

# Data Mining Pengelompokan Dataset Bus Biskita Di Kota Bogor Menggunakan Metode K-Means

Anjar Permadi<sup>a,1,\*</sup>

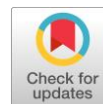
<sup>a</sup> Program Studi Manajemen Informatika STMIK IKMI Cirebon, Jl. Galunggung, Kecapi, Harjamukti, Kota Cirebon dan 45142, Indonesia

<sup>1</sup> [anjarpermadi300@gmail.com](mailto:anjarpermadi300@gmail.com).

\* Penulis Korespondensi

## ABSTRAK

*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*. Metode ini juga sudah banyak digunakan dalam penyelesaian masalah terkait klusterisasi pejualan, kebakaran hutan, pertanian, transportasi, dan sebagainya. 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 akan 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 penelitian yang telah dilakukan, dapat diketahui bahwa algoritma *K-Means* dapat digunakan untuk mengelompokkan dataset bus biskita. Yang nantinya bisa dimanfaatkan oleh perusahaan sebagai gambaran, penelitian juga ini bisa digunakan sebagai masukan bagi perusahaan/penyedia jasa.



## Kata Kunci

Clustering  
K-Means  
Data Mining  
DBI



This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

## 1. Pendahuluan

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].

Dalam penelitian ini penulis menggunakan acuan dari penelitian sebelumnya dengan menggunakan pengelompokan dan algoritma yang sama yaitu *K-Means*. Berdasarkan penelitian yang dilakukan oleh Erni Dianawati, Putri Previa Yanti dan Yulia Suryandari pada tahun 2019 dengan judul “Kluster Jumlah

Penumpang pada Halte Bus Rapid Transit Kota Tangerang”. Beberapa kendala yang dihadapi Bus Rapid Transit Trans Tangerang seperti jumlah penumpang yang perlu ditingkatkan untuk mengurangi kemacetan dan penempatan halte bus yang tepat untuk menarik penumpang. Oleh karena itu, untuk mengatasi masalah tersebut, diperlukan data dan informasi tentang BRT Trans Tangerang untuk menemukan pola distribusi penumpang. Proses pencarian informasi ini dapat dilakukan dengan menggunakan metode data mining. Dalam metode ini terdapat Teknik *clustering* yang berguna untuk melakukan pengelompokan semua data dalam kelompok yang memiliki kesamaan data. Hasil dari penelitian tersebut adalah dapat mengetahui jumlah total lintasan yang dihasilkan perbulan, sehingga dengan melakukan clustering karakteristik, pengetahuan tentang jadwal sibuk penumpang dalam satu hari dapat membantu pihak terkait mengantisipasi kepadatan penumpang pada hari-hari tertentu [8].

Menurut jurnal yang ditulis oleh Syahputra dan Ulfa, pertumbuhan penduduk di suatu wilayah akan berbanding lurus dengan kebutuhan sarana transportasi [9]. Berdasarkan hal itu, tidak menutup kemungkinan akan terjadi penumpukan penumpang, dikarenakan ketidaktahuan dalam menentukan jumlah penumpang di waktu tertentu. Untuk mengatasi masalah tersebut dibutuhkan data dan informasi terkait arus jumlah penumpang bus untuk menemukan pola distribusi penumpang.

Tujuan dari penelitian ini adalah proses penambahan data yang akan menghasilkan *cluster* sesuai dengan pengetahuan atau pola yang tersimpan pada data. Yang nantinya didapat hasil pengelompokan dataset bus Biskita Bogor. Diharapkan memberikan gambaran untuk perusahaan penyedia jasa yang nantinya hasil penelitian ini bisa digunakan sebagai masukan bagi perusahaan/penyedia jasa bus Biskita Bogor.

## 2. Metode

### 2.1. Data Mining

Data mining juga bisa disebut *knowledge discovery* adalah proses pengambilan pola pada data yang akan di proses lalu keluaran tersebut berupa informasi yang sangat penting. Dimana data merupakan kumpulan fakta dan dapat memberikan gambaran, jadi setiap kita melakukan sesuatu dalam pengambilan data maka data tersebut tersimpan dalam pola-pola data, yang nantinya akan diteliti secara manual sehingga kita bisa mengetahui hal-hal yang akan terjadi. *Data mining* sering juga disebut *Knowledge Discovery in Database (KDD)*, yaitu kegiatan yang meliputi pengumpulan data, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam dataset berukuran besar [10].

### 2.2. K-Means

Algoritma *k-means* adalah algoritma pengelompokan tipikal berdasarkan jarak. Pada algoritma *k-means*, semakin kecil jarak antar sampel, semakin tinggi kesamaannya [11]. Algoritma ini efisien dan banyak digunakan dalam kumpulan data yang tidak berlabel untuk tugas pengelompokan [12]. Algoritma ini sederhana untuk diimplementasikan, relatif cepat, mudah diadaptasi, dan umum digunakan dalam praktik. Metode *K-Means* membagi data menjadi beberapa kelompok sehingga data dengan karakteristik yang sama dimasukkan ke dalam kelompok yang sama dan data dengan karakteristik yang berbeda dikelompokkan ke dalam kelompok lain [13].

### 2.3. Clustering

*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* [14]. *Clustering* merupakan teknik untuk mengelompokkan data berdasarkan kesamaan karakteristik data [15]. 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 [16].

## 2.4. Davies-Bouldin Index

*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 [17]. Metode *DBI* dipilih karena dapat diimplementasikan dalam semua data baik ukuran besar dan kecil sehingga sangat cocok dan dapat diterima untuk perhitungan akurasi *clustering* [18].

## 2.5. Metode Pengumpulan Data

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 pada Tabel 1 yaitu data selama tahun 2022, dengan total jumlah data sebanyak 3724 *records*.

**Tabel 1.** Dataset Biskita Bogor dari salah satu koridor

No	Nomor Kendaraan	Kode Kendaraan	Jenis Bus	Total Penumpang	Total Ritase	KM Tempuh Koridor
1	F 7602 AC	TP002	Bus Sedang Swasta	636	11	225,50
2	F 7602 AC	TP002	Bus Sedang Swasta	596	10	205,00
3	F 7602 AC	TP002	Bus Sedang Swasta	528	10	205,00
4	F 7602 AC	TP002	Bus Sedang Swasta	507	11	225,50
5	F 7602 AC	TP002	Bus Sedang Swasta	524	10	205,00
6	F 7602 AC	TP002	Bus Sedang Swasta	1.172	20	410,00
7	F 7602 AC	TP002	Bus Sedang Swasta	4.494	120	2.460,00
8	F 7602 AC	TP002	Bus Sedang Swasta	1.011	20	410,00
9	F 7602 AC	TP002	Bus Sedang Swasta	426	11	225,50
10	F 7602 AC	TP002	Bus Sedang Swasta	15.298	304	6.232,00
...	...	...	...	...	...	...
721	F 7650 AC	TP038	Bus Sedang Swasta	341	11	225,50
722	F 7650 AC	TP038	Bus Sedang Swasta	331	11	225,50

## 2.6. Tahap Perancangan

Penerapan algoritma K-Means Clustering ini dirancang dengan menggunakan tool Rapidminer. Pada penelitian ini menggunakan tahapan *Knowledge Discovery in Database (KDD)*, adapun tahapan *KDD* adalah sebagai berikut :

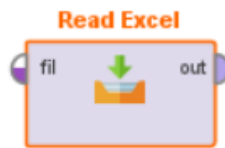
- *Data Selection*, pada tahap ini dilakukan proses seleksi atribut mana yang akan digunakan dalam proses data mining. Dari 6 atribut yaitu Nomor Kendaraan, Kode Kendaraan, Jenis Bus, Total Penumpang, Total Ritase, dan KM Tempuh Koridor. Dataset tersebut akan diseleksi menjadi 4 atribut yaitu KM Tempuh Koridor, Kode Kendaraan, Total Penumpang, dan Total Ritase. Kemudian ditambahkan operator *generate id*, dikarenakan data mentahnya belum memiliki *id*.
- *Preprocessing*, merupakan proses pembersihan data dari missing value atau noise yaitu data yang tidak relevan atau tidak konsisten. Pada tahap *preprocessing* ini, data yang diperoleh akan dibersihkan dari kesalahan, *missing*, atau ketidak lengkapan kriteria data.
- *Data Transformation*, merupakan proses perubahan tipe data menjadi tipe data numerik, agar data dapat diolah menggunakan algoritma *K-Means*.
- *Data Mining*, tahapan ini dilakukan dengan cara menerapkan algoritma atau pencarian pengetahuan. Pada penelitian ini, algoritma yang digunakan adalah algoritma *k-means clustering*.
- *Interpretation/Evaluation*, pada tahap ini dilakukan dengan menganalisa hasil eksperimen yang telah dilakukan.

### 3. Hasil dan Pembahasan

Penerapan algoritma *k-means clustering* ini dirancang dengan menggunakan tool Rapidminer. Pada penelitian ini, menggunakan tahapan *Knowledge Discovery in Database (KDD)*. Penerapan KDD dalam penelitian ini yaitu sebagai berikut :

#### 3.1. Data Selection

- *Read Excel* pada Gambar 1, operator ini digunakan untuk memasukan data excel yang berada di komputer pengguna ke dalam proses pada rapidminer. 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 1. Read Excel

Parameter pada operator *Read Excel* menggunakan parameter default. Dari hasil pembacaan operator *Read Excel* didapat informasi pada Tabel 2,3,4 dan 5.

Tabel 2. 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 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 Koridor	<i>Real, missing 0</i>

Tabel 3. 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 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	Integer, missing 0
KM Tempuh Koridor	Real, missing 0

**Tabel 4.** Statik dataset file ketiga

No	Uraian	Isi
1	Record	1362
2	Special Attribute	0
3	Regular Attribute	6
4	Attributes :	
	Nomor Kendaraan	Polynomial, missing 0
	Kode Kendaraan	Polynomial, missing 0
	Jenis Bus	Polynomial, missing 0
	Total Penumpang	Integer, missing 0
	Total Ritase	Integer, missing 0
	KM Tempuh Koridor	Real, missing 0

**Tabel 5.** Statik dataset file keempat

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

- *Append*, operator ini berfungsi menggabungkan beberapa dataset menjadi satu seperti terlihat pada Gambar 2. 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 2.** *Append*

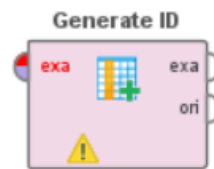
Untuk parameter yang digunakan pada operator ini adalah parameter *default*. Tabel 6 merupakan hasil dataset menggunakan operator *append*.

**Tabel 6.** 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 Koridor	<i>Real, missing 0</i>

- Gambar 3 merupakan *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.

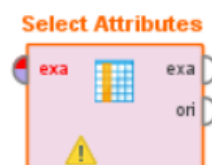


Gambar 3. *Generate ID*

Tabel 7. 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 Koridor	<i>Real, missing 0</i>

- Gambar 4 merupakan *Select Attributes*, operator ini dipergunakan untuk memfilter 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 4.** *Select Attributes*

Parameter pada operator *Select Attributes* yang digunakan tampak pada tabel 8.

**Tabel 8.** 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 Attribute* didapat informasi pada Tabel 9.

**Tabel 9.** 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>

### 3.2. Preprocessing

Dataset bus biskita tidak ditemukan *missing value* atau data yang tidak memiliki nilai, sehingga *preprocessing* tidak dilakukan.

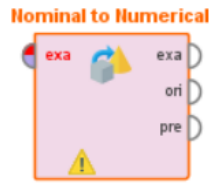
id	Integer	0	Min 1	Max 3724	Average 1862.500
Kode Kendaraan	Polynomial	0	Least TP007 (25)	Most TP020 (166)	Value TP020 (166), TP019 (160), ...[47 more]
Total Penumpang	Integer	0	Min 0	Max 31506	Average 1163.020
Total Ritase	Integer	0	Min 0	Max 913	Average 25.688
KM Tempuh Koridor	Real	0	Min 0	Max 25031.700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

**Gambar 5.** Hasil Statistik

Setelah dilihat dari hasil statistik pada Gambar 5, tidak ditemukan adanya *missing value*. Maka proses dapat dilanjutkan ke tahap selanjutnya.

### 3.3. Data Transformation

- *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 numeric seperti pada Gambar 6.



**Gambar 6.** *Nominal to Numerical*

Parameter pada operator *Nominal to Numerical* yang digunakan tampak pada tabel 10.

**Tabel 10.** Parameter pada operator *Nominal to Numerical*

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

Dari hasil pembacaan operator *Nominal to Numerical* didapat informasi pada Tabel 11.

**Tabel 11.** 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>
	Kode Kendaraan	<i>Numerical, missing 0</i>
	Total Penumpang	<i>Integer, missing 0</i>
	Total Ritase	<i>Integer, missing 0</i>
	KM Tempuh Koridor	<i>Real, missing 0</i>

- Gambar 7 merupakan *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 7.** *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 seperti terlihat pada Tabel 12.

**Tabel 12.** 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>

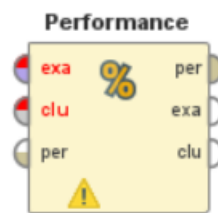
Dari hasil pembacaan operator *Normalize* didapat informasi pada Tabel 13.

**Tabel 13.** 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>
	Kode Kendaraan	<i>Numerical, missing 0</i>
	Total Penumpang	<i>Real, missing 0</i>
	Total Ritase	<i>Real, missing 0</i>
	KM Tempuh Koridor	<i>Real, missing 0</i>

### 3.4. Data Mining

Dalam tahap ini algoritma yang digunakan adalah *k-means clustering*. Parameter pada operator *K-Means Clustering* menggunakan parameter default. Dengan mengubah nilai *k* dimulai dari 2 sampai dengan 20. Selanjutnya ditambahkan operator *Performance* dengan metode *Davies-Bouldin Index (DBI)*. Dengan tujuan untuk mengetahui nilai *DBI* nya seperti terlihat pada Gambar 8. Parameter pada operator *Performance* yang digunakan adalah *Davies Bouldin Index (DBI)*. Dari hasil pembacaan operator *Performance* didapat informasi pada Tabel 14.



**Gambar 8.** Cluster Distance Performance

**Tabel 14.** 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

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

### 3.5. 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* seperti terlihat pada Tabel 15.

**Tabel 15.** 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

## 4. Kesimpulan

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] Krishnamoorthy, M., and R. Karthikeyan. 2022. "Pattern Mining Algorithms for Data Streams Using Itemset." *Measurement: Sensors* 24 (June): 100421. <https://doi.org/10.1016/j.measen.2022.100421>.
- [2] Utomo, Dito Putro, and Mesran Mesran. 2020. "Analisis Komparasi Metode Klasifikasi Data Mining Dan Reduksi Atribut Pada Data Set Penyakit Jantung." *Jurnal Media Informatika Budidarma* 4 (2): 437. <https://doi.org/10.30865/mib.v4i2.2080>.
- [3] Sari, Yanti Puspita, Aji Primajaya, and Agung Susilo Yuda Irawan. 2020. "Implementasi Algoritma K-Means Untuk Clustering Penyebaran Tuberkulosis Di Kabupaten Karawang." *INOVTEK Polbeng - Seri Informatika* 5 (2): 229. <https://doi.org/10.35314/isi.v5i2.1457>.

- [4] Herlinda, Vera, and Dedi Darwis. 2021. "Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means." *Darwis, Dartono 2* (2): 94–99. <http://jim.teknokrat.ac.id/index.php/JTISI>.
- [5] Wibowo, Arief, Moh Makruf, Inge Virdyna, and Farah Chikita Venna. 2021. "Penentuan Klaster Koridor TransJakarta Dengan Metode Majority Voting Pada Algoritma Data Mining." *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)* 5 (3): 565–75. <https://doi.org/10.29207/resti.v5i3.3041>.
- [6] Hutagalung, Juniar, and Fifin Sonata. 2021. "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah." *Jurnal Media Informatika Budidarma* 5 (3): 1187. <https://doi.org/10.30865/mib.v5i3.3113>.
- [7] Aditya, Agil, Ivan Jovian, and Betha Nurina Sari. 2020. "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama Di Indonesia Tahun 2018/2019." *Jurnal Media Informatika Budidarma* 4 (1): 51. <https://doi.org/10.30865/mib.v4i1.1784>.
- [8] Dianawati, Erni, Putri Previa Yanti, and Yulia Suryandari. 2019. "Klustering Jumlah Penumpang Pada Halte Bus Rapid Transit Kota Tangerang." *Jurnal Sistem Cerdas* 2 (3): 163–72. <https://doi.org/10.37396/jsc.v2i3.34>.
- [9] Syahputra, M R, and M Ulfa. 2021. "Penerapan Data Mining Dalam Pengelompokan Jumlah Penumpang Trans Musi Di Kota Palembang Menggunakan Metode K-Means." *Bina Darma Conference on ...* <https://conference.binadarma.ac.id/index.php/BDCCS/article/view/2160>.
- [10] Anggraeni, N.L. 2019. "Teknik Clustering Dengan Algoritma K-Medoids Untuk Menangani Strategi Promosi Di Politeknik Tedc Bandung." *Jurnal Teknologi Informasi Dan Pendidikan*.
- [11] Shao, Kaixuan, Gang Mei, and Yinghan Wu. 2022. "Investigating Changes in Global Distribution of Ozone in 2018 Using k -Means Clustering Algorithm." *Journal of Computational Mathematics and Data Science* 3 (March): 100028. <https://doi.org/10.1016/j.jcmds.2022.100028>.
- [12] Liu, Wei, Peng Zou, Dingguo Jiang, Xiufeng Quan, and Huichao Dai. 2022. "Zoning of Reservoir Water Temperature Field Based on K-Means Clustering Algorithm." *Journal of Hydrology: Regional Studies* 44 (June): 101239. <https://doi.org/10.1016/j.ejrh.2022.101239>.
- [13] Sulistiyawati, Ari, and Eko Supriyanto. 2021. "Implementasi Algoritma K-Means Clustering Dalam Penentuan Siswa Kelas Unggulan." *Jurnal Tekno Kompak* 15 (2): 25. <https://doi.org/10.33365/jtk.v15i2.1162>.
- [14] Priyatman, Hendro, Fahmi Sajid, and Dannis Haldivany. 2019. "Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Waktu Kelulusan Mahasiswa." *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)* 5 (1): 62. <https://doi.org/10.26418/jp.v5i1.29611>.
- [15] Aditya, Agil, Ivan Jovian, and Betha Nurina Sari. 2020. "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama Di Indonesia Tahun 2018/2019." *Jurnal Media Informatika Budidarma* 4 (1): 51. <https://doi.org/10.30865/mib.v4i1.1784>.
- [16] Mahartika, Indah Rizky, and Arief Wibowo. 2019. "Data Mining Klasterisasi Dengan Algoritme K-Means Untuk Pengelompokkan Provinsi Berdasarkan Konsumsi Bahan Bakar Minyak Nasional." *Prosiding Seminar Nasional SISFOTEK (Sistem Informasi Dan Teknologi)* 3 (1): 87–91. <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/108>.
- [17] Muningsih, Elly, Ina Maryani, and Vembria Rose Handayani. 2021. "Penerapan Metode K-Means Dan Optimasi Jumlah Cluster Dengan Index Davies Bouldin Untuk Clustering Propinsi Berdasarkan Potensi Desa." *Jurnal Sains Dan Manajemen* 9 (1): 95–100. <https://ejournal.bsi.ac.id/ejournal/index.php/evolusi/article/view/10428/4839>.
- [18] Murpratiwi, Santi Ika, I Gusti Agung Indrawan, and Arik Aranta. 2021. "Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail." *Jurnal Pendidikan Teknologi Dan Kejuruan* 18 (2): 152. <https://doi.org/10.23887/jptk-undiksha.v18i2.37426>.