



PENERAPAN ALGORITMA *K-MEANS* UNTUK *CLUSTERING* DATA JUMLAH PENDUDUK MISKIN BERDASARKAN KOTA/KABUPATEN DI JAWA BARAT MENGGUNAKAN *RAPIDMINER*

Nova Novitasari¹, Nisa Dienwati Nuris², Ruli Herdiana³

¹ Komputerisasi Akuntansi, STMIK IKMI Cirebon

^{2,3} Teknik Informatika, STMIK IKMI Cirebon

Cirebon, Jawa Barat, Indonesia 45131

nova.novitasari555@gmail.com, nisadienwatinuris@gmail.com, ruliherdiana43@gmail.com

Abstract

Poverty is currently the only problem every region faces, as happened in the city/district of Responsibilities, which is still visible. Poverty is a condition where a person's standard of living cannot meet the basic needs of his life. According to information obtained by the West Java open data website, there are 513 data in West Java cities/regencies from 2002 to 2020. The number of poor people in many cities/regencies in West Java has increased. Thus, the method used in this study to group data using the K-means clustering method will be implemented in Rapidminer because the k-means algorithm is effective for analyzing large amounts of data. It can be seen from the data that will be generated that there are several variables needed, such as the Province code, Province name, City/Regency code, City/Regency name, number of poor people, unit, and year. From the variable data, it will be easier to run it. So that this study will produce several clusters, namely the results obtained by the researcher are cluster 4 because this cluster is the result taken from the smallest and best DBI value of several clusters that the researcher is testing. It is from the cluster results that data on the number of poor people will be seen by city/district which has an increasing level of collapse or by period each year.

Keywords: Clustering, Data Mining, K-Means, Poverty, Rapidminer

Abstrak

Kemiskinan saat ini menjadi satu-satunya masalah yang dihadapi setiap daerah. Seperti yang terjadi di Kota/Kabupaten Jawa Barat yang masih terlihat. Kemiskinan sendiri merupakan suatu kondisi dimana standar hidupnya seseorang benar-benar kurang mampu dalam memenuhi kebutuhan pokok hidupnya. Menurut informasi yang diperoleh *website* open data jabar, terdapat 513 data di Kota/Kabupaten Jawa Barat pada tahun 2002 s/d 2020, jumlah penduduk miskin di banyak Kota/Kabupaten di Jawa Barat semakin meningkat. Dengan demikian, metode yang digunakan dalam penelitian ini untuk melakukan pengelompokan data menggunakan metode K-Means *Clustering* yang akan diimplementasikan pada Rapidminer. Dikarenakan Algoritma K-Means merupakan Algoritma yang efektif untuk menganalisis data dalam jumlah besar. Jadi, bisa dilihat dari data yang akan dihasilkan ada beberapa variabel yang dibutuhkan, seperti kode Provinsi, nama Provinsi, kode Kota/Kabupaten, nama Kota/Kabupaten, jumlah penduduk miskin, satuan, dan tahun. Dari variabel data tersebut akan lebih mudah untuk menjalankannya. Sehingga penelitian ini akan menghasilkan beberapa *cluster*, yaitu hasil yang didapat oleh peneliti adalah *cluster* 4 karena *cluster* ini adalah hasil yang diambil dari nilai DBI yang terkecil dan terbaik dari beberapa *cluster* yang peneliti uji coba. Dari hasil *cluster* itulah akan terlihat data jumlah penduduk miskin berdasarkan kota/kabupaten yang memiliki tingkat kemiskinannya meningkat atau menurun dalam periode setiap tahunnya.

Kata kunci: Clustering, Data Mining, Kemiskinan, K-Means, Rapidminer

1. PENDAHULUAN

Menurut Undang-Undang No. 24 Tahun 2004, Kemiskinan adalah kondisi ekonomi sosial individu atau kelompok ketika mereka tidak memiliki sarana untuk menegakkan dan memajukan kehidupan martabat mereka [1]. Kemiskinan adalah masalah yang dihadapi banyak negara, bahkan yang

berkembang pesat, seperti Indonesia, yang terletak di Provinsi Jawa Barat [2].

Kemiskinan yang terjadi pada Provinsi Jawa Barat yang terdapat 27 jumlah Kota dan Kabupaten di Jawa Barat. Menurut Badan Pusat Statistik, ada 513 data mengenai

jumlah penduduk miskin di Jawa Barat dari tahun 2002 s/d 2020 berdasarkan Kabupaten/Kota. Data tersebut dikeluarkan dalam periode setiap 1 tahun sekali [3].

Secara umum, kemiskinan ini terus – menerus merupakan hasil dari tingginya tingkat keterbelakangan dan pengangguran. Penanggulangan kemiskinan dapat didukung dengan salah satu strategi seperti adanya data yang akurat. Data yang disebutkan diatas dapat dikelompokkan dengan menggunakan Algoritma K-Means untuk menghasilkan cluster. Cluster adalah pengelompokan data jumlah penduduk miskin berdasarkan Kota/Kabupaten yang selalu meningkat atau apakah sudah menurun [4]. Berikut ini terdapat gambar 1 menjelaskan tentang data jumlah penduduk miskin berdasarkan Kota/Kabupaten di Jawa Barat.

id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jumlah_penduduk_miskin	satuan	tahun
1	32	JAWA BARAT	3201	KABUPATEN BOGOR	451,3	RIBU JIWA	2002
2	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	352,2	RIBU JIWA	2002
3	32	JAWA BARAT	3203	KABUPATEN CIAMILIR	368,6	RIBU JIWA	2002
4	32	JAWA BARAT	3204	KABUPATEN BANDUNG	543,3	RIBU JIWA	2002
5	32	JAWA BARAT	3205	KABUPATEN GARUT	323,7	RIBU JIWA	2002
6	32	JAWA BARAT	3206	KABUPATEN TASIKMALAYA	341,1	RIBU JIWA	2002
7	32	JAWA BARAT	3207	KABUPATEN CIAMIS	265,8	RIBU JIWA	2002
8	32	JAWA BARAT	3208	KABUPATEN KUNINGAN	203,3	RIBU JIWA	2002
9	32	JAWA BARAT	3209	KABUPATEN CIREBON	388,4	RIBU JIWA	2002
10	32	JAWA BARAT	3210	KABUPATEN MAJALENGKA	214,3	RIBU JIWA	2002
11
505	32	JAWA BARAT	3271	KOTA BOGOR	75,04	RIBU JIWA	2020
506	32	JAWA BARAT	3272	KOTA SUKABUMI	25,42	RIBU JIWA	2020
507	32	JAWA BARAT	3273	KOTA BANDUNG	100,02	RIBU JIWA	2020
508	32	JAWA BARAT	3274	KOTA CIREBON	30,61	RIBU JIWA	2020
509	32	JAWA BARAT	3275	KOTA BEKASI	134,01	RIBU JIWA	2020
510	32	JAWA BARAT	3276	KOTA DEPOK	60,43	RIBU JIWA	2020
511	32	JAWA BARAT	3277	KOTA CIMAHI	31,64	RIBU JIWA	2020
512	32	JAWA BARAT	3278	KOTA TASHIMALAYA	86,13	RIBU JIWA	2020
513	32	JAWA BARAT	3279	KOTA BANJAR	11,16	RIBU JIWA	2020

Gambar 1. Data Jumlah Penduduk Miskin Di Jawa Barat

Penyebab masalah kemiskinan yang terjadi di Kota/Kabupaten Jawa Barat biasanya terjadi karena bertambahnya jumlah penduduk akan terjadinya tingkat kemiskinan, kurangnya lapangan pekerjaan yang berdampak pada tingginya tingkat pengangguran, tingkat pendidikan yang rendah, dan kemiskinan juga berarti kurangnya kebutuhan sosial, termasuk ketidakmampuan untuk berpartisipasi dalam kebutuhan masyarakat yang layak. Sebagaimana kemiskinan dapat dipahami sebagai keadaan kekurangan uang atau barang untuk menjamin kelangsungan hidup [5].

Melihat dari permasalahan diatas, pengelompokan data menggunakan metode K-Means akan digunakan dalam pengumpulan data peneliti. Pengelompokan data tersebut pada Algoritma K-Means akan dibagi menjadi 4 cluster sesuai dengan hasil nilai DBI. Pada pengelompokan data tersebut sebagai solusi untuk mempermudah pencarian/perolehan data jumlah penduduk miskin berdasarkan Kota/Kabupaten dalam golongan peningkatan hingga penurunan tingkat kemiskinan [6].

Adapun yang menjadi alasan dilakukannya peneliti ini dengan judul “penerapan algoritma k-means untuk clustering data jumlah penduduk miskin berdasarkan Kota/Kabupaten di Jawa Barat” adalah untuk dapat menerapkan Algoritma K-Means pada pengelompokan data jumlah penduduk miskin tersebut. Dari penerapan tersebut peneliti akan mengelompokkan berdasarkan tingkat tinggi rendahnya jumlah penduduk miskin berdasarkan Kota/Kabupaten dalam 4 cluster sesuai dengan hasil yang

diperoleh dari nilai DBI yaitu nilai yang terkecil dan terbaik [7].

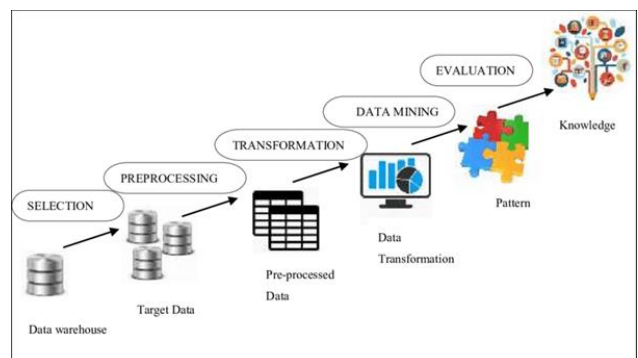
2. METODE PENELITIAN

2.1 Tahapan Pengumpulan Data

Tahapan pengumpulan data yang digunakan dalam penelitian ini menggunakan data sekunder berdasarkan sumbernya. Terdapat 513 data dari tahun 2002 s/d 2020 yang berasal dari Badan Pusat Statistik Provinsi Jawa Barat, yang merupakan sumber data dari situs resmi open data jabar [8].

2.2 Tahapan Perancangan

Pada tahap ini, peneliti mengambil tema *Artificial Intelligence* bisa *Merefrence ke Knowledge Discovery In Databases* [9]. Berikut ini adalah langkah-langkah proses KDD dalam penambahan data dan dibawah ini terdapat gambar 2 tentang tahap perancangan KDD yaitu sebagai berikut:



Gambar 2. Tahap Perancangan KDD

1. *Selection*: Pilih sub kumpulan data yang relevan untuk dianalisis.
2. *Preprocessing*: Bersihkan dan ubah data agar siap untuk dianalisis.
3. *Transformation*: Mengubah data menjadi formal yang cocok untuk penambahan data.
4. *Data mining*: Terapkan teknik dan Algoritma penambahan data ke data untuk mengekstrak informasi dan wawasan yang bermanfaat.
5. *Evaluation*: Evaluasi hasil untuk memastikan bahwa pengetahuan yang diekstraksi bermanfaat, akurat, dan bermakna.

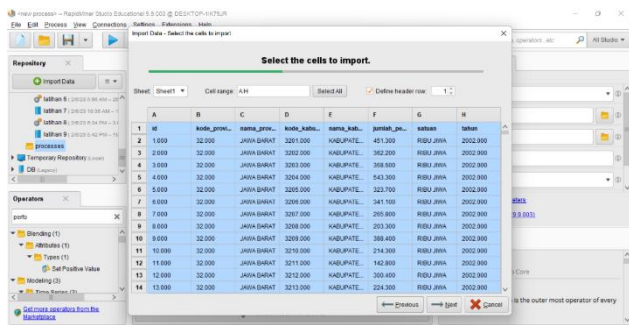
3. HASIL DAN PEMBAHASAN

3.1 Hasil Tujuan Mengelompokkan Data Jumlah Penduduk Miskin berdasarkan Kota/Kabupaten di Jawa Barat menggunakan Algoritma K-Means

3.1.1 Selection

Langkah pertama adalah mengumpulkan data. Mengenai data yang akan digunakan untuk di analisis, penting untuk dicatat bahwa data tersebut berasal dari sumber terbuka. Data yang akan digunakan berasal dari website resmi Badan Pusat Statistik (BPS) dan merupakan data terbuka jumlah

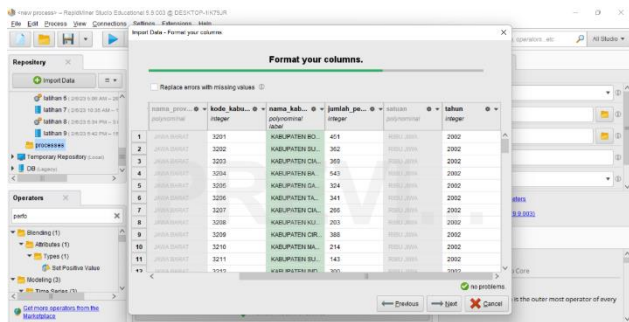
penduduk miskin menurut Kota/Kabupaten di Jawa Barat dari tahun 2002 s/d 2020 sebanyak 513 data. Di bawah ini terdapat gambar 3 yang menerapkan tentang data yang akan di Clustering dalam Rapidminer yaitu tahap pertama tentang Operator *Selection* [10].



Gambar 3. Selection

3.1.2 Preprocessing

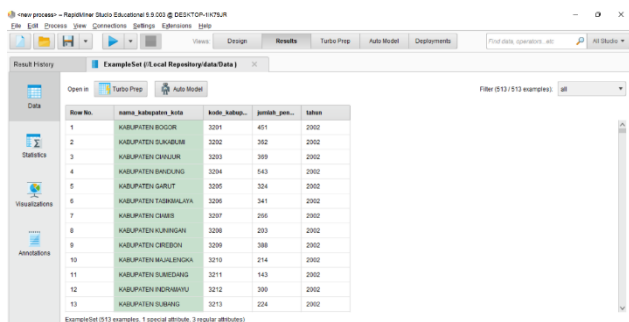
Tahap selanjutnya yaitu data preprocessing. Dalam tahapan ini mencangkup beberapa hal yaitu menghilangkan duplikat data, memperbaiki kesalahan data dan sebagainya. Dimana tahap ini akan menghasilkan atribut yang sudah cukup bersih dan baik untuk dapat melakukan ketahap selanjutnya transformation. Berikutnya terdapat gambar 4 menerapkan tahap ke-2 pada proses menjalankan Rapidminer yaitu Operator *Preprocessing* [11].



Gambar 4. Preprocessing

3.1.3 Transformation

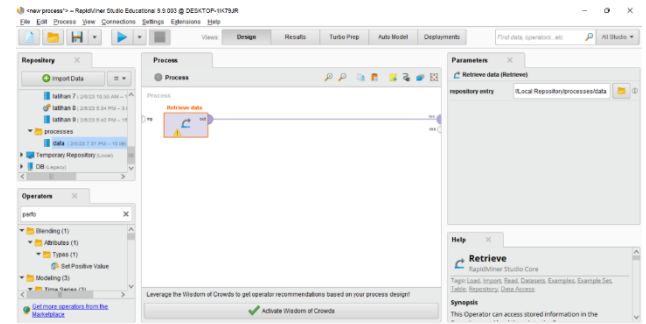
Pada tahap transformation merupakan tahap dimana data tersebut sudah berubah menjadi data yang peneliti inginkan untuk dapat menjalankan ketahap selanjutnya. Berikutnya pada gambar 5 menjelaskan tentang tahap ke-3 pada proses Rapidminer yaitu Operator *Transformation*.



Gambar 5. Transformation

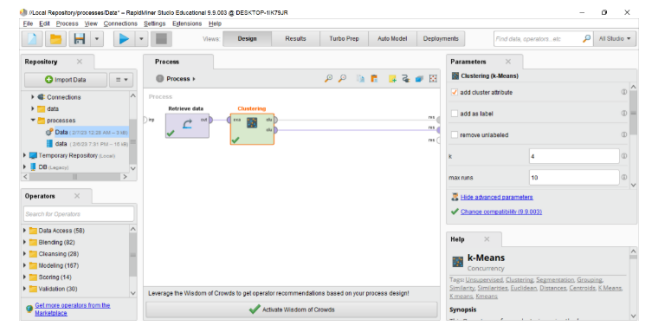
3.1.4 Data Mining

Teknik data mining disajikan pada tahap ini, dimana peneliti menggunakan Algoritma Clustering *data mining* untuk melakukan proses data tersebut. Langkah pertama adalah meminta operator mengambil data dari lokasi yang telah diimport dan siap untuk melanjutkan proses. Berikutnya pada gambar ini terdapat beberapa bagian diantaranya gambar 6(a) menerapkan tahap ke-4 pada proses Rapidminer yaitu Operator *Retrieve Data* [12].



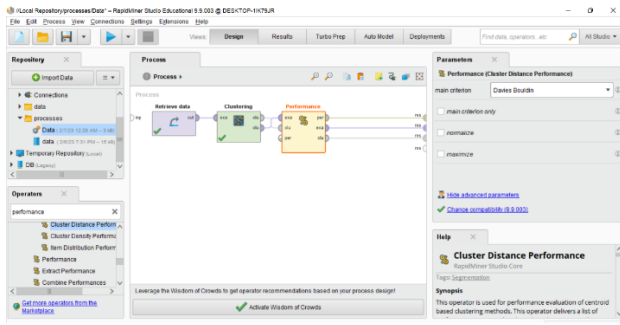
Gambar 6. (a) Operator Retrieve Data

Langkah selanjutnya adalah membuat model clustering dengan menggunakan metode k-means yang akan menghasilkan jumlah cluster yang diinginkan. Selanjutnya pada gambar 6(b) menerapkannya proses Rapidminer tentang Operator K-Means.



Gambar 6. (b) Operator K-means

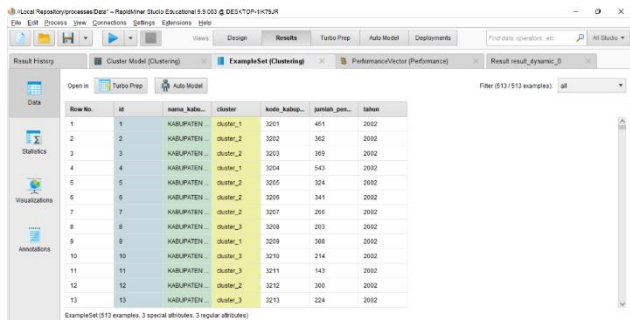
Dan operator terakhir dalam cluster *distance performance* adalah operator untuk menurunkan *Index Davies Bouldin* (DBI). Carilah angka DBI terkecil untuk menentukan jumlah cluster yang ukurannya paling ideal. Dan selanjutnya pada gambar 6(c) menerapkan proses pada Rapidminer tentang Operator *Performance*.



Gambar 6. (c) Operator Performance

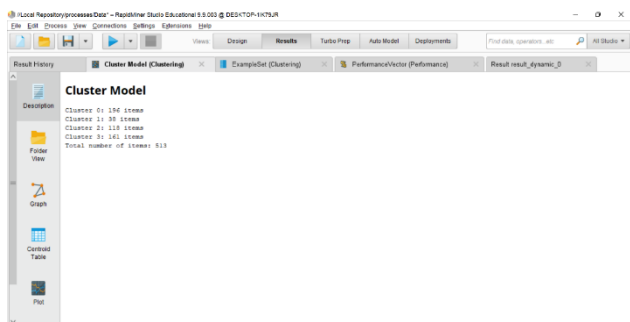
3.1.5 Evaluation

Setelah melakukan teknik data mining dan mendapatkan hasil clustering. Selanjutnya adalah tahap evaluation, tahap ini dilakukan untuk melihat keberhasilan model yang telah diterapkan [13]. Di bawah ini terdapat beberapa gambar yang menjelaskan tentang penerapannya Clustering pada proses Rapidminer yaitu gambar 7(a) tentang Operator Performance.



Gambar 7. (a) Operator Performance

Dan selanjutnya terdapat gambar 7(b) menjelaskan tentang Operator Cluster model yaitu hasil dari Cluster 4.



Gambar 7. (b) Cluster model

Pada tahap ini dimana hasil analisis data terkait setiap cluster tersedia. Cluster 0 berisi 196 anggota, cluster 1 memiliki 38 anggota, cluster 2 memiliki 118 anggota, dan cluster 3 memiliki 161 anggota dari total 513 dataset yang dinilai [14].

1. Cluster 0

Pada tabel 1 ini menjelaskan tentang beberapa data jumlah penduduk miskin berdasarkan Kota/Kabupaten yang termasuk dalam Cluster 0.

Tabel 1. Cluster 0

No	Kode Kabupaten / Kota	Nama Kabupaten / Kota	Jumlah penduduk miskin	Tahun
1	3217	Kabupaten Bandung Barat	0	2002
2	3218	Kabupaten Pangandaran	0	2002
3	3271	Kota Bogor	65,4	2002
4	3272	Kota Sukabumi	21,7	2002
5	3273	Kota Bandung	75,3	2002
6	3274	Kota Cirebon	24,7	2002
7	3275	Kota Bekasi	66,2	2002
8...
196	3297	Kota Banjar	11,16	2020

2. Cluster 1

Pada tabel 2 ini menjelaskan tentang beberapa data jumlah penduduk miskin berdasarkan Kota/Kabupaten yang termasuk dalam Cluster 1.

Tabel 2. Cluster 1

No	Kode Kabupaten / Kota	Nama Kabupaten / Kota	Jumlah penduduk miskin	Tahun
1	3201	Kabupaten Bogor	451,3	2002
2	3204	Kabupaten Bandung	543,3	2002
3	3209	Kabupaten Cirebon	388,4	2002
4	3201	Kabupaten Bogor	476,4	2002
5	3203	Kabupaten Cianjur	388,8	2002
6	3204	Kabupaten Bandung	515,5	2002
7	3201	Kabupaten Bogor	453,4	2002
8...
38	3201	Kabupaten Bogor	465,67	2020

3. Cluster 2

Pada tabel 3 ini menjelaskan tentang beberapa data jumlah penduduk miskin berdasarkan Kota/Kabupaten yang termasuk dalam Cluster 2.

Tabel 3. Cluster 2

No	Kode Kabupaten / Kota	Nama Kabupaten / Kota	Jumlah penduduk miskin	Tahun
1	3202	Kabupaten Sukabumi	362,2	2002
2	3203	Kabupaten Cianjur	368,6	2002
3	3205	Kabupaten Garut	323,7	2002

No	Kode Kabupaten / Kota	Nama Kabupaten / Kota	Jumlah penduduk miskin	Tahun
4	3206	Kabupaten Tasikmalaya	341,1	2002
5	3207	Kabupaten Ciamis	265,8	2002
6	3212	Kabupaten Indramayu	300,4	2002
7	3215	Kabupaten Karawang	267,4	2002
8...
118	3209	Kabupaten Cirebon	247,94	2020

4. Cluster 3

Pada tabel 4 ini menjelaskan tentang beberapa data jumlah penduduk miskin berdasarkan Kota/Kabupaten yang termasuk dalam Cluster 3.

Tabel 4. Cluster 3

No	Kode Kabupaten / Kota	Nama Kabupaten / Kota	Jumlah penduduk miskin	Tahun
1	3208	Kabupaten Kuningan	203,3	2002
2	3210	Kabupaten Majalengka	214,3	2002
3	3211	Kabupaten Sumedang	142,8	2002
4	3213	Kabupaten Subang	224,3	2002
5	3214	Kabupaten Purwakarta	101,4	2002
6	3216	Kabupaten Bekasi	118,2	2002
7	3208	Kabupaten Kuningan	201,7	2002
8...
161	3275	Kota Bekasi	134,01	2020

3.2 Hasil Tujuan Menentukan Cluster Optimum dan Nilai Dbi dari Hasil Mengelompokkan Data Jumlah Penduduk Miskin berdasarkan Kota/Kabupaten di Jawa Barat menggunakan Algoritma K-Means

Davies Bouldin Index (DBI) adalah satu-satunya metode yang digunakan untuk mengurangi validitas cluster dalam metode pengumpulan data apapun [15]. Berikut ini terdapat tabel 5 yang menjelaskan beberapa nilai DBI pada setiap clusternya.

Tabel 5. Nilai DBI tiap cluster

Cluster	Nilai DBI
2	-0,618
3	-0,571
4	-0,563
5	-0,594

Cluster	Nilai DBI
6	-0,655
7	-0,659
8	-0,608
9	-0,573
10	-0,572

Jumlah cluster yang dimodelkan pada model clustering di atas dengan menggunakan metode K-means adalah 2-10 cluster. Terdapat batasan nilai DBI dari setiap cluster untuk mengidentifikasi hasil terbaik dan mengidentifikasi jumlah cluster yaitu cluster 4 dengan nilai DBI sebesar -0,563. Dan dibawah ini terdapat gambar 8 menjelaskan tentang hasil dari Cluster 4 dengan nilai DBI -0,563



Gambar 8. Nilai DBI Cluster 4

4. KESIMPULAN

Berdasarkan hasil penelitian data mining yang dilakukan dengan menggunakan informasi jumlah penduduk miskin menurut Kota/Kabupaten di Jawa Barat, dapat ditarik kesimpulan sebagai berikut:

1. Dengan menggunakan teknik data mining Algoritma K-Means, peneliti menghitung jumlah penduduk miskin di Kota/Kabupaten Jawa Barat berdasarkan atribut yang tersedia: Kode Kota/Kabupaten, Nama Kota/Kabupaten, Jumlah Penduduk Miskin, dan Tahun.
2. Dari total 513 *dataset* yang dianalisis, hasil cluster 0 menunjukkan 196 anggota, hasil cluster 1 menunjukkan 38 anggota, hasil cluster 2 menunjukkan 118 anggota, dan hasil cluster 3 menunjukkan 161 anggota.
3. Hasil menunjukkan nilai *Davies Bouding Index* (DBI) yang rendah untuk mengidentifikasi hasil cluster 4 dengan menggunakan nilai DBI -0,563 sebagai hasil yang cukup baik.

DAFTAR PUSTAKA

- [1] F. Febriansyah and S. Muntari, "Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam," vol. 8, no. 1, pp. 66–77, 2023.
- [2] E. Fammaldo and L. Hakim, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar," *J. Ilm. Teknol. Infomasi*

- Terap.*, vol. 5, no. 1, pp. 23–31, 2019, doi: 10.33197/jitter.vol5.iss1.2018.249.
- [3] N. N. F. R, D. S. Anggraeni, and U. Enri, “Pengelompokan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient”, *TEMATIK*, vol. 9, no. 1, pp. 29-35, Jun. 2022..
- [4] T. R. Melati, I. Nawangsih, and I. Afriantoro, “Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Berdasarkan Kabupaten/Kota Di Indonesia.”
- [5] E. Muningsih, and S. Kiswati, “Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang,” *J. Bianglala Informatika*, vol. 3, no. 1, 2015.
- [6] M. S. Zain, S. Defit, and Sumijan, “Data Mining dengan Metode K-Means untuk Pengelompokan Mahasiswa yang Mengunjungi Perpustakaan berdasarkan Data Kunjungan dan IPK,” 2018.
- [7] S. J. Latupeirissa, N. Lewaherilla, and A. Hiariey, “Pengelompokan Kabupaten/Kota di Provinsi Maluku Berdasarkan Data Kemiskinan Tahun 2021 menggunakan Metode K-Means Cluster (*Grouping District/City In Maluku Province Based On Poverty Data For 2021 Using The K-Means Cluster Method*),” *Variance*, vol. 4, no. 1, 2022, Doi: 10.30598/Variancevol4iss1page15-22.
- [8] F. Ilena Putri And R. Damayanti, “Prosiding Seminar Nasional Matematika, Statistika, Dan Aplikasinya Terbitan Ii,” 2022.
- [9] Y. F. S. Y. Damanik, S. Sumarno, I. Gunawan, D. Hartama, And I. O. Kirana, “Penerapan Data Mining Untuk Pengelompokan Penyebaran Covid-19 Di Sumatera Utara Menggunakan Algoritma K-Means,” *J. Ilmu Komput. Dan Inform.*, Vol. 1, No. 2, Nov. 2021, Doi: 10.54082/Jiki.13.
- [10] P. Algoritma K-Means dalam Pengelompokan Jumlah Penduduk Berdasarkan Kelurahan Di Kota Pematangsiantar, “Widodo Saputra 5) 1)2)3)4)5) Program Studi Teknik Informatika,” Vol. 2, No. 2, Pp. 20–26, 2021, [Online]. Available: [Http://Creativecommons.Org/Licences/By/4.0/](http://Creativecommons.Org/Licences/By/4.0/)
- [11] B. Nurseptia, A. Voutama, N. Haeryana, And J. Hsronggo Waluyo, “Penerapan Algoritma K-Means Untuk Mengelompokkan Kabupaten/Kota Dalam Upaya Pemetaan Lapangan Pekerjaan Baru,” *J. Teknol. Informasi*, Vol. 6, No. 2, 2022.
- [12] Y. Ratna Sari, A. Sudewa, D. Ayu Lestari, And T. Ika Jaya, “Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer,” 2020.
- [13] M. M. K-Means, “Penerapan Data Mining Untuk Clustering Data Penduduk Miskin,” Vol. 3, No. 1, Pp. 31–37, 2021.
- [14] T. D. Desa, “Sistem Pendukung Keputusan Penerima Bantuan Langsung Tunai Dana Desa Menggunakan Metode Algoritma K-Means Clustering,” Vol. 14, No. 1, Pp. 150–160, 2022.
- [15] I. Nasution, A. P. Windarto, And M. Fauzan, “Penerapan Algoritma K-Means dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi,” Vol. 2, No. 2, Pp. 76–83, 2020.