2018). Net Zero by 2050 - a roadmap for the global energy sector. *International Renewable Energy Agency*. https://www.irena.org/-/media/Files/IRENA/Agency/Publica tion/2018/Apr/IRENA_Report_GET_2018.pdf.

Islam, M. S., Zisad, S. N., Kor, A.-L., & Hasan, M. H. (2023). Sustainability of machine learning models: an energy consumption centric evaluation. *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1–6. https://doi.org/10.1109/ECCE57851.2023.10101532

Jafari, Y., Othman, J., & Nor, A. H. S. M. (2012). Energy consumption, economic growth and environmental pollutants in Indonesia. *Journal of Policy Modelling*, *34*(6), 879–889. https://doi.org/10.1016/j.jpolmod.2012.05.020

Jian, L., Sohail, M. T., Ullah, S., & Majeed, M. T. (2021). Examining the role of non-economic factors in energy consumption and CO2 emissions in China: policy options for the green economy. *Environmental Science and Pollution Research*, *28*(47), 67667–67676. https://doi.org/10.1007/s11356-021-15359-3

Jones, D. W. (1991). How urbanization affects energy-use in developing countries. *Energy Policy, 19(7), 621–630*. https://doi.org/10.1016/0301-4215(91)90094-5

Kusumadewi, T. V., & Limmeechokchai, B. (2015). Energy efficiency improvement and CO2 mitigation in residential sector: comparison between Indonesia and Thailand. *Energy Procedia, 79*, 994–1000. https://doi.org/10.1016/j.egypro.2015.11.599

Lee, P., Lam, P. T. I., Yik, F. W. H., & Chan, E. H. W. (2013). Probabilistic risk assessment of the energy saving shortfall in energy performance contracting projects-A case study. *Energy and Buildings*, *66*, 353–363. https://doi.org/10.1016/j.enbuild.2013.07.018

Lin, H., & Zhai, X. (2023). Energy efficiency through user adoption of the sharing economy leading to environmentally sustainable development. *Journal of Innovation and Knowledge*, *8*(1). https://doi.org/10.1016/j.jik.2023.100315

Lloyd, P. J. (2017). The role of energy in development. *Journal of Energy in Southern Africa*, *28*(1), 54. https://doi.org/10.17159/2413-3051/2017/v28i1a1498

Longa, F. D., Fragkos, P., Pupo Nogueira, L., & van der Zwaan, B. (2022). System-level effects of increased energy efficiency in global low-carbon scenarios: A model comparison. *Computers & Industrial Engineering*, *167*, 108029. https://doi.org/10.1016/j.cie.2022.108029

Ma, J., Heppenstall, A., Harland, K., & Mitchell, G. (2014). Synthesising carbon emission for mega-cities: A static spatial microsimulation of transport CO2 from urban travel in Beijing. *Computers, Environment and Urban Systems*, *45*, 78–88. https://doi.org/10.1016/j.compenvurbsys.2014.02.006

Mazur, A. (1994). How does population growth contribute to rising energy consumption in America? *Population and Environment*, *15*(5), 371–378. https://doi.org/10.1007/BF02208318

Mentaschi, L., Besio, G., Cassola, F., & Mazzino, A. (2013). Problems in RMSE-based wave model validations. *Ocean Modelling*, *72*, 53–58. https://doi.org/10.1016/j.ocemod.2013.08.003

Milovanovic, D., Babic, M., Jovicic, N., & Gordic, D. (2012). Energy efficiency in buildings, industry and transportation. *AIP Conference Proceedings*, *1496*, 71–82. https://doi.org/10.1063/1.4768971

Min, J., Yan, G., Abed, A. M., Elattar, S., Amine Khadimallah, M., Jan, A., & Elhosiny Ali, H. (2022). The effect of carbon dioxide emissions on the building energy efficiency. *Fuel, 326*. https://doi.org/10.1016/j.fuel.2022.124842

Miškinis, V., Baublys, J., Konstantinavičiūtė, I., & Lekavičius, V. (2014). Aspirations for sustainability and global energy development trends. *Journal of Security and Sustainability Issues, 3*(4), 17–26. https://doi.org/10.9770/jssi.2014.3.4(2)

Mohapatra, S. K., Mishra, S., Tripathy, H. K., Bhoi, A. K., & Barsocchi, P. (2021). A pragmatic investigation of energy consumption and utilization models in the urban sector using predictive intelligence approaches. *Energies, 14*(13), 3900. https://doi.org/10.3390/en14133900

Moriasi, D. N., Arnold, J. G., Liew, M. W. Van, Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. In *Transactions of the ASABE 50* (3).

Nejat, P., Jomehzadeh, F., Taheri, M. M., Gohari, M., & Abd. Majid, M. Z. (2015). A global review of energy consumption, CO 2 emissions and policy in the residential sector (with an overview of the top ten CO 2 emitting countries). *Renewable and Sustainable Energy Reviews, 43*, 843–862. https://doi.org/10.1016/j.rser.2014.11.066

Nuţă, F. M., Nuţă, A. C., Zamfir, C. G., Petrea, S.-M., Munteanu, D., & Cristea, D. S. (2021). National Carbon Accounting—analyzing the impact of urbanization and energy-related factors upon CO2 emissions in Central–Eastern European Countries by using machine learning algorithms and panel data analysis. *Energies*, *14*(10), 2775. https://doi.org/10.3390/en14102775

Papa, R., Gargiulo, C., Zucaro, F., Angiello, G., & Carpentieri, G. (2014). Urban energy consumptions: its determinants and future research. *WIT Transactions on Ecology and the Environment*, *191*. 561–570. https://doi.org/10.2495/SC140471

Peters, G. P., Weber, C. L., Guan, D., & Hubacek, K. (2007). China's growing CO2 emissions: a race between increasing consumption and efficiency gains. *Environmental Science & Technology, 41*(17), 5939–5944. https://doi.org/10.1021/es070108f

Pfeiffer, P. E. (1990). Variance and standard deviation. *In Springer texts in statistics* (355–370). https://doi.org/10.1007/978-1-4615-7676-1_19

Polydoros, A., & Cartalis, C. (2015). Use of earth observation based indices for the monitoring of built-up area features and dynamics in support of urban energy studies. *Energy and Buildings*, *98*, 92–99. https://doi.org/10.1016/j.enbuild.2014.09.060

Romero-García, A. G., Ramírez-Corona, N., Sánchez-Ramírez, E., Alcocer-García, H., De Blasio, C., & Segovia-Hernández, J. G. (2022). Sustainability assessment in the CO2 capture process: Multi-objective optimization. *Chemical Engineering and Processing - Process Intensification*, *182*. https://doi.org/10.1016/j.cep.2022.109207

Santika, W. G., Anisuzzaman, M., Bahri, P. A., Shafiullah, G. M., Rupf, G. V., & 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

Singh, J., Knapp, H. V., Arnold, J., & Demissie, M. (2005). Hydrological modelling of the Iroquois River watershed using HSPF and SWAT. *JAWRA Journal of the American Water Resources Association*, *41*(2), 343–360. https://doi.org/10.1111/j.1752-1688.2005.tb03740.x

Singla, S., & Eldawy, A. (2020). raptor zonal statistics: fully distributed zonal statistics of big raster + vector data. *2020 IEEE International Conference on Big Data (2020)*, 571–580. https://doi.org/10.1109/BigData50022.2020.9377907

Soltangazinov, A., Smagulova, Z., Amirova, M., Kashuk, L., Karimbergenova, M., Kadyrova, A., & Zhaltyrova, O. (2020). Energy efficiency as a factor of sustainable development In Kazakhstan. *International Journal of Energy Economics and Policy*, *10*(1), 325–330. https://doi.org/10.32479/ijeep.8618

Su, Y., Yu, Y., & Zhang, N. (2020). Carbon emissions and environmental management based on Big Data and Streaming Data: A bibliometric analysis. *Science of the Total Environment*, *733*. https://doi.org/10.1016/j.scitotenv.2020.138984

Sun, W., & Huang, C. (2020). How does urbanization affect carbon emission efficiency? Evidence from China. Journal of Cleaner Production, *Journal of Cleaner Production*, *272*, 122828. https://doi.org/10.1016/j.jclepro.2020.122828

Svenfelt, Å., Engström, R., & Svane, Ö. (2011). Decreasing energy use in buildings by 50% by 2050 — A backcasting study using stakeholder groups. *Technological Forecasting and Social Change, 78*(5), 785–796. https://doi.org/10.1016/j.techfore.2010.09.005

Swardika, I. K., & Santiary, P. A. W. (2023). Modelling spatio-temporal energy consumption from nighttime radiance satellite dataset. *MATRIX : Jurnal Manajemen Teknologi Dan Informatika, 13*(3), 166–175. https://doi.org/10.31940/matrix.v13i3.166-175

Tajudeen, I. A., Wossink, A., & Banerjee, P. (2018). How significant is energy efficiency to mitigate CO2 emissions? Evidence from OECD countries. *Energy Economics, 72*, 200–221. https://doi.org/10.1016/j.eneco.2018.04.010

Townsend, A. C., & Bruce, D. A. (2010). The use of night-time lights satellite imagery as a measure of Australia's regional electricity consumption and population distribution. *International Journal of Remote Sensing, 31*(16), 4459–4480. https://doi.org/10.1080/01431160903261005

Tu, Z., Feng, C., & Zhao, X. (2022). Revisiting energy efficiency and energy related CO 2 emissions: Evidence from RCEP economies. *Economic Research-Ekonomska Istraživanja, 35*(16), 5858–5878. https://doi.org/10.1080/1331677X.2022.2038651

United Nations. (2016). Progress towards the Sustainable Development Goals. *Report of the Secretary-General*. In *Economic and Social Council* (Issue June).

Ürge-Vorsatz, D., Graham, P., & Harvey, D. (2012). Energy end-use: buildings. In *Global Energy Assessment - Toward a Sustainable Future* (pp. 649–760). Cambridge University Press and IIASA. https://www.researchgate.net/publication/274632710

Vieira, M. A., & Ceretta, P. S. (2021). CO2 emissions: a dynamic structural analysis. *Revista de Administração Da UFSM, 14*, 949–966. https://doi.org/10.5902/1983465963884

Wang, S., & Liu, X. (2017). China's city-level energy-related CO2 emissions: Spatiotemporal patterns and driving forces. *Applied Energy, 200*, 204–214. https://doi.org/10.1016/j.apenergy.2017.05.085

Wang, W., Melnyk, L., Kubatko, O., Kovalov, B., & Hens, L. (2023). Economic and technological efficiency of renewable energy technologies implementation. *Sustainability, 15*(11), 8802. https://doi.org/10.3390/su15118802

Wu, Z., & Chu, W. (2021). Sampling strategy analysis of machine learning models for energy consumption prediction. *2021 IEEE 9th International Conference on Smart Energy Grid Engineering (SEGE)*, 77–81. https://doi.org/10.1109/SEGE52446.2021.9534987

Zakari, A., Khan, I., Tan, D., Alvarado, R., & Dagar, V. (2022). Energy efficiency and sustainable development goals (SDGs). *Energy, 239*, 122365. https://doi.org/10.1016/j.energy.2021.122365

Zarco-Periñán, P. J., Zarco-Soto, I. M., & Zarco-Soto, Fco. J. (2021). Influence of the population density of cities on energy consumption of their households. *Sustainability, 13*(14), 7542. https://doi.org/10.3390/su13147542

Zhang, C., & Lin, Y. (2012). Panel estimation for urbanization, energy consumption and CO2 emissions: A regional analysis in China. *Energy Policy, 49*, 488–498. https://doi.org/10.1016/j.enpol.2012.06.048

Zhao, P., & Zhang, M. (2018). The impact of urbanisation on energy consumption: a 30-year review in China. *Urban Climate, 24*, 940–953. https://doi.org/10.1016/j.uclim.2017.11.005

Appendix

Appendix A (Data)

Province	Average of CO2 Emission					Average of Built Up Area				
	CO2_2 019	CO2_2 020	CO2_2 021	CO2_2 022	CO2_2 023	BuiltUp_2 019	BuiltUp_2 020	BuiltUp_2 021	BuiltUp_2 022	BuiltUp_2 023
Nusa Tenggara Timur	0,00041	0,00041	0,00041	0,00041	0,00042	0,01662	0,01707	0,01767	0,01830	0,01879
Kalimantan Timur	0,00041	0,00041	0,00041	0,00041	0,00042	0,00445	0,00490	0,00534	0,00576	0,00614
Jambi	0,00041	0,00041	0,00041	0,00042	0,00042	0,01051	0,01126	0,01216	0,01283	0,01367
Kalimantan Barat	0,00041	0,00041	0,00041	0,00042	0,00042	0,00474	0,00495	0,00522	0,00545	0,00576
Bengkulu	0,00041	0,00041	0,00041	0,00041	0,00042	0,01387	0,01456	0,01516	0,01620	0,01695
Sumatera Utara	0,00041	0,00041	0,00041	0,00042	0,00042	0,03226	0,03449	0,03632	0,03858	0,04050
Jawa Barat	0,00041	0,00041	0,00041	0,00041	0,00042	0,16095	0,16571	0,17122	0,17673	0,18232
Kalimantan Tengah	0,00041	0,00041	0,00041	0,00041	0,00042	0,00230	0,00247	0,00265	0,00287	0,00310
Sulawesi Utara	0,00041	0,00041	0,00041	0,00041	0,00042	0,02461	0,02613	0,02799	0,02972	0,03151
Papua Tengah	0,00041	0,00041	0,00041	0,00041	0,00042	0,00202	0,00205	0,00206	0,00216	0,00226
Jawa Timur	0,00041	0,00041	0,00041	0,00041	0,00042	0,14705	0,15484	0,16228	0,17034	0,17876
Kalimantan Selatan	0,00041	0,00041	0,00041	0,00041	0,00042	0,02736	0,02982	0,03281	0,03607	0,03955
Gorontalo	0,00041	0,00041	0,00041	0,00041	0,00042	0,01224	0,01349	0,01549	0,01749	0,01932
Papua Pegunungan	0,00041	0,00041	0,00041	0,00041	0,00042	0,00074	0,00077	0,00077	0,00077	0,00079
Kepulauan Bangka Belitung	0,00041	0,00041	0,00041	0,00042	0,00042	0,02021	0,02129	0,02351	0,02543	0,02849
Daerah Istimewa Yogyakarta	0,00041	0,00041	0,00041	0,00041	0,00042	0,20845	0,22201	0,23494	0,24629	0,26269
Sumatera Selatan	0,00041	0,00041	0,00041	0,00042	0,00042	0,01187	0,01263	0,01353	0,01455	0,01554
Maluku Utara	0,00041	0,00041	0,00041	0,00041	0,00042	0,00645	0,00709	0,00744	0,00827	0,00865
Banten	0,00041	0,00041	0,00041	0,00041	0,00042	0,18008	0,18307	0,18585	0,18864	0,19163
Lampung	0,00041	0,00041	0,00041	0,00041	0,00042	0,04864	0,05186	0,05558	0,05928	0,06354
Sulawesi Barat	0,00041	0,00041	0,00041	0,00041	0,00042	0,00718	0,00820	0,00904	0,00959	0,01055
DKI Jakarta	0,00041	0,00041	0,00041	0,00041	0,00042	0,94825	0,95129	0,95282	0,95282	0,95282
Jawa Tengah	0,00041	0,00041	0,00041	0,00041	0,00042	0,18193	0,19196	0,20189	0,21140	0,22078
Kalimantan Utara	0,00041	0,00041	0,00041	0,00041	0,00042	0,00161	0,00170	0,00181	0,00190	0,00204
Nusa Tenggara Barat	0,00041	0,00041	0,00041	0,00041	0,00042	0,05474	0,05814	0,06109	0,06450	0,06791
Papua	0,00041	0,00041	0,00041	0,00041	0,00042	0,00125	0,00133	0,00142	0,00153	0,00163
Sumatera Barat	0,00041	0,00041	0,00041	0,00042	0,00042	0,04193	0,04376	0,04550	0,04726	0,04868
Riau	0,00041	0,00041	0,00041	0,00042	0,00042	0,01230	0,01340	0,01457	0,01600	0,01748
Sulawesi Tenggara	0,00041	0,00041	0,00041	0,00041	0,00042	0,01054	0,01248	0,01483	0,01743	0,02078
Kepulauan Riau	0,00041	0,00041	0,00041	0,00042	0,00042	0,04176	0,04335	0,04469	0,04590	0,04809
Sulawesi Tengah	0,00041	0,00041	0,00041	0,00041	0,00042	0,00634	0,00688	0,00735	0,00780	0,00853
Aceh	0,00041	0,00041	0,00041	0,00042	0,00042	0,01836	0,01905	0,02009	0,02118	0,02210
Papua Selatan	0,00041	0,00041	0,00041	0,00041	0,00042	0,00041	0,00041	0,00041	0,00041	0,00042
Papua Barat Daya	0,00041	0,00041	0,00041	0,00041	0,00042	0,00215	0,00231	0,00244	0,00262	0,00285
Sulawesi Selatan	0,00041	0,00041	0,00041	0,00041	0,00042	0,02533	0,02657	0,02831	0,02966	0,03103
Maluku	0,00041	0,00041	0,00041	0,00041	0,00042	0,00520	0,00546	0,00585	0,00630	0,00661
Bali	0,00041	0,00041	0,00041	0,00041	0,00042	0,11763	0,12265	0,12838	0,13430	0,13843
Papua Barat	0,00041	0,00041	0,00041	0,00041	0,00042	0,00119	0,00125	0,00130	0,00130	0,00136

Province	Average of Light Intensity					Average of Population				
	NTL_2019	NTL_2020	NTL_2021	NTL_2022	NTL_2023	Pop_2019	Pop_2020	Pop_2021	Pop_2022	Pop_2023
Nusa Tenggara Timur	0,05	0,06	0,07	0,08	0,13	93,41	95,36	97,42	99,52	101,57
Kalimantan Timur	0,14	0,13	0,14	0,18	0,26	24,69	25,59	26,55	27,53	28,51
Jambi	0,22	0,19	0,19	0,25	0,41	63,68	65,36	67,09	68,88	70,66
Kalimantan Barat	0,05	0,04	0,05	0,07	0,10	28,30	28,82	29,36	29,93	30,51
Bengkulu	0,09	0,08	0,09	0,14	0,24	75,10	75,85	76,52	77,28	78,02
Sumatera Utara	0,42	0,42	0,43	0,51	0,65	164,58	166,63	168,87	171,01	173,18
Jawa Barat	2,36	2,23	2,48	2,59	3,03	1112,72	1135,72	1159,22	1184,08	1209,18
Kalimantan Tengah	0,03	0,03	0,03	0,04	0,08	14,09	14,41	14,74	15,08	15,42
Sulawesi Utara	0,25	0,25	0,25	0,27	0,37	148,41	151,96	155,45	159,28	163,20
Papua Tengah	0,02	0,02	0,02	0,02	0,04	23,67	26,42	29,62	33,14	37,16
Jawa Timur	1,93	1,91	2,11	2,32	2,73	686,25	691,58	697,09	702,65	708,19
Kalimantan Selatan	0,25	0,23	0,24	0,29	0,42	94,22	96,37	98,50	100,84	103,25
Gorontalo	0,16	0,16	0,17	0,20	0,26	82,42	84,33	86,26	88,24	90,22
Papua Pegunungan	0,00	0,00	0,00	0,00	0,01	22,70	24,98	27,34	30,01	32,96
Kepulauan Bangka Belitung	0,14	0,13	0,15	0,18	0,24	73,96	76,25	78,73	81,19	83,75
Daerah Istimewa Yogyakarta	2,49	2,32	2,54	2,74	3,49	978,96	990,16	1001,28	1012,53	1023,56
Sumatera Selatan	0,29	0,24	0,28	0,38	0,47	77,70	78,37	79,07	79,74	80,42
Maluku Utara	0,03	0,05	0,05	0,08	0,14	33,37	34,92	36,59	38,41	40,18
Banten	3,21	3,07	3,23	3,51	4,02	1145,29	1177,30	1211,84	1246,51	1283,87
Lampung	0,63	0,53	0,64	0,70	0,88	209,62	212,73	215,97	219,22	222,46
Sulawesi Barat	0,03	0,03	0,03	0,05	0,08	72,36	75,14	78,10	81,19	84,35
DKI Jakarta	37,17	33,71	33,60	35,60	38,59	13148,83	13337,30	13522,91	13714,68	13912,00
Jawa Tengah	1,79	1,69	1,91	2,16	2,61	817,67	821,54	824,81	828,92	832,57
Kalimantan Utara	0,03	0,02	0,02	0,04	0,10	8,42	8,85	9,32	9,81	10,32
Nusa Tenggara Barat	0,33	0,31	0,36	0,40	0,49	207,50	210,98	214,62	218,09	221,46
Papua	0,02	0,02	0,02	0,03	0,04	10,07	10,59	11,17	11,78	12,40
Sumatera Barat	0,20	0,18	0,20	0,24	0,33	106,08	107,52	109,04	110,51	112,04
Riau	0,29	0,25	0,25	0,29	0,42	66,79	69,56	72,50	75,49	78,61
Sulawesi Tenggara	0,08	0,08	0,10	0,13	0,18	61,44	63,37	65,44	67,63	69,79
Kepulauan Riau	0,87	0,87	0,93	1,13	1,28	219,88	232,14	245,71	260,09	274,51
Sulawesi Tengah	0,08	0,07	0,08	0,12	0,17	41,08	42,06	43,10	44,20	45,25
Aceh	0,21	0,22	0,21	0,25	0,34	69,43	69,99	70,50	71,13	71,77
Papua Selatan	0,00	0,00	0,00	0,00	0,01	4,34	4,65	5,02	5,40	5,78
Papua Barat Daya	0,03	0,03	0,03	0,03	0,04	13,69	14,77	15,86	17,11	18,39
Sulawesi Selatan	0,28	0,24	0,28	0,33	0,45	165,27	167,71	170,28	172,88	175,53
Maluku	0,02	0,02	0,02	0,03	0,05	31,61	32,61	33,82	34,98	36,13
Bali	1,45	1,25	1,36	1,60	1,97	645,09	660,31	676,61	692,51	709,38
Papua Barat	0,02	0,02	0,03	0,03	0,05	6,57	6,92	7,36	7,78	8,24

Appendix B (Python Script)

```
# -*- coding: utf-8 -*-
"""Energy Efficiency Revision.ipynb

import pandas as pd

# Load the Excel file
file_path = '/content/Zonal Provinsi.xlsx'
data = pd.read_excel(file_path)
```

```
# Display the first few rows of the dataframe and the column names
data.head(), data.columns

import seaborn as sns
import matplotlib.pyplot as plt

# Compute correlation for each year
correlation_2019 = data[['CO2_2019', 'BuiltUp_2019', 'NTL_2019', 'Pop_2019']].corr()
correlation_2020 = data[['CO2_2020', 'BuiltUp_2020', 'NTL_2020', 'Pop_2020']].corr()
correlation_2021 = data[['CO2_2021', 'BuiltUp_2021', 'NTL_2021', 'Pop_2021']].corr()
correlation_2022 = data[['CO2_2022', 'BuiltUp_2022', 'NTL_2022', 'Pop_2022']].corr()
correlation_2023 = data[['CO2_2023', 'BuiltUp_2023', 'NTL_2023', 'Pop_2023']].corr()

# Visualize correlation matrices for each year using heatmap
fig, ax = plt.subplots(1, 5, figsize=(25, 5), sharey=True)

sns.heatmap(correlation_2019, annot=True, fmt=".2f", cmap='coolwarm', ax=ax[0])
ax[0].set_title('Correlation 2019')
ax[0].set_yticklabels(ax[0].get_yticklabels(), rotation=0)

sns.heatmap(correlation_2020, annot=True, fmt=".2f", cmap='coolwarm', ax=ax[1])
ax[1].set_title('Correlation 2020')
ax[1].set_yticklabels(ax[1].get_yticklabels(), rotation=0)

sns.heatmap(correlation_2021, annot=True, fmt=".2f", cmap='coolwarm', ax=ax[2])
ax[2].set_title('Correlation 2021')
ax[2].set_yticklabels(ax[2].get_yticklabels(), rotation=0)

sns.heatmap(correlation_2022, annot=True, fmt=".2f", cmap='coolwarm', ax=ax[3])
ax[3].set_title('Correlation 2022')
ax[3].set_yticklabels(ax[3].get_yticklabels(), rotation=0)

sns.heatmap(correlation_2023, annot=True, fmt=".2f", cmap='coolwarm', ax=ax[4])
ax[4].set_title('Correlation 2023')
ax[4].set_yticklabels(ax[4].get_yticklabels(), rotation=0)

plt.tight_layout()
plt.show()

# Save the heatmap as a PNG file
heatmap_path = "/content/CO2_correlations_2019_to_2023.png"
fig.savefig(heatmap_path)

heatmap_path

# Prepare data for plotting the trends over the years
years = ['2019', '2020', '2021', '2022', '2023']
co2_columns = [f'CO2_{year}' for year in years]
builtup_columns = [f'BuiltUp_{year}' for year in years]
ntl_columns = [f'NTL_{year}' for year in years]
pop_columns = [f'Pop_{year}' for year in years]

# Calculate mean values for each variable over the years
mean_co2 = data[co2_columns].mean()
mean_builtup = data[builtup_columns].mean()
mean_ntl = data[ntl_columns].mean()
mean_pop = data[pop_columns].mean()

# Re-running the plotting code with a command to save the figure as a PNG file
```

```
fig, axs = plt.subplots(4, 1, figsize=(10, 15))

# CO2 Emissions
axs[0].plot(years, mean_co2, marker='o', linestyle='-')
axs[0].set_title('Average CO2 Emissions (2019-2023)')
axs[0].set_ylabel('CO2 Emissions')
axs[0].grid(True)

# Built-Up Areas
axs[1].plot(years, mean_builtup, marker='o', color='orange', linestyle='-')
axs[1].set_title('Average Built-Up Areas (2019-2023)')
axs[1].set_ylabel('Built-Up Area')
axs[1].grid(True)

# Night-time Lights
axs[2].plot(years, mean_ntl, marker='o', color='green', linestyle='-')
axs[2].set_title('Average Night-time Light (NTL) Intensity (2019-2023)')
axs[2].set_ylabel('NTL Intensity')
axs[2].grid(True)

# Population
axs[3].plot(years, mean_pop, marker='o', color='red', linestyle='-')
axs[3].set_title('Average Population (2019-2023)')
axs[3].set_ylabel('Population')
axs[3].grid(True)

# Layout adjustment
plt.tight_layout()

# Save the figure
trends_png_path = "/content/Average_Trends_2019_to_2023.png"
fig.savefig(trends_png_path)

# Show the plot
plt.show()

trends_png_path

from sklearn.model_selection import train_test_split

# Prepare features and target
# Features from 2019 to 2022 (excluding CO2_2023)
features_columns = ['BuiltUp_2019', 'NTL_2019', 'Pop_2019',
                    'BuiltUp_2020', 'NTL_2020', 'Pop_2020',
                    'BuiltUp_2021', 'NTL_2021', 'Pop_2021',
                    'BuiltUp_2022', 'NTL_2022', 'Pop_2022']
target_column = 'CO2_2023'

X = data[features_columns]
y = data[target_column]

# Splitting the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Checking the shape of the training and testing sets
X_train.shape, X_test.shape, y_train.shape, y_test.shape

from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import cross_val_score
import numpy as np
```

```
# Train the Random Forest model
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# Perform cross-validation
rf_scores = cross_val_score(rf_model, X_train, y_train, cv=5, scoring='neg_mean_squared_error')
rf_rmse_scores = np.sqrt(-rf_scores)

# Calculate mean and standard deviation of the RMSE scores
rf_rmse_mean = np.mean(rf_rmse_scores)
rf_rmse_std = np.std(rf_rmse_scores)

rf_rmse_mean, rf_rmse_std

from sklearn.svm import SVR

# Train the Support Vector Machine model
svm_model = SVR(kernel='rbf')
svm_model.fit(X_train, y_train)

# Perform cross-validation
svm_scores = cross_val_score(svm_model, X_train, y_train, cv=5, scoring='neg_mean_squared_error')
svm_rmse_scores = np.sqrt(-svm_scores)

# Calculate mean and standard deviation of the RMSE scores
svm_rmse_mean = np.mean(svm_rmse_scores)
svm_rmse_std = np.std(svm_rmse_scores)

svm_rmse_mean, svm_rmse_std

from sklearn.neural_network import MLPRegressor

# Train the Artificial Neural Network model
ann_model = MLPRegressor(hidden_layer_sizes=(100,), activation='relu', solver='adam',
                         max_iter=500, random_state=42)
ann_model.fit(X_train, y_train)

# Perform cross-validation
ann_scores = cross_val_score(ann_model, X_train, y_train, cv=5, scoring='neg_mean_squared_error')
ann_rmse_scores = np.sqrt(-ann_scores)

# Calculate mean and standard deviation of the RMSE scores
ann_rmse_mean = np.mean(ann_rmse_scores)
ann_rmse_std = np.std(ann_rmse_scores)

ann_rmse_mean, ann_rmse_std

# Calculate the annual growth rate for each feature
growth_rates = {}

# Calculating growth rate as (value_end / value_start) ** (1 / years) - 1
for feature in ['BuiltUp', 'NTL', 'Pop']:
    start_value = data[f'{feature}_2019']
    end_value = data[f'{feature}_2023']
    years = 2023 - 2019
    growth_rate = (end_value / start_value) ** (1 / years) - 1
    growth_rates[feature] = growth_rate

# Calculate future values for each year
```

```
future_years = [2025, 2030, 2035, 2040, 2045]
future_data = pd.DataFrame({'Provinsi': data['Provinsi']})

for year in future_years:
    years_ahead = year - 2023
    for feature in ['BuiltUp', 'NTL', 'Pop']:
        future_data[f'{feature}_{year}'] = data[f'{feature}_2023'] * ((1 + growth_rates[feature]) ** years_ahead)

future_data.head()

print(rf_model.feature_names_in_)  # Hanya berfungsi pada scikit-learn versi 1.0 ke atas

# Adjust the preparation of feature projections to align correctly with the prediction years
for year in future_years:
    years_prior = list(range(year - 4, year))  # Last 4 years up to the target year
    for yr in years_prior:
        for feature in ['BuiltUp', 'NTL', 'Pop']:
            if f'{feature}_{yr}' not in future_data.columns:  # Only calculate if not already done
                future_data[f'{feature}_{yr}'] = data[f'{feature}_2023'] * ((1 + growth_rates[feature]) ** (yr - 2023))

# Now preparing the feature sets correctly for each future year and predict
corrected_predictions = {'Provinsi': data['Provinsi']}
for year in future_years:
    years_prior = list(range(year - 4, year))  # Last 4 years up to the target year
    feature_columns = []
    for yr in years_prior:
        for feature in ['BuiltUp', 'NTL', 'Pop']:
            feature_columns.append(f'{feature}_{yr}')
    # Predict CO2 emissions for the year using the corresponding feature set
    corrected_predictions[f'CO2_{year}'] = rf_model.predict(future_data[feature_columns])

# Convert corrected predictions to a DataFrame for easier visualization
corrected_predictions_df = pd.DataFrame(corrected_predictions)
corrected_predictions_df.head()

# Save the predictions to an Excel file
predictions_file_path = '/content/CO2_Emissions_Predictions_2025_to_2045.xlsx'
corrected_predictions_df.to_excel(predictions_file_path, index=False)

# Prepare for visualization
import matplotlib.pyplot as plt

# Plotting the predictions for each province for future years
plt.figure(figsize=(12, 8))
for index, row in corrected_predictions_df.iterrows():
    plt.plot(future_years, row[1:], marker='o', label=row['Provinsi'])

plt.title('Predicted CO2 Emissions per Province from 2025 to 2045')
plt.xlabel('Year')
plt.ylabel('Predicted CO2 Emissions')
plt.legend(loc='upper right', fontsize='small')
plt.grid(True)
plt.show()

predictions_file_path

# Calculate total predicted CO2 emissions for each future year
total_predictions = corrected_predictions_df.drop('Provinsi', axis=1).sum()

# Plotting the total predicted CO2 emissions for future years
```

```
plt.figure(figsize=(10, 6))
plt.plot(total_predictions.index, total_predictions.values, marker='o', linestyle='-')
plt.title('Total Predicted CO2 Emissions from 2025 to 2045')
plt.xlabel('Year')
plt.ylabel('Total Predicted CO2 Emissions')
plt.grid(True)
plt.show()

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Assuming 'corrected_predictions_df' is your DataFrame
# If you need to load it from a file, use: corrected_predictions_df = pd.read_excel('path_to_your_file.xlsx')

# Define future_years explicitly for clarity
future_years = [2025, 2030, 2035, 2040, 2045]

# Categorize trends of CO2 emissions
categories = {'Increasing': [], 'Fluctuating/Stagnant': [], 'Decreasing': []}
for index, row in corrected_predictions_df.iterrows():
    emissions = row[1:].values  # Skip 'Provinsi' column which is the first column
    if np.all(np.diff(emissions) > 0):
        categories['Increasing'].append(row['Provinsi'])
    elif np.all(np.diff(emissions) < 0):
        categories['Decreasing'].append(row['Provinsi'])
    else:
        categories['Fluctuating/Stagnant'].append(row['Provinsi'])

# Plotting
plt.figure(figsize=(15, 10))
colors = {'Increasing': 'green', 'Fluctuating/Stagnant': 'blue', 'Decreasing': 'red'}
for category, provinsi_list in categories.items():
    for provinsi in provinsi_list:
        data_to_plot = corrected_predictions_df[corrected_predictions_df['Provinsi'] == provinsi]
        plt.plot(future_years, data_to_plot.iloc[0, 1:], marker='o', color=colors[category], label=f"{provinsi}
({category})")

plt.title('Predicted CO2 Emissions Per Province Categorized by Trends')
plt.xlabel('Year')
plt.ylabel('Predicted CO2 Emissions')
plt.xticks(future_years, future_years)  # Set x-axis ticks to be explicit years
plt.grid(True)
plt.legend(loc='center left', bbox_to_anchor=(1, 0.5))  # Moving the legend to the side
plt.tight_layout(rect=[0, 0, 0.75, 1])  # Adjust layout

# Save the plot
plt.savefig('CO2_Emissions_Trends_by_Province.png')
plt.show()

# Save the DataFrame to Excel
corrected_predictions_df.to_excel('Corrected_CO2_Emissions_Predictions.xlsx', index=False)

import pandas as pd
import matplotlib.pyplot as plt

# Assuming 'corrected_predictions_df' has been loaded or defined
# Let's assume corrected_predictions_df has columns like 'Provinsi', 'CO2_2025', 'CO2_2030', etc.

# Calculate the average increase in emissions for each province
```

```
average_increase = corrected_predictions_df.set_index('Provinsi').diff(axis=1).mean(axis=1)

# Create a sorted DataFrame for visualization
sorted_provinces = average_increase.sort_values(ascending=False).reset_index()
sorted_provinces.columns = ['Provinsi', 'Average Increase']

# Plotting
plt.figure(figsize=(12, 8))
plt.bar(sorted_provinces['Provinsi'], sorted_provinces['Average Increase'], color='skyblue')
plt.xlabel('Province')
plt.ylabel('Average Increase in CO2 Emissions')
plt.title('Average Increase in CO2 Emissions Per Province')
plt.xticks(rotation=90)  # Rotate the province names to make them readable
plt.tight_layout()

# Save the plot
plt.savefig('Average_Increase_CO2_Emissions_Per_Province.png')
plt.show()
```