Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 10, No. 2, June 2024, pp. 359-372
ISSN: 2338-3070, DOI: 10.26555/jiteki.v10i2.29032

Exploring Energy Data through Clustering: A Hyperparameter Approach to Mapping Indonesia's Primary Energy Supply

Agus Perdana Windarto[1], Yerika Puspa Rosanti[2], Mesran[3]
[1]Informatics Engineering Program, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
[2]Master of Physics Teaching, Institut Teknologi Bandung, West Java, Indonesia
[3]Department of Computer Science, Universitas Budi Darma, Medan, North Sumatra, Indonesia

ARTICLE INFO

Article history:

Received June 10, 2024
Revised June 28, 2024
Published July 11, 2024

Keywords:

Energy Data;
Clustering;
Hyperparameter Tuning;
K-Means;
Indonesia Energy Supply;
Davies-Bouldin Index;
Data Analysis

ABSTRACT

The rapid economic growth and population development in Indonesia have significantly increased the demand for energy, presenting complex challenges in managing the primary energy supply due to geographical variability and dispersed natural resources. This study addresses these challenges by applying clustering techniques with a hyperparameter approach to explore and map Indonesia's primary energy supply. The research contributes to the field by offering an effective method for analyzing energy data patterns and optimizing energy management. Secondary data on energy production, consumption, and distribution from reliable sources such as the Ministry of Energy and Mineral Resources were collected and analyzed. Various clustering algorithms, including K-Means, Fast K-Means, X-Means, and K-Medoids, were applied to identify energy supply patterns across different regions. The Davies-Bouldin Index was used to evaluate the effectiveness of the clustering algorithms. The results indicate that distance measures such as Euclidean Distance and Chebychev Distance consistently show excellent clustering performance. The study found that the choice of distance measure significantly impacts the clustering quality. The insights gained from this analysis provide valuable information for stakeholders involved in energy planning and policy-making, enhancing the efficiency and sustainability of energy management in Indonesia. This research establishes a foundation for further detailed and holistic energy data analysis, supporting better decision-making in energy planning and development.

Corresponding Author:

Agus Perdana Windarto, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
Email: agus.perdana@amiktunasbangsa.ac.id

1. INTRODUCTION

The rapid economic growth and population development in Indonesia have significantly increased energy demand over the past few decades [1], [2], [3]. As the fourth most populous country in the world with a rapidly growing economy, Indonesia relies on its primary energy supply to support economic activities, infrastructure, and the needs of its society [4], [5]. However, managing the primary energy supply presents complex challenges. The geographical variability and dispersed natural resources result in different energy supply structures across various regions of Indonesia [6], [7]. Additionally, the swift economic growth puts pressure on energy infrastructure, creating a need for efficient and measurable management [8], [9]. In this context, research on energy data exploration becomes critically important [10], [11], [12]. Through careful data analysis, crucial information about energy supply patterns, resource availability, and consumer needs can be uncovered. One analytical tool used is clustering techniques, which allow for the grouping of data based on similar characteristics to achieve a more even mapping of primary energy supply [13], [14], [15], [16]. Several robust algorithms are available for clustering [9], [17]. Clustering algorithms are essential in data analysis for grouping data with similar characteristics into distinct clusters [18], [19]. Popular clustering algorithms include

K-Means, which optimizes the position of cluster centroids; Hierarchical Clustering, which builds a hierarchy of clusters; DBSCAN, which identifies dense regions within the data; Mean Shift, which focuses on centroid movement; Gaussian Mixture Model, which assumes normal distribution within clusters; OPTICS, which creates a cluster hierarchy based on data density; Agglomerative Clustering, which merges the closest clusters; and Spectral Clustering, which uses a graph representation of data [20], [21], [22], [23]. The choice of the appropriate clustering algorithm largely depends on the structure and characteristics of the data being processed and the analysis goals to achieve optimal and meaningful results [24], [25].

The study [26] proposes a data deduplication solution based on the numerical conversion of datasets and the use of Dynamic K-Means Clustering (DKMEAN) to cluster and identify duplicate chunks. This method has experimentally demonstrated improvements in dataset quality and reduced resource consumption, achieving a 65.78% reduction in large datasets. However, there is a lack of explanation on how this data deduplication method can be applied to real-world use cases outside experimental settings. The article focuses more on developing data deduplication techniques using a K-Means Clustering-based Data Mining approach but provides insufficient concrete examples or case studies on implementing these techniques in real-world scenarios. Subsequent research [27] utilizes K-Means Clustering in data mining to develop sales strategies at S&R Baby Store. This article offers deep insights into using the K-Means algorithm to cluster sales transaction data to determine highly demanded, moderately demanded, and less demanded products. However, it lacks a more in-depth explanation of the evaluation process related to the product clustering results using the K-Means Clustering method. Although the article mentions using the Davies-Bouldin Index (DBI) to evaluate clustering quality, it does not detail how DBI is calculated and how these evaluation results can be interpreted.

Previous studies in this domain often face challenges in determining optimal hyperparameters for clustering, particularly in the context of complex and varied energy data such as that of Indonesia [27], [28]. Hyperparameters are variables that need to be set at the beginning of clustering analysis and can significantly influence the results and interpretation of the clustering [29], [30], [31], [32]. Therefore, this research aims to bridge this knowledge gap by proposing an effective and precise hyperparameter approach for mapping Indonesia's primary energy supply through clustering techniques [33], [34], [35], [36]. By understanding the emerging patterns from energy data, stakeholders such as the government, industry, and research institutions can make better decisions in energy planning, development, and policy-making [44], [45], [46]. Through this innovative approach, it is hoped that this article will contribute significantly to the understanding and management of energy in Indonesia and serve as a foundation for further research in holistic and detailed energy data analysis.

2. METHODS

This study employs a secondary data analysis approach, involving the collection of primary energy data for Indonesia from reliable sources such as the Ministry of Energy and Mineral Resources (ESDM) and other relevant institutions. The data used includes information on the production, consumption, distribution, and types of energy resources available across various regions of Indonesia [37]. The research design aims to address the issue of data imbalance in the primary energy dataset of Indonesia by applying clustering techniques with a hyperparameter approach in the exploration of primary energy data. The goal is to map energy supply patterns, identify the geographical distribution of energy resources, and provide a deeper understanding of the energy structure in different regions of Indonesia [38], [39], [40], [41].

The research framework (Fig. 1) outlined in the provided diagram represents a structured approach to addressing the imbalance in Indonesia's primary energy dataset through the application of clustering techniques with a hyperparameter approach. The process begins with identifying the problem, where the challenges related to the imbalance in energy data are recognized. Following this, a comprehensive literature review is conducted, gathering data from credible sources such as the Ministry of Energy and Mineral Resources (ESDM) and other relevant institutions [47], [48], [49], [50]. This step ensures that the research is grounded in reliable and pertinent information about energy production, consumption, distribution, and resource types across various regions of Indonesia. The next phase involves data collection, where primary energy data is sourced from the identified repositories. This data is then prepared, which includes cleaning, handling missing values, and normalizing the data to ensure it is suitable for analysis. Subsequently, the model development phase focuses on applying clustering techniques. This step involves selecting appropriate clustering algorithms and determining the optimal hyperparameters to enhance the clustering results, thus enabling effective exploration of the primary energy data.

Once the model is developed, it undergoes testing, and the results are measured to ensure accuracy and meaningfulness. The clustering quality is then evaluated using various metrics, such as the Davies-Bouldin Index, to validate the clusters formed. The final phase involves analyzing the clustering results to identify geographical patterns in the energy supply, providing a deeper understanding of the distribution and structure of energy resources across different regions of Indonesia.

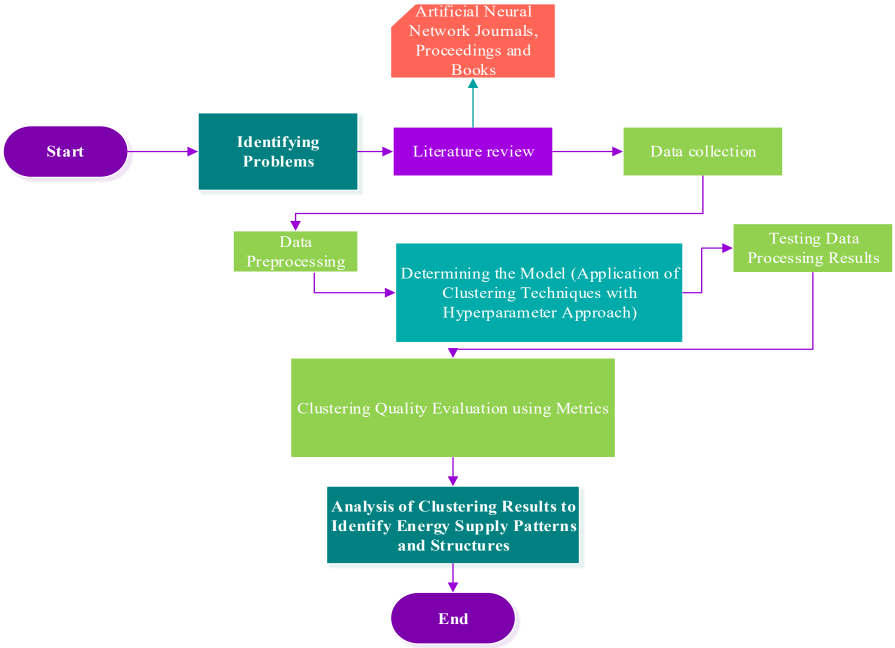

Fig. 1. The research framework

This systematic approach concludes with the findings and insights gained from the clustering analysis, offering valuable information for stakeholders involved in energy planning, development, and policy-making. The framework not only addresses the immediate research problem but also lays the groundwork for further studies in energy data analysis.

2.1. Primary Energy Supply Proportion Dataset

The proportion of primary energy supply in Indonesia illustrates the relative distribution of various energy sources used to meet the country's primary energy needs. Oil and natural gas have long been the main components of energy supply, given that Indonesia is the largest oil producer in Southeast Asia and has significant natural gas reserves. Additionally, coal also plays a crucial role in the energy supply, particularly for coal-fired power plants. In recent years, efforts to develop renewable energy have also gained more attention, including the use of hydro and geothermal energy. However, the proportion of energy supply from renewable sources remains relatively small compared to conventional energy sources. A careful analysis is needed to understand the shift in the proportion of primary energy supply in Indonesia and its implications for sustainability and future energy policies. Table 1 shows the data on Indonesia's primary energy supply proportions.

Table 1. Indonesia's Primary Energy Supply Proportion Data

Energy Type	Total Energy Proportion in Percentage Form Every year				
	2015	2016	2012	2013	2014
Oil	35.25	35.19	38.45	38.89	38.37
Coal	23.47	23.43	22.35	19.82	20.61
Gas	18	17.97	16.81	17.69	17.48
Hydroelectric Power Plant	2.27	2.27	1.89	2.52	2.45
Geothermal	1.05	1.05	0.98	1	1.04
Biomass	19.93	20.06	19.49	20.04	19.96
Biofuels	0.04	0.04	0.02	0.04	0.08

2.2. Data Preprocessing

This stage outlines the steps taken to ensure that the data is well-prepared for the modeling process. These steps include removing empty or duplicate data entries, scaling the data to a range of 0 to 1 using a min-max scaler, and assigning appropriate labels to the data. Labeling the data is crucial in the subsequent clustering process. By labeling, data samples can be categorized and grouped based on shared characteristics, which is essential for effective cluster formation. This becomes particularly relevant in the data balancing process using clustering algorithms optimized with adjusted hyperparameters, which will be discussed in the next section. For example, the data preprocessing shown in Table 1 is presented in Table 2.

Table 2. Preprocessing Results

Energy Type	Total Energy Proportion in Percentage Form Every year				
	2015	2016	2012	2013	2014
Oil	0.9064	0.9048	0.9887	1.0000	0.9866
Coal	0.6033	0.6023	0.5745	0.5094	0.5297
Gas	0.4626	0.4618	0.4320	0.4546	0.4492
Hydroelectric Power Plant	0.0579	0.0579	0.0481	0.0643	0.0625
Geothermal	0.0265	0.0265	0.0247	0.0252	0.0262
Biomass	0.5122	0.5156	0.5009	0.5151	0.5130
Biofuels	0.0005	0.0005	0.0000	0.0005	0.0015

2.3. Clustering Using K-Means

The K-Means algorithm partitions data into clusters such that similar data points are grouped within the same cluster, while dissimilar data points are placed in different clusters. Sarwono elaborates on the K-Means algorithm in more detail as follows [42]. The flowchart illustrates (**Fig. 2**) the iterative process of the K-Means clustering algorithm, a popular method in data mining and machine learning for partitioning datasets into K distinct, non-overlapping clusters. The process begins with the initialization step, where the user specifies the number of clusters, K, to divide the dataset into. Following this, initial centroids for the K clusters are determined, which can be done randomly or through more sophisticated methods like K-Means++ for better initial placement.

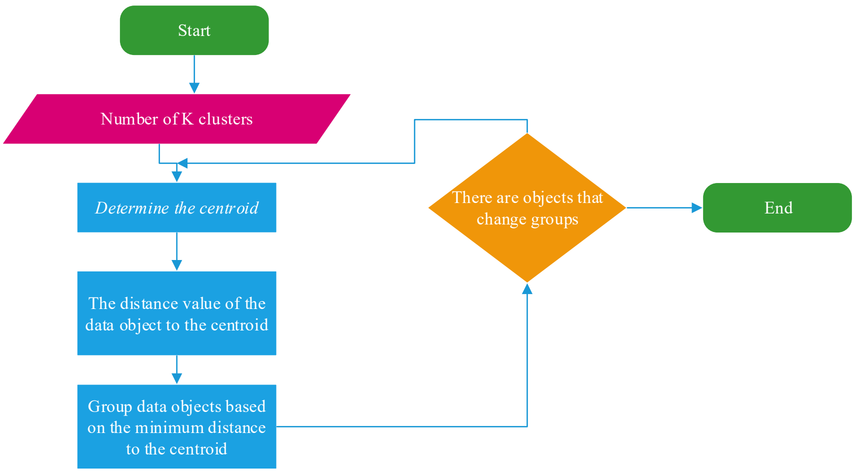

Fig. 2. The flowchart illustrates of K-Means

Next, the algorithm calculates the distance of each data object to the centroids using a chosen distance metric such as Euclidean distance. Based on these distance calculations, data objects are assigned to the nearest centroid, grouping them into clusters. The algorithm then checks if any data objects have changed their assigned clusters during the current iteration. If changes are detected, the algorithm continues to iterate, recalculating distances and reassigning data points to the nearest centroids, refining the clusters further.This iterative process repeats until no data objects change their cluster assignments, indicating that the clusters have stabilized. At this point, the algorithm terminates, and the final clusters are established with their respective centroids representing the central points. The decision point to check for changes ensures that the algorithm iterates enough times to achieve optimal clustering, minimizing variance within each cluster and ensuring that the clusters are as distinct and well-separated as possible.

2.4. Clustering Using X-Means

The X-Means algorithm is employed to cluster data into k groups based on their attributes. The process begins with the initialization of centroids, where the desired number of clusters, k, is selected. Randomly chosen k points serve as the initial centroids, which act as the center points for each cluster. Following this, the algorithm calculates the distance between each data point and the centroids using a distance metric such as Euclidean distance. The Euclidean distance between two points A(x1, y1) and B(x2, y2) in a two-dimensional space is calculated as the square root of the sum of the squared differences between the corresponding coordinates.

Once the distances are calculated, each data point is assigned to the cluster with the nearest centroid, meaning the cluster whose centroid has the smallest Euclidean distance to the data point. After all data points are assigned to clusters, the centroids are updated by recalculating their positions as the average of all data points within each cluster. This new centroid position ensures that the centroids accurately represent the center of their respective clusters.

The algorithm then iterates through these steps—calculating distances, assigning data points to clusters, and updating centroids—until a stopping condition is met. The stopping condition could be a predefined number of iterations or minimal changes in centroid positions between iterations. Typically, the K-Means algorithm halts when the centroids' positions stabilize and no longer change significantly, indicating that the clusters have converged to stable points. Through these iterative refinements, the X-Means algorithm ensures that each data point is optimally assigned to the most appropriate cluster based on its proximity to the centroid, thus achieving a stable and accurate clustering of the data.

2.5. Clustering Using K-Medoids

The flowchart represents the process of the K-Medoids clustering algorithm, which is similar to K-Means but uses medoids instead of centroids to define the clusters. The algorithm begins with the initialization step, where the desired number of clusters, k, is specified. Randomly selected k data points serve as the initial medoids, acting as representative points for each cluster.

Following this Fig. 3 , the algorithm calculates the distance between each data point and the medoids using a distance metric such as the Euclidean distance. Each data point is then assigned to the cluster with the nearest medoid. After all data points are assigned, the algorithm updates the medoids by recalculating them. The new medoid for each cluster is determined by selecting the data point within the cluster that minimizes the total distance to all other points in the same cluster.

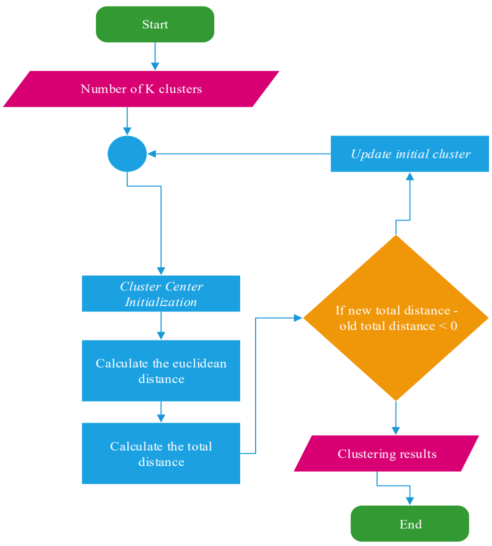

Fig. 3. The flowchart illustrates of K-Medoids

The algorithm then checks whether any data points have changed clusters. If the change in the positions of medoids is less than a predefined threshold (Δ), the algorithm considers the clusters to have converged and stops further iterations. If there are significant changes, the process of calculating distances, assigning data points to clusters, and updating medoids is repeated.

This iterative process continues until the medoids stabilize, indicating that the clusters have reached their optimal configuration. The final output is the clusters with their respective medoids. This approach ensures that the algorithm minimizes the dissimilarity within clusters, providing a robust and efficient clustering solution that is less sensitive to outliers compared to K-Means.

2.6. Model Evaluation

In the evaluation model, the process involves several critical steps to assess the effectiveness and accuracy of the clustering model. Fig. 4 illustrates the evaluation model, capturing the key components and flow of the evaluation process.

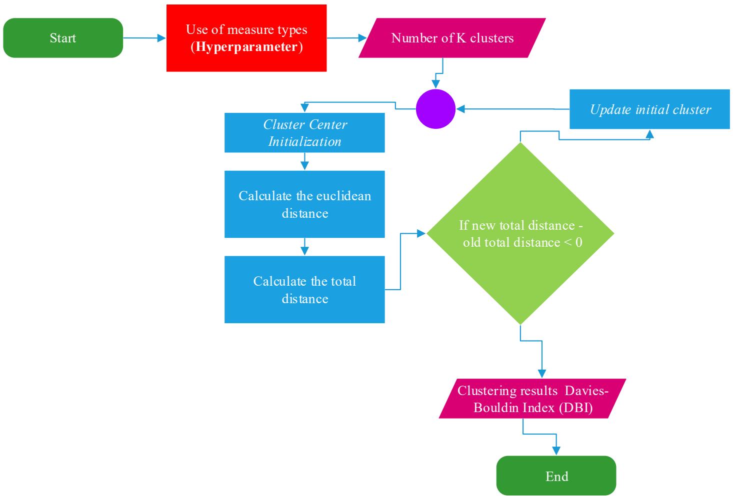

Fig. 4. Hyperparameter Model Evaluation Workflow

The flowchart illustrates the model evaluation process for clustering (Fig. 4). The evaluation process begins by specifying the type of evaluation metric to be used, either the Silhouette Score or the Davies-Bouldin Index, which are essential for assessing the quality of the clusters. The clusters are then calculated based on the chosen clustering algorithm and hyperparameters. Next, the selected evaluation metric is calculated for each cluster. The Silhouette Score measures how similar a data point is to its own cluster compared to other clusters, with higher scores indicating better-defined clusters. The Davies-Bouldin Index measures the average similarity ratio of each cluster with respect to other clusters, with lower values indicating better clustering performance.

A decision point then determines whether new metrics or hyperparameters need to be evaluated. If a new evaluation is needed, the clustering process is repeated with the updated settings. The calculated metrics are compared to identify the best clustering performance. This iterative process ensures that the evaluation is thorough, allowing for the identification of the optimal clustering configuration. The process includes a stop criterion, such as the Davies-Bouldin Index being below a certain threshold, to ensure that the clustering continues until satisfactory quality is achieved. Once the stop criteria are met, the optimal clustering results are identified, indicating that the best possible clustering configuration has been achieved. The evaluation process concludes, providing the final assessment of the clustering model. Through this comprehensive and iterative evaluation process, the flowchart ensures that the clustering model is rigorously assessed, leading to the identification of the best clustering configuration for mapping Indonesia's primary energy supply.

3. RESULTS AND DISCUSSION

3.1. Results

In this stage, we present experimental results and explore the primary energy supply proportions in Indonesia using clustering techniques and the Davies-Bouldin Index (DBI) evaluation metric. The research method involves applying various clustering algorithms, such as K-Means, Fast K-Means, X-Means, and K-Medoids, with diverse measure types including Mixed Measures and Numerical Measures. These algorithms use different metrics such as Euclidean Distance, Canberra Distance, and Chebychev Distance, among others. The evaluation is conducted by measuring the DBI of the clustering results to assess the quality of the clusters produced by each algorithm. This context allows us to understand how the subsequent results and discussions will address the effectiveness and suitability of different clustering algorithms in mapping the proportions of primary energy supply in Indonesia.

The objective of this research is to apply clustering techniques with a hyperparameter approach to explore Indonesia's primary energy data. The aim is to map energy supply patterns, identify the geographical distribution of energy resources, and provide a deeper understanding of the energy structure across various regions of Indonesia. Thus, this study aims to significantly contribute to decision-making related to strategic planning, infrastructure development, and the formulation of sustainable energy policies in Indonesia through the use of clustering techniques and the Davies-Bouldin Index (DBI) evaluation metric. The processed results can be seen in Table 3 (K-Means), Table 4 (Fast K-Means), Table 5 (X-Means), and Table 6 (K-Medoids).

Table 3 (Fig. 5) shows excellent clustering performance with several distance measures. Measures like Mixed Measures, Euclidean Distance, Canberra Distance, Chebychev Distance, Dynamic Time Warping Distance, Kernel Euclidean Distance, Manhattan Distance, Generalized Divergence, and Squared Euclidean Distance all achieve a low DBI of 0.065, indicating tight and well-separated clusters. Correlation Similarity, with a high DBI of 2.137, performs poorly, suggesting it may not be suitable for this dataset. Cosine Similarity and Jaccard Similarity show moderate performance with DBIs of 1.021 and 0.400, respectively. Some metrics have unknown DBI values, indicating a need for further analysis.

Table 3. Results from the K-Means Analysis

Measures Types		Davies Bouldin
Mixed Measures	**Mixed Measures**	**0.065**
Nurmerical Measures	**EuclideanDistance**	**0.065**
	CanberraDistance	**0.065**
	ChebychevDistance	**0.065**
	CorrelationSimilarity	2.137
	CosineSimilarity	0.121
	DiceSimilarity	unknwon
	DynamicTimeWarpingDistance	**0.065**
	InnerProductSimilarity	unknwon
	JaccardSimilarity	0.400
	KernelEuclideanDistance	**0.065**
	ManhattanDistance	**0.065**
	MaxProductSimilarity	unknwon
	OverlapSimilarity	unknwon
Bregman Divergences	**GeneralizedDivergence**	**0.065**
	KLDivergence	unknwon
	Mahalanobis Distance	0.151
	SquaredEuclideanDistance	**0.065**

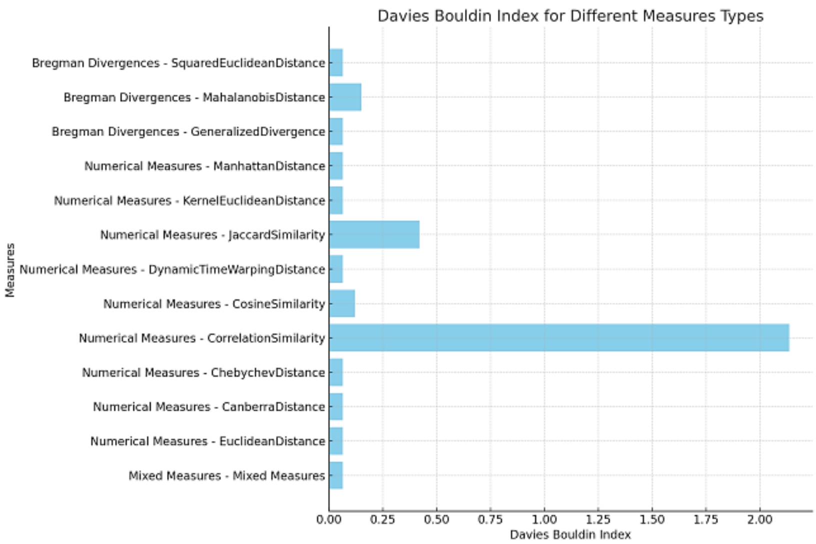

Fig. 5. Graph of K-Means Analysis

Table 4 (Fig. 6) shows the K-Means (fast) algorithm also demonstrates excellent clustering quality with several distance measures, such as Mixed Measures, Euclidean Distance, Canberra Distance, Chebychev Distance, Dynamic Time Warping Distance, Manhattan Distance, Generalized Divergence, and Squared Euclidean Distance, all achieving a low DBI of 0.065. Cosine Similarity performs relatively well with a DBI of 0.121. Jaccard Similarity and Mahalanobis Distance provide moderate clustering quality with DBIs of 0.400 and 0.396, respectively. Kernel Euclidean Distance shows less effective clustering with a DBI of 0.473. Some measures have unknown DBI values.

Table 4. Results from K-Means (fast) Analysis

Measures Types		Davies Bouldin
Mixed Measures	**Mixed Measures**	**0.065**
Nurmerical Measures	**EuclideanDistance**	**0.065**
	CanberraDistance	**0.065**
	ChebychevDistance	**0.065**
	CorrelationSimilarity	Unknown
	CosineSimilarity	0.121
	DiceSimilarity	Unknown
	DynamicTimeWarpingDistance	**0.065**
	InnerProductSimilarity	Unknown
	JaccardSimilarity	0.400
	KernelEuclideanDistance	0.475
	ManhattanDistance	**0.065**
	MaxProductSimilarity	Unknown
	OverlapSimilarity	Unknown
Bregman Divergences	**GeneralizedDivergence**	**0.065**
	KLDivergence	Unknown
	Mahalanobis Distance	0.396
	SquaredEuclideanDistance	**0.065**

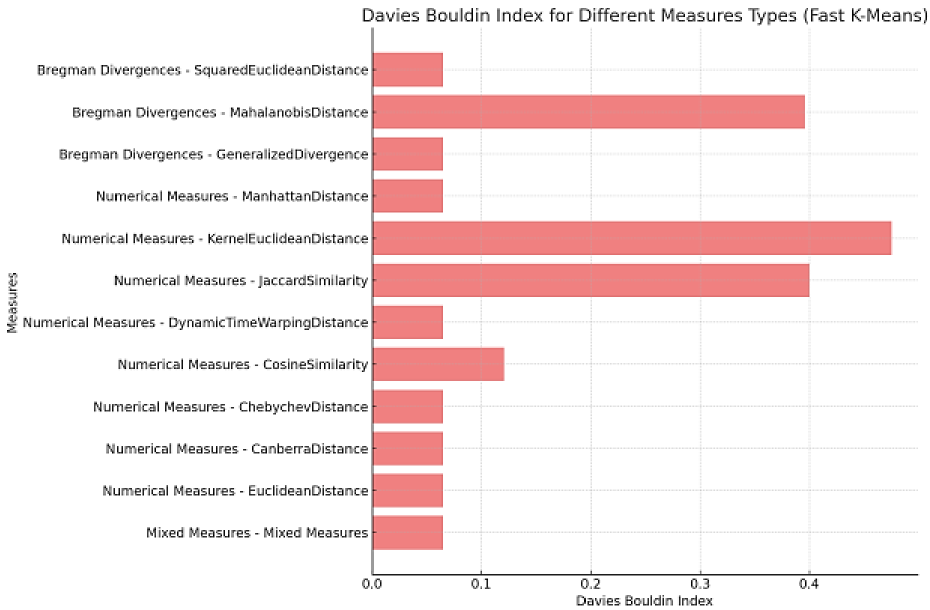

Fig. 6. Graph of K-Means (fast) Analysis

Table 5 (Fig. 7) shows the X-Means algorithm performs excellently with several distance measures, including Mixed Measures, Euclidean Distance, Canberra Distance, Chebychev Distance, Dynamic Time Warping Distance, Kernel Euclidean Distance, Manhattan Distance, Generalized Divergence, and Squared Euclidean Distance, all achieving a low DBI of 0.065. Cosine Similarity shows relatively good performance with a DBI of 0.121. Jaccard Similarity and Mahalanobis Distance exhibit moderate clustering quality with DBIs of 0.400 and 0.396, respectively. Correlation Similarity performs poorly with a high DBI of 2.137. Some metrics have unknown DBI values.

Table 6 (Fig. 8) shows the K-Medoids algorithm shows excellent clustering quality with several distance measures, such as Mixed Measures, Euclidean Distance, Chebychev Distance, Dynamic Time Warping Distance, Manhattan Distance, Generalized Divergence, and Squared Euclidean Distance, all achieving a low DBI of 0.067. Canberra Distance and Cosine Similarity have slightly higher DBIs of 0.099, still showing relatively good clustering quality. Correlation Similarity and Kernel Euclidean Distance exhibit moderate clustering quality with a DBI of 0.118. Mahalanobis Distance has a higher DBI of 0.156, indicating less effective clustering. Some measures have unknown DBI values.

Table 5. Results from X-Means Analysis

Measures Types		Davies Bouldin
Mixed Measures	**Mixed Measures**	**0.065**
Nurmerical Measures	**Euclideandistance**	**0.065**
	Canberradistance	**0.065**
	Chebychevdistance	**0.065**
	Correlationsimilarity	2.137
	Cosinesimilarity	0.121
	Dicesimilarity	Unknown
	Dynamictimewarpingdistance	**0.065**
	Innerproductsimilarity	Unknown
	Jaccardsimilarity	0.400
	Kerneleuclideandistance	0.065
	Manhattandistance	0.065
	Maxproductsimilarity	Unknown
	Overlapsimilarity	Unknown
Bregman Divergences	**Generalizeddivergence**	**0.065**
	Kldivergence	Unknown
	Mahalanobis Distance	0.396
	Squaredeuclideandistance	**0.065**

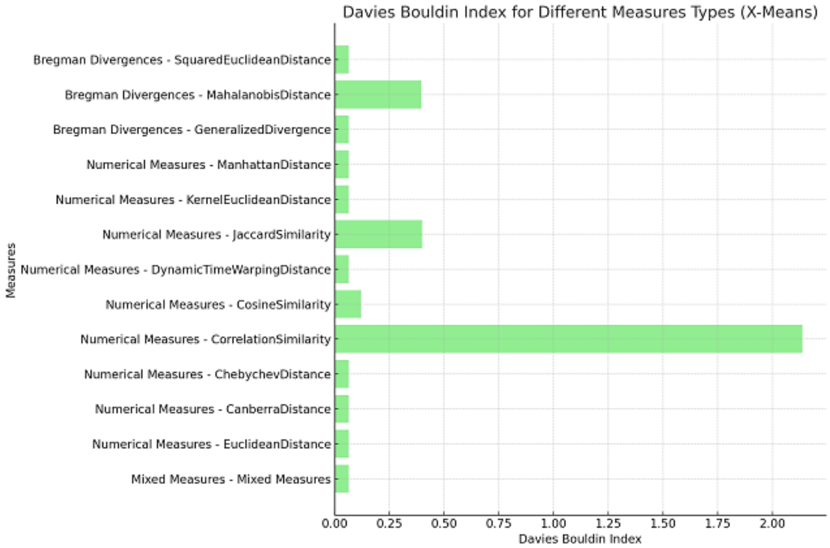

Fig. 7. Graph of X-Means Analysis

Table 6. Results from K-Medoids Analysis

Measures Types		Davies Bouldin
Mixed Measures	**Mixed Measures**	**0.067**
Nurmerical Measures	**EuclideanDistance**	**0.067**
	CanberraDistance	0.099
	ChebychevDistance	**0.067**
	CorrelationSimilarity	0.118
	CosineSimilarity	0.099
	DiceSimilarity	Unknown
	DynamicTimeWarpingDistance	**0.067**
	InnerProductSimilarity	Unknown
	JaccardSimilarity	Unknown
	KernelEuclideanDistance	0.118
	ManhattanDistance	**0.067**
	MaxProductSimilarity	Unknown
	OverlapSimilarity	Unknown
Bregman Divergences	**GeneralizedDivergence**	**0.067**
	KLDivergence	Unknown
	Mahalanobis Distance	0.156
	SquaredEuclideanDistance	**0.067**

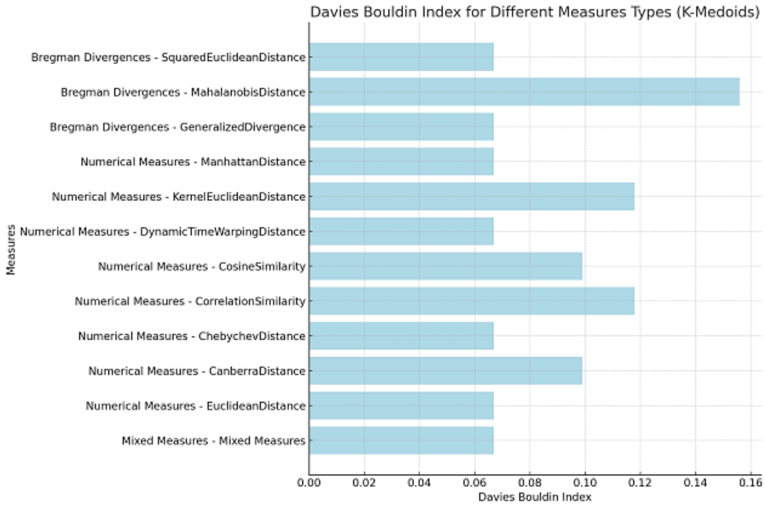

Fig. 8. Graph of K-Medoids Analysis

3.2. Discussions

From the analysis across all tables, it is evident that the choice of distance measure significantly impacts the clustering quality for each algorithm. Measures like Euclidean Distance, Chebychev Distance, Dynamic Time Warping Distance, Manhattan Distance, Generalized Divergence, and Squared Euclidean Distance consistently show excellent performance across different algorithms. These measures indicate that the clusters are well-defined and separated.

The Fig. 9 shows the comparison graph of the Davies Bouldin Index values for various measurement types using K-Means, K-Means (fast), X-Means, and K-Medoids clustering methods reveals some interesting findings. Firstly, Correlation Similarity consistently shows the highest Davies Bouldin Index values in the K-Means and X-Means methods, indicating the worst clustering performance for this measurement type. Additionally, Mahalanobis Distance also exhibits relatively high values in the K-Means (fast), X-Means, and K-Medoids methods, indicating suboptimal performance.

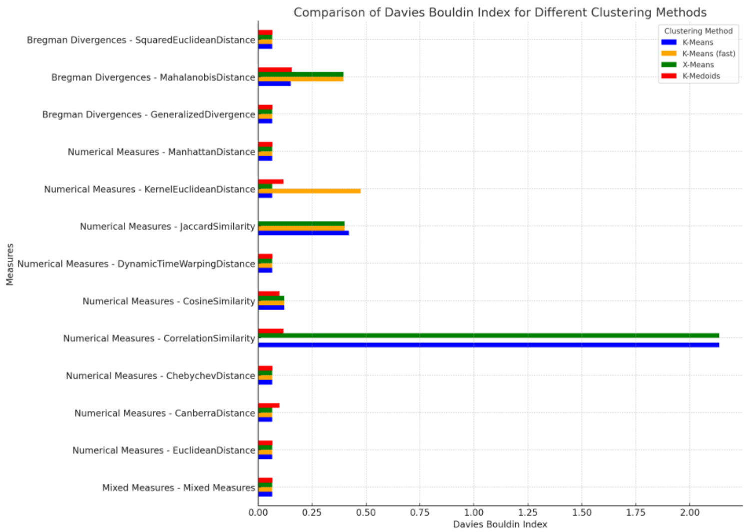

Fig. 9. Comparative Analysis of the Davies Bouldin Index for Various Clustering Methods

Kernel Euclidean Distance in the K-Means (fast) method stands out with significantly higher values compared to other methods, indicating poor performance. Meanwhile, Canberra Distance and Cosine Similarity have slightly higher values than most other measurements but remain lower than Correlation Similarity and Mahalanobis Distance. On the other hand, measurements such as Euclidean Distance, Chebychev Distance, Dynamic Time Warping Distance, Manhattan Distance, Generalized Divergence, and Squared Euclidean Distance consistently show low Davies Bouldin Index values across all clustering methods, indicating better clustering performance. Although the K-Medoids method tends to show slightly higher values compared to other methods for some measurements, these values remain within an acceptable range.

Selecting the appropriate distance measure is crucial for achieving optimal clustering quality. The consistent performance of certain measures across different algorithms highlights their robustness and suitability for a wide range of datasets. This analysis underscores the importance of tailoring the distance measure to the specific characteristics of the data and the clustering task at hand, ensuring robust and meaningful clustering

4. CONCLUSION

This research demonstrates the effectiveness of clustering techniques with a hyperparameter approach in mapping Indonesia's primary energy supply. By applying various clustering algorithms and evaluating them with the Davies-Bouldin Index, the study identifies the most suitable methods for analyzing complex and varied energy data. The findings highlight the importance of selecting appropriate distance measures to achieve optimal clustering quality. The insights gained from this analysis can significantly enhance decision-making in energy planning, infrastructure development, and the formulation of sustainable energy policies. This study lays a foundation for further research in detailed and holistic energy data analysis, supporting the efficient management of energy resources in Indonesia.

REFERENCES

[1] A. Sapitri, "The Impact Of Population Factors On Economic Growth Of Bangka Belitung Islands - Indonesia," *J. Ilmu Ekon. Terap.*, vol. 6, no. 2, p. 280, 2021, https://doi.org/10.20473/Jiet.V6i2.28689.
[2] N. W. Handayani, H. Z. Hadibasyir, A. A. Sigit, *Analysis Of The Population Density Correlation With The Temperature Changes In The Semarang Area From 1999 To 2019*, vol. 2035, 2023, https://doi.org/10.2991/978-2-38476-066-4_10.
[3] H. Pratiwi *et al.*, "Sigmoid Activation Function In Selecting The Best Model Of Artificial Neural Networks," *J. Phys. Conf. Ser.*, vol. 1471, no. 1, 2020, https://doi.org/10.1088/1742-6596/1471/1/012010.
[4] N. I. El-Sawalhi, "Support Vector Machine Cost Estimation Model For Road Projects," *J. Civ. Eng. Archit.*, vol. 9, no. 9, pp. 1115–1125, 2015, https://doi.org/10.17265/1934-7359/2015.09.012.
[5] A. Wanto *et al.*"Analysis Of The Accuracy Batch Training Method In Viewing Indonesian Fisheries Cultivation Company Development," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, https://doi.org/10.1088/1742-6596/1255/1/012003.
[6] C. Li, S. Lin, F. Xu, D. Liu, J. Liu, "Short-Term Wind Power Prediction Based On Data Mining Technology And Improved Support Vector Machine Method: A Case Study In Northwest China," *Journal of Cleaner Production*, vol. 205, 2018, https://doi.org/10.1016/J.Jclepro.2018.09.143.
[7] S. Yana, M. Nizar, D. Mulyati, "Biomass waste as a renewable energy in developing bio-based economies in Indonesia: A review," *Renewable and Sustainable Energy Reviews*, vol. 160, p. 112268, 2022, https://doi.org/10.1016/j.rser.2022.112268.
[8] F. Yu, X. Xu, "A Short-Term Load Forecasting Model Of Natural Gas Based On Optimized Genetic Algorithm and Improved Bp Neural Network," *Appl. Energy*, vol. 134, pp. 102–113, 2014, https://doi.org/10.1016/J.Apenergy.2014.07.104.
[9] S. Kumar, V. Kumar-Solanki, S. K. Choudhary, A. Selamat, R. Gonzalez-Crespo, "Comparative Study On Ant Colony Optimization (Aco) and K-Means Clustering Approaches For Jobs Scheduling and Energy Optimization Model In Internet Of Things (Iot)," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 6, no. 1, p. 107, 2020, https://doi.org/10.9781/Ijimai.2020.01.003.
[10] S. M. Ashhar, "Comparison of Deep Learning Convolutional Neural Network (Cnn) Architectures For Ct Lung Cancer Classification," *Int. J. Adv. Technol. Eng. Explor.*, vol. 8, no. 74, pp. 126–134, 2021, https://doi.org/10.19101/Ijatee.2020.S1762126.
[11] X. Ji, "Exploration Of Supply Chain Financing Model And Virtual Economic Risk Control Using The Backpropagation Neural Network," *J. Glob. Inf. Manag.*, vol. 31, no. 9, 2023, https://doi.org/10.4018/Jgim.333605.
[12] S. Yang, "Aero: Design Space Exploration Framework For Resource-Constrained Cnn Mapping On Tile-Based Accelerators," *Ieee J. Emerg. Sel. Top. Circuits Syst.*, vol. 12, no. 2, pp. 508–521, 2022,: https://doi.org/10.1109/Jetcas.2022.3171826.
[13] Y. Cheng, Y. Yu, "K-Means Clustering Algorithm-Based Functional Magnetic Resonance for Evaluation Of Regular Hemodialysis on Brain Function of Patients with End-Stage Renal Disease," *Comput. Math. Methods Med.*, p. 1181030. 2022, https://doi.org/10.1155/2022/1181030.
[14] Y. Ji, "Deep Mining of Mobile Learning Data Based on Multi-Scale Clustering Analysis," *Int. J. Contin. Eng. Educ. Life-Long Learn.*, vol. 33, no. 4, pp. 456–467, 2023, https://doi.org/10.1504/Ijceell.2023.132414.
[15] Y. Wang, F. Sibaii, K. Lee, M. J. Gill, And J. L. Hatch, "Clusa: Clustering-Based Spatial Analysis Framework Through Graph Neural Network For Chronic Kidney Disease Prediction Using Histopathology Images," *Medrxiv*, vol. 1, no. 165, pp. 1–13, 2021, https://www.nature.com/articles/s41598-023-39591-8.

[16] M. Moradi Fard, T. Thonet, E. Gaussier, "Deep K-Means: Jointly Clustering With K-Means and Learning Representations," *Pattern Recognit. Lett.*, vol. 138, no. 2016, pp. 185–192, 2020, https://doi.org/10.1016/J.Patrec.2020.07.028.
[17] S. F. Waris, S. Koteeswaran, "Early Prediction of Heart Conditions by K-Means Consensus Clustering and Convolution Neural Network," *Annals of the Romanian Society for Cell Biology*, vol. 25, no. 3, pp. 6623–6640, 2021, http://annalsofrscb.ro/index.php/journal/article/view/2178.
[18] H. Zhang, "Clustering Mining Method of Large-Scale Network Abnormal Data Based on Selective Collaborative Learning," *J. Comput. Methods Sci. Eng.*, vol. 23, no. 1, pp. 9–21, 2023, https://doi.org/10.3233/Jcm-226537.
[19] E. Bakhshizadeh, "Customer Clustering Based on Factors Of Customer Lifetime Value With Data Mining Technique (Case Study: Software Industry)," *Int. J. Ind. Eng. Prod. Res.*, vol. 33, no. 1, 2022, https://doi.org/10.22068/Ijiepr.33.1.1.
[20] A. J. -Piadlo *et al.*, "Data-Driven Versus A Domain-Led Approach To K-Means Clustering On An Open Heart Failure Dataset," *Int. J. Data Sci. Anal.*, vol. 15, no. 1, pp. 49–66, 2023, https://doi.org/10.1007/S41060-022-00346-9.
[21] D. Novaliendry, "Optimizing Patient Medical Records Grouping Through Data Mining and K-Means Clustering Algorithm: A Case Study at Rsud Mohammad Natsir Solok," *Int. J. Online Biomed. Eng.*, vol. 19, no. 12, pp. 144–155, 2023, https://doi.org/10.3991/Ijoe.V19i12.42147.
[22] V. Duhoon, "Canopy, Cobweb, Dbscan Clustering Data Mining Techniques: Study and Analysis," *Ital. J. Pure Appl. Math.*, vol. 47, pp. 438–448, 2022, https://ijpam.uniud.it/online_issue/202247/29%20Duhoon-Bhardwaj.pdf.
[23] F. Y. Wattimena, "Data Mining Application for The Spread Of Endemic Butterfly Cenderawasih Bay Using The K-Means Clustering Algorithm," *Int. J. Online Biomed. Eng.*, vol. 19, no. 9, pp. 108–121, 2023, https://doi.org/10.3991/Ijoe.V19i09.40907.
[24] S. Basar, M. Ali, G. Ochoa-Ruiz, M. Zareei, A. Waheed, And A. Adnan, "Unsupervised Color Image Segmentation: A Case Of Rgb Histogram Based K-Means Clustering Initialization," *Plos One*, vol. 15, no. 10, pp. 1–21, 2020, https://doi.org/10.1371/Journal.Pone.0240015.
[25] H. Li, "Summary of Clustering Research in Time Series Data Mining," *Dianzi Keji Daxue Xuebao/Journal Univ. Electron. Sci. Technol. China*, vol. 51, no. 3, pp. 416–424, 2022, https://doi.org/10.12178/1001-0548.2022055.
[26] S. T. Ahmed, "Data Mining For Non-Redundant Big Data Using Dynamic Kmean Clustering," *Int. J. Comput. Digit. Syst.*, vol. 14, no. 1, pp. 243–251, 2023, https://doi.org/10.12785/Ijcds/140121.
[27] T. Wahyudi, "Implementation Of Data Mining Using K-Means Clustering Method To Determine Sales Strategy In S&Amp;R Baby Store," *J. Appl. Eng. Technol. Sci.*, vol. 4, no. 1, pp. 93–103, 2022, https://doi.org/10.37385/Jaets.V4i1.913.
[28] N. V Bondarev, "Clustering And Classification Of Red Wines According To Physical-Chemical Properties Using Data Mining Methods," *Russ. J. Gen. Chem.*, vol. 93, no. 9, pp. 2325–2339, 2023, https://doi.org/10.1134/S1070363223090141.
[29] J. H. -Rodríguez, "Convolutional Neural Networks for Multi-Scale Lung Nodule Classification In Ct: Influence Of Hyperparameter Tuning On Performance," *Tem J.*, vol. 11, no. 1, pp. 297–306, 2022, https://doi.org/10.18421/Tem111-37.
[30] K. S. Chong, "Comparison of Naive Bayes and Svm Classification in Grid-Search Hyperparameter Tuned and Non-Hyperparameter Tuned Healthcare Stock Market Sentiment Analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 12, pp. 90–94, 2022, https://doi.org/10.14569/Ijacsa.2022.0131213.
[31] S. Kiziloluk, E. Sert, "Covid-Ccd-Net: Covid-19 And Colon Cancer Diagnosis System With Optimized Cnn Hyperparameters Using Gradient-Based Optimizer," *Med. Biol. Eng. Comput.*, vol. 60, no. 6, pp. 1595–1612, 2022, https://doi.org/10.1007/S11517-022-02553-9.
[32] E. H. Houssein, M. M. Emam, and A. A. Ali, "An Optimized Deep Learning Architecture for Breast Cancer Diagnosis Based on Improved Marine Predators Algorithm," *Neural Comput. Appl.*, vol. 34, no. 20, pp. 18015–18033, 2022, https://doi.org/10.1007/S00521-022-07445-5.
[33] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," *Ieee Access*, vol. 7, pp. 41525–41550, 2019, https://doi.org/10.1109/Access.2019.2895334.
[34] J. M. Ede, R. Beanland, "Adaptive Learning Rate Clipping Stabilizes Learning," *Mach. Learn. Sci. Technol.*, vol. 1, no. 1, 2020, https://doi.org/10.1088/2632-2153/Ab81e2.
[35] M. Ahmad *et al.*, "A Lightweight Convolutional Neural Network Model for Liver Segmentation In Medical Diagnosis," *Computational Intelligence and Neuroscience*, no. 1, p. 7954333, 2022, https://doi.org/10.1155/2022/7954333.
[36] M. Hosseinzadeh, O. Hassan, A. Marwan, and Y. Ghafour, "A Multiple Multilayer Perceptron Neural Network with An Adaptive Learning Algorithm for Thyroid Disease Diagnosis In The Internet Of Medical Things," *J. Supercomput.*, no. 0123456789, 2020, https://doi.org/10.1007/S11227-020-03404-W.
[37] A. D. W. M. Sidik, Z. Akbar, "Analyzing the potential for utilization of new renewable energy to support the electricity system in the cianjur regency region," *Fidelity: Jurnal Teknik Elektro*, vol. 3, no. 3, pp. 46-51, 2021, https://doi.org/10.52005/fidelity.v3i3.66.
[38] M. Bordoloi, S. K. Biswas, "Sentiment Analysis: A Survey on Design Framework, Applications And Future Scopes," *Artificial intelligence review,* vol. 56, no. 11, 2023, https://doi.org/10.1007/S10462-023-10442-2.

[39] C. H. Chan, M. Sun, B. Huang, "Application of Machine Learning For Advanced Material Prediction and Design," *Ecomat*, vol. 4, No. 4, pp. 1–28, 2022, https://doi.org/10.1002/Eom2.12194.
[40] I. Tlemsani, M. A. Mohamed Hashim, R. Matthews, "Portfolio Replication: Islamic Vs Conventional," *J. Islam. Account. Bus. Res.*, vol. 14, no. 1, pp. 1 – 20, 2023, https://doi.org/10.1108/Jiabr-09-2021-0261.
[41] P. Golpour *et al.*, "Comparison Of Support Vector Machine, Naïve Bayes And Logistic Regression For Assessing The Necessity For Coronary Angiography," *Int. J. Environ. Res. Public Health*, vol. 17, no. 18, pp. 1–9, 2020, https://doi.org/10.3390/Ijerph17186449.
[42] S. Sutikno, S. D. Putra, F Wijitrisnanto, M. E. Aminanto, "Detecting Unknown Hardware Trojans in Register Transfer Level Leveraging Verilog Conditional Branching Features," *IEEE Access*, vol. 11, pp. 46073-46083, 2023, https://doi.org/10.1109/ACCESS.2023.3272034.
[43] O. Renn, P. J. Schweizer, "Inclusive governance for energy policy making: Conceptual foundations, applications, and lessons learned," *In The role of public participation in energy transitions*, pp. 39-79, 2020, https://doi.org/10.1016/B978-0-12-819515-4.00003-9.
[44] B. Hunter, "Radiomics-based decision support tool assists radiologists in small lung nodule classification and improves lung cancer early diagnosis," *Br. J. Cancer*, vol. 129, no. 12, pp. 1949–1955, 2023, https://doi.org/10.1038/s41416-023-02480-y.
[45] Y. Guan *et al.*, "Willingness to decrease mammogram frequency among women at low risk for hereditary breast cancer," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, 2019, https://doi.org/10.1038/s41598-019-45967-6.
[46] S. H. Zanakis, A. Solomon, N. Wishart, and S. Dublish, "Multi-attribute decision making: A simulation comparison of select methods," *Eur. J. Oper. Res.*, vol. 107, no. 3, pp. 507–529, 1998, https://doi.org/10.1016/S0377-2217(97)00147-1.
[47] N. Hai, "An improved random drift particle swarm optimization-feed forward backpropagation neural network for high-precision state-of-charge estimation of lithium-ion batteries," *J. Energy Storage*, vol. 73, 2023, https://doi.org/10.1016/j.est.2023.109286.
[48] C. Y. Lee, K. Y. Huang, Y. X. Shen, and Y. C. Lee, "Improved weighted k-nearest neighbor based on PSO for wind power system state recognition," *Energies*, vol. 13, no. 20, 2020, https://doi.org/10.3390/en13205520.
[49] F. Al-Turjman, H. Zahmatkesh, and L. Mostarda, "Quantifying uncertainty in internet of medical things and big-data services using intelligence and deep learning," *IEEE Access*, vol. 7, pp. 115749–115759, 2019, doi: https://doi.org/10.1109/ACCESS.2019.2931637.
[50] K. Fatmawati *et al.*, "Analysis of Promothee II Method in the Selection of the Best Formula for Infants under Three Years," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, https://doi.org/10.1088/17426596/1255/1/012009.

BIOGRAPHY OF AUTHORS

Agus Perdana Windarto the authors were born in Pematangsiantar on August 30 1986. They completed their Masters in Computer Science and Doctoral (Ph.D.) degrees at Putra Indonesia University 'YPTK' Padang in 2014 and 2024. The author has been an active lecturer at STIKOM Tunas Bangsa since 2012, teaching in the Information Systems program. Their research focuses on artificial intelligence. Additionally, the author has served as a reviewer for various nationally accredited journals (SINTA 2 - SINTA 6) and manages a community called "Pemburu Jurnal" at STIKOM Tunas Bangsa. They have won multiple research grant proposals from DIKTI (twice in 2018-2019), DIKTI Community Service Grant (once in 2019), PKM-P Grant (as a student advisor in 2018), and PKM-AI Grant (as a student advisor in 2019). The author is also part of the Relawan Jurnal Indonesia (RJI) community in North Sumatra, the Data Science Indonesia Researchers Association (PDSI), the Forum of Higher Education Communities (FKPT), and is a co-founder of the Yayasan Adwitiya Basurata Inovasi (Yayasan Abivasi) foundation with fellow professors. He can be contacted at email: agus.perdana@amiktunasbangsa.ac.id.

Yerika Puspa Rosanti was born in Salatiga on January 8, 1997. She is currently studying Master Degree of Physics Teaching at Institut Teknologi Bandung. Her research focuses on collaborating physics education with technology, especially development of interactive simulation. She can be reach via e- mail: yerikapuspa@gmail.com.

Mesran the author was born in Medan on August 24, 1978, he completed his master’s degree in Computer Science in 2008 at Universitas Putra Indonesia. Currently, he is actively teaching at STMIK Budi Darma since 2005 as a permanent lecturer in the Informatics Engineering program. He can be contacted at email: mesran.skom.mkom@gmail.com .