

# ALGORITMA K-MEANS UNTUK MENINGKATKAN SEGMENTASI POLA KEKERASAN

Nurul Muna Fithryani<sup>1\*</sup>, Martanto<sup>2</sup>, Arif Rinaldi Dikananda<sup>3</sup>, Dede Rohman<sup>4</sup>

<sup>1,2,3,4</sup> STMIK IKMI Cirebon; Jl. Perjuangan No. 10B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat, Telp: 0231-490480

Received: 18 Desember 2024

Accepted: 14 Januari 2025

Published: 20 Januari 2025

## Keywords:

K-Means, Data Mining,

DBI, Kekerasan

## Correspondent Email:

nurulmuna168@gmail.com

**Abstrak.** Peningkatan angka kekerasan terhadap perempuan dan anak di berbagai wilayah telah menimbulkan kebutuhan mendesak akan strategi yang efektif untuk mengidentifikasi dan mengelompokkan daerah-daerah yang rawan kekerasan. Penelitian ini bertujuan untuk mengembangkan metode klasterisasi berbasis algoritma K-Means dalam upaya meningkatkan segmentasi pola kekerasan, khususnya dalam kasus yang melibatkan perempuan dan anak. Dengan menggunakan data kekerasan dari berbagai wilayah, algoritma K-Means diterapkan untuk mengelompokkan kasus berdasarkan karakteristik tertentu yang relevan. Hasil penelitian menunjukkan bahwa algoritma K-Means memiliki potensi yang kuat dalam segmentasi data kekerasan dan mampu memberikan hasil yang lebih optimal dibandingkan metode lain pada kasus yang dipelajari. Penelitian ini memberikan wawasan baru dalam pengelompokan data sosial menggunakan pendekatan klasterisasi, yang pada akhirnya dapat meningkatkan upaya penanganan kasus kekerasan di berbagai wilayah. Penelitian ini menggunakan tahapan Knowledge Discovery in Database (KDD). Data yang di dapatkan bersumber dari portal website <https://www.kaggle.com/datasets>. Metode *k-means* clustering yang berfungsi untuk memecah dataset menjadi beberapa kelompok / cluster. Berdasarkan hasil dari penelitian ini terdapat 2 *cluster* yaitu *cluster* 0 dengan jumlah anggota 1573 dan *cluster* 1 dengan jumlah anggota 3431. pengukuran *performance* menggunakan DBI, K=2 dengan tingkat *performance* 0,459 maka tingkat *performance* yang terbaik karena tingkat *index* mendekati 0.133

**Abstract.** The increasing number of violence against women and children in various regions has created an urgent need for effective strategies to identify and cluster areas prone to violence. This study aims to develop a clustering method based on the K-Means algorithm in an effort to improve the segmentation of violence patterns, especially in cases involving women and children. Using violence data from various regions, the K-Means algorithm is applied to cluster cases based on certain relevant characteristics. This research is expected to make a significant contribution to the field of data analysis, especially in the context of violence data clustering. The results of the study indicate that the K-Means algorithm has strong potential in violence data segmentation and is able to provide more optimal results than other methods in the cases studied. This study provides new insights into grouping social data using a clustering approach, which can ultimately improve efforts to handle cases of violence in various regions. This study uses the Knowledge Discovery in Database (KDD) stage. The data obtained comes from the website portal <https://www.kaggle.com/datasets>. The *k-means* clustering method functions to break the dataset into several groups / clusters. Based on the results of this study, there are 2 clusters, namely cluster 0 with a total of 1573 members and cluster 1 with a total of 3431 members. Performance measurement using DBI, K = 2 with a performance level of 0.459, the best performance level because the index level is close to 0.133.

## 1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mengalami peningkatan pesat dalam dua dekade terakhir, membawa dampak signifikan pada berbagai aspek kehidupan, termasuk bisnis, pendidikan, dan layanan masyarakat [1]. Identifikasi pola kekerasan dan segmentasi daerah rawan kekerasan memegang peran penting dalam pencegahan dan penanganan kasus kekerasan, khususnya pada perempuan dan anak. Dalam kajian terbaru, isu ini sangat relevan karena meningkatnya jumlah kasus kekerasan yang membutuhkan penanganan secara cepat dan tepat [2].

Namun, keterbatasan model analisis yang ada seringkali menghasilkan klasifikasi yang tidak cukup akurat, terutama ketika menangani data dengan variabilitas tinggi dan distribusi yang tidak seragam. Penelitian ini bertujuan untuk mengatasi kesenjangan tersebut melalui pengembangan dan penerapan algoritma K-Means dalam segmentasi kasus kekerasan yang lebih efektif.

Memberikan perlindungan serta pengawasan pada kasus ini maka dibutuhkan suatu metode untuk mengklusterisasi kasus kekerasan berdasarkan pemetaan wilayah. Adapun pengertian metode algoritma *k-means* adalah teknik dalam analisis *clustering* yang digunakan untuk mengelompokkan data ke dalam kelompok – kelompok yang disebut *cluster* berdasarkan kemiripan dari atribut data tersebut [3]. Dalam konteks ini, algoritma klusterisasi, khususnya K-Means, memiliki aplikasi luas dalam mengelompokkan data berdasarkan karakteristik tertentu. Misalnya, penerapan algoritma K-Means dapat membantu mengidentifikasi pola dalam data kekerasan untuk mendukung pengambilan kebijakan preventif dan reaktif [1]. Penelitian sebelumnya tentang klusterisasi kasus kekerasan menggunakan algoritma K-Means telah dilakukan dengan berbagai pendekatan untuk mengelompokkan kasus kekerasan terhadap anak dan perempuan di daerah tertentu berdasarkan karakteristik korban [4]. mengaplikasikan algoritma K-Means dalam klusterisasi lokasi berdasarkan tingkat kejahatan di suatu wilayah, namun metode ini memiliki

keterbatasan dalam menangani data dengan dimensi tinggi. Keterbatasan-keterbatasan ini membuka peluang untuk penelitian lebih lanjut yang dapat meningkatkan akurasi dan fleksibilitas metode klusterisasi. [5].

Tujuan utama penelitian ini adalah mengembangkan metode klusterisasi menggunakan algoritma K-Means untuk meningkatkan segmentasi pola kekerasan terhadap perempuan dan anak. Algoritma K-Means dipilih karena efektivitasnya dalam mengidentifikasi pola yang sulit terlihat dalam analisis. Penelitian ini diharapkan dapat memberikan kontribusi dalam bidang analisis data dengan menyediakan metode yang lebih akurat dan adaptif untuk segmentasi data kekerasan. Secara praktis, hasil penelitian ini diharapkan dapat mendukung pengambilan keputusan pihak berwenang dalam pencegahan kekerasan dan alokasi sumber daya secara efektif di wilayah yang teridentifikasi sebagai rawan kekerasan.

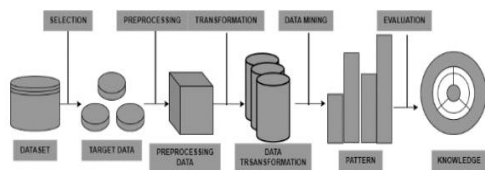
## 2. TINJAUAN PUSTAKA

### 2.1 Data Kekerasan

Analisis ini dapat menghasilkan pengetahuan baru tentang pola, tren, dan faktor-faktor yang menyebabkan kekerasan. Pengetahuan baru ini dapat digunakan untuk menentukan langkah-langkah yang tepat dalam menekan kasus kekerasan di Jawa Barat. Penelitian ini bertujuan untuk melakukan klusterisasi kasus kekerasan di Jawa Barat berdasarkan kelompok usia menggunakan algoritma K-Means Clustering [6].

### 2.2 Clustering

Data mining adalah suatu proses yang mengekstrak informasi berguna dari sejumlah besar data dengan menggunakan berbagai teknik dan algoritma untuk menemukan pola atau pengetahuan yang tersembunyi dalam data, yang membantu membuat keputusan yang lebih baik dan mendapatkan pemahaman yang lebih baik tentang apa yang ada di dalam data [5]. Teknik data mining clustering akan mengelompokkan data dengan karakteristik yang sama dan mengelompokkan data dengan karakteristik yang berbeda ke kelompok lain.



Metode yang digunakan dalam penelitian ini yaitu dengan menggunakan algoritma K-Means yaitu mengelompokkan data berdasarkan titik pusat kluster atau centroid terdekat dengan data [7]. Untuk mengelompokkan kumpulan data menjadi sejumlah kelompok yang terdiri dari entitas atau objek yang memiliki karakteristik yang serupa, teknik data mining yang dikenal sebagai clustering digunakan [8]. Penelitian ini akan memberikan kontribusi dengan menggunakan pendekatan clustering yang lebih canggih yaitu algoritma k-means.

### 2.3 K-Means

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah studi dokumentasi dengan memanfaatkan data sekunder. Data yang digunakan berupa jumlah tingkat kekerasan yang diambil dari platform Kaggle. Berdasarkan deskripsi yang terdapat di situs Kaggle, data ini ditranskripsi dari laporan kejahatan asli yang diketik di atas kertas. Data tersebut mencakup jumlah tingkat kekerasan di kota Los Angeles sejak tahun 2020.[9] Metode clustering yang dievaluasi dengan K-Means dapat memberikan hasil yang optimal dan akurat untuk menyatukan data yang memiliki ciri yang serupa dan berbeda dari yang lain[10].

### 2.4 DBI

Penelitian ini menggunakan teknik kinerja Davies Bouldin Index [11]. Nilai DBI yang rendah, atau sama sekali tidak ada, menunjukkan kualitas kluster yang dihasilkan [12].

## METODE PENELITIAN

Metode penelitian yang diusulkan dalam judul “Algoritma K-Means Untuk Meningkatkan Segmentasi Pola Kekerasan” cenderung bersifat kuantitatif. Penelitian ini menggunakan pendekatan *Knowledge Discovery in Databases* (KDD) untuk

membangun model pengelompokan data kekerasan. Adapun tahapannya sebagai berikut:

### 3.1 Data Selection

Data Kekerasan Kota Los Angeles. Data sekunder yang digunakan dalam penelitian ini berasal dari laporan warga setempat dan ditulis tangan yang memiliki 7500 baris dan dipotong menjadi 5004 dan 6 atribut.

### 3.2 Preprocessing

Preprocessing data, juga dikenal sebagai data cleaning, adalah proses membersihkan data dari elemen yang tidak penting seperti data duplikat, nilai yang hilang, dan outlier. Proses ini dilakukan dengan mengubah tipe data dari sumber data sebelumnya untuk menghapus data duplikat, nilai yang hilang, dan typo.

### 3.3 Data Transformation

Transformasi data adalah proses mengubah jenis data sehingga sesuai dengan analisis yang akan dilakukan. Pada atribut status, data dalam penelitian ini diubah dari non-numerik menjadi numerik.

### 3.4 Data Mining

Tahap awal proses data mining, yang merupakan inti dari proses KDD, dilakukan dengan algoritma clustering K-means. Selain itu, untuk melakukan evaluasi kinerja, Mengenai berapa banyak cluster yang ideal, Davies Bouldin Index (DBI) digunakan. Selanjutnya, proses iteratif dilakukan untuk menghasilkan nilai keanggotaan yang stabil untuk setiap kelompok data mining.

### 3.5 Evaluation

Merupakan proses melihat atau menilai pola yang sudah ada sebelumnya. Ini dilakukan untuk menentukan apakah pola memberikan hasil yang diharapkan dan membantu dalam pengambilan keputusan.

### 3.6 Knowledge

Tahap ini melibatkan analisis menyeluruh terhadap pola, tren, atau hubungan

dalam data. Tujuan dari tahap ini adalah untuk membuat kesimpulan yang relevan dan membuat keputusan yang didukung oleh informasi yang diperoleh dari analisis yang dilakukan pada tahap sebelumnya.

### 3. HASIL DAN PEMBAHASAN

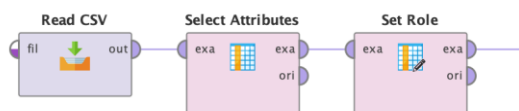
Dalam penelitian ini, metodologi Knowledge Discovery in Databases (KDD) digunakan.

#### 4.1 Data Selection

Atribut pertama adalah Nomor Urut (NO), yang menjelaskan urutan data. Atribut berikutnya adalah Area Name yang menjelaskan nama area dimana kejadian itu terjadi, diikuti oleh CRM CD Desc (Criminal Code Description), yang menjelaskan deskripsi kode kejahatan kriminal, selanjutnya ada Vict Age yang menjelaskan usia dan Vict Sex adalah jenis kelamin “F” artinya Female dan “M” artinya Male, lalu ada date occ yang menjelaskan waktu kejadian. Nilai-nilai ini akan digunakan sebagai dasar dalam proses *clustering* dengan menggunakan algoritma K-Means.

#### 4.2 Preprocessing

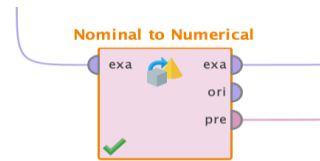
Untuk memastikan bahwa data akurat dan relevan dengan tujuan penelitian, berbagai prosedur dilakukan selama tahap pre-processing data. Berikut ini adalah penjelasan tentang proses pemrosesan data sebelum menggunakannya. Data akan diperiksa untuk nilai yang salah, typo, atau duplikat setelah proses seleksi data.



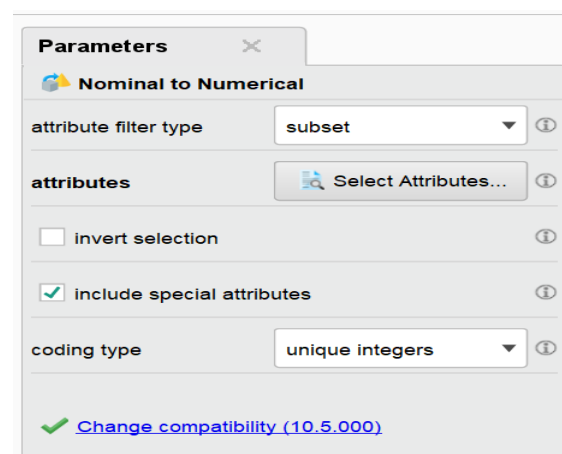
Gambar 4. 1 Operator Read CSV, Select Attributes Dan Set Role.

#### 4.3 Transformation

Selanjutnya, transformasi data dilakukan untuk mengonversi atribut nominal menjadi numerik menggunakan operator Nominal to Numerical. Konversi ini bertujuan agar data kategori dapat dianalisis secara optimal. Gambar 4.2 di bawah ini menunjukkan proses transformasi data.



Gambar 4. 2 Operator Nominal To Numerical menunjukkan penggunaan operator Nominal to Numerical dengan pengaturan parameter tertentu untuk mengonversi atribut nominal menjadi bentuk numerik. Gambar 4.3 di bawah ini menunjukkan parameter dari operator Nominal To

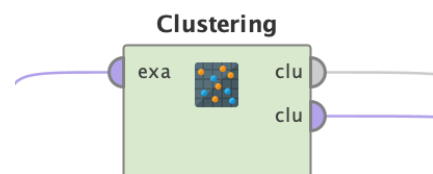


Numerical.

Gambar 4. 3 Parameter Operator Nominal To Numerical.

#### 4.4 Data Mining

Pada penelitian ini, algoritma K-Means digunakan untuk mengelompokkan data kekerasan.



Gambar 4. 4 Algoritma K-Means.

Gambar 4.4 menunjukkan penggunaan operator Algoritma K-Means untuk mengelompokkan data kekerasan. Dalam proses ini, operator Cluster Distance Performance digunakan untuk mengevaluasi kualitas cluster dan menentukan jumlah cluster

(K) terbaik. Pada kasus ini, nilai K diatur sebanyak 10 untuk melihat variasi pengelompokan yang mungkin memberikan hasil optimal.

#### 4.5 Interpretation/Evaluation

Pada tahap evaluasi ini, kluster yang telah dibangun akan diperiksa kualitasnya menggunakan pendekatan DBI ini untuk menentukan kekuatan atau kualitasnya.

#### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 747.169
Avg. within centroid distance_cluster_0: 1014.918
Avg. within centroid distance_cluster_1: 624.415
Davies Bouldin: 0.133
```

Gambar 8. Proses Menampilkan K Terbaik

Dalam perhitungan DBI di atas, kami menemukan nilai 0,133. Menurut aturan DBI, cluster yang dihasilkan lebih baik jika nilai DBI mendekati 0 atau lebih rendah.

#### 4.6 Knowledge

Hasil penelitian menunjukkan bahwa algoritma K-Means memiliki potensi yang kuat dalam segmentasi data kekerasan dan mampu memberikan hasil yang lebih optimal dibandingkan metode lain pada kasus yang dipelajari. Penelitian ini memberikan wawasan baru dalam pengelompokan data sosial menggunakan pendekatan klasterisasi, yang pada akhirnya dapat meningkatkan upaya penanganan kasus kekerasan di berbagai wilayah.

Penelitian ini menggunakan tahapan Knowledge Discovery in Database (KDD). Data yang di dapatkan bersumber dari portal website <https://www.kaggle.com/datasets>. Metode *k-means* clustering yang berfungsi untuk memecah dataset menjadi beberapa kelompok / cluster. Berdasarkan hasil dari penelitian ini terdapat 2 cluster yaitu cluster 0 dengan jumlah anggota 1573 dan cluster 1 dengan jumlah anggota 3431. pengukuran *performance* menggunakan DBI, K=2 dengan tingkat *performance* 0,459 maka tingkat *performance*

yang terbaik karena tingkat *index* mendekati 0.133

Tidak hanya relevansi harus dibahas secara menyeluruh, tetapi pembahasan juga harus menjelaskan bagaimana hasil penelitian dibandingkan dengan temuan yang relevan. Akibatnya, kutipan penting harus dimasukkan ke dalam bagian diskusi. Pada bagian terakhir, hasil penelitian keilmuan harus dibahas.

Dengan menggunakan data kekerasan dari berbagai wilayah, algoritma K-Means diterapkan untuk mengelompokkan kasus berdasarkan karakteristik tertentu yang relevan. Metode ini dipilih karena kemampuannya dalam menangani data berukuran besar serta efisiensi dalam proses komputasi. Sebagai bagian dari analisis, algoritma K-Means juga dibandingkan dengan metode klasterisasi lain, seperti K-Medoids, guna menilai efektivitasnya dalam menghasilkan segmentasi yang lebih akurat. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam bidang analisis data, khususnya dalam konteks klasterisasi data kekerasan.

#### 4. KESIMPULAN

Pada hasil penelitian ini, algoritma K-Means memberikan gambaran yang jelas mengenai pengimplementasian algoritma K-Means untuk meningkatkan segmentasi pola kekerasan. Dua klaster utama terbentuk berdasarkan pola kekerasan yang berbeda. Cluster 0 dengan 1573 anggota, Cluster 1 beranggotakan 3431. Selain itu, nilai Davies-Bouldin Index (DBI) sebesar 0.133 pada hasil clustering ini menunjukkan kualitas klaster yang cukup baik. DBI yang lebih rendah dari 1 mengindikasikan bahwa pemisahan antar klaster sudah cukup baik, meskipun masih ada peluang untuk peningkatan. Berdasarkan hasil ini, dapat disimpulkan bahwa algoritma K-Means berhasil mengelompokkan ribuan data kekerasan menjadi 2 cluster.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada :

1. Cinta pertama dan panutanku, **Bapak H.M.Saprudin P.** Beliau yang mendidik, memotivasi, memberikan dukungan dan

- doa. Hingga penulis mampu menyelesaikan studi sampai sarjana.
2. Pintu surgaku, **Mamah Hj. Neti Roneti**. Terimakasih sebesar-besarnya penulis sampaikan kepada beliau atas segala bentuk bantuan, semangat dan doa yang diberikan selama ini. Terimakasih atas nasehat yang diberikan meski terkadang fikiran kita tidak sejalan. Terimakasih atas kesabaran hati menghadapi penulis yang keras kepala ini. Terimakasih sudah menjadi penguat dan pengingat paling hebat.
  3. Kepada cinta kasih ketiga kakak- kakakku, **Sri Yuliati Prihandini, S.Pd.I.** , **Nurhidayah Tinia Lestari ST., MT.** , **Muhammad Irhamunnai'em, S.Pd.** Terimakasih atas kasih sayang, doa dan motivasi yang telah diberikan kepada adik terakhir ini.
  4. Kepada Pemilik **NIM 41215664**, yang juga sebentar lagi akan menyandang gelar yang sama. Terimakasih sudah banyak membantu dalam proses penulisan skripsi ini, menjadi tempat bertukar fikiran, dan tempat berkeluh kesah. Terimakasih sudah memberikan keluarga yang kehangatan dan supportnya sangat berarti bagi penulis.
  5. Kepada teman-teman terdekat semasa kuliah yang tidak bisa saya sebutkan satu persatu. Terimakasih sudah menemani penulis dari masa awal kuliah sampai penulisan skripsi ini selesai. Terimakasih atas kebersamaan yang tak ternilai harganya.
- ### DAFTAR PUSTAKA
- [1] R. Fauziah and A. I. Purnamasari, "Implementasi Algoritma K-Means pada Kasus Kekerasan Anak dan Perempuan Berdasarkan Usia," *Hello World J. Ilmu Komput.*, vol. 2, no. 1, pp. 34–41, 2023, doi: 10.56211/helloworld.v2i1.232.
  - [2] L. Awaliyah, N. Rahaningsih, and R. Danar Dana, "Implementasi Algoritma K-Means Dalam Analisis Cluster Korban Kekerasan Di Provinsi Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 188–195, 2024, doi: 10.36040/jati.v8i1.8332.
  - [3] M. Veronika, B. Tarigan, I. M. S. S, and P. S. M. Lumbanraja, "Penerapan Algoritma K-Medoids Untuk Mengelompokkan Penyebaran Tindak Kriminal Di Wilayah Hukum Polres Tanah Karo," vol. 3, no. 1, pp. 77–86, 2023.
  - [4] N. Adawiyah, N. Sulistiyowati, and M. Jajuli, "Klasterisasi Kasus Kekerasan Terhadap Anak dan Perempuan Berdasarkan Algoritma K-Means," *Gener. J.*, vol. 5, no. 2, pp. 69–80, 2021, doi: 10.29407/gj.v5i2.15995.
  - [5] Fl. F. W. Daeli, D. Nofriansyah, and S. Yakub, "Implementasi Data Mining Untuk Mengelompokkan Lokasi Berdasarkan Tingkat Kejahatan Pada Kabupaten Nias Barat Menggunakan Metode K-Means Clustering," *Sains dan Komput.*, pp. 1–14, 2020.
  - [6] M. Riskandi, M. Martanto, and U. Hayati, "Klasterisasi Korban Kekerasan Menggunakan Algoritma K-Means Di Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 820–826, 2024, doi: 10.36040/jati.v8i1.8457.
  - [7] Y. Sopyan, A. D. Lesmana, and C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1464–1470, 2022, doi: 10.47065/bits.v4i3.2697.
  - [8] R. Buaton, Y. Sundari, and Y. Maulita, "Clustering Tindak Kekerasan Pada Anak Menggunakan Algoritma K-Means Dengan Perbandingan Jarak Kedekatan Manhattan City Dan Euclidean," *MEANS (Media Inf. Anal. dan Sist.*, vol. 1, no. 2, pp. 47–53, 2016, doi: 10.54367/means.v1i2.8.
  - [9] N. Azizah, A. Fauzi, T. Rohana, and S. Faisal, "Perbandingan Metode K-Means dan K-Medoids Untuk Clustering Jenis Kriminalitas," vol. 6, no. 2, pp. 1011–1019, 2024, doi: 10.47065/bits.v6i2.5723.
  - [10] Ni Putu Viona Viandari, I Made Agus Dwi Suarjaya, and I Nyoman Piarsa, "Pemetaan Pelanggan dengan LRFM dan Two Stage Clustering untuk Memenuhi Strategi Pengelolaan," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 130–139, 2022, doi: 10.29207/resti.v6i1.3778.
  - [11] M. Sholeh and K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 8, no. 1, p. 56, 2023, doi: 10.30998/string.v8i1.16388.
  - [12] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021.

