Framework and Clustering Dashboard for Analysing Temporal-Based Parameters of Solar PV Output Model

Yusak Tanoto
Electrical Engineering Department
Petra Christian University
Surabaya, Indonesia
tanyusak@petra.ac.id

Gregorius Satia Budhi
Informatics Department
Petra Christian University
Surabaya, Indonesia
greg@petra.ac.id

2023 11th International Conference on Smart Grid and Clean Energy Technologies (ICSGCE) | 979-8-3503-0536-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSGCE59477.2023.10419975

***Abstract*—This paper presents a framework and dashboard designed to analyze and present long-term hourly temporal-based parameters of a solar photovoltaic (PV) output model. The framework uses time series and artificial intelligence-based methods to analyze direct and diffuse irradiance, as well as temperature data, while the web-based dashboard aims to present the characterization and prediction results of these parameters. The framework's first part clusters the three parameters and solar PV output over a ten-year period and beyond, focusing on the Java-Bali region in Indonesia. The paper emphasizes the importance of a dashboard that visualizes areas across regions resulting from the k-means clustering analysis. This framework provides vital insights into the potential of solar photovoltaics, specifically regarding spatial issues that may influence investment decision-making.**

Keywords—framework, hourly temporal, dashboard, solar photovoltaic, machine learning

I. Introduction

Solar photovoltaic (solar PV) has been widely used around the world by many utilities for electricity generation in various capacities and system configurations. Similarly, there is a growing interest in the energy customer side to implement solar PV systems. Capacity deployment in many developing countries and emerging economies, including Indonesia, however, appears to be noticeably slower than targeted and comparable to those that occurred in many developed countries, despite declining systems costs over the decade [1].

Apart from non-technical factors such as renewable energy regulations and coal dependency in the electricity industries, one of the challenges in making investment decisions for PV capacity deployment remains the intermittency of PV power output, which causes fluctuations in voltage and frequency at the interconnected grid [2]. Stakeholders, nevertheless, can take advantage of long-term historical data on temporal and spatial parameters related to PV power output by developing specific models that can be useful, for example, in determining a suitable location for a solar PV plant. The model would particularly be useful in helping planners characterise, classify, or predict solar PV output or related parameters in the future.

The hourly temporal-based solar PV output with wide area coverage can be used as a basis for investment decision-making purposes for solar PV plants and infrastructure, as part of the electricity industry planning. The temporal and spatial data of solar PV output and related parameters are useful for conducting studies of the future energy generation mix considering the uncertainty of energy resources [3]. In addition, this temporal-based solar resources big data is particularly critical, not only for studying and predicting short-term operational capacity mix and reserves but also in cost-reliability trade-off analysis of future generation mix in developing country context [4].

This paper introduces a framework for analysing and presenting long-term hourly temporal-based parameters of a solar PV output model, along with a dashboard for visualising the clustered parameters. The methods applied in the overall framework consist of clustering analysis methods, time series and artificial intelligence-based methods, i.e., machine learning and deep learning techniques.

Several machine learning regression methods are specifically employed in the framework. All these methods are used to analyse direct and diffuse irradiance and temperature data, as well as solar PV output data, and eventually develop an electricity supply-demand model that incorporates a novel machine learning method. The methods are applied over a period of ten years and beyond, taking the Java-Bali region, Indonesia, as the case study. This paper particularly focuses on several clustering, as developed in the first stage of the framework, in the Java-Bali region, Indonesia, using the k-means clustering method. The optimal number of clusters is evaluated using the Davies-Bouldin index [5] and Silhouette Score [6].

Many studies of course have used clustering analysis in problems related to solar PV resources and output. In this study, the framework puts together unsupervised machine learning-based methods to classify solar PV resources and output based on three parameters related to output, i.e., direct irradiation, diffuse irradiation, and ambient temperature, over a choice of wide (spatial) area of coverage and various long-term period.

This paper is structured as follows: Solar PV resources and deployment in Indonesia with a focus on Java-Bali are briefly described in Section 2. A framework for analysing and presenting hourly temporal-based parameter datasets of a solar PV output model is presented in Section 3. Results of the analysis using the 2005 dataset, as shown in the dashboard, are presented in Section 4, followed by conclusions and future work in Section 5.

II. SOLAR PV RESOURCE AND DEPLOYMENT IN INDONESIA WITH FOCUS IN JAVA-BALI

Indonesia has an excellent solar resource throughout the country, averaging 4.8 kWh/m^2/day [7]. The islands of Java and Bali together are by far the economic and population powerhouses in Indonesia. Of the 278 million population in 2022, Java and Bali are inhabited by around 160 million people. Every year, this region has contributed around 70% of national electricity energy consumption, which shows the great opportunity to utilise rooftop PV, especially for the purpose of saving electricity bills.

In addition to the great potential on other islands, the Java-Bali region only has a solar PV potential of around 39 GW [8]. The Indonesian government has set an ambitious target of 4.6 GW solar PV capacity by the end of 2030 in the "Green" National Electricity Supply Business Plan for 2021-2030 [9]. Nevertheless, the total national capacity of utility-scale solar PV plants has achieved only 201 MW in 2021, or equivalent to 0.27% of the total capacity mix.

Utility-scale PV deployment has been particularly slow in Java-Bali, with only 73 MW until the end of 2021 [9]. While solar rooftop PV is one of the solutions for utilising renewable energy in cities, including in Indonesia, the actual national installed capacity has only reached 62 MW as of July 2022 even though the potential for rooftop PV is quite significant at 32.5 GW [10].

Long-term Java-Bali solar resources can be further explored and analysed by mapping, for example, the solar PV output and capacity factor. This study constructs maps of inter-annual average daily output (kWh/kWp/day) and monthly inter-annual (hourly average) capacity factor (%) during 2005-2016, as presented in Fig.1 and Fig. 2, respectively.

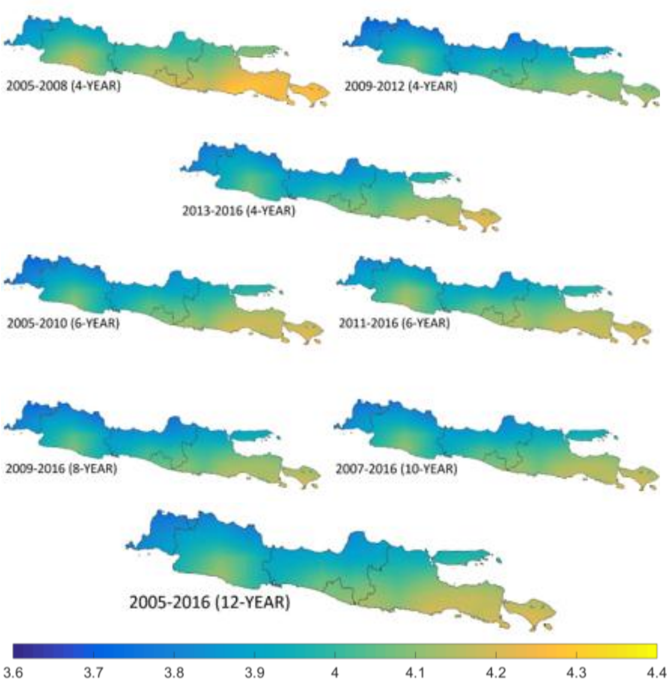

Fig. 1. Inter-annual daily average output (kWh/kWp/day) of the Java-Bali region during 2005-2016

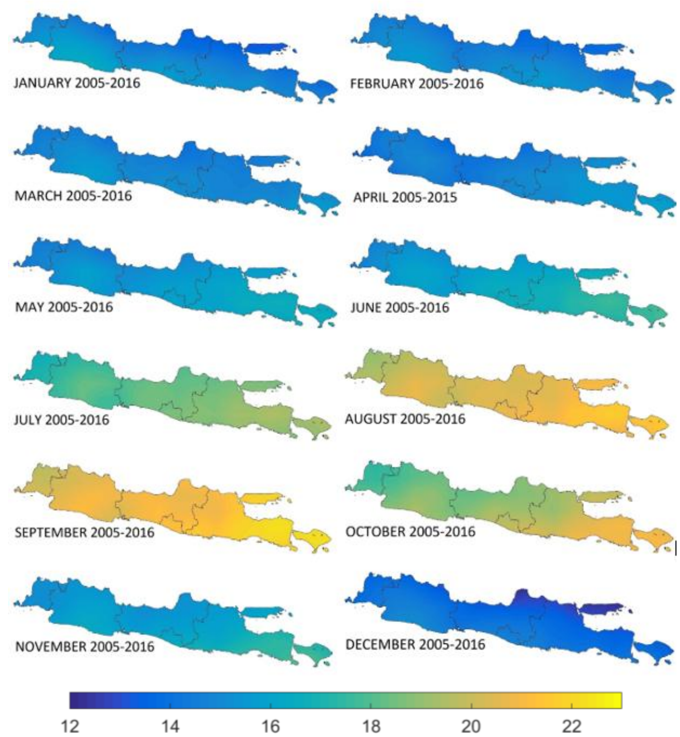

Fig. 2. Monthly inter-annual (average hourly) capacity factor (%) of the Java-Bali region during 2005-2016

The mapping of the solar PV output model for the Java-Bali region presented in Fig. 1 and Fig. 2 was built based on an online renewable energy tool of the solar PV output model called Global Solar Energy Estimator established at Renewables Ninja [11]. For each specified latitude and longitude, solar PV output is modelled in Renewable Ninja for hourly time steps at a specific tilt angle based on NASA-MERRA2 direct and diffuse irradiation, and ambient temperature data. In this study, all the data for several years (2005-2016) were obtained in the form of gridded hourly with a spatial resolution of 0.05° x 0.05° or 5 km x 5 km using an API provided by Renewables Ninja.

As presented in Fig. 1, the southern and eastern regions of East Java province, together with Bali Island, have exhibited relatively higher average daily yields compared to the western part of Java. However, over all time periods of the year, the daily average output has reached more than 4 kWh/kWp/day across the region.

Meanwhile, the Java-Bali region showed different monthly capacity factors during 2005-2016, ranging from 12%-23%. August and September were the highest capacity factor among other months during this period, followed by October and July. The capacity factor reached 19%-23% during August-September, when the Java-Bali region experienced dry season.

In contrast, the lowest capacity factor was achieved in December, ranging from 12%-14%. The monthly capacity factor values, as presented in Fig. 2, reflect the climatic conditions in the Java-Bali region, where the rainy season usually occurs extensively in December-April. The mapping results indicate that the placement of solar PV plants, especially utility-scale ones in the East Java and Bali regions, will produce better output compared to the western regions.

III. THE FRAMEWORK

The overall framework for analysing and presenting long-term hourly temporal-based parameters of a solar PV output model considers multi-methods. As illustrated in Fig. 3 below, it consists of three stages. The first stage of the framework is applied to this paper. It uses a multi-year dataset consisting of direct and diffuse irradiation and ground temperature to analyse the clustering of these parameters.

Two unsupervised machine learning clustering methods, namely k-means clustering [5] and DBSCAN [12], were employed in this study to obtain the results of these parameters' characterisation over different periods and to identify different areas that refer to different clusters.

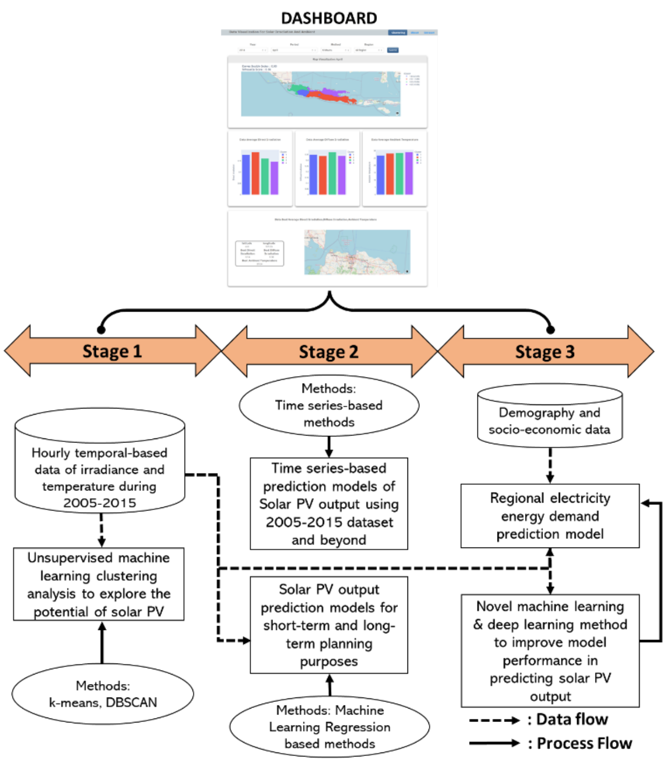

Fig. 3. The overall framework for analysing and presenting hourly temporal-based data of solar PV output model

Stage 2 of the framework mainly focuses on modelling and predicting the solar PV output along with the three parameters associated with it. The framework applies two different approaches, i.e., time series-based methods, and machine learning-based methods. The main purpose of stage 2 is to explore various possible methods leading to either short-term or long-term generation expansion planning.

The time-series-based methods use ARIMA, SARIMA, and ARIMAX, among other econometric-based approaches to model and predict solar PV output using the hourly temporal-based long-term solar irradiations and temperature dataset for the Java-Bali region. Meanwhile, the second approach employs several supervised machine learning-based methods, such as Multiple Linear Regression (MLR), Multi-Layer Perceptron, Gradient Boosting (GB), and Random Forest (RM) algorithms.

The final stage of the framework aims to utilise novel machine learning and deep learning-based techniques to improve the performance of the models that were established in the previous stage. Specifically, this stage will focus on analysing the supply and demand of electricity in the Java-Bali region, as a case study.

The goal of the analysis is to identify how much solar power can be used to meet the energy needs of the areas under investigation. Furthermore, the results of this analysis could help in generating more accurate outcomes, like a portfolio of electricity generation sources that accounts for future uncertainties. To achieve this, a stochastic optimization-based system or a similar method could be used.

The hardware specifications used in this research are CPU-i5-1135G7; GPU-UHD Graphics 630; RAM Capacity-32 GB; and Storage Capacity of 512 GB. Meanwhile, software used in this research includes I/O: Windows 10; Python programming language; and Python library that consists of dash, pandas, sklearn, numpy, plotly, math, matplotlib, time, openpyxl, and calendar.

The research methodology used in this study involves four main steps. First, the dataset retrieved from Renewables Ninja undergoes pre-processing. Next, the processed dataset is clustered and saved based on its data frame. Finally, the clustered dataset is presented using maps for better visualisation.

IV. THE CLUSTERING DASHBOARD

The dashboard particularly informs the user of the results of clustering analysis of direct irradiation, diffuse irradiation, and temperature. The clustering can be carried out simultaneously for all parameters or individually, using either k-means or DBSCAN. In addition, the dashboard can be used to display the clustering results of solar PV output.

A yearly-based clustering looks at results on an annual basis, where the dataset involved at each location ranges from January to December on an hourly temporal basis. Moreover, the dashboard also displays groupings based on monthly periods throughout the Java-Bali region as well as in each province. This paper, however, focuses on displaying the clustering analysis using the k-means technique via the dashboard. The dashboard features clustering visualisation, useful numerical information with respect to geographic locations, and bar charts. The following sub-sections show examples that are displayed in the dashboard.

A. Clustering Visualisation Maps

Applying the k-means clustering technique, the 2005 results of average hourly direct irradiation and diffuse irradiation over the entire Java-Bali region are presented in Fig. 4 and Fig. 5, accordingly. It should be noted that the calculation of solar PV output in the model is based on 1 MW of solar PV capacity.

Of the two results of direct and diffuse irradiation clustering analysis using the k-means technique, as shown in Fig. 4 and Fig. 5, some parts of the East Java and whole Bali areas are clustered as the area having the highest range of average direct irradiation values, i.e., 0.32-0.35 kW/m^2, that is in orange colour. Conversely, as expected, East Java and all areas of Bali have also been identified as having the lowest average diffuse irradiation values, shown in blue, among other provinces or regions, ranging from 0.27-0.29 kW/m^2.

The clustering results for 2005 show that most areas of East Java province and all of Bali have demonstrated a moderately high solar PV output compared to the rest of the Java-Bali region. This implies most of East Java and Bali are better candidate locations for placing solar PV power plants than other areas in western Java.

It is understandable since more direct sunlight and less diffuse sunlight fall on the area in one year, as preferred by planners. Of course, planners should be aware that the location chosen to place a solar PV plant should not be determined based on a single year's analysis, but instead incorporates annual data sets over a long period, and the analysis cannot be used as the only resources for making decision on the PV placement.

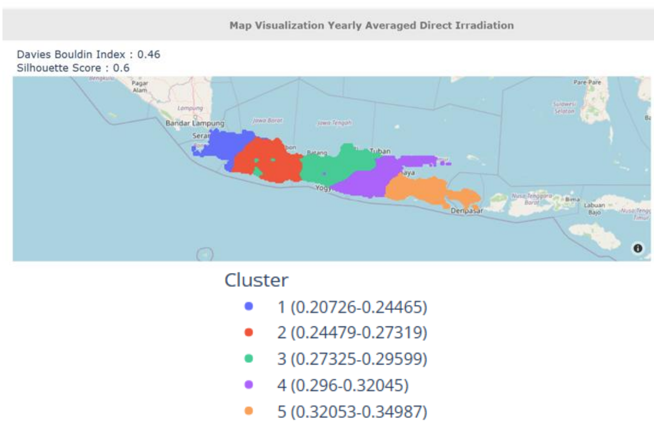

Fig. 4. Mapping of k-means clustering (5 clusters) of average hourly direct irradiation throughout the Java-Bali region in 2005.

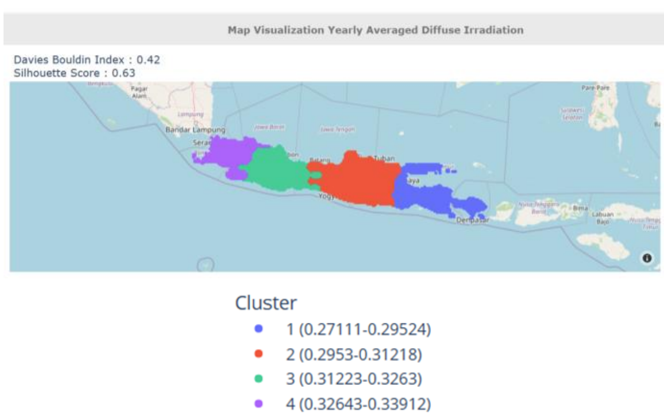

Fig. 5. Mapping of k-means clustering (4 clusters) of average hourly diffuse irradiation throughout the Java-Bali region in 2005

Calculating solar PV output becomes complex when temperature clustering analysis is taken into account. The efficiency of solar PV modules is known to be affected by temperature, making it one of the crucial variables to consider. Notwithstanding, clustering analysis of the temperature dataset is particularly beneficial when determining the optimal solar PV output, as it helps to factor in the impact of temperature on the PV performance.

The clustering results of the 2005 hourly average temperature (°C) across the Java-Bali region are presented in Fig. 6, in terms of the Java-Bali region's average temperature clustering map.

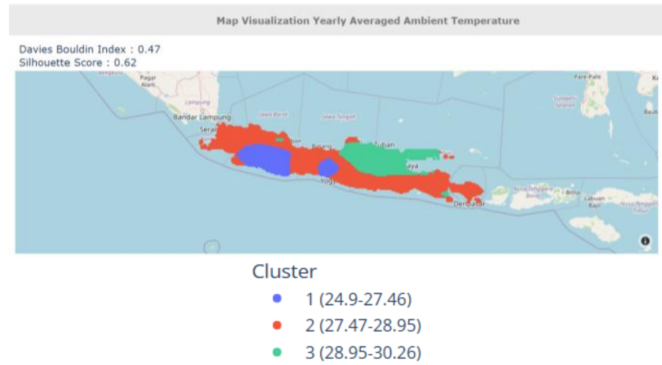

Fig. 6. Mapping of k-means clustering (3 clusters) of average hourly temperature throughout the Java-Bali region in 2005.

Based on Fig. 6, it can be observed that the average temperature in cluster 2, which covers the southern and eastern regions of Java and Bali, is not the highest when compared to cluster 3. Additionally, the areas depicted in blue indicate the lowest average temperature among all three clusters.

B. Location Identification

The geographical location with the highest average hourly direct irradiation within a cluster in a specific year can be identified through the clustering results.

In addition, similar features can be used to identify locations with the highest hourly average diffuse irradiance in a given year, possibly indicating regions that are frequently covered by clouds, and the highest hourly average temperature, which may reduce the efficiency of solar PV panels. Fig. 7 displays the latitude and longitude of the location that exhibits the highest average direct irradiation value in cluster 5, as shown in Fig. 4.

Fig. 7. The latitude and longitude of the location show the highest average direct irradiation value in cluster 5 (as shown in Fig. 4, i.e., circled red) at the Southeastern tip of Nusa Penida Island, Bali).

The dashboard allows users to view numerical information regarding the parameters across different locations within each cluster. The bar charts provide an overview of the highest average value of the parameters, which can be useful for policy studies and further analysis. The dataset can also be used by potential users for conducting various data analyses and statistical studies to complement their research. Additionally, users can carry out a more detailed study of the entire region or province (which includes six provinces in Java besides Bali) by clustering the parameters on a monthly basis in different years.

V. CONCLUSIONS

The paper introduced a framework and dashboard to analyse and cluster the long-term temporal-based parameters of a solar PV output model, on an hourly basis. The case study was conducted in the Java-Bali region of Indonesia. The framework aimed to provide crucial insights into the potential of solar PV, especially in terms of temporal and spatial issues that could affect investment decisions.

The paper presents a clustering analysis that shows the mapping of locations based on averaged hourly direct and diffuse irradiation, and hourly temperature. This analysis is presented on an annual basis through a dashboard. The dashboard has been developed using Python and provides a clear visualization of the clustering of locations. Additionally, it displays some numerical information on the geographical locations as well as bar charts.

The results obtained from clustering analysis can help in identifying the most suitable areas for the installation of PV solar plants. The suitability can be determined based on several criteria such as the average direct irradiation values, the average diffuse irradiation values (for areas that are mostly covered by clouds throughout the year), and the areas with the highest value of hourly average temperature. It is important to note that locations with a higher average temperature can potentially reduce the efficiency of the solar PV modules installed in those areas.

Despite challenges in determining an appropriate hyperparameter value that involves Silhouette Score evaluation criteria, this study found that, based on k-means clustering results using the 2005 dataset, most of East Java and Bali have the highest range of average direct irradiation while these same areas, as expected, exhibit the lowest average diffuse irradiation. Therefore, these areas can be considered better candidate locations for placing solar PV power plants than other locations in the Java-Bali regions.

For future work, this study will further explore the application of k-means clustering on all three parameters simultaneously while comparing the results with solar PV output clustering. Additionally, the study will utilise the DBSCAN clustering method on longer hourly-based data while taking into account the dataset spanning from 2005 to 2022. The performance of k-means clustering will be compared to that of the DBSCAN method. Furthermore, the study will employ time series methods and machine learning regression-based methods to forecast solar PV output accurately.

While a one-year temporal-based parameter analysis can give users a rough estimate of the potential for solar PV deployment, as shown in the examples of this paper, further work on exploring more approaches to clustering solar irradiations and temperature would be of great value to stakeholders.

A comprehensive analysis can involve examining detailed monthly clustering of the observed regions, comparing the outcomes obtained from year-by-year clustering for an extended annual period, and analysing the clustering of all three parameters simultaneously to generate 3-dimensional clustering outcomes, while also comparing the results with the solar PV output clustering.

ACKNOWLEDGMENT

The authors are grateful for the funding provided by The Ministry of Education, Culture, Research, and Technology of The Republic of Indonesia through the Fundamental Research Scheme 2023, with contract no. 002/SP2H/PT/LL7/2023, 11/SP2H/PT/LPPM-UKP/2023. We also thank Sean Mingardi, Dick Jovian, and Jimlee Widjaya, graduates of Petra Christian University Informatics Department, who helped the authors in compiling the data.

REFERENCES

[1] IRENA, *Renewable capacity statistics 2023*. Abu Dhabi: International Renewable Energy Agency, 2023.

[2] X. Liang, "Emerging power quality challenges due to integration of renewable energy sources," *IEEE Transactions on Industry Applications*, vol. 53, issue. 2, pp. 855-866, March-April 2017.

[3] R. Go, F. Kahrl, and C. Kolster, "Planning for low-cost renewable energy," *The Electricity Journal*, vol. 33, issue. 2, pp. 106698, March 2020.

[4] Y. Tanoto, N. Haghdadi, A. Bruce, and I. MacGill, "Reliability-cost trade-offs for electricity industry planning with high variable renewable energy penetrations in emerging economies: A case study of Indonesia's Java-Bali grid," Energy, vol. 227, pp. 120474, March 2021.

[5] A. K. Singh, S. Mittal, P. Malhotra, and Y. V. Srivastava, "Clustering evaluation by Davies-Bouldin index (DBI) in cereal data using K-Means," In *Proc. of the 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 11-13 March 2020, Erode, India* [Online]. Available: IEEE Xplore, http://www.ieee.org. [Accessed: Jul. 20, 2023].

[6] K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," In. *Proc. of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 6-9 October 2020, Sydney, Australia* [Online]. Available: IEEE Xplore, http://www.ieee.org. [Accessed: Jul. 20, 2023].

[7] PT. PLN (Persero), *Electricity supply business plan 2017-2026*. Jakarta: PT. Perusahaan Listrik Negara (Persero), 2017.

[8] IRENA, *Renewable energy prospects: Indonesia*. Abu Dhabi: International Renewable Energy Agency, 2017.

[9] DGE. MEMR. *Electricity statistics 2021 (in Bahasa Indonesia).* Jakarta: Directorate General Electricity, Ministry of Energy and Mineral Resources Republic of Indonesia. 2022.

[10] AHK Indonesien, In decentralizing energy in Indonesia, the future of solar power remains bright. Jakarta: AHK Indonesien; 2022. [Online]. Available: https://indonesien.ahk.de/en/ [Accessed: Jul. 19, 2023].

[11] S. Pfenninger and I. Staffel, "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data," Energy, vol. 114, pp. 1251-1265, November 2016.

[12] D. Deng, "DBSCAN Clustering Algorithm Based on Density," In. *Proc. of the 2020 7th International Forum on Electrical Engineering and Automation (IFEEA), 25-27 September 2020, Hefei, China* [Online]. Available: IEEE Xplore, http://www.ieee.org. [Accessed: Jul. 21, 2023].