

Clustering Koridor Transjakarta Berdasarkan Jumlah Penumpang Dengan Algoritma K-Means

Adi Supriyatna¹, Irmawati Carolina², Suhar Janti³, Ali Haidir⁴

Universitas Bina Sarana Informatika

Jl. Kramat Raya No.98, Senen, Jakarta Pusat, 10450

adi.asp@bsi.ac.id, irmawati.imc@bsi.ac.id, suharjanti.shj@bsi.ac.id, ali.alh@bsi.ac.id

Abstract

Transportation is one of the facilities that make it easy for humans to carry out activities to move places using vehicles that are driven by humans or machines. Based on data obtained from data.jakarta.go.id, the number of Transjakarta bus passengers in corridors 1 to 13 of 2017 amounted to 114,239,960, and in 2018 there were 121,918,964 passengers. The algorithm used in this research is K-Means Cluster, which is implemented using Microsoft Excel and Rapidminer Studio. The purpose of this study is to cluster Transjakarta corridors based on the number of passengers divided into 3 clusters: high, medium, and low. The results of data processing show that the Transjakarta corridor data cluster is based on the number of passengers using the K-Means cluster algorithm using Microsoft Excel and Rapidminer Studio to produce 3 clusters, namely cluster 1 with the highest number of passengers, one corridor, cluster 2 with the number of passengers being nine corridors and cluster 3 or 0 with a low number of passengers there are three corridors. The highest number of passengers is corridor one which serves the Blok M - Kota route, indicating that the Blok M - Kota route is the most used by Transjakarta passengers.

Keywords: Algoritma K-Means, Clustering, Koridor Transjakarta.

Abstrak

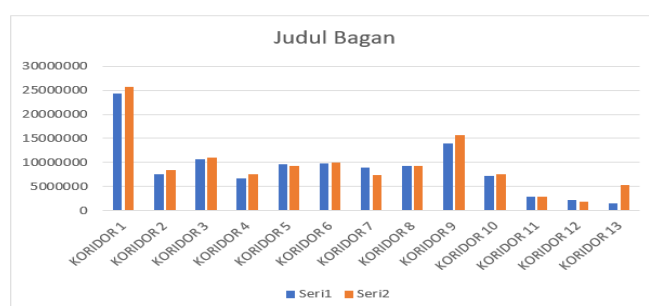
Transportasi merupakan salah satu sarana yang memberikan kemudahan kepada manusia dalam melakukan kegiatan untuk berpindah tempat dengan menggunakan kendaraan yang digerakkan oleh manusia atau mesin. Berdasarkan data yang diperoleh dari data.jakarta.go.id jumlah penumpang bus transjakarta pada koridor 1 sampai dengan 13 tahun 2017 berjumlah 114.239.960 dan tahun 2018 berjumlah 121.918.964 penumpang. Algoritma yang digunakan pada penelitian ini yaitu K-Means Cluster yang diimplementasikan menggunakan Microsoft Excel dan Rapidminer Studio. Tujuan dari penelitian ini adalah untuk mengelompokkan koridor-koridor transjakarta berdasarkan jumlah penumpang yang dibagi menjadi 3 cluster yaitu tinggi, sedang dan rendah. Hasil pengolahan data menunjukkan Pengelompokan data koridor transjakarta berdasarkan jumlah penumpang dengan algoritma K-Means cluster menggunakan Microsoft Excel dan Rapidminer Studio menghasilkan 3 cluster yaitu cluster 1 dengan jumlah penumpang terbanyak terdapat 1 koridor, cluster 2 dengan jumlah penumpang sedang terdapat 9 koridor dan cluster 3 atau 0 dengan jumlah penumpang rendah terdapat 3 koridor. Jumlah penumpang tertinggi yaitu koridor 1 yang melayani rute Blok M - Kota, hal tersebut menunjukkan bahwa trayek Blok M - Kota paling banyak digunakan oleh penumpang transjakarta.

Kata kunci: K-Means Algorithm, Clustering, Transjakarta Corridor.

1. PENDAHULUAN

Transportasi merupakan salah satu sarana yang memberikan kemudahan kepada manusia dalam melakukan kegiatan untuk berpindah tempat dengan menggunakan kendaraan yang digerakkan oleh manusia atau mesin. Transjakarta merupakan salah satu sistem transportasi *Bus Rapid Transit (BRT)* yang pertama kali di Asia Tenggara dan Selatan yang mulai

beroperasi sejak tahun 2004 di Indonesia khususnya DKI Jakarta[1]. Berdasarkan data dari situs resmi transjakarta.co.id saat ini koridor bus transjakarta yang sudah tersedia sebanyak 13 koridor utama dan 25 rute lintas koridor. Berdasarkan data yang diperoleh dari data.jakarta.go.id jumlah penumpang bus transjakarta pada koridor 1 sampai dengan 13 tahun 2017 berjumlah 114.239.960 dan tahun 2018 berjumlah 121.918.964 penumpang[2], gambar 1. Merupakan grafik jumlah penumpang transjakarta pada tahun 2017 dan 2018.



Sumber: data.jakarta.go.id

Gambar 1. Jumlah Penumpang Transjakarta 2017 dan 2018

Clustering merupakan metode pengelompokan yang sudah dikenal[3], yang digunakan untuk pengolahan data dalam mengelompokkan objek yang memiliki kemiripan[4][5][6]. Metode *clustering* yang digunakan pada penelitian ini adalah *K-Means* yang termasuk ke dalam golongan algoritma *Partitioning Clustering*[7][8]. Metode *K-Means clustering* juga banyak digunakan dalam penyelesaian masalah tentang pengelompokkan seperti penjualan[9], kebakaran hutan[6], pertanian[10], transportasi[11][12].

Hasil *clustering* jumlah penumpang pada halte BRT kota Tangerang memiliki karakteristik yang mirip antara data-data yang berada dalam satu *cluster*, yaitu jumlah penumpang secara total perbulannya, sehingga jadwal padatnya penumpang dalam satu hari dapat membantu pihak terkait untuk mengantisipasi padatnya penumpang pada hari tertentu[7]. pengelompokan data jumlah penumpang pada Bandar Udara Hang Nadim Batam dengan menggunakan data mining *clustering* algoritma *K-means* menghasilkan 3 *cluster* yang dimana data di dalamnya memiliki sifat yang sejenis. Sifat yang sejenis yang dimaksud ada terletak pada rata-rata jumlah penumpang yang ada pada setiap bulannya[13]. Hasil proses *clustering* pada jumlah penumpang bus Trans Jogja didapatkan 3 *cluster* dengan jumlah *cluster* 0 sebanyak 15 data, *cluster* 1 sebanyak 44 data dan *cluster* 2 sebanyak 54 data, sehingga dapat dibuktikan bahwa metode *K-Means* memiliki hasil yang lebih baik dari pada metode AHC[3].

Tujuan dari penelitian ini adalah untuk mengelompokkan koridor-koridor transjakarta berdasarkan jumlah penumpang yang dibagi menjadi 3 *cluster* yaitu tinggi, sedang dan rendah. Diharapkan hasil penelitian ini dapat digunakan untuk bahan pertimbangan dalam peningkatan jumlah armada untuk koridor yang termasuk ke dalam *cluster* jumlah penumpangnya tinggi.

2. METODOLOGI PENELITIAN

Penelitian pengelompokan koridor transjakarta berdasarkan jumlah penumpang dengan algoritma k-means ini dilakukan dengan beberapa tahapan, yaitu:

a) Pengumpulan Data.

Tahapan ini dilakukan untuk mencari data-data yang akan digunakan dalam penelitian terkait pengelompokan koridor bus transjakarta berdasarkan jumlah penumpang. data diambil dari situs resmi Jakarta Open Data (data.jakarta.go.id).

b) *Preprocessing Data*.

Preprocessing data yang dilakukan menggunakan Teknik normalisasi, data yang dinormalisasi digunakan untuk skala nilai atribut dari data sehingga mereka cocok dalam rentang tertentu. Rumus yang digunakan dalam normalisasi data adalah *Min-Max* yang merupakan metode normalisasi dengan melakukan transformasi linier terhadap data yang asli.

c) Pengolahan dataset menggunakan *Microsoft Excel*.

Merupakan tahap pengolahan data menggunakan algoritma *K-Means* untuk mendapatkan kelompok data koridor transjakarta berdasarkan jumlah penumpang. Aplikasi yang digunakan tahapan ini adalah *Microsoft Excel*. Tahapan yang dilakukan dalam algoritma *K-Means* adalah sebagai berikut[10]:

1) Tentukan jumlah *cluster* (*c*), tetapkan pusat *cluster* sembarang.

2) Hitung jarak setiap data kepusat *cluster* menggunakan persamaan:

$$d_{ik} = \sqrt{\sum_{j=1}^m C_{ij} - C_{kj}^2} \quad (1)$$

3) Kelompokkan data ke dalam *cluster* dengan jarak yang paling pendek menggunakan persamaan:

$$\text{Min } \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m C_{ij} - C_{kj}^2} \quad (2)$$

4) Hitung pusat *cluster* yang baru menggunakan persamaan:

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (3)$$

5) Ulangi langkah kedua sampai tidak ada lagi data yang berpindah ke *cluster* lain.

d) Implementasi *K-Means Cluster* dengan Rapidminer.

Sebagai bahan perbandingan pengolahan data yang sudah dilakukan menggunakan *Microsoft Excel* juga dilakukan pengolahan data kembali menggunakan aplikasi Rapidminer Studio.

e) Analisa Hasil Pengolahan Data.

Berdasarkan hasil pengolahan data yang telah dilakukan menggunakan *Microsoft Excel* dan Rapidminer Studio, kemudian data tersebut dianalisa dalam bentuk *cluster model* dan *performance vector* atas pengelompokan koridor transjakarta berdasarkan jumlah penumpang.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data jumlah penumpang pada 13 koridor transjakarta di tahun 2017 dan 2018.

Tabel 1. Data Koridor Transjakarta Dan Jumlah Penumpang.

No	Trayek	2017	2018
1	Koridor 1	24257058	25809337
2	Koridor 2	7498917	8409829
3	Koridor 3	10619387	10979625
4	Koridor 4	6683555	7521168
5	Koridor 5	9597209	9234312
6	Koridor 6	9836477	9980206
7	Koridor 7	8966521	7429375
8	Koridor 8	9238704	9347184
9	Koridor 9	13956227	15606290
10	Koridor 10	7123737	7543560
11	Koridor 11	2917261	2888342
12	Koridor 12	2124478	1902479
13	Koridor 13	1418412	5267257

3.2. Preprocessing Data

Teknik normalisasi data yang digunakan adalah *Min-Max* dengan persamaan[6]:

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (4)$$

Dimana :

X = Data yang dinormalisasi

Min(X) = Nilai minimal dari range data yang akan dinormalisasi

Max(X) = Nilai maximal dari range data yang akan dinormalisasi.

Hasil *preprocessing data* dengan Teknik normalisasi, maka didapatkan data seperti pada tabel 2 dibawah ini:

Tabel 2. Data Normalisasi Koridor Transjakarta Dan Jumlah Penumpang.

Trayek	2017	2018
Koridor 1	1,000	1,000
Koridor 2	0,243	0,272
Koridor 3	0,384	0,380
Koridor 4	0,206	0,235
Koridor 5	0,338	0,307
Koridor 6	0,348	0,338
Koridor 7	0,309	0,231
Koridor 8	0,321	0,311
Koridor 9	0,535	0,573
Koridor 10	0,226	0,236

Trayek	2017	2018
Koridor 11	0,036	0,041
Koridor 12	0,000	0,000
Koridor 13	-0,032	0,141

3.3. Pengolahan Dataset dengan *Microsoft Excel*

Setelah dilakukan proses pengolahan dataset koridor transjakarta berdasarkan jumlah penumpang menggunakan *Microsoft Excel*, untuk mendapatkan pusat *cluster* dan penempatan data dalam *cluster* dilakukan dengan dua kali perulangan atau iterasi, karena pada iterasi kedua pengelompokan data kedalam *cluster* sama dengan iterasi pertama.

a) Iterasi Pertama

1. Menentukan jumlah *cluster*.

Jumlah *cluster* yang digunakan pada penelitian ini berjumlah 3 *cluster* yang terdiri dari *cluster* 1 (C1) merupakan koridor dengan jumlah penumpang terbanyak, *cluster* 2 (C2) merupakan koridor dengan jumlah penumpang sedang dan *cluster* 3 (C3) merupakan koridor dengan jumlah penumpang sedikit. *Cluster* yang ditentukan menggunakan rumus pengukuran jarak/ *Euclidian Distance* seperti yang terlihat pada tabel 3 dibawah ini.

Tabel 3. Pusat *Cluster* Iterasi Pertama

<i>Cluster</i> 1 (Tinggi)	1	1
<i>Cluster</i> 2 (Sedang)	0,309139	0,231185
<i>Cluster</i> 3 (Rendah)	0	0

2. Menghitung jarak setiap data ke pusat *cluster*.

Tabel 4. Jarak Data Ke Pusat *Cluster* Iterasi Pertama

Trayek	C1	C2	C3	Jarak Terdekat
Koridor 1	0,000	1,282	2,000	0,000
Koridor 2	1,287	0,068	0,317	0,068
Koridor 3	1,001	0,097	0,528	0,097
Koridor 4	1,379	0,103	0,261	0,103
Koridor 5	1,143	0,034	0,432	0,034
Koridor 6	1,090	0,051	0,463	0,051
Koridor 7	1,282	0,000	0,363	0,000
Koridor 8	1,153	0,019	0,418	0,019
Koridor 9	0,648	0,342	0,863	0,342
Koridor 10	1,358	0,083	0,282	0,083
Koridor 11	1,883	0,309	0,038	0,038
Koridor 12	2,000	0,363	0,000	0,000
Koridor 13	1,770	0,349	0,052	0,052

3. Kelompokkan data ke dalam *cluster* dengan jarak yang paling pendek.

Tabel 5. Pengelompokan Data Iterasi Pertama

Trayek	C1	C2	C3	Jarak Terdekat	Cluster
Koridor 1	0,000	1,282	2,000	0,000	C1
Koridor 2	1,287	0,068	0,317	0,068	C2
Koridor 3	1,001	0,097	0,528	0,097	C2
Koridor 4	1,379	0,103	0,261	0,103	C2
Koridor 5	1,143	0,034	0,432	0,034	C2
Koridor 6	1,090	0,051	0,463	0,051	C2
Koridor 7	1,282	0,000	0,363	0,000	C2
Koridor 8	1,153	0,019	0,418	0,019	C2
Koridor 9	0,648	0,342	0,863	0,342	C2
Koridor 10	1,358	0,083	0,282	0,083	C2
Koridor 11	1,883	0,309	0,038	0,038	C3
Koridor 12	2,000	0,363	0,000	0,000	C3
Koridor 13	1,770	0,349	0,052	0,052	C3

4. Hitung pusat *cluster* yang baru.

$$C1(2017) = \frac{1}{1} = 1$$

$$C2(2017) = \frac{0,243 + 0,384 + 0,206 + 0,338 + 0,348 + 0,309 + 0,321 + 0,535 + 0,226}{9} = 0,323$$

$$C3(2017) = \frac{0,036 + 0,000 + (-0,032)}{3} = 0,001$$

$$C1(2018) = \frac{1}{1} = 1$$

$$C2(2018) = \frac{0,272 + 0,380 + 0,235 + 0,307 + 0,338 + 0,231 + 0,311 + 0,573 + 0,236}{9} = 0,320$$

$$C3(2018) = \frac{0,041 + 0,000 + (0,141)}{3} = 0,061$$

b) Iterasi Kedua

1. Menentukan Jumlah *Cluster*

Hasil perhitungan pusat *cluster* yang baru pada iterasi pertama dijadikan sebagai pusat *cluster* pada iterasi kedua.

Tabel 6. Pusat *Cluster* Iterasi Kedua

Cluster 1 (Tinggi)	1	1
Cluster 2 (Sedang)	0,323	0,320
Cluster 3 (Rendah)	0,0001	0,061

2. Menghitung jarak setiap data ke pusat *cluster*.

Tabel 7. Jarak Data Ke Pusat *Cluster* Iterasi Kedua

Trayek	C1	C2	C3	Jarak Terdekat
Koridor 1	0,000	1,139	1,881	0,000

Trayek	C1	C2	C3	Jarak Terdekat
Koridor 2	1,287	0,083	0,286	0,083
Koridor 3	1,001	0,064	0,484	0,064
Koridor 4	1,379	0,125	0,235	0,125
Koridor 5	1,143	0,015	0,397	0,015
Koridor 6	1,090	0,025	0,424	0,025
Koridor 7	1,282	0,022	0,337	0,022
Koridor 8	1,153	0,002	0,383	0,002
Koridor 9	0,648	0,275	0,796	0,275
Koridor 10	1,358	0,105	0,255	0,105
Koridor 11	1,883	0,365	0,035	0,035
Koridor 12	2,000	0,426	0,005	0,005
Koridor 13	1,770	0,387	0,040	0,040

3. Kelompokkan data ke dalam *cluster* dengan jarak yang paling pendek.

Tabel 8. Pengelompokan Data Iterasi Kedua

Trayek	C1	C2	C3	Jarak Terdekat	Cluster
Koridor 1	0,000	1,139	1,881	0,000	C1
Koridor 2	1,287	0,083	0,286	0,083	C2
Koridor 3	1,001	0,064	0,484	0,064	C2
Koridor 4	1,379	0,125	0,235	0,125	C2
Koridor 5	1,143	0,015	0,397	0,015	C2
Koridor 6	1,090	0,025	0,424	0,025	C2
Koridor 7	1,282	0,022	0,337	0,022	C2
Koridor 8	1,153	0,002	0,383	0,002	C2
Koridor 9	0,648	0,275	0,796	0,275	C2
Koridor 10	1,358	0,105	0,255	0,105	C2
Koridor 11	1,883	0,365	0,035	0,035	C3
Koridor 12	2,000	0,426	0,005	0,005	C3
Koridor 13	1,770	0,387	0,040	0,040	C3

4. Hitung pusat *cluster* yang baru

$$C1(2017) = \frac{1}{1} = 1$$

$$C2(2017) = \frac{0,243 + 0,384 + 0,206 + 0,338 + 0,348 + 0,309 + 0,321 + 0,535 + 0,226}{9} = 0,323$$

$$C3(2017) = \frac{0,036 + 0,000 + (-0,032)}{3} = 0,001$$

$$C1(2018) = \frac{1}{1} = 1$$

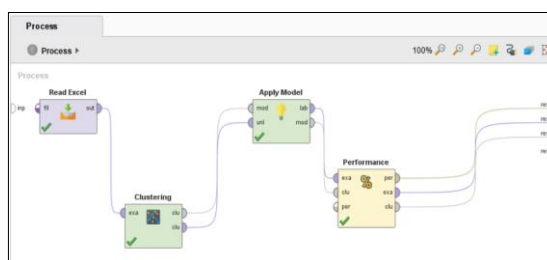
$$C2(2018) = \frac{0,272 + 0,380 + 0,235 + 0,307 + 0,338 + 0,231 + 0,311 + 0,573 + 0,236}{9} = 0,320$$

$$C3(2018) = \frac{0,041 + 0,000 + (0,141)}{3} = 0,061$$

3.4. Implementasi *K-Means Cluster* dengan *Rapid Miner*

Dataset jumlah penumpang transjakarta berdasarkan koridor dengan *centroid data* awal yang telah ditentukan, kemudian dilakukan pengolahan data dengan algoritma *K-Means cluster* menggunakan aplikasi Rapidminer studio.

a) Pemodelan pengolahan data dengan algoritma *K-Means cluster* pada Rapidminer studio.

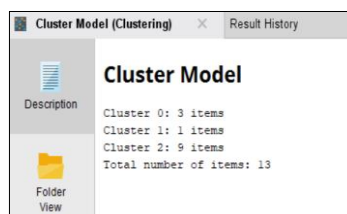


Gambar 2. Pemodelan Algoritma *K-Means Cluster*

Perancangan pengolahan data dengan algoritma *K-Means* pada Rapidminer studio menggunakan 4 operator, *Read Excel* digunakan untuk pembacaan dataset jumlah penumpang transjakarta berdasarkan koridor yang sudah di normalisasi, kemudian dataset tersebut di klaster atau dikelompokkan menggunakan operator *K-Means clustering* dengan nilai $K=3$ dan $\text{max runs}=10$, operator *Apply Model* digunakan agar dapat membaca dataset yang di klaster berdasarkan data yang sudah dipelajari. Operator *Cluster Performance Distance* digunakan Untuk mengukur kinerja algoritma *K-Means clustering* terhadap dataset yang dibaca.

b) *Cluster Model*.

Hasil pengolahan data pengelompokkan koridor transjakarta berdasarkan penumpang dengan algoritma *K-Means* pada Rapidminer studio dapat dilihat pada gambar dibawah ini:

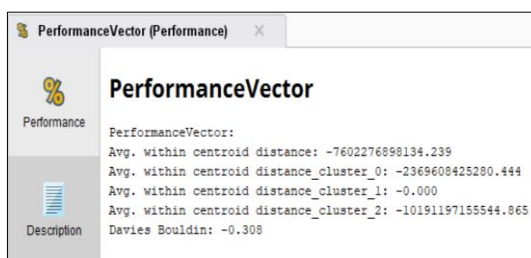


Gambar 3. *Cluster Model* Yang Dihasilkan Algoritma *K-Means*

Gambar 3 menunjukkan hasil pengolahan data pengelompokkan koridor transjakarta berdasarkan penumpang menggunakan algoritma *K-Means cluster* terbagi menjadi 3 *cluster* yaitu, 3 provinsi yang termasuk kedalam *cluster* 0, 1 koridor termasuk kedalam *cluster* 1 dan 9 koridor termasuk kedalam *cluster* 2.

c) Performance Vector.

Parameter yang digunakan untuk mengukur kinerja algoritma K-Means dilakukan dengan menghitung *avg_within_centroid_distance* dan *davies_bouldin*. *avg_within_centroid_distance* merupakan rata-rata dalam jarak *cluster* yang dihitung dengan rata-rata jarak antara *centroid* dengan semua contoh *cluster*. Sedangkan *davies_bouldin* merupakan algoritma yang menghasilkan *cluster* dengan jarak intra-*cluster* rendah (tingkat kesamaan intra-*cluster* tinggi) dan jarak inter-*cluster* tinggi (tingkat kesamaan inter-*cluster* rendah) akan memiliki *Davies-Bouldin index* yang rendah. algoritma pengelompokan yang menghasilkan kumpulan *cluster* dengan indeks *Davies-Bouldin* terkecil dianggap sebagai algoritma terbaik berdasarkan kriteria. Hasil pengukuran kinerja algoritma *K-Means cluster* dalam melakukan pengelompokan data koridor transjakarta menggunakan operator *cluster performance distance* pada rapidminer studio, dapat dilihat pada gambar 4. dibawah ini:



Gambar 4. Performance Vector Yang Dihasilkan Algoritma K-Means

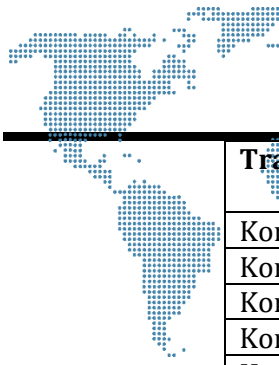
Gambar 4. merupakan *performance vector* pengolahan data koridor transjakarta yang menunjukkan rata-rata dalam jarak *centroid* sebesar -760227689134.239, rata-rata dalam *centroid distance_cluster_0* sebesar -2369608425280.444, rata-rata dalam *centroid distance_cluster_1* sebesar -0.000, rata-rata dalam *centroid distance_cluster_2* sebesar -10191197155544.865 dan Davies Bouldin sebesar -0.308.

3.5. Analisis Hasil Pengolahan Data.

Pengelompokan data koridor transjakarta berdasarkan jumlah penumpang dengan algoritma K-Means *Cluster* menggunakan *Microsoft Excel* dan Aplikasi *Rapidminerstudio* keduanya menghasilkan 3 *cluster*, gambar 5. dibawah ini merupakan pengelompokan 13 koridor transjakarta kebeberapa *cluster* yang terbentuk.

Tabel 9. Pengelompokan Data Menggunakan Ms. Excel

Trayek	Jumlah Penumpang		Cluster
	2017	2018	
Koridor 1	24257058	25809337	Cluster 1
Koridor 2	7498917	8409829	Cluster 2
Koridor 3	10619387	10979625	Cluster 2
Koridor 4	6683555	7521168	Cluster 2



Trayek	Jumlah Penumpang		Cluster
	2017	2018	
Koridor 5	9597209	9234312	Cluster 2
Koridor 6	9836477	9980206	Cluster 2
Koridor 7	8966521	7429375	Cluster 2
Koridor 8	9238704	9347184	Cluster 2
Koridor 9	13956227	15606290	Cluster 2
Koridor 10	7123737	7543560	Cluster 2
Koridor 11	2917261	2888342	Cluster 3
Koridor 12	2124478	1902479	Cluster 3
Koridor 13	1418412	5267257	Cluster 3

Open in  Turbo Prep  Auto Model

Row No.	Trayek	cluster	2017.0	2018.0
1	KORIDOR 1	cluster_1	24257058	25809337
2	KORIDOR 2	cluster_2	7498917	8409829
3	KORIDOR 3	cluster_2	10619387	10979625
4	KORIDOR 4	cluster_2	6683555	7521168
5	KORIDOR 5	cluster_2	9597209	9234312
6	KORIDOR 6	cluster_2	9836477	9980206
7	KORIDOR 7	cluster_2	8966521	7429375
8	KORIDOR 8	cluster_2	9238704	9347184
9	KORIDOR 9	cluster_2	13956227	15606290
10	KORIDOR 10	cluster_2	7123737	7543560
11	KORIDOR 11	cluster_0	2917261	2888342
12	KORIDOR 12	cluster_0	2124478	1902479
13	KORIDOR 13	cluster_0	1418412	5267257

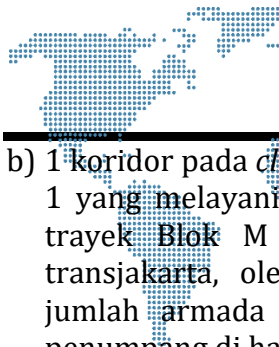
Gambar 5. Pengelompokan Data Menggunakan Rapidminer

Berdasarkan hasil pengolahan data dalam pengelompokan koridor transjakarta berdasarkan jumlah penumpang dengan algoritma *K-Means* menggunakan aplikasi *Microsoft Excel* dan *rapidminer studio* didapatkan 1 koridor termasuk ke *cluster 1* yaitu koridor 1 (Blok M – Kota), 9 koridor termasuk ke *cluster 2* yaitu koridor 2 (Pulogadung – Harmoni), koridor 3 (Kalideres – Pasar Baru), koridor 4 (Pulogadung 2 – Dukuh Atas 2), koridor 5 (Kp. Melayu – Ancol), koridor 6 (Ragunan – Dukuh Atas 2), koridor 7 (Kp. Rambutan – Kp. Melayu), koridor 8 (Lebak Bulus – Harmoni), koridor 9 (Pinang Ranti – Pluit) dan koridor 10 (PGC 2 – Tanjung Priok) sedangkan 3 koridor lainnya termasuk ke *cluster 3* (Excel) atau *cluster 0* (Rapidminer) yaitu koridor 11 (Kp. Melayu – Pulo Gebang), koridor 12 (Penjaringan – Sunter Kelapa Gading) dan koridor 13 (Ciledug – Tendea).

4. SIMPULAN

Berdasarkan hasil pengolahan data yang telah dilakukan, maka dapat diambil kesimpulan sebagai berikut:

- Pengelompokan data koridor transjakarta berdasarkan jumlah penumpang dengan algoritma *K-Means cluster* menggunakan *Microsoft Excel* dan *Rapidminer Studio* menghasilkan 3 *cluster* yaitu *cluster 1* dengan jumlah penumpang terbanyak terdapat 1 koridor, *cluster 2* dengan jumlah penumpang sedang terdapat 9 koridor dan *cluster 3* atau 0 dengan jumlah penumpang rendah terdapat 3 koridor.



- b) 1 koridor pada *cluster* 1 dengan jumlah penumpang tertinggi yaitu koridor 1 yang melayani rute Blok M - Kota, hal tersebut menunjukkan bahwa trayek Blok M - Kota paling banyak digunakan oleh penumpang transjakarta, oleh karena itu pemerintah harus dapat meningkatkan jumlah armada pada trayek tersebut untuk mengatasi penumpukan penumpang di halte trayek tersebut.

Untuk penelitian lebih lanjut tentang *cluster* koridor transjakarta berdasarkan jumlah penumpang, maka terdapat beberapa saran yang telah dirangkum yaitu:

- a) Perlu penambahan jumlah dataset dan dilakukannya *preprocessing data* menggunakan teknik yang lain agar data yang diolah dan dihasilkan lebih berkualitas.
- b) Lakukan pengolahan data dengan menggunakan algoritma *clustering* yang lain seperti K-Medoid, Fuzzy, dsb agar mendapatkan perbandingan *performance* model algoritma yang digunakan.
- c) Lakukan pengujian model dengan menggunakan aplikasi lain seperti Phyton, R Programming, dsb.

DAFTAR PUSTAKA

- [1] Wikipedia, "Transjakarta." [Online]. Available: <https://id.wikipedia.org/wiki/Transjakarta>. [Accessed: 25-Jan-2020].
- [2] D. Perhubungan, "Data Jumlah Penumpang Transjakarta," Jakarta, 2018.
- [3] L. Zahrotun, "Analisis Pengelompokan Jumlah Penumpang Bus Trans Jogja Menggunakan Metode *Clustering* K-Means Dan Agglomerative Hierarchical *Clustering* (Ahc)," *J. Inform.*, vol. 9, no. 1, pp. 1039–1047, 2015.
- [4] A. Hazman, F. T. Elektro, and U. Telkom, "Penerapan Metode K-Means *Clustering* Untuk Mengelompokkan Data Pelabuhan Dan Bongkar Muat Barang Di Indonesia," *e-Proceeding Eng.*, vol. 6, no. 1, pp. 1450–1454, 2019.
- [5] M. I. Firdaus, R. D. Octaviani, and I. Yusnita, "Clastering Calon Penumpang Kereta Cepat Jakarta-Bandung," *J. Manaj. Transp. DAN LOGISTIK*, vol. 4, no. 2, p. 193, Sep. 2017.
- [6] S. Sukamto, I. D. Id, and T. R. Angraini, "Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan *Clustering* Algoritma K-Means," *JUITA J. Inform.*, vol. 6, no. 2, p. 137, 2018.
- [7] E. Dianawati, P. P. Yanti, and Y. Suryandari, "Klustering Jumlah Penumpang pada Halte Bus Rapid Transit Kota Tangerang," *J. Sist. Cerdas*, vol. 2, no. 3, pp. 163–172, 2019.
- [8] K. Handoko and L. S. Lesmana, "Data Mining Pada Jumlah Penumpang Menggunakan Metode *Clustering*," *Snistek*, no. 1, pp. 97–102, 2018.
- [9] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA J. Inform.*, vol. 7, no. 2, p. 109, 2019.



- [10] A. Supriyatna, I. Carolina, W. Widiati, and C. Nuraeni, "Rice Productivity Analysis by Province Using K-Means *Cluster* Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 771, p. 012025, 2020.
- [11] G. Qiu, R. Song, S. He, W. Xu, and M. Jiang, "Clustering Passenger Trip Data for the Potential Passenger Investigation and Line Design of Customized Commuter Bus," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3351–3360, 2019.
- [12] V. Adikariwattage, A. G. de Barros, S. C. Wirasinghe, and J. Ruwanpura, "Airport classification criteria based on passenger characteristics and terminal size," *J. Air Transp. Manag.*, vol. 24, pp. 36–41, 2012.
- [13] K. Handoko and L. S. Lesmana, "Pengelompokkan Data Mining Pada Jumlah Penumpang Di Bandara Hang Nadim," *Comput. Based Inf. Syst. J.*, vol. 6, no. 2, p. 60, 2018.