

***DATA MINING METODE K-MEANS CLUSTERING***  
**DALAM PENGELOMPOKAN JUMLAH**  
**BAGASI DAN BARANG KARGO**

SKRIPSI



Oleh:  
Lastari Indah Yani  
150210144

**PROGRAM STUDI TEKNIK INFORMATIKA**  
**FAKULTAS TEKNIK DAN KOMPUTER**  
**UNIVERSITAS PUTERA BATAM**  
**TAHUN 2019**

***DATA MINING METODE K-MEANS CLUSTERING***  
**DALAM PENGELOMPOKAN JUMLAH**  
**BAGASI DAN BARANG KARGO**

**SKRIPSI**

**Untuk memenuhi salah satu syarat  
guna memperoleh gelar Sarjana**



**Oleh:**  
**Lastari Indah Yani**  
**150210144**

**PROGRAM STUDI TEKNIK INFORMATIKA**  
**FAKULTAS TEKNIK DAN KOMPUTER**  
**UNIVERSITAS PUTERA BATAM**  
**TAHUN 2019**

## PERNYATAAN

Dengan ini saya menyatakan bahwa :

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana, dan/atau magister), baik di Universitas Putera Batam maupun di perguruan tinggi lain.
2. Skripsi ini adalah murni gagasan, rumusan dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arah pembimbing.
3. Dalam skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di perguruan tinggi.

Batam, 7 Agustus 2019

Yang membuat pernyataan,

Materai 6000

Lastari Indah Yani  
150210144

***DATA MINING METODE K-MEANS CLUSTERING***  
**DALAM PENGELOMPOKAN JUMLAH**  
**BAGASI DAN BARANG KARGO**

Oleh:  
**Lastari Indah Yani**  
**150210144**

**SKRIPSI**  
Untuk memenuhi salah satu syarat  
guna memperoleh gelar Sarjana

Telah disetujui oleh Pembimbing pada tanggal  
seperti tertera dibawah ini

**Batam, 7 Agustus 2019**

**Sunarsan Sitohang, S.Kom., M.TI.**  
Pembimbing

## ABSTRAK

Transportasi sudah menjadi bagian dari kehidupan manusia dalam melakukan kesehariannya. Tidak hanya manusia saja yang berpindah tempat dengan memanfaatkan transportasi, melainkan barang pengiriman juga dikirimkan dengan memanfaatkan transportasi yang ada dalam kehidupan kita. Pengiriman barang atau kargo sering memanfaatkan transportasi udara berupa pesawat disebabkan kecepatan pengiriman. Di kota Batam, pengiriman barang dengan menggunakan transportasi udara sangat sering dilakukan dan dalam jumlah yang banyak setiap harinya. Banyaknya pengiriman yang dilakukan menghasilkan data yang sangat banyak yang bisa memperlambat pekerjaan dalam pendataan tersebut. Dengan memanfaatkan *data mining* pada pendataan pengiriman barang, diharapkan dapat membantu pihak bandar udara yang di kota Batam bernama Bandar Udara Hang Nadim. Salah satu algoritma yang dapat digunakan untuk melakukan *data mining* adalah penggunaan *K-means Clustering*. *K-means clustering* adalah salah satu cara data mining yang bertujuan untuk mengelompokkan data ke dalam kelompok-kelompok data yang memiliki sifat yang berdekatan. Dari Penelitian ini didapatkan 3 cluster, dimana cluster tersebut dibagi berdasarkan jumlah rata-rata berat dari kargo dari tahun 2015-2017 dari bandara Hang Nadim. Cluster 0 yang didapat memiliki jumlah data sebanyak 5 data, cluster 1 memiliki jumlah data sebanyak 15 data, dan cluster 2 memiliki jumlah data sebanyak 16 data. Dengan menggunakan *data mining K-means clustering* akan membantu pihak bandar udara Hang Nadim mempermudah dalam melakukan pendataan.

**Kata Kunci :** Pengelompokan, *Data mining*, *K-means clustering*

## ABSTRACT

*Transportation has become a part of human life in doing their daily lives. Not only humans utilize transportation, but shipping goods are also sent by utilizing the transportation that is in our lives. Delivery of goods or cargo often utilizes air transportation in the form of aircraft due to the speed of delivery. In the city of Batam, shipping goods by air transportation is very often done and in large quantities every day. The large number of shipments that are made produces very much data that can slow down the work in the data collection. By utilizing data mining in data collection, it is hoped that it can help airports in Batam city called Hang Nadim Airport. One algorithm that can be used to do data mining is the use of K-means Clustering. K-means clustering is one way of data mining that aims to group data into groups of data that have adjacent properties. Result from this research is 3 cluster model, which that cluster divided by their own weight. On cluster 0, it have 5 data, cluster 1 have 15 data and cluster 2 have 16 data.. By using K-means clustering data mining will help Hang Nadim airport dealers make data collection easier.*

*Keywords: Grouping, Data mining, K-means clustering*

## KATA PENGANTAR

Puji dan syukur kepada Allah SWT yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati.

Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Putera Batam.
2. Ketua Program Studi bapak Andi Maslan, S.T., M.SI.
3. Sunarsan Sitohang, S.Kom., M.TI. selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
4. Dosen dan Staf Universitas Putera Batam.
5. Orang tua penulis yang selalu mendoakan dan memberi semangat.
6. Adhitya Dewanta, Suwanda, Liok Habri, Ongki wibowo, Sofar, Andika restu, Suci Alfi, Erika Damanik sebagai sahabat yang selalu memberi support.
7. Teman-teman seangkatan yang berjuang bersama dan saling memberi semangat dan motivasi.
8. Serta berbagai pihak yang telah membantu dalam menyelesaikan skripsi ini secara langsung maupun tidak langsung.

Semoga Tuhan membalas kebaikan dan selalu mencurahkan hidayah serta taufikNya, Aamin.

Batam, Agustus 2019

Penulis



## DAFTAR ISI

	HALAMAN
<b>HALAMAN SAMPUL DEPAN</b>	
<b>HALAMAN JUDUL</b> .....	ii
<b>HALAMAN PERNYATAAN</b> .....	iii
<b>ABSTRAK</b> .....	v
<b>ABSTRACT</b> .....	vi
<b>KATA PENGANTAR</b> .....	vii
<b>DAFTAR ISI</b> .....	ix
<b>DAFTAR TABEL</b> .....	xi
<b>DAFTAR GAMBAR</b> .....	xii
<b>DAFTAR RUMUS</b> .....	xiii
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang Penelitian .....	1
1.2 Identifikasi Masalah.....	3
1.3 Pembatasan Masalah.....	3
1.4 Perumusan Masalah .....	4
1.5 Tujuan Penelitian .....	4
1.6 Manfaat Penelitian .....	4
<b>BAB II KAJIAN PUSTAKA</b>	
2.1 <i>Knowledge Discovery in Database (KDD)</i> .....	6
2.2 <i>Data Mining</i> .....	9
2.3 <i>Metode Data Mining</i> .....	11
2.4 <i>Software Pendukung</i> .....	15
2.5 Penelitian Terdahulu .....	16
2.6 Kerangka Pemikiran .....	18
<b>BAB III METODE PENELITIAN</b>	
3.1 Desain Penelitian .....	19
3.2 Teknik Pengumpulan Data .....	21
3.3 Operasional Variabel .....	21
3.4 Metode Analisis dan Rancangan Sistem.....	24
3.5 Lokasi Dan Jadwal Penelitian .....	25
<b>BAB IV HASIL DAN PEMBAHASAN</b>	
4.1 Analisis Data .....	27
4.2 Hasil Pengujian.....	41
<b>BAB V SIMPULAN DAN SARAN</b>	
5.1 Simpulan .....	51
5.2 Saran .....	52

**DAFTAR PUSTAKA  
DAFTAR RIWAYAT HIDUP  
SURAT KETERANGAN PENELITIAN  
LAMPIRAN**

## DAFTAR GAMBAR

	HALAMAN
<b>Gambar 2.1</b> Proses <i>Knowledge Discovery in Database</i> .....	7
<b>Gambar 2.2</b> Kerangka Pemikiran .....	18
<b>Gambar 3.1</b> Desain Penelitian.....	19
<b>Gambar 3.2</b> Rancangan Sistem.....	24
<b>Gambar 4.1</b> <i>Welcome To RapidMiner Studio</i> .....	41
<b>Gambar 4.2</b> <i>Insert Operator</i> .....	42
<b>Gambar 4.3</b> <i>Import Configuration Wizard</i> .....	43
<b>Gambar 4.4</b> <i>Select the Cells to Import</i> .....	44
<b>Gambar 4.5</b> <i>Format your columns</i> .....	45
<b>Gambar 4.6</b> <i>Insert Operator k-means</i> .....	45
<b>Gambar 4.7</b> Pengisian jumlah <i>k-cluster</i> .....	46
<b>Gambar 4.8</b> Proses penyambungan garis <i>read excel ke clustering</i> .....	47
<b>Gambar 4.9</b> Hasil <i>RapidMiner</i> Pengelompokan Cargo.....	47
<b>Gambar 4.10</b> Hasil <i>RapidMiner</i> Pengelompokan Cargo.....	48
<b>Gambar 4.11</b> <i>Cluster 0</i> .....	48
<b>Gambar 4.12</b> <i>Cluster 1</i> .....	49
<b>Gambar 4.13</b> <i>Cluster 2</i> .....	50

## DAFTAR TABEL

	HALAMAN
<b>Tabel 3.1</b> Data Bongkar Muat Barang/Kargo .....	21
<b>Tabel 4.1</b> Titik Pusat <i>Cluster</i> .....	28
<b>Tabel 4.2</b> <i>Cluster</i> Baru Iterasi ke-1 .....	34
<b>Tabel 4.3</b> Hasil Perhitungan Jarak Dan Pengelompokan Data Iterasi ke-2.....	34
<b>Tabel 4.4</b> <i>Cluster</i> Baru Iterasi ke-2 .....	36
<b>Tabel 4.5</b> Hasil Perhitungan Jarak Dan Pengelompokan Data Iterasi ke-3.....	37
<b>Tabel 4.6</b> Hasil Pengelompokan Dengan Jumlah Cargo Banyak .....	39
<b>Tabel 4.7</b> Hasil Pengelompokan Dengan Jumlah Cargo Sedikit .....	39
<b>Tabel 4.8</b> Hasil Pengelompokan Dengan Jumlah Cargo Menengah.....	36

## DAFTAR RUMUS

	HALAMAN
<b>Rumus 4.1</b> <i>Euclidian Distance</i> .....	16

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang Penelitian**

Transportasi merupakan bagian penting dalam kehidupan manusia. Bukan hanya transportasi seperti kendaraan saja yang dibutuhkan oleh manusia, tetapi ada juga transportasi yang berfokus untuk pengiriman barang yang sering disebut dengan kargo. Proses pengiriman kargo dapat menghasilkan data yang sangat banyak yang masih dapat diolah untuk keperluan tertentu. Jika data tersebut tidak dapat diolah dengan baik dan dibiarkan saja, maka akan menimbulkan penumpukan data dalam jumlah yang sangat besar. Namun, jika data yang banyak ini dapat diolah dengan baik, maka akan menghasilkan pengetahuan yang sangat bermanfaat (Setiawan & Tes, 2016).

Pada lokasi penelitian yang ditetapkan oleh peneliti pada penelitian ini, data kargo pada bandara Hang Nadim Batam yang dimiliki masih dalam kondisi tidak digunakan secara efektif dan efisien sehingga menjadi penumpukan data. Penumpukan data adalah keadaan dimana data berjumlah banyak tetapi tidak dimanfaatkan dengan baik mengakibatkan kelambatan kerja. Dicontohkan sebagai berikut, sebuah folder dalam komputer berisi banyak dokumen yang tidak diatur dengan rapi dan baik, ketika akan dilakukan pencarian, maka akan memakan waktu untuk menemukan dokumen yang kita inginkan. Sama seperti yang terjadi di kehidupan nyata, data yang menumpuk dapat berakibat karyawan kesulitan dalam melakukan pencarian data yang diinginkan dan berdampak kerja yang tidak efektif.

Data yang telah menumpuk dapat diolah untuk mendapatkan sebuah hasil pengetahuan baru. Dalam penelitian ini, hasil pengetahuan yang ingin didapatkan adalah pengelompokan data berdasarkan berat bagasi dan kargo agar nantinya pengetahuan ini dapat dimanfaatkan oleh pihak yang bersangkutan untuk memperkirakan kuantiti berat bagasi dan kargo pada bulan-bulan tertentu.

Data yang akan digunakan pada penelitian adalah data arus lalu lintas angkutan udara Bandar Udara Hang Nadim - Batam. Pada data penelitian terdapat berat barang bagasi dan kargo dalam satuan Kilogram (Kg) yang dibongkar dan dimuat dalam kategori per bulan dimana data yang diambil adalah data yang dimulai pada tahun 2015 sampai 2017. Pada penelitian ini, untuk mendapatkan pengetahuan yang diperlukan tersebut, peneliti akan digunakan teknik *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 berukuran besar (Handoko, 2016). Salah satu teknik yang terdapat pada *data mining* adalah *clustering*. *Clustering* merupakan keilmuan pada *data mining* yang dilakukan dengan cara mengelompokkan sejumlah data ke dalam kelompok (*cluster*) dengan sifat yang semirip mungkin dengan tujuan untuk mendapatkan sebuah pengetahuan baru yang dapat dimanfaatkan lagi untuk keperluan yang lain (Alfina & Santosa, 2012).

Namun, *data mining* yang dilakukan harus bersifat benar dan tepat sehingga dibutuhkan aplikasi perangkat lunak yang dapat membuktikan dan mengevaluasi *clustering* yang dilakukan. Pada penelitian ini, peneliti akan menggunakan

*software RapidMiner* yang merupakan salah satu *software* terkenal dalam membantu penelitian *data mining*.

Atas dasar yang telah dinyatakan diatas, maka peneliti tertarik untuk mengangkat penelitian yang berjudul “*Data Mining Metode K-Means Clustering Dalam Pengelompokan Jumlah Bagasi dan Barang Kargo*”.

## **1.2. Identifikasi Masalah**

Masalah yang akan dibahas pada penelitian ini adalah:

1. Data jumlah kargo pada tahun 2015 hingga data jumlah kargo pada tahun 2017 yang ada belum dimanfaatkan sepenuhnya.
2. Terjadi penumpukan data disebabkan banyaknya data yang ada.

## **1.3. Pembatasan Masalah**

Penelitian dibatasi oleh hal-hal sebagai berikut :

1. Data penelitian yang akan digunakan diambil dari Bandar Udara Hang Nadim Batam.
2. Data yang diambil adalah data lalu lintas barang atau kargo dalam satuan kilogram dalam jangka tahun 2015 – 2017.
3. Penelitian ini menggunakan program *RapidMiner* untuk menguji hasil penelitian.
4. Hasil dari penelitian hanya melakukan pengelompokkan data tanpa membangun *software* atau aplikasi.



#### **1.4. Perumusan Masalah**

Berdasarkan masalah yang ada, dapat diambil rumusan masalah sebagai berikut :

1. Bagaimana penerapan *data mining* metode *clustering* untuk mengelompokan data kargo?
2. Bagaimana implementasi *data mining clustering* dengan menggunakan *software RapidMiner*?

#### **1.5. Tujuan Penelitian**

Berdasarkan perumusan masalah di atas, maka tujuan dari penelitian sebagai berikut :

1. Untuk mengetahui bagaimana cara penerapan *data mining clustering* dalam membantu mencari pengetahuan yang baru.
2. Untuk mengimplementasikan *data mining clustering* dalam *software RapidMiner*.

#### **1.6. Manfaat Penelitian**

Penelitian ini diharapkan membawa manfaat bagi yang membaca, adapun sebagai berikut :

### **1.6.1 Aspek Teoritis**

Hasil penelitian diharapkan dapat memberikan manfaat dalam aspek teoritis seperti berikut :

1. Meningkatkan wawasan dan pengetahuan mengenai *data mining clustering*.
2. Menjadi sumber referensi bagi penelitian selanjutnya.

### **1.6.2 Aspek Praktis**

Hasil penelitian diharapkan dapat memberikan manfaat dalam aspek praktis seperti berikut :

1. Bagi Bandar Udara Hang Nadim Batam, dapat membantu dalam melakukan pencarian pengetahuan yang baru sehingga data dapat bermanfaat dan mengurangi penumpukan data yang tidak digunakan.
1. Bagi Peneliti, dapat dimanfaatkan menjadi bahan pembandingan bagi penelitian selanjutnya dengan topik pembahasan yang terkait serta sebagai salah satu syarat untuk mendapatkan gelar sarjana.

## **BAB II**

### **KAJIAN PUSTAKA**

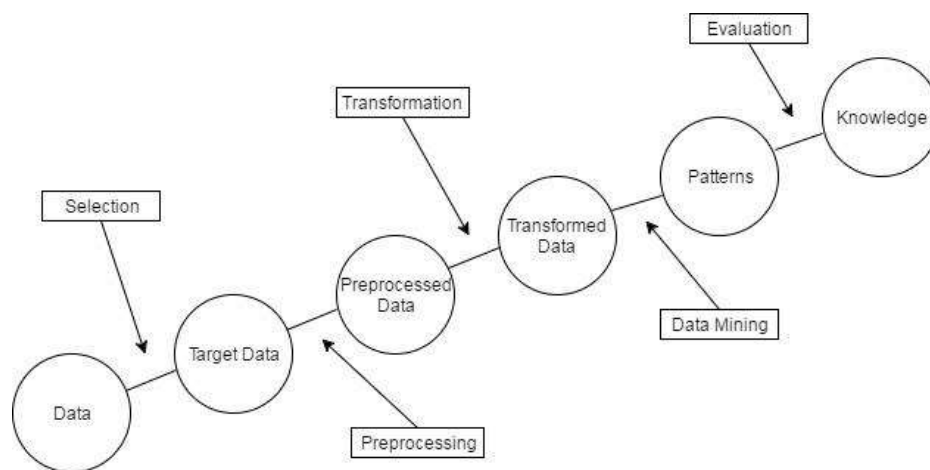
#### **2.1. *Knowledge Discovery in Database (KDD)***

Seperti yang telah dijelaskan pada bab 1, data bukan lagi hal yang aneh untuk didengar di zaman sekarang. Bahkan data sudah menjadi salah satu hal yang penting di dunia. Data dihasilkan oleh seluruh tindakan yang dilakukan manusia di dunia. Maka tidak heran jika data di dunia ini dapat dibilang sangat banyak. Data yang banyak di dunia ini akan menjadi tidak berguna jika tidak diolah dengan baik. Kondisi dimana data yang banyak tapi tidak diolah dengan baik inilah yang dapat menyusahkan pekerjaan. Jumlah data yang berlebihan dapat menjadi beban saat akan melakukan pencarian ulang. Penumpukan data sering disebut kondisi “Banyak Data tapi Sedikit Informasi”. Hal-hal yang seperti ini yang harus dihindari sehingga digunakanlah *data mining* (Suyanto, 2017).

Menurut Undang-Undang Nomor 1 Tahun 2009 tentang Penerbangan yang dituangkan di Peraturan Pemerintah Nomor 3 Tahun 2001 tentang Keamanan dan Keselamatan Penerbangan bahwa semua operasional penerbangan telah diatur mengikuti ketentuan yang ada pada Program Nasional Pengamanan Penerbangan Sipil. Namun pada faktanya, pada saat ini, Indonesia masih belum memiliki aturan terkait dengan keamanan dan keselamatan untuk layanan kargo udara (Siahaan, 2011). Semua data yang dihasilkan dari pendaftaran barang bagasi dan kargo hanya sekedar disimpan sebagai referensi tanpa digunakan lagi. Data yang banyak ini dapat dimanfaatkan lagi untuk mendapatkan pengetahuan yang bermanfaat lainnya.

Namun data yang banyak ini harus lebih dulu melalui proses sehingga data menjadi layak untuk diolah menggunakan *data mining*. Proses ini bernama proses *Knowledge Discovery in Database (KDD)*.

*Data mining* merupakan bagian dari proses KDD. Proses KDD merupakan proses untuk menganalisa dan mencari informasi yang dapat berguna dan bermanfaat pada sebuah kumpulan data. Proses KDD mengalami iterasi yang digambarkan seperti (Kasus, Dehasen, Haryati, Sudarsono, & Suryana, 2015):



**Gambar 2.1** Proses *Knowledge Discovery in Database*  
Sumber : Data Olahan Penelitian, 2019

1. *Selection*

Diawali dengan data yang masuk dilakukan pemilihan agar data yang akan diolah tepat dengan metode yang digunakan yang mengubah *data* menjadi *target data*.

2. *Preprocessing / Cleaning*

*Target data* yang telah diseleksi akan dilakukan pra-proses berupa pembersihan data dari data redundan dan data yang sama sehingga proses selanjutnya akan menjadi lebih cepat.

3. *Transformation*

Data yang telah melalui proses pembersihan akan melalui proses transformasi jenis data jika diperlukan. Transformasi data dari kualitatif menjadi kuantitatif dan sebaliknya. Jika data tidak perlu dilakukan transformasi, maka langkah ini boleh dilewati.

4. *Data Mining*

Setelah data siap untuk diproses, maka dilanjutkan tahap *data mining* yang memproses data dengan metode dan algoritma yang tepat dan menghasilkan sebuah pengetahuan baru yang bermanfaat.

5. *Evaluation*

Proses *data mining* baik *input*, *process* dan *output* harus melalui proses evaluasi agar tidak terjadi kesalahan pada proses. Jika proses evaluasi sudah selesai, maka telah selesai juga proses KDD. Informasi atau pengetahuan sudah dapat disajikan.

Proses *Knowledge Discovery in Database* dapat disimpulkan menjadi keseluruhan proses yang berhubungan penemuan hasil informasi atau pengetahuan yang terjadi saat proses mengidentifikasi pola yang ditemukan nantinya bersifat bermanfaat serta dapat dimengerti. (Sari & Sindunata, 2014).

## 2.2. *Data Mining*

*Data mining* sebenarnya telah diperkenalkan sejak tahun 1990an saat data yang ada pada waktu itu dimanfaatkan menjadi sebuah pengetahuan bermanfaat untuk berbagai bidang, baik dari akademis hingga statistik dan medis. Semakin berkembangnya zaman, *data mining* semakin diperlukan. Hal ini disebabkan jumlah data yang ada di dunia semakin banyak jumlahnya (Nur, Zarlis, & Nasution, 2015).

*Data mining* merupakan bagian dari proses *Knowledge Discovery in Database* (KDD) dan juga bagian dari *Pattern Recognition* (Nur et al., 2015). *Data mining* berarti sebuah istilah untuk menjelaskan proses mendapatkan sebuah informasi atau pengetahuan tersembunyi dari sebuah kumpulan data yang sangat besar yang disimpan dalam sebuah database dengan tujuan untuk menemukan dan memahami pengetahuan dari data tersebut (Setiawan & Tes, 2016). *Data mining* juga dapat diartikan sebagai suatu rangkaian atau urutan proses yang dimana *data mining* menggali sebuah hal tambah dari kumpulan data yang banyak dimana hal tambah ini berupa pengetahuan bermanfaat yang selama ini tidak diketahui jika dicari secara manual (Handoko, 2016). *Data mining* sendiri merupakan salah satu bagian akar bidang ilmu lainnya seperti *database*, *statistic*, *machine learning* dan *artificial learning* (Kusrini & Luthfi, 2009).

*Data mining* memiliki dampak yang besar untuk bidang ilmu yang mencakup pengolahan data, baik secara positif ataupun negatif. Jika tidak digunakan secara baik dan melanggar etika-etika yang ada, maka *data mining* dapat membawa efek negatif disebabkan pada proses pengolahan data, data-data pribadi yang bersifat rahasia dapat dibaca dan digunakan untuk hal-hal yang tidak baik. Namun jika *data mining* digunakan secara baik, maka dapat berefek positif. Selain untuk bidang keilmuan, *data mining* juga dapat digunakan dalam bidang komersial seperti kontrol jumlah data dan menghasilkan profit dengan memanfaatkan data yang ada menjadi aset penting seperti memanfaatkan data penjualan untuk memprediksi kondisi bisnis di masa depan (Vulandari, 2017).

*Data mining* dibagi ke dalam kategori sesuai dengan spesifikasi yang dimiliki masing-masing kategori, yaitu (Vulandari, 2017):

a. Prediktif

Kategori prediktif adalah kategori dimana data mining dimanfaatkan dengan tujuan memprediksi suatu nilai dari atribut tertentu dibandingkan dengan nilai dari atribut lainnya.

b. Deskriptif

Kategori deskriptif adalah kategori dimana *data mining* dimanfaatkan dengan tujuan untuk menjelaskan hubungan antar data yang ada dan pola-pola yang dihasilkan sehingga fokus dari deskriptif adalah untuk melakukan penyelidikan data yang sering diikuti dengan penjelasan hasil.

### 2.3. Metode *Data Mining*

*Data mining* dibagi menjadi beberapa kelompok berdasarkan tugasnya yang dapat dilakukan sebagai berikut (Setiawan & Tes, 2016):

#### 1. Deskripsi

Untuk mencari cara yang menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola dan kecenderungan yang sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

#### 2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.

#### 3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.

#### 4. Klasifikasi

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



## 5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan atau memperhatikan dan membentuk kelas objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam *cluster* lain. Algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

## 6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

### 2.3.1. Metode *Clustering*

*Clustering* merupakan salah satu bagian dari ilmu *data mining* yang bersifat *unsupervised*. Dikatakan *unsupervised*, disebabkan pada metode *clustering*, kelompok tidak dikenal dari awal sehingga tujuan dari permodelan ini adalah untuk menentukan kelompok pada data yang ada (Bianglala et al., 2013). *Clustering* bertujuan untuk memisahkan data yang berjumlah besar ke dalam sejumlah kelompok dengan karakteristiknya masing-masing yang dimana setiap kelompoknya memiliki tingkat hubungan yang sama antar satu anggota *cluster* dengan anggota lain dalam *cluster* yang sama (Jannah & Yulianto, 2016).

Dengan mengartikan *clustering* merupakan pengelompokan data dan membentuk kelas objek-objek yang memiliki kemiripan, maka arti dari *cluster* adalah kumpulan data yang memiliki kemiripan antar satu dengan yang lainnya dan memiliki perbedaan dengan data di *cluster* lain. *Clustering* berbeda dengan metode klasifikasi yaitu tidak ada variabel target pada *clustering* yang berarti *clustering* tidak berusaha untuk melakukan klasifikasi ataupun memprediksi, namun hanya melakukan pembagian data menjadi kelompok yang bersifat homogen (Kusrini & Luthfi, 2009). Selain untuk mengelompokkan data yang memiliki sifat yang sama, *clustering* juga dapat dimanfaatkan untuk tujuan yang lain seperti prediksi dan klasifikasi. Prediksi seperti peramalan harga emas per hari dapat memanfaatkan metode *clustering* sehingga dapat membantu investor dalam mengambil keputusan dalam membeli emas (Sitohang & Siringo, 2018).

### **2.3.2. Algoritma *K-Means***

*K-Means* merupakan salah satu metode data *clustering* non hierarki yang mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lainnya. (Metisen & Sari, 2015).

*K-Means* merupakan salah satu algoritma dalam data mining yang bisa digunakan untuk melakukan pengelompokan/*clustering* suatu data. Ada banyak pendekatan untuk membuat *cluster*, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam group yang sama berdasarkan tingkat persamaan diantara anggota-anggotanya. Pendekatan lainnya adalah dengan membuat

sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah *clustering*. Metode *K-Means* adalah metode yang termasuk dalam algoritma *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan algoritma ini hanya bekerja pada atribut numerik (Dhuhita, 2015).

*K-Means Clustering* merupakan metode yang termasuk ke dalam golongan algoritma *Partitioning Clustering*. Langkah-langkah dari metode *K-Means* adalah sebagai berikut (Handoyo et al., 2014) :

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*) yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X_{1i} - (X_{1j}))^2 + (X_{2i} - (X_{2j}))^2 + \dots + (X_{ki} - (X_{kj}))^2}$$

Dimana:

$D(i,j)$  = jarak data ke  $i$  ke pusat  $j$

$X_{ki}$  = Data ke  $i$  pada atribut data ke  $k$

$X_{kj}$  = Titik pusat ke  $j$  pada atribut  $k$

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*nya.
5. Tentukan posisi *centroid* baru ( $k$  C) dengan cara menghitung nilai rata-rata dari data yang ada pada *centroid* yang sama. Dimana  $n_k$  adalah jumlah dokumen dalam *cluster*  $k$  dan  $d_i$  adalah dokumen dalam *cluster*

k. 6. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama, tidak sama.

## **2.4. Software Pendukung**

### **2.4.1 RapidMiner**

*Data mining* mempunyai beberapa perangkat lunak aplikasi atau *software* pendukung seperti WEKA, *Orange*, *Microsoft Analysis Service*, *Oracle Data Mining*, dan *RapidMiner* dimana aplikasi-aplikasi ini dapat membantu dalam pengolahan *data mining*. Pada penelitian ini, peneliti akan menggunakan aplikasi *RapidMiner* untuk melakukan pengujian.

*RapidMiner* merupakan salah satu perangkat lunak atau aplikasi yang diciptakan oleh Dr. Markus Hofman dari *Institute of Technology Blanchardtown* dan Ralf Klinkenberg dengan tampilan GUI (*Graphical User Interface*) yang memudahkan pengguna dalam menggunakan aplikasi untuk pengolahan *data mining*. Perangkat lunak ini bersifat *open source* dan dibuat dengan menggunakan program Java. Walaupun diciptakan dengan menggunakan program Java, pengguna tidak perlu melakukan koding dalam penggunaan *RapidMiner* disebabkan semua fasilitas telah disediakan. *RapidMiner* adalah suatu aplikasi yang ditujukan secara khusus dalam pengelolaan *data mining* dengan fasilitas model-model *data mining* seperti Model *Bayes*, *Decision Tree*, *Neural Network*, dan banyak lagi (Kasus et al., 2015).

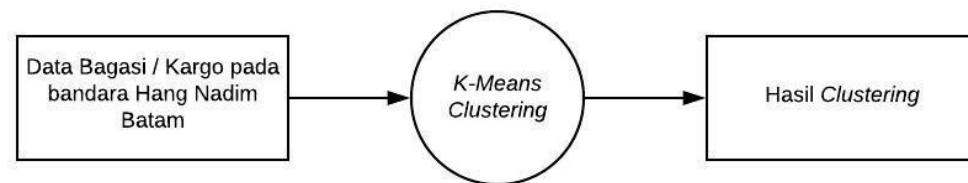
## 2.5. Penelitian Terdahulu

Terdapat beberapa penelitian yang berupa jurnal digunakan peneliti sebagai referensi dalam penelitian yang dilakukan. Referensi-referensi tersebut diambil karena adanya permasalahan yang berhubungan dengan penelitian yang dibahas di dalamnya, Berikut adalah penelitian terdahulu berupa jurnal yang terkait dengan penelitian, yakni:

1. Pada jurnal referensi yang ditulis oleh Benri Melpa Metisen yang berjudul **ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA**, meneliti tentang pemanfaatan aplikasi Tanagra dalam melakukan *data mining clustering K-means* yang menghasilkan dua jenis kelompok data yaitu jenis barang yang laku terjual dan tidak laku terjual sehingga dapat membantu Swalayan Fadhila dalam melakukan proses *stocking* barang penjualan.
2. Pada jurnal referensi yang ditulis oleh Nur Jannah yang berjudul **MENGELOMPOKKAN SISWA BERPRESTASI AKADEMIK DENGAN MENGGUNAKAN METODE K-MEANS KELAS VII MTS HIDAYATUL MUDTAP'IN PANCORAN KEDIRI**, meneliti tentang pengelompokan peserta didik yang memiliki kesamaan dan perbedaan di bidang nilai akademis, nilai kognitif, psikomotorik dan afektif dengan jumlah sampel 30 orang. Penelitian dilakukan dengan menggunakan metode *k-means* yang menghasilkan hasil validasi sebesar 33.3333% valid dan 66.6667% tidak valid.

3. Pada jurnal referensi yang ditulis oleh Femi Dwi Astuti yang berjudul **PENERAPAN DATA MINING UNTUK CLUSTERING DATA PENDUDUK MISKIN MENGGUNAKAN ALGORITMA HARD C-MEANS**, meneliti tentang pengelompokan penduduk miskin di provinsi Daerah Istimewa Yogyakarta Kecamatan Bantul. Penelitian dilakukan dengan tujuan agar bantuan yang tersalurkan dapat dilakukan dengan tepat. Hasil penelitian menunjukkan penduduk dibagi ke dalam 3 *cluster* dengan total penduduk yang dijadikan sampel penelitian berjumlah 1313 sampel.
4. Pada jurnal referensi yang ditulis oleh Johan Oscar Ong yang berjudul **IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN STRATEGI MARKETING PRESIDENT UNIVERSITY**, meneliti tentang pemanfaatan *data mining clustering* metode *k-means* dengan variabel berupa jurusan, kota asal dan nilai mahasiswa dalam menentukan fokus pemasaran yang akan dilakukan.
5. Pada jurnal referensi yang ditulis oleh Rendy Handoyo yang berjudul **PERBANDINGAN METODE CLUSTERING MENGGUNAKAN METODE SINGLE LINKAGE DAN K-MEANS PADA PENGELOMPOKAN DOKUMEN**, meneliti tentang perbandingan antara penggunaan metode *Single Linkage* dan *K-means* dalam melakukan pengelompokan yang menghasilkan jawaban berupa metode *Single Linkage* memiliki performansi yang lebih baik dalam pengelompokan dibandingkan metode *K-means*, namun metode *K-means* melakukan pengelompokan secara sederhana dan dapat digunakan serta dimengerti dengan mudah.

## 2.6. Kerangka Pemikiran



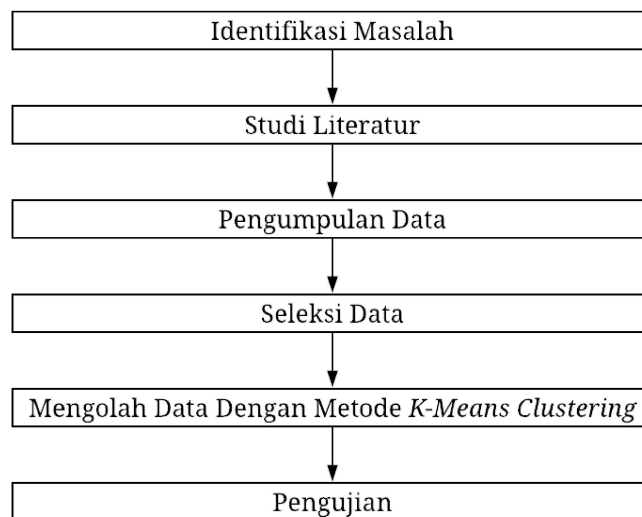
**Gambar 2.2** Kerangka Pemikiran  
**Sumber :** Data Olahan penelitian, 2019

Pada gambar 2.2, Penelitian ini menggunakan data bagasi / kargo yang didapat dari bandara Hang Nadim Batam dari tahun 2015-2017, dimana data tersebut akan dijadikan input pada Penelitian ini. Dimana data tersebut akan diseleksi dan dibersihkan (*cleaning*) kemudian data-data yang telah diseleksi dan *cleaning*, data tersebut akan diproses dengan menggunakan Metode *K-mean clustering* dan setelah data tersebut telah proses maka akan mendapatkan hasil yang berupa pengelompokan data.

## BAB III METODE PENELITIAN

### 3.1 Desain Penelitian

Adapun bentuk desain penelitian yang terdiri dari beberapa tahap atau langkah sebagai berikut:



**Gambar 3.1** Desain Penelitian  
**Sumber :** Data Olahan Penelitian, 2019

Penjelasan dari desain penelitian pada Gambar 3.1 sebagai berikut:

1. Identifikasi Masalah

Menentukan pokok masalah dalam penelitian. Dimana permasalahan data kargo yang ada belum dimanfaatkan sepenuhnya dan terjadinya penumpukan data yang disebabkan banyaknya data yang ada.



## 2. Studi *Literature*

Studi *literature* atau kajian pustaka dilakukan dengan tujuan untuk mencari jurnal-jurnal dan buku-buku yang akan dijadikan sebagai referensi untuk membantu peneliti dalam menyelesaikan permasalahan pada penelitian.

## 3. Pengumpulan Data

Pada penelitian ini peneliti melakukan pengumpulan data-data penelitian dari Bandara Udara Hang Nadim Batam dimana data yang diambil adalah data lalu lintas barang atau kargo dalam satuan kilogram.

## 4. Seleksi Data

Setelah data-data penelitian telah dikumpulkan, data-data tersebut akan diseleksi dan setelah dilakukan seleksi data-data tersebut, maka akan mendapatkan sampel yang akan diolah, setelah dilakukannya seleksi data pada data-data yang dikumpulkan, maka tahap selanjutnya adalah melakukan tahapan *cleaning*. Menghilangkan atribut-atribut atau menghilangkan variabel yang tidak berguna.

## 5. Mengolah Data dengan Metode *K-means Clustering*

Setelah data-data tersebut telah diseleksi dan *cleaning*, maka selanjutnya adalah mengolah data dengan menggunakan metode *K-means Clustering*, data-data tersebut akan diolah sehingga akan menghasilkan yaitu pengelompokan data.

## 6. Pengujian Hasil

Pengujian Hasil dilakukan Untuk memastikan bahwa hasil yang dikelompokan tersebut benar, maka peneliti menggunakan aplikasi *Rapid Miner* untuk memastikan bahwa hasil yang telah diolah telah dikelompokan itu sesuai atau sama.

### 3.2 Teknik Pengumpulan Data

Pada penelitian ini adapun cara dalam pengumpulan atau teknik pengumpulan data yang dilakukan oleh peneliti yaitu sebagai berikut:

#### 1. Observasi

Untuk mendapatkan data-data yang diperlukan dalam penelitian, peneliti melakukan observasi langsung kelapangan untuk mengumpulkan data data mengenai data lalu lintas barang atau kargo.

#### 2. Wawancara

Peneliti melakukan wawancara dengan bapak Rizki yang bertugas sebagai staff pengelola bagasi dan kargo pada Bandar Udara Hang Nadim – Batam.

#### 3. Kajian Pustaka

Agar peneliti dapat menjelaskan dengan tepat, maka diperlukan kajian pustaka dari media seperti dari buku, jurnal dan referensi skripsi yang berhubungan dengan penelitian yang dilakukan.

### 3.3 Operasional Variabel

Berikut adalah data-data Bongkar Muat barang/kargo pada tabel 3.1:

**Tabel 3.1** Data Bongkar Muat Barang/Kargo

No	Data Bulan	Bongkar (bgs)	Muat(bgs)	Bongkar(brg)	Muat(brg)
1	15-Jan	1,615,493	1,554,331	2,385,891.00	1,144,380.00
2	15-Feb	1,298,032	1,363,293	2,168,091.00	976,890.00
3	15-Mar	1,354,522	1,341,041	2,279,866.00	1,052,940.00

**Sumber :** Data Olahan Penelitian, 2019

**Tabel 3.1** Lanjutan

No	Data Bulan	Bongkar (bgs)	Muat(bgs)	Bongkar(brg)	Muat(brg)
4	15-Apr	1,391,630	1,344,203	2,207,586.00	1,038,112.00
5	15-May	1,437,226	1,575,520	2,293,286.00	1,061,185.00
6	15-Jun	1,422,526	1,708,083	2,341,039.00	1,213,953.00
7	15-Jul	2,129,468	2,574,476	1,922,490.00	811,191.00
8	15-Aug	2,030,155	1,594,825	2,036,476.00	905,056.00
9	15-Sep	1,264,881	1,396,846	1,902,423.00	978,006.00
10	15-Oct	1,731,033	1,514,823	2,063,718.00	882,711.00
11	15-Nov	1,385,589	1,499,960	2,126,011.00	872,055.00
12	15-Dec	1,934,594	2,093,048	2,516,928.00	1,154,985.50
13	16-Jan	2,004,546	1,901,279	2,434,684.00	971,913.50
14	16-Feb	1,579,179	1,700,132	2,038,258.00	931,481.00
15	16-Mar	1,630,160	1,759,522	2,220,727.00	1,136,599.00
16	16-Apr	1,547,165	1,658,495	1,893,738.00	1,009,355.00
17	16-May	1,767,176	2,081,359	2,454,420.00	1,295,113.00
18	16-Jun	1,502,470	2,282,088	2,555,435.00	1,603,245.00
19	16-Jul	2,861,101	2,865,419	2,121,141.00	1,047,345.50
20	16-Aug	1,877,949	2,044,721	2,378,266.00	1,411,718.50
21	16-Sep	1,887,937	1,953,957	2,350,924.00	1,444,352.00
22	16-Oct	1,813,086	1,929,812	2,332,901.00	1,301,384.50
23	16-Nov	1,816,892	2,118,083	2,438,462.00	1,541,222.50
24	16-Dec	2,295,567	2,586,779	2,543,180.00	1,714,343.50
25	17-Jan	2,290,267	2,317,261	2,552,877.00	1,568,617.00
26	17-Feb	1,608,986	1,759,788	2,013,113.00	1,226,223.50
27	17-Mar	1,767,988	2,023,985	2,344,558.00	1,551,776.50
28	17-Apr	1,722,730	2,160,886	2,088,504.50	1,476,573.50
29	17-May	1,629,224	2,016,575	2,218,719.00	1,605,049.00
30	17-Jun	1,564,631	2,459,049	1,998,764.00	1,619,007.00
31	17-Jul	2,618,545	2,394,004	2,090,964.00	1,558,929.50
32	17-Aug	1,803,973	2,179,078	2,417,752.00	1,614,847.80
33	17-Sep	2,058,118	2,074,788	2,293,631.00	1,962,946.50

**Sumber :** Data Olahan Penelitian, 2019



**Tabel 3.1** Lanjutan

No	Data Bulan	Bongkar (bgs)	Muat(bgs)	Bongkar(brg)	Muat(brg)
34	17-Oct	1,693,054	1,874,876	2,396,529.00	2,294,509.50
35	17-Nov	1,520,961	1,797,474	2,362,423.00	2,036,155.00
36	17-Dec	1,851,811	2,213,280	2,691,942.00	2,339,083.00

**Sumber :** Data Olahan Penelitian, 2019

Operasional variabel yang digunakan dalam penelitian ini adalah data kargo tahun 2015 hingga 2017 pada bandar udara Hang Nadim Batam, yaitu sebagai berikut:

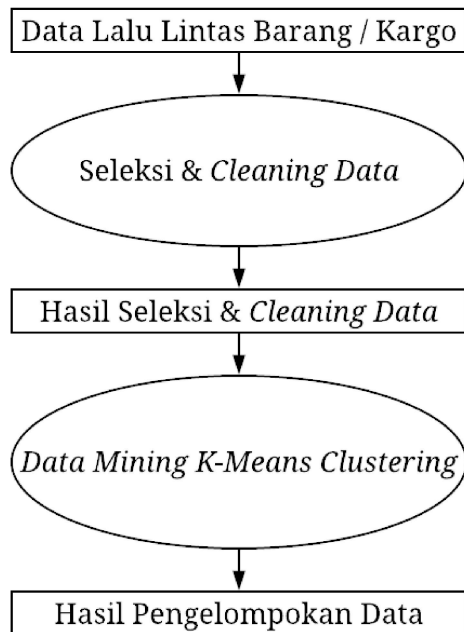
1. Bongkar barang/kargo

Data-data barang/kargo yang digunakan dalam operasional variabel ini adalah data Bongkar sebanyak 36 data. kegiatan bongkar adalah proses menurunkan barang dari kapal lalu menyusunnya di dalam gudang di pelabuhan atau *stock pile/container yard*.

2. Muat barang/kargo

Data-data barang/kargo yang digunakan dalam operasional variabel ini adalah data muat sebanyak 36 data. Yang dimaksud dengan kegiatan muat adalah proses memindahkan barang dari gudang, menaikkan lalu menumpuknya di atas kapal.

### 3.4 Metode Analisis dan Rancangan Sistem



**Gambar 3.2** Rancangan Sistem  
**Sumber :** Data olahan penelitian, 2019

Pada gambar 3.2 langkah pertama yaitu mengumpulkan data-data tersebut yang diperlukan pada penelitian, kemudian data data yang telah dikumpulkan tersebut akan dilakukan seleksi data sehingga mendapatkan hasil berupa sampel data. Kemudian sampel data tersebut akan dibersihkan atau *cleaning* yang bertujuan untuk menghilangkan variabel yang tidak berguna atau atribut-atribut yang tidak digunakan. Setelah sampel tersebut telah dibersihkan atau *cleaning*, sampel-sampel tersebut kemudian diolah dengan menggunakan *data mining k-means clustering* dimana langkah dari metode *k-means* adalah

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 ( $kC$ ) dengan cara menghitung nilai rata-rata dari data yang ada pada *centroid* yang sama. Dimana  $n_k$  adalah jumlah dokumen dalam *cluster k* dan  $d_i$  adalah dokumen dalam *cluster k*.
6. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama, tidak sama. Setelah proses pengolahan data menggunakan metode *k-means clustering* telah selesai maka akan mendapatkan hasil berupa pengelompokan.

### **3.5 Lokasi dan Jadwal Penelitian**

Berikut adalah lokasi dan jadwal pelaksanaan penelitian yang dilakukan oleh peneliti yaitu sebagai berikut:

#### **3.5.1 Lokasi Penelitian**

Lokasi penelitian yang dilakukan di Bandara Udara Internasional Hang Nadim Batam yang beralamat di Jl. Hang Nadim No.01, Batu Besar, Nongsa, Kota Batam, Kepulauan Riau.

### 3.5.2 Jadwal Penelitian

Penelitian ini dilaksanakan mulai dari Maret 2019 sampai dengan Juli 2019 yang dapat dilihat pada jadwal penelitian di bawah ini:

**Tabel 3.2** Jadwal Penelitian

No	Kegiatan	MARET 2019				APRIL 2019				MEI 2019				JUNI 2019				JULI 2019			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Input Judul	■																			
2	Pengumpulan Data		■	■																	
3	Pembuatan BAB 1				■	■															
4	Pembuatan BAB 2						■	■													
5	Pembuatan BAB 3								■	■	■										
6	Pembuatan BAB 4												■	■	■	■	■	■			
7	Pembuatan BAB 5																		■	■	
8	Penyerahan Skripsi																				■

**Sumber :** Data Olahan Penelitian, 2019