

## Data Mining Pada Jumlah Penumpang Menggunakan Metode *Clustering*

Koko Handoko\*, Lido Sabda Lesmana

Universitas Putera Batam, Batam

\*kokohandoko01@gmail.com

---

### **Abstract**

Currently, the concept of Data Mining is increasingly recognized as an important tool in information management because of the increasing amount of information. This study aims to classify the number of passengers at the airport Hang Nadim Batam. A very large sum of these huge amounts of data wants data for the new knowledge of the data. Data drilling is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify relevant information and large databases. One technique known in data mining is grouping, which is the process of grouping a number of data or objects into a cluster (group). Each of these clusters will contain data that is as similar as possible and different from the objects in the other clusters. In this case we can select the price data as the initial cluster center, then calculate the distance between each data in the cluster center and determine the nearest cluster, then average averages of all groups, so that the existing process is not the same. After the process is done quickly by using RapidMiner, create clusters in grouping the number of passengers. Where the variable used is the first variable, namely: the number of passengers coming, the number of passengers departing, the number of transits that transit. Where will present data on the number of passengers a lot, medium and small.

**Keywords:** *Clustering; Data mining; RapidMiner; Total passenger.*

### **Abstrak**

Saat ini, konsep Data Mining semakin dikenal sebagai *tools* penting dalam manajemen informasi karena jumlah informasi yang semakin besar jumlahnya. Penelitian ini bertujuan untuk mengelompokkan jumlah penumpang di bandara hang nadim batam. Jumlah penumpang yang setiap harinya banyak maka dari itu dari jumlah data yang sangat besar ini ingin menggali data untuk mendapatkan suatu pengetahuan yang baru dari data tersebut. Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari *database* yang besar. Salah satu teknik yang dikenal dalam data mining adalah *clustering*, berupa proses pengelompokan sejumlah data atau objek ke dalam *cluster* (group) sehingga setiap dalam *cluster* tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya. algoritma ini kita terlebih dahulu kita harus memilih data nilai  $k$  sebagai pusat *cluster* awal, kemudian menghitung jarak antara setiap nilai data pada pusat *cluster* dan menetapkan *cluster* terdekat, selanjutnya memperbarui rata-rata dari semua kelompok, ulangi proses ini sampai kriteria tersebut tidak ada perbandingan. Setelah mealkukan proses pngelompokkan dilanjutkan dengan pengujian dilakukan dengan aplikasi *RapidMiner* sehingga menghasilkan *cluster-cluster* dalam mengelompokkan jumlah penumpang. Dimana variabel yang pengujian pertama ditentukan sebanyak tiga variabel, yaitu: jumlah penumpang yang datang, jumlah penumpang yang berangkat, jumlah penumpang yang transit. Di mana akan mempresentasikan data jumlah penumpang banyak, sedang dan sedikit.

**Kata Kunci:** *Clustering; Data mining; Jumlah Penumpang; RapidMiner.*

---

## 1. Pendahuluan

Jumlah penumpang di bandara hang nadim batam merupakan jumlah penumpang yang dikatakan banyak, penumpang setiap harinya ribuan baik penumpang yang berangkat, datang dan transit. Penggalian data dari jumlah yang sangat besar ini dapat dilakukan dengan *Data Mining*. Data jumlah penumpang merupakan data yang sangat besar di bandara hang nadim, oleh karna itu peneliti ingin mengolah data yang besar itu dengan data mining dengan metode *k-means clustering* sehingga mendapatkan suatu pengetahuan baru yang bisa dimanfaatkan.

*Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar. *Clustering* merupakan proses mengelompokkan item data ke dalam sejumlah kecil grup sedemikian rupa sehingga masing-masing grup memiliki suatu persamaan antar satu sama lain. Maka daripada itu, tujuan yang akan dicapai pada penelitian ini adalah untuk mengelompokkan data penumpang dalam pengolahan data yang besar sehingga mendapatkan informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar. Data yang akan digunakan pada penelitian ini adalah data sampel hasil rekap data penumpang pada bandar udara yang terletak di Batam yaitu Bandar Udara Hang Nadim.

## 2. Kajian Literatur

### 2.1 Knowledge Discovery in Database (KDD)

Saat ini, konsep *Data Mining* semakin dikenal sebagai tools penting dalam manajemen informasi karena jumlah informasi yang semakin besar jumlahnya. *Data Mining* sendiri sering disebut sebagai Knowledge Discovery in Database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam set data berukuran besar (Handoko, 2016).

Pada dasarnya ada enam elemen yang paling esensial dalam teknik pencarian informasi / pengetahuan dalam KDD yaitu: (1) Mengerjakan sejumlah besar data; (2) Diperlukan efisiensi berkaitan dengan *volume* data; (3) Mengutamakan ketetapan/keakuratan; (4) Membutuhkan pemakaian bahasa tingkat tinggi; (5) Menggunakan beberapa bentuk dari pembelajaran otomatis; (6) Menghasilkan hasil yang menarik.

### 2.2 Data Mining

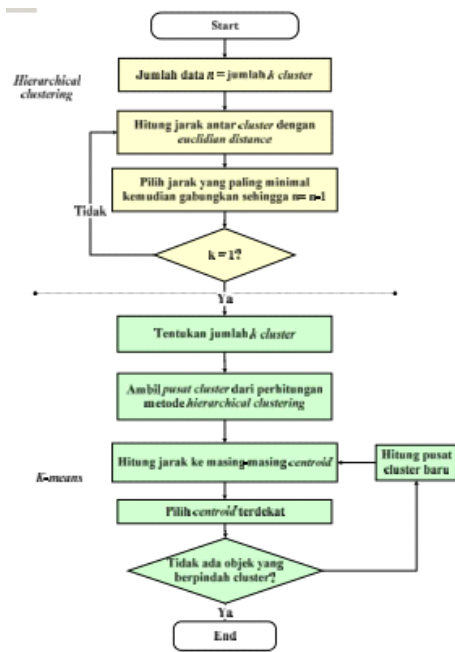
*Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari *database* yang besar. *Data Mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan basis data. Beberapa teknik yang sering disebut-sebut dalam literatur *data mining* antara lain : *clustering*, *classification*, *association rule mining*, *neural network*, dan *genetic algorithm* (Handoko, 2016).

Sampai saat ini, para ilmuwan masih terus melakukan berbagai usaha untuk melakukan perbaikan model *cluster* dan menghitung jumlah *cluster* yang optimal sehingga dapat dihasilkan *cluster* yang paling baik (Alfina, Santosa, & Barakbah, 2012).

Tahap-tahap pada *data mining* dapat diilustrasikan seperti berikut: (Bisilisin, Herdiyeni, & Silalahi, n.d.): (1) Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*); (2) Integrasi data (penggabungan data dari beberapa sumber); (3) Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-*mining*); (4) Aplikasi teknik DM; (4) Evaluasi pola yang ditemukan (untuk menemukan yang menarik/bernilai); (5) Presentasi pengetahuan (dengan teknik visualisasi).

### 2.3 K-means Clustering

Pada tahap ini dilakukan pengelompokan data menggunakan kombinasi dua algoritma *clustering*, yaitu *hierarchical clustering* dan metode *K-means*. Selanjutnya, pusat *cluster* yang diperoleh *hierarchical clustering* tersebut digunakan untuk proses pengelompokan data dengan menggunakan metode *K-means*. Gambar 1 adalah *flowchart* yang menjelaskan urutan pengerjaan penelitian dengan menggunakan metode *hierarchical clustering* dan *K-means* (Alfina et al., 2012). Lakukan langkah tersebut hingga nilai *centroid* tidak berubah (stabil) (Rismawan, 2008):



Gambar 1. Hierarchical clustering dan k-means

*K-Means Clustering* merupakan metode yang termasuk ke dalam golongan algoritma *Partitioning Clustering*. Langkah-langkah dari metode *K-Means* adalah sebagai berikut (Handoko, 2016): (1) Tentukan nilai *k* sebagai jumlah *cluster* yang ingin dibentuk; (2) Bangkitkan *k centroid* (titik pusat cluster) awal secara acak; (3) Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antar dua objek (*Euclidean Distance*); (4) Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya; (5) Tentukan posisi *centroid* baru ( $k C$ ) dengan cara menghitung nilai rata-rata dari data yang ada pada *centroid* yang sama.

Sebelum algoritma *k-means* menyatu, perhitungan jarak dan sentroid dilakukan saat *loop* dieksekusi beberapa kali, katakanlah *l*, di mana bilangan bulat positif *l* dikenal sebagai jumlah iterasi *k-means*. Nilai yang tepat dari saya bervariasi tergantung pada *centroid center* awal bahkan pada dataset yang sama ("Application of *k-Means Clustering* algorithm for prediction of Students' Academic Performance," 2010).

2.4 Terkait Penelitian Terdahulu Yang Dijadikan Acuan

Penelitian yang terkait yang sebelumnya telah dilakukan oleh peneliti lainnya berkaitan pada penggunaan metode *K-Means* untuk melakukan *selection* untuk menentukan pelanggan berpotensi (Putra & Wadisman, 2018). Letak perbedaan penelitian ini dengan penelitian yang terkait terletak pada objek yang diteliti dan juga metode yang digunakan. Pada penelitian yang terkait, objek yang

digunakan adalah data pelanggan pada sebuah *laundry* yang bernama MC Laundry sedangkan pada penelitian ini, objek yang diteliti adalah lalu lintas jumlah penumpang pesawat udara di Bandar Udara Hang Nadim Batam Metode yang digunakan adalah metode *clustering* yang bertujuan untuk mengelompokkan penumpang pesawat udara.

3. Metode Penelitian

3.1 Analisis Data

Analisis data adalah mengelompokkan data berdasarkan variabel dan jenis responden, mentabulasi data berdasarkan variabel dari seluruh responden, menyajikan data tiap variabel yang diteliti, melakukan perhitungan untuk menjawab rumusan masalah, dan melakukan perhitungan untuk menguji hipotesis yang telah diajukan.

Berikut bentuk analisis data yang akan dilakukan dalam penelitian ini.

Input:

- Jumlah penumpang yang datang
- Jumlah penumpang yang berangkat
- Jumlah penumpang yang transit

Proses

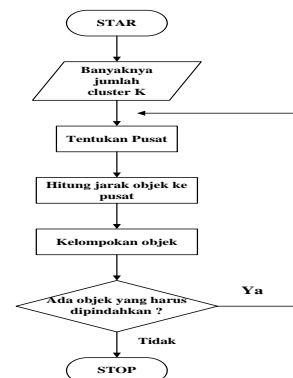
: Metode *Rapid Miner*

Keputusan

: Pengelompokan jumlah penumpang

3.2 Analisa Clustering Dengan Algoritma K-Means

Berikut ini merupakan gambar *flowchart* dari algoritma *K-Means* yang digunakan dan menggambarkan langkah-langkah dalam algoritma *K-Means* dengan anggapan bahwa input adalah jumlah data set sebanyak *n* data dan jumlah inialisasi *centroid*  $K=2$  sesuai dengan penelitian.



Gambar 2. Flowchart Proses K-Means

4. Hasil dan Pembahasan

4.1 Pengumpulan Data

Pada tahap pengumpulan data yang digunakan adalah dengan pengambilan data yang digunakan langsung oleh administrasi pada Bandar Udara Hang Nadim Kota Batam. Data tersebut merupakan data arus lalu lintas

angkutan udara Bandar Udara Hang Nadim dalam kurun waktu tahun 2015 hingga tahun 2017.

#### 4.2 Sampel Data Yang Digunakan

Tahap ini akan menjelaskan sampel data rekapitulasi arus lalu lintas angkutan udara Bandar Udara Hang Nadim Batam yang digunakan untuk proses analisa metode *K-Means*. Seluruh data atribut pada *dataset* tersebut selanjutnya akan diseleksi untuk mendapatkan atribut-atribut yang bernilai relevan, tidak *missing value* dan tidak *redundance*, sehingga data tersebut dapat memenuhi syarat awal yang harus dilakukan dalam *data mining*, dimana diperoleh *dataset* yang bersih untuk digunakan pada tahap *mining*. Atribut pengujian pertama ditentukan sebanyak 3 atribut, yaitu: penumpang berangkat, penumpang datang, dan penumpang transit, sedangkan pada atribut kode diambil dari *no-index* supaya kita dapat mengetahui *no-index* mana saja yang diambil menjadi data sampel. Berikut contoh sampel yang telah diambil:

Tabel 1. Sampel Rekapitulasi Arus Lalu Lintas Angkutan Udara Bandar Udara Hang Nadim – Batam

No	Bulan	Datang	Berangkat	Transit
1	Jan-16	258,573	222,315	28,320
2	Feb-16	229,837	208,857	24,609
3	Mar-16	247,098	216,113	34,879
4	Apr-16	232,176	214,189	23,473
5	May-16	259,825	248,723	26,459
6	Jun-16	198,840	238,925	19,957
7	Jul-16	324,215	286,773	19,741
8	Aug-16	264,791	219,815	25,919
9	Sep-16	245,484	218,048	26,138
10	Oct-16	241,841	207,866	26,107
11	Nov-16	243,331	220,177	34,516
12	Dec-16	290,670	274,730	33,051
13	Jan-17	270,214	269,101	54,937
14	Feb-17	215,501	210,834	56,419
15	Mar-17	241,834	240,471	68,153
16	Apr-17	227,325	238,245	56,779
17	May-17	220,896	237,202	62,176
18	Jun-17	196,315	250,074	44,076
19	Jul-17	293,348	265,493	56,313
20	Aug-17	242,978	251,641	50,258
21	Sep-17	254,795	208,632	54,312
22	Oct-17	240,473	204,826	51,562
23	Nov-17	220,241	194,607	45,721
24	Dec-17	257,948	234,201	38,882

#### 4.3 Analisa Proses Algoritma K-Means

Pada tahap ini akan digunakan algoritma *K-Means* untuk mengelompokkan data yang ada. Untuk pengujian pertama, data yang ada akan dikelompokkan menjadi 3 kelompok sebagian yang telah dijelaskan pada sub bab sebelumnya. Adapun prosesnya adalah sebagai berikut: (1) Menentukan jumlah *cluster*. Sebagaimana telah dijelaskan pada sub bab sebelumnya bahwa untuk mengelompokkan data pada pengujian yang pertama ini adalah sebanyak 3 *cluster*, sehingga dapat ditentukan untuk nilai *k* adalah *k=3*. (2) Menentukan titik pusat *cluster*. Menentukan *centroid* awal dilakukan secara acak dari data/objek yang tersedia sebanyak jumlah *cluster k*. Nilai *centroid* awal pada penelitian ini dilakukan pemilihan secara acak, dimana jumlah *centroid* awal dilakukan sebanyak tiga *centroid* awal, nilai untuk C1 diambil dari baris data ke-12, nilai C2 diambil dari baris ke-15, nilai C3 diambil dari baris data ke-9. Berikut ini nilai *centroid* awal pada penelitian:

$$C1 = (290.670 ; 274.730 ; 33.051)$$

$$C2 = (241.834 ; 240.471 ; 68.153)$$

$$C3 = (245.484 ; 218.048 ; 26.138)$$

(3) Hitung jarak setiap data yang ada terhadap setiap pusat *cluster*. Untuk menghitung jarak setiap data yang ada terhadap pusat *cluster* terdapat beberapa cara, yaitu dengan menggunakan rumus *Manhattan/City Block*, dan *Euclidean Distance*. Sedangkan dalam penelitian ini penulis menggunakan rumus *Euclidean Distance* untuk melakukan perhitungan jarak setiap data terhadap titik pusat *cluster*. Berikut ini adalah contoh perhitungan jarak dengan *Euclidean Distance* untuk iterasi 0.

Perhitungan jarak dari data ke-1 terhadap pusat *cluster*:

$$A(C1,A1) =$$

$$\sqrt{(290.670-258.573)^2 + (274.730-222.315)^2 + (33.051-28.320)^2}$$

$$= 61.644$$

$$A(C2,A1) =$$

$$\sqrt{(241.834-258.573)^2 + (240.471-222.315)^2 + (68.153-28.320)^2}$$

$$= 48.867$$

$$A(C3,A1) =$$

$$\sqrt{(245.484-258.573)^2 + (218.048-222.315)^2 + (26.138-28.320)^2}$$

$$= 13.939$$

(4) Alokasikan masing-masing data ke dalam *centroid* yang paling terdekat.

Setelah dilakukan proses perhitungan maka akan didapatkan data selengkapnya adalah:

Tabel 2. Perhitungan Jarak Data Dengan Centroid Iterasi -0

	C1	C2	C3	C1	C2	C3
61.644	46.867	13.939	0	0	1	
90.062	55.131	18.211	0	0	1	
73.060	41.571	9.097	0	0	1	
84.726	52.729	14.110	0	0	1	
40.881	46.154	33.863	0	0	1	
99.429	64.604	51.475	0	0	1	
38.045	106.180	104.703	1	0	0	
61.125	52.320	19.389	0	0	1	
72.818	47.764	0	0	0	1	
83.086	53.207	10.814	0	0	1	
72.244	39.313	8.908	0	0	1	
0	69.215	72.818	1	0	0	
30.482	42.424	63.619	1	0	0	
101.386	41.346	43.220	0	1	0	
69.215	0	47.764	0	1	0	
76.855	18.570	40.946	0	1	0	
84.410	22.018	47.646	0	1	0	
98.144	52.382	61.360	0	1	0	
25.172	58.481	73.841	1	0	0	
55.711	21.126	41.431	0	1	0	
78.154	37.058	31.131	0	0	1	
88.028	39.341	29.091	0	0	1	
107.427	55.434	39.626	0	0	1	
52.415	33.997	24.056	0	0	1	

(5) Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan cara menghitung rata-rata dari data-data yang berada pada *centroid* yang sama.

- Cluster pertama = 7, 12, 13, 19
- Cluster kedua = 14, 15, 16, 17, 18, 20
- Cluster ketiga = 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 21, 22, 23, 24

Perhitungan pusat *cluster* baru yang dapat diperoleh dari hasil *cluster* di atas adalah:

$$C1,1 = (324.215+290.670+270.214+293.348) / 4 = 294.612$$

$$C1,2 = (286.773+274.730+269.101+265.493) / 4 = 274.024$$

$$C1,3 = (19.741+33.051+54.937+56.313) / 4 = 41.011$$

Dari perhitungan pusat *cluster* di atas dapat diperoleh *centroid* baru untuk iterasi - 1 ini sebagai berikut:

Tabel 3. Cluster Baru Iterasi ke-0

C1	294.612	274.024	41.011
C2	224.142	238.078	56.310
C3	242.518	218.378	32.918

(6) Ulangi langkah ke-3 hingga nilai *centroid* tidak mengalami perubahan. Karena proses yang dilakukan baru pada iterasi ke-0, maka perlu dilakukan beberapa iterasi lagi untuk dapat membandingkan nilai dari dua iterasi terakhir. Jika nilai dua iterasi terakhir tersebut telah sama, maka proses iterasi telah selesai, dan jika tidak maka ulangi lagi langkah berikutnya.

Tabel 4. Hasil Perhitungan Jarak Dan Pengelompokan Data Iterasi ke-2

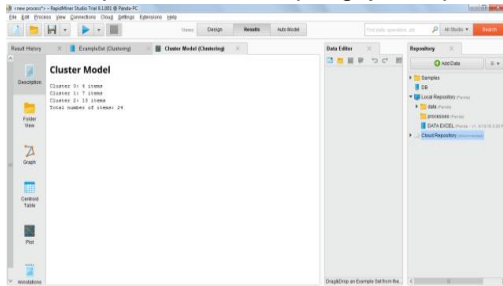
	C1	C2	C3	C1	C2	C3
64.294	47.111	14.930	0	0	1	
93.336	40.624	20.173	0	0	1	
75.159	38.177	1.699	0	0	1	
88.239	38.423	17.424	0	0	1	
45.409	47.572	35.628	0	0	1	
104.151	37.971	53.824	0	1	0	
38.617	118.723	105.991	1	0	0	
63.684	54.150	20.754	0	0	1	
75.948	40.655	7.887	0	0	1	
82.929	44.720	12.531	0	0	1	
74.642	33.472	4.274	0	0	1	
8.910	81.123	73.234	1	0	0	
28.521	58.637	61.399	1	0	0	
102.416	28.323	38.272	0	1	0	
68.176	27.375	41.821	0	1	0	
77.822	8.847	36.427	0	1	0	
85.076	11.110	42.885	0	1	0	
101.219	27.871	60.556	0	1	0	
17.565	77.941	71.598	1	0	0	
57.031	26.182	38.595	0	1	0	
77.707	45.373	23.711	0	0	1	
88.491	38.882	21.999	0	0	1	
108.905	43.925	35.903	0	0	1	
54.172	39.573	21.754	0	0	1	

Dapat terlihat pada iterasi ke-1 dan iterasi ke-2 tidak lagi mengalami perubahan pada titik *cluster*, sehingga dapat disimpulkan bahwa iterasi dapat dihentikan pada iterasi ke-2 dengan hasil:

- Cluster pertama = 4 item
- Cluster kedua = 7 item
- Cluster ketiga = 13 item

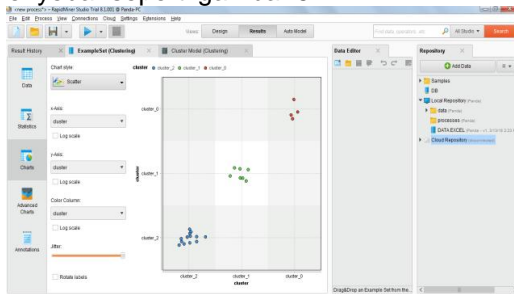
#### 4.4 Pengujian Dengan Aplikasi RapidMiner

Penelitian ini dilakukan dengan menggunakan data yang nantinya akan menghasilkan *cluster* atau pengelompokan data penumpang yang akan diuji kebenarannya. Aplikasi yang digunakan dalam pengujian *clustering* dengan algoritma *K-Means* adalah aplikasi *RapidMiner*. Gambar ini merupakan hasil dari pengujian *rapidMiner*.



Gambar 3. Tampilan *Cluster Model (Clustering) Text View*

Hasil *clustering* juga dapat dilihat dalam format *charts* yang berupa titik-titik yang menyebar seperti gambar 5.



Gambar 4. Tampilan Hasil *Clustering*

Hasil perhitungan dari algoritma *k-means clustering* dengan aplikasi *RapidMiner* adalah sama.

## 5. Kesimpulan dan Saran

### Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah diuraikan, maka penulis menarik kesimpulan adalah sebagai berikut: (1) Hasil *clustering* memiliki sifat yang hampir sama antara data-data yang berada dalam satu *cluster*, yaitu jumlah penumpang secara total perbulannya, sehingga dengan melakukan karakteristik *clustering*, pengetahuan mengenai jadwal padatnya penumpang perbulannya dapat membantu bandar udara untuk mengantisipasi padatnya penumpang pada bulan tertentu; (2) Hasil ekstraksi pengetahuan dari *data mining* yang dilakukan adalah mengetahui jadwal padatnya lalu lintas penumpang bandar udara hang nadim diamana pengujian manual dengan 24 sampel data penumpang dan *software RapidMiner*, hasil *cluster* pertama terdiri dari 4 bulan data penumpang yang datang, *cluster*

kedua terdiri dari 7 bulan data penumpang yang berangkat, dan *cluster* ketiga terdiri dari 13 bulan data penumpang yang transit. Sehingga *software RapidMiner* ini sangat mempermudah pengelompokan pelanggan dengan menggunakan data sedikit maupun data yang lebih banyak dari sebelumnya.

### Saran

Berdasarkan hasil penelitian dan kesimpulan, maka dapat diambil saran sebagai berikut: (1) Untuk metode *K-Means* dapat dikembangkan lebih lanjut dengan data dan *cluster* yang lebih banyak; (2) Pada penelitian selanjutnya bisa menerapkan algoritma-algoritma lainnya dan *software-software* yang digunakan untuk metode *clustering*.

### Ucapan Terima Kasih

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu baik secara langsung ataupun tidak secara langsung dan yang telah memberikan dukungan terhadap penyusunan penelitian ini.

### Daftar Pustaka

- Alfina, T., Santosa, B., & Barakbah, R. (2012). Analisa Perbandingan Metode Hierarchical Clustering, K-Means Dan Gabungan Keduanya Dalam Cluster Data ( Studi Kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS ), 1.
- Application Of K-Means Clustering Algorithm For Prediction Of Students ' Academic Performance. (2010), 7, 292–295.
- Bisilisin, F. Y., Herdiyeni, Y., & Silalahi, B. I. B. P. (N.D.). Optimasi K-Means Clustering Menggunakan Particle Swarm Optimization Pada Sistem Identifikasi Tumbuhan Obat Berbasis Citra K-Means Clustering Optimization Using Particle Swarm Optimization On Image Based Medicinal Plant Identification System, 3(2002).
- Handoko, K. (2016). PENERAPAN DATA MINING DALAM MENINGKATKAN MUTU PEMBELAJARAN PADA INSTANSI PERGURUAN TINGGI MENGGUNAKAN METODE K-MEANS CLUSTERING ( STUDI KASUS DI PROGRAM STUDI TKJ AKADEMI KOMUNITAS SOLOK SELATAN ), 2(3).
- Putra, R. R., & Wadisman, C. (2018). IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA K-MEANS, 1(1), 72–77.
- Rismawan, T. (2008). APLIKASI K-MEANS UNTUK PENGELOMPOKKAN MAHASISWA, 2008(Snati).