

ALGORITMA *K-MEDOIDS CLUSTERING* UNTUK MENGELOMPOKKAN TINGKAT KEMISKINAN PADA KABUPATEN DAN KOTA DI KEPULAUAN MALUKU DAN PAPUA

(K-Medoids Clustering Algorithm for Classification of Poverty Levels at Districts and Cities in Maluku Islands and Papua)

G. Haumahu^{1*}, M. Y. Matdoan², Samsul B. Loklomin

^{1,2,3}Program Studi Statistika, FMIPA Universitas Pattimura,
Jl. M. J. Putuhena, Ambon, 97233, Maluku, Indonesia

e-mail: ghaumahu@gmail.com

Abstrak: Kepulauan Maluku dan Papua merupakan daerah yang memiliki tingkat kemiskinan tinggi jika dibandingkan dengan kepulauan lain di Indonesia. Penelitian ini bertujuan untuk mengelompokkan tingkat kemiskinan di Kepulauan Maluku dan Papua dengan menggunakan algoritma *k-medoids clustering*. Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS). Diperoleh hasil bahwa terdapat 4 kelompok dalam clusterisasi tingkat kemiskinan di Kepulauan Maluku dan Papua. *Cluster 0* terdiri atas 15 kabupaten/kota, *cluster 1* terdiri atas 18 kabupaten/kota, *cluster 2* terdiri atas 8 kabupaten/kota dan *cluster 3* terdiri atas 22 kabupaten/kota.

Kata Kunci: K-Medoids, Kemiskinan, Maluku, Papua.

Abstract: Maluku and Papua Islands are areas that have a high poverty rate when compared to other islands in Indonesia. This study aims to classify poverty levels in Maluku Islands and Papua using the *k-medoids clustering* algorithm. This study uses secondary data sourced from the Central Bureau of Statistics (BPS). The results show that there are 4 groups in the poverty level clustering in the Maluku and Papua islands. *Cluster 0* consists of 15 regencies/cities, *cluster 1* consists of 18 regencies/cities, *cluster 2* consists of 8 regencies/cities and *cluster 3* consists of 22 regencies/cities.

Keywords: K-Medoids, Poverty, Maluku, Papua.

1. PENDAHULUAN

Kemiskinan adalah kondisi seseorang atau sekelompok orang yang tidak mampu memenuhi kebutuhan dasarnya, antara lain pangan, kesehatan, pendidikan, pekerjaan, perumahan, air bersih, tanah, sumber daya alam, lingkungan hidup dan rasa aman [1]. Berdasarkan data dari BPS tahun 2022 menunjukkan bahwa Kepulauan Maluku dan Papua merupakan daerah yang memiliki tingkat kemiskinan tinggi jika dibandingkan dengan kepulauan lain di Indonesia. Persentase penduduk miskin di Provinsi Papua sebesar 26,56%, Provinsi Papua Barat sebesar 21,33%, Provinsi Maluku sebesar 15,97%, dan Provinsi Maluku Utara sebesar 6,23%. Selain itu, berdasarkan data ketertinggalan daerah menunjukan bahwa terdapat 36 dari 63 kabupaten/kota masih dikategorikan sebagai daerah tertinggal [2]. Oleh karena itu, perlu menjadi perhatian utama pemerintah guna meminimalisir angka kemiskinan dan kesenjangan di dalam masyarakat.

Data mining merupakan proses analitis yang dirancang untuk memeriksa data dalam jumlah besar sehingga diperoleh informasi tersembunyi yang berharga dan konsisten. *Data mining* memiliki manfaat diantaranya yaitu dapat dilakukan untuk deskripsi, estimasi, prediksi, klasifikasi, *clustering*, dan asosiasi

[3]. *Data mining* dapat disebut juga sebagai *Knowledge Discovery Databases* (KDD) yaitu tahapan yang dilakukan dalam menggali pengetahuan dari sekumpulan data [4] [5].

Analisis *cluster* merupakan salah satu metode dalam *data mining* yang dapat digunakan untuk mengelompokkan sekumpulan objek (seri data) menjadi beberapa *cluster* berdasarkan kesamaan sifat atribut yang dimiliki oleh objek data tersebut, sehingga objek data tersebut berada dalam satu *cluster* yang sama [6] [7].

Algoritma K-Medoids atau sering dikenal sebagai *Partitioning Around Medoids* (PAM), merupakan algoritma yang menerapkan objek sebagai perwakilan di setiap *cluster* yaitu medoid [8]. Algoritma K-Medoids dapat meminimalkan jumlah perbedaan antara titik data dalam *cluster* dengan titik data terpilih di setiap *cluster* sebagai pusat (medoid). K-Medoids memiliki karakteristik dimana pusat *cluster* berada di antara titik-titik datanya [9] [10].

Algoritma K-Medoids adalah algoritma yang mirip dengan algoritma K-Means karena keduanya merupakan sub-algoritma yang mempartisi dataset menjadi beberapa grup (kelompok) [11]. Perbedaannya yaitu pada penentuan pusat *cluster*, dimana algoritma K-Means menggunakan rata-rata (*mean*) dari setiap *cluster* sebagai pusat *cluster*, sedangkan algoritma K-Medoids menggunakan objek data untuk perwakilan (medoids) sebagai pusat *cluster* [8]. Kelebihan algoritma K-Medoids dapat mengatasi kelemahan dari K-Means, yaitu sifat sensitif terhadap *noise* dan *outlier* serta menyebabkan objek besar menyimpang dari distribusi data. Keuntungan lain dari K-Medoids yaitu hasil dari proses *clustering* tidak bergantung pada urutan dataset yang dimasukkan. Selain itu, berdasarkan nilai validitas algoritma K-Medoids lebih baik dalam melakukan *clustering* data jika dibandingkan algoritma K-Means [12].

2. METODOLOGI

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS). Unit sampel yang digunakan dalam penelitian ini bersumber dari kabupaten dan kota di Provinsi Maluku, Maluku Utara, Papua, dan Papua Barat. Variabel yang digunakan dalam penelitian ini terdiri dari variabel Kepadatan Penduduk (X1), Jumlah Angkatan Kerja (X2), Tingkat Partisipasi Angkatan Kerja (X3), Tingkat Pengangguran Terbuka (X4), Rata-Rata Lama Sekolah (X5), Angka Harapan Hidup (X6), Pengeluaran Per kapita (X7) dan variabel Persentase Penduduk Miskin (X8).

Analisa data dalam penelitian ini terdiri atas statistika deskriptif kemudian dilanjutkan dengan algoritma *K-Medoids Clustering*. Algoritma K-Medoids merupakan metode yang dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw pada tahun 1987. Metode ini merupakan teknik partisi *clustering* klasik yang mengklasifikasikan n objek dataset ke dalam kelompok k yang diketahui secara apriori. Metode ini menggunakan objek dari sekumpulan objek untuk mewakili sebuah *cluster*. Objek yang dipilih untuk mewakili sebuah *cluster* disebut medoid [13] [14]. Tahapan-tahapan algoritma K-Medoids sebagai berikut [15].

1. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*).
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)} \quad (1)$$

dengan

$$i = 1 \cdots n$$

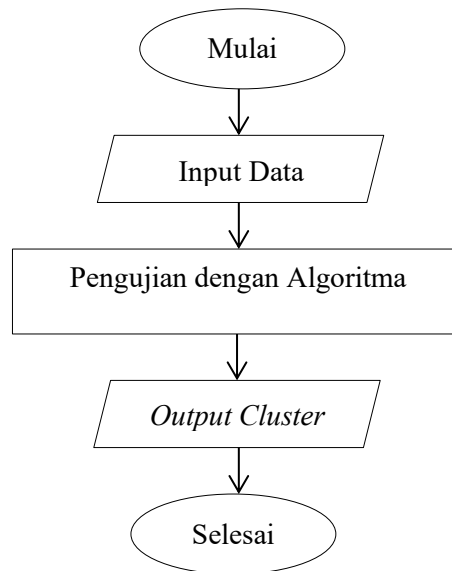
$$j = 1 \cdots n$$

$$p = \text{banyaknya variabel}$$

V = matriks varian kovarian

Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat medoid baru.

3. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat medoid baru.
4. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai medoid.
5. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoid*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

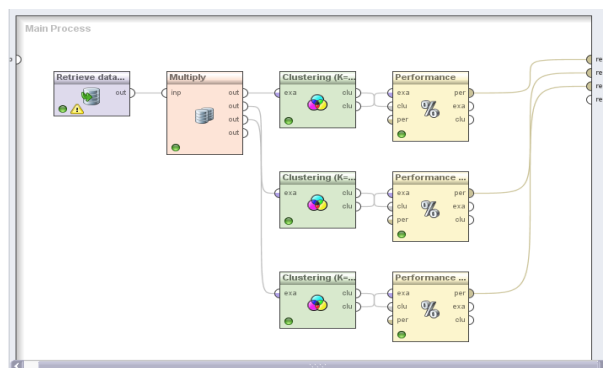


Gambar 1. Prosedur Algoritma K-Medoids

3. HASIL DAN PEMBAHASAN

a. Pengujian Menggunakan Rapdminer

Proses pengujian data ini dilakukan dengan menggunakan aplikasi Rapidminer. Pengujian dilakukan menggunakan algoritma k-medoids seperti digambarkan pada Gambar 2.



Gambar 2. Pemodelan *Clustering K-Medoids*

b. *Davies Bouldin Index (DBI)*

Davies Bouldin Index (DBI) digunakan untuk melakukan validasi data pada setiap *cluster*. Pengukuran menggunakan DBI bertujuan untuk memaksimalkan jarak *inter-cluster*. Dengan menggunakan DBI suatu *cluster* akan dianggap memiliki skema *clustering* yang optimal jika yang memiliki *Index Davies* minimum. Hasil pengujian *Davies Bouldin* dengan $k = 2$, $k = 3$ dan $k = 4$ dapat dilihat pada Tabel 1.

Tabel 1. Hasil Pengujian *Davies Bouldin*

<i>Cluster (K)</i>	<i>Davies Bouldin</i>
2	3,764
3	2,432
4	-0,942

Berdasarkan Tabel 1, ditunjukkan bahwa nilai *Davies Bouldin* untuk $k = 2$ sebesar 3,767; $k = 3$ sebesar 2,432 dan $k = 4$ sebesar -0,492. Oleh karena itu, dapat disimpulkan bahwa penelitian ini menggunakan $k=4$, karena memiliki nilai *Davies Bouldin* yang paling minimum.

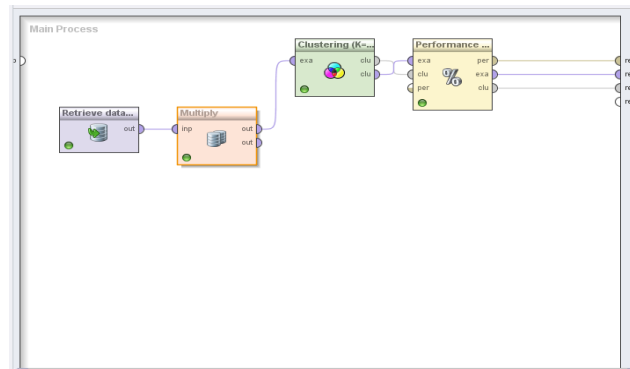
PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -323186024.525
Avg. within centroid distance_cluster_0: -218080343.032
Avg. within centroid distance_cluster_1: -77367440.131
Avg. within centroid distance_cluster_2: -1767802955.442
Avg. within centroid distance_cluster_3: -70658037.894
Davies Bouldin: -0.942
```

Gambar 3. *Davies Buldin* untuk $k = 4$

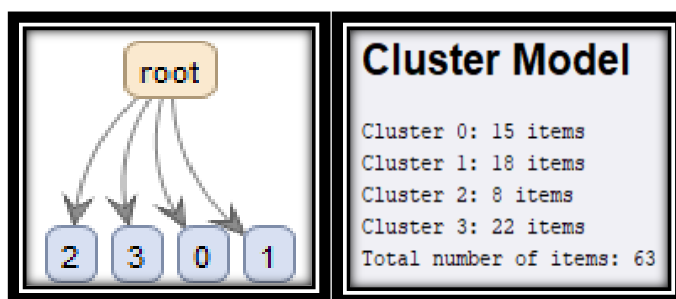
c. Pengujian K-Medoids Clustering dengan $k=4$

Selanjutnya diuji clusterisasi tingkat kemiskinan kabupaten/kota dengan menggunakan algoritma *k-medoids* berdasarkan $k = 4$.



Gambar 4. Pengujian Algoritma K-Medoids

Berdasarkan pemodelan *k-medoids clustering* seperti pada Gambar 4, dengan jumlah *cluster* (k) sebanyak 4, maka *cluster* yang terbentuk sebanyak 4 *cluster*. Untuk *cluster* 0 terdapat 15 *items*, *cluster* 1 terdapat 18 *items*, *cluster* 2 terdapat 8 *items*, *cluster* 3 terdapat 22 *items* dengan total *items* sebanyak 63 *items*. Adapun penjelasan bisa dilihat pada Gambar 5.



Gambar 5. Hasil Data Cluster K-Medoids

Selanjutnya pada Gambar 5, dapat dirincikan hasil clusterisasi sebagai berikut.

Tabel 2. Hasil Clusterisasi

<i>Cluster 0</i>	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
Buru	Buru Selatan	Maluku Tengah	Kepulauan Tanimbar
Seram Bagian Barat	Yahukimo	Ambon	Maluku Tenggara
Kepulauan Sula	Halmahera Barat	Merauke	Kepulauan Aru
Halmahera Selatan	Sarmi	Tidore Kepulauan	Seram Bagian Timur
Halmahera Utara	Keerom	Tual	Maluku Barat Daya
Halmahera Timur	Waropen	Ternate	Halmahera Tengah
Nabire	Supiori	Kota Jayapura	Pulau Morotai
Biak Numfor	Mamberamo Raya	Kota Sorong	Pulau Taliabu
Paniai	Intan Jaya		Lanny Jaya
Puncak Jaya	Kaimana		Jayawijaya
Mimika	Teluk Wondama		Jayapura
Tolikara	Teluk Bintuni		Kepulauan Yapen
Nduga	Sorong Selatan		Boven Digoel
Puncak	Raja Ampat		Mappi
Manokwari	Tambrau		Asmat
	Maybrat		Pegunungan Bintang
	Manokwari Selatan		Mamberamo Tengah
	Pegunungan Arfak		Yalimo
			Dogiyai
			Deiyai
			Fakfak
			Sorong

Berdasarkan Tabel 2, ditunjukkan bahwa hasil uji algoritma K-Medoids *clustering* dapat dirincikan sebagai berikut.

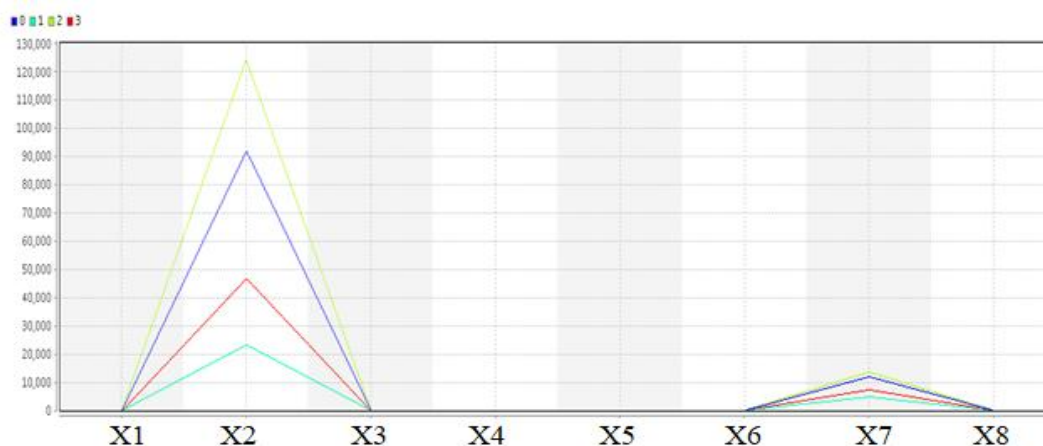
- Cluster 0* : Kabupaten Buru, Seram Bagian Barat, Kepulauan Sula, Halmahera Selatan, Halmahera Utara, Halmahera Timur Nabire, Biak Numfor, Paniai, Puncak Jaya, Mimika, Tolikara, Nduga, Puncak dan Kota Manokwari.
- Cluster 1* : Kabupaten Buru Selatan, Yahukimo, Halmahera Barat, Sarmi, Keerom, Waropen, Supiori, Mamberamo Raya, Intan Jaya, Kaimana, Teluk Wondama, Teluk Bintuni, Sorong Selatan, Raja Ampat, Tambrau, Maybrat, Manokwari Selatan dan Kabupaten Pegunungan Arfak.
- Cluster 2* : Kabupaten Maluku Tengah, Kota Ambon, Merauke, Tidore Kepulauan, Kota Tual, Ternate, Kota Jayapura dan Kota Sorong.
- Cluster 3* : Kabupaten Kepulauan Tanimbar, Maluku Tenggara, Kepulauan Aru, Seram Bagian Timur, Maluku Barat Daya, Halmahera Tengah, Pulau Morotai, Pulau Taliabu, Lanny Jaya, Jayawijaya, Jayapura, Kepulauan Yapen, Boven Digoel, Mappi, Asmat, Pegunungan Bintang, Mamberamo Tengah, Yalimo, Dogiyai dan Kabupaten Deiyai.

Selanjutnya pada Tabel 3, diberikan informasi tentang titik *centroid* dalam setiap *cluster*, dari hasil proses data yang dilakukan menggunakan aplikasi *rapidminer*.

Tabel 3. Hasil Tabel Centroid

<i>Attribute</i>	<i>Cluster 0</i>	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
Kepadatan Penduduk (X1)	61,170	14,040	441,290	18,640
Jumlah Angkatan Kerja (X2)	92201	23594	124491	46881
Tingkat Partisipasi Angkatan Kerja (X3)	71,490	94,710	61,400	68,890
Tingkat Pengangguran Terbuka (X4)	7,040	0,700	9,950	3,360
Rata-Rata Lama Sekolah (X5)	8,340	5,120	11,190	8,330
Angka Harapan Hidup (X6)	68.820	67,260	70,930	66,220
Pengeluaran Per kapita (X7)	11977	5009	13744	7350
Persentase Penduduk Miskin (X8)	20,560	34,700	15,350	27.780

Berdasarkan Tabel 3, ditunjukkan bahwa kepadatan penduduk (X1) tertinggi berada pada *cluster* 2, dan terendah berada pada *cluster* 1. Jumlah Angkatan Kerja (X2) tertinggi berada pada *cluster* 2 dan terendah berada pada *cluster* 1. Tingkat Partisipasi Angkatan Kerja (X3) tertinggi berada pada *cluster* 1 dan terendah berada pada *cluster* 2. Tingkat Pengangguran Terbuka (X4) tertinggi berada pada *cluster* 2 dan terendah berada pada *cluster* 1. Rata-rata Lama Sekolah (X5) tertinggi berada pada *cluster* 2 dan terendah berada pada *cluster* 1. Angka Harapan Hidup (X6) tertinggi berada pada *cluster* 2 dan terendah berada pada *cluster* 3. Pengeluaran Per Kapita (X7) tertinggi berada pada *cluster* 2 dan terendah berada pada *cluster* 1. Persentase Penduduk Miskin (X8) tertinggi berada pada *cluster* 1 dan terendah berada pada *cluster* 2.



Gambar 6. K-Medoids Clustering dengan Grafik

4. KESIMPULAN

Berdasarkan hasil dan pembahasan maka penelitian ini dapat disimpulkan bahwa terdapat 4 *cluster* dalam clusterisasi tingkat kemiskinan di Kepulauan Maluku dan Papua. *Cluster* 0 terdiri atas 15 kabupaten/kota, *cluster* 1 terdiri atas 18 kabupaten/kota, *cluster* 2 terdiri atas 8 kabupaten/kota dan *cluster* 3 terdiri atas 22 kabupaten/kota. Selanjutnya Kepadatan Penduduk (X1) tertinggi berada pada *cluster* 2, dan terendah berada pada *cluster* 1. Jumlah Angkatan Kerja (X2) tertinggi berada pada *cluster* 2 dan terendah berada pada *cluster* 1. Tingkat Partisipasi Angkatan Kerja (X3) tertinggi berada pada *cluster* 1 dan terendah berada pada *cluster* 2. Tingkat Pengangguran Terbuka (X4) tertinggi berada pada *cluster* 2 dan terendah berada pada *cluster* 1. Rata-rata Lama Sekolah (X5) tertinggi berada pada *cluster* 2 dan terendah berada pada *cluster* 1. Angka Harapan Hidup (X6) tertinggi berada pada *cluster* 2 dan terendah berada pada *cluster* 3. Pengeluaran Per Kapita (X7) tertinggi berada pada *cluster* 2 dan terendah berada pada *cluster* 1. Persentase Penduduk Miskin (X8) tertinggi berada pada *cluster* 1 dan terendah berada pada *cluster* 2.

DAFTAR PUSTAKA

- [1] Sembiring, R, “Dampak Perubahan Budaya Sosial Ekonomi Terhadap Kemiskinan Dan Kesejahteraan Pada Masyarakat Desa Pahlawan,” *Jepa*, vol. 3, no. 1, pp. 75-82, 2018.
- [2] Statistik, B. P., Badan pusat statistik, Badan pusat statistik, 2012.
- [3] Defiyanti, S., Jajuli, M., & Rohmawati, N., “Optimalisasi K-Medoid Dalam Pengklasteran Mahasiswa Pelamar Beasiswa Dengan Cubic Clustering Criterion,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 3, no. 1, pp. 211-218, 2017.
- [4] Mardi, Y., “Data Mining: Klasifikasi Menggunakan Algoritma C4. 5,” *Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, vol. 2, no. 2, pp. 213-219, 2017.
- [5] Hormarudin, A. N., “Teknik Data Mining: Algoritma K-Means Clustering,” *J. Ilmu Komputer*, pp. 1-12, 2016.
- [6] Han, J., & Kamber, M., *Data mining: concepts and techniques*, 2 penyunt., Urbana Champaign: University of Illinois : Morgan Kaufmann, 2006.
- [7] Aditya, A., Jovian, I., & Sari, B. N., “Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019.,” *Jurnal Media Informatika Budidarma*, vol. 4, no. 1, pp. 51-58, 2020.
- [8] Sulistyawati, A. A. D., “Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan,” *Sistemasi: Jurnal Sistem Informasi*, vol. 10, no. 3, pp. 516-526, 2021.
- [9] Lesmana, A., *Perbandingan algoritma k-means dan k-medoids dalam penclustering data penjualan Pt. United teknologi integrasi*, Jakarta-Menteng: Doctoral dissertation, Universitas Mercu Buana , 2022.
- [10] Pratama, D. Y., *Analisis Permasalahan Perangkat Jaringan Menggunakan Metode Algoritma K-Means dan Kmedoids*, Jakarta: Doctoral dissertation, Universitas Mercu Buana, 2022.
- [11] Sangga, V. A. P., *Perbandingan algoritma K-Means dan algoritma K-Medoids dalam pengelompokan komoditas peternakan di provinsi Jawa Tengah tahun 2015*, 2018.
- [12] Darma, S., “Klasterisasi Teknik Promosi dalam Meningkatkan Mutu Kampus Menggunakan Algoritma K-Medoids,” *Jurnal Informatika Ekonomi Bisnis*, pp. 89-94, 2021.
- [13] Purwanto, A., *Pembangunan Perangkat Lunak Bantu Klasterisasi dengan Metode Fuzzy Subtractive Clustering*, Doctoral dissertation, UAJY, 2008.
- [14] Faradilla, S. B., *Komparasi Analisis K-Medoids Clustering dan Hierarchical Clustering (Studi Kasus: Data Kriminalitas di Indonesia Tahun 2020)*, Doctoral dissertation, Universitas Islam Indonesia, 2022.
- [15] Fauzy, A., *Perbandingan Algoritma K-Means dan Algoritma K-Medoids dalam Pengelompokan Komoditas Tanaman Biofarmaka di Provinsi Jawa Tengah Tahun 2018*, 2020.