

**KLASTERISASI PENYAKIT MENULAR DI INDONESIA MENGGUNAKAN METODE K-MEANS CLUSTERING****Muhammad Hafiz\*, Sylvia Anggraini**

Mahasiswa Prodi Sistem Informasi

Sekolah Tinggi Manajemen Informatika dan Komputer Royal

\*email: [muhammad11hafiz09@gmail.com](mailto:muhammad11hafiz09@gmail.com)

**Abstract:** Disease refers to an abnormal state in which the body or mind experiences discomfort or dysfunction. Every day, the number of people affected by infectious diseases continues to increase with various types of diseases. Therefore, grouping is needed to help the government obtain information about the most common diseases in society. In this research, patient disease data was grouped using multidimensional clustering data mining techniques. One method used is K-Means Clustering, a non-hierarchical approach that groups data into one or more clusters. Data that has similar characteristics is grouped together in one cluster, while data that has different characteristics is placed in different clusters, thus allowing clusters to be formed with minimal variation. This research aims to make the government pay more attention to areas that have high rates of infectious diseases both from the environment and other things. The results of research from 34 provinces and 8 infectious diseases show data where 32 areas are very vulnerable, 2 areas are vulnerable and 2 areas are quite vulnerable.

**Keywords:** Data Mining, Disease Cases, K-Means

**Abstrak:** Penyakit merujuk pada keadaan abnormal di mana tubuh atau pikiran mengalami ketidaknyamanan atau disfungsi. Setiap hari, jumlah orang yang terkena penyakit menular terus bertambah dengan berbagai jenis penyakit. Oleh karena itu, diperlukan pengelompokan untuk membantu pemerintah memperoleh informasi tentang penyakit yang paling umum di masyarakat. Dalam penelitian ini, dilakukan pengelompokan data penyakit pasien menggunakan teknik penggalian data clustering multidimensi. Salah satu metode yang digunakan adalah K-Means Clustering, sebuah pendekatan non-hirarkis yang mengelompokkan data ke dalam satu atau lebih cluster. Data-data yang memiliki karakteristik serupa dikelompokkan bersama dalam satu cluster, sementara data yang memiliki karakteristik berbeda ditempatkan dalam cluster yang berbeda, sehingga memungkinkan terbentuknya cluster dengan variasi yang minimal. Penelitian ini bertujuan agar pemerintah lebih perhatian terhadap daerah yang memiliki angka tinggi terhadap penyakit menular baik dari lingkungan maupun dari hal lainnya. Hasil dari penelitian dari 34 Provinsi dan 8 Penyakit menular menunjukkan data dimana 32 daerah dengan sangat rawan, 2 daerah rawan dan 2 daerah cukup rawan.

**Kata kunci:** Data Mining, Kasus Penyakit, *K-Means*

**PENDAHULUAN**

Di Indonesia, masalah polusi dan lingkungan merupakan isu yang sangat penting dan harus segera diatasi. Lingkungan saat ini mengalami berbagai kerusakan, seperti pencemaran air, tanah, sungai, kerusakan hutan, dan juga ancaman kepunahan bagi beberapa spesies hewan [1]. Kerusakan dan pencemaran di Indonesia berkembang

dengan cepat. Ini disebabkan oleh sering diabaikannya kepentingan pelestarian lingkungan dalam pengambilan keputusan di tingkat pusat dan daerah. Kondisi sanitasi lingkungan yang buruk menyebabkan penyebaran penyakit yang beragam, menjadi penyebab utama munculnya masalah kesehatan terkait lingkungan yang tidak memadai [2].

Penyakit merupakan kondisi tidak normal di mana tubuh atau pikiran seseorang mengalami ketidaknyamanan atau gangguan. Setiap hari, jumlah pasien di rumah sakit terus meningkat dengan berbagai jenis penyakit yang berbeda [3]. Kesehatan menjadi salah satu penunjuk tingkat pembangunan manusia. Karena pembangunan di bidang kesehatan di Indonesia saat ini mengalami beban ganda [4]. Masalah kesehatan tidak hanya timbul karena kesalahan individu, tetapi juga bisa disebabkan oleh kurangnya pengetahuan masyarakat karena minimnya informasi yang akurat tentang suatu penyakit [5].

Di Indonesia juga sangat banyak di jumpai berbagai penyakit-penyakit, seperti dikutip dari data Badan Statistik Indonesia pada tahun 2015. Data penyakit di Indonesia menjadi kunci utama dalam upaya pencegahan dan penanggulangan kesehatan masyarakat. Melalui pengumpulan, analisis, dan interpretasi data yang akurat, kita dapat mengidentifikasi tren penyakit, mengukur dampaknya, dan merancang strategi intervensi yang efektif. Terdapat data penyakit seperti penyakit Malaria, Pneumonia, Kusta, TB Paru, Tetanus Neonatorum, Campak, DBD dan Diare. Penyakit tersebut banyak di jumpai di berbagai provinsi. Karena hal tersebut penyakit-penyakit tersebut harus segera ditangani agar tidak terjadi kasus yang lebih banyak lagi.

Malaria disebabkan oleh parasit Plasmodium yang kemudian disebarkan melalui gigitan nyamuk Anopheles [6]. Tuberkulosis paru adalah penyakit menular yang disebabkan oleh bakteri Mycobacterium tuberculosis. Penularannya terjadi ketika seseorang yang menderita TBC paru dengan hasil tes BTA positif batuk atau bersin, dan secara tidak sengaja menyebarkan bakteri ke udara dalam bentuk percikan dahak [7]. Pneumonia adalah inflamasi pada jaringan paru-paru yang disebabkan oleh infeksi bakteri atau virus [8]. Kusta adalah penyakit yang diakibatkan oleh bakteri Mycobacterium leprae. [9]. Tetanus terjadi karena neurotoksin yang dihasilkan oleh bakteri anaerob Clostridium tetani [10]. Campak merupakan penyakit infeksi disebabkan oleh virus campak dan sangat menular. Virus campak adalah jenis virus RNA tunggal negatif yang berselubung, tidak bersegmen, dan termasuk dalam genus Morbillivirus dalam famili Paramyxoviridae [11]. Diare adalah kondisi di mana terjadi perubahan pada bentuk dan konsistensi tinja serta peningkatan frekuensi buang air besar yang berlebihan, yaitu lebih dari tiga kali dalam satu hari [12]. Demam Berdarah Dengue (DBD) adalah penyakit infeksi akut yang menular kepada manusia melalui gigitan nyamuk Aedes yang terinfeksi virus dengue [13].

Data mining merupakan sebuah alat yang memungkinkan pengguna untuk dengan cepat mengakses data dalam jumlah besar. Secara lebih spesifik, data mining adalah alat dan aplikasi yang menggunakan analisis statistik pada data. Proses data mining adalah ekstraksi atau penggalian data dan informasi yang besar, yang sebelumnya tidak diketahui, namun dapat dipahami dan bermanfaat dari database besar untuk membuat keputusan bisnis penting. Data mining mengacu pada kumpulan teknik dengan tujuan untuk menemukan pola-pola yang tidak diketahui dalam data yang telah

dikumpulkan. Ini memungkinkan pengguna untuk menemukan pengetahuan dalam database yang tidak dapat diketahui oleh pengguna sebelumnya [14].

Salah satu teknik pengelompokan data adalah Algoritma K-means, yang termasuk dalam kategori algoritma clustering. K-means clustering merupakan metode non-hirarki yang mengorganisir data menjadi satu atau lebih kelompok. Tujuan utama K-means adalah untuk mengurangi kesalahan kuadrat antara rata-rata dari setiap kelompok dengan titik data di dalam kelompok tersebut. Algoritma K-means termasuk dalam kategori metode partisi yang menggunakan titik pusat sebagai acuan, berbeda dengan algoritma k-Medoids yang menggunakan objek sebagai acuan [15].

Clustering merupakan proses pengelompokan data ke dalam beberapa kelompok, di mana kesamaan antara data dalam satu kelompok lebih besar daripada kesamaannya dengan data dalam kelompok lain. Ini merupakan salah satu masalah utama dalam pembelajaran tanpa pengawasan, di mana tujuannya adalah untuk menemukan struktur dalam kumpulan data yang tidak memiliki label. Secara sederhana, clustering adalah proses mengorganisir objek menjadi kelompok yang memiliki kesamaan tertentu. Sebuah kelompok adalah kumpulan objek yang memiliki kesamaan di antara mereka, sementara berbeda dengan objek dalam kelompok lain [15].

Dari penjelasan tersebut, dapat dilakukan studi tentang analisis kluster dengan fokus pada penyakit yang sering mengancam kesehatan masyarakat di Indonesia, seperti Malaria, Tuberkulosis Paru, Pneumonia, Kusta, Tetanus Neonatorum, Campak, Diare, dan Demam Berdarah Dengue (DBD). Judul dalam penelitian ini adalah “Klasterisasi Penyakit Menular di Indonesia Menggunakan Metode K-Means Clustering”.

## **METODE**

K-Means Clustering adalah teknik pengelompokan data yang bukan bersifat hirarkis, yang mengorganisir data menjadi satu atau lebih kelompok. Data dengan karakteristik serupa ditempatkan dalam satu kelompok, sementara data yang berbeda ditempatkan dalam kelompok yang berbeda, sehingga kelompok memiliki variasi yang minim [18].

Berikut adalah tahapan-tahapan dalam melakukan pengelompokan menggunakan metode K-Means:

1. Pilih jumlah kelompok K
2. Pengaturan awal pusat kluster K dapat dilakukan dengan berbagai metode, tetapi yang umum dilakukan adalah dengan pendekatan acak. Nilai awal pusat kluster ditentukan secara acak.
3. Bagikan semua data atau objek ke kluster yang paling dekat. Kedekatan antara dua data atau objek diukur berdasarkan jarak di antara keduanya. Sementara itu, dekatnya suatu data atau objek ke kluster tertentu diukur berdasarkan jarak antara data atau objek tersebut dengan pusat kluster. Dalam langkah ini, perlu dilakukan perhitungan jarak dari setiap data atau objek ke setiap pusat

kluster. Jarak terdekat antara suatu data atau objek dengan kluster tertentu akan menentukan di mana data atau objek tersebut akan ditempatkan. Untuk menghitung jarak dari data ke setiap pusat kluster, dapat digunakan rumus jarak Euclidean

$$D_{i,j} = \sqrt{(X_{1i}-X_{1j})^2+(X_{2i}-X_{2j})^2+\dots+(X_{ki}-X_{kj})^2} \quad (1)$$

Keterangan:

$D(i,j)$  = jarak data ke  $i$  ke pusat *cluster*  $j$

$X(k,i)$  = data ke  $i$  pada atribut data ke  $k$

$X(k,j)$  = titik pusat ke  $j$  pada atribut ke  $k$

- Lakukan perhitungan kembali untuk menentukan pusat kluster dengan anggota kluster saat ini. Pusat kluster adalah nilai rata-rata dari semua data atau objek dalam kluster tertentu. Alternatifnya, dapat pula menggunakan nilai median dari kluster tersebut. Oleh karena itu, nilai rata-rata (median) bukanlah satu-satunya parameter yang bisa digunakan. Kemudian, alokasikan kembali setiap objek menggunakan pusat kluster yang baru. Jika pusat kluster tidak berubah lagi, maka proses pengelompokan telah selesai.

**Tabel 1. Data Jumlah Kasus Penyakit (Kasus)**

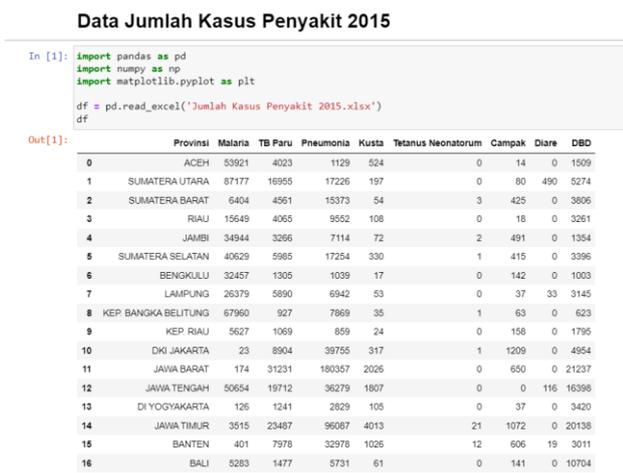
Provinsi	Malaria	TB Paru	Pneumonia	Kusta	Tetanus	Campak	Diare	DBD
Aceh	53921	4023	1129	524	0	14	0	1509
Sumut	87177	16955	17226	197	0	80	490	5274
Sumbar	6404	4561	15373	54	3	425	0	3806
...	...	...	...	...	...	...	...	...
Papua Barat	121257	793	0	712	0	0	0	66
Papua	327113	2538	0	1084	1	418	0	647

## HASIL DAN PEMBAHASAN

Implementasi algoritma K-Means dapat dilihat pada gambar 3 di bawah ialah diagram alur metode K-Means yang digunakan kedalam pengelompokan daerah potensi kasus penyakit menular. Hasil analisis dapat diberikan informasi berharga terkait fakta-fakta yang mempengaruhi penyebaran penyakit menular di Indonesia. Implikasi tersebut dapat membantu dalam perumusan kebijakan kesehatan yang lebih efektif di tingkat

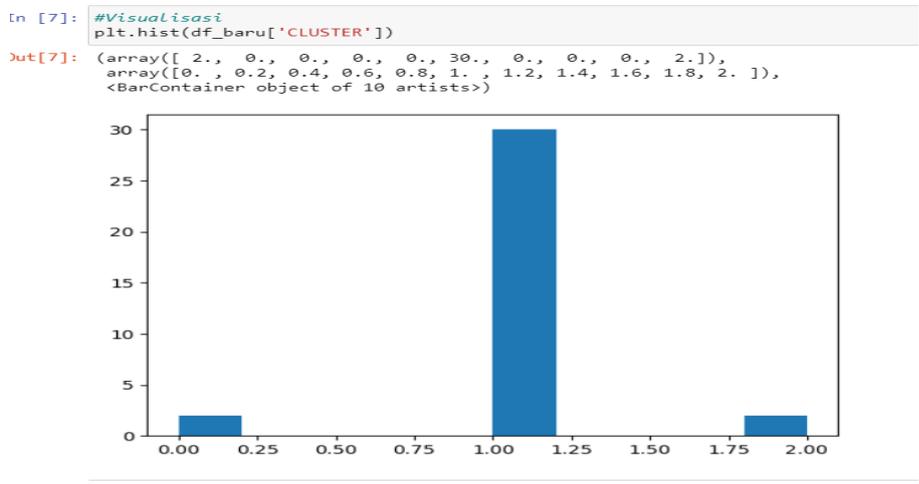
provinsi.

Provinsi-provinsi yang tergabung dalam satu cluster memiliki kesamaan pola penyebaran penyakit menular tertentu. Interpretasi hasil akan membantu dalam pemahaman faktor-faktor yang mempengaruhi penyebaran penyakit di setiap kelompok.



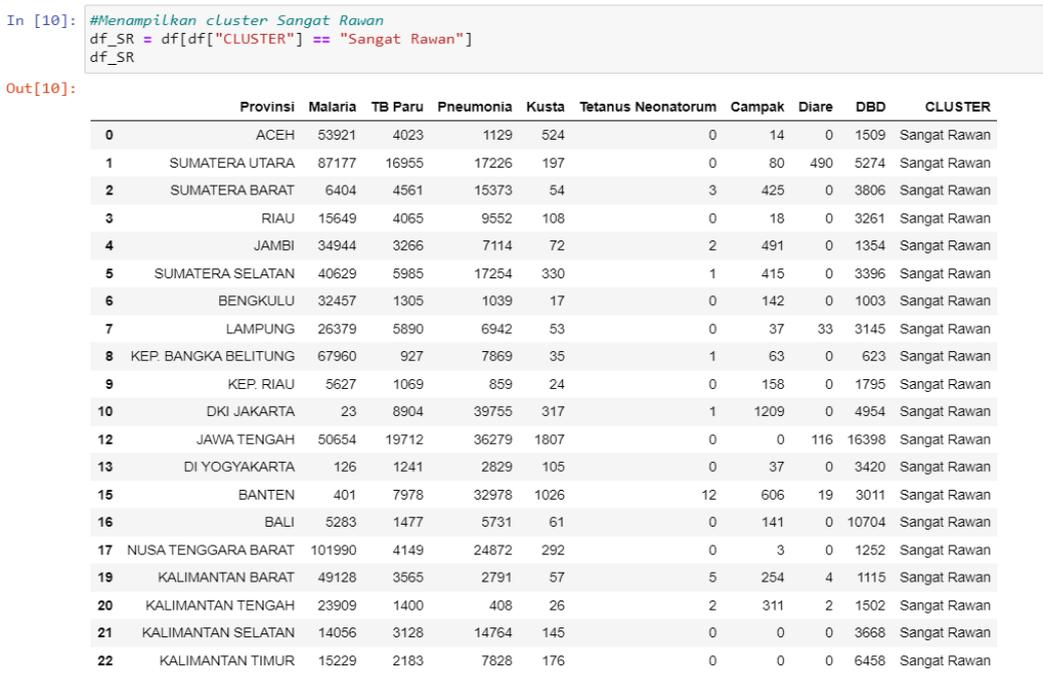
**Gambar 1 Dataset Jumlah Kasus Penyakit**

Penentuan penyebaran cluster sebagai berikut:



**Gambar 2 Visualisasi Data**

Pada penelitian ini terdapat 3 (tiga) cluster yang digunakan yaitu data kasus penyakit rawan, cukup rawan dan sangat rawan.



**Gambar 3 Hasil Cluster**

## SIMPULAN

Dalam melakukan analisis cluster terhadap provinsi-provinsi di Indonesia berdasarkan variabel penyakit menular menggunakan metode K-Means Clustering, ditemukan pola-pola yang signifikan dalam penyebaran penyakit di berbagai wilayah. Hasil analisis ini memberikan wawasan yang berharga terkait dengan keragaman geografis penyakit menular di Indonesia.

K-Means Clustering berhasil mengelompokkan provinsi-provinsi dengan kesamaan pola penyebaran penyakit, memungkinkan pemahaman yang lebih mendalam tentang faktor-faktor yang memengaruhi prevalensi penyakit menular di tingkat regional. Kelompok-kelompok yang terbentuk dapat menjadi dasar bagi perumusan kebijakan kesehatan yang lebih tepat sasaran dan efektif di setiap wilayah.

Implikasi hasil analisis ini mencakup potensi perbaikan strategi pencegahan, pengendalian, dan penanggulangan penyakit menular di tingkat provinsi. Pemahaman yang lebih baik terhadap pola cluster juga dapat mendukung alokasi sumber daya kesehatan dengan lebih efisien dan tepat guna, memastikan respons yang cepat dan efektif terhadap situasi kesehatan yang berkembang.

Meskipun K-Means Clustering memberikan kontribusi yang berharga dalam analisis ini, perlu diingat bahwa pendekatan ini memiliki keterbatasan tertentu, seperti kepekaan terhadap inisialisasi pusat cluster. Oleh karena itu, penggunaan metode ini perlu disesuaikan dengan kebutuhan dan karakteristik data yang dihadapi.

Secara keseluruhan, analisis cluster dengan K-Means Clustering memberikan kontribusi yang berarti dalam memahami dinamika penyakit menular di Indonesia,

mendukung upaya pencegahan dan pengendalian penyakit, serta memberikan landasan bagi perbaikan kebijakan kesehatan di tingkat provinsi.

## DAFTAR PUSTAKA

- [1] S. Maesaroh, B. Bahagia, and K. Kamalludin, "Strategi Menumbuhkan Literasi Lingkungan Pada Siswa," *J. Basicedu*, vol. 5, no. 4, pp. 1998–2007, 2021, [Online]. Available: <https://jbasic.org/index.php/basicedu/article/view/1048>
- [2] U. Maisarah Rasyidah, "Diare Sebagai Konsekuensi Buruknya Sanitasi Lingkungan," *KELUWIH J. Kesehat. dan Kedokt.*, vol. 1, no. 1, pp. 31–36, 2019, [Online]. Available: <https://doi.org/10.24123/jkkd.v1i1.19><http://journal.ubaya.ac.id/index.php/jkkd>
- [3] A. D. Andini and T. Arifin, "Implementasi Algoritma K-Medoids Untuk Klasterisasi Data Penyakit Pasien Di Rsud Kota Bandung," *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 2, pp. 128–138, 2020, doi: 10.51977/jti.v2i2.247.
- [4] U. Rahardja, Q. Aini, and M. Iqbal, "InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan," *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 5, no. 1, pp. 40–43, 2020, [Online]. Available: <https://doi.org/10.30743/infotekjar.v5i1.2464%0A><https://doi.org/10.30743/infotekjar.v5i1.1997>
- [5] H. Akbar and F. F. Tumiwa, "Edukasi Upaya Pencegahan Hipertensi pada Masyarakat di Kecamatan Passi Barat Kabupaten Bolaang Mongondow," *JPKMI (Jurnal Pengabd. Kpd. Masy. Indones.*, vol. 1, no. 3, pp. 154–160, 2020.
- [6] R. Rafdinal, D. Lestari, A. M. Ashari, R. K. Apindiati, and A. B. Aritonang, "Usaha Pencegahan Penularan Penyakit Malaria melalui Sosialisasi Bersih Lingkungan Ekowisata Mangrove Kuala Mempawah," *Lambung Inov. J. Pengabd. Kpd. Masy.*, vol. 7, no. 4, pp. 670–675, 2022, doi: 10.36312/linov.v7i4.981.
- [7] T. Kristini and R. Hamidah, "Potensi Penularan Tuberculosis Paru pada Anggota Keluarga Penderita," *J. Kesehat. Masy. Indones.*, vol. 15, no. 1, p. 24, 2020, doi: 10.26714/jkmi.15.1.2020.24-28.
- [8] N. P. Sari, R. Angelina, and L. Fauziah, "Pengaruh Edukasi melalui Media Video terhadap Pengetahuan dan Sikap Keluarga tentang Pneumonia pada Balita," *J. Ilmu Keperawatan Anak*, vol. 2, no. 2, p. 69, 2019, doi: 10.32584/jika.v0i0.357.
- [9] Luhung Mustika Budiharti and S. Sunendiari, "Pemodelan dan Pemetaan Jumlah Penderita Kusta di Jawa Barat dengan Regresi Binomial Negatif dan Flexibly Shaped Spatial Scan Statistic," *J. Ris. Stat.*, vol. 1, no. 2, pp. 99–106, 2021, doi: 10.29313/jrs.v1i2.409.
- [10] H. Tjung and I. W. Aryabiantara, "Perawatan pasien dengan penyakit Tetanus yang menjalani perawatan di ruang Intensif," *Medicina (B. Aires)*, vol. 52, no. 1, p. 36, 2021, doi: 10.15562/medicina.v52i1.1045.
- [11] Maryati Sutarno and Noka Ayu Putri Liana, "Faktor-Faktor Yang Berhubungan Dengan Kejadian Ispa," *J. Antara Keperawatan*, vol. 2, no. 2, pp. 44–50, 2019,

- doi: 10.37063/antaraperawat.v2i2.76.
- [12] D. D. Prawati, “Faktor Yang Mempengaruhi Kejadian Diare Di Tambak Sari, Kota Surabaya,” *J. PROMKES*, vol. 7, no. 1, p. 34, 2019, doi: 10.20473/jpk.v7.i1.2019.34-45.
- [13] H. Akbar and E. Maulana Syaputra, “Faktor Risiko Kejadian Demam Berdarah Dengue (DBD) di Kabupaten Indramayu,” *MPPKI (Media Publ. Promosi Kesehat. Indones. Indones. J. Heal. Promot.*, vol. 2, no. 3, pp. 159–164, 2019, doi: 10.31934/mppki.v2i3.626.
- [14] I. Ahmad, S. Samsugi, and Y. Irawan, “Implementasi Data Mining Sebagai Pengolahan Data,” *J. Teknoinfo*, vol. 16, no. 1, p. 46, 2022, [Online]. Available: <http://portaldata.org/index.php/portaldata/article/view/107>
- [15] M. Herviany, S. Putri Delima, T. Nurhidayah, and K. Kasini, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021, doi: 10.57152/malcom.v1i1.60.
- [16] P. Alkhairi and A. P. Windarto, “Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara,” *Semin. Nas. Teknol. Komput. Sains*, pp. 762–767, 2019.
- [17] A. Yudhistira and R. Andika, “Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering,” *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, 2023, doi: 10.58602/jaiti.v1i1.22.
- [18] D. Sartika and J. Jumadi, “Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: Universitas Dehasen Bengkulu),” pp. 703–709, 2019, [Online]. Available: <https://seminar-id.com/semnas-sainteks2019.html>