



Comparison of Unsupervised Learning Techniques for Clustering Data on the Number of Villages in Indonesia

Perbandingan Teknik Unsupervised Learning untuk Pengelompokan Data Jumlah Desa Di Indonesia

Akhiril Anwar Harahap^{1*}, Muhammad Raihan², Nailul Amani³, Putri Risma Andini⁴

^{1,2,3,4}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

E-Mail: ¹akhirilanwar5@gmail.com, ²raihan123723@gmail.com,
³nailula253@gmail.com, ⁴putririsma1803@gmail.com

Corresponding Author: Akhiril Anwar Harahap

Abstract

Education is a place to develop and improve the quality of human resources. A good quality of education can be achieved by improving the quality of education. However, the quality of education in Indonesia is still quite concerning. The quality of education is influenced by good and adequate school facilities. The fact is that there are still many areas in Indonesia that cannot enjoy access and good educational facilities. In this study, the number of villages with school facilities was grouped by the province to provide information to the government about areas with low school facilities. The aim is for the government to prioritize development in the area. This study uses data processing with 3 clustering algorithms, including the K-Means algorithm, K-Medoids algorithm, and Fuzzy C-Means algorithm. Each algorithm was tested with experiments $K = 3$ to $K = 10$ so that the cluster validity comparison value obtained using DBI on the K-Means algorithm was 0.451, K-Medoids was 0.638, and Fuzzy-C-Means was 0.491. K-Means is proven to be superior to the K-Medoids algorithm and the Fuzzy C-Means algorithm for grouping data on villages that have school facilities according to provinces in Indonesia.

Keyword: Clustering, DBI, Fuzzy C-Means, K-Means, K-Medoids

Abstrak

Pendidikan adalah wadah untuk berkembang dan meningkatkan kualitas sumber daya manusia. Kualitas pendidikan yang baik dapat dicapai dengan perbaikan mutu pendidikan. Namun kualitas pendidikan di Indonesia untuk saat ini masih cukup memprihatikan. Kualitas pendidikan salah satunya dipengaruhi oleh fasilitas sekolah yang baik dan memadai. Faktanya masih banyak daerah Indonesia yang tidak dapat menikmati akses dan fasilitas pendidikan yang baik. Pada penelitian ini dilakukan pengelompokan jumlah desa yang memiliki fasilitas sekolah berdasarkan provinsi untuk memberikan informasi kepada pemerintah tentang daerah-daerah yang memiliki fasilitas sekolah rendah. Tujuannya adalah agar pemerintah dapat memprioritaskan pembangunan pada daerah tersebut. Pada penelitian ini menggunakan pengolahan data dengan 3 algoritma clustering, diantaranya adalah algoritma K-Means, algoritma K-Medoids, dan algoritma Fuzzy C-Means. Masing-masing algoritma diuji dengan percobaan $K=3$ sampai $K=10$ sehingga diperoleh nilai perbandingan validitas cluster menggunakan DBI pada algoritma K-Means yaitu 0,451, K-Medoids yaitu sebesar 0,638, dan Fuzzy-C-Means yaitu sebesar 0,491. K-Means terbukti lebih unggul daripada algoritma K-Medoids dan algoritma Fuzzy C-Means untuk melakukan pengelompokan data desa yang memiliki fasilitas sekolah menurut provinsi di Indonesia

Kata Kunci: Clustering, DBI, Fuzzy C-Means, K-Means, K-Medoids

1. PENDAHULUAN

Pendidikan adalah wadah untuk berkembang dan meningkatkan kualitas sumber daya manusia. Pendidikan diadakan untuk membentuk sumber daya manusia yang mampu berfikir kreatif, aktif, dan bermoral baik [1]. Mutu Pendidikan adalah komponen penting untuk mencapai kualitas pendidikan yang baik. Peningkatan kualitas pendidikan bertujuan untuk mewujudkan Indonesia sebagai negara maju [2]. Namun kualitas pendidikan di Indonesia untuk saat ini masih cukup memprihatinkan. Hal ini dapat terjadi akibat dari adanya permasalahan pada sistem pendidikan di Indonesia sehingga mempengaruhi kualitas dari pendidikan. Salah satunya adalah kesenjangan sarana dan prasarana sekolah di berbagai daerah [3]. Kualitas Pendidikan dapat dilihat dari hasil dan banyaknya serapan lulusan pada dunia usaha/industri. Faktanya masih banyak daerah di Inodnesia yang tidak dapat menikmati akses dan fasilitas pendidikan yang baik dan merata. Tentu saja hal ini tidak adil bagi masyarakat yang tinggal di daerah dengan fasilitas sekolah rendah. Padahal sejatinya pendidikan merupakan hak bagi setiap rakyat Indonesia tanpa membedakan agama, suku, budaya dan ras [4].

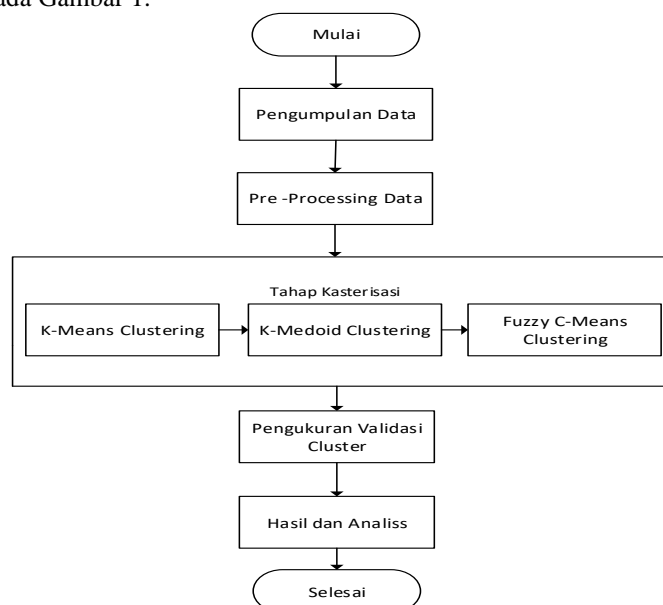
Indonesia merupakan negara maritim sekaligus negara kepulauan yang sangat luas. Hal ini memungkinkan diperlukannya pengelompokan daerah-daerah berdasarkan fasilitas sekolah untuk memberikan informasi dan rekomendasi kepada pemerintah dalam prioritas pemberian dan pembangunan fasilitas sekolah yang ada di Indonesia. Ada beberapa cara untuk melakukan pengelompokan ini yaitu dengan data mining menggunakan teknik clustering. Ada banyak algoritma clustering, namun yang paling sering digunakan diantaranya yaitu algoritma K-Means, K-Medoids, dan Fuzzy C-Means [5].

Menurut Hamdi Syukron dan kawan-kawan (2022) pada pengelompokan data pelanggan memperoleh hasil bahwa K-Means adalah algoritma yg lebih baik daripada algoritma K-Medoids dan algoritma Fuzzy C-Means. Dimana K-Means memiliki nilai DBI terkecil mendekati 0 yaitu 0,167 pada K=6 [6]. Berdasarkan penelitian yang dilakukan oleh Anissa Enggar Pramitasari dan Yessica Nataliani (2021) tentang pengelompokan data karyawan menggunakan algoritma K-Means dan algoritma Fuzzy C-Means. Dimana algoritma Fuzzy C-Means memperoleh nilai akurasi 76% dan algoritma K-Means memperoleh nilai akurasi sebesar 44%. Sehingga terbukti bahwa Fuzzy C-Means lebih unggul daripada K-Means [7]. Menurut Nanda Try Luciha dan kawan-kawan (2022) tentang pengelompokan data miskin di Indonesia memperoleh hasil bahwa K-Means terbukti lebih baik dibandingkan dengan K-Medoids dimana nilai DBI Pada K-Means sebesar 0,041 dan DBI pada K-Medoids sebesar 0,052 [8].

Berdasarkan uraian diatas, penelitian ini dilakukan untuk membandingkan 3 algoritma, yaitu algoritma K-Means, algoritma K-Medoids, dan algoritma fuzzy C-Means dalam melakukan clustering data jumlah desa yang memiliki fasilitas sekolah menurut provinsi di Indonesia. Penelitian ini diharapkan dapat membantu pemerintah untuk menentukan prioritas pemberian dan pembangunan fasilitas sekolah yang ada di Indonesia untuk meratakan fasisitas pendidikan demi memajukan kualitas pendidikan di Indonesia untuk indonesia yang gemilang.

2. METODOLOGI PENELITIAN

Ada 5 tahapan pada penelitian ini, dimulai dari pengumpulan data, kemudian pre-processing data, klasterisasi data, pengukuran validasi cluster pada algoritma serta hasil dan analisis. Metodologi pada penelitian ini adalah pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1 Data Mining

Data mining merupakan metode untuk mencari dan menggali informasi yang penting dan tersembunyi dari data yang sangat besar [9]. Tujuan data mining yaitu menemukan trend tertentu atau pola yang unik yang dapat membantu dalam hal memprediksi dan dalam hal pengambilan keputusan [10]. Ada beberapa metode pada data mining, diantaranya adalah clustering. Clustering merupakan metode pengelompokan data atau objek yang ada pada database ke dalam beberapa cluster [11]. Clustering termasuk ke dalam teknik unsupervised learning yang paling penting terutama dalam data mining. Clustering dilakukan dengan cara membagi data menjadi beberapa cluster atau kelompok tergantung pada objek data yang mirip satu sama lain, sehingga objek yang berbeda masuk ke dalam kelompok yang berbeda dan objek serupa masuk ke dalam kelompok atau cluster yang sama [12].

2.2 K-Means

K-Means merupakan metode dalam unsupervised learning yang terkenal akan kesederhanaan dan efektifitasnya dalam clustering [13]. Algoritma ini melakukan pengelompokan data ke dalam kelompok K dimana K dinyatakan sebagai parameter masukan. Masing-masing objek dimasukkan ke dalam cluster tergantung pada kedekatan nilai dari rata-rata cluster tersebut [14]. Dalam melakukan perhitungan jarak antara data dan titik pada pusat cluster menggunakan persamaan Euclidian Distance. Persamaan Euclidian Distance adalah sebagai berikut:

$$D(a,b) = \sqrt{(X1p - X1q)^2 + (X2p - X2q)^2 + \dots + (Xnp - Xnq)^2} \quad (1)$$

Dimana:

D(p,q) = Jarak data ke p ke pusat cluster q
 Xnp = Data ke p atribut data ke n
 Xnq = Titik pusat ke q pada atribut ke n

2.3 K-Medoids

K-Medoids adalah algoritma unsupervised learning yang dapat digunakan dalam mencari medoid pada cluster [15]. K-Medoids juga sering disebut dengan nama Partitioning Around Medoids (PAM). K-Medoids lebih mudah diimplementasikan daripada K-Means [16]. Untuk mengelompokkan data pada algoritma Medoids dilakukan dengan sistem partisi. Caranya adalah dengan menghitung jarak antar data non-medoid dan data medoid. Untuk mengalokasikan data ke cluster terdekat, K-Medoids dapat dihitung menggunakan persamaan Euclidian Distance [17].

2.4 Fuzzy C-Means (FCM)

Algoritma Fuzzy C-Means merupakan algoritma pengelompokan yang dapat diterapkan pada bidang analisis fitur dan juga clustering [18]. Pengelompokan data pada algoritma Fuzzy C-Means ditentukan menurut derajat keanggotaan yakni antara 0 sampai dengan 1. Untuk mencari klaster dapat menggunakan persamaan berikut [19]:

$$jm = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m d^2(y_{1z_1}) \quad (2)$$

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{m-1}} \right]^{-1} \quad 1 \leq i \leq c; 1 \leq j \leq n \quad (3)$$

$$z_i = \frac{\sum_{j=1}^n (u_{ij})^m y_j}{\sum_{j=1}^n (u_{ij})^m} \quad (4)$$

2.5 Davies-Bouldin Index (DBI)

Metode Davies-Bouldin Index atau yang disingkat dengan DBI dikenalkan pada tahun 1979 oleh David L. Davies bersama Donald W. Bouldin. Pada metode DBI, hasil cluster dilihat dari kedekatan antar masing-masing data hasil cluster. Pengukuran DBI yaitu meminimalkan jarak intra cluster dan memaksimalkan jarak inter cluster. Dimana semakin kecil nilai pada DBI maka semakin baik cluster tersebut [20]. Rumus DBI adalah:

$$R_{j,k} = \frac{MAE_j - MAE_k}{d(c_j, c_k)} \quad (5)$$

$$DBI = \frac{1}{M} \sum_{j=1}^M \max_{j \neq k} R_{j,k} \quad (6)$$

3. HASIL DAN ANALISA

3.1 Pengumpulan Data

Pada penelitian ini, data berasal dari website Badan Pusat Statistik (BPS). Dataset yang digunakan adalah dataset jumlah desa yang memiliki fasilitas sekolah di Indonesia berdasarkan provinsi pada tahun 2011 hingga 2018. Ada 16 atribut yang digunakan yaitu sebagai berikut:

Table 1. Dataset

NO	PROV	SD2011	SMP2011	SMU2011	SMK2011	PT2011	SD2014	...	PT2018
1	ACEH	3227	1023	497	127	120	3358	...	115
2	SUMATERA UTARA	4725	1912	923	504	181	4957	...	193
3	SUMATERA BARAT	987	611	279	137	113	1100	...	116
4	RIAU	1602	957	425	163	63	1779	...	79
5	JAMBI	1326	651	262	99	33	1457	...	50
6	SUMATERA SELATAN	2875	1143	532	150	84	2938	...	90
7	BENGGKULU	1148	403	134	66	24	1180	...	30
8	LAMPUNG	2350	1242	517	232	60	2499	...	70
...
34	PAPUA	1640	408	142	79	41	1979	...	54

Keterangan Atribut:

Prov	= Nama Provinsi di Indonesia
SD2011	= Jumlah fasilitas Sekolah Dasar tahun 2011
SMP2011	= Jumlah fasilitas Sekolah Menengah Pertama tahun 2011
SMU2011	= Jumlah fasilitas Sekolah Menengah Umum tahun 2011
SMK2011	= Jumlah fasilitas Sekolah Menengah Kejuruan tahun 2011
PT2011	= Jumlah fasilitas Perguruan Tinggi tahun 2011
SD2014	= Jumlah fasilitas Sekolah Dasar tahun 2014
SMP2014	= Jumlah fasilitas Sekolah Menengah Pertama tahun 2014
SMU2014	= Jumlah fasilitas Sekolah Menengah Umum tahun 2014
SMK2014	= Jumlah fasilitas Sekolah Menengah Kejuruan tahun 2014
PT2014	= Jumlah fasilitas Perguruan Tinggi tahun 2014
SD2018	= Jumlah fasilitas Sekolah Dasar tahun 2018
SMP2018	= Jumlah fasilitas Sekolah Menengah Pertama tahun 2018
SMU2018	= Jumlah fasilitas Sekolah Menengah Umum tahun 2018
SMK2018	= Jumlah fasilitas Sekolah Menengah Kejuruan tahun 2018
PT2018	= Jumlah fasilitas Perguruan Tinggi tahun 2018

3.2 Pre-Processing Data

Sebelum melakukan proses clustering, data harus dipre-processing terlebih dahulu. Hasil dari pre-processing data dapat dilihat pada tabel 2.

Table 2. Data Hasil Pre-Processing

NO	PROV	SD2011	SMP2011	SMU2011	SMK2011	PT2011	SD2014	...	PT2018
1	ACEH	0,361	0,209	0,238	0,088	0,313	0,377	...	0,264
2	SUMATERA UTARA	0,544	0,426	0,474	0,450	0,493	0,573	...	0,470
3	SUMATERA BARAT	0,088	0,108	0,117	0,098	0,292	0,102	...	0,266
4	RIAU	0,163	0,193	0,198	0,123	0,145	0,185	...	0,169
5	JAMBI	0,129	0,118	0,107	0,061	0,056	0,146	...	0,092
6	SUMATERA SELATAN	0,318	0,238	0,257	0,110	0,206	0,326	...	0,198
7	BENGGKULU	0,108	0,057	0,037	0,030	0,029	0,112	...	0,040
8	LAMPUNG	0,254	0,262	0,249	0,189	0,136	0,273	...	0,145
...
33	PAPUA	0,168	0,058	0,041	0,042	0,080	0,209	...	0,103

3.3 Proses Clustering dan Validasi Cluster

Ada 3 algoritma yang digunakan dalam pengelompokan data jumlah desa yang memiliki fasilitas di Indonesia yaitu algoritma K-Means, algoritma K-Medoids, dan algoritma Fuzzy C-Means. Berdasarkan hasil pre-processing pada tahapan sebelumnya, terdapat 33 data jumlah desa yang dapat digunakan pada pengolahan data dengan tiga algoritma tersebut.

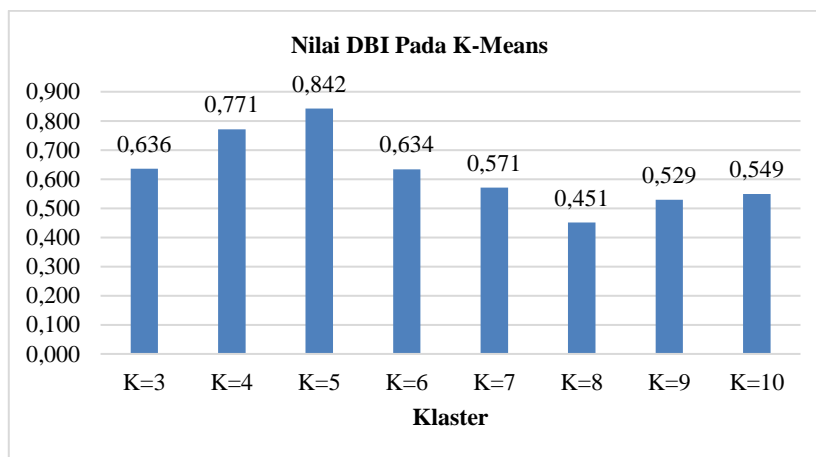
3.3.1 K-Means

Setelah tahapan pre-processing data selesai dilakukan, kemudian dilakukan pengelompokan dengan percobaan pada K=3 sampai dengan K=10 menggunakan algoritma K-Means. Berikut hasil klusterisasi dari algoritma K-Means yang bisa dilihat dari tabel berikut:

Tabel 3. Hasil Klusterisasi Pada Algoritma K-Means

Percobaan	Klaster									
	1	2	3	4	5	6	7	8	9	10
K=3	23	7	3	-	-	-	-	-	-	-
K=4	2	10	2	19	-	-	-	-	-	-
K=5	6	12	11	2	2	-	-	-	-	-
K=6	2	12	6	1	11	1	-	-	-	-
K=7	2	11	7	9	1	2	1	-	-	-
K=8	10	1	1	10	7	1	2	1	-	-
K=9	4	1	1	11	2	11	1	10	2	-
K=10	9	1	1	1	1	4	10	3	2	1

Tahapan selanjutnya setelah proses klusterisasi dilakukan adalah tahap validitas cluster dengan metode DBI untuk menemukan cluster terbaik. Hasil dari validitas kluster K-Means dengan menggunakan DBI adalah sebagai berikut:



Gambar 2. Nilai DBI Pada Algoritma K-Means

Rata-rata dari DBI pada algoritma K-Means yang diperoleh setelah menjumlahkan nilai DBI masing-masing cluster adalah 0,623. Nilai DBI terkecil K-Means diperoleh pada K=8 yaitu sebesar 0,451.

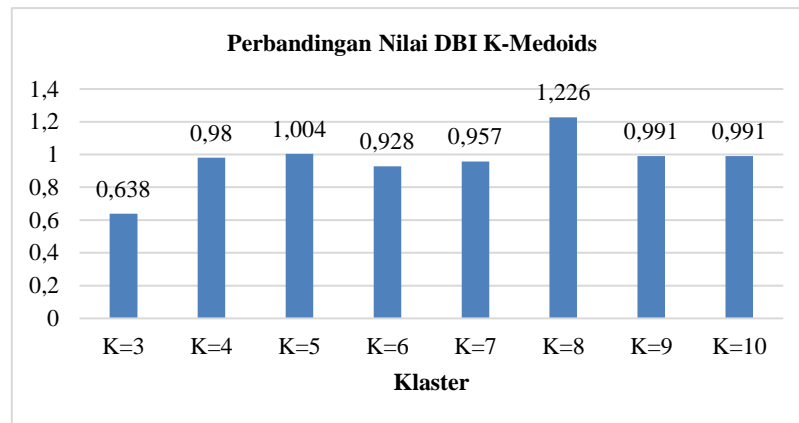
3.3.2 K-Medoids

Pada tahap ini dilakukan percobaan yang sama dengan jumlah percobaan pada K-Means. Berikut hasil klusterisasi dari algoritma K-Medoids yang dapat dilihat dari tabel dibawah:

Tabel 4. Hasil Klusterisasi Pada Algoritma K-Medoids

Percobaan	Klaster									
	1	2	3	4	5	6	7	8	9	10
K=3	6	24	3	-	-	-	-	-	-	-
K=4	13	11	6	3	-	-	-	-	-	-
K=5	5	3	11	11	3	-	-	-	-	-
K=6	2	5	10	4	9	3	-	-	-	-
K=7	4	8	4	11	3	2	1	-	-	-
K=8	4	4	3	11	3	5	2	1	-	-
K=9	3	5	6	6	7	2	1	1	2	-
K=10	4	5	4	1	7	2	2	1	3	4

Tahapan selanjutnya setelah proses klusterisasi dilakukan adalah tahap validitas cluster dengan metode DBI untuk menemukan cluster terbaik. Hasil dari validitas kluster K-Medoids dengan menggunakan DBI adalah sebagai berikut:



Gambar 3. Nilai DBI Pada Algoritma K-Medoids

Rata-rata dari DBI pada algoritma K-Medoids yang diperoleh setelah menjumlahkan nilai DBI masing-masing cluster adalah 0,964. Nilai DBI K-Medoids terkecil diperoleh pada K=3 yaitu 0,638.

3.3.3 Fuzzy C-Means

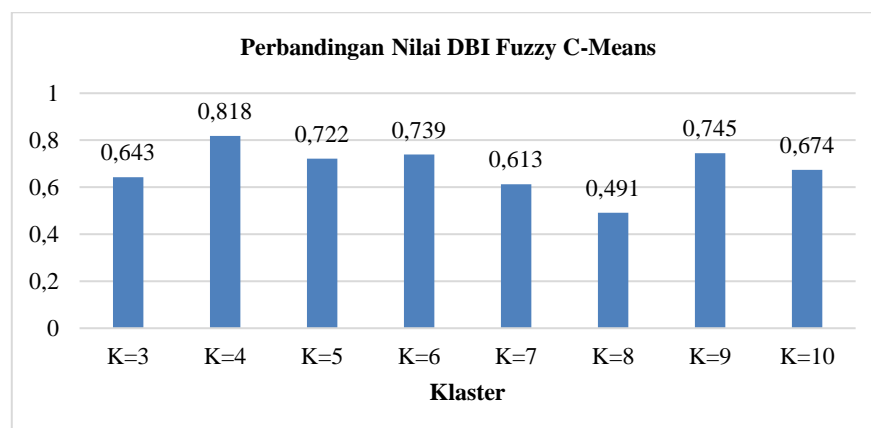
Tahapan berikutnya adalah clustering dengan algoritma Fuzzy C-Means dengan melakukan percobaan yang sama yaitu K=3 sampai K=10. Berikut hasil klusterisasi dari algoritma Fuzzy C-Means yang dapat dilihat dari tabel dibawah:

Tabel 4. Hasil Klusterisasi Pada Algoritma Fuzzy C-Means

Percobaan	Kluster									
	1	2	3	4	5	6	7	8	9	10
K=3	3	22	8							
K=4	11	12	7	3						
K=5	6	11	12	1	3					
K=6	6	2	1	2	11	11				
K=7	2	10	1	1	7	3	9			
K=8	1	9	10	1	1	3	7	1		
K=9	1	3	7	5	2	7	1	5	2	
K=10	5	2	2	4	7	2	1	8	1	1

Setelah proses klusterisasi dilakukan adalah tahap validitas cluster dengan metode DBI untuk menemukan cluster terbaik. Hasil dari validitas kluster Fuzzy C-Means dengan menggunakan DBI adalah dapat dilihat pada Gambar 4.

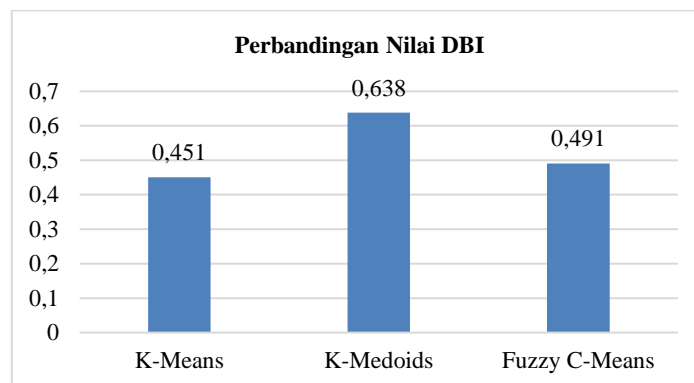
Rata-rata DBI pada algoritma Fuzzy C-Means yang diperoleh setelah menjumlahkan nilai DBI masing-masing cluster adalah 0,681. Nilai DBI Fuzzy C-Means terkecil diperoleh pada K=8 yaitu sebesar 0,491.



Gambar 4. Nilai DBI Pada Algoritma Fuzzy C-Means

3.3.4 Perbandingan Hasil DBI Algoritma

Setelah pengelompokan dari masing-masing algoritma selesai dilakukan, tahapan berikutnya adalah membandingkan nilai DBI pada masing-masing algoritma. Tujuannya adalah untuk mengetahui algoritma terbaik dalam melakukan pengelompokan data. Perbandingan nilai DBI adalah sebagai berikut:



Gambar 5. Perbandingan Nilai DBI Unsupervised Learning

Dari perbandingan 3 algoritma yang telah dilakukan, terbukti bahwa K-Means memiliki nilai DBI terkecil yaitu 0,451 kemudian diikuti oleh Fuzzy C-Means yaitu 0,491 dan terakhir algoritma K-Medoids yaitu 0,638. Rata-rata nilai DBI pada 3 algoritma diatas adalah sebesar 0,527.

4. KESIMPULAN

Berdasarkan hasil dan analisis yang telah dilakukan, Nilai DBI yang paling kecil diperoleh pada algoritma K-Means yakni ditemukan pada K=8 sebesar 0,451. Dapat diambil kesimpulan bahwa K-Means terbukti lebih baik daripada 2 algoritma lainnya, yaitu algoritma K-Medoids dan juga Fuzzy C-Means dalam melakukan pengelompokan data desa yang Memiliki Fasilitas Sekolah Menurut Provinsi di Indonesia. Maka diperoleh hasil dari K=8 yang terdiri dari yaitu kluster 1 sebanyak 10 provinsi, kluster 2 sebanyak 1 provinsi, kluster 3 sebanyak 1 provinsi, kluster 4 sebanyak 10 provinsi, kluster 5 sebanyak 7 provinsi, kluster 6 sebanyak 1 provinsi, kluster 7 sebanyak 2 provinsi dan kluster 8 sebanyak 1 provinsi. Dari hasil pengelompokan tersebut, daerah yang memiliki fasilitas sekolah rendah yaitu pada kluster 1 yakni provinsi Papua Barat, Maluku, Kepulauan Bangka Belitung, Maluku Utara, Bengkulu, Gorontalo, DI Yogyakarta, Kepulauan Riau, Bali, dan Sulawesi Barat. Saran kepada pemerintah adalah supaya lebih memprioritaskan pembangunan fasilitas sekolah yang ada pada daerah-daerah tersebut untuk meningkatkan pemerataan akses dan fasilitas pendidikan di Indonesia demi kemajuan pendidikan di Indonesia.

REFERENSI

- [1] A. Faiz and I. Kurniawaty, "Faiz, Aiman Kurniawaty, Imas," vol. 12, no. 2, pp. 155–164, 2020.
- [2] D. A. Nurhayati and Ambari, "Jurnal Pendidikan Kewarganegaraan Undiksha Vol. 8 No. 2 (Mei, 2020)," *J. Pendidik. Kewarganegaraan Undika*, vol. 8, no. 2, pp. 157–167, 2020, [Online]. Available: <https://www.google.com/url?sa=t&source=web&rct=j&url=https://ejournal.undiksha.ac.id/index.php/JJPP/article/view/25434&ved=2ahUKEwic5fep2PHsAhUSeYKHU8fBGQQFjAAegQICRAC&usg=AOvVaw3Tezzh5erg5X3r51XrDwtW>
- [3] S. F. N. Fitri, "Problematika Kualitas Pendidikan di Indonesia," *J. Pendidik. Tambusai*, vol. 5, no. 1, pp. 1617–1620, 2021.
- [4] T. J. V. Aristo, "Analisis permasalahan pemerataan pendidikan di Kabupaten Sintang," *J. Akuntabilitas Manaj. Pendidik.*, vol. 7, no. 1, pp. 25–34, 2019, doi: 10.21831/amp.v7i1.10923.
- [5] M. T. Islam, P. K. Basak, P. Bhowmik, and M. Khan, "Data Clustering Using Hybrid Genetic Algorithm with k-Means and k-Medoids Algorithms," *ICSEC 2019 - 23rd Int. Comput. Sci. Eng. Conf.*, pp. 123–128, 2019, doi: 10.1109/ICSEC47112.2019.8974797.
- [6] 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," vol. 2, no. October, pp. 76–83, 2022.
- [7] A. E. Prमितasari and Y. Nataliani, "Perbandingan Clustering Karyawan Berdasarkan Nilai Kinerja Dengan Algoritma K-Means Dan Fuzzy C-Means," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 3, pp. 1119–1132, 2021, doi: 10.35957/jatisi.v8i3.957.
- [8] 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.
- [9] D. S. Maylawati, T. Priatna, H. Sugilar, and M. A. Ramdhani, “Data science for digital culture improvement in higher education using K-means clustering and text analytics,” *Int. J. Electr. Comput. Eng.*, vol. 10, no. 5, pp. 4569–4580, 2020, doi: 10.11591/IJECE.V10I5.PP4569-4580.
- [10] D. F. Pramesti, Lahan, M. Tanzil Furqon, and C. Dewi, “Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 9, 2017.
- [11] R. A. Farissa, R. Mayasari, and Y. Umaidah, “Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangasambung,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.
- [12] F. H. Kuwil, F. Shaar, A. E. Topcu, and F. Murtagh, “A new data clustering algorithm based on critical distance methodology,” *Expert Syst. Appl.*, vol. 129, pp. 296–310, 2019, doi: 10.1016/j.eswa.2019.03.051.
- [13] 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.
- [14] I. Kamila, U. Khairunnisa, and M. Mustakim, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau,” *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.
- [15] D. Kurmiati, M. Zakiy Fauzi, Ripangi, A. Falegas, and Indria, “Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm Klasterisasi Daerah Rawan Gempa Bumi di Indonesia Menggunakan Algoritma K-Medoids,” *Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 47–57, 2021.
- [16] X. Luo, “Digital Art Design Effectiveness Model System Based on K-Medoids Algorithm,” *Adv. Multimed.*, vol. 2022, 2022, doi: 10.1155/2022/2901815.
- [17] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R.H.Zer, and D. Hartama, “Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia,” *J. Teknol. Inf.*, vol. 4, no. 1, pp. 166–173, 2020, doi: 10.36294/jurti.v4i1.1296.
- [18] A. K. Dubey, U. Gupta, and S. Jain, “Comparative study of K-means and fuzzy C-means algorithms on the breast cancer data,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 1, pp. 18–29, 2018, doi: 10.18517/ijaseit.8.1.3490.
- [19] S. Fitri Octavia, “Penerapan K-Means dan Fuzzy C-Means untuk Pengelompokan Data Kasus Covid-19 di Kabupaten Indragiri Hilir,” *Technol. Sci.*, vol. 3, no. 2, pp. 88–94, 2021, doi: 10.47065/bits.v3i2.1005.
- [20] Widiarina and R. S. Wahono, “Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam Pemetaan Nasabah Potensial,” *J. Intell. Syst.*, vol. 1, no. 1, pp. 33–35, 2015.