



ANALISIS KLASSTER PROVINSI DI INDONESIA BERDASARKAN AKSES DAN INFRASTRUKTUR LISTRIK MENGGUNAKAN ALGORITMA K-MEANS

Risbuwono Heru Cokro ^{1*}, Aryaputra Jagaddatri ², Reyner Archi Sanchia ³,
Alfan Rizaldy Pratama ⁴, Aviolla Terza Damaliana ⁵

^{1,2,3,4,5}Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa Timur

E-mail: 23083010104@student.upnjatim.ac.id ^{1*}, 23083010102@student.upnjatim.ac.id ²,
23083010085@student.upnjatim.ac.id ³, alfan.fasilkom@upnjatim.ac.id ⁴,
aviolla.terza.sada@upnjatim.ac.id ⁵

ABSTRACT

Access to electricity is one of the main indicators of regional development in Indonesia. However, disparities in electrification remain significant among provinces, both in terms of infrastructure capacity and energy production. This study aims to classify Indonesian provinces based on electricity access, generation capacity, and energy output using the K-Means clustering algorithm. The analysis used secondary data from BPS and PLN for the year 2023 and was processed using Python on Google Colab. Prior to clustering, all numerical data were standardized. The optimal number of clusters was determined through the Elbow Method and Silhouette Score, which indicated three clusters ($k = 3$). The results show that Cluster 0 includes highly developed provinces such as West Java and East Java with the highest capacity and energy generation. Cluster 1 consists mostly of eastern and rural provinces with low infrastructure, although electrification rates are generally high. Cluster 2 contains strategic provinces with high access but moderate capacity, including Jakarta and North Sumatra. The spatial and PCA visualizations further highlight the geographic concentration and disparity of electricity infrastructure. These findings are expected to support policymakers in designing more equitable and targeted regional energy development strategies.

Keywords: *electricity access, clustering, K-Means, regional disparity, infrastructure*

ABSTRAK

Akses listrik merupakan salah satu indikator utama pembangunan wilayah di Indonesia. Namun, hingga kini masih terdapat ketimpangan signifikan antar provinsi, baik dari sisi kapasitas infrastruktur maupun produksi energi. Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan tiga indikator utama: persentase rumah tangga berlistrik, kapasitas pembangkit listrik, dan jumlah energi yang dibangkitkan, dengan menggunakan algoritma K-Means Clustering. Data yang digunakan bersifat sekunder dan bersumber dari BPS

Article History

Received: Juni 2025

Reviewed: Juni 2025

Published: Juni 2025

Plagiarism Checker No
235

Prefix DOI :

[10.8734/Koehesi.v1i2.365](https://doi.org/10.8734/Koehesi.v1i2.365)

Copyright : Author
Publish by : Koehesi



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



dan PLN tahun 2023. Seluruh variabel numerik distandarisasi sebelum dilakukan klasterisasi. Penentuan jumlah klaster optimal dilakukan menggunakan metode Elbow dan Silhouette Score, yang menunjukkan jumlah klaster terbaik sebanyak tiga ($k = 3$). Hasil klasterisasi menunjukkan bahwa Klaster 0 mencakup provinsi dengan infrastruktur listrik paling maju, seperti Jawa Barat dan Jawa Timur. Klaster 1 mencakup mayoritas provinsi di Indonesia, khususnya wilayah timur, dengan kapasitas pembangkitan yang rendah meskipun tingkat elektrifikasi cukup tinggi. Sementara itu, Klaster 2 berisi provinsi strategis dengan akses listrik tinggi namun kapasitas menengah, seperti DKI Jakarta dan Sumatera Utara. Visualisasi spasial dan PCA menguatkan sebaran geografis antar klaster. Temuan ini diharapkan dapat menjadi masukan strategis dalam perencanaan kebijakan energi yang lebih adil dan tepat sasaran.

Kata kunci: akses listrik, klasterisasi, K-Means, ketimpangan wilayah, infrastruktur

1. PENDAHULUAN

Ketersediaan listrik merupakan indikator utama dalam pembangunan wilayah, karena listrik tidak hanya menjamin penerangan, tetapi juga menjadi bahan bakar produktivitas ekonomi dan peningkatan kualitas hidup masyarakat. Pemerataan akses listrik ke seluruh provinsi di Indonesia telah lama menjadi perhatian utama pemerintah, sebagai bagian dari visi keadilan sosial dan keberlanjutan pembangunan nasional. Namun di lapangan, terdapat disparitas yang cukup mencolok antar wilayah—mulai dari persentase rumah tangga yang berlistrik, kapasitas pembangkit, hingga total energi yang dikonsumsi atau dibangkitkan. Misalnya, daerah perkotaan di Pulau Jawa dengan pusat-pusat industri besar menunjukkan tingkat konsumsi listrik yang jauh lebih besar dibanding wilayah perdesaan atau di luar pulau utama. Sebaliknya, di wilayah terpencil, meskipun secara teori rasio elektrifikasi sudah tinggi, namun distribusi dan kualitas suplai masih sering bermasalah.

Seiring dengan berkembangnya teknologi informasi dan ketersediaan data terbuka dari PLN dan BPS, penelitian berbasis statistik dan data mining semakin bisa dimanfaatkan untuk membantu pemetaan serta analisis wilayah berdasarkan indikator ketenagalistrikan. Salah satu teknik populer dalam pendekatan ini adalah K-Means clustering, yang efektif dalam mengelompokkan objek wilayah dengan karakteristik serupa berdasarkan sejumlah variabel terukur. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi kelompok provinsi yang memiliki karakteristik ketenagalistrikan yang serupa, menyediakan wawasan berbasis



data yang dapat dimanfaatkan oleh para pembuat kebijakan energi, terutama dalam merancang intervensi berbasis wilayah, serta menyoroti daerah-daerah dengan kebutuhan prioritas, baik dari sisi peningkatan infrastruktur pembangkitan, distribusi, maupun penguatan program kelistrikan rumah tangga. Hasil dari klasterisasi ini diharapkan dapat mendukung penyusunan kebijakan energi yang lebih adaptif, tepat sasaran, dan berkeadilan antar wilayah.

2. METODE PENELITIAN

2. 1 Data dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder bersifat kuantitatif. Data sekunder adalah data yang diperoleh secara tidak langsung melalui pihak lain dan telah dikumpulkan serta dipublikasikan untuk keperluan tertentu. Dalam konteks penelitian ini, data diperoleh dari publikasi resmi lembaga pemerintah terkait yang memuat informasi statistik ketenagalistrikan rumah tangga di Indonesia. Adapun data yang digunakan mencakup:

1. Persentase rumah tangga dengan akses listrik per provinsi
2. Kapasitas pembangkit listrik per provinsi (dalam MW)
3. Jumlah energi listrik yang dibangkitkan per provinsi (dalam GWh)
4. Wilayah: Seluruh provinsi di Indonesia
5. Periode: Tahun 2023

Sumber utama data dalam penelitian ini adalah Badan Pusat Statistik (BPS), lembaga resmi pemerintah yang menyediakan data statistik nasional, termasuk data tentang akses rumah tangga terhadap sumber penerangan utama dari listrik. Data kapasitas pembangkit dan energi yang dibangkitkan berasal dari publikasi resmi lembaga penyedia data ketenagalistrikan. Data telah melalui proses verifikasi dan validasi untuk memastikan keakuratannya dalam merepresentasikan kondisi ketenagalistrikan di setiap provinsi.

2.2 Tahapan Analisis

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis kluster, yaitu K-Means Clustering, untuk melakukan segmentasi provinsi-provinsi di Indonesia berdasarkan karakteristik kelistrikan rumah tangga pada tahun 2023. Metode K-Means dipilih karena efisien dalam mengelompokkan data numerik ke dalam sejumlah kluster berdasarkan kesamaan karakteristik antar data.



Menurut Siregar & Nursal (2020), K-Means merupakan salah satu algoritma partisi paling umum karena kemampuannya dalam membagi data ke dalam k kelompok berdasarkan jarak Euclidean sebagai fungsi objektif. Selain itu, Wardhani et al. (2021) menyatakan bahwa K-Means sangat efektif untuk analisis segmentasi regional berbasis data multivariat, karena hasil klusterisasinya mudah divisualisasikan dan diinterpretasikan. Dalam penelitian ini, proses analisis dilakukan menggunakan Google Colab dengan bahasa pemrograman Python. Tahapan analisis meliputi:

1. Pra-pemrosesan data, termasuk normalisasi nilai numerik
2. Penentuan jumlah kluster optimal menggunakan metode evaluasi internal Elbow Method, dengan mengamati perubahan nilai within-cluster sum of squares (WCSS) untuk berbagai nilai k
3. Penerapan algoritma K-Means untuk klusterisasi provinsi
4. Visualisasi hasil klusterisasi menggunakan scatter plot dan peta wilayah provinsi Indonesia

Menurut Herdiana et al. (2025), Elbow Method efektif digunakan untuk mengidentifikasi jumlah kluster optimal melalui titik “siku” pada grafik WCSS. Berdasarkan hasil evaluasi dari metode tersebut, ditetapkan bahwa jumlah kluster optimal dalam penelitian ini adalah tiga kluster ($k = 3$).

3. HASIL DAN PEMBAHASAN

3.1 Statistika Deskriptif

Analisis awal dilakukan menggunakan statistik deskriptif untuk memahami sebaran data antar provinsi. Statistik yang digunakan mencakup nilai minimum, maksimum, rata-rata, dan standar deviasi dari masing-masing variabel: Persentase Rumah Tangga Berlistrik, Kapasitas Pembangkit (MW), dan Energi Dibangkitkan (GWh). Hasilnya disajikan pada Tabel 1 berikut:



Tabel 1. Statistik Deskriptif

Variabel	Min	Max	Rata-rata	Standar Deviasi
Persentase Listrik (%)	14,53	100,00	93,56	17,81
Kapasitas Pembangkit (MW)	0,78	12.121,30	1.933,18	3.152,22
Energi Dibangkitkan (GWh)	3,48	68.376,00	8.575,15	15.830,55

Berdasarkan Tabel 1, diketahui bahwa rata-rata persentase rumah tangga yang memiliki akses listrik di Indonesia adalah 93,56%, namun terdapat ketimpangan cukup besar antar wilayah, terlihat dari nilai minimum 14,53% dan standar deviasi 17,81. Demikian pula, kapasitas pembangkit listrik sangat bervariasi antar provinsi, dengan nilai maksimum mencapai lebih dari 12.000 MW dan rata-rata sebesar 1.933 MW. Hal yang sama juga terjadi pada variabel energi dibangkitkan, yang memiliki rata-rata 8.575 GWh namun dengan rentang sangat luas, dari hanya 3,48 GWh hingga lebih dari 68.000 GWh. Kondisi ini menunjukkan ketimpangan infrastruktur kelistrikan yang signifikan antar provinsi di Indonesia.

3.2 Hasil Klasterisasi Menggunakan K-Means

Proses klasterisasi dilakukan menggunakan algoritma K-Means. Untuk menentukan jumlah klaster optimal, digunakan pendekatan, yaitu Elbow Method. Menunjukkan bahwa jumlah klaster optimal adalah tiga ($k = 3$). Hasil klasterisasi menghasilkan tiga kelompok provinsi dengan karakteristik yang berbeda. Rata-rata dari tiap variabel untuk masing-masing klaster ditampilkan pada Tabel 2 berikut :

Tabel 2. Rata-rata Tiap Variabel Berdasarkan Klaster

Variabel	Klaster 0 (n= 4)	Klaster 1 (n = 31)	Klaster 2 (n = 3)
Persentase Listrik (%)	99,94	92,14	99,74
Kapasitas Pembangkit (MW)	10.363,05	623,93	4.222,22
Energi Dibangkitkan (GWh)	51.217,68	2.362,45	15.916,34

Proses klasterisasi dilakukan menggunakan algoritma K-Means. Untuk menentukan jumlah klaster optimal, digunakan pendekatan, yaitu Elbow Method. Menunjukkan bahwa jumlah klaster optimal adalah tiga ($k = 3$) Klaster 0 terdiri dari 4 provinsi yang memiliki infrastruktur kelistrikan paling maju, ditandai dengan kapasitas pembangkit dan produksi energi yang sangat tinggi, serta akses listrik yang nyaris menyeluruh. Klaster ini mencakup



Jawa Barat, Jawa Tengah, Jawa Timur, dan Banten, yang merupakan pusat industri dan ekonomi nasional.

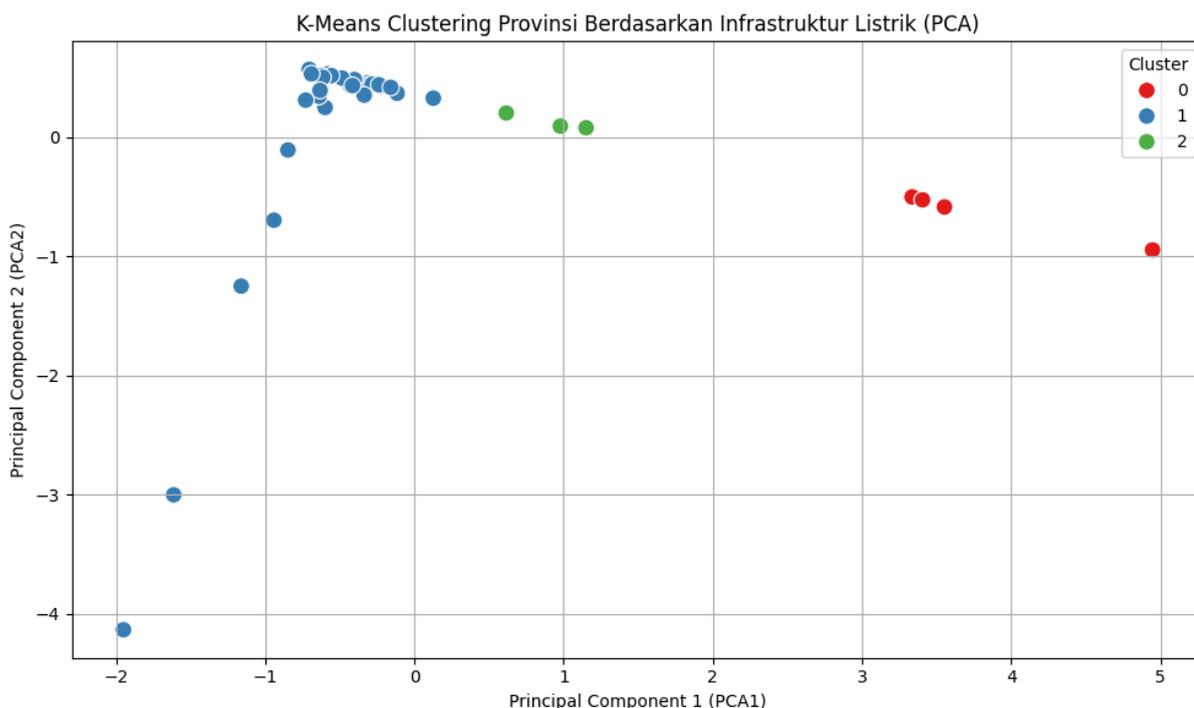
Klaster 1 berisi 31 provinsi, sebagian besar berada di wilayah Indonesia timur, Kalimantan, serta beberapa di Sumatera. Klaster ini ditandai dengan kapasitas pembangkit dan energi yang dibangkitkan relatif rendah, meskipun sebagian besar sudah memiliki persentase elektrifikasi yang tinggi. Hal ini mencerminkan bahwa banyak provinsi telah mendapat sambungan listrik, tetapi masih bergantung pada sumber energi yang terbatas atau kecil skalanya.

Klaster 2 terdiri dari 3 provinsi, yakni DKI Jakarta, Sumatera Utara, dan Sumatera Selatan, yang memiliki akses listrik tinggi dan infrastruktur pembangkit yang tergolong sedang. Provinsi-provinsi ini memiliki peran strategis dan pusat distribusi listrik di wilayah masing-masing.

3.3 Visualisasi Hasil Klastering

Untuk mempermudah interpretasi, dilakukan reduksi dimensi menggunakan Principal Component Analysis (PCA) sehingga hasil klasterisasi dapat divisualisasikan dalam bidang dua dimensi.

Gambar 1. Visualisasi PCA



Grafik scatter plot menunjukkan bahwa masing-masing klaster membentuk kelompok yang cukup jelas. Klaster 0 mengelompok di sisi kanan atas (mewakili provinsi dengan

pembangkitan dan kapasitas besar), sementara Klaster 1 lebih tersebar dengan konsentrasi di bagian bawah grafik. Klaster 2 berada di antara keduanya, menunjukkan posisi transisi antara daerah maju dan daerah yang masih berkembang dalam hal infrastruktur kelistrikan.

3.4 Visualisasi Clustering

Selanjutnya, hasil klasterisasi divisualisasikan dalam bentuk peta spasial interaktif menggunakan library Folium, sebagaimana ditunjukkan pada Gambar 2. Peta ini menyajikan sebaran provinsi di Indonesia berdasarkan hasil klaster KMeans, di mana warna marker pada peta menunjukkan klaster masing-masing provinsi. Visualisasi ini memberikan pemahaman geografis yang lebih kuat mengenai sebaran infrastruktur dan akses kelistrikan di Indonesia.

Gambar 2. Visualisasi Clustering



Pada peta, terlihat bahwa klaster dengan infrastruktur listrik sangat tinggi (ditandai dengan warna biru) terkonsentrasi di pulau Jawa, terutama pada provinsi-provinsi dengan populasi dan aktivitas industri besar. Sementara itu, klaster dengan infrastruktur rendah (ditandai dengan warna hijau) tersebar luas di wilayah timur Indonesia seperti Nusa Tenggara, Maluku, dan Papua. Sedangkan klaster dengan akses listrik tinggi namun infrastruktur sedang (warna oranye) muncul di beberapa pusat kota besar di luar Jawa seperti DKI Jakarta dan Sumatera Utara.

KESIMPULAN

Penelitian ini berhasil mengelompokkan 38 provinsi di Indonesia ke dalam tiga klaster berdasarkan indikator kelistrikan, yaitu: persentase rumah tangga berlistrik, kapasitas pembangkit listrik (MW), dan energi listrik yang dibangkitkan (GWh). Proses klasterisasi



menggunakan metode KMeans yang telah dioptimalkan melalui pendekatan Elbow Method dan Silhouette Score, dengan hasil jumlah kluster optimal sebanyak tiga ($k = 3$).

Hasil klasterisasi menunjukkan adanya ketimpangan signifikan dalam infrastruktur kelistrikan antar provinsi. Kluster 0 mencakup provinsi-provinsi maju seperti Jawa Barat dan Jawa Timur yang memiliki infrastruktur listrik sangat tinggi. Kluster 1 meliputi sebagian besar wilayah Indonesia, dengan karakteristik kapasitas dan produksi energi yang masih rendah, meskipun secara akses sudah cukup baik. Sementara itu, Kluster 2 terdiri dari provinsi-provinsi dengan akses listrik tinggi namun infrastruktur yang tergolong sedang.

Visualisasi hasil klasterisasi dalam bentuk scatter plot PCA dan peta spasial interaktif memperjelas distribusi geografis dan pola ketimpangan antar wilayah. Temuan ini diharapkan dapat menjadi referensi bagi pengambil kebijakan dalam merumuskan strategi pembangunan dan pemerataan infrastruktur kelistrikan, khususnya untuk memperkuat wilayah-wilayah yang tergolong dalam kluster dengan kapasitas rendah.

DAFTAR PUSTAKA

- [1] Wardhani, D. K., Astuti, D. M., & Huda, M. (2021). Implementasi Metode K-Means untuk Segmentasi Wilayah Berdasarkan Indikator Sosial Ekonomi. *Jurnal Sains dan Informatika*, 7(2), 55-62. <https://doi.org/10.21009/jsi.007.2.05>
- [2] Siregar, M. F., & Nursal, D. (2020). Segmentasi Wilayah Menggunakan K-Means Clustering untuk Perencanaan Pembangunan Daerah. *Jurnal Ilmiah Informatika KOMPUTA*, 9(2), 107-116. <https://doi.org/10.37676/komputa.v9i2.347>
- [3] A. R. Pratama, B. S. B. Dewantara, D. M. Sari, dan D. Pramadihanto, Improvement of DBSCAN Algorithm Involving Automatic Parameters Estimation and Curvature Analysis in 3D Point Cloud of Piled Pipe, *Journal of Image and Graphics*, vol. 12, no. 2, pp. 175-185, 2024. <https://www.joig.net/2024/JOIG-V12N2-176.pdf>
- [4] Herdiana, I., Kamal, M. A., Triyani, M. N. E., & Renny. (2025). A More Precise Elbow Method for Optimum K-means Clustering. arXiv preprint arXiv:2502.00851.
- [5] S. Mulyadi, F. Insani, S. Agustian, dan L. Afriyanti, Grouping Electricity Distribution Data Using The Mini Batch K-Means Clustering Algorithm, *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 1051-1062, 2024. [Online]. Tersedia: <https://doi.org/10.57152/malcom.v4i3.1425>
- [6] R. F. Utari, F. Insani, S. Agustian, dan L. Afriyanti, Clustering Electricity Distribution Data Using the Mean Shift Algorithm, *MALCOM: Indonesian Journal of Machine Learning and*



Computer Science, vol. 4, no. 3, pp. 1015-1023, 2024. [Online]. Tersedia: <https://doi.org/10.57152/malcom.v4i3.1428>

[7] R. A. Faujia, E. S. Setianingsih, dan H. Pratiwi, Analisis Klaster K-Means dan Agglomerative Nesting Pada Indikator Stunting Balita Di Indonesia, Seminar Nasional Official Statistics, vol. 2022, pp. 1249-1251, 2022.

[8] N. Nugroho dan F. D. Adhinata, Penggunaan Metode K-Means dan K-Means++ Sebagai Clustering Data Covid-19 di Pulau Jawa, TEKNIKA, vol. 11, no. 3, pp. 170-179, 2022. [Online]. Tersedia: <https://doi.org/10.34148/teknika.v11i3.502>

[9] B. Olanda dan D. Susilo, Desain dan Rancang Instalasi Listrik Sederhana Skala Rumah Tangga, Jurnal ELECTRA: Electrical Engineering Articles, vol. 1, no. 2, pp. 7-12, 2021.

[10] S. A. Perdana, S. F. Florentin, dan A. Santoso, Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfragift, Sebatik, vol. 26, no. 2, pp. 446-448, 2022. [Online]. Tersedia: <https://doi.org/10.46984/sebatik.v26i2.2134>

[11] A. R. Simbolon, Kebijakan Kendaraan Listrik dalam Perspektif Pasar dan Infrastruktur: Studi Reviu Komparasi Bilateral Korea Selatan dan Indonesia, Jurnal Penelitian Transportasi Darat, vol. 24, no. 2, pp. 83-91, 2022. <https://doi.org/10.25104/jptd.v24i2.1943>

[12] N. Selle, N. Yudistira, dan C. Dewi, Perbandingan Prediksi Penggunaan Listrik dengan Menggunakan Metode Long Short Term Memory (LSTM) dan Recurrent Neural Network (RNN), Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 9, no. 1, pp. 155-162, 2022. <https://doi.org/10.25126/jtiik.202295585>

[13] J. D. Sari, N. A. Sufiawan, B. Rizky, dan Weriantoni, Analisis Ketimpangan Pendapatan Terhadap Indeks Pembangunan Manusia Melalui Akses Listrik di Indonesia, Equilibrium, vol. 11, no. 2, pp. 35-49, 2022.

[14] N. A. Adistia, R. A. Nurdiansyah, J. Fariko, Vincent, dan J. W. Simatupang, Potensi Energi Panas Bumi, Angin, dan Biomassa Menjadi Energi Listrik di Indonesia, TESLA, vol. 22, no. 2, 2020.

[15] Bahauddin, A., Fatmawati, A., & Permata Sari, F. (2021). Analisis clustering provinsi di Indonesia berdasarkan tingkat kemiskinan menggunakan algoritma K-Means. MISI: Jurnal Manajemen Informatika & Sistem Informasi, 4(1), 1-9. <https://doi.org/10.36595/misi.v4i1.216>