



Computer Based Information System Journal

ISSN (Print) : 2337-8794 | E- ISSN : 2621-5292
 web jurnal : <http://ejournal.upbatam.ac.id/index.php/cbis>



PENGELOMPOKKAN *DATA MINING* PADA JUMLAH PENUMPANG DI BANDARA HANG NADIM

Koko Handoko, Lido Sabda Lesmana

Universitas Putera Batam, Indonesia.

INFORMASI ARTIKEL

KATA KUNCI

Data Mining, K-Means, Clustering,
Data Penumpang

KORESPONDENSI

E-mail:
kokohandoko263@gmail.com

ABSTRACT

The concept of data mining becomes one of the important tools in information management because the existing information has an increasing number. Data mining has many techniques in practice, one of which is the clustering technique which is the process of grouping data into groups so that data exist in the same group have properties as closely as possible. Clustering has many different methods, one of which is K-Means. By using data mining clustering on traffic activity data taken from Hang Nadim Airport Batam, it can be obtained by grouping passenger based on clusters according to the nature of each data. The data taken include the number of passengers coming, departing, and transiting. In the process of performing data mining clustering, existing sample data must go through several important stages in order to get the correct cluster results. Stages that must be passed the Stages of Data Processing, Clustering Stage and Stage Algorithm. Based on the results of research that has been done on the existing sample data, it can be concluded the results of data grouping of passengers at Hang Nadim Airport Batam.

I. Latar Belakang

Banyaknya data yang dipunyai oleh sebuah bandara penerbangan dapat menyebabkan kesulitan dalam pengklasifikasian data tersebut. Kegiatan pengklasifikasian yang dilakukan oleh manusia masih terbatas, terutama pada kemampuan manusia dalam menampung jumlah data yang ingin diklasifikasikan. Selain itu bisa juga sering terjadi kesalahan dalam pengklasifikasian yang dilakukan atau sering disebut dengan *human errors*. Penggalan data dari jumlah yang sangat besar ini dapat dilakukan dengan *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. Metode K-means merupakan metode clustering yang paling sederhana dan umum. Hal ini dikarenakan K-

means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien. Namun, metode *K-Means* memiliki kelemahan yang diakibatkan oleh penentuan pusat awal *cluster*. (Handoko, 2016).

Clustering merupakan proses mengelompokkan item data ke dalam sejumlah kecil grup sedemikian rupa sehingga masing-masing grup memiliki suatu persamaan antar satu sama lain [2].

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.

II. Kajian Literatur

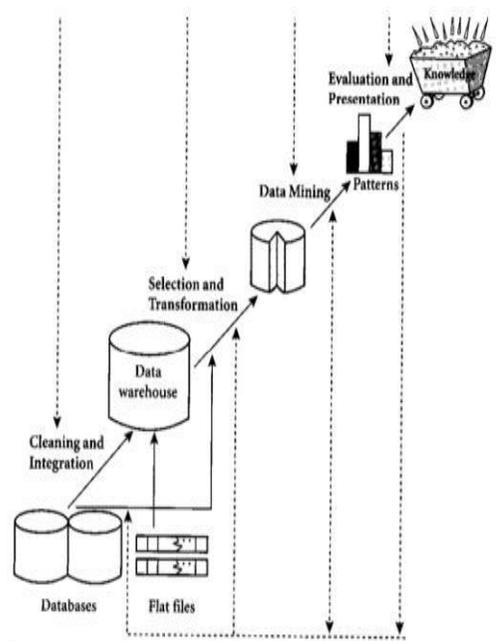
A. Knowledge Discovery in Database (KDD)

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 [1].

Ada beberapa macam pendekatan yang berbeda yang diklasifikasikan sebagai teknik pencarian informasi / pengetahuan dalam KDD. Ada pendekatan kuantitatif, seperti pendekatan probabilistik seperti logika induktif, pencarian pola, dan analisis pohon keputusan.

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.



Gambar 1. Tahapan pada KDD

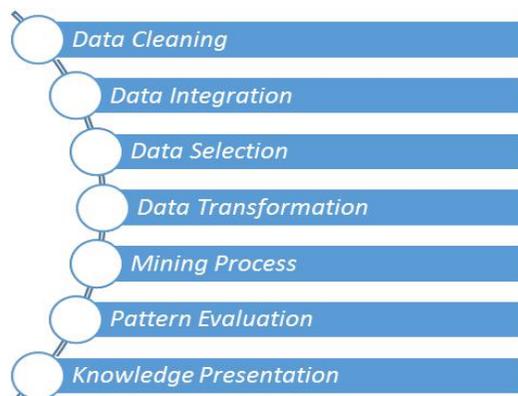
B. Data Mining

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* [1].

Data Mining adalah suatu rangkaian proses yang akan menghasilkan keluaran / *output* berupa sebuah pengetahuan baru yang diharapkan dapat dimanfaatkan di masa depan. Tahap-tahap pada *data mining* dapat diilustrasikan seperti berikut : [4]:

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
5. Evaluasi pola yang ditemukan (untuk menemukan yang menarik/bernilai)
6. Presentasi pengetahuan (dengan teknik visualisasi)

Jika dilihat pada gambar 1 tahapan yang ada pada KDD, dapat dilihat *data mining* merupakan salah satu tahapan didalamnya. Namun, dalam tahapan *data mining*, masih terbagi ke dalam beberapa tahapan yang harus dilakukan, yakni sebagai berikut : [5]



Gambar 2. Tahapan pada *Data Mining*

- *Data Cleaning* (Pembersihan Data)

Data cleaning atau pembersihan data merupakan proses menghilangkan data yang tidak relevan dalam sebuah *dataset*. Data yang tidak relevan seperti data hilang, data tidak valid dan data yang salah ketik akan dibuang agar performansi *data mining* menjadi lebih ringan karena jumlah yang akan diproses berkurang dan kompleksitas data menjadi lebih gampang.

- **Data Integration (Integrasi Data)**
Data integration atau integrasi data merupakan hasil persatuan dari beberapa data dari berbagai *database* yang dimasukkan ke dalam *database* yang baru. Integrasi harus dilakukan dengan teliti karena kesalahan pada tahap ini dapat menyekat dalam pengambilan aksi selanjutnya.
- **Data Selection (Seleksi Data)**
Data selection atau seleksi data adalah tahap pemilihan data yang akan digunakan karena tidak semua data pada *database* akan digunakan dalam penelitian. Hanya data yang sesuai dengan kriteria analisis yang akan diambil dari *database* untuk diteliti.
- **Data Transformation (Transformasi Data)**
Data transformation atau transformasi data adalah tahap perubahan format data ke dalam format yang cocok untuk diproses. Beberapa metode *data mining* membutuhkan jenis format data yang berbeda. Seperti pada *clustering*, data input harus dalam format data kategorial agar dapat diproses, sehingga pada tahap inilah proses perubahan format data tersebut dilakukan.
- **Mining Process (Proses mining)**
Mining process atau proses mining merupakan tahapan proses yang utama. Semua tahapan sebelumnya adalah untuk mendukung tahapan proses ini. Tahapan ini adalah proses menggunakan metode dan algoritma yang ada untuk menemukan pengetahuan dari data yang ada.
- **Pattern Evaluation (Evaluasi Pola)**
Pattern evaluation atau evaluasi pola adalah tahapan dalam mengidentifikasi pola – pola yang ada berdasarkan hasil

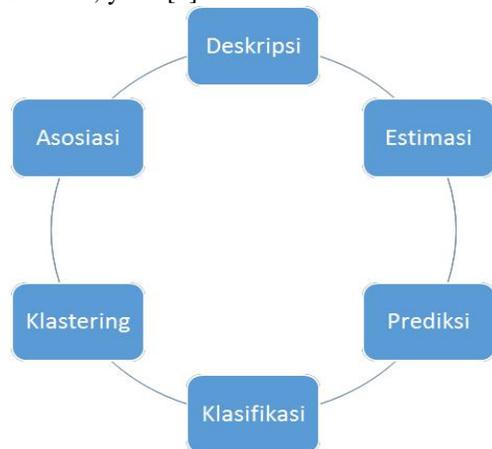
dari proses mining yang telah dilakukan dan menarik kesimpulan berupa pengetahuan untuk menilai apakah hipotesis yang ada telah terpenuhi atau belum.

- **Knowledge Presentation (Presentasi Pengetahuan)**

Knowledge presentation atau presentasi pengetahuan adalah tahapan visualisasi atau penyajian hasil pengetahuan yang telah didapat ke dalam deskripsi agar dipahami oleh semua orang.

C. Pengelompokan Data Mining

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu [6]



Gambar 3. Pengelompokan Data Mining

- **Deskripsi**

Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelesaian untuk suatu pola atau kecenderungan.

- **Estimasi**

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik dari pada kearah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit

berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

- **Prediksi**

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

- **Klasifikasi**

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

- **Klastering**

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (*homogeny*), yang mana kemiripan dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

- **Asosiasi**

Tugas asosiasi dalam *Data Mining* adalah menemukan *attribut* yang muncul dalam satu waktu. Dalam dunia

bisnis lebih umum disebut analisis keranjang belanja.

D. *K-means Clustering*

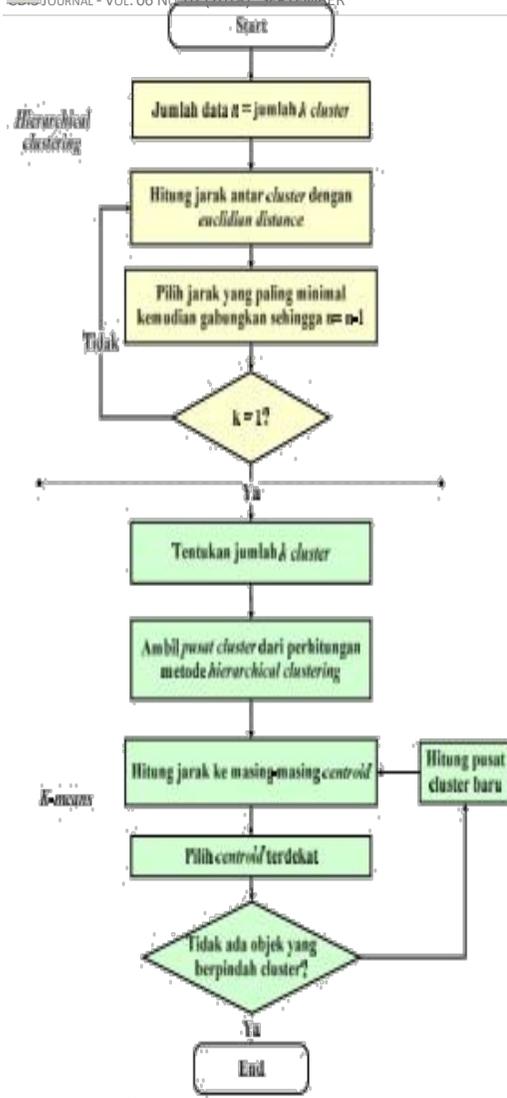
Pada penelitian ini, *data mining* yang dilakukan adalah *clustering* dengan menggunakan gabungan antara metode *K-means* dan *hierarchical clustering*. Dari algoritma *hierarchical clustering* ini digunakan untuk menentukan pusat cluster. Selanjutnya, pusat *cluster* yang diperoleh *hierarchical clustering* tersebut digunakan untuk proses pengelompokan data dengan menggunakan metode *K-means*.

Pada dasarnya *clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*) serta tidak memerlukan target output.

Hierarchical clustering adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga *cluster* akan membentuk semacam pohon di mana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah *cluster*.

Hierarchical clustering adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga *cluster* akan membentuk semacam pohon di mana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah *cluster*.

Berikut langkah – langkah agar nilai *centroid* tidak berubah [7]:



Gambar 4. Langkah mencari nilai centroid

K-Means Clustering merupakan metode yang termasuk ke dalam golongan algoritma *Partitioning Clustering*. Langkah-langkah dari metode *K-Means* adalah sebagai berikut:

1. Tentukan *k* sebagai jumlah cluster yang di bentuk.
Untuk menentukan banyaknya cluster *k* dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak cluster.
2. Bangkitkan *k Centroid* (titik pusat cluster) awal secara *random*.
Penentuan *centroid* awal dilakukan secara *random/acak* dari objek-objek yang tersedia sebanyak *k cluster*, kemudian untuk menghitung *centroid cluster ke-i*

berikutnya, digunakan rumus sebagai berikut :

$$v = \frac{\sum_{i=1}^n x_i}{n} \quad ; i = 1, 2, 3, \dots, n$$

(1)

3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing cluster.
Untuk menghitung jarak antara objek dengan *centroid* penulis menggunakan *Euclidian Distance*.

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; i = 1, 2, 3, \dots, n$$

(2)

4. Alokasikan masing-masing objek ke dalam *centroid* yang paling terdekat. Untuk melakukan pengalokasian objek ke dalam masing-masing cluster pada saat iterasi secara umum dapat dilakukan dengan dua cara yaitu dengan *hard k-means*, dimana secara tegas setiap objek dinyatakan sebagai anggota cluster dengan mengukur jarak kedekatan sifatnya terhadap titik pusat cluster tersebut, cara lain dapat dilakukan dengan *fuzzy C-Means*.
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan.
6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama.

Pengecekan *konvergensi* dilakukan dengan membandingkan matriks *group assignment* pada iterasi sebelumnya dengan matrik *group assignment* pada iterasi yang sedang berjalan. Jika hasilnya sama maka *algoritma k-means cluster analysis* sudah *konvergen*, tetapi jika berbeda maka belum *konvergen* sehingga perlu dilakukan iterasi berikutnya.

III. Metodologi

Dalam penelitian ini, data yang digunakan adalah hasil rekap data aktifitas lalu lintas penumpang pada Bandar Udara Hang Nadim Batam. Data tersebut terdiri dari data penumpang

yang datang, penumpang berangkat dan penumpang transit.

A. Variabel

Dengan menggunakan data tersebut, akan didapatkan variabel penelitian. Variabel penelitian yang digunakan adalah sebagai berikut:

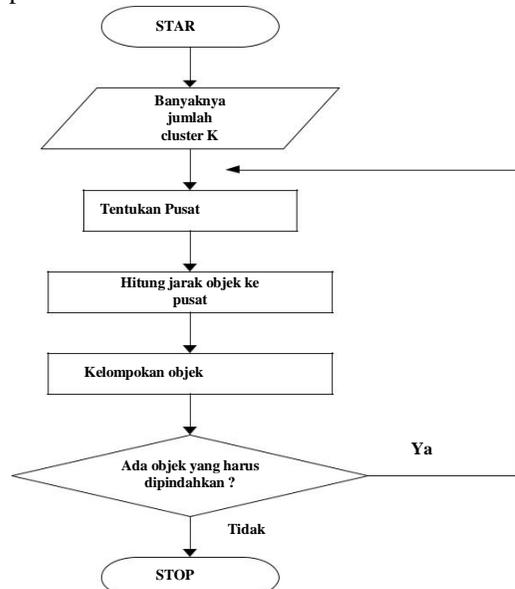
1. Variabel Input
 - a. Jumlah penumpang yang datang
 - b. Jumlah penumpang yang berangkat
 - c. Jumlah penumpang yang transit
2. Variabel Output : Pengelompokan jumlah penumpang

IV. Pembahasan

A. Analisa Clustering dengan Algoritma K-Means

K-Means merupakan salah satu dalam metode data mining partitioning clustering yaitu dimana setiap data harus masuk dalam cluster tertentu dan memungkinkan untuk setiap data yang termasuk dalam cluster tertentu melalui suatu tahapan proses, yang pada tahap berikutnya berpindah ke cluster lainnya. K-Means memisahkan data ke K daerah bagian terkenal dan kemudian mampu melakukan klasifikasi data besar dan outlier dengan sangat cepat.

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 5. Flowchart Algoritma K-Means

Penjelasan dari gambar 5 yaitu :

1. Penentuan Jumlah Cluster
 Penentuan jumlah cluster berdasarkan data yang kita peroleh akan menjadi langkah pertama dalam melakukan metode algoritma ini.
2. Penentuan Pusat Cluster Awal
 Dalam menentukan n buah pusat cluster awal dilakukan pembangkitan bilangan random yang merepresentasikan urutan data input. Pusat awal cluster didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan merandom pusat awal dari data.
3. Perhitungan Jarak Objek ke Pusat Cluster
 Untuk mengukur jarak antara data dengan pusat cluster digunakan Euclidean Distance yaitu algoritma perhiungan jarak data dengan pusat cluster :
 1. Ambil nilai data dan nilai pusat cluster
 2. Hitung Euclidean Distance data dengan tiap pusat cluster.
4. Pengelompokan Objek Data
 Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat cluster, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat cluster terdekat.
 Adapun algoritma pengelompokan data adalah sebagai berikut :
 1. Ambil nilai jarak tiap pusat cluster dengan data.
 2. Cari nilai jarak terkecil.
 3. Kelompokkan data dengan pusat cluster yang memiliki jarak terkecil.
5. Penentuan Pusat Cluster Baru
 Untuk mendapatkan pusat cluster baru bisa dihitung dari rata-rata nilai anggota cluster dan pusat cluster. Pusat cluster yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen, proses iterasi akan berhenti jika telah memenuhi maksimum iterasi yang dimasukkan oleh user atau hasil yang dicapai sudah konvergen (pusat cluster baru sama dengan pusat cluster lama).
6. Perhitungan Jarak Pusat Cluster
 Hitung Euclidean Distance dari semua data ketitik pusat yang baru (C1 dan C2) seperti yang telah dilakukan pada tahap 2. Setelah hasil perhitungan kita

dapatkan, kemudian bandingkan hasil tersebut.

B. Tahap Pengolahan Data

Data yang digunakan dalam penelitian ini adalah data yang bertipe *record* dengan 36 *record*. Data yang digunakan dalam penelitian ini adalah data Rekapitulasi Arus Lalu Lintas Angkutan Udara Bandar Udara Hang Nadim Batam. Berikut ini adalah beberapa rekapitulasi data yang diambil :

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

N O	BULAN	DATAN	BERANGKA	TRANSI	
		G	T	T	
1	JAN	15	199,703	182,996	9,137
2	FEB	15	177,622	167,652	9,720
3	MAR	15	186,019	174,511	9,475
4	APR	15	187,733	176,805	12,929
5	MEI	15	200,033	198,265	15,229
6	JUNI	15	192,572	200,533	16,629
7	JULI	15	237,443	260,933	18,154
	AGTS				
8		15	261,001	188,385	21,583
9	SEPT	15	179,800	168,567	12,477
10	OKT	15	211,832	175,035	16,799
11	NOV	15	212,929	193,730	17,688
12	DES	15	266,438	260,608	19,820
13	JAN	16	258,573	222,315	28,320
14	FEB	16	229,837	208,857	24,609
15	MAR	16	247,098	216,113	34,879
16	APR	16	232,176	214,189	23,473
17	MEI	16	259,825	248,723	26,459
18	JUNI	16	198,840	238,925	19,957
19	JULI	16	324,215	286,773	19,741
	AGTS				
20		16	264,791	219,815	25,919
21	SEPT	16	245,484	218,048	26,138
22	OKT	16	241,841	207,866	26,107
23	NOV	16	243,331	220,177	34,516
24	DES	16	290,670	274,730	33,051
25	JAN	17	270,214	269,101	54,937
26	FEB	17	215,501	210,834	56,419
27	MAR	17	241,834	240,471	68,153
28	APR	17	227,325	238,245	56,779
29	MEI	17	220,896	237,202	62,176

30	JUNI	17	196,315	250,074	44,076
31	JULI	17	293,348	265,493	56,313
	AGTS				
32		17	242,978	251,641	50,258
33	SEPT	17	254,795	208,632	54,312
34	OKT	17	240,473	204,826	51,562
35	NOV	17	220,241	194,607	45,721
36	DES	17	257,948	234,201	38,882

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. Sampel data yang digunakan pada penelitian ini adalah 24 *record*, yaitu :

Tabel 2. Sampel Rekapitulasi Arus Lalu Lintas Angkutan Udara Bandar Udara Hang Nadim – Batam

N O	BULAN	DATAN	BERANGKA	TRANSI	
		G	T	T	
1	JAN	17	270,214	269,101	54,937
2	FEB	17	215,501	210,834	56,419
3	MAR	17	241,834	240,471	68,153
4	APR	17	227,325	238,245	56,779
5	MEI	17	220,896	237,202	62,176
6	JUNI	17	196,315	250,074	44,076
7	JULI	17	293,348	265,493	56,313
	AGTS				
8		17	242,978	251,641	50,258
9	SEPT	17	254,795	208,632	54,312
10	OKT	17	240,473	204,826	51,562
11	NOV	17	220,241	194,607	45,721
12	DES	17	257,948	234,201	38,882
13	JAN	16	258,573	222,315	28,320
14	FEB	16	229,837	208,857	24,609
15	MAR	16	247,098	216,113	34,879
16	APR	16	232,176	214,189	23,473
17	MEI	16	259,825	248,723	26,459

18	JUNI ,,16	198,840	238,925	19,957
19	JULI ,,16	324,215	286,773	19,741
20	AGTS ,,16	264,791	219,815	25,919
21	SEPT ,,16	245,484	218,048	26,138
22	OKT ,,16	241,841	207,866	26,107
23	NOV ,,16	243,331	220,177	34,516
24	DES ,,16	290,670	274,730	33,051

a. Analisa Pengelompokan Data

Dalam menentukan pengelompokan data atau banyaknya jumlah *cluster* yang akan dibuat, dapat dilakukan dengan beberapa pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan beberapa banyak *cluster* yang harus dibentuk. Pada penelitian ini penentuan jumlah kelompok atau *cluster* ditentukan berdasarkan pada pengelompokan data rekapitulasi arus lalu lintas angkutan udara Bandar Udara Hang Nadim Batam. Jadi pada penelitian ini *cluster* yang akan dibentuk adalah sebanyak tiga kelompok atau nilai $k=3$. Di mana atribut yang digunakan adalah sebanyak 3 buah atribut.

b. 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 :

- 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 $k=3$
- Menentukan titik pusat *cluster*. Menentukan *centroid* awal dilakukan secara acak dari data/objek yang tersedia sebanyak jumlah *cluster* k .
- 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*.

- Alokasikan masing-masing data ke dalam *centroid* yang paling terdekat.
Dalam mengalokasikan kembali objek ke dalam masing-masing *cluster* didasarkan pada perbandingan jarak antara data dengan *centroid* setiap *cluster* yang ada, objek dialokasikan secara tegas ke dalam *cluster* yang mempunyai jarak ke *centroid* terdekat dengan data tersebut.
- Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan cara menghitung rata-rata dari data-data yang berada pada *centroid* yang sama.
- Ulangi langkah ke-3 hingga nilai *centroid* tidak mengalami perubahan.
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.

c. Tahap Hasil Perhitungan

Setelah melakukan langkah-langkah yang telah diuraikan pada sub bab sebelumnya, maka dari data yang ada, akan mendapatkan iterasi terakhir sebagai hasil seperti berikut :

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

Dengan hasil perhitungan iterasi terakhir seperti pada tabel di bawah ini :

Tabel 3. 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

V. Kesimpulan

Kesimpulan yang dapat diambil pengelompokan data jumlah penumpang pada Bandar Udara Hang Nadim Batam dengan menggunakan *data mining clustering* metode *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. Dengan memanfaatkan aplikasi pendukung *data mining* seperti WEKA, *RapidMiner*, atau Tanagra, hasil dapat dibuktikan kebenarannya.

Ucapan Terima Kasih

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

Daftar Pustaka

[1] K. Handoko, “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),” vol. 2, no. 3, 2016.

[2] R. D. Ramadhani, “Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi,” *Ind. Mark. Manag.*, vol. 1, no. 1, pp. 1–9, 2014.

[3] R. R. Putra and C. Wadisman, “IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA K-MEANS,” vol. 1, no. 1, pp. 72–77, 2018.

[4] F. Y. Bisilisin, Y. Herdiyeni, and B. I. B. P. Silalahi, “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,” vol. 3, no. 2002.

[5] Y. Asriningtias *et al.*, “APLIKASI DATA MINING UNTUK MENAMPILKAN INFORMASI,” vol. 8, no. 1, pp. 837–848, 2014.

[6] F. Sebagai and I. Enzim, “Universitas Sumatera Utara,” 2014.

[7] T. Rismawan, “APLIKASI K-MEANS UNTUK PENGELOMPOKKAN MAHASISWA,” vol. 2008, no. Snati, 2008.