



Clustering Model K-Means pada Kasus Angka Putus Sekolah Tingkatan Sekolah Dasar di Provinsi Jawa Tengah

Laila Khoirun Nisa¹, Tari Fitri Ningsih² dan Burhanuddin Izzul Salam³, Fatkhurokhman Fauzi^{*4}, Eny Winaryati⁵

^{1, 2, 3, 4}Program Studi Statistika, Fakultas Sains dan Teknologi Pertanian, Universitas Muhammadiyah Semarang, Indonesia

⁵Program Studi Pendidikan Kimia, Fakultas Ilmu Pendidikan dan Humaniora, Universitas Muhammadiyah Semarang, Indonesia

E-mail : fatkhurokhmanf@unimus.ac.id*

*Penulis Korespondensi

Received 30 May 2024; Revised 12 June 2024; Accepted 19 June 2024

Abstrak - Pendidikan dasar bertujuan untuk membekali anak dengan keterampilan dasar yang mereka butuhkan untuk mengarungi kehidupan mereka sebagai individu, elemen masyarakat, elemen warga negara dan juga sebagai umat manusia, serta mempersiapkan mereka untuk pendidikan tinggi dimasa mendatang. Kasus putus sekolah tampaknya menjadi sebuah masalah yang belum bisa diatasi. Dampak putus sekolah jika tidak dikelola dengan baik tentunya akan merugikan, termasuk berdampak pada kualitas sumber daya di masa depan. Oleh karena itu, perlu dilakukan suatu tindakan guna mengurangi angka putus sekolah jenjang SD. Penelitian ini dilakukan dengan algoritma K-Means clustering guna mengetahui wilayah kabupaten/kota di Jawa Tengah yang memiliki angka putus sekolah tingkat tinggi, sedang, dan rendah. Hasil pengelompokan (clustering) dengan algoritma K-Means melalui 3 metode memperoleh nilai K optimal adalah 3 maka dari itu terbentuk 3 cluster dari keseluruhan wilayah kabupaten/kota di Jawa Tengah yang berjumlah 35 wilayah, cluster 1 dengan kecenderungan angka putus sekolah dasar tinggi terdapat 14 wilayah kabupaten/kota, cluster 2 dengan kecenderungan angka putus sekolah dasar sedang terdapat 15 wilayah kabupaten/kota, dan cluster 3 dengan kecenderungan angka putus sekolah dasar rendah terdapat 6 kabupaten/kota.

Kata Kunci: Clustering, K-Means, Putus Sekolah, Sekolah Dasar.

Abstract - Basic education aims to equip children with the basic skills they need to navigate their lives as individuals, elements of society, elements of citizens and also as human beings, and prepare them for higher education in the future. The case of dropping out of school seems to be a problem that cannot be overcome. The impact of dropping out of school if not managed properly will certainly be detrimental, including having an impact on the quality of resources in the future. Therefore, it is necessary to take action to reduce the dropout rate at the elementary school level. This research was conducted using the K-Means clustering algorithm to find out which districts/cities in Central Java have high, medium, and low dropout rates. The results of clustering using the K-Means algorithm through 3 methods obtained an optimal K value of 3, therefore 3 clusters were formed from all 35 regencies/cities in Central Java, cluster 1 with a tendency for high elementary school dropout rates to be 14 regencies/cities, cluster 2 with a moderate trend of dropout rates from elementary school there are 15 regencies/cities, and cluster 3 with a low trend of dropout rates from elementary school there are 6 regencies/cities.

Keywords: Clustering, K-Means, Dropout, Elementary School.



1. PENDAHULUAN

Pendidikan memegang peranan yang krusial dalam kehidupan, tidak hanya bertujuan untuk membuat seseorang menjadi cerdas tetapi pendidikan juga digunakan dalam proses pembentukan karakter, penerapan akhlak, budi pekerti, dan pengembangan keterampilan. Pendidikan memiliki peran vital sebagai salah satu sarana untuk meningkatkan mutu kehidupan seseorang. Tidaklah mengherankan bahwa kegiatan pendidikan merupakan bagian integral dari kehidupan manusia. Sejak lahir, setiap individu mengalami proses belajar dan pendidikan. Mulai dari usia dini, orang tua mengajarkan anak-anak untuk berbicara, makan, berjalan, dan memahami dunia sekitar mereka. Ketika anak-anak tumbuh dewasa, mereka umumnya mengikuti pendidikan formal di sekolah (Makkawaru, 2019).

Jenis pendidikan dibagi menjadi 3 macam yang pada umumnya dikenal yaitu pendidikan formal, pendidikan nonformal, dan pendidikan informal. Pendidikan formal adalah pendidikan yang dilaksanakan oleh suatu lembaga pendidikan yang dibedakan sesuai tingkatan atau jenjang yaitu SD, SMP, dan SMA. Sekolah dasar (SD) memberikan pendidikan selama enam tahun bagi anak-anak berusia 6-12 tahun (Nurlaila et al., 2017). Sekolah dasar atau SD bagi anak sangat penting guna mendorong pemerintah untuk terus berupaya melaksanakan pendidikan dasar yang bermutu dan berkualitas (Lestari et al., 2020). Pendidikan sekolah dasar memiliki tujuan untuk memberikan anak-anak keterampilan dasar yang diperlukan untuk mengembangkan kehidupan mereka sebagai individu, anggota masyarakat, warga negara, dan anggota umat manusia. Selain itu, tujuan pendidikan ini juga adalah mempersiapkan mereka untuk melanjutkan pendidikan ke tingkat yang lebih tinggi di masa depan (Sinaga et al., 2016).

Masalah mengenai putus sekolah sering terdengar di masyarakat, bahkan jumlahnya semakin meningkat setiap tahunnya. Kasus putus sekolah tampaknya menjadi sebuah masalah yang belum bisa diatasi (Niyogisubizo et al., 2022). Akibat dari putus sekolah tentunya berdampak negatif jika tidak ditangani dengan baik, salah satunya berdampak pada kualitas sumber daya manusia dimasa depan. Berdasarkan data yang diperoleh dari Kementerian Pendidikan dan Kebudayaan (KEMENDIKBUD), jumlah anak yang putus sekolah pada jenjang sekolah dasar (SD) di provinsi Jawa Tengah terus mengalami peningkatan untuk tiap tahun ajarannya. Selain itu, provinsi Jawa Tengah menempati posisi ke tiga pada kasus angka putus sekolah jenjang SD terbanyak di Indonesia untuk tahun ajaran 2020/2021 setelah provinsi Jawa Barat dan DKI Jakarta. Jumlah anak putus sekolah jenjang SD di Jawa Tengah tahun ajaran 2016/2017 sebanyak 2.205 orang, tahun ajaran 2017/2018 tercatat 2.238 orang, kemudian pada tahun ajaran 2018/2019 berjumlah 3.800 orang, dan pada tahun ajaran 2020/2021 sebanyak 3.992 orang. Oleh karena itu, perlu dilakukan pengelompokan angka putus sekolah untuk mengelompokkan kabupaten/kota di Jawa Tengah yang perlu dilakukan *treatment*. Salah satu metode pengelompokan yang bisa digunakan untuk pengelompokan tersebut adalah metode *clustering*.

Berdasarkan informasi yang diperoleh dari KEMENDIKBUD mengenai jumlah anak putus sekolah pada jenjang sekolah dasar (SD) di provinsi Jawa Tengah, penelitian ini bertujuan untuk mengelompokkan tingkat putus sekolah SD menjadi kategori rendah, sedang, dan tinggi. Adapun metode atau teknik yang diambil untuk penelitian ini ialah metode *clustering* dengan algoritma K-Means. *Clustering* merupakan suatu metode yang digunakan untuk mengelompokkan data ke dalam beberapa *cluster* atau grup, dimana data pada satu *cluster* mempunyai tingkat kesamaan yang tinggi dan mempunyai tingkat kesamaan rendah dengan *cluster* lainnya (Ramadhan et al., 2017). Teknik *clustering* ini digunakan tanpa ada training dan testing, dan tanpa perlu adanya target outcome (Julianti Hablum et al., 2019). Metode pengelompokan non-hierarki diawali dengan cara menentukan berapa banyak cluster yang diharapkan. Proses pengelompokan



dilakukan tanpa mengikuti hierarki, setelah jumlah cluster ditentukan. Proses ini sering disebut sebagai *K-Means clustering* (Ong, 2013).

Metode K-Means adalah salah satu metode non-hierarki yang digunakan untuk mengelompokkan item ke dalam *K cluster*. Metode ini dilakukan dengan cara meminimalkan SS (*sum of square*) jarak dengan centroid (titik pusat) dari masing-masing *cluster* (Das et al., 2023). Penentuan nilai *K* optimal pada *cluster* dapat menggunakan *Elbow Method*, metode ini melibatkan pengamatan terhadap perbandingan antara jumlah cluster dan varians dalam setiap *cluster*. Pada titik tertentu, terlihat adanya siku pada grafik yang menunjukkan penurunan yang signifikan dalam varians. Pada titik ini, kita dapat menambahkan jumlah *cluster* tersebut untuk membentuk model data yang paling baik dalam menentukan jumlah *cluster* yang optimal (Sari et al., n.d.).

Penelitian mengenai K-Means telah banyak dilakukan, dimana beberapa diantaranya yaitu penelitian yang dilakukan oleh Widyadhana *et al* yang berjudul Perbandingan Analisis Kluster K-Means dan *Average Linkage* untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah, setelah membandingkan kedua metode, penelitian ini menyimpulkan bahwa metode *Average Linkage* memberikan hasil yang lebih baik. Hal ini didasarkan pada fakta bahwa nilai *Silhouette Coefficient* pada metode *Average Linkage* lebih tinggi dibandingkan dengan metode K-Means, yaitu sebesar 0,35, sedangkan untuk metode K-Means hanya sebesar 0,2 (Widyadhana et al., 2021). Penelitian lainnya yang dilakukan oleh Nurhidayati dengan judul Penggunaan Metode K-Means *Cluster* untuk mengklasifikasikan Kemampuan 4C Mahasiswa, menyatakan bahwa hasil klasifikasi kompetensi 4C mahasiswa dengan metode K-Means menunjukkan nilai *R2* sebesar 47,1%. Hal ini mengindikasikan bahwa *cluster* yang dihasilkan belum optimal, karena terdapat satu anggota *cluster 2* yang seharusnya termasuk dalam *cluster 3* (M Nurhidayati & N Khasanah, 2021).

Metode K Means bertujuan untuk memaksimalkan kesamaan antara data dalam suatu kelompok dan meminimalkan kesamaan data antar kelompok. Metode ini merupakan salah satu teknik pengelompokan (*clustering*) berbasis jarak yang mengelompokkan data menjadi dua atau lebih kelompok berdasarkan karakteristik yang sama. Namun, metode K-Means hanya bekerja dengan atribut angka (Aditya et al., 2018). Tujuan penelitian ini untuk mengidentifikasi wilayah kabupaten atau kota di Provinsi Jawa Tengah yang termasuk dalam kategori kelompok putus sekolah tahapan SD dengan tingkat rendah, sedang, dan tinggi. Hal ini diharapkan dapat memberikan kontribusi untuk pemerintah dan masyarakat dalam menangani masalah putus sekolah tahapan SD di Provinsi Jawa Tengah.

2. METODE PENELITIAN

Penelitian ini dilakukan di Provinsi Jawa Tengah. Data penelitian ialah data kuantitatif yang merupakan data sekunder pada tahun 2021, diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Tengah. Teknik yang digunakan dalam penelitian ini yaitu *clustering* dengan algoritma K-Means. Variabel yang digunakan yaitu:

Tabel 1. Variabel Indikator Putus Sekolah

No	Variabel Putus Sekolah
1	Rasio jenis kelamin
2	Jumlah kepadatan penduduk
3	Rasio siswa terhadap guru
4	Rasio siswa terhadap sekolah
5	Jumlah kelurahan/desa yang memiliki fasilitas sekolah dasar
6	Tingkat kemiskinan



Langkah-langkah umum dari algoritma K-Means ialah (Sarwono, n.d.)(Hervisari et al., 2014)(Liu et al., 2023):

1. Tetapkan jumlah K cluster yang diinginkan, pilih secara acak K titik pusat awal (centroid) sebagai representasi awal dari setiap cluster.
2. Pilih secara acak K titik pusat awal (*centroid*) sebagai representasi awal dari setiap cluster.
3. Hitung jarak dari semua data input dengan setiap centroid menggunakan rumus jarak *Euclidean (Euclidian Distance)*. Kemudian, identifikasi centroid yang memiliki jarak terdekat dengan setiap data input.

$$d(x_j, \mu_k) = \sqrt{(x_j - \mu_k)^2} \quad (1)$$

Dimana:

x_j : data kriteria

μ_k : centroid pada *cluster* ke-k

4. Setiap data dikelompokkan berdasarkan tingkat kedekatannya dengan *centroid*.
5. Hitung rata-rata setiap *cluster* untuk memperbarui nilai centroid. Rumus yang digunakan untuk menghitung *centroid* baru adalah sebagai berikut:

$$\mu_k(t+1) = \frac{1}{N_{sk}} \sum_{k \in S_k} x_k \quad (2)$$

Dimana:

$\mu_k(t+1)$: centroid baru pada iterasi ke (t+1)

N_{sk} : banyak data pada *cluster* S_k

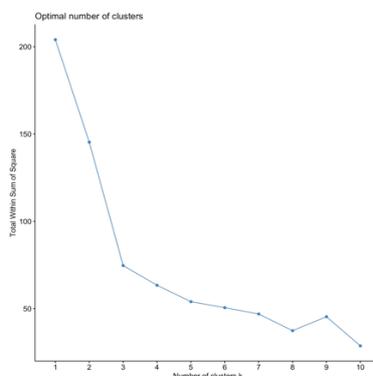
6. Ulangi langkah-langkah 2 hingga 5 secara iteratif hingga tidak ada lagi perubahan dalam anggota setiap *cluster* atau hingga batas iterasi maksimum tercapai.

3. HASIL DAN PEMBAHASAN

Penentuan K optimal *cluster* menggunakan beberapa metode antara lain metode *elbow*, metode *silhouette*, dan metode *gap statistics*.

Metode *Elbow*

Metode *elbow* (siku) berfungsi dalam menentukan nilai K optimal dalam *clustering* dengan mempertimbangkan nilai total WSS (*within sum square*). Nilai WSS ialah jumlah dari jarak kuadrat antara setiap data dalam *cluster* dengan pusat clusternya.



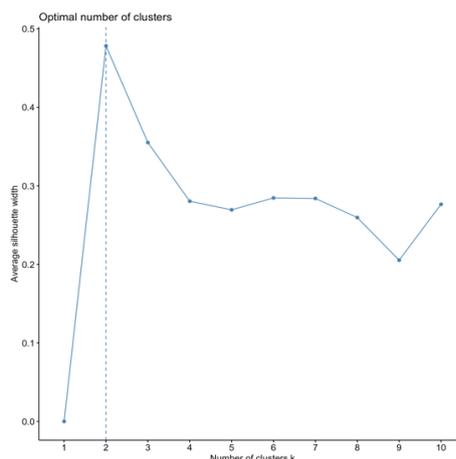
Gambar 1. Output *Elbow*



Berdasarkan gambar di atas, garis membentuk siku pada saat nilai K adalah 3. Menggunakan *elbow method*, menunjukkan bahwa penambahan cluster setelah $K = 3$ tidak memberikan penurunan nilai total WSS yang signifikan. Oleh karena itu, dengan menggunakan metode elbow dapat disimpulkan bahwa nilai K optimal adalah 3.

Metode *Silhouette*

Metode *silhouette* dalam penentuan nilai K tidak dilakukan dengan mempertimbangkan garis tertinggi, melainkan dengan melihat garis yang menghasilkan nilai silhouette yang paling optimum.

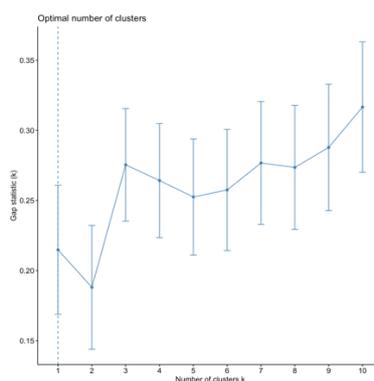


Gambar 2. Output *Silhouette*

Gambar di atas menunjukkan bahwa nilai K optimal berdasarkan metode *silhouette* yaitu 2, tetapi terdapat opsi alternatif dalam menentukan nilai K dengan grafik selanjutnya yang memiliki nilai tertinggi setelah puncak pertama. Dalam kasus ini, grafik tertinggi kedua terjadi pada angka 3. Oleh karena itu, nilai K optimal menggunakan metode *silhouette* yaitu 3.

Metode *GAP Statistic*

Menentukan nilai K optimal atau terbaik menggunakan cara GAP Statistic atau biasa disebut *gap_stat* dengan melihat dari garis putus-putus yang menghubungkan garis.



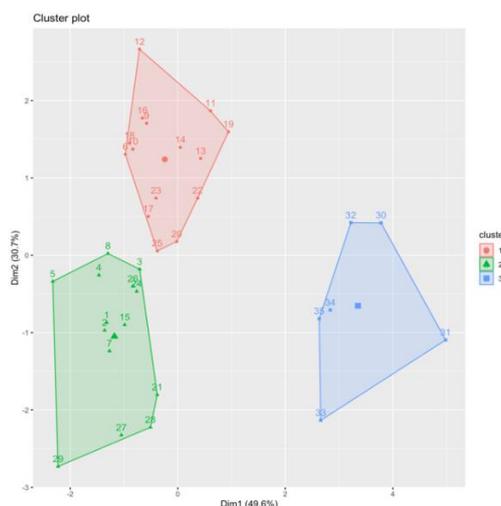
Gambar 3. Output *GAP Statistic*

Hasil dari metode *gap_stat* menunjukkan bahwa sebuah garis pada angka 10, namun jika dianalisis secara logis, dengan jumlah data sebanyak 210 melakukan pengklasteran sebanyak 10 itu terlalu banyak. Oleh karena itu, kita bisa melihat ulang garis tertinggi lainnya dan ternyata



terletak pada angka 3 sehingga menunjukkan bahwa nilai K optimal atau jumlah *cluster* yang optimal menurut metode *gap_stat* adalah 3.

Setelah dilakukan untuk menentukan K optimal menggunakan 3 metode di atas maka dapat disimpulkan bahwa bahwa nilai K optimalnya adalah 3. Hasil uji K-Means dengan K = 3 diperoleh kluster-kluster putus sekolah jenjang SD di Jawa Tengah tahun 2021. Kluster 1 berisikan putus sekolah tinggi terdapat 14 kabupaten/kota, kluster 2 berisikan putus sekolah sedang terdapat 15 kabupaten/kota, dan kluster ke 3 berisikan putus sekolah rendah terdapat 6 kabupaten/kota.



Gambar 4. *Clustering Plot* Putus Sekolah SD

Tabel 2. *Clustering* Kabupaten/Kota di Jawa Tengah pada Putus Sekolah SD

Cluster	Kabupaten/Kota	Identifikasi
1	Kab. Cilacap, Kab. Banyumas, Kab. Purbalingga, Kab. Banjarnegara, Kab. Kebumen, Kab. Wonosobo, Kab. Magelang, Kab. Grobogan, Kab. Demak, Kab. Kendal, Kab. Pekalongan, Kab. Pemasang, Kab. Tegal, Kab. Brebes	<i>Cluster</i> ini memiliki rata-rata tingkat kemiskinan dan rasio siswa terhadap guru yang paling tinggi dibandingkan <i>cluster</i> lain. Jumlah guru yang terbatas akan membuat rasio siswa terhadap guru juga besar sehingga akan ada siswa yang tidak diperhatikan oleh guru dan dapat membuat anak tersebut putus sekolah.
2	Kab. Purworejo, Kab. Boyolali, Kab. Klaten, Kab. Sukoharjo, Kab. Wonogiri, Kab. Karanganyar, Kab. Sragen, Kab. Blora, Kab. Rembang, Kab. Pati, Kab. Kudus, Kab. Jepara, Kab. Semarang, Kab. Temanggung, Kab. Batang	Tingkat kemiskinan dan jumlah kepadatan penduduk <i>cluster</i> ini lebih rendah dibandingkan <i>cluster</i> 1.
3	Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, Kota Tegal	Berbanding terbalik dengan <i>cluster</i> 1, untuk <i>cluster</i> 3 ini tingkat kemiskinannya lebih rendah dibandingkan <i>cluster</i> lain.



4. KESIMPULAN

Hasil pengelompokan (*clustering*) menggunakan algoritma K-Means dengan 3 cara yaitu *elbow method*, *silhouette method*, dan *GAP Statistic method* memperoleh nilai K terbaik adalah 3. Oleh karena itu, 35 kabupaten/kota di Jawa Tengah terbagi menjadi 3 *cluster* yang menyatakan adanya perbedaan atau karakteristik yang membedakan antara cluster-cluster tersebut. Sebagai hasilnya bahwa *cluster* 1 dengan kecenderungan angka putus sekolah dasar tinggi terdapat 14 wilayah kabupaten/kota, *cluster* 2 dengan kecenderungan angka putus sekolah dasar sedang terdapat 15 wilayah kabupaten/kota, dan *cluster* 3 dengan kecenderungan angka putus sekolah dasar rendah terdapat 6 kabupaten/kota.

DAFTAR PUSTAKA

- Aditya, K. B., Setiawan, Y., & Puspitaningrum, D. (2018). Sistem Informasi Geografis Pemetaan Faktor-faktor yang Mempengaruhi Angka Kematian Ibu (AKI) dan Angka Kematian Bayi (AKB) dengan Metode K-Means Clustering (Studi Kasus: Provinsi Bengkulu). *Jurnal Teknik Informatika*, 10(1), 59–66.
- Das, D., Kayal, P., & Maiti, M. (2023). A K-means clustering model for analyzing the Bitcoin extreme value returns. *Decision Analytics Journal*, 6. <https://doi.org/10.1016/j.dajour.2022.100152>
- Hervisari, S., Musdalifah, & Sudarsana. (2014). Implementasi Algoritma K-Means Untuk Diagnosa Penyakit Gagal Ginjal Kronis. *Jurnal Ilmiah Matematika Dan Terapan*, 11(1), 27–25.
- Julianti Hablum, R., Khairan, A., Jati Metro, J., & Ternate Selatan, K. (2019). Clustering Hasil Tangkap Ikan di Pelabuhan Perikanan Nusantara (PPN) Ternate Menggunakan Algoritma K-Means. *Jurnal Informatika Dan Komputer*, 02 no 1.
- Lestari, A. Y., Kurniawan, F., & Ardi, R. B. (2020). Penyebab Tingginya Angka Anak Putus Sekolah Jenjang Sekolah Dasar (SD). *Jurnal Ilmiah Sekolah Dasar*, 4 no 2, 299–308.
- Liu, G., Ji, F., Sun, W., & Sun, L. (2023). Optimization design of short-circuit test platform for the distribution network of integrated power system based on improved K-means clustering. *Energy Reports*, 9, 716–726. <https://doi.org/10.1016/j.egy.2023.04.319>
- M Nurhidayati, & N Khasanah. (2021). Penggunaan Metode K-Means Cluster Untuk Mengklasifikasikan Kemampuan 4C Mahasiswa. *Jurnal Ilmiah Matematika Dan Terapan*, 18(2), 160–169. <https://doi.org/10.22487/2540766x.2021.v18.i2.15615>
- Makkawaru, M. (2019). Pentingnya Pendidikan Bagi Kehidupan dan Pendidikan Karakter dalam Dunia Pendidikan. *Jurnal Konsepsi*, 8(3), 116–119. <https://p3i.my.id/index.php/konsepsi/article/view/87>
- Niyogisubizo, J., Liao, L., Nziyumva, E., Murwanashyaka, E., & Nshimyumukiza, P. C. (2022). Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. *Computers and Education: Artificial Intelligence*, 3. <https://doi.org/10.1016/j.caeai.2022.100066>
- Nurlaila, S., Dadi, S., & Karjiati, V. (2017). Pengaruh Model Cooperative Learning Tipe Teams Games Tournaments (TGT) Terhadap Hasil Belajar Siswa Pada Mata Pelajaran IPS Di Kelas IV SDN Gugus X kota Bengkulu. *Jurnal Riset Pendidikan Dasar*, 1, 43–49.
- Ong, J. O. (2013). Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing President University. *Jurnal Ilmiah Teknik Industri*, 12 no 1.
- Ramadhan, A., Efendi, Z., & Mustakim. (2017). Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI)* 9, 219–226.
- Sari, R. Y., Oktavianto, H., & Sulistyono, H. W. (n.d.). *Algoritma K-Means Dengan Metode Elbow Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Komponen Pembentuk Indeks Pembangunan Manusia*.
- Sarwono, Y. T. (n.d.). *Aplikasi Model Jaringan Syaraf Tiruan dengan Radial Basis Function untuk Mendeteksi Kelainan Otak (Stroke Infark)*.
- Sinaga, N. D., Rais, & Sain, H. (2016). Model Regresi Logistik Biner Untuk Menentukan Faktor Yang Berpengaruh Terhadap Anak Putus Sekolah Di Sulawesi Tengah. *Jurnal Ilmiah Matematika Dan Terapan*, 13 no.1, 24–37.



Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 584–594. <https://journal.unnes.ac.id/sju/index.php/prisma/>