

Pengelompokan Terbaik Menggunakan Algoritma K-Means Pada Dataset Bus Biskita Bogor

Anjar Permadi¹, Yudhistira Arie Wiyaja²

^{1,2}Prodi Manajemen Informatika, STMIK IKMI Cirebon, Indonesia

anjarpermadi300@gmail.com

Info Artikel

Sejarah artikel:

Diterima Mei 2023

Direvisi Juli 2023

Disetujui Juli 2023

Diterbitkan Juli 2023

ABSTRACT

In the city of Bogor, as one of the growing metropolitan cities, the bus transportation system is a major concern in order to increase population mobility and reduce traffic congestion. With the aim of the data mining process that will produce clusters according to the knowledge or patterns stored in the data. And also to find out the grouping of Biskita bus datasets from each corridor, which later is expected to help companies/providers of Biskita bus mass transit services. Data mining is the process of finding information by identifying patterns from datasets. The process of finding this information can be done by grouping data into several groups from a dataset which in data mining is called the clustering method. Clustering is the process of partitioning a dataset into several subsets or groups based on the similarity of the characteristics of each data in the existing groups. The clustering method used in this research is K-Means which belongs to the Partition Clustering algorithm group. In this study, the k-means algorithm was used to classify the Bisikita Bogor bus dataset based on data collected during 2022. In the process of converting the raw dataset into useful information, the Knowledge Discovery in Database (KDD) process was used. In the early stages, data cleaning was carried out, then data selection, data transformation, and data mining were carried out using the Rapidminer software. Modeling results were evaluated using the Davies Bouldin Index (DBI) instrument. Based on the results of grouping data that has been done using the K-Means method using default parameters, the smallest result is obtained, namely in cluster k-3 with a DBI value of 0.530. So cluster k-3 is the best cluster out of a total of 20 k. With a total of 1972 items of cluster 0, 1612 items of cluster 1, and 140 items of cluster 2.

Keywords : Data mining, Clustering, K-Means, DBI

ABSTRAK

Di kota Bogor, sebagai salah satu kota metropolitan yang terus berkembang, sistem transportasi bus menjadi perhatian utama dalam rangka meningkatkan mobilitas penduduk dan mengurangi kemacetan lalu lintas. Dengan tujuan proses penambangan data yang akan menghasilkan cluster sesuai dengan pengetahuan atau pola yang tersimpan pada data. Dan juga untuk mengetahui pengelompokan dataset bus biskita dari setiap koridor, yang nanti nya diharapkan dapat membantu perusahaan/penyedia jasa angkutan massal bus biskita. Data mining merupakan proses penemuan informasi dengan mengidentifikasi pola dari dataset. Proses penemuan informasi tersebut dapat dilakukan dengan metode pengelompokan data ke dalam beberapa kelompok dari sebuah dataset yang dalam data mining disebut metode clustering. Clustering merupakan proses mempartisi dataset menjadi beberapa subset atau kelompok berdasarkan kesamaan karakteristik masing-masing data pada kelompok yang ada. Metode Clustering yang digunakan dalam penelitian ini adalah K-Means yang termasuk ke dalam golongan algoritma Partition Clustering. Pada penelitian digunakan algoritma k-means untuk mengelompokkan dataset bus Bisikita Bogor berdasarkan data yang diambil selama tahun 2022. Dalam melakukan proses mengubah dataset mentah menjadi suatu informasi yang bermanfaat, maka digunakan proses Knowledge Discovery in Database (KDD). Pada tahap awal dilakukan pembersihan data, selanjutnya dilakukan seleksi data, transformasi data, dan data mining dengan menggunakan perangkat lunak Rapidminer. Hasil pemodelan dievaluasi menggunakan instrumen Davies Bouldin Index (DBI). Berdasarkan hasil dari pengelompokkan data yang sudah dilakukan dengan metode K-Means menggunakan parameter default diperoleh hasil terkecil yaitu pada cluster k-3 dengan nilai DBI sebesar 0,530. Maka cluster k-3 merupakan cluster terbaik dari total k sebanyak 20. Dengan jumlah cluster 0 sebanyak 1972 items, cluster 1 sebanyak 1612 items, dan cluster 2 sebanyak 140 items.

Kata Kunci : Data Mining, Clustering, DBI, K-Means

PENDAHULUAN

Transportasi umum, terutama bus, merupakan sarana vital dalam sistem transportasi kota modern. Kota-kota besar di dunia sering menghadapi tantangan dalam mengelola sistem transportasi bus mereka agar dapat berjalan efisien, nyaman, dan sesuai dengan kebutuhan masyarakat. Di kota Bogor, sebagai salah satu kota metropolitan yang terus berkembang, sistem transportasi bus menjadi perhatian utama dalam rangka meningkatkan mobilitas penduduk dan mengurangi kemacetan lalu lintas.

Perencanaan dan pengelolaan sistem transportasi bus yang efisien memerlukan pemahaman yang mendalam tentang karakteristik dan perilaku bus di kota tersebut. Untuk tujuan ini, kumpulan data tentang bus yang beroperasi di Bogor diperlukan. Data tersebut dapat mencakup informasi seperti jadwal, rute, jumlah penumpang, waktu perjalanan, dan variabel lain yang relevan.

Pada tahap awal pengelolaan sistem transportasi bus, analisis data dan pemahaman mengenai pola pergerakan bus dan aktivitasnya menjadi hal penting. Salah satu pendekatan yang umum digunakan untuk mengelompokkan data seperti ini adalah dengan menggunakan algoritma K-Means. Algoritma K-Means adalah metode klustering yang populer digunakan untuk mengelompokkan data berdasarkan kesamaan atribut. Pendekatan ini dapat membantu mengidentifikasi pola dan karakteristik tertentu yang mungkin tersembunyi dalam dataset bus, yang kemudian dapat digunakan untuk pengambilan keputusan yang lebih baik dalam perencanaan dan pengelolaan sistem transportasi.

Namun, untuk mendapatkan hasil pengelompokan terbaik, beberapa aspek perlu dipertimbangkan. Pertama, pemilihan jumlah kluster (cluster) yang optimal harus ditentukan agar kluster yang dihasilkan dapat merepresentasikan karakteristik yang berarti dari data. Selain itu, penting untuk memilih atribut atau fitur yang tepat dari dataset bus untuk mencapai hasil klustering yang akurat dan informatif.

Teknik memperoleh atau menambang pengetahuan dari sejumlah besar data dikenal sebagai *data mining* [1]. *Data mining* adalah proses yang memakai metode statistik, matematika, kecerdasan buatan, serta *machine learning* untuk mengidentifikasi informasi yang berguna dan pengetahuan yang terikat dari bermacam database [2]. Salah satu algoritma dalam *data mining* yaitu K-Means. K-Means didefinisikan sebagai algoritma yang bertugas membagi data menjadi kelompok bersumber pada tingkat kemiripannya [3]. *Clustering* merupakan proses pengelompokan titik data menjadi dua atau lebih kelompok sehingga titik data yang tercantum dalam kelompok yang sama lebih mirip satu sama lain daripada di kelompok yang berbeda, hanya berdasarkan informasi yang tersedia dengan titik data [4].

Penelitian ini mengadopsi pendekatan *Knowledge Discovery In Database (KDD)*. Tahapan KDD meliputi kegiatan-kegiatan yang dimulai dari *Data Selection* hingga *Interpretation/Evaluation* [5]. Sedangkan untuk pengelompokan datanya menggunakan algoritma K-Means. K-Means merupakan metode pengelompokan data non-hierarkis yang berupaya mempartisi data yang ada menjadi dua atau lebih kelompok. [6]. Metode ini akan membagi data ke dalam kelompok dimana data dengan karakteristik yang sama akan dimasukkan ke dalam kelompok yang sama,

dan data dengan karakteristik yang berbeda akan dikelompokkan ke dalam kelompok lain [7]. Metode ini dipilih karena lebih sederhana, mudah diimplementasikan, tidak lambat, mudah disesuaikan, dan paling sering digunakan dalam proses data mining [3]. Tujuan dari penelitian ini adalah proses penambangan data yang akan menghasilkan *cluster* sesuai dengan pengetahuan atau pola yang tersimpan pada data. Dan juga untuk mengetahui pengelompokan dataset bus biskita dari setiap koridor, yang nantinya diharapkan dapat membantu perusahaan/penyedia jasa angkutan massal bus biskita.

Dengan melakukan penelitian ini, diharapkan dapat diperoleh pemahaman yang lebih baik tentang pola pergerakan dan perilaku bus Biskita di kota Bogor. Hasil klastering yang optimal dapat membantu pihak terkait dalam perencanaan dan pengelolaan sistem transportasi bus, serta memberikan manfaat bagi masyarakat dalam meningkatkan kualitas transportasi publik di kota tersebut.

METODE

Data mining merupakan teknik untuk menemukan informasi baru yang berguna dari sekumpulan data yang besar dan dapat membantu pada saat mengambil keputusan [6]. *Data mining* merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari *machine learning*, pengenalan pola, *statistic*, *database*, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar [2].

Algoritma *k-means* adalah algoritma pengelompokan tipikal berdasarkan jarak. Pada algoritma *k-means*, semakin kecil jarak antar sampel, semakin tinggi kesamaannya [8]. Algoritma ini efisien dan banyak digunakan dalam kumpulan data yang tidak berlabel untuk tugas pengelompokan [9].

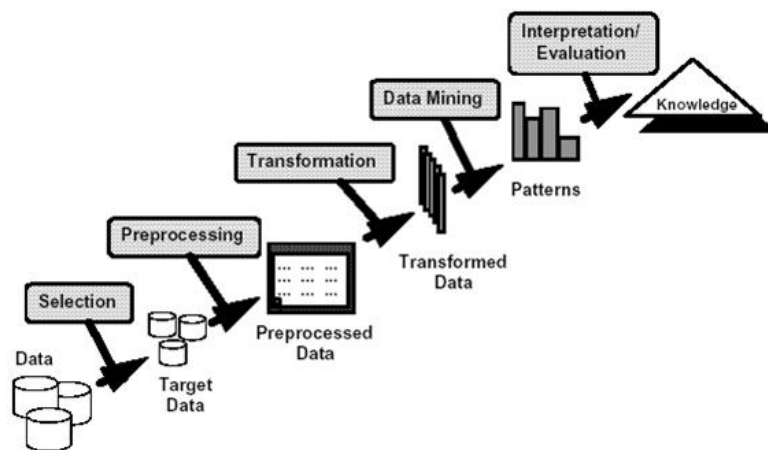
Clustering adalah proses pembentukan kelompok data (*cluster*) dari kumpulan data yang tidak diketahui kelompok atau kelasnya, dan merupakan proses penentuan data mana yang termasuk dalam cluster mana. Metode *clustering* membagi data menjadi beberapa kelompok sehingga data dengan karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama. Tujuan dari clustering ini adalah untuk meminimalisasi fungsi tujuan yang ditetapkan dalam proses *clustering*, yang biasanya bertujuan untuk meminimalisasi variasi dalam suatu *cluster* dan memaksimalkan variasi antar *cluster* [10]. Cara kerja teknik ini mengelompokkan sekumpulan data ke dalam kelas-kelas atau *cluster-cluster*, yang mana objek-objek yang ada pada kelas tersebut memiliki similaritas yang tinggi jika dibandingkan dengan objek lain yang ada dalam kelas tersebut, namun memiliki similaritas yang rendah jika dibandingkan dengan objek yang ada di *cluster* lain [11].

Davies-Bouldin Index (DBI) merupakan salah satu metode yang digunakan untuk mengukur validitas atau jumlah *cluster* paling optimal pada suatu metode pengelompokan dimana kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat *cluster* dari *cluster* yang diikuti [12]. Metode *DBI* dipilih karena dapat diimplementasikan dalam semua data baik ukuran besar dan kecil sehingga sangat cocok dan dapat diterima untuk perhitungan akurasi clustering [13].

Sumber data dalam penelitian ini bersumber dari Website Executive Dashboard Kementerian Perhubungan. Data memiliki 6 field atau

atribut yaitu : Nomor Kendaraan, Kode Kendaraan, Jenis Bus, Total Penumpang, Total Ritase, dan KM Tempuh Koridor. Data yang digunakan yaitu data selama tahun 2022, dengan total jumlah data sebanyak 3724 records.

Metode yang digunakan yaitu *Knowledge Discovery in Database (KDD)*. KDD yaitu kegiatan yang meliputi pengumpulan data, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [14]. Proses *Knowledge Discovery in Database* diawali dengan menetapkan tujuan dan diakhiri dengan evaluasi [15].



Gambar 1. Tahapan proses KDD

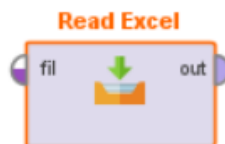
Sumber: Isal, 2019

HASIL DAN PEMBAHASAN

Penerapan KDD dalam penelitian ini yaitu sebagai berikut :

Data Selection :

1. *Read Excel*, operator ini digunakan untuk mengimport atau memasukan data excel yang berada di komputer pengguna ke dalam proses pada rapidminer. Karena dataset yang digunakan adalah dataset bus Biskita Bogor yang terdiri dari 4 file excel berbeda. Maka dari itu dibutuhkan juga 4 operator *read excel* di dalam proses Rapidminer. Untuk parameter yang digunakan adalah parameter *default*.



Gambar 2. Read Excel

Parameter pada operator *Read Excel* menggunakan parameter default. Dari hasil pembacaan operator *Read Excel* didapat informasi sebagai berikut.

Tabel 1. Statik dataset file pertama

No	Uraian	Isi
1	<i>Record</i>	538
2	<i>Special Attribute</i>	0
3	<i>Regular Attribute</i>	6
4	<i>Attributes :</i>	
	Nomor Kendaraan	<i>Polynomial, missing</i> 0
	Kode Kendaraan	<i>Polynomial, missing</i> 0
	Jenis Bus	<i>Polynomial, missing</i> 0
	Total Penumpang	<i>Integer, missing</i> 0
	Total Ritase	<i>Integer, missing</i> 0
	KM Tempuh	<i>Real, missing</i> 0
	Koridor	

Tabel 2. Statik dataset file kedua

No	Uraian	Isi
1	<i>Record</i>	1102
2	<i>Special Attribute</i>	0
3	<i>Regular Attribute</i>	6
4	<i>Attributes :</i>	
	Nomor Kendaraan	<i>Polynomial, missing</i> 0
	Kode Kendaraan	<i>Polynomial, missing</i> 0
	Jenis Bus	<i>Polynomial, missing</i> 0
	Total Penumpang	<i>Integer, missing</i> 0
	Total Ritase	<i>Integer, missing</i> 0
	KM Tempuh	<i>Real, missing</i> 0
	Koridor	

Tabel 3. Statik dataset file ketiga

No	Uraian	Isi
1	<i>Record</i>	1362
2	<i>Special Attribute</i>	0
3	<i>Regular Attribute</i>	6
4	<i>Attributes :</i>	
	Nomor Kendaraan	<i>Polynomial, missing</i> 0
	Kode Kendaraan	<i>Polynomial, missing</i> 0
	Jenis Bus	<i>Polynomial, missing</i> 0
	Total Penumpang	<i>Integer, missing</i> 0

Total Ritase	<i>Integer, missing 0</i>
KM Tempuh	<i>Real, missing 0</i>
Koridor	

Tabel 4. Statik dataset file keempat

No	Uraian	Isi
1	<i>Record</i>	722
2	<i>Special Attribute</i>	0
3	<i>Regular Attribute</i>	6
4	<i>Attributes :</i>	
	Nomor Kendaraan	<i>Polynomial, missing 0</i>
	Kode Kendaraan	<i>Polynomial, missing 0</i>
	Jenis Bus	<i>Polynomial, missing 0</i>
	Total Penumpang	<i>Integer, missing 0</i>
	Total Ritase	<i>Integer, missing 0</i>
	KM Tempuh	<i>Real, missing 0</i>
	Koridor	

2. *Append*, operator ini berfungsi menggabungkan beberapa dataset menjadi satu. Dataset yang digunakan terdiri dari 4 file excel yang berisi data koridor bus yang berbeda kemudian dijadikan satu, dengan catatan atribut yang digunakan harus sama.

Gambar 3. *Append*

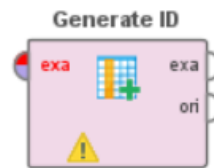
Untuk parameter yang digunakan pada operator ini adalah parameter *default*. Dari hasil pembacaan operator *Append* didapat informasi sebagai berikut.

Tabel 5. Statik dataset setelah menggunakan *append*

No	Uraian	Isi
1	<i>Record</i>	3724
2	<i>Special Attribute</i>	0
3	<i>Regular Attribute</i>	6
4	<i>Attributes :</i>	
	Nomor Kendaraan	<i>Polynomial, missing 0</i>
	Kode Kendaraan	<i>Polynomial, missing 0</i>
	Jenis Bus	<i>Polynomial, missing 0</i>

Total Penumpang	<i>Integer, missing 0</i>
Total Ritase	<i>Integer, missing 0</i>
KM Tempuh	<i>Real, missing 0</i>
Koridor	

3. *Generate ID*, operator ini menambahkan atribut baru dengan peran id di input dataset. Karena pada dataset bus biskita tidak ada atribut yang dapat dijadikan *id* maka ditambahkan operator *Generate ID*. Untuk parameter yang digunakan adalah parameter default.



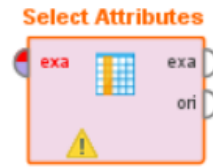
Gambar 4. *Generate ID*

Untuk parameter yang digunakan pada operator ini adalah parameter *default*. Dari hasil pembacaan operator *Generate ID* didapat informasi sebagai berikut.

Tabel 6. Statik dataset

No	Uraian	Isi
1	<i>Record</i>	3724
2	<i>Special Attribute</i>	1
3	<i>Regular Attribute</i>	6
4	<i>Attributes :</i>	
	<i>ID</i>	<i>Integer, missing 0</i>
	Nomor Kendaraan	<i>Polynomial, missing 0</i>
	Kode Kendaraan	<i>Polynomial, missing 0</i>
	Jenis Bus	<i>Polynomial, missing 0</i>
	Total Penumpang	<i>Integer, missing 0</i>
	Total Ritase	<i>Integer, missing 0</i>
	KM Tempuh	<i>Real, missing 0</i>
	Koridor	

4. *Select Attributes*, operator ini dipergunakan untuk memisahkan data apa saja yang akan dipergunakan dalam pemrosesan data. Atribut dalam dataset bus Biskita di *filter* dari 7 menjadi 5 atribut yang digunakan. Atribut yang digunakan yaitu : *id*, KM Tempuh Koridor, Kode Kendaraan, Total Penumpang, dan Total Ritase.



Gambar 5. *Select Attributes*

Parameter pada operator *Select Attributes* yang digunakan tampak pada tabel dibawah ini.

Tabel 7. Parameter dan atribut yang dipilih pada operator *Select Attributes*

No	Parameter	Isi
1	<i>Attribute Filter Type</i>	<i>Subset</i>
2	<i>Selected Attributes</i>	ID, Kode Kendaraan, Total Penumpang, Total Ritase, dan KM Tempuh Koridor

Dari hasil pembacaan operator *Select Attribut* didapat informasi sebagai berikut.

Tabel 8. Statik dataset

No	Uraian	Isi
1	<i>Record</i>	3724
2	<i>Special Attributes</i>	1
3	<i>Regular Attributes</i>	4
4	<i>Attributes :</i>	
	<i>ID</i>	<i>Integer, missing 0</i>
	<i>Kode Kendaraan</i>	<i>Polynomial, missing 0</i>
	<i>Total Penumpang</i>	<i>Integer, missing 0</i>
	<i>Total Ritase</i>	<i>Integer, missing 0</i>
	<i>KM Tempuh Koridor</i>	<i>Real, missing 0</i>

Preprocessing

Dikarenakan pada dataset bus biskita tidak ditemukan *missing value* atau data yang tidak memiliki nilai maka *preprocessing* tidak dilakukan.

id	Integer	0	Min 1	Max 3724	Average 1862.500
Kode Kendaraan	Polynomial	0	Least TP007 (25)	Most TP020 (166)	Values TP020 (166), TP019 (160)... [47 more]
Total Penumpang	Integer	0	Min 0	Max 31506	Average 1163.020
Total Ritase	Integer	0	Min 0	Max 913	Average 25.688
KM Tempuh Koridor	Real	0	Min 0	Max 25031.700	Average 705.583

Gambar 6. Hasil Statik

Setelah dilihat dari hasil statistik, tidak ditemukan adanya *missing value*. Maka proses dapat dilanjutkan ke tahap selanjutnya.

Data Transformation

1. *Nominal to Numerical*, digunakan untuk mengubah tipe atribut yang bersifat non-numerik menjadi tipe numerik. Dalam dataset bus Biskita atribut Kode Kendaraan akan dirubah menjadi numerik.



Gambar 7. Nominal to Numerical

Parameter pada operator *Nominal to Numerical* yang digunakan tampak pada tabel dibawah ini.

Tabel 9. Parameter pada operator *Nominal to Numerical*

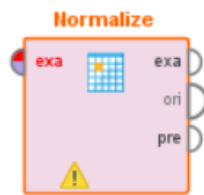
No	Parameter	Isi
1	Attribute Filter Type	Subset
2	Selected Attributes	Kode Kendaraan
3	Coding Type	Unique Integer

Dari hasil pembacaan operator *Nominal to Numerical* didapat informasi sebagai berikut.

Tabel 10. Statik dataset

No	Uraian	Isi
1	Record	3724
2	Special Attributes	1
3	Regular Attributes	4
4	Attributes :	
	ID	Integer, missing 0
	Kode Kendaraan	Numerical, missing 0
	Total Penumpang	Integer, missing 0
	Total Ritase	Integer, missing 0
	KM Tempuh Koridor	Real, missing 0

2. *Normalize*, digunakan untuk menskalakan nilai agar cocok dalam rentang tertentu. Menyesuaikan rentang nilai sangat penting saat berhadapan dengan atribut unit dan skala yang berbeda. Dalam dataset bus Biskita atribut yang akan dirubah menggunakan *normalize* yaitu KM Tempuh Koridor, Total Penumpang, dan Total Ritase.



Gambar 8. Normalize

Parameter pada operator *Normalize* yang digunakan tampak pada tabel dibawah ini. Operator ini mengambil nilai rata-rata pada atribut. Dengan tujuan membuat jarak antar data semakin dekat.

Tabel 11. Parameter dan atribut yang dipilih pada operator *Normalize*

No	Parameter	Isi
1	<i>Attribute Filter Type</i>	<i>Subset</i>
2	<i>Selected Attributes</i>	KM Tempuh Koridor, Total Penumpang, dan Total Ritase
3	<i>Method</i>	<i>Range Transformation</i>

Dari hasil pembacaan operator *Normalize* didapat informasi sebagai berikut.

Tabel 12. Statik dataset

No	Uraian	Isi
1	<i>Record</i>	3724
2	<i>Special Attributes</i>	1
3	<i>Regular Attributes</i>	4
4	<i>Attributes :</i>	
	<i>ID</i>	<i>Integer, missing 0</i>
	<i>Kode Kendaraan</i>	<i>Numerical, missing 0</i>
	<i>Total Penumpang</i>	<i>Real, missing 0</i>
	<i>Total Ritase</i>	<i>Real, missing 0</i>
	<i>KM Tempuh Koridor</i>	<i>Real, missing 0</i>

Data Mining

Dalam tahap ini algoritma yang digunakan adalah *k-means clustering*.



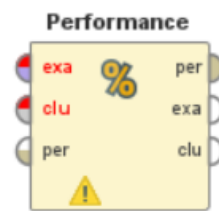
Gambar 9. K-Means Clustering

Parameter pada operator *K-Means Clustering* menggunakan parameter *default*. Dengan mengubah nilai *k* dimulai dari 2 sampai dengan 20.

Tabel 13. Parameter pada operator *K-Means Clustering*

No	Parameter	Isi
1	<i>K</i>	2-20

Selanjutnya ditambahkan operator *Performance* dengan metode *Davies-Bouldin Index (DBI)*. Dengan tujuan untuk mengetahui nilai *DBI* nya.



Gambar 10. *Cluster Distance Performance*

Parameter pada operator *Performance* yang digunakan tampak pada tabel dibawah ini.

Tabel 14. Parameter pada operator *Performance*

No	Parameter	Isi
1	<i>Main Criterion</i>	<i>Davies Bouldin</i>

Dari hasil pembacaan operator *Performance* didapat informasi sebagai berikut.

Tabel 15. Hasil operator *Performance (DBI)*

No	K	DBI
1	2	0,607
2	3	0,530
3	4	0,644
4	5	0,692
5	6	0,714
6	7	0,885
7	8	0,931
8	9	0,899
9	10	0,886
10	11	0,871
11	12	0,896
12	13	0,884
13	14	0,916
14	15	0,872
15	16	0,889
16	17	0,866
17	18	0,898

No	K	DBI
18	19	0,844
19	20	0,863

Interpretation/Evaluation

Setelah melakukan perbandingan *DBI* dengan metode *K-Means* dari k-2 sampai dengan k-20 yang terdapat pada tabel diatas, dapat dilihat bahwa *cluster* terkecil yang mendekati 0 yaitu k-3, dengan nilai *DBI* sebesar 0,530. Karena nilai k-3 merupakan nilai terkecil dibandingkan k lainnya, maka dapat disimpulkan bahwa k-3 dengan nilai 0,530 yang paling mendekati 0 merupakan hasil *cluster* terbaik. Dengan jumlah *cluster* 0 sebanyak 1972 *items*, *cluster* 1 sebanyak 1612 *items*, dan *cluster* 2 sebanyak 140 *items*.

Tabel 16. Hasil *Clustering* k-3 yang merupakan nilai *DBI* terbaik

No	Cluster	Keterangan
1	Cluster 0	1972
2	Cluster 1	1612
3	Cluster 2	140
	Total Number of Items	3724

PENUTUP

Berdasarkan hasil dari pengelompokkan data yang sudah dilakukan dengan metode *K-Means* menggunakan parameter *default* diperoleh hasil terkecil yaitu pada *cluster* k-3 dengan nilai *DBI* sebesar 0,530. Maka *cluster* k-3 merupakan *cluster* terbaik dari total k sebanyak 20. Dengan jumlah *cluster* 0 sebanyak 1972 *items*, *cluster* 1 sebanyak 1612 *items*, dan *cluster* 2 sebanyak 140 *items*.

DAFTAR PUSTAKA

- [1] M. Krishnamoorthy and R. Karthikeyan, "Pattern mining algorithms for data streams using itemset," *Meas. Sensors*, vol. 24, no. June, p. 100421, 2022, doi: 10.1016/j.measen.2022.100421.
- [2] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [3] Y. P. Sari, A. Primajaya, and A. S. Y. Irawan, "Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang," *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, p. 229, 2020, doi: 10.35314/isi.v5i2.1457.
- [4] V. Herlinda and D. Darwis, "Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means," *Darwis, Dartono*, vol. 2, no. 2, pp. 94-99, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [5] A. Wibowo, Moh Makruf, Inge Virdyna, and Farah Chikita Venna, "Penentuan Klaster Koridor TransJakarta dengan Metode Majority Voting pada Algoritma Data Mining," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*,

- vol. 5, no. 3, pp. 565–575, 2021, doi: 10.29207/resti.v5i3.3041.
- [6] J. Hutagalung and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1187, 2021, doi: 10.30865/mib.v5i3.3113.
- [7] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.
- [8] K. Shao, G. Mei, and Y. Wu, "Investigating changes in global distribution of Ozone in 2018 using k -means clustering algorithm," *J. Comput. Math. Data Sci.*, vol. 3, no. March, p. 100028, 2022, doi: 10.1016/j.jcmds.2022.100028.
- [9] W. Liu, P. Zou, D. Jiang, X. Quan, and H. Dai, "Zoning of reservoir water temperature field based on K-means clustering algorithm," *J. Hydrol. Reg. Stud.*, vol. 44, no. June, p. 101239, 2022, doi: 10.1016/j.ejrh.2022.101239.
- [10] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, p. 62, 2019, doi: 10.26418/jp.v5i1.29611.
- [11] I. R. Mahartika and A. Wibowo, "Data Mining Klasterisasi dengan Algoritme K-Means untuk Pengelompokan Provinsi Berdasarkan Konsumsi Bahan Bakar Minyak Nasional," *Pros. Semin. Nas. SISFOTEK (Sistem Inf. dan Teknol.*, vol. 3, no. 1, pp. 87–91, 2019, [Online]. Available: <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/108>
- [12] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95–100, 2021, [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839>
- [13] S. I. Murpratiwi, I. G. Agung Indrawan, and A. Aranta, "Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail," *J. Pendidik. Teknol. dan Kejuru.*, vol. 18, no. 2, p. 152, 2021, doi: 10.23887/jptk-undiksha.v18i2.37426.
- [14] N. . Anggraeni, "Teknik Clustering Dengan Algoritma K-Medoids Untuk Menangani Strategi Promosi Di Politeknik Tedc Bandung," *J. Teknol. Inf. dan Pendidik.*, 2019.
- [15] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.