



Comparison of K-Means and K-Medoids Algorithms for Temperature Grouping in Riau Province

Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Suhu di Provinsi Riau

Fahikra Faisal¹, Luna Aliska Giopani², Ma'idatul Fitriah³, Zaira Cindya Dwyne⁴, Siti Syahidatul Helma⁵, Mustakim⁶

^{1,2,3,4} Information System, Faculty of Science and Technology

^{1,2,4,5,6} Puzzle Research Data Technology (Predatech), Faculty of Science and Technology
Sultan Syarif Kasim State Islamic University, Pekanbaru, Indonesia

E-Mail: ¹fahikra23@gmail.com, ²lunaaliskagiopani@gmail.com, ³maidatulfitria2002@gmail.com,
⁴wynnezaira@gmail.com, ⁵syahidah313@gmail.com, ⁶mustakim.telematika@gmail.com.

Makalah: Diterima 29 Agustus 2022; Diperbaiki 5 September 2022; Disetujui 11 September 2022
Corresponding Author: Siti Syahidatul Helma

Abstrak

Indonesia merupakan negara yang beriklim tropis dimana setiap daerahnya memiliki suhu yang berbeda-beda. Suhu di Indonesia memiliki tingkatan mulai dari panas, sedang, dingin dengan suhu yang selalu berubah-ubah setiap bulan bahkan setiap harinya, peneliti melakukan penelitian dalam pengelompokan data suhu di Provinsi Riau yang bersumber dari Badan Pusat Statistik (BPS) Riau mulai dari tahun 2019, 2020, 2021. Penelitian ini dilakukan untuk membandingkan algoritma yang digunakan serta mengetahui algoritma yang tepat untuk pengelompokan suhu berdasarkan hasil uji validitas DBI dimana semakin kecil nilai DBI maka semakin baik cluster yang dihasilkan. Berdasarkan hasil percobaan, algoritma K-Means memperoleh nilai DBI terbaik sebesar 0,2 pada K=6 dengan 10 iterasi sedangkan Algoritma K-Medoids memperoleh nilai terbaik sebesar 0,279 pada K=8 dengan 100 iterasi. Hal ini mengartikan bahwa algoritma K-Means lebih baik jika dibandingkan dengan K-Medoids dalam pengelompokan suhu di Provinsi Riau. Penelitian ini diharapkan dapat memberikan informasi mengenai segmentasi suhu di Provinsi Riau kepada *stakeholder* sebagai bahan pertimbangan atau analisis suhu di waktu yang akan datang.

Keyword: DBI, K-Means, K-Mendoids, Suhu.

Abstract

Indonesia is a country with a tropical climate where each region has different temperatures. Temperatures in Indonesia have levels ranging from hot, medium, cold or minimum, medium and maximum. With temperatures that always change every month and even every day, researchers conducted research in grouping temperature data in Riau Province sourced from the Riau Central Statistics Agency (BPS) starting from 2019, 2020, 2021. The purpose of this research was to facilitate the Agency Meteorology, Climatology and Geophysics (BMKG) in Riau Province. This research was conducted to compare the algorithms used and to find out the right algorithm for temperature grouping. Based on the results of research using the K-Means and K-Medoids algorithm, the best value was obtained as evidenced by the K-Means DBI in the k=6 experiment with a value of 0.2. This means that the K-Means algorithm is better than K-Medoids in temperature grouping in Riau Province.

Keyword: DBI, K-Means, K-Mendoids, Temperature.

1. Pendahuluan

Skala satuan mengenai panas dan dinginnya benda disebut dengan suhu. Temperatur suhu dapat bertukar dari temperatur yang tinggi ke temperatur yang lebih rendah. Indonesia menggunakan derajat Celcius (C) untuk

memaparkan tingkatan suhu[1]. Indonesia tercatat sebagai negara yang beriklim tropis, yang memiliki suhu hangat dan kelembaban yang cukup tinggi [2]. Riau adalah salah satu provinsi di Indonesia yang beriklim tropis. Provinsi Riau mempunyai suhu udara tertinggi antara 35,10 Co dan terendah antara 21,80 Co. Kawasan Provinsi Riau termasuk daerah yang tidak mudah terbakar karena indikator titik panas antara 0 - 330 (rendah - tinggi). Akan tetapi, setiap tiga bulan ditemukan banyak titik api yang menyebabkan bencana kebakaran di provinsi Riau. Sementara itu, intensitas curah hujan 1700 mm - 4000 mm / Tahun. Menurut Forecaster on Duty dari Badan Meteorologi, Klimatologi dan Geofisika (BMKG) Riau menyatakan salah satu faktor penyebab suhu ekstrim di Provinsi Riau karena gangguan tekanan rendah dilaut selat cina disebelah pulau Kalimantan bagian utara. Banyaknya faktor yang mempengaruhi perubahan suhu. Sehingga temperature suhu selalu berubah tidak menentu.

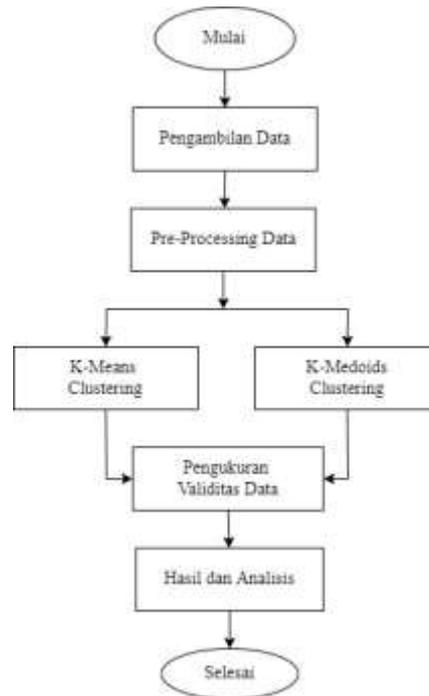
Penelitian ini mengacu pada penelitian yang dilakukan oleh Rahmawati,dkk pada tahun 2019 yang menerapkan salah satu algoritma Data Mining yaitu Fuzzy K-Means untuk mengelompokan Suhu menjadi tiga kategori di Kota Pekanbaru menggunakan data pada tahun 2016-2018. Hasil penelitian menunjukkan bahwa *cluster* pertama dikategorikan bersuhu tinggi terdapat 9 bulan ditahun 2016, 2 bulan ditahun 2017, dan 3 bulan ditahun 2018. *Cluster* kedua dikategorikan bersuhu sedang, terdapat 1 bulan ditahun 2016, pada tahun 2017 juga terdapat 1 bulan, sementara terdapat 5 bulan pada tahun 2018. *Cluster* ketiga dikategorikan bersuhu terendah berturut-turut pada tahun 2016-2018 terdapat 2 bulan, 8 bulan dan 4 bulan [1]. Data mining adalah bagian dalam pengelompokan data dengan menghubungkan pola pada setiap *dataset* yang berukuran besar melalui kuantitas data yang besar pula [3]. Teknik data mining dalam pengelompokan kumpulan data menjadi beberapa kelompok yang memiliki kesamaan disebut dengan *clustering*. K-Means dan K-Medoids merupakan salah satu algoritma bagian dalam *clustering*[4]. K-Means merupakan pengklasteran data yang diperoleh melalui proses berulang berdasarkan jarak data dengan pusat *cluster* (centroid) terdekat. Sedangkan K-Medoids adalah objek yang dipilih dari pusat *cluster* untuk mewakili himpunan objek [11].

Kamila, dkk (2019) melakukan Perbandingan Algoritma *Clustering* pada Transaksi Bongkar Muat selama tahun 2017 di Provinsi Riau. Penelitian ini menerapkan algoritma K-Means dan K-Medoids kemudian divalidasi menggunakan *Davies Boldin Index* (DBI) hasilnya algoritma K-Means lebih baik dengan nilai 0,112 sementara K-Medoids bernilai 0,119 [5]. Farahdinna, dkk tahun 2019 melakukan penelitian mengenai algoritma *clustering* dengan metode K-Means dan K-Medoids dalam pengelompokan Produk Asuransi Perusahaan Nasional. Penelitian ini menggunakan 3 atribut yaitu premi, jumlah nasabah dan tahun rilis pada setiap produk. Hasilnya adalah penggunaan metode K-Means lebih optimal dari pada metode K-Medoids dilihat dari nilai DBI terendah 0,018 pada k=5 [6]. Penelitian lainnya, Harahap F, pada tahun 2021, tentang *Clustering* siswa Tunagrahita dengan membandingkan Algoritma K-Means dan K-Medoids Penelitian ini menghasilkan 3 jenis *cluster*. *Cluster* tunagrahita ringan, tunagrahita sedang dan tuna grahita berat. Penelitian ini juga menghasilkan *Clustering* yang lebih baik jika menggunakan metode K-Means untuk pengelompokan data Siswa Tunagrahita daripada metode K-Medoids *Clustering*, karena memperoleh nilai DBI yang lebih kecil yaitu 0,16 dimana semakin kecil nilai DBI maka semakin baik *cluster* yang dihasilkan [3].

Berdasarkan beberapa penelitian terdahulu, Penelitian ini akan membandingkan algoritma clustering K-Means dan K-Medoids pada *dataset* suhu di Provinsi Riau tahun 2019-2021. Sebagai acuan pengelompokan *cluster* akan dilakukan validasi dan evaluasi menggunakan nilai DBI. Penelitian ini diharapkan mampu memberikan informasi mengenai segmentasi suhu di Provinsi Riau kepada *stakeholder* sebagai bahan pertimbangan atau analisis suhu diwaktu yang akan datang.

2. Bahan dan Metode

Tahap awal pada penelitian ini adalah melakukan perolehan data dari Badan Pusat Statistik (BPS) Riau yang tersedia pada tautan berikut <https://riau.bps.go.id/indicator/151/138/1/suhu.html>. Data yang diperoleh berisikan data suhu di provinsi Riau pada tahun 2019, 2020, 2021. Algoritma yang digunakan untuk pengelompokan data pada penelitian ini adalah algoritma K-Means dan K-Medoids. Perhitungan dilakukan menggunakan aplikasi rapid miner. Data terlebih dahulu di normalisasikan menggunakan tools phyton sebelum diproses lebih lanjut menggunakan algoritma clustering. Kerangka kerja penelitian ini dibuat secara sistematis. Tahapan proses yang akan ditempuh pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1 in-Max Normalization

Min-Max normalization ialah metode yang menormalisasikan data dengan mentransformasikan linear terhadap data asli yang akan menghasilkan nilai perbandingan antara data sebelum proses dan data yang sudah diproses. Persamaan untuk MMN dapat ditunjukkan pada persamaan 1 [7].

$$normalized_{(x)} = \frac{minRange + (x - minValue)(maxRange - minRange)}{maxValue - minValue} \quad (1)$$

2.2 Clustering

Clustering ialah metode yang mengelompokkan kumpulan data kedalam beberapa golongan atau cluster kelas yang memiliki kesamaan objek atau mempunyai kemiripan cluster yang tinggi ke dalam wilayah yang sama dan data yang tidak mempunyai kemiripan objek akan di masukkan wilayah yang berbeda [8][4][3][9]. Metode clustering memiliki fungsi pengelompokan data berdasarkan karakteristik lalu memperkirakan jarak antar data yang memiliki kemiripan dalam satu kelompok walaupun setiap clustering memiliki kelebihan dan kekurangan masing-masing [10]. Didalam proses clustering hal pertama dilakukan adalah mengidentifikasi data yang akan diklasterisasikan menggunakan persamaan 2 [6].

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

Keterangan

D_{ij} = jarak antara I pusat data cluster j

X_{ki} = data i pada atribut data k

X_{kj} = titik pusat j ke atribut k

2.3 K-Means

K-means ialah salah satu metode dari data mining yang mempunyai fungsi proses dengan permodelan tanpa variable kontrol serta menjadi metode pengelompokkan data dalam partisi [11]. K-means juga menjadi algoritma clustering yang membagi setiap objek yang ada ke dalam masing-masing cluster [12]. Data yang dikumpulkan didalam satu cluster adalah data yang mempunyai kemiripan. Untuk menetapkan titik pusat cluster pada data dilakukan dengan menggunakan persamaan 3 [13].

$$C_{kj} = \frac{x_{1j} + x_{2j} + \dots + x_{nj}}{n} \quad (3)$$

Dimana

C_{kj} = pusat *cluster* k pada variabel j (j = 1,2,...,p)

n = banyak data pada *cluster* k

2.4 K-Medoids

Algoritma K-Medoids atau Partitioning Around Medoids (PAM) merupakan salah satu metode clustering yang menggunakan perwakilan (medoid) untuk membagi himpunan (n) beberapa objek menjadi(k) cluster [14][3][15]. Objek yang dipilih dari beberapa pusat cluster disebut medoid [15]. Metode K-Medoids akan mengukur jarak kemiripan pada objek medoids dengan objek non-medoids untuk membuat sebuah cluster [14]. Untuk mengalokasikan data ke dalam cluster yang paling dekat digunakan ukuran jarak Euclidean Distance dengan persamaan 4 [11].

$$\sum_{i=1}^n (x_i - y_i)^2 ; 1,2,3, \dots, n \tag{4}$$

2.5 Davies-Bouldin Index (DBI)

David L. Davies W. Bouldin memperkenalkan DBI pada tahun 1979 yang bertujuan untuk mempertimbangkan algoritma serta memastikan kuantitas cluster yang lebih unggul sesudah pengklasteran selesai.[16][17]. Ini merupakan kerangka penilaian internal, pada tahap validasi menilai keunggulan clustering yang telah dilakukan menggunakan jumlah dan fitur pada dataset [15]. Pendekatan DBI memiliki tujuan memaksimalkan jarak antara kluster dengan kluster yang berbeda dan mengusahakan meminimalisasikan jarak antar objek dalam kluster. Semakin rendah nilai DBI (non-negatif ≥ 0) maka semakin baik kluster yang didapatkan dari pengelompokan K-Modes clustering yang telah dipakai [18]. Untuk mencari nilai DBI digunakan persamaan 5 [19].

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_i, j, \dots, k) \tag{5}$$

Keterangan

$R_{i,j}$ = ratio jumlah cluster

3. Hasil dan Pembahasan

Data yang digunakan bersumber dari situs web Badan Pusat Statistic Provinsi Riau (BPS) mulai dari tahun 2019 hingga 2021. Algoritma terbaik dapat diketahui setelah melalui perbandingan nilai DBI.

3.1 Praproses Data

Data suhu di normalisasi karena semua data atau persamaan sudah numeric. Adapun data suhu yang telah di normalisasi disajikan pada tabel 1.

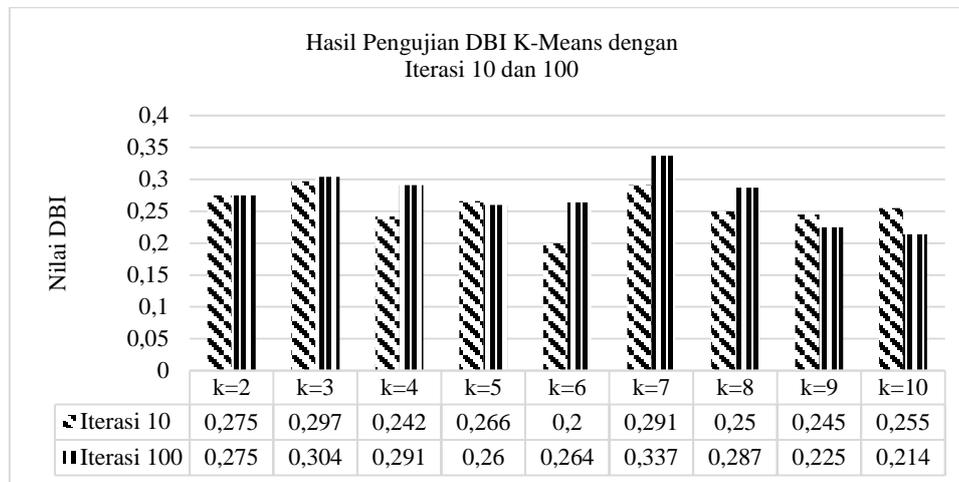
Table 1. Hasil Normalisasi Data Suhu

TAHUN	BULAN	MAKS	MIN	SEDANG
2019	JANUARI	0,633333333	0,555555556	0,44
	FEBRUARI	0,583333333	0,277777778	0,64
	MARET	0,866666667	0,444444444	0,84
2020
	JANUARI	0,5	0,5	0,56
	FEBRUARI	0,616666667	0,583333333	0,68
2021	MARET	0,783333333	0,666666667	0,84

	JANUARI	0	0,722222222	0
2021	FEBRUARI	0,25	0,777777778	0,44
	MARET	0,216666667	0,75	0,28

3.2 Hasil Pengujian DBI K-Means

Setelah dilakukannya normalisasi pada data suhu, tahapan selanjutnya pengolahan data untuk menunjukkan nilai DBI sebagai model pengelompokan cluster terbaik. Pengujian dilakukan pada cluster $k=2$ sampai dengan $k=10$ dengan iterasi 10 dan 100. Untuk mengetahui grafik hasil pengujian disajikan pada Gambar 1.

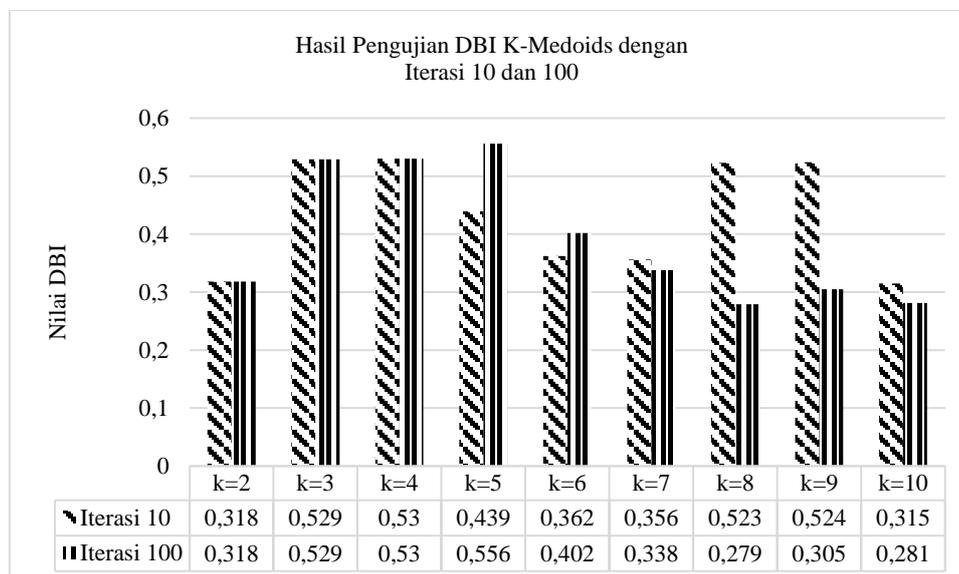


Gambar 1. Hasil Pengujian DBI K-Means dengan iterasi 10 dan 100

Berdasarkan Gambar 1. dapat dilihat bahwa nilai yang sesuai dengan DBI terletak pada percobaan $k=6$ dengan nilai 0,2 pada iterasi 10.

3.3 Hasil Pengujian DBI K-Medoids

Setelah dilakukannya pengujian DBI K-Means, tahapan selanjutnya yakni mencari hasil pengujian DBI K-Medoids dimulai dari percobaan $k=2$ hingga $k=10$ dengan iterasi 10 dan 100. Hasil pengujian tersebut terdapat pada Gambar 2.

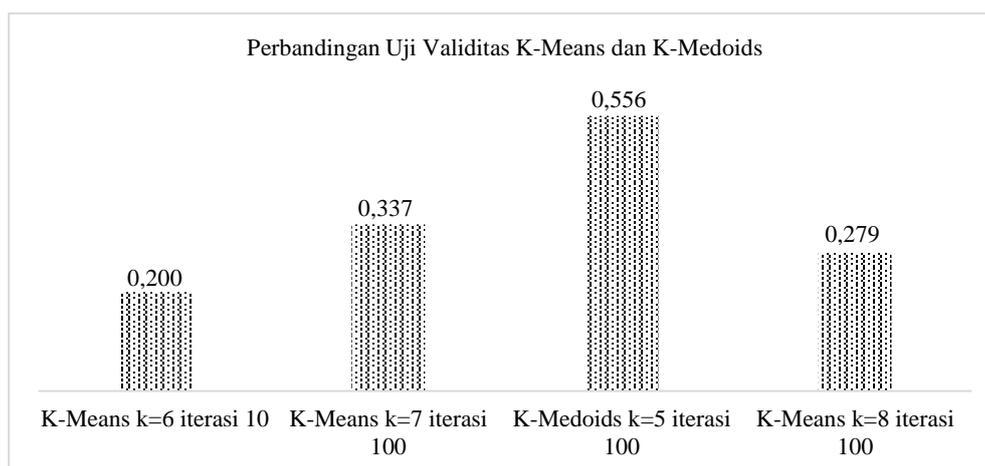


Gambar 2. Hasil Pengujian DBI K-Medoids dengan iterasi 10 dan 100

Berdasarkan Gambar 2 dapat dilihat bahwa nilai yang sesuai dengan DBI terletak pada percobaan $k=8$ dengan nilai 0,279 pada iterasi 100.

3.4 Perbandingan Uji Validitas K-Means dan K-Medoids

Untuk mengetahui perbandingan nilai DBI terbaik disajikan pada Gambar 3.



Berdasarkan diagram batang pada Gambar 3 dapat disimpulkan bahwa nilai terendah pada DBI K-Means terdapat pada percobaan k=6 iterasi 10 sedangkan nilai tertinggi pada DBI K-Medoids terdapat pada percobaan k=5 iterasi 100.

4. Kesimpulan

Berdasarkan hasil analisis dan pengolahan data menggunakan DBI terhadap algoritma K-Means dan K-Medoids. Maka diperoleh *cluster* terbaik di algoritma K-Means pada k=6 dengan nilai 0,2 dibandingkan dengan algoritma K-Medoids. Sehingga dalam penelitian ini menerapkan algoritma K-Means dalam pengelompokan suhu di Provinsi Riau. Hasil algoritma K-Means membagi data menjadi 6 *cluster* yaitu 10 bulan pada *cluster* 1, 10 bulan pada *cluster* 2, 1 bulan pada *cluster* 3, 1 bulan pada *cluster* 4, 3 bulan pada *cluster* 5, dan 11 bulan pada *cluster* 6.

References

- [1] S. Saputri and A. N. Rahma, "Pengelompokan Suhu Di Kota Pekanbaru Menggunakan Metode Fuzzy K-Means," no. November, pp. 505–512, 2019.
- [2] S. Wirohamidjojo and Y. S. Swarinoto, "Iklim kawasan Indonesia," *Jakarta Badan Meteorol. dan Geofis.*, 2010.
- [3] F. Harahap, "Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita," vol. 2, no. 4, pp. 191–197, 2021.
- [4] S. F. Octavia, "Penerapan K-Means dan Fuzzy C-Means untuk Pengelompokan Data Kasus Covid-19 di Kabupaten Indragiri Hilir," vol. 3, no. 2, pp. 88–94, 2021, doi: 10.47065/bits.v3i2.1005.
- [5] I. Kamila *et al.*, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," vol. 5, no. 1, pp. 119–125, 2019.
- [6] P. Algoritma and K. D. A. N. K. Dalam, "Perbandingan algoritma k-means dan k-medoids dalam klasterisasi produk asuransi perusahaan nasional," vol. XI, no. 2, pp. 208–214, 2019.
- [7] "PERBANDINGAN NORMALISASI DATA UNTUK KLASIFIKASI WINE MENGGUNAKAN ALGORITMA K-NN," vol. 4, no. 1, pp. 78–82, 2019.
- [8] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.
- [9] D. T. Kusuma and N. Agani, "Prototipe Komparasi Model Clustering Menggunakan Metode K-Means Dan FCM untuk Menentukan Strategi Promosi: Study Kasus Sekolah Tinggi Teknik-PLN Jakarta," *J. TICom*, vol. 3, no. 3, p. 93460, 2015.
- [10] Y. hilda Susanti, "PERBANDINGAN K-MEANS DAN K-MEDOIDS CLUSTERING TERHADAP KELAYAKAN PUSKESMAS DI DIY TAHUN 2015," 2017.
- [11] P. Penyebaran *et al.*, "Analisis algoritma k-medoids clustering dalam pengelompokan penyebaran covid-19 di indonesia," vol. 4, no. 1, pp. 166–173, 2020.
- [12] H. Firdaus and A. Sofro, "Analisa Cluster Menggunakan K-Means Dan Fuzzy C-Means Dalam Pengelompokan Provinsi Menurut Data Intesitas Bencana Alam Di Indonesia Tahun 2017-2021," *MATHunesa J. Ilm. Mat.*, vol. 10, no. 1, pp. 50–60, 2022, doi: 10.26740/mathunesa.v10n1.p50-60.
- [13] D. R. Ningrat, D. A. I. Maruddani, and T. Wuryandari, "Analisis cluster dengan algoritma K-Means dan Fuzzy C-Means clustering untuk pengelompokan data obligasi korporasi," *None*, vol. 5, no. 4, pp. 641–650, 2016.
- [14] D. Suyanto, "Data Mining untuk klasifikasi dan klasterisasi data," *Bandung Inform. Bandung*, 2017.

- [15] D. Listiyanti, Y. A. Syahbana, and S. R. Henim, "Perancangan Dan Implementasi Aplikasi Android Penentu Salient Area Pada Video Dengan Algoritma," in *Annual Research Seminar*, 2016, vol. 2016, pp. 96–101.
- [16] V. A. P. Sangga, "Perbandingan algoritma K-Means dan algoritma K-Medoids dalam pengelompokan komoditas peternakan di provinsi Jawa Tengah tahun 2015," 2018.
- [17] A. F. Khairati, A. A. Adlina, G. F. Hertono, and B. D. Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA," in *PRISMA, Prosiding Seminar Nasional Matematika*, 2019, vol. 2, pp. 161–170.
- [18] F. Tempola, M. Muhammad, and A. Mubarak, "Penggunaan Internet Dikalangan Siswa SD di Kota Ternate: Suatu Survey, Penerapan Algoritma Clustering dan Validasi DBI," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1153–1160, 2020.
- [19] S. Butsianto and N. Saepudin, "Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 1, pp. 51–59, 2020, doi: 10.32672/jnkti.v3i1.2008.