

Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Penduduk Kota Bekasi Berdasarkan Tingkat Pendidikan (2021-2023)

Hamdun Sulaiman^{1*}, Aifah Nurul Faizah², Muhamad Abdul Ghani³, Ramdhan Saepul Rohman⁴,
Wawan Kurniawan⁵

^{1,2,3,5}Program Studi Sistem Informasi, Universitas Bina Sarana Informatika

⁴Program Studi Informatika, Universitas Bina Sarana Informatika

*hamdun.hsl@bsi.ac.id

Abstrak

Pendidikan merupakan hal penting dalam upaya meningkatkan kualitas sumber daya manusia. Penelitian ini membandingkan algoritma K-Means dan K-Medoids untuk menganalisis distribusi tingkat pendidikan penduduk Kota Bekasi. Dataset yang digunakan berasal dari 56 kelurahan selama tahun 2021–2023 yang telah dirata-rata untuk memberikan gambaran yang lebih stabil. Proses clustering dilakukan menggunakan aplikasi RapidMiner, dengan evaluasi hasil berdasarkan Davies Bouldin Index (DBI) untuk mengukur validitas kelompok data. Hasil evaluasi menunjukkan bahwa algoritma K-Means menghasilkan performa yang lebih baik dibandingkan K-Medoids, ditunjukkan oleh nilai DBI terendah sebesar 0,558 pada $k = 4$, sedangkan algoritma K-Medoids menghasilkan nilai DBI terbaik sebesar 0,803 pada $k = 3$. Nilai DBI yang lebih rendah pada K-Means menunjukkan kualitas cluster yang lebih baik, sehingga K-Means dipilih sebagai metode terbaik untuk pemetaan tingkat pendidikan di Kota Bekasi. Selain itu, hasil K-Means juga konsisten dengan metode Elbow yang menunjukkan jumlah cluster optimal pada $k = 4$, sedangkan K-Medoids menunjukkan ketidaksesuaian antara hasil Elbow Method ($k = 5$) dan DBI ($k = 3$). Hasil clustering K-Means membentuk empat cluster dengan karakteristik tingkat pendidikan yang berbeda. Cluster 0 didominasi oleh pendidikan menengah sehingga memerlukan program beasiswa dan pelatihan kejuruan. Cluster 1 memiliki tingkat pendidikan tinggi sehingga perlu difokuskan pada pengembangan infrastruktur pendidikan lanjutan dan penciptaan lapangan kerja berkualitas. Cluster 2 memiliki tingkat pendidikan rendah hingga menengah sehingga membutuhkan peningkatan akses pendidikan dan program penyuluhan. Sementara itu, Cluster 3 menunjukkan distribusi pendidikan menengah hingga tinggi yang relatif merata sehingga memerlukan peningkatan kualitas pendidikan dan pelatihan profesional guna mendukung pengembangan sumber daya manusia di Kota Bekasi.

Kata kunci : Clustering, K-Means, K-Medoids, Tingkat Pendidikan, Kota Bekasi.

Abstract

Education is crucial for improving human resource quality. This study compares the K-Means and K-Medoids algorithms to analyze the distribution of education levels in Bekasi City. The dataset used was derived from 56 urban villages during 2021–2023, which were averaged to provide a more stable picture. The clustering process was performed using the RapidMiner application, with the results evaluated based on the Davies Bouldin Index (DBI) to measure the validity of the data groups. The evaluation results showed that the K-Means algorithm produced better performance than K-Medoids, indicated by the lowest DBI value of 0.558 at $k = 4$, while the K-Medoids algorithm produced the best DBI value of 0.803 at $k = 3$. The lower DBI value in K-Means indicates better cluster quality, thus K-Means was chosen as the best method for mapping education levels in Bekasi City. In addition, the K-Means results are also consistent with the Elbow method which shows the optimal number of clusters at $k = 4$, while K-Medoids shows a discrepancy between the results of the Elbow Method ($k = 5$) and DBI ($k = 3$). The K-Means clustering results form four clusters with different educational level characteristics. Cluster 0 is dominated by secondary education so that it requires scholarship programs and vocational training. Cluster 1 has a higher level of education so that it needs to focus on developing further education infrastructure and creating quality jobs. Cluster 2 has a low to middle level of education so that it requires increasing access to education and extension programs. Meanwhile, Cluster 3 shows a relatively even distribution of secondary to higher education so

that it requires improving the quality of education and professional training to support human resource development in Bekasi City.

Keywords : Clustering, K-Means, K-Medoids, Education Levels, Bekasi City.

1. Pendahuluan

Salah satu cara untuk meningkatkan kualitas sumber daya manusia adalah melalui pendidikan^{[1][2]}. Peningkatan kualitas pendidikan pada sebuah daerah tertentu diharapkan dapat meningkatkan kesejahteraan manusia^{[3][4]}. Pendidikan juga merupakan faktor terpenting dalam kehidupan, karena berkaitan dengan kualitas sumber daya manusia sebagai kunci pertumbuhan ekonomi pada suatu wilayah atau negara. Pemerintah Indonesia telah merancang program untuk memajukan pendidikan di Indonesia yang mencakup pendidikan dasar dan menengah, yaitu program pendidikan wajib belajar 12 tahun yang mana merupakan kelanjutan dari program 9 tahun wajib belajar sebelumnya^{[5][6]}.

Terdapat tiga tingkatan pendidikan formal di Indonesia. Tingkatan dasar, yaitu Sekolah Dasar (SD), dan Sekolah Lanjutan Tingkat Pertama (SLTP). Tingkat menengah yang terdiri dari Sekolah Lanjutan Tingkat Atas (SLTA) serta Sekolah Menengah Kejuruan (SMK), diikuti tingkat pendidikan tinggi yang berupa program pendidikan diploma, sarjana, magister, dan doktor spesialis ^[7].

Kota Bekasi merupakan salah satu kota besar di Indonesia yang terdiri dari 56 kelurahan dan berdiri dengan luas wilayah sekitar 210,49 km². Menurut statistik Open Data Kota Bekasi yang dikelola oleh Diskominfo Jawa Barat sebagai walidata seluruh data yang berasal dari Pemda Kota Bekasi, jumlah penduduk di Kota Bekasi pada tahun 2021 terdapat ± 2.46 juta jiwa, 2022 ± 2.48 juta jiwa, dan 2023 adalah ± sebanyak 2.51 juta jiwa.

Penelitian ini bertujuan untuk membandingkan performa kedua algoritma dalam pemetaan kelompok penduduk berdasarkan tingkat penyelesaian pendidikan. Data dalam penelitian ini dikumpulkan melalui teknik pengamatan (observasi). Jumlah data dalam dataset yang akan diolah dalam penelitian ini yaitu 168 data dengan 10 atribut tingkat penyelesaian pendidikan penduduk Kota Bekasi berdasarkan kelurahan, dengan rentan waktu mulai dari tahun 2021-2023 dengan kebaruan, telah diperbarui pada bulan September 2024. Penggunaan dataset yang relevan dengan kondisi saat ini (terbaru) dapat memberikan wawasan yang lebih tepat dan menghasilkan pertimbangan keputusan yang lebih baik^[3]. Dataset didapatkan dengan kriteria public melalui riset website resmi Open

Data Kota Bekasi, dengan bentuk format dataset xlxs, serta satuan angka “Orang”.

Metode analisis seperti K-Means dan K-Medoids telah banyak digunakan pada berbagai penelitian sebelumnya, namun hingga saat ini, jika ditinjau melalui kepustakaan, belum ditemukan penelitian spesifik yang menguji identifikasi pola distribusi pendidikan di Kota Bekasi. Penelitian perbandingan antara kedua metode ini dilakukan sebagai upaya untuk mengisi kesenjangan tersebut, dan mengetahui mana algoritma yang lebih efektif dalam mengidentifikasi wilayah (kelurahan) dengan tingkat penyelesaian pendidikan yang berbeda, baik rendah, maupun tinggi. Proses clustering dilakukan menggunakan aplikasi RapidMiner, dengan evaluasi hasil berdasarkan Davies Bouldin Index (DBI) untuk mengukur validitas kelompok data. Untuk evaluasi hasil klaster yang sudah terbentuk oleh K-Means dan K-Medoids digunakan Davies Bouldin Index (DBI). DBI adalah metode pengukuran validitas internal yang menghitung tingkat klasterisasi dari hasil perhitungan jumlah kriteria dan fitur lainnya dari dataset yang ada^[14]. Hasil akhir yang menunjukkan DBI terbaik dan relevan antara kedua algoritma tersebut dapat digunakan untuk memberikan saran evaluasi distribusi pendidikan, serta memberikan insight (wawasan) yang bermanfaat bagi masyarakat luas, khususnya Kota Bekasi.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Penulis melakukan analisa terhadap penelitian sebelumnya, melihat metode apa saja yang telah digunakan, dan melihat bagaimana penelitian sebelumnya pernah dilakukan yang berhubungan dengan penelitian ini, terdapat beberapa penelitian terkait:

1. Penelitian pertama yang membandingkan algoritma K-Means dan K-Medoids di antaranya penelitian dengan judul “The Comparison Of K-Means And K-Medoids Algorithms For Clustering The Spread Of The Covid-19 Outbreak In Indonesia”, dengan menggunakan dataset Covid-19 menghasilkan nilai DBI sebesar pada 0,064 pada K-Means dan nilai DBI sebesar 0,411 pada algoritma K-Medoids ^[15].

2. Penelitian kedua tentang, “*Comparative Analysis of K-Means and K-Medoids Clustering Methods on Weather Data of Denpasar City*” yang dilakukan terkait perbandingan algoritma K-means dan K-Medoids untuk menganalisis data cuaca di Kota Denpasar, Bali. Algoritma K-Means terbukti lebih efektif dalam pengelompokan cuaca dibandingkan K-Medoids, dengan hasil validitas optimal sebesar 1,795 untuk periode curah hujan 1 jam. Penelitian ini menunjukkan bahwa penggunaan algoritma K-Means dapat memberikan pemahaman yang lebih baik tentang variasi cuaca dan informasi yang lebih akurat

untuk perencanaan dan pengambilan keputusan terkait cuaca^[9].

3. Penelitian berikutnya dengan judul “Implementasi Metode K-Medoids clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan” yang mengelompokkan provinsi di Indonesia berdasarkan indikator pendidikan menggunakan algoritma K-Medoids dengan hasil data indikator pendidikan di Indonesia tahun ajaran 2019/2020 menggunakan metode K-Medoids dan Silhouette Coefficient menghasilkan tiga cluster pendidikan. Cluster 1 terdiri dari 6 provinsi dengan karakteristik indikator pendidikan kategori tinggi, cluster 2 mencakup 13 provinsi dengan indikator pendidikan sedang, dan cluster 3 berisi 15 provinsi dengan indikator pendidikan rendah^[5].

4. Pada penelitian selanjutnya yang berjudul “Penerapan Algoritma K-Medoids pada Pengelompokan Wilayah Provinsi di Indonesia Berdasarkan Indikator Pendidikan”, mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator pendidikan tahun 2020. Hasil klaster divalidasi menggunakan Silhouette Coefficient (SC) untuk menilai kualitas dan kekuatan klaster tersebut. Dari penelitian ini ditemukan bahwa klaster optimal terdiri dari 2 cluster dengan nilai SC sebesar 0.464. Klaster 1 terdiri dari 14 provinsi, sementara klaster 2 mencakup 20 provinsi^[12]

2.2. Landasan Teori

1. Data Mining

Data mining adalah serangkaian teknik ekstraksi pengetahuan yang berguna untuk menemukan pola khusus dari kumpulan data sehingga menghasilkan informasi nilai tambah yang sebelumnya tidak diketahui ^[7].

2. Model CRISP-DM

CRISP-DM (Cross Industry Standard for Data mining) merupakan metode yang banyak digunakan oleh para ahli dengan menggunakan proses pemodelan data di dalamnya, Tujuan dari metode CRISP-DM ini yaitu untuk menemukan pola yang menarik dan memiliki makna pada data yang digunakan ^[5].



Gambar 1. Model CRISP-DM

3. Metode Clustering

Clustering adalah suatu metode dalam pengolahan data yang digunakan untuk memisahkan sekumpulan data ke dalam beberapa kelompok yang berbeda, di mana setiap kelompok atau himpunan data disebut sebagai klaster yang memiliki karakteristik yang sama ^[2]

4. Algoritma K-Means

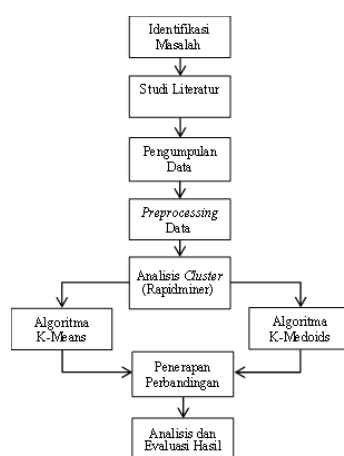
Algoritma pertama yang digunakan untuk mengatasi permasalahan penelitian adalah algoritma K-Means. Algoritma K-Means merupakan metode pengelompokan non-hirarki yang digunakan dalam analisis data atau metode data mining yang bersifat unsupervised learning^[1].

5. Algoritma K-Medoids

Algoritma K-Medoids adalah metode clustering partisional yang bertujuan untuk menemukan set k-cluster antara data yang paling mencirikan objek pada kumpulan data ^[8].

2.3. Tahapan Penelitian

Berikut adalah untuk menerapkan perbandingan K-Means dan K-Medoids untuk clustering penduduk Kota Bekasi berdasarkan tingkat Pendidikan :



Gambar 2. Tahap Penelitian

Penelitian diawali dengan identifikasi masalah terkait pengelompokan wilayah Kota Bekasi

berdasarkan tingkat pendidikan menggunakan algoritma K-Means dan K-Medoids. Selanjutnya dilakukan studi literatur untuk memperoleh landasan teori dan referensi penelitian. Data diperoleh dari Open Data Kota Bekasi berupa data tingkat pendidikan penduduk pada 56 kelurahan tahun 2021–2023. Data kemudian melalui tahap preprocessing untuk membersihkan data yang tidak lengkap atau tidak valid. Proses clustering dilakukan menggunakan aplikasi RapidMiner dengan algoritma K-Means dan K-Medoids. Hasil clustering selanjutnya dibandingkan dan dievaluasi menggunakan Davies Bouldin Index (DBI) untuk menentukan metode dan jumlah cluster yang paling optimal. Tahap akhir adalah interpretasi hasil cluster serta penyusunan rekomendasi berdasarkan karakteristik tingkat pendidikan pada masing-masing kelompok wilayah

3. Metode Penelitian

3.1. Lokasi dan Waktu penelitian

Penelitian ini dilakukan di Kota Bekasi, Jawa Barat, dengan menggunakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik Kota Bekasi. Data yang digunakan berupa jumlah penduduk berdasarkan tingkat pendidikan pada 56 kelurahan selama periode 2021–2023. Untuk memperoleh data yang lebih stabil, dilakukan perhitungan rata-rata dari data ketiga tahun tersebut sebelum dianalisis. Teknik

analisis data menggunakan metode data mining dengan pendekatan clustering melalui algoritma K-Means dan K-Medoids yang diimplementasikan menggunakan RapidMiner. Penentuan jumlah cluster optimal dilakukan menggunakan Elbow Method, sedangkan kualitas cluster dievaluasi menggunakan Davies Bouldin Index (DBI). Hasil evaluasi kemudian digunakan untuk membandingkan performa kedua algoritma dan menentukan metode terbaik dalam mengelompokkan tingkat pendidikan penduduk di Kota Bekasi.

3.2. Sumber Data

Data dalam penelitian ini dikumpulkan melalui teknik pengamatan (observasi). Jumlah data dalam dataset yang akan diolah dalam penelitian ini yaitu 168 data dengan 10 atribut tingkat penyelesaian pendidikan penduduk Kota Bekasi berdasarkan kelurahan, dengan rentan waktu mulai dari tahun 2021-2023 dengan kebaruan, telah diperbarui pada bulan September 2024. Dataset didapatkan dengan kriteria public melalui riset website resmi Open Data Kota Bekasi, dengan bentuk format dataset xlxs, serta satuan angka "Orang".

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

Pada bagian ini disajikan hasil dari setiap tahapan penelitian berdasarkan metode CRISP-DM,

meliputi Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Hasil yang ditampilkan mencakup proses pengolahan data tingkat pendidikan penduduk Kota Bekasi, pembentukan cluster menggunakan algoritma K-Means dan K-Medoids, penentuan jumlah cluster optimal menggunakan Elbow Method, serta evaluasi kualitas cluster menggunakan Davies Bouldin Index (DBI).

1. Business Understanding

Pada tahap Business Understanding, fokus penelitian adalah memahami tujuan utama yang ingin dicapai, yaitu melakukan clustering penduduk Kota Bekasi berdasarkan tingkat pendidikan menggunakan algoritma K-Means dan K-Medoids. Penelitian ini bertujuan untuk membandingkan performa kedua algoritma dalam pemetaan kelompok penduduk berdasarkan tingkat penyelesaian pendidikan.

Analisis awal pada tahap ini terhadap dataset, yaitu dilakukan pemahaman struktur data, mendeteksi inkonsistensi, dan mengidentifikasi atribut yang relevan. Hasil dari fase ini akan menjadi dasar bagi Data Preparation, ketika dataset diproses lebih lanjut melalui pembersihan, normalisasi, dan agregasi. Dataset yang telah disiapkan kemudian digunakan pada fase Modeling, saat algoritma K-Means dan K-Medoids diaplikasikan untuk membentuk cluster optimal berdasarkan kemiripan data. Ketiga fase ini

bekerja secara berkesinambungan untuk memastikan hasil clustering yang akurat dan representatif terhadap kondisi pendidikan di Kota Bekasi saat ini.

2. Data Understanding

Tahap Data Understanding dimulai dengan mengumpulkan dataset penduduk Kota Bekasi yang diperoleh dari website resmi Open Data Kota Bekasi untuk periode 2021-2023. Data ini berisi variabel-variabel berupa wilayah Kota Bekasi yang terbagi menjadi 56 kelurahan, serta berbagai tingkat pendidikan "tidak atau belum sekolah", "belum tamat sekolah dasar sederajat", "tamat sekolah dasar sederajat", "sekolah lanjutan tingkat pertama sederajat", "sekolah lanjutan tingkat atas sederajat", "diploma satu atau dua", "diploma tiga", "strata satu", strata dua", dan "strata tiga".

3. Data Preparation

Pada tahap Data Preparation, dilakukan preprocessing data untuk memastikan kualitas data yang digunakan dalam analisis. Eksplorasi awal dilakukan menggunakan Microsoft Excel 2019 untuk memahami distribusi data, serta mengidentifikasi missing values dan inkonsistensi dalam variabel-variabel yang digunakan dalam analisis clustering pada penelitian ini. Penanganan inkonsistensi data ini dilakukan penghapusan record data yang tidak sesuai.

JATIRANGGON	5099	2904	2301	2765	10648
JATIRANGGA	3594	1879	2703	1985	5340
JATIRADEN	4321	2203	2442	2322	6904
PADURENAN	11274	6364	5660	5801	17652
CIMUNING	8739	5009	3198	3909	14874
MUSTIKAJAYA	14151	8675	5997	75161	27567
MUSTIKASARI	7029	4214	3686	4036	13889
JATIRAHAYU	10159	5176	4623	5514	22274
JATIWARNA	4045	2221	2102	2393	8845
JATIMELATI	3560	2157	1567	1933	7044
JATIMURNI	4775	2597	2589	2903	9856

Gambar 3. Inkonsistensi Pada Dataset

Data yang terdapat inkonsistensi atau kesalahan baca pada program dinormalisasi, kemudian dipisahkan menjadi dataset siap uji untuk memastikan validitas hasil clustering. Proses persiapan ini juga melibatkan penghitungan rata-rata dari jumlah penduduk berdasarkan tingkat pendidikan selama periode tiga tahun (2021-2023). Dataset 'mentah' yang diolah pada fase Data Preparation akan menghasilkan data bersih yang nantinya digunakan langsung untuk proses modeling.

	AK	AG	AS	AT	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF
data_jumlahpenduduk	jumlah_penduduk_tahun_2021	jumlah_penduduk_tahun_2022	jumlah_penduduk_tahun_2023	jumlah_penduduk_tahun_2024	jumlah_penduduk_tahun_2025	jumlah_penduduk_tahun_2026	jumlah_penduduk_tahun_2027	jumlah_penduduk_tahun_2028	jumlah_penduduk_tahun_2029	jumlah_penduduk_tahun_2030	jumlah_penduduk_tahun_2031	jumlah_penduduk_tahun_2032	jumlah_penduduk_tahun_2033	jumlah_penduduk_tahun_2034	jumlah_penduduk_tahun_2035
1	BEKASIREA	12479	4431	4909	6338	24886	418	2397	6972	535	NAVER				
2	MARGARET	13210	6104	5149	6831	27680	362	2142	8489	428	AXEL				
3	JERONIA	14587	6886	6774	7881	36129	439	2814	7551	638	YU				
4	ABENALIA	12461	2803	3229	8528	28826	413	2890	4893	345	WZ				
5	SEVIA	12786	5366	4387	8846	24432	386	2181	4659	512	YU				
6	ARAB	8891	4786	3413	8123	21880	392	1388	2897	192	YU				
7	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
8	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
9	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
10	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
11	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
12	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
13	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
14	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
15	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
16	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
17	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
18	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
19	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
20	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
21	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				
22	SEVIA	8771	5366	4387	8846	24432	386	2181	4659	512	YU				

Gambar 4. Proses Agregasi Data 3 Tahun Menjadi Rata-rata.

Penggunaan data rata-rata bertujuan untuk memberikan gambaran yang lebih stabil dan representatif mengenai tingkat pendidikan di setiap wilayah Kota Bekasi ke dalam tiga kategori pendidikan (tinggi, menengah, dan rendah), tanpa dipengaruhi oleh perubahan tidak tetapan pada variabel yang mungkin terjadi tiap tahunnya.

Ketiga kategori kriteria pendidikan ini juga digunakan untuk mempermudah interpretasi hasil clustering.

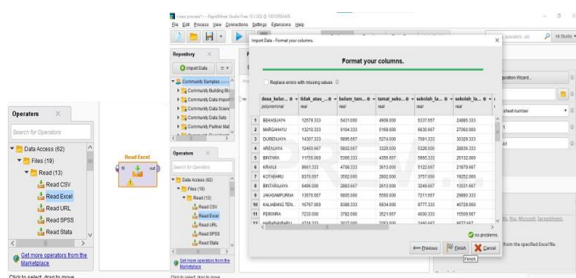
Tabel 1 Hasil Standarisasi Atribut Dataset Penelitian (2021-2023)

Desa/kelurahan	tidak atau belum sekolah	belum tamat sekolah dasar	tamat sekolah dasar sederajat	belum tamat sekolah lanjutan pertama sederajat	tamat sekolah lanjutan pertama sederajat	Diploma satu atau dua	Diploma tiga	strata satu	strata dua	strata tiga
BEKASIJAYA	12479	6431	4905	6338	24886	418	2397	6572	936	25
MARGABAYU	12319	6184	5169	6631	27069	342	2142	5492	428	38
DURENJAYA	14397	6894	5274	7981	39329	435	2614	7331	529	21
ARENJAYA	12461	5893	3329	5328	28836	413	2650	6583	345	17
JATINARNA	3991	2321	2397	2487	8925	81	759	2393	193	19
JATINELATI	3669	2139	1847	1997	7153	74	936	3136	347	11
JATINENGG	4887	2427	2357	2964	10975	118	842	2316	177	13

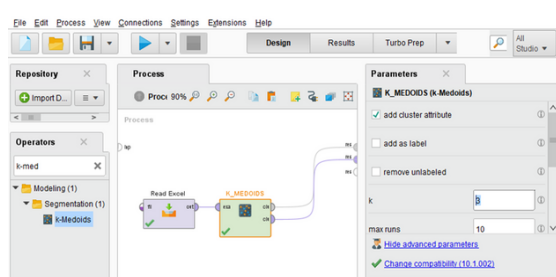
Sumber: Open Data Kota Bekasi, 2024 (Diolah oleh penulis)

4. Modelling

Fase modeling dilakukan dengan menguji perbandingan algoritma K-Means dan K-Medoids yang diterapkan pada dataset yang telah diproses sebelumnya terhadap jumlah kluster dan atribut yang ditentukan.



Gambar 5. Proses Import Dataset Ke RapidMiner
a. Menentukan jumlah cluster k secara acak pada parameter RapidMiner



Gambar 6. Proses Pemilihan Algoritma K-Means
b. Menentukan titik pusat (centroid & medoid) awal secara acak kemudian menghitung centroid dan medoid cluster k-1.

Tabel 1 Centroid Awal (RapidMiner)

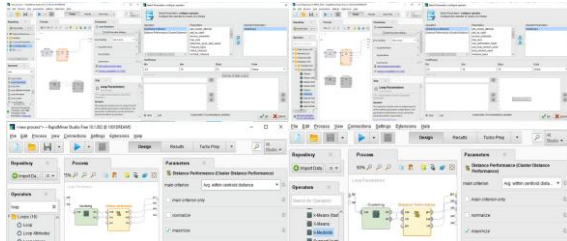
Atribut	C0	C1	C2
Tidak atau belum sekolah	8325.792	13363.643	4942.222
Belum tamat sekolah dasar sederajat	4273.208	6671.071	2575.944
Tamat sekolah dasar sederajat	3399.542	4907.214	2440.222
Sekolah lanjutan tingkat pertama sederajat	4137.583	8353.214	2676.444
Sekolah lanjutan tingkat atas sederajat	16440.833	29132.786	9167.728
Diploma satu atau dua	208.125	439.929	90.111
Diploma tiga	1527.542	2755.643	651.833
Strata satu	4521.167	7332.500	1924.167
Strata dua	293.583	524.643	60.667
Strata tiga	131.125	60.500	124

Tabel 2 Medoid Awal (RapidMiner)

Atribut	C0	C1	C2
Tidak atau belum sekolah	10197.667	4807.333	14701.667
Belum tamat sekolah dasar sederajat	5187.333	2627	8627.333
Tamat sekolah dasar sederajat	4507.667	2557	6022
Sekolah lanjutan tingkat pertama sederajat	5497.667	2964	30137
Sekolah lanjutan tingkat atas sederajat	22249.333	10075	27940
Diploma satu atau dua	324.667	118.333	456.667
Diploma tiga	2364.667	842.333	2976
Strata satu	6563	2214.667	7176.667
Strata dua	513	177.333	444
Strata tiga	29	13	19.333

c. Proses percobaan berlanjut dan dimulai lagi dari menggunakan 2 cluster dan dilanjutkan hingga 10 cluster untuk mendapatkan hasil cluster optimal yang tepat dengan menggunakan operator "Cluster Distance Performance" untuk membantu evaluasi kualitas hasil clustering untuk

setiap percobaan jumlah cluster. Proses menggunakan tambahan operator loop untuk mempersingkat waktu perhitungan.



Gambar 7. Variasi Jumlah Cluster K-Means dan K-Medoids Yang Diuji (dari 2 hingga 10)

Hasil rata-rata jarak dalam cluster untuk setiap nilai k yang diuji percobaan clustering algoritma K-Means dan K-Medoids pada Gambar 8 dapat dilihat pada tabel berikut:

Tabel 3 Rata-rata Jarak dalam Centroid untuk Variasi Jumlah Cluster

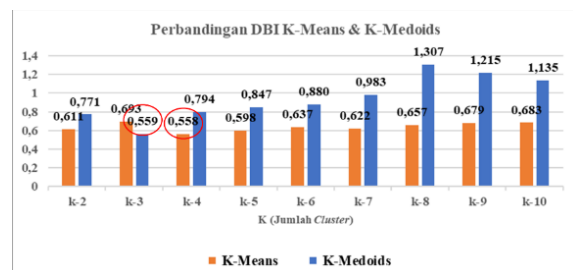
K	Avg. within centroid distance
2	37020692.070
3	25322907.415
4	15983990.167
5	10730733.710
6	7879895.796
7	6594820.967
8	5465386.476
9	4898351.875
10	4119865.545

Tabel 4 Rata-rata Jarak dalam Medoid untuk Variasi Jumlah Cluster

K	Avg. within centroid distance
2	48328370.036
3	36280765.583
4	28732401.865
5	14537558.425
6	13100752.992
7	10420288.506
8	11461667.185
9	10958671.206
10	8330336.254

5. Evaluasi

Tahap Evaluation melibatkan analisis hasil clustering dari algoritma K-Means dan K-Medoids. Hasil dari kedua model dibandingkan menggunakan DBI (Davies Bouldin Index) untuk mengevaluasi apakah objek berada dalam cluster sudah tepat. Dalam penggunaan RapidMiner, untuk mempermudah analisis hasil, digunakan visualisasi cluster dalam bentuk diagram batang yang dapat menunjukkan distribusi data dalam masing-masing cluster.



Gambar 8. Grafik Perbandingan Nilai DBI pada K-Means dan K-Medoids

6. Deployment

Fase Deployment merupakan hasil clustering untuk memberikan saran berdasarkan algoritma K-Means dengan k = 4. Pemilihan K-Means didasarkan pada evaluasi kualitas cluster yang konsisten, dengan DBI terendah (0,558) dan keselarasan dengan elbow method. K-Medoids tetap digunakan untuk perbandingan, namun hasilnya kurang relevan karena ketidaksesuaian jumlah cluster optimal (k = 5 pada elbow method, k = 3 pada DBI).

4.2. Pembahasan

Pada bagian ini dilakukan interpretasi dan analisis terhadap hasil clustering yang diperoleh. Pembahasan mencakup perbandingan performa algoritma K-Means dan K-Medoids berdasarkan nilai DBI, analisis karakteristik setiap cluster yang terbentuk, serta implikasi hasil penelitian terhadap pemetaan tingkat pendidikan di Kota Bekasi. Selain itu, dijelaskan alasan pemilihan algoritma K-Means sebagai metode terbaik karena menghasilkan nilai DBI yang lebih rendah dan jumlah cluster yang lebih konsisten dibandingkan K-Medoids

5. Kesimpulan

Berdasarkan hasil penelitian, algoritma K-Means menghasilkan performa yang lebih baik dibandingkan K-Medoids dalam mengelompokkan tingkat pendidikan penduduk Kota Bekasi. Hal ini ditunjukkan oleh nilai Davies Bouldin Index (DBI) yang lebih rendah, yaitu sebesar 0,558 pada $k = 4$, sedangkan K-Medoids menghasilkan nilai DBI terbaik sebesar 0,803 pada $k = 3$. Selain memiliki nilai DBI yang lebih baik, hasil K-Means juga konsisten dengan Elbow Method yang menunjukkan jumlah cluster optimal sebanyak 4 cluster. Cluster yang terbentuk menggambarkan karakteristik wilayah dengan tingkat pendidikan rendah, menengah, hingga tinggi, sehingga dapat digunakan sebagai dasar

dalam penyusunan kebijakan dan program peningkatan kualitas pendidikan di Kota Bekasi. Penelitian ini memiliki beberapa keterbatasan, yaitu data yang digunakan hanya berasal dari satu indikator, yaitu tingkat pendidikan penduduk pada periode 2021–2023, sehingga belum mempertimbangkan faktor lain yang dapat memengaruhi kondisi pendidikan, seperti tingkat ekonomi, jumlah sekolah, atau kepadatan penduduk. Selain itu, penelitian hanya membandingkan dua algoritma clustering, yaitu K-Means dan K-Medoids, sehingga hasil yang diperoleh belum dapat menggambarkan performa metode clustering lainnya. Oleh karena itu, penelitian selanjutnya dapat menggunakan variabel yang lebih beragam serta membandingkan lebih banyak algoritma clustering untuk memperoleh hasil yang lebih komprehensif.

6. Daftar Pustaka

- [1] Agustini, K., & Pradnyana, G. A. (2022). Konsep Dasar Data Mining (1st ed.). Penerbit Universitas Terbuka. <https://pustaka.ut.ac.id/lib/msim4403-data-mining/#tab-id-3>.
- [2] Amri Muliawan Nur, Yahya, Suhartini, & Rodhiyah Filkhaer (2026). Penerapan Algoritma K-Means Dalam Mengelompokkan Ruang Pasien BPJS Untuk Menentukan Pola Perawatan Kesehatan Yang Efektif. Infotek: Jurnal Informatika Dan Teknologi, 9(1), 174-185. <https://dx.doi.org/10.29408/jit.v9i1.33079>

- [3] Alexandropoulos, S.-A. N., Kotsiantis, S. B., & Vrahatis, M. N. (2019). Data preprocessing in predictive data mining. *The Knowledge Engineering Review*, 34, e1.
<https://doi.org/10.1017/S026988891800036X>
- [4] Dewi, E. Y. T. P., & Kamila, I. (2022). Pengelompokan Wilayah Berdasarkan Faktor Pendukung Pendidikan Dengan Jumlah Sekolah Dan Jumlah Guru Menggunakan Algoritma K-Means. *Interval: Jurnal Ilmiah Matematika*, 2(1), 1–12.
<https://journal.unpak.ac.id/index.php/intv/article/view/5161/3017>
- [5] Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., Jajuli, M., HSRonggo Waluyo, J., Telukjambe Timur, K., Karawang, K., & Barat, J. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*.
<https://doi.org/10.34010/jati.v12i1>
- [6] Fialine, A. P., Alodia, D. A., Endriani, D., & Widodo, E. (2021). Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan. *SEPREN: Journal of Mathematics Education and Applied*, 02(02), 1–13
- [7] Hasim Azari, Dwi Hartanti, & Aprilisa Arum Sari. (2024). Pengelompokan Produksi Padi dan Beras Provinsi Jawa Timur dengan Metode Agglomerative Hierarchical Clustering. *Infotek: Jurnal Informatika Dan Teknologi*, 7(2), 379–389.
<https://doi.org/10.29408/jit.v7i2.26016>
- [8] Kurniawan, W., Rifai, A., Gata, W., & Gunawan, D. (2020). Analisis Algoritma K-Medoids Clustering Dalam Menentukan Pemesanan Hotel. *JURNAL SWABUMI*, 8(2), 182–187.
<https://doi.org/https://doi.org/10.31294/swabumi.v8i2.9393>
- [9] Prihasto, B., Darmansyah, D., Yuda, D. P., Alwafi, F. M., Ekawati, H. N., & Sari, Y. P. (2023). Comparative Analysis of K-Means and K-Medoids Clustering Methods on Weather Data of Denpasar City. *Jurnal Pendidikan Multimedia (Edsence)*, 5(2), 91–114.
<https://doi.org/10.17509/edsence.v5i2.65925>
- [10] Ramadhana, B., & Meitasari, I. (2023). Kajian Tingkat Pendidikan Terhadap Kualitas Hidup Masyarakat. *Jurnal Penelitian Pendidikan Geografi.*, 8(2), 38–45.
<https://jppg.uho.ac.id/index.php/journal/article/download/1/13/150>
- [11] Ramdhan, D., Dwilestari, G., Dana, R. D., Ajiz, A., & Kaslani. (2022). Clustering Data Persediaan Barang dengan Menggunakan Metode K-Means. *MEANS (Media Informasi Analisa Dan Sistem)*, 7(1), 1–9.
http://ejournal.ust.ac.id/index.php/Jurnal_Means/
- [12] Septian, R., Syarippudin, & Nohe, D. A. (2023). Penerapan Algoritma K-Medoids pada Pengelompokan Wilayah Provinsi di Indonesia Berdasarkan Indikator Pendidikan. *Jurnal EKSPONENSIAL*, 14(2).
<http://jurnal.fmipa.unmul.ac.id/index.php/ekponensial>
- [13] Supriyadi, A., Triyudi, A., & Sholihati, I. D. (2021). Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 2, 229–240
- [14] Tempola, F., Muhammad, M., & Mubarak, A. (2020). Penggunaan Internet Dikalangan Siswa SD di Kota Ternate: Suatu Survey, Penerapan Algoritma Clustering Dan Validasi DBI. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 7(6), 1153–1160.
<https://doi.org/10.25126/jtiik.202072370>

- [15] Utomo, W. (2021). The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia. ILKOM Jurnal Ilmiah, 13(1), 31–35.
<https://doi.org/10.33096/ilkom.v13i1.763.3>
1-35