



Application of K-Means and K-Medoids Algorithms for Grouping Country Data Based on Socio-Economic and Health Factors

Penerapan Algoritma K-Means dan K-Medoids untuk Mengelompokkan Data Negara Berdasarkan Faktor Sosial-Ekonomi dan Kesehatan

Delvi Hastari^{1*}, Fadilah Nurunnisa², Salsa Winanda³, Dhea Dwi Aprillia⁴

^{1,2,3,4}Department of Information System, Faculty of Science and Technology,
Sultan Syarif Kasim State Islamic University Riau, Indonesia

E-Mail: ¹12050320385@students.uin-suska.ac.id, ²12050322891@students.uin-suska.ac.id,
³12050322147@students.uin-suska.ac.id, ⁴12050326788@students.uin-suska.ac.id

Corresponding Author: Delvi Hastari

Abstract

Increased development for a country can be assessed from the welfare of its people, therefore people's welfare is one of the goals for a country. The increasing welfare of the people in a country, shows the increasing development of development in that country. Assessment of development progress in a country can be seen from various factors, for example socio-economic factors and health. Non-Governmental Organizations (NGOs) have managed to raise around \$10 million or the equivalent of 156 billion to be given to countries in need. However, they find it difficult to determine which countries need the assistance the most. As a result, this study employs data mining to categorize country datasets based on economic-social and health aspects by applying the K-Means algorithm compared to the K-Medoids algorithm. The purpose of the grouping is to show countries that need assistance, so that existing costs can be used strategically and effectively. According to the study, K-Means performs better than K-Medoids when clustering nation data using the K-Means and K-Medoids algorithms. The cluster results are determined using the Davies Bouldin Index (DBI), and K-Means has a DBI value of 0.095 and $K = 5$, where the validity value is near to 0.

Keyword: Country, Healths, K-Means, K-Medoids, Socio-Economic

Abstrak

Peningkatan pembangunan bagi suatu negara dapat dinilai dari faktor kesejahteraan masyarakatnya, oleh karena itu kesejahteraan rakyat merupakan salah satu tujuan bagi suatu negara. Semakin meningkatnya kesejahteraan rakyat pada suatu negara, menunjukkan semakin meningkat pula perkembangan pembangunan di negara tersebut. Penilaian perkembangan pembangunan di suatu negara dapat terlihat dari berbagai faktor contohnya faktor sosial-ekonomi serta kesehatan. Lembaga Swadaya Masyarakat (LSM) telah berhasil mengumpulkan sekitar \$10 juta atau setara dengan 156 milyar untuk diberikan kepada negara-negara yang membutuhkan. Namun mereka kesulitan untuk menentukan negara mana yang paling membutuhkan bantuan tersebut. Oleh karena itu, penelitian ini menggunakan data mining untuk mengkategorikan kumpulan data negara berdasarkan aspek ekonomi-sosial dan kesehatan dengan menerapkan algoritma K-Means yang dibandingkan dengan algoritma K-Medoids. Tujuan pengelompokkan adalah untuk menunjukkan negara-negara yang membutuhkan bantuan, sehingga biaya yang ada dapat digunakan secara strategis dan efektif. Berdasarkan penelitian ini, K-Means bekerja lebih baik daripada K-Medoids ketika melakukan clustering data negara menggunakan algoritma K-Means dan K-Medoids. Hasil cluster ditentukan menggunakan Davies-Bouldin Index (DBI), dan K-Means memiliki nilai DBI sebesar 0,095 dan $K = 5$ di mana nilai validitasnya mendekati 0.

Kata Kunci: Kesehatan, K-Means, K-Medoids, Negara, Sosial-Ekonomi

1. PENDAHULUAN

Peningkatan pembangunan bagi suatu negara dapat dinilai dari faktor kesejahteraan masyarakatnya, oleh karena itu kesejahteraan rakyat merupakan salah satu tujuan bagi suatu negara, agar dapat berkembang secara keseluruhan ke arah yang lebih baik dan positif. Kesejahteraan dapat diartikan sebagai suatu keadaan di mana rakyatnya selalu berada dalam kondisi berkecukupan segala kebutuhannya, baik material maupun spiritual. Semakin meningkatnya kesejahteraan rakyat pada suatu negara, menunjukkan semakin meningkat pula perkembangan pembangunan di negara tersebut [1]. Penilaian perkembangan pembangunan di suatu negara dapat terlihat dari berbagai macam faktor contohnya faktor sosial-ekonomi, serta faktor kesehatan.

Lembaga Swadya Masyarakat (LSM) atau Non-Governmental Organization (NGO) yaitu organisasi non-pemerintah dengan jangkauan operasi internasional yang saat ini bekerja untuk mengentas kemiskinan dan memberikan fasilitas dan donasi kepada orang-orang di negara terbelakang selama masa bencana alam. LSM berhasil mengumpulkan sekitar 10 juta dolar atau setara dengan 156 milyar untuk diberikan kepada negara-negara yang membutuhkan. Namun mereka kesulitan untuk menentukan negara mana yang paling membutuhkan bantuan tersebut. Oleh karena itu peneliti menggunakan data mining dengan menerapkan algoritma K-Means yang dibandingkan dengan algoritma K-Medoids untuk mengelompokkan dataset negara berdasarkan faktor ekonomi-sosial dan kesehatan. Tujuan pengelompokkan adalah untuk menunjukkan negara-negara yang membutuhkan bantuan, sehingga biaya yang ada dapat digunakan secara strategis dan efektif.

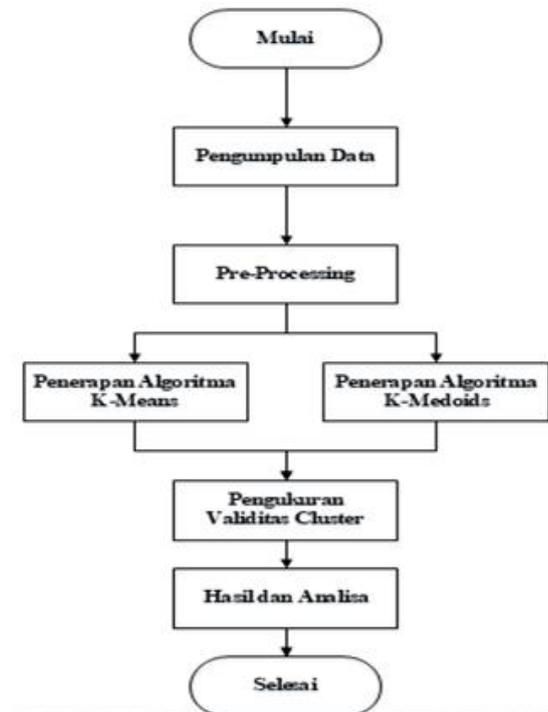
Teknik data mining berperan penting dalam analisis data. Data mining adalah suatu proses komputasi yang digunakan untuk menciptakan suatu model yang menarik dalam dataset dengan jumlah besar yang melibatkan metode kecerdasan buatan, machine learning, statistik, dan sistem database. Terdapat dua jenis teknik dalam data mining yaitu, *Clustering* (Pengelompokan) dan *Classification* (Klasifikasi) [2]. Pada penelitian ini akan menggunakan teknik *clustering* atau pengelompokan. Teknik *Clustering* adalah teknik pengelompokan data yang sama ke dalam suatu cluster dan data yang tidak sama ke cluster yang berbeda dalam data mining [3]. Ada dua pendekatan yang digunakan untuk mengevaluasi kualitas kluster yaitu akurasi cluster dan analisis validitas kluster. Penelitian ini akan menggunakan analisis validitas kluster dengan metode Davies Bouldin Index (DBI). Penelitian ini menggunakan algoritma K-Means dan K-Medoid yang merupakan bagian dari algoritma *clustering*. K-Means merupakan algoritma metode analisis kluster yang bertujuan mempartisi n observasi menjadi k kluster yang dimana masing-masing observasi tersebut termasuk kedalam kluster dengan rata-rata yang terdekat. Sedangkan algoritma K-Medoid merupakan teknik dalam *clustering* yang memiliki konsep dasar untuk mengidentifikasi dokumen dalam sebuah kluster menggunakan a kluster k yang dihasilkan secara acak. Setiap dokumen yang tersisa akan dikelompokkan dengan Medoid yang paling dekat hubungannya [4].

Pemilihan penggunaan algoritma K-Means dan K-Medoid berdasarkan penelitian terdahulu yang telah dikerjakan oleh Santosh Nirmal (2019) dengan judul *Comparative Study between K-Means and K-Medoids Clustering Algorithms*, hasil yang diperoleh membuktikan bahwa penggunaan algoritma K-Medoids dinilai lebih unggul untuk pengelompokan data ketika terdapat outlier jika dibandingkan dengan penggunaan algoritma K-Means, selain itu algoritma K-Medoids bekerja lebih baik dalam skalabilitas untuk kumpulan data dengan jumlah besar dan dinilai lebih efisien dibanding algoritma K-Means [5]. Sedangkan menurut Mohamad, dkk (2022), yang melakukan penelitian dengan judul *Performance analysis in text clustering using K-Means and K-Medoids algorithms for Malay Crime document* menunjukkan bahwa algoritma K-Means mendapatkan hasil terbaik dalam mengelompokkan dokumen yang relevan dengan tingkat akurasi 78%. Pengelompokan K-Means juga mencapai kinerja terbaik untuk evaluasi cluster ketika membandingkan jarak rata-rata dalam cluster dengan algoritma K-Means dibandingkan penggunaan algoritma K-Medoids [4].

Adanya hasil kesimpulan dari kedua penelitian terdahulu membuat perbedaan pendapat tentang algoritma mana yang dianggap lebih baik. Agar perbedaan tersebut tidak semakin meluas, oleh karena itu penelitian ini akan melakukan perbandingan algoritma K-Means dan K-Medoid. Diantara sekian banyak metode analisis kluster yang ada, terdapat dua jenis metode yang masih memiliki saling keterkaitan satu dengan lainnya, yaitu K-Means dan K-Medoids [6]. Dikatakan bahwa K-means sensitif terhadap outlier karena objek dengan nilai multipel yang sangat besar dapat mendukung perubahan distribusi data namun k-means mampu mengelompokkan data dalam jumlah besar dengan waktu komputasi yang relatif cepat dan efisien [7] [8]. Walaupun algoritma k-medoids membutuhkan waktu komputasi yang lebih lama daripada algoritma k-means, namun algoritma k-medoids dapat mengatasi noise pada algoritma k-means dengan mengurangi outlier cluster yang dihasilkan [9] [8] [10]. Dengan adanya kekurangan dan kelebihan dari kedua algoritma sehingga seringkali menyebabkan hasil yang berbeda. Maka dari itu, pada penelitian ini akan melakukan perbandingan algoritma K-Means dan K-Medoids sebagai perbandingan hasil dengan uji validitas cluster. Yang mana dari hasil perbandingan diperoleh model terbaik untuk mengelompokkan dataset negara berdasarkan faktor sosial-ekonomi dan kesehatan.

2. BAHAN DAN METODE

Terdapat 5 fase yang dikerjakan pada penelitian ini yaitu, fase mengumpulkan data, fase pre-processing data, fase clustering menggunakan algoritma K-Means dan K-medoids, fase uji validitas cluster, kemudian hasil dan analisis. Gambar 1 menunjukkan tahap-tahap dari penelitian yang akan dilakukan.



Gambar 1. Metodologi Penelitian

Penggunaan dataset pada penelitian ini berasal dari website Kaggle. Website kaggle menyediakan berbagai dataset yang dapat digunakan untuk penelitian yang berkaitan dengan data mining. Penelitian ini menggunakan dataset dari Lembaga Swadaya Masyarakat (LSM) Internasional atau yang biasa disebut Non-Government Organization (NGO). Dataset LSM berasal dari data gabungan negara berdasarkan faktor sosial-ekonomi dan kesehatan, pada tahun 2020. Selanjutnya melakukan tahapan preprocessing terhadap data yang akan digunakan, adapun tahapan dalam proses pre-processing yaitu cleaning data, dan normalisasi data dengan menerapkan min max normalization. Tahapan selanjutnya dilakukan clustering menggunakan algoritma K-Means dan juga K-Medoids. Setelah didapatkan hasil dari kedua algoritma tersebut, akan dilakukan proses validitas cluster menggunakan *Davies Bouldin Index* (DBI), kemudian akan dianalisis dari hasil algoritma terbaik yang didapatkan.

2.1 Clustering

Clustering merupakan tipe berdasarkan *unsupervised learning* yang rata-rata dipakai pada data mining [11]. Clustering adalah langkah-langkah yang bertujuan mengelompokkan data yang menunjukkan kesamaan antara data satu dengan data lain ke dalam kluster sebagai akibatnya data pada satu kluster mempunyai taraf kesamaan maksimum dan kesamaan minimum antar kluster [12].

2.2 K-Means

Algoritma k-means adalah metode clustering data iteratif untuk mempartisi kumpulan data menjadi sejumlah kluster yang telah ditentukan sebelumnya [13]. Pada algoritma K-Means data dapat dibagi menjadi dua atau lebih kelompok berdasarkan kelompok data nonhierarki. Metode ini bertujuan untuk mengelompokkan suatu dataset tertentu ke dalam sejumlah cluster tertentu berdasarkan data dengan karakteristik yang sejenis, sedangkan data yang memiliki karakteristik berbeda di tempatkan pada kelompok yang lainnya [14]. Pusat kluster K harus dipilih secara acak. Pengukuran jarak memainkan aturan yang sangat penting tentang bagaimana algoritma ini bekerja. Ada berbagai metode untuk mengukur jarak pada algoritma K-Means yaitu *Euclidean Distance*, *Manhattan Distance*, *Chebyshev Distance*, dan lain-lainnya. Tetapi memilih teknik perhitungan jarak yang tepat tergantung sepenuhnya pada jenisnya data yang dikumpulkan. Namun, menggunakan *Euclidean Distance* sebagai pengukuran jarak dirasa lebih cepat, efisien dan mudah dipahami [15]. Berikut tahapan dalam algoritma k-means:

Pertama, tentukan jumlah kelompok (k), kemudian pilih pusat kelompok secara acak. Kedua, hitung jarak antara setiap titik ke semua titik pusat menggunakan Euclidean Distance. Ketiga, kelompokkan data ke suatu kelompok yang memiliki jarak paling dekat. Keempat, menghitung jumlah pusat cluster baru, kemudian ulangi langkah-langkah sebelumnya (langkah 2 - langkah 4) sehingga tidak ada lagi data yang berpindah ke cluster lainnya [16] [17].

2.3 K-Medoids

K-Medoids merupakan salah satu teknik yang paling populer untuk analisis cluster, di mana titik-titik data dipilih sebagai medoid daripada memilih rata-rata sebagai rata-rata dalam K-Means [18]. Pada algoritma K-Medoids berupaya untuk meminimalkan jarak antar titik data dalam sebuah cluster dengan titik data yang representatif di klaster yang sama [19]. Algoritma K-medoid dimulai dengan menghitung K medoid dan memetakan setiap objek dataset ke medoid terdekat menggunakan metrik jarak. Setelah itu, k -medoids menghitung biaya penukaran objek P_i dan M_i menggunakan persamaan 1 [20]:

$$COST_{PM} = \sum_{M_i} \sum_{P_i \in m_i} |P_i - M_i| \quad (1)$$

Ketika nilai jatuh ke ambang batas yang ditetapkan, algoritma melakukan langkah-langkah berikut. Untuk setiap medoid M , titik data P sehingga $P \neq M$.

1. Pertimbangkan pertukaran M dan P , dan hitung perubahan nilai
2. Jika perubahan nilai adalah yang terbaik saat ini, ingat pertukaran M dan P ini
3. Lakukan penukaran M dan P . Jika, ini mengurangi nilai, ulangi langkah 1 dan 2. Jika tidak, algoritma akan berhenti.

2.4 Davies-Bouldin Index

David L. Davies dan Donald W. Bouldin menerbitkan sebuah teknik dan menamakannya *Davies-Bouldin Index* (DBI) yang dipergunakan untuk menilai klaster. Skor *Indeks Davies-Bouldin* memiliki skema penilaian klaster internal yang menentukan skor klaster baik atau buruk berdasarkan jumlah skor klaster dan kedekatan antar skor klaster [21]. Indeks Davies-Bouldin ialah salah satu teknik yang diaplikasikan untuk memperkirakan kualitas pengelompokkan dengan mengevaluasi jarak intra-cluster (jarak rata-rata semua titik data dalam sebuah cluster dari centroid) ke jarak antar-cluster (jarak antara dua centroid). Semakin kecil nilai DBI, semakin baik kekompakan atau pemisahannya, sebaliknya untuk nilai DBI yang besar [22]. Davies-Bouldin Index [23] rata-rata kesamaan R dari setiap cluster i sehubungan dengan cluster $j \neq i$ yang paling mirip secara maksimal.

$$DB = \frac{1}{k} \sum_{i=0}^k R_i \quad (1)$$

2.5 Euclidean Distance

Untuk representasi berbasis distribusi, jarak antara dua peringkat didefinisikan sebagai jarak Euclidean antara masing-masing distribusi probabilitas [24]. Dengan menggunakan Euclidean Distance Theory pada persamaan, dapat menghitung jarak data ke setiap titik pusat cluster menggunakan persamaan 3 [25].

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

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan dan Pre-Processing Data

Data yang dipakai pada penelitian ini berjumlah 167 data. Atribut yang terdapat pada data antara lain country, child_mort, export, health, import, income, inflation, life_expec, total_fer dan gdp. Pada tahap preprocessing data, dilakukan pembersihan dan normalisasi data-data yang dipakai. Data hasil cleaning dapat dilihat pada Tabel 1.

Tabel 1. Data Hasil Cleaning

No	Country	Child_mort	Export	Health	Import	Income	Inflation	Life_expec	Total_fer	Gdp
1	Afghanistan	90,200	10,000	7,580	44,900	1610,000	9,440	56,200	5,820	553,000
2	Albania	16,600	28,000	6,550	48,600	9930,000	4,490	76,300	1,650	4090,000
3	Algeria	27,300	38,400	4,170	31,400	12900,000	16,100	76,500	2,890	4460,000

No	Country	Child_mort	Export	Health	Import	Income	Inflation	Life_exp ec	Total_ fer	Gdpp
4	Angola	119,00	62,300	2,850	42,900	5900,000	22,400	60,100	6,160	3530,000
5	Antigua and Barbuda	10,300	45,500	6,030	58,900	19100,000	1,440	76,800	2,130	12200,000
6	Argentina	14,500	18,900	8,100	16,000	18700,000	20,900	75,800	2,370	10300,000
...
167	Zambia	83,100	37,000	5,890	30,900	3280,000	14,000	52,000	5,400	1460,000

3.2 Normalisasi Data

Dalam penelitian ini normalisasi yang akan digunakan adalah min-max normalization. Tujuan dari normalisasi yang dilakukan yaitu agar menghasilkan data maksimum dan minimum untuk mendapatkan nilai yang seimbang. Berikut hasil normalisasi data pada Tabel 2.

Tabel 2. Data Hasil Normalisasi

No	Country	Child_mort	Export	Health	Import	Income	Inflation	Life_e xpec	Total_ fer	Gdpp
1	Afghanistan	0,426	0,049	0,359	0,258	0,008	0,126	0,475	0,737	0,003
2	Albania	0,068	0,140	0,295	0,279	0,075	0,080	0,872	0,079	0,037
3	Algeria	0,120	0,192	0,147	0,180	0,099	0,188	0,876	0,274	0,040
4	Angola	0,567	0,311	0,065	0,246	0,043	0,246	0,552	0,790	0,031
5	Antigua and Barbuda	0,037	0,227	0,262	0,338	0,149	0,052	0,882	0,155	0,114
6	Argentina	0,058	0,094	0,391	0,092	0,145	0,232	0,862	0,192	0,096
...
167	Zambia	0,392	0,185	0,254	0,177	0,021	0,168	0,393	0,670	0,012

3.3 Klasterisasi dengan Algoritma

Untuk melakukan pengelompokan dua algoritma yaitu K-Means dan K-Medoids, penelitian ini memakai data yang sudah di normalisasikan dari fase sebelumnya. Pada fase ini akan dilakukan beberapa percobaan untuk menemukan jumlah kluster yang optimal menggunakan algoritma K-Means dan K-Medoids.

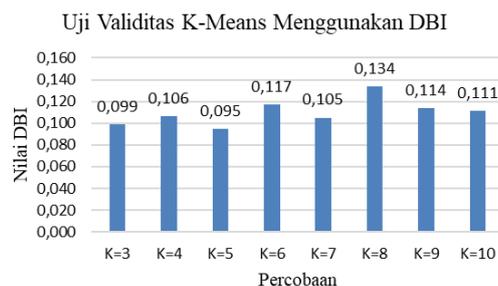
3.1.1 K-Means

Proses ini dilakukan dengan uji coba terhadap 3 sampai 10 kluster untuk mendapatkan kluster yang optimal. Hasil klasterisasi dapat terlihat di Tabel 3.

Tabel 3. Hasil Klasterisasi K-Means

Percobaan	Kluster									
	1	2	3	4	5	6	7	8	9	10
K=3	51	3	113	-	-	-	-	-	-	-
K=4	46	30	88	3	-	-	-	-	-	-
K=5	88	4	3	26	46	-	-	-	-	-
K=6	67	3	24	29	40	4	-	-	-	-
K=7	85	28	3	3	5	1	42	-	-	-
K=8	38	25	47	23	24	4	3	3	-	-
K=9	12	48	3	6	23	1	30	43	1	-
K=10	38	23	53	7	1	4	3	1	8	29

Setelah melakukan proses clustering dengan algoritma K-Means, selanjutnya melakukan tes validitas cluster untuk menemukan kluster yang terbaik. DBI digunakan untuk mencari nilai kepercayaan suatu himpunan k-means. Nilai DBI yang diperoleh dari klasterisasi menggunakan algoritma K-Means ditunjukkan pada Gambar 2.



Gambar 2. Uji Validitas K-Means Menggunakan DBI

Pada Gambar 2. nilai DBI terbaik untuk algoritma k-means adalah 0,095, nilai DBI diperoleh pada percobaan dengan k = 5, yang membagi data menjadi 5, dengan 88 negara di kluster 1, kluster 2 dengan 4 negara yang disertakan., 3 negara di kluster 3, 26 negara di kluster 4, dan 46 negara di kluster 5.

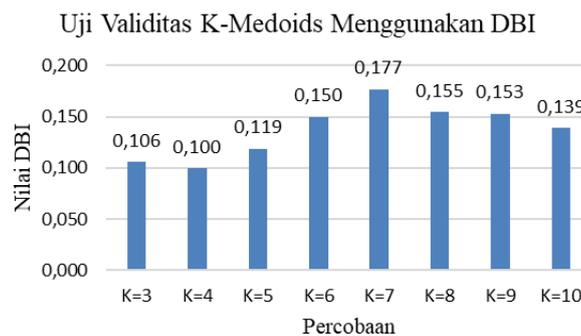
3.1.2 K-Medoids

Proses clustering pada tahapan ini menerapkan algoritma k-medoids dengan eksperimen clustering yang serupa dengan algoritma dengan algoritma yang sebelumnya, yaitu jumlah cluster adalah 3-10. Hasil clusterisasi ditunjukkan pada Tabel 4.

Tabel 4. Hasil Klasterisasi K-Medoids

Percobaan	Kluster									
	1	2	3	4	5	6	7	8	9	10
K=3	101	47	19	-	-	-	-	-	-	-
K=4	18	98	4	47	-	-	-	-	-	-
K=5	27	18	3	73	46	-	-	-	-	-
K=6	25	18	64	22	35	3	-	-	-	-
K=7	25	18	54	29	6	25	10	-	-	-
K=8	5	20	51	6	18	25	29	-	-	-
K=9	20	17	26	41	6	13	25	18	1	-
K=10	29	3	35	19	18	25	10	22	1	5

Setelah dilakukan klasterisasi data dengan algoritma k-medoids, selanjutnya dilakukan validasi kluster untuk mencari kluster yang optimal menggunakan uji validitas Davies -Bouldin Index (DBI). Hasil nilai DBI yang didapatkan pada percobaan ini dapat terlihat pada Gambar 3.



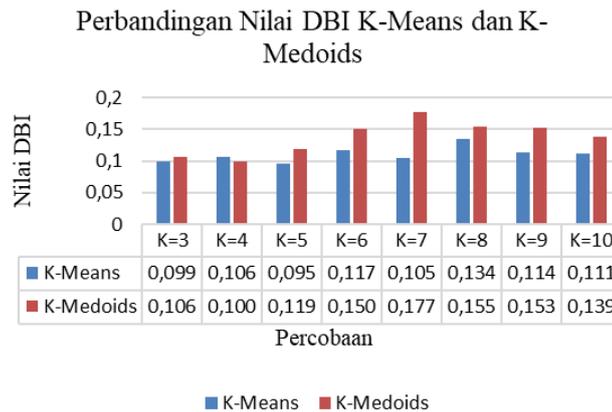
Gambar 3. Uji Validitas K-Medoids Menggunakan DBI

Pada Gambar 3. nilai DBI terbaik untuk algoritma K-Medoid pada k=7, diperoleh nilai DBI sebesar 0,100 membagi data menjadi empat, 18 negara di kluster 1, 98 negara di kluster 2, 14 negara berada di kluster 3 dan 47 di kluster 4.

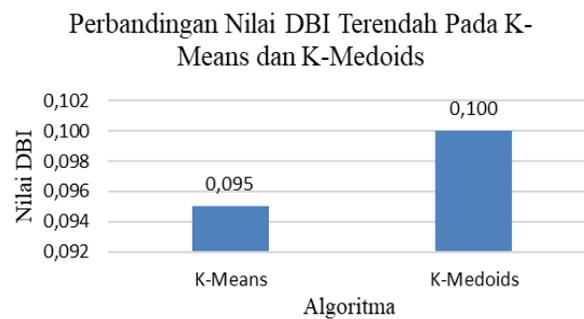
3.4 Perbandingan Validitas Cluster

Setelah dilakukan percobaan dengan dua algoritma yaitu K-Means dan K-Medoid, mulai dari percobaan 3 hingga 10 cluster, langkah berikutnya adalah menentukan cluster terbaik dengan mengevaluasi nilai validitas dalam penelitian ini yaitu DBI, dari algoritma k-means dan k-medoids di setiap percobaan kluster. Gambar 4 menunjukkan perbandingan dari dua algoritma yang sudah dilakukan pengujian.

Dari gambar 4 dilihat bahwa diagram perbandingan nilai DBI terbaik untuk algoritma K-Means, khususnya pada K=5 dengan nilai DBI 0,095 dan pada algoritma K-Medoids khususnya pada K=4 dengan nilai DBI 0,100.



Gambar 4. Perbandingan Nilai DBI K-Means dan K-Medoids



Gambar 5. Perbandingan Nilai DBI Terendah Pada K-Means dan K-Medoids

Hasil pada Gambar 5 menunjukkan bahwa algoritma K-Means memiliki skor validitas kluster yang lebih baik daripada algoritma K-Medoids pada penelitian ini. Oleh karena itu, pada penelitian ini kluster terbaik diperoleh dengan menggunakan algoritma K-Means, nilai eksperimennya adalah $k=5$.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengelompokkan data negara-negara berdasarkan faktor ekonomi-sosial dan kesehatan yang membutuhkan bantuan, sehingga biaya yang ada dapat digunakan secara strategis dan efektif. Dari penelitian yang sudah dilakukan, bisa diambil kesimpulan ternyata pengelompokkan data negara dengan algoritma K-Means dan K-Medoids, clustering terbaik didapatkan oleh algoritma K-Means dengan nilai $K=5$, sedangkan algoritma K-Medoids $K=4$. Hasil cluster dihitung dengan menggunakan *Davies Bouldin Index* (DBI). Oleh karena itu, dalam penelitian K-Means mengungguli K-Medoids dalam mengelompokkan data negara berdasarkan faktor sosial-ekonomi dan kesehatan, dengan nilai DBI 0,095 dan $K=5$, yang mana nilai validitas dari algoritma K-Means mendekati 0. Dalam uji K-Means dengan $k=5$, bagi data jadi 5 kluster, yaitu 88 negara di kluster 1, 4 negara di kluster 2, 3 negara di kluster 3, 26 negara di kluster 4, dan 46 negara di kluster 5.

REFERENSI

- [1] N. Dwitiyanti, N. Selvia, and F. R. Andrari, "Penerapan Fuzzy C-Means Cluster dalam Pengelompokkan Provinsi Indonesia Menurut Indikator Kesejahteraan Rakyat," *Fakt. Exacta*, vol. 12, no. 3, p. 201, 2019, doi: 10.30998/faktorexacta.v12i3.4526.
- [2] J. Majumdar, S. Naraseyappa, and S. Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," *J. Big Data*, vol. 4, no. 1, 2017, doi: 10.1186/s40537-017-0077-4.
- [3] M. Verma, M. Srivastava, N. Chack, A. K. Diswar, and N. Gupta, "A Comparative Study of Various Clustering Algorithms in Data Mining," *Int. J. Eng. Res. Appl. www.ijera.com*, vol. 2, no. 3, pp. 1379–1384, 2012.
- [4] R. Mohamad, N. N. M. Muhait, N. M. M. Noor, and Z. A. Othman, "Performance analysis in text clustering using k-means and k-medoids algorithms for Malay crime documents," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 5, pp. 5014–5026, 2022, doi: 10.11591/ijece.v12i5.pp5014-5026.
- [5] S. Nirmal, "Comparative study between k-means and k-medoids clustering algorithms," *Int. Res. J. Eng. Technol.*, vol. 839, pp. 839–844, 2019, [Online]. Available: <https://www.irjet.net/archives/V6/i3/IRJET-V6I3154.pdf>

-
- [6] Y. H. Susanti and E. Widodo, "Perbandingan K-Means dan K-Medoids Clustering terhadap Kelayakan Puskesmas di DIY Tahun 2015," *Pros. SI MaNIs (Seminar Nas. Integr. Mat. dan Nilai Islam.*, vol. 1, no. 1, pp. 116–122, 2017.
- [7] Athifaturrofifah, R. Goejantoro, and D. Yuniarti, "Comparison of K-Means And K-Medoids Grouping Data on Potential Forest / Land Fires Based on Hotspots Distribution(Case study: Hotspots Data in Indonesia on April 28, 2018)," *J. EKSPONENSIAL*, vol. 10, no. 2, pp. 143–152, 2019.
- [8] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *JIFI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 6, no. 2, pp. 229–240, 2021, doi: 10.29100/jipi.v6i2.2008.
- [9] R. D. Ramadhani and D. J. AK, "Evaluasi K-Means dan K-Medoids pada Dataset Kecil," *Semin. Nas. Inform. dan Apl.*, no. September, pp. 20–24, 2017.
- [10] S. Sundari, I. S. Damanik, A. P. Windarto, H. S. Tambunan, J. Jalaluddin, and A. Wanto, "Analisis K-Medoids Clustering Dalam Pengelompokan Data Imunisasi Campak Balita di Indonesia," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 687, 2019, doi: 10.30645/senaris.v1i0.75.
- [11] R. R. A. Rahman and A. W. Wijayanto, "Pengelompokan Data Gempa Bumi Menggunakan Grouping Earthquakes Data Using," pp. 31–38, 2021.
- [12] S. Dewi, S. Defit, and Y. Yuhandri, "Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 28–33, 2021, doi: 10.37034/jsisfotek.v3i1.40.
- [13] D. Syaputri, P. H. Noprita, and S. Romelah, "Implementasi Algoritma K-Means untuk Pengelompokan Distribusi Sosial Ekonomi Masyarakat Berdasarkan Demografi Kependudukan," *Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 1–6, 2021.
- [14] H. Syukron, M. F. Fayyad, and F. J. Fauzan, "Comparison K-Means K-Medoids and Fuzzy C-Means for Clustering Customer Data with LRFM Model " Perbandingan K-Means K-Medoids dan Fuzzy C-Means untuk Pengelompokan Data Pelanggan dengan Model LRFM," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 76–83, 2022.
- [15] M. S. Alam *et al.*, "Automatic human brain tumor detection in mri image using template-based k means and improved fuzzy c means clustering algorithm," *Big Data Cogn. Comput.*, vol. 3, no. 2, pp. 1–18, 2019, doi: 10.3390/bdcc3020027.
- [16] W. Utomo, "The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia," *Ilk. J. Ilm.*, vol. 13, no. 1, pp. 31–35, 2021, doi: 10.33096/ilkom.v13i1.763.31-35.
- [17] M. Moghadaszadeh and H. Shokrzadeh, "An Overview of Expectation Maximization and K-Means family Clustering Algorithms in Data Mining Applications," pp. 17–20, 2018, doi: 10.15242/dirpub.dir1017002.
- [18] M. M. Madbouly, S. M. Darwish, N. A. Bagi, and M. A. Osman, "Clustering Big Data Based on Distributed Fuzzy K-Medoids: An Application to Geospatial Informatics," *IEEE Access*, vol. 10, pp. 20926–20936, 2022, doi: 10.1109/ACCESS.2022.3149548.
- [19] A. Prahara, D. P. Ismi, and A. Azhari, "Parallelization of Partitioning Around Medoids (PAM) in K-Medoids Clustering on GPU," *Knowl. Eng. Data Sci.*, vol. 3, no. 1, pp. 40–49, 2020, doi: 10.17977/um018v3i12020p40-49.
- [20] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi, "K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir," *IEEE Access*, vol. 8, pp. 151847–151855, 2020, doi: 10.1109/ACCESS.2020.3014021.
- [21] M. Mughnyanti, S. Efendi, and M. Zarlis, "Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, 2020, doi: 10.1088/1757-899X/725/1/012128.
- [22] M. B. Agbaje, A. E. Ezugwu, and R. Els, "Automatic data clustering using hybrid firefly particle swarm optimization algorithm," *IEEE Access*, vol. 7, pp. 184963–184984, 2019, doi: 10.1109/ACCESS.2019.2960925.
- [23] L. Enzo, B. Da, and N. M. A. X. Melton, "Incremental Cluster Validity Indices for Online Learning of Hard Partitions : Extensions and Comparative Study," no. i, 2020.
- [24] F. Cabitza, A. Campagner, and M. Mattioli, "The unbearable (technical) unreliability of automated facial emotion recognition," 2022, doi: 10.1177/20539517221129549.
- [25] N. T. Luchia, H. Handayani, F. S. Hamdi, D. Erlangga, and S. Fitri Octavia, "Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 35–41, 2022.
-