



## Pengelompokan Provinsi Di Indonesia Berdasarkan *Index* Kesehatan Masyarakat Menggunakan Algoritma *Partitioning Around Medoids* (PAM) Dan Metode *Davies Bouldin Index* (DBI)

Dedi Setiawan<sup>1</sup>, Agung Nilogiri<sup>2</sup>, Moh. Dasuki<sup>3</sup>

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember<sup>1,2,3</sup>

Email: [sdedi900@gmail.com](mailto:sdedi900@gmail.com)<sup>1\*</sup>, [agungnilogiri@unmuhjember.ac.id](mailto:agungnilogiri@unmuhjember.ac.id)<sup>2</sup>, [moh.dasuki22@unmuhjember.ac.id](mailto:moh.dasuki22@unmuhjember.ac.id)<sup>3</sup>

### ABSTRAK

Rendahnya akses terhadap air minum dan sanitasi yang layak di Indonesia disebabkan oleh kesenjangan dan ketimpangan antara Indonesia Bagian Barat dan Timur. Sebagai contoh, di Provinsi Papua, sekitar separuh dari penduduk memiliki akses air minum yang kurang memadai dan dua dari tiga rumah tangga memiliki akses sanitasi yang buruk. Dalam hal tersebut, diperlukan upaya untuk mengetahui tingkat sanitasi suatu wilayah atau provinsi guna mengatasi ketimpangan kesehatan lingkungan. Provinsi yang akan dijadikan prioritas dapat diketahui dengan melakukan pengelompokan, satu diantaranya adalah menggunakan algoritma *Partitioning Around Medoids* dengan metode *Davies Bouldin Index Optimization* untuk menghitung *cluster* optimal dalam rangka menentukan *cluster* yang paling baik. Hasil dari penerapan metode *Partitioning Around Medoids* dalam mengelompokkan provinsi di Indonesia pada indikator kesehatan lingkungan diperoleh *cluster* optimum dengan hasil 5 *cluster* berdasarkan indeks nilai *Davies Bouldin* sebesar 1,3473 dengan skenario *cluster* 2 sampai dengan 10 *cluster*.

**Kata Kunci:** Sanitasi, *Cluster*, *Partitioning Around Medoids*, *Davies Bouldin Index*, Optimal

### ABSTRACT

*Low access to drinking water and proper sanitation in Indonesia is caused by gaps and disparities between Western and Eastern Indonesia. For example, in Papua Province, about half of the population has inadequate access to drinking water and two out of three households have poor access to sanitation. In this case, efforts are needed to determine the sanitation level of a region or province in order to address environmental health disparities. Provinces that will be prioritized can be identified by grouping, one of which is using the Partitioning Around Medoids algorithm with the Davies Bouldin Index Optimization method to calculate optimal clusters in order to determine the best clusters. The results of applying the Partitioning Around Medoids method in classifying provinces in Indonesia on environmental health indicators obtained optimum clusters with 5 cluster results based on the Davies Bouldin index score of 1.3473 with a cluster scenario of 2 to 10 clusters*

**Keywords:** Sanitation, *Cluster*, *Partitioning Around Medoids*, *Davies Bouldin Index*, Optimum

## 1. PENDAHULUAN

Sanitasi adalah upaya untuk membina dan menciptakan suatu situasi yang baik dibidang kesehatan, berfokus pada gaya hidup bersih dan sehat, terutama kesehatan masyarakat. Sanitasi berkaitan dengan kesehatan lingkungan yang memiliki manfaat mengurangi dampak dari malnutrisi, meningkatkan taraf khususnya kesehatan masyarakat serta menjamin ketersediaan air bersih dan terbebas dari pencemaran limbah rumah tangga dan industri. Sanitasi layak, air minum layak, dan rumah layak huni merupakan bagian dari kebutuhan dasar manusia (Lubis dkk., 2019).

Perumahan dan permukiman yang layak sama halnya dengan kesehatan serta pendidikan yang kebutuhannya dijamin oleh konstitusi. Upaya ini semakin ditingkatkan setiap tahunnya. Komitmen pemerintah untuk memenuhi kebutuhan yang mendasar tadi juga berbanding lurus dengan agenda global dengan tujuan pembangunan berkelanjutan pada tahun 2030 (*Sustainable Development Goals*) atau disebut SDGs. Tujuan utamanya pada sektor lingkungan hidup, yaitu akses air minum, akses sanitasi, dan akses rumah layak huni dengan dipastikannya warga dapat mengakses sanitasi serta air bersih secara merata (Lubis dkk., 2019).

Rendahnya akses terhadap air minum dan sanitasi yang layak di Indonesia disebabkan oleh kesenjangan dan ketimpangan antara Indonesia Bagian Barat dan Timur. Sebagai contoh, di Provinsi Papua, sekitar separuh dari penduduk memiliki akses air minum yang kurang memadai dan dua dari

tiga rumah tangga memiliki akses sanitasi yang buruk. Sementara itu, hampir semua rumah tangga di Jakarta memiliki akses terhadap air minum yang layak. Mempertimbangkan kesenjangan dan ketidasetaraan tersebut, diperlukan kajian yang mendalam terhadap indikator kesehatan lingkungan yang berasal dari variabel akses air minum yang layak, sanitasi yang layak dan perumahan yang layak huni. Kemudian diperlukan penelitian lebih lanjut terhadap ketiga variabel tersebut. Dengan begitu, dapat ditentukan provinsi mana yang perlu mendapat perhatian dalam rangka peningkatan indikator kesehatan lingkungan. Oleh karena itu, sangat penting memerhatikan pengelompokan provinsi di Indonesia berdasarkan indikator kesehatan lingkungan (Mayasari, 2020).

Algoritma *Partitioning Around Medoids* berkaitan dengan *K-Means*, tetapi ada perbedaan besar antara kedua algoritma antara kedua algoritma yaitu pada *K-Means cluster* diwakili oleh pusat *cluster*, sedangkan di *K-Medoids cluster* diwakili oleh obyek yang paling dekat dengan pusat *cluster*. *Medoid* merupakan objek yang letaknya terpusat di dalam suatu *cluster* sehingga kuat terhadap *outlier*. *Cluster* dibangun dengan menghitung kedekatan yang dimiliki antara *medoids* dengan objek non *medoids*. *Partitioning Around Medoids* (PAM) lebih tahan terhadap *noise* atau *outlier* daripada algoritma *K-Means*. Ini disebabkan *medoids* kurang rentan terhadap *outlier* dan nilai ekstrem lainnya daripada sebuah *mean*. Dari sekian banyak metode yang dapat menggunakan metode *Davies Bouldin Index* (DBI). Pengujian metode DBI ini dengan melakukan beberapa perhitungan pada beberapa *cluster* yang sudah ditentukan, kemudian hasil dari nilai DBI beberapa *cluster* yang telah didapat, nilai DBI yang terkecil merupakan nilai yang optimal (Abdurrahman dkk, 2021).

## 2. KAJIAN PUSTAKA

Pada bagian ini merupakan bagian yang membahas teori dasar dari penelitian serta penelitian terdahulu sebelumnya yang relevan yang dianggap penting untuk dikemukakan.

### A. Data Mining

*Data mining* merupakan proses menemukan hubungan, pola, dan tren yang memiliki makna dengan cara memilih data dalam jumlah besar yang nantinya akan disimpan dalam repositori. *Data mining* juga berkaitan dengan bidang ilmu yang bervariasi, seperti sistem basis data, statistik dan komputasi tingkat tinggi. *Data Mining* menggambarkan sebuah pengumpulan teknik-teknik yang bertujuan untuk menemukan pola-pola yang tidak diketahui pada data yang tidak diketahui pada data yang telah dikumpulkan (Han, 2006).

### B. Clustering

Mengelompokkan data yang mempertimbangkan untuk menggali persamaan pada data serta menemposisikan data yang mirip dalam beberapa kelompok. *Clustering* membagi *dataset* kedalam beberapa kelompok yang mana memiliki kesamaan satu dengan yang lain daripada kelompok yang lain. Tujuan *clustering* adalah meminimalisir jarak di dalam *cluster* serta memaksimalkan jarak antar *cluster*. *Clustering* adalah proses mengelompokkan, mengamati, atau mengelompokkan kelas-kelas yang mempunyai obyek yang sama. Oleh karena itu, pengertian *clustering* adalah sekumpulan dari semua kelompok. Perbedaan antara *clustering* dan klasifikasi adalah tidak ada variabel target saat mengelompokkan data pada proses *clustering* (Abdillah, 2016).

### C. *Partitioning Around Medoids* (PAM)

Algoritma *Partitioning Around Medoids* (PAM), biasa disebut dengan *K-Medoids*, adalah algoritma pengelompokan yang terkait dengan algoritma *K-Means* yang dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw pada tahun 1987. Algoritma ini sangat mirip dengan *K-Means*, terutama karena kedua algoritma dipartisi. Artinya, kedua algoritma ini menguraikan kumpulan data menjadi beberapa kelompok, dan kedua algoritma tersebut mencoba meminimalkan kesalahan. Namun, algoritma *Partitioning Around Medoids* (PAM) bekerja dengan *medoids*, entitas dari kumpulan data yang mewakili kelompok yang akan dimasukkan. Perbedaannya adalah algoritma k-

*medoids* menggunakan objek sebagai *medoid* pada pusat setiap *cluster*, sedangkan *K-Means* menggunakan *mean* atau rata-rata sebagai pusat *cluster* (Kaur & Singh, 2014).

Algoritma *K-Medoids* mempunyai keunggulan dalam mengatasi kelemahan algoritma *K-Means* yakni sensitif terhadap *noise / outlier*, serta obyek dengan nilai yang besar sehingga memungkinkan untuk menyimpang dari distribusi data, tapi keunggulan lainnya adalah hasil urutan proses *clustering* tidak bergantung pada urutan dataset yang dimasukkan. Algoritma yang digunakan dalam metode ini yaitu *clustering partitioning* untuk mengelompokkan kumpulan  $n$  objek menjadi sejumlah  $k$  *cluster*. Obyek yang digunakan dalam algoritma ini merupakan kumpulan objek untuk mewakili sebuah *cluster*. Dimana *medoids* adalah objek yang dipilih untuk mewakili *cluster*. *Cluster* yang dibentuk dengan menghitung jarak antara objek *Medoids* dan *Non-Medoids* (Pramesti, 2017).

Langkah-langkah dari algoritma PAM (Agarwal, 2014), yaitu:

1. Menetapkan terlebih dahulu  $k$  objek dalam sekumpulan  $n$  objek sebagai *medoids* dengan memilihnya secara acak
2. Kemudian secara acak memilih  $o_{random}$ : sebuah obyek non-*medoids*.
3. Kemudian menemposisikan obyek non-*medoids* dalam *cluster* yang dekat dengan *medoids*.
4. Menghitung total *cost*,  $S$  (selisih), dari pertukaran *medoids*  $o_j$  dengan  $o_{random}$ . Total nilai ditunjukkan dalam persamaan:

$$\sum \sqrt{\sum_{k=1}^n} (x_k - y_k)^2 \quad (1)$$

Dengan:

- $n$  = jumlah suatu data
- $k$  = indeks data
- $x_k$  = nilai atribut ke- $k$  dari  $x$
- $y_k$  = nilai atribut ke- $k$  dari  $y$

Nilai  $S$  dinyatakan dalam persamaan:

$$S = \text{Total Cost baru} - \text{Total Cost Lama} \quad (2)$$

Dengan:

- $S$  = Selisih
- Total Cost Baru* = jumlah *cost* non-*medoids*
- Total Cost Lama* = jumlah *cost medoids*

5. Jika  $S < 0$ , maka tukar  $o_j$  dengan  $o_{random}$  untuk membentuk sekumpulan  $k$  objek baru sebagai *medoids*. Ulangi langkah 2 sehingga hasil dari langkah 5 tidak ada perubahan

#### D. Metode *Davies Bouldin Index*

Merupakan salah satu metode validasi *cluster* yang digagas oleh D.L.Davies. Metode ini merupakan fungsi rasio dari jumlah persebaran dalam *cluster* berguna sebagai pembatas antar *cluster*. Pada penelitian ini, DBI sebagai validasi data pada setiap *cluster*. Berikut ini merupakan cara menghitung DBI (Sujacka, 2019), yaitu:

1. Menghitung *Sum of Square Within cluster* (SSW). Menggunakan persamaan guna memahami matriks kohesi dalam sebuah *cluster* ke- $i$  menggunakan rumus:

$$SSW_i = \frac{1}{m_i} \sum_j = im_i d(x_j C_j) \quad (3)$$

Keterangan:

- $m_i$  = jumlah data dalam *cluster* ke- $i$

$c_i$  = disimbolkan sebagai *centroid cluster* ke- $i$   
 $d(x_j C_j)$  = merupakan jarak setiap data terhadap *centroid*

- Mengitung *Sum of Square Between cluster* (SSB). Persamaan yang dapat digunakan untuk mengetahui pemisah antar *cluster* yang dihitung menggunakan persamaan dibawah:

$$SSB_{i,j} = d(c_i, c_j) \quad (4)$$

- Menghitung jumlah rasio. Setelah nilai SSW dan SSB didapat, maka selanjutnya melakukan pengukuran pada rasio ( $R_{ij}$ ) guna mengetahui nilai perbandingan antara *cluster* ke- $i$  dan *cluster* ke- $j$ . persamaan untuk menghitung nilai rasio:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

- Menghitung nilai *Davies Bouldin Index* (DBI). Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *Davies Bouldin Indeks* (DBI) dapat menggunakan persamaan dibawah:

$$DBI = \frac{1}{k} \sum_i = 1k \max_{i \neq j} (R_{ij}) \quad (6)$$

Keterangan:

$k$  = total *cluster* yang ditentukan  
 $R_{ij}$  = rasio *cluster* ke- $i$  dan *cluster* ke- $j$

### 3. METODE PENELITIAN

#### A. Pengumpulan Data

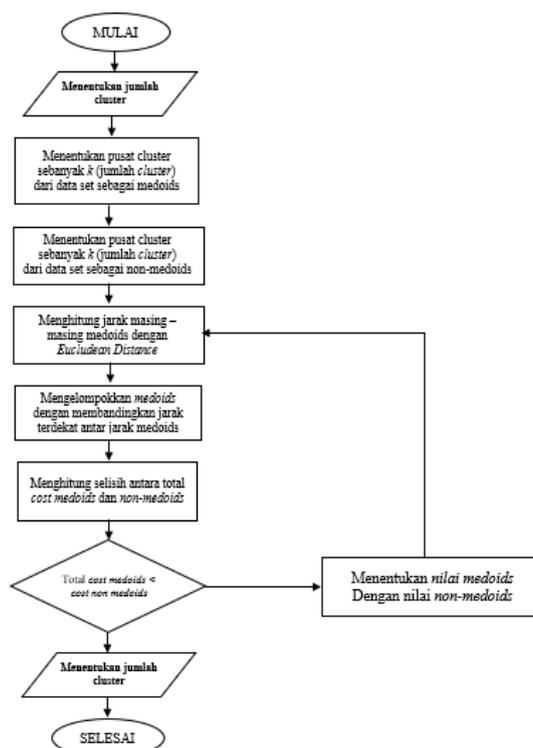
Menggunakan data set sebanyak 34 Provinsi di Indonesia berdasar pada prosentase rumah tangga yang memiliki akses terhadap sanitasi layak, sumber air minum layak, dan rumah layak huni tahun 2020-2021. Pada data tersebut memiliki enam atribut sanitasi layak (SL), air minum layak (AML), rumah layak huni (RLH). Hal ini dapat dilakukan dengan melakukan proses *clustering* data menggunakan algoritma *Partitioning Around Medoids* (PAM) dan menggunakan metode *Davies Bouldin Index* (DBI) sebagai metode optimasi *cluster* untuk menentukan jumlah *cluster* optimum yang akan digunakan pada proses *clustering* (Triyanto, 2015).

Tabel 1. Data Provinsi di Indonesia

Provinsi	SL 2020	SL 2021	AML 2020	...	RLH 20214
Jambi	77,82	80,36	78,86	...	62,54
Belitung	92,58	92,24	75,06	...	27,6
Sulut	85,49	84,85	90,31	...	69,5
Bengkulu	78,1	79,81	62,47	...	51,36
Jatim	80,98	80,97	95,56	...	66,93
Jakarta	93,04	95,17	99,84	...	40
Sulteng	82,38	85,62	92,49	...	70,45
Aceh	77,06	77,55	87,66	...	64,43
Lampung	78,81	83,89	74,97	...	61,56
...	...	...	...	...	...
Papua B	78,81	77,89	79,56	...	57,9

## B. Metode Analisis Data

Dalam penelitian ini menggunakan metode analisis data *Partitioning Around Medoids* (PAM). Dimana *K-Medoids* merupakan teknik *clustering*, yaitu teknik untuk mengelompokkan sekumpulan objek sebagai perwakilan pusat (*medoid*) dari setiap *cluster*. Dalam penelitian ini menerapkan algoritma *Partitioning Around Medoids* guna mengelompokkan data sekunder sebanyak 34 Provinsi di Indonesia berdasarkan prosentase rumah tangga yang dapat mengakses sanitasi yang layak, air minum yang layak. Dalam mengelompokkan tersebut menggunakan teknik *Davies Bouldin Indeks* sebagai menentukan *cluster* optimal. Metode penelitian yang dilakukan dengan menggunakan algoritma *clustering Partitioning Around Medoids* (PAM). *Clustering* disertai dengan performa menggunakan *Davies Bouldin Index* melalui beberapa tahapan yaitu studi awal, pengumpulan data, proses *clustering* dalam menentukan *cluster* terbaik pada kasus sanitasi layak, air minum layak, dan rumah layak. Algoritma *Partitioning Around Medoids* (PAM) memiliki tahapan alur yang pada umumnya tahapan pada *flowchart* adalah sebagai berikut:



Gambar 1. Flowchart Algoritma PAM

## 4. HASIL DAN PEMBAHASAN

Pada bab ini membahas mengenai hasil yang diperoleh dari pengujian. Data yang telah diperoleh akan diolah menggunakan metode *Partitioning Around Medoids*. Pengujian hasil kinerja dari metode *Partitioning Around Medoids* menggunakan indeks pengukuran kinerja *Davies Bouldin* dalam menentukan *cluster* optimum. Implementasi metode *Davies Bouldin Index* (DBI) Pada Algoritma *Partitioning Around Medoids* (PAM). Data yang digunakan sebanyak 34 Provinsi di Indonesia berdasarkan sanitasi layak, air minum layak, dan rumah layak huni tahun 2020-2021 yang diambil dari [Badan Pusat Statistik \(bps.go.id\)](https://bps.go.id).

### A. Penentuan Cluster Optimum

Setelah melakukan proses *cluster* menggunakan algoritma *Partitioning Around Medoids* (PAM), kemudian dilakukan proses metode *Davies Bouldin-Index* untuk menentukan *cluster optimum* dalam penentuan *cluster* terbaik. Berikut tabel 2 merupakan hasil metode *Davies Bouldin-Index*:

Tabel 2. Perhitungan *Davies Bouldin-Index*

<i>Cluster</i>	<i>Silhouette Coefficient</i>
2	1,6837
3	2,5181
4	1,8873
5	1,3473
6	4,9553
7	2,2941
8	2,1649
9	2,4721
10	2,2973

Berdasarkan hasil pengujian yang telah dilakukan, telah didapatkan *cluster optimum* untuk pengelompokan provinsi di Indonesia berdasarkan *index* kesehatan masyarakat menggunakan algoritma *partitioning around medoids* (PAM) dan metode *Davies Bouldin Index* yaitu pada *cluster* 5.

## 5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka didapatkan kesimpulan bahwa dengan menggunakan metode *Partitioning Around Medoids* hasil dari proses *clustering* diuji dengan menggunakan teknik *Davies Bouldin Indeks* yang berfungsi untuk mengetahui jumlah *cluster optimum* yang terbentuk dari pengelompokan provinsi di Indonesia pada indikator kesehatan lingkungan. Hasil dari penerapan metode *Partitioning Around Medoids* dalam mengelompokkan provinsi di Indonesia pada indikator kesehatan lingkungan diperoleh *cluster optimum* dengan hasil 5 *cluster* berdasarkan indeks nilai *Davies Bouldin* sebesar 1,3473 dengan skenario *cluster* 2 sampai dengan 10 *cluster*.

Hasil dari pengelompokan dari 5 *cluster* pada *cluster* 1 terdapat 1 provinsi yaitu Bengkulu. Pada *cluster* 2 terdapat 4 provinsi terdiri dari Kep. Bangka Belitung, DKI Jakarta, Jawa Barat, dan Nusa Tenggara Timur. Pada *cluster* 3 terdapat 1 provinsi yaitu Papua. Pada *cluster* 4 terdapat 5 provinsi terdiri dari Lampung, Kep. Riau, Kalimantan Tengah, Kalimantan Selatan, dan Sulawesi Barat. Sedangkan pada *cluster* 5 terdapat 23 Provinsi terdiri dari Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Timur, Kalimantan Utara, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, dan Papua Barat.

Berdasarkan penelitian yang telah dilakukan maka didapatkan saran bahwa penelitian ini hanya sebagai penerapan dari algoritma *Partitioning Around Medoids*, agar lebih bermanfaat maka disarankan membangun sistem Informasi sebagai media implementasi algoritma *Partitioning Around Medoids*. Penelitian ini dapat dikembangkan menggunakan metode *clustering* yang berbeda, untuk uji validasi *cluster* dapat menggunakan teknik yang berbeda seperti metode *elbow*, *silhouette coefficient*, dan dapat meningkatkan jumlah atribut dan variabel pada data yang digunakan.

## 6. DAFTAR PUSTAKA

- Abdillah, G., Putra, F. A., Renaldi, F., Informatika, P. S., Jenderal, U., Yani, A., Barat, J., & Cimahi, K. (2016). Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means. *Sentika 2016*, 18–19.
- Abdurrahman, D. D., Agus, F. & Putra, G. M. (2021). Implementasi Algoritma Partitioning Around Medoids (PAM) untuk Mengelompokkan Hasil Produksi Komoditi Perkebunan (Studi Kasus: Dinas Perkebunan Provinsi Kalimantan Timur). *Jurnal Ilmiah Ilmu Komputer, Informatika*

*Mulawarman*. Vol. 16, No. 2 September 2021.

- Agarwal, S. (2014). Data mining: Data Mining Concepts and Techniques. *Proceedings – 2013 International Conference on Machine Intelligence Research and Advancement*. <https://doi.org/10.1109/ICMIRA.2013.45>
- Badan Pusat Statistik. (2018). Indikator Perumahan dan Kesehatan Lingkungan Tahun 2018. Badan Pusat Statistik. Jakarta.
- Han, J., & Kamber, M. (2006). *Data Mining: Concept and Techniques, 2nd ed.* San Fransisco: Morgan Kauffman.
- Kaur, N. U., & Singh, D. (2014). K-Medoids Clustering Algorithm – A Review. *International Journal of Computer Application and Technology (IJCAT)*. Vol. 1, Issue 1.
- Lubis, I., Aisyah, N., & Mardikanto, A. K. (2019). *Pedoman Pengukuran Capaian Pembangunan Perumahan dan Permukiman Berbasis Hasil (Outcome)*. Jakarta: Kementerian Perencanaan Pembangunan Nasional.
- Mayasari, T. R. (2020). Pengelompokan Berdasarkan Variabel Kesehatan Lingkungan dan Pengaruhnya Terhadap Kemiskinan di Indonesia Tahun 2018. Lampung: *Jurnal Siger Matematika*.
- Pramesti, D. F., Furqon, M. T., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(9), 723–732.
- Sujacka, R. (2019.) *Peningkatan Akurasi Algortima K-Means dengan Clustering Purity sebagai Titik Pusat Cluster Awal (Centroid)*. Universitas Sumatera Utara.
- Triyanto, W. A. (2015). Algoritma K-Medoids untuk Penentuan Strategi Pemasaran Produk. *Jurnal SIMETRIS*. Vol. 6 No. 1, 183 – 188.