

**PENGELOMPOKAN DATA  
MATERIAL PROYEK MV DOULOS PHOS HOTEL  
MENGUNAKAN ALGORITMA K-MEANS  
CLUSTERING**

**SKRIPSI**



**Oleh:  
Larasanti  
160210183**

**PROGRAM STUDI TEKNIK DAN KOMPUTER  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
TAHUN 2020**

**PENGELOMPOKAN DATA  
MATERIAL PROYEK MV DOULOS PHOS HOTEL  
MENGUNAKAN ALGORITMA K-MEANS  
CLUSTERING**

**SKRIPSI**

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



**Oleh:  
Larasanti  
160210183**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
TAHUN 2020**



## SURAT PERNAYTAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Larasanti  
NPM : 160210183  
Fakultas : Teknik dan Komputer  
Program Studi : Teknik Informatika

Menyatakan bahwa skripsi yang saya buat dengan judul:

### **Pengelompokan data material proyek mv doulos phos hotel menggunakan algoritma k-means clustering**

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, didalam naskah “Pengelompokan data material proyek mv doulos phos hotel menggunakan algoritma k-means clustering“ ini tidak terdapat kerja ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah “Pengelompokan data material proyek mv doulos phos hotel menggunakan algoritma k-means clustering“ ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah “Pengelompokan data material proyek mv doulos phos hotel menggunakan algoritma k-means clustering“ ini digugurkan dan sarjana komputer yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun

Batam, 19 Februari 2020  
Yang membuat pernyataan,

**Larasanti**  
160210183

**PENGELOMPOKAN DATA  
MATERIAL PROYEK MV DOULOS PHOS HOTEL  
MENGUNAKAN ALGORITMA K-MEANS  
CLUSTERING**

**SKRIPSI**

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

**Oleh:  
Larasanti  
160210183**

**Telah disetujui oleh Pembimbing pada tanggal  
seperti tertera di bawah ini**

**Batam, 19 Februari 2020**

**Rahmat Fauzi, S.Kom., M.Kom  
Pembimbing**

## ABSTRAK

PT. Starindo Acetech Perkasa merupakan sebuah perusahaan yang bergerak dibidang *maintenance, general contractor* dan *supplier* setiap harinya memiliki banyak data transaksi mulai dari data pembelian hingga data penjualan. Dengan banyaknya transaksi data tersebut namun belum memiliki sebuah solusi sistem yang digunakan untuk mengolah data transaksi tersebut menjadi informasi yang bermanfaat dalam mengambil keputusan untuk penyetokan material proyek sehingga untuk penyetokan material selama ini masih mengira-ngira karena data transaksi hanya berfungsi sebagai arsip dan tidak dimanfaatkan untuk kedepannya. Selain itu barang yang dibeli pihak perusahaan untuk kebutuhan proyek dapat terjadi secara berlebihan dan kekurangan, kelebihan pembelian material dapat menyebabkan ketidakgunaan barang tersebut sehingga menumpuk di gudang menyebabkan kerugian dan kekurangan pembelian material dapat menyebabkan keterlambatan kerja sehingga dapat menyebabkan ketidakefektifan kerja. *Data mining* merupakan salah satu solusi yang dapat diterapkan oleh pihak perusahaan dalam menganalisis data perusahaan dalam jumlah yang banyak, metode *data mining* yang digunakan dalam penelitian ini yaitu *k-means clustering* yang dapat dimanfaatkan pihak perusahaan untuk membantu manajemen penyetokan material proyek yang akan datang sehingga lebih teratur dan mengantisipasi mengalami kerugian. Data yang diolah penelitian ini yaitu data transaksi pembelian proyek *MV Doulos Phos* hotel dari Juli 2017 hingga Oktober 2019, menggunakan aplikasi *RapidMiner*. Hasil pengujian menggunakan *RapidMiner* dengan 23 record data pembelian, hasil dari evaluasi menetapkan hasil *cluster* pertama terdiri dari 14 item (tinggi), *cluster* kedua terdiri dari 6 item (sedang), *cluster* ketiga terdiri dari 3item (rendah). Dari hasil tersebut diharapkan memberikan manfaat yang membantu perusahaan lebih teratur dalam manajemen penyetokan barang-barang dan mengantisipasi mengalami kerugian di proyek yang akan datang.

Kata Kunci: Penyetokan, *Data Mining*, *K-Means Clustering*

## **ABSTRACT**

*PT. Starindo Acetech Perkasa is a company engaged in maintenance, general contractor and supplier every day has a lot of transaction data ranging from purchase data to sales data. With so many data transactions but do not yet have a system solution that is used to process the transaction data into useful information in making decisions for the provision of project material so that for the provision of material so far it is still guessing because the transaction data only functions as an archive and is not utilized for going forward Besides the goods purchased by the company for project needs can occur excessively and shortages, excess material purchases can cause uselessness of the goods so that piling up in warehouses causes losses and lack of material purchases can cause work delays that can cause work ineffectiveness. Data mining is one of the solutions that can be applied by companies in analyzing company data in large numbers, the data mining method used in this research is k-means clustering that can be utilized by the company to help management of future project material sourcing so that it is more regularly and anticipate loss. The data processed by this research is the transaction data purchase of the MV Doulos Phos hotel project from July 2017 to October 2019, using the RapidMiner application. The test results using RapidMiner with 23 purchase data records, the results of the evaluation set the results of the first cluster consisting of 14 items (high), the second cluster consisting of 6 items (medium), the third cluster consisting of 3 items (low). The results are expected to provide benefits that help companies be more organized in the management of goods supply and anticipate loss in future projects.*

*Keyword: Depositing, Data Mining, K-means Clustering*

## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan YME 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, Ibu Dr. Nur Elfi Husda, S.Kom., M.SI.
2. Ketua Program Studi Teknik Informatika, Bapak Andi Maslan, S.T., M.SI.
3. Rahmat Fauzi, S.Kom., M.Kom selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
4. Joni Eka Candra, S.T., M.T. selaku pembimbing akademik selama mejalani perkuliahan.
5. Dosen dan Staff Universitas Putera Batam.
6. Orang tua dan saudara saya tercinta, yang telah memberikan dukungan, pengertian dan doa selama penyusunan skripsi ini dan selama saya menjalani perkuliahan.
7. Pimpinan dan staff PT. Starindo Acetech Perkasa
8. Teman-teman seperjuang yang telah saling membantu dan memotivasi dalam penyusunan skripsi dan selama menjalani perkuliahan.
9. Semua pihak yang telah bersedia meluangkan waktu, tenaga dan pikirannya dalam memberikan data/informasi selama penulis membuat makalah yang tidak dapat saya sebutkan satu-persatu.

Semoga Tuhan YME membalas kebaikan dan selalu mencurahkan hidayah serta taufik-Nya, Amin.

Batam, 19 Februari 2020

Larasanti

# DAFTAR ISI

	Halaman
<b>HALAMAN SAMPUL DEPAN</b>	
<b>HALAMAN JUDUL</b>	
<b>HALAMAN PERNYATAAN.....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>iii</b>
<b>ABSTRAK.....</b>	<b>iv</b>
<b>ABSTRACK.....</b>	<b>v</b>
<b>KATA PENGANTAR.....</b>	<b>vi</b>
<b>DAFTAR ISI.....</b>	<b>vii</b>
<b>DAFTAR GAMBAR.....</b>	<b>ix</b>
<b>DAFTAR TABEL.....</b>	<b>x</b>
<b>DAFTAR RUMUS.....</b>	<b>xi</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>xii</b>
<b>BAB 1.....</b>	<b>1</b>
<b>PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang Masalah.....	1
1.2. Identifikasi Masalah.....	3
1.3. Batasan Masalah.....	4
1.4. Rumusan Masalah.....	4
1.5. Tujuan Penelitian.....	5
1.6. Manfaat Penelitian.....	5
<b>BAB II.....</b>	<b>7</b>
<b>KAJIAN PUSTAKA.....</b>	<b>7</b>
2.1. <i>Knowledge Discovery in Database (KDD)</i> .....	7
2.2. <i>Data Mining</i> .....	10
2.2.1. Sejarah <i>Data Mining</i> .....	10
2.2.2. Asal Ilmu <i>Data Mining</i> .....	11
2.2.3. Pengertian <i>Data Mining</i> .....	12
2.2.4. Algoritma <i>Data mining</i> .....	13
2.3. Metode <i>Data Mining</i> .....	23
2.3.1. <i>Clustering</i> .....	23
2.3.2. Estimasi.....	23

2.3.3. Prediksi .....	24
2.3.4. Klasifikasi .....	24
2.3.5. Asosiasi .....	24
2.4. <i>Software Data Mining</i> .....	24
2.5. Tujuan Umum Penelitian.....	29
2.6. Penelitian Terdahulu.....	29
2.7. Kerangka Pemikiran .....	35
<b>BAB III.....</b>	<b>37</b>
<b>METODE PENELITIAN .....</b>	<b>37</b>
3.1. Desain Penelitian.....	37
3.2. Pengumpulan Data.....	39
3.3. Operasional Data .....	40
3.4. Metode Perancangan Sistem.....	40
3.5. Lokasi Dan Jadwal Penelitian.....	46
3.5.1. Lokasi Penelitian.....	46
3.5.2. Jadwal Penelitian.....	47
<b>BAB IV .....</b>	<b>48</b>
<b>HASIL DAN PEMBAHASAN .....</b>	<b>48</b>
4.1. Arsitektur Sistem .....	48
4.2. Sampel Data yang Digunakan .....	49
4.3. Analisa Data .....	51
4.4. Analisa Proses Algoritma .....	55
4.5. Pengujian Algoritma.....	68
<b>SIMPULAN DAN SARAN .....</b>	<b>79</b>
5.1. Simpulan.....	79
5.2. Saran .....	80
<b>DAFTAR PUSTAKA .....</b>	<b>82</b>

## DAFTAR GAMBAR

	Halaman
<b>Gambar 2. 1</b> Proses <i>Knowledg Discovery In Database</i> .....	7
<b>Gambar 2. 2</b> Asal Ilmu <i>Data Mining</i> .....	11
<b>Gambar 2. 3</b> Proses Mencari Nilai <i>Centroid</i> .....	14
<b>Gambar 2. 4</b> <i>Flowchart</i> Proses <i>K-means</i> .....	16
<b>Gambar 2. 5</b> Logo <i>RapidMiner</i> .....	25
<b>Gambar 2. 6</b> Logo <i>Weka</i> .....	26
<b>Gambar 2. 7</b> Logo <i>Orange</i> .....	26
<b>Gambar 2. 8</b> Logo <i>Rattel GUI</i> .....	27
<b>Gambar 2. 9</b> Logo <i>R Studio</i> .....	28
<b>Gambar 2. 10</b> Kerangka Pemikiran .....	36
<b>Gambar 3. 1</b> Desain Penelitian .....	37
<b>Gambar 3. 2</b> Lokasi Penelitian .....	46
<b>Gambar 4. 1</b> Menu <i>RapidMiner</i> .....	48
<b>Gambar 4. 2</b> Sumber Data Mentah <i>Statement</i> Pembelian 2018 .....	51
<b>Gambar 4. 3</b> Sampel Data Mentah <i>Invoice</i> Pembelian 2018.....	52
<b>Gambar 4. 4</b> Hasil <i>Cluster</i> .....	55
<b>Gambar 4. 5</b> <i>Interface RapidMiner</i> .....	68
<b>Gambar 4. 6</b> Tampilan Awal <i>RapidMiner</i> .....	69
<b>Gambar 4. 7</b> Halaman Kosong .....	69
<b>Gambar 4. 8</b> Halaman <i>Import Data</i> .....	70
<b>Gambar 4. 9</b> Memasukkan Data <i>Input</i> .....	70
<b>Gambar 4. 10</b> Halaman Data Terinput.....	71
<b>Gambar 4. 11</b> Halaman Proses Akhir <i>Import Data</i> .....	71
<b>Gambar 4. 12</b> Halaman <i>Design</i> .....	72
<b>Gambar 4. 13</b> Proses <i>Drag Data</i> .....	73
<b>Gambar 4. 14</b> Proses <i>Input</i> Metode <i>K-means</i> .....	73
<b>Gambar 4. 15</b> Proses <i>Input</i> Nilai <i>K</i> .....	74
<b>Gambar 4. 16</b> Tahap Akhir Proses <i>K-means</i> .....	74
<b>Gambar 4. 17</b> Hasil <i>Clustering</i> Data .....	75
<b>Gambar 4. 18</b> Hasil Tampilan <i>Text View</i> .....	76
<b>Gambar 4. 19</b> Hasil Tampilan <i>Pie 3D</i> .....	76
<b>Gambar 4. 20</b> Tampilan Hasil <i>Cluster</i> Pertama.....	77
<b>Gambar 4. 21</b> Tampilan Hasil <i>Cluster</i> Kedua .....	77
<b>Gambar 4. 22</b> Tampilan Hasil <i>Cluster</i> Ketiga .....	78

## DAFTAR TABEL

	Halaman
<b>Tabel 2. 1</b> Simbol-Simbol <i>Flowchart</i> Proses <i>K-means</i> .....	18
<b>Tabel 3. 1</b> Operasional Data Pembelian Material.....	40
<b>Tabel 3. 2</b> Data Rekapulasi Pembelian Material <i>MV Doulos Phos Hotel</i> .....	41
<b>Tabel 3. 3</b> Data Setelah <i>Cleaning</i> .....	43
<b>Tabel 3. 4</b> Hasil Pengelompokan <i>Cluster 1</i> .....	44
<b>Tabel 3. 5</b> Hasil Pengelompokan <i>Cluster 1</i> .....	45
<b>Tabel 3. 6</b> Hasil Pengelompokan <i>Cluster 2</i> .....	45
<b>Tabel 3. 7</b> Jadwal Penelitian .....	47
<b>Tabel 4. 1</b> Sampel Data .....	50
<b>Tabel 4. 2</b> Data <i>Cleaning</i> .....	53
<b>Tabel 4. 3</b> Data Hasil Transformasi.....	54
<b>Tabel 4. 4</b> Perhitungan Jarak Data dengan <i>Centroid</i> Iterasi ke-0.....	61
<b>Tabel 4. 5</b> <i>Cluster</i> Baru Iterasi ke-0 .....	62
<b>Tabel 4. 6</b> Perhitungan Jarak Data dengan <i>Centroid</i> Iterasi ke-1 .....	67

## DAFTAR RUMUS

	Halaman
<b>Rumus 2.1</b> <i>Centorid Cluster</i> .....	15
<b>Rumus 2.2</b> <i>Euclidean Distance</i> .....	15

## **DAFTAR LAMPIRAN**

- Lampiran 1. Pendukung Penelitian
- Lampiran 2. Daftar Riwayat Hidup
- Lampiran 3. Surat Keterangan Penelitian
- Lampiran 4. Surat Balasan Permohonan Izin Penelitian

# **BAB 1**

## **PENDAHULUAN**

### **1.1. Latar Belakang Masalah**

Teknologi adalah gambaran perkembangan sebuah negara, adanya teknologi informasi secara global informasi dapat diterima dan disalurkan dengan cepat dan tepat sehingga menghasilkan sebuah informasi terkini yang membawa pengaruh dan tindakan kedepanya. Dalam masa digital ini teknologi membawa peranan penting dan sejumlah dampak dalam aktivitas manusia baik dalam dampak yang bersifat positif ataupun sebaliknya, dampak bersifat positif dapat dilihat dan rasakan dengan dimanfaatnya teknologi jika ditemukan beberapa permasalahan yang timbul dalam aktivitas manusia hingga dipermudahahkan dalam penyelesaian masalah dengan solusi-solusi yang tepat.

Dalam sebuah kasus dapat dimanfaatkan teknologi untuk mencapai penyelesaian masalah tersebut, yaitu mengenai pengelompokan data material pada proyek MV Doulos Phos Hotel, diperlukan pengolahan lebih lanjut karena data pembelian material untuk proyek tersebut selama awal pengerjaan hingga selesai pengerjaan proyek belum dimanfaatkan secara baik, data-datanya hanya sebagai *file* dan data belum/tidak diolah menjadi informasi strategi manajemen yang berguna untuk proyek yang akan datang. Data dalam jumlah yang besar tersebut dalam bentuk *statement* bulanan dari toko penyuplai material. Dan saat ini *statement* bulanan tersebut disimpan dalam sebuah gudang berkas, untuk

menemukan data-data tersebut membutuhkan waktu yang cukup lama. Solusi yang dapat dilakukan yakni perlunya pengolahan data-data material tersebut menggunakan cabang ilmu *data mining*. *Data mining* atau penambangan data adalah teknik yang relatif cepat dan mudah untuk menemukan pengetahuan, pola atau relasi antar data, secara otomatis. Suyanto (dalam Fayyad et al. 1996). *Data mining* diketahui secara umum ialah sebuah kegiatan menggali informasi dari data berjumlah besar kemudian diolah secara otomatis untuk memperoleh informasi baru yang bermanfaat untuk menentukan tindakan keputusan untuk kedepannya.

*Data mining* terdapat beberapa metode, diantaranya merupakan metode *clustering*. *Clustering* ialah kegiatan mengelompokkan data yang telah diperoleh ke sejumlah *cluster* yang sudah di tentukan sehingga dalam satu *cluster* memiliki data dengan kemiripan yang minimum. Dengan *clustering* data dalam material proyek MV Doulos Phos Hotel akan dikelompokkan ke dalam beberapa *cluster* yaitu *cluster* tinggi (C1), sedang (C2) dan rendah (C3). Kemudian terdapat tiga variabel yaitu variabel besi, pvc dan ppr. Berdasarkan variabel tersebut *quantity* material dijadikan sebagai hasil acuan prediksi, selanjutnya data-data diolah dengan algoritma *k-means*. Algoritma tersebut merupakan metode untuk perhitungan secara manual, hasil perhitungan diimplementasikan menggunakan sebuah *software data mining* yaitu *RapidMiner*. Dengan memanfaatkan aplikasi *RapidMiner* pihak perusahaan dapat memperoleh hasil pengelompokan material yang tertinggi, sedang dan rendah. Apabila terdapat material dengan skala tinggi, sedang dan rendah pemilik perusahaan dapat memanfaatkan hasil pengelompokan tersebut sebagai acuan dalam menambah dan mengurangi penyetokan material

pada bulan tertentu untuk digunakan proyek hotel yang akan datang, dengan ketersediaan stok material yang mencukupi dan tidak berlebih maka akan membantu pengerjaan proyek selesai dengan tepat waktu dan perusahaan dapat menghemat sejumlah pengeluaran biaya untuk material proyek.

Mv. Doulos Phos ini merupakan kapal bersejarah seberat 6.81GRT yang dibangun pada tahun 1914 oleh *Newport News Ship Building and Drydock Company*, kapal ini tercatat di *Guinness Book of Record* sebagai kapal penumpang tertua di dunia yang masih aktif dari tahun 1914 hingga 2009. Pemilik kapal *BizNaz Resources International Pte Ltd* memutuskan menjadikan MV. Doulos Phos menjadi hotel bintang 5 di darat yang terletak di Lagoi. PT. Starindo Acetech Perkasa merupakan perusahaan yang bergerak dibidang *maintenance, general contractor* dan *supplier*, berlokasi di Batam Center. Pada tahun 2017 mendapatkan proyek pembuatan hotel di MV. Doulos Phos dengan perancangan pipa air (pipa pvc dan pipa ppr), ducting hingga pemasangan air conditioner.

Berdasarkan latar belakang yang terpapar di atas maka penulis tertarik mengangkat penelitian ini yang berjudul **“PENGELOMPOKAN DATA MATERIAL PROYEK MV DOULOS PHOS HOTEL MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING”**

## **1.2. Identifikasi Masalah**

Berdasarkan latar belakang masalah yang terpapar di atas untuk itu identifikasi masalah yang terdapat pada penelitian ini dihasilkan :

1. Banyaknya data pembelian material proyek MV Doulos Phos Hotel yang belum dimanfaatkan dengan maksimal.

2. Data-data dalam skala besar tersebut hanya sebagai arsip, tidak dimanfaatkan untuk strategi kedepannya.
3. Penyetokkan material masih belum teratur karena tidak ada pengolahan data informasi untuk penyetokkan material.

### **1.3. Batasan Masalah**

Batasan masalah dalam penelitian ini terdapat beberapa permasalahan yang perlu dibatasi agar penelitian lebih fokus yaitu sebagai berikut:

1. Tempat pengambilan data di PT. Starindo Acetech Perkasa.
2. Transaksi pembelian material sebagai acuan pengolahan data
3. Pengambilan data melalui *statement* hasil pembelian bulanan dari toko penyuplai material.
4. Data yang diolah adalah data pembelian tahun 2017-2019.
5. Variabel yang digunakan *quantity* material dan harga maaterial.
6. Mengolah data menggunakan algoritma *k-means clustering*
7. *Software data mining* yang digunakan adalah *RapidMiner*.

### **1.4. Rumusan Masalah**

Dari penejelasan latar belakang dapat di rumuskan beberapa masalah, yaitu:

1. Bagaimana algoritma *k-means clustering* mengelompokan data pembelian material proyek MV. Doulos Phos hotel ?

2. Bagaimana algoritma *k-means clustering* mengetahui kelompok material yang paling sering dipakai, sering dipakai dan jarang dipakai dari data-data *cluster* yang telah ditentukan ?
3. Bagaimana mengimplementasikan *RapidMiner* untuk membantu mengelompokan data pembelian ?

### **1.5. Tujuan Penelitian**

Berdasarkan rumusan masalah yang terpapar di atas dapat diperoleh beberapa tujuan penelitian, yaitu:

1. Untuk mengelompokan data pembelian material proyek MV. Doulos Phos hotel dengan algoritma *k-means clustering*.
2. Untuk menentukan kelompok material yang paling sering dipakai, sering dipakai dan jarang dipakai menggunakan algoritma *k-means clustering*.
3. Untuk membantu memudahkan pengelompokan material dengan menggunakan *RapidMiner*.

### **1.6. Manfaat Penelitian**

Manfaat yang diharapkan penelitian ini menjadi dua jenis yaitu manfaat secara teoritis dan manfaat secara praktis, yaitu sebagai berikut:

1. Manfaat Teoritis:
  - a. Secara umum pada penelitian ini dapat digunakan sebagai referensi bagi penelitian yang bersifat sejenis untuk kebutuhan dimasa yang akan datang.

- b. Penelitian ini bisa memberikan sumbangan pemikiran yang sesuai dalam bidang ilmu pada penelitian kedepannya.
- c. Penelitian ini dapat dikembangkan untuk peneliti kedepannya dan sejenisnya.

2. Manfaat secara praktis :

a. Manfaat bagi akademik

Manfaat yang dapat diperoleh dalam akademik, berdasarkan proses penelitian pada pembuatan skripsi ini merupakan sumbangsih laporan proposal skripsi untuk pihak kampus, yang di harapkan dapat digunakan dengan sebaiknya sesuai kebutuhan, serta sebagai tambahan buku di perpustakaan kampus.

b. Bagi Mahasiswa

Skripsi ini dapat dijadikan sumber referensi dan menambah wawasan kepada mahasiswa yang mengambil topik atau melakukan penelitian sejenisnya.

c. Bagi Universitas

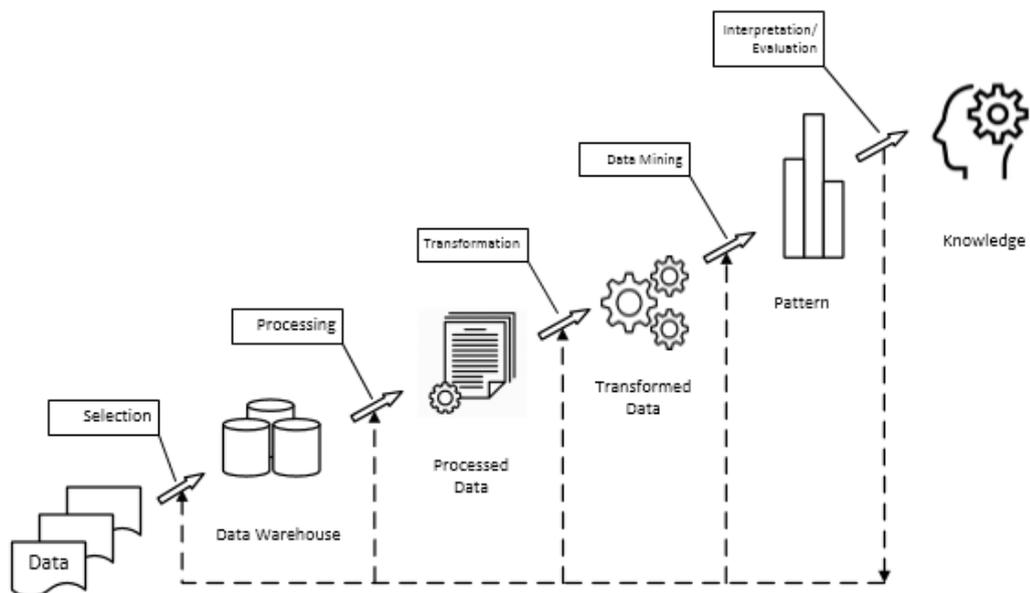
Dari skripsi ini diharapkan dapat menambah referensi untuk bahan penelitian yang lebih lanjut dan lebih mendalam pada masa kedepannya.

## BAB II

### KAJIAN PUSTAKA

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

Istilah *data mining* dan *knowledge discovery in database (KDD)* sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan dengan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Secara garis besar proses KDD sebagai berikut (Nofriansyah, 2014)



**Gambar 2. 1** Proses *Knowledge Discovery In Database*  
**Sumber:** Gambar Peneliti (2020)

1. *Data Selection*, sebelum melakukan tahap penggalian data memerlukan tahap pemrosesan pemilihan data terlebih dahulu dari kumpulan-kumpulan data operasional. Kemudian data dari hasil seleksi/pemilihan akan digunakan untuk proses *data mining*, data tersebut disimpan ke sebuah berkas tersendiri yang berbeda dari berkas basis data operasional agar mempermudah pada penggunaan ke tahapan berikutnya.
2. *Pre-processing/cleansing*, proses di tahap ini berupa memeriksa data yang bersifat non-konsisten, membuang data ganda dan atribut-atribut data yang non-relevan serta memperbaiki kesalahan yang terdapat di data tersebut. Kemudian data yang telah diperoleh melalui database atau berkas di suatu perusahaan, terdapat keterangan atau isian yang kurang lengkap seperti data tersebut tidak valid, data tersebut hilang, atau data tersebut mengalami kesalahan pengetikan, keberadaan data-data tersebut dapat mengakibatkan kurangnya akurasi atau kualitas dari hasil *data mining* berikutnya. *Cleansing* data akan berpengaruh pada performansi melalui sistem *data mining* sebab data akan mengalami pengurangan kompleksitas dan jumlahnya berdasarkan data yang ditangani.
3. *Transformation*, dalam metode *data mining* sebelum bisa diaplikasikan ada beberapa metode yang membutuhkan format data yang khusus. Seperti beberapa metode *clustering* dan metode analisis, metode tersebut hanya bisa menerima data-data input kategorikal. Oleh karena itu, data berupa angka numerik yang berlanjut memerlukan pembagian menjadi beberapa interval. Di tahap ini akan melakukan proses pemilihan data yang dibutuhkan oleh

metode *data mining* yang digunakan, proses pemilihan dan transformasi data ini akan diperoleh kualitas melalui hasil *data mining* sebab terdapat beberapa karakteristik di metode-metode *data mining* tertentu yang bergantung dari tahapan ini.

4. *Data mining*, tahapan ini berupa proses menemukan pola atau mencari informasi menarik yang terdapat pada data-data pilihan menggunakan metode maupun teknik tertentu. Algoritma maupun metode yang terdapat dalam *data mining* memiliki berbagai jenis dan sangat bervariasi, penggunaan algoritma atau metode yang tepat bergantung pada tujuan serta dengan keseluruhan bergantung pada proses KDD. Untuk data yang bisa digunakan menjadi sebuah model yang baik data harus tercukupi sebagai data riset, apabila semakin banyaknya data yang digunakan maka semakin sedikit kesalahan atau *error* sehingga semakin bagus model yang dijadikan sebagai acuan.
5. *Evaluation/Interpretation*, pada tahapan akhir ini proses *data mining* yang menghasilkan pola informasi memerlukan penampilan dalam bentuk yang mudah dimengerti oleh pihak yang membutuhkan atau pihak yang berkepentingan. Tahapan ini juga termasuk cakupan pemeriksaan pola maupun informasi yang ditemukan apakah menimbulkan tentangan dengan hipotesis dan fakta yang telah ada pada sebelumnya.

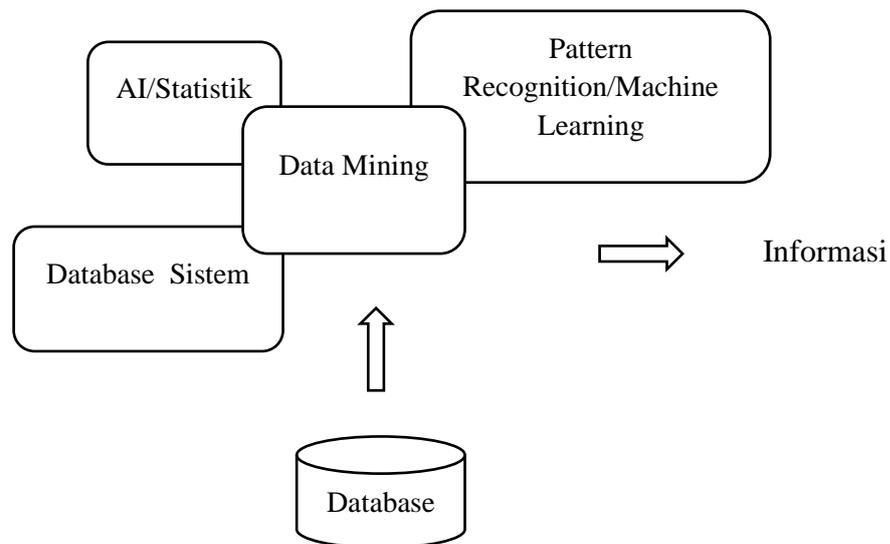
## **2.2. Data Mining**

### **2.2.1. Sejarah Data Mining**

Istilah *data mining* populer sejak tahun 1990-an di komunitas pengguna basis data. Namun, metode dan teori dasar dari *data mining* telah lahir sebelum era 90. *Data mining* berasal melalui berbagai macam disiplin ilmu, disiplin ilmu data mining memiliki dua ilmu yang paling mendasari yakni *machine learning* dan statistik. Teori-teori statistik yang berbasal dari teori matematika menitikberatkan pada pembentukan model. Model merupakan pendekatan struktur atau asumsi yang mendekati data yang sesungguhnya, melainkan *machine learning* ini lebih memprioritaskan pengembangan algoritme. Awal perkembangan *data mining* dimulai tahun (1763) saat Thomas Bayes mempublikasikan Teorema Bayes. Teori ini merupakan teori yang sangat penting dalam *data mining* karena mengestimasi kemungkinan suatu kejadian berdasarkan sebuah kejadian yang telah terjadi atau sedang berlangsung. Pada tahun (1805) mulai berkembangnya teori regresi yang mempelajari hubungan antar variabel, regresi juga menjadi salah satu bagian sebagai alat penting didalam data mining. Penggunaan komputer untuk mengolah data-data dalam jumlah yang besar dimulai saat Alan Turing memperkenalkan ide mesin pengolah data-data yang bersifat universal pada tahun (1936).

### 2.2.2. Asal Ilmu *Data Mining*

Asal mula ilmu *data mining* berawal dari irisan berbagai macam disiplin ilmu pengetahuan antara lain : sistem basis data, kecerdasan buatan atau statistik, serta pattern recognition atau machine learning.



**Gambar 2. 2** Asal Ilmu *Data Mining*

**Sumber:** Gambar Peneliti (2020)

1. *Artificial Intelligence (AI)* atau kecerdasan buatan dan statistik, *AI* adalah salah satu disiplin ilmu yang sangat penting di dalam pembangunan *data mining* dengan teknik pengolahan informasi didasari pada pola pikir atau penalaran manusia. Dengan adanya statistik data yang telah diolah dapat dirangkum ke EDA atau *exploratory data analysis*, *exploratory data analysis* ini berfungsi sebagai pengidentifikasi hubungan antarvariabel yang sistematis atau fitur apabila tidak mencukupi informasi alami yang telah dibawanya.

2. *Pattern Recognition/Machine Learning*, pada dasarnya *data mining* merupakan pengenalan pola tetapi hanya terbatas pada pola basis data. Data yang akan diambil tidak dalam bentuk relasi, melainkan dalam bentuk normal pertama.
3. Sistem basis data, sistem basis data ini menyediakan sejumlah informasi berupa data-data yang akan diolah.

### **2.2.3. Pengertian *Data Mining***

Berbagai macam pendefinisian *data mining*, sebagai berikut:

(Lailil & Dkk, 2018)

- a. Penguraian yang tidak sederhana dari sekumpulan data menjadi informasi yang memiliki potensi secara implisit tidak nyata atau jelas yang sebelumnya tidak diketahui.
- b. Penggalan dan analisis, dengan menggunakan peranti otomatis atau otomatis, dari sejumlah besar data yang bertujuan untuk menemukan pola yang memiliki arti.
- c. *Data mining* juga merupakan bagian dari *knowledge discovery* dalam database *Knowledge Discovery in Database* atau KDD.

*Data mining* ialah sebuah proses penggalan atau penambangan data terhadap data yang tidak dapat diolah dengan cara manual karena jumlah data terlalu besar, menggunakan satu atau beberapa teknik untuk menganalisis sehingga memperoleh informasi yang dapat dimengerti serta berguna untuk

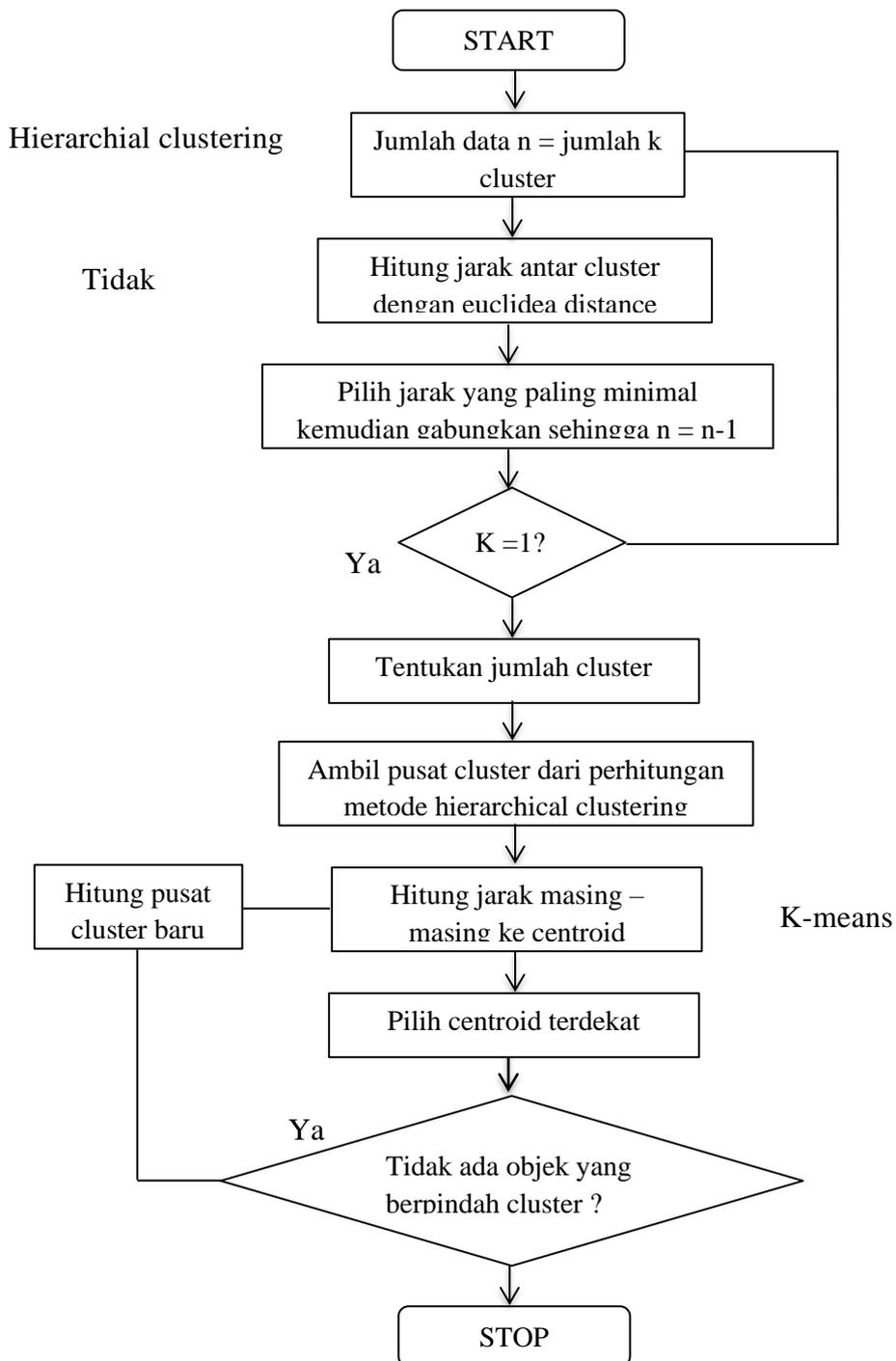
pemilik data yang sebelumnya tidak disadari keberadaan informasi tersebut yang mengandung nilai positif untuk kedepannya.

#### **2.2.4. Algoritma *Data Mining***

##### **2.2.4.1. Algoritma *k-means clustering***

*K-means* merupakan sebuah algoritma *clustering* yang mengalami pengulangan, yang mengelompokkan sejumlah objek menggunakan variabel tertentu ke dalam kelompok-kelompok, pada algoritma *k-means* jumlah *cluster* harus ditentukan terlebih dahulu. Algoritma ini menetapkan nilai-nilai cluster secara random dan sementara nilai cluster tersebut akan menjadi sebuah pusat dari *centroid* atau *means*. Untuk menghitung jarak setiap data pada beberapa *centroid* menggunakan sebuah rumus *Euclidian Distance*, rumus ini digunakan hingga ditemukan jarak paling dekat pada setiap data dengan *centroid* dan perhitungan ini akan dilakukan terus menerus hingga nilai *centroid* tidak lagi berubah.

Berikut merupakan langkah - langkah untuk proses nilai centroid agar nilai centroid tidak berubah:



**Gambar 2. 3** Proses Mencari Nilai *Centroid*  
**Sumber:** Gambar Peneliti (2020)

1. Tentukan nilai  $k$  sebagai jumlah *cluster* yang akan di bentuk, temukan titik pusat *cluster* /  $k$  *centroid* awal secara acak / random dari objek-objek yang ada sebanyak  $k$  *cluster*. Untuk menghitung centroid cluster ke- $i$  selanjutnya menggunakan rumus berikut:

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

**Rumus 2. 1 Centorid Cluster**

Keterangan :

$v$  = *centroid* pada *cluster*

$x_i$  = objek ke- $i$

$n$  = jumlah / banyaknya objek yang menjadi bagian / anggota *cluster*

2. Gunakan rumus *Euclidean Distance* hitung jarak *centroid* setiap objek dari masing-masing *cluster* :

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

**Rumus 2. 2 Euclidean Distance**

Keterangan :

$x_i$  = objek  $x$  ke- $i$

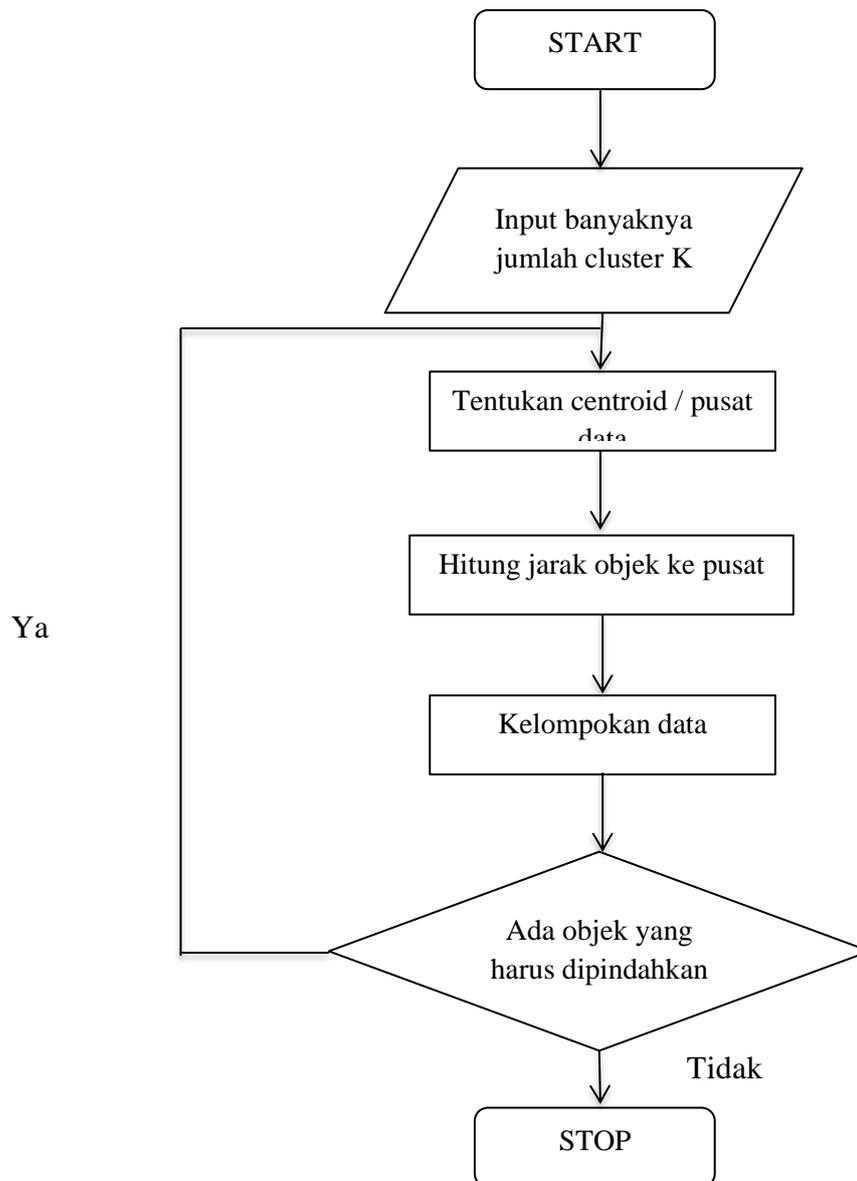
$y_i$  = daya  $y$  ke- $i$

$n$  = banyaknya objek

3. Selanjutnya melakukan pengalokasikan masing-masing objek ke *centroid* yang terdekat, lakukan iterasi dan tentukan posisi *centroid* baru dengan menggunakan persamaan. Lakukan pengulangan perhitungan langkah ke 3, apabila posisi *centroid* baru memiliki nilai yang tidak sama dengan iterasi

sebelumnya. Apabila nilai yang dihasilkan sama dengan iterasi sebelumnya maka iterasi berhenti.

Berikut merupakan *flowchart* dari algoritma *k-means clustering*, sebagai berikut:



**Gambar 2. 4** *Flowchart* Proses *K-means*

**Sumber:** Gambar Peneliti (2020)

Penjelasan langkah-langkah dari gambar 2.4 *flowchart* proses *k-means*, yaitu:

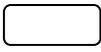
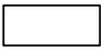
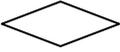
1. Tentukan jumlah *cluster*, langkah pertama pada metode algoritma *k-means* ialah penentuan jumlah *cluster* didasarkan data yang telah diperoleh.
2. Tentukan pusat *cluster* awal, pusat *cluster* awal diperoleh dari data sendiri dengan mereandom atau acak pusat *cluster* awal melalui data yang telah diperoleh.
3. Hitung jarak objek ke pusat *cluster*, menggunakan rumus *Euclidean Distance* berfungsi mengukur jarak antara data dengan pusat *cluster*.

Tahap algoritma perhitungan jarak data dengan pusat cluster yaitu :

1. Ambil nilai data dan nilai pusat *cluster* yang sudah diperoleh
2. Melakukan hitungan menggunakan *Euclidean Distance* data dengan tiap pusat *cluster*
4. Mengelompokan objek-objek data, lakukan perbandingan jarak dari hasil perhitungan kemudian lakukan pemilihan jarak yang terdekat antara data dengan pusat *cluster*, pemilihan jarak ini memperlihatkan data tersebut berada dalam kelompok yang sama atau satu kelompok dengan pusat *cluster* yang terdekat. Tahap algoritma mengelompokan objek-objek data yaitu :
  1. Ambil nilai jarak pada tiap pusat *cluster* pada data-data
  2. Temukan nilai jarak terkecil dari data
  3. Kelompokkan data-data dengan pusat *cluster* yang memiliki jarak terkecil.

5. Tentukan pusat *cluster* yang baru, pusat *cluster* yang baru berfungsi untuk melakukan iterasi berikutnya (apabila hasil tidak konvergen). Proses iterasi akan dihentikan apabila hasil telah konvergen antara pusat *cluster* baru dengan pusat *cluster* lama.
6. Hitung jarak pusat *cluster*, melakukan perhitungan menggunakan *Euclidean Distance* dari semua data ketitik pusat yang baru atau ke C1 dan C2 (yang telah dilakukan pada tahap 2). Setelah memperoleh hasil perhitungan berikutnya bandingkan hasil tersebut.
7. Berikut merupakan penjelasan simbol-simbol yang terdapat pada *flowchart* proses *k-means*, sebagai berikut:

**Tabel 2. 1** Simbol-Simbol *Flowchart* Proses *K-means*

Simbol	Nama	Deskripsi
	<i>Flow Direction Symbol</i>	<i>Flow direction symbol</i> ialah simbol yang digunakan untuk mengkoneksikan atau menghubungkan simbol yang satu dengan simbol yang lainnya.
	Simbol <i>Input-Output</i>	<i>Input-Output symbol</i> ialah simbol yang menandakan proses input dan output data.
	<i>Terminator Symbol</i>	Terminar ialah simbol untuk menyatakan permulaan begin/mulai dan stop/akhir dari suatu proses.
	<i>Processing Symbol</i>	<i>Processing symbol</i> ialah simbol yang menandakan pengolahan yang dijalankan oleh pc.
	<i>Decision Symbol</i>	<i>Decision symbol</i> ialah simbol pemilihan proses kegiatan berdasarkan kondisi yang telah ada.

**Sumber:** Tabel Peneliti (2020)

#### **2.2.4.2. Algoritma apriori**

Algoritma apriori merupakan algoritma mencari pola hubungan atau kombinasi item satu dengan item lainnya dalam sebuah data, kombinasi ini mengandung sebuah nilai keseringan pengambilan antara item satu dengan item yang lain. Hasil dari algoritma apriori ini dapat dimanfaatkan untuk membantu pihak yang menyediakan suatu barang dalam pengambilan keputusan dalam manajemen penempatan barang mau pun strategi pemasaran berupa diskon untuk kombinasi barang tersebut.

#### **2.2.4.3. Algoritma *nearest neighbor***

Algoritma *Nearest Neighbor* merupakan algoritma yang mencari kasus dengan cara menghitung kedekatan antara kasus yang lama dengan kasus yang baru, jadi kasus lama digunakan sebagai pedoman untuk menghasilkan sebuah hasil untuk pemecahan masalah pada kasus baru, berdasarkan pencocokan-pencocokan bobot dari sejumlah fitur yang telah ada.

#### **2.2.4.4. Operasi dasar *data mining***

Operasi dasar dalam *data mining* dikategorikan menjadi dua yaitu metode prediktif dan metode deskriptif. Metode prediktif merupakan metode yang mempunyai tujuan dalam memprediksi atau memperkirakan nilai suatu variabel tertentu berdasarkan pada nilai variabel-variabel lain. Variabel tak bebas merupakan variabel yang dijadikan target untuk diprediksi, sedangkan variabel

bebas merupakan variabel-variabel yang dijadikan untuk membantu prediksi. Regresi dan klasifikasi termasuk dalam metode prediktif.

Regresi bertujuan untuk melakukan tugas prediksi, regresi ini memiliki konsep dasar pada *data mining* yang berasal dari teori statistika. Regresi akan mengidentifikasi beberapa relasi antar variabel terikat dengan variabel bebas, setelah menghasilkan suatu model matematika berdasarkan identifikasi relasi tersebut, hasil dari model matematika digunakan untuk memperkirakan nilai melalui suatu variabel terikat berdasarkan nilai variabel bebasnya. Dalam *data mining* klasifikasi mempunyai tujuan untuk mengelompokkan data-data menjadi sejumlah kelompok, pengelompokan ini memanfaatkan proses acuan data yang sudah diketahui kelas atau kelompoknya. Apabila terdapat data yang belum memiliki kelompok maka dapat ditentukan kelompoknya menggunakan proses perbandingan melalui data yang sudah diketahui kelompoknya.

Metode deskriptif merupakan metode yang memiliki tujuan untuk menemukan relasi atau pola melalui data yang mudah dipahami oleh manusia, *association rule* dan *clustering* termasuk dalam metode deskriptif.

*association rule* adalah metode pencarian relasi serta pola antar data dalam sekumpulan data-data, pola yang dihasilkan suatu data dapat diprediksi kemunculannya berdasarkan kemunculan data lainnya. Dan *clustering* bertujuan untuk membagi data-data ke dalam beberapa kelompok, pada *clustering* proses pengelompokkan data tidak membanding data-data lain yang telah diketahui kelompoknya. Data-data pada *clustering* dikelompokkan dengan cara membandingkan seluruh data yang belum memiliki kelompok kemudian membagi

data-data tersebut kedalam beberapa kelompok yang memiliki kemiripan antar data.

#### **2.2.4.5. Tantangan dalam *data mining***

Dalam data mining memiliki tantangan yang akan dihadapi peneliti meliputi:

1. Data mining memiliki data yang kompleks, dalam kesatuan data terdiri dari beberapa bagian. Bagian tersebut saling bergantung dan saling berhubungan.
2. Heterogen data dalam data mining, memiliki beraneka ragam data dan karakteristik yang tidak sama diantara satu data dengan data yang lainnya.
3. Data mining memiliki jumlah variabel dan *cluster* yang banyak dalam data yang akan diproses.
4. Skalabilitas dalam data mining, kemampuan menampung data yang berskala besar.
5. Distribusi dalam data mining, data dalam jumlah yang besar akan dilakukan proses pembagian data dengan metode data mining.
6. Kepemilikan data, data akan diperoleh dari pihak yang memiliki data sesuai dengan topik peneliti.
7. Data mining membutuhkan proteksi data apabila terdapat data yang sensitif.

#### **2.2.4.6. Etika dalam *data mining***

*Data mining* akan memberikan dampak positif maupun sebaliknya yang sangat bergantung pada penggunaannya, dampak negatif akan muncul apabila tidak diperhatikan etika dalam penggunaan data khususnya data-data yang berhubungan dengan data pelanggan yang bersifat pribadi. Seperti klasterisasi pelanggan berdasarkan golongan, ras, agama, suku bangsa, adat, usia maupun gender jika tidak bijak dalam penggunaan data tersebut akan menimbulkan masalah diskriminasi dan bisa berakibat merugikan kelompok-kelompok tertentu. Akan tetapi, pembeda masalah usia atau gender tertentu akan memberikan efek yang positif jika diterapkan dalam masalah medis seperti beberapa jenis penyakit yang rentan diderita oleh kelompok usia tertentu atau kaum pria maupun wanita.

#### **2.2.4.7. Manfaat *data mining***

Pemanfaatan data mining dilihat dari dua sudut pandang, yaitu sudut pandang komersial dan sudut pandang keilmuan. (Vulandari, 2017)

Sudut pandang di lihat dari sudut komersial, data mining dimanfaatkan untuk mengatasi penumpukan data yang berlebihan, bagaimana memanfaatkan dan menyimpan data-data tersebut. Adanya teknik komputasi dapat dimanfaatkan untuk menghasilkan informasi yang digunakan sebagai aset untuk meningkatkan kompetitif suatu perusahaan. Seperti, bagaimana memprediksi tingkat pembelian dan penjualan, mengetahui konsumen dan produk yang memiliki kesamaan atau kemiripan karakteristik, mengetahui bagaimana mengidentifikasi produk-produk yang telah terjual secara bersamaan dalam satu waktu dengan produk lainnya,

mengetahui menyusutnya pelanggan akibat persaingan, memprediksi perilaku bisnis dimasa depan, memprediksi tingkatan resiko dalam menentukan jumlah produksi.

Data mining dalam sudut pandang keilmuan dapat dimanfaatkan untuk menyimpan dan menganalisa data yang sangat besar serta bersifat *real time*. Seperti, pemindaian langit dengan menggunakan teleskop, *remote sensor* yang ditempatkan pada *remote TV*.

### **2.3. Metode Data Mining**

Terdapat sejumlah fungsi atau metode dalam *data mining* yang dapat digunakan dalam menemukan, menambang dan menggali informasi pengetahuan. Dalam metode *data mining* terdapat 5 metode utama, yaitu:

#### **2.3.1. Clustering**

*Clustering* yaitu metode pengelompokan yang membentuk kelompok objek dengan memiliki kesamaan atau kemiripan dengan objek-objek lainnya, dan membentuk kelompok objek lainnya jika memiliki ketidaksamaan atau ketidakmiripan. *Clustering* ini melakukan proses pembagian keseluruhan data-data menjadi kelompok yang terdapat kesamaan atau kemiripan.

#### **2.3.2. Estimasi**

Metode estimasi ini melakukan estimasi sebuah data atau nilai baru yang belum diketahui atau yang tidak memiliki keputusan berdasarkan histori data atau nilai yang telah ada, seperti melakukan estimasi terhadap pendapatan atau

penghasilan seseorang apabila informasi mengenai orang tersebut telah diketahui.

### **2.3.3. Prediksi**

Metode prediksi ini digunakan untuk memperkirakan suatu kejadian atau sebuah peristiwa tertentu terjadi atau memperkirakan suatu kejadian atau sebuah peristiwa yang memiliki nilai masa yang akan datang, seperti memprediksi saham satu tahun ke depan.

### **2.3.4. Klasifikasi**

Metode klasifikasi merupakan proses penemuan model atau sebuah fungsi yang membedakan maupun menjelaskan kelas data atau konsep, yang memiliki tujuan untuk memperoleh perkiraan kelas melalui suatu objek yang labelnya tidak diketahui.

### **2.3.5. Asosiasi**

Metode ini berfungsi untuk menemukan hubungan antar variable-variabel pada sebuah database yang berjumlah banyak. Metode ini secara umum disebut *Market Basket Analysis* atau analisis keranjang belanja, dimana mengidentifikasi hubungan yang bersangkutan kuat antar berbagai jenis produk yang akan diambil bersamaan dalam setiap pembelian.

## **2.4. *Software Data Mining***

Pada saat ini telah banyak *software* atau aplikasi *data mining* yang dapat dimanfaatkan untuk mempermudah pengguna untuk mendapatkan informasi, mulai

dari *software* atau aplikasi yang gratis hingga yang berbayar. Dibawah ini merupakan *software* atau aplikasi gratis terdiri dari :

1. *RapidMiner*



**Gambar 2. 5** Logo *RapidMiner*  
**Sumber:** rapidminer.com

Pada awalnya *RapidMiner* dikenal dengan *Yet Another Learning Environment* atau YALE. Pada tahun 2001 *RapidMiner* dikembangkan bersifat *Open Source* oleh Simon Fischer, Ralf Klinkenberg dan Ingo Mierswa. *RapidMiner* mampu bekerja di semua SO/sistem operasi yang ditulis dalam bahasa *Java*. *RapidMiner* sebagai *software open source* untuk *data mining* yang sudah terkemuka di dunia. *RapidMiner* menempati peringkat pertama sebagai *software* data mining pada *polling* oleh KDnuggets, sebuah portal *data mining* pada 2010-2011. *RapidMiner* memberikan solusi dalam melakukan analisis prediksi, dan menggunakan beberapa teknik prediksi dan deskriptif dalam memberikan wawasan untuk para pengguna sehingga dapat mengambil keputusan yang terbaik dari hasil analisa. (Aprillia c & Dkk, 2013)

2. *Weka (Waikato Enviroment for Knowledge Analysis)*



**Gambar 2. 6** Logo *Weka*  
**Sumber :** [analyticsinsight.net](http://analyticsinsight.net)

*Weka* merupakan *software* atau aplikasi *data mining* dengan tampilan yang sederhana bersifat *open source*, ditulis menggunakan bahasa Java. *Weka* pertama kali rilis pada 14 april 2016 oleh Universitas Waikato, *software* ini mendukung OS X, Linux dan *Windows*. *Weka* memiliki lingkungan untuk membandingkan beberapa algoritma pembelajaran, *weka* juga memiliki fasilitas untuk menganalisis data seperti perangkat-perangkat *pre-processing* data, metode-metode evaluasi serta algoritma pembelajaran.

3. *Orange*



**Gambar 2. 7** Logo *Orange*  
**Sumber:** [github.com](http://github.com)

*Orange* merupakan *software data mining* yang bersifat *open source* ditulis dalam bahasa *python*, *cython*, C++ dan C. *Orange* rilis pada 10 oktober 1996 oleh Universitas Ljubljana, yang mendukung pada *macOS*, *Linux* dan *Windows*. *Orange* memiliki keunggulan dalam hal visualisasi dapat digunakan dengan mudah tidak hanya untuk digunakan oleh para ahli, tetapi dapat digunakan oleh kalangan umum dan para pemula. Dengan menggunakan *software orange* dapat menganalisis suatu data-data penelitian, teks opini dari masyarakat, teks program kerja, membaca data, membandingkan algoritma pembelajaran, memvisualisasikan elemen data dan seterusnya.

#### 4. *Rattle GUI*



**Gambar 2. 8** Logo *Rattle GUI*  
**Sumber:** redbubble.com

*Rattle GUI* dikembangkan oleh Graham Williams, dirilis pada 5 september 2017, *Rattle GUI* adalah *software open source* yang terdapat *user interface* dengan menggunakan bahasa pemrograman statistik R. *Rattle GUI* mampu menyajikan ringkasan-ringkasan data serta menghasilkan visualisasi

data melalui antarmuka pengguna grafis, *Rattle GUI* juga bisa dimanfaatkan sebagai fasilitas pengajaran untuk mempelajari dan mendalami bahasa perangkat lunak R. *Rattle GUI* menawarkan kemudahan untuk pengguna dalam melakukan penambangan data tanpa harus *coding*. Saat ini memiliki sejumlah perusahaan yang telah menggunakan aplikasi *Rattle GUI* untuk penambangan data salah satunya *Bank Commonwealth*.

5. *R studio*



**Gambar 2. 9** Logo *R Studio*  
**Sumber:** [rstudio.com](http://rstudio.com)

*R studio* merupakan *software open source* yang tersedia untuk *Linux*, *Windows*, *MacOS* yang dikembangkan oleh *RStudio, Inc.* *R studio* merupakan *Integrated Development Environment* (IDE) dari bahasa pemrograman R yang berbahasa pemrograman bersifat standar untuk komputasi statistik dan grafik. *R studio* mewajibkan pengguna untuk menggunakan baris kode dalam melakukan analisis, berbeda dengan *software – software* seperti *Orange*, *Weka* dan lain sebagainya yang hanya mengklik untuk melakukan analisis.

## 2.5. Tujuan Umum Penelitian

Dari penelitian ini memiliki beberapa tujuan yaitu untuk menambah pengetahuan dan wawasan mengenai *data mining k-means clustering* yang bisa diterapkan untuk memberikan solusi dalam memecahkan masalah pada dunia kerja, kehidupan sehari-hari dan sebagainya.

## 2.6. Penelitian Terdahulu

Penelitian terdahulu sebagai salah satu alat bantu atau referensi penulis yang berguna untuk memperluas kajian-kajian yang dilakukan dalam penelitian.

Berikut beberapa referensinya :

1. (Windarto, 2017) dengan jurnal ISSN 1412-2693 yang berjudul **“Penerapan Datamining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan *K-Means Clustering Method*”**, Indonesia adalah salah satu negara pengekspor ke negara-negara maju dan berkembang. Tujuan dari eksportir adalah untuk dapat memperoleh keuntungan. Penelitian ini membahas tentang Penerapan Datamining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan *K-Means Clustering Method*. Sumber data penelitian ini dikumpulkan berdasarkan dokumen-dokumen keterangan ekspor impor yang dihasilkan oleh Direktorat Jenderal Bea dan Cukai. Data yang digunakan dalam penelitian ini adalah data dari tahun 2002- 2015 yang terdiri dari 11. Variable yang digunakan (1) jumlah ekspor berat bersih (*netto*) dan (2) nilai *Free On Board (FOB)*. Data akan diolah dengan melakukan clustering dalam 3 *cluster* yaitu *cluster* tingkat

ekspor tinggi, *cluster* tingkat ekspor sedang dan cluster tingkat ekspor rendah. *Centroid* data untuk *cluster* tingkat ekspor tinggi 904.276,5, *Centroid* data untuk *cluster* tingkat ekspor sedang 265.501 dan *Centroid* data untuk *cluster* tingkat ekspor rendah 34.280,1. Sehingga diperoleh penilaian berdasarkan indeks ekspor buah-buahan dengan 2 negara cluster tingkat ekspor tinggi yakni India dan Pakistan, 3 negara *cluster* tingkat ekspor sedang yakni Singapura, Bangladesh dan Negara lainnya dan 6 negara *cluster* tingkat ekspor rendah yakni Hongkong, Tiongkok, Malaysia, Nepal, Vietnam dan Iran. Hal ini dapat menjadi masukan kepada pemerintah, negara yang menjadi prioritas tertinggi pada kegiatan ekspor buah-buahan berdasarkan klaster yang telah dilakukan.

2. (Novita Sari & Dkk, 2018) dengan jurnal ISSN 2407-1811 yang berjudul **“Penerapan Metode K-Means Clustering Dalam Menentukan Predikat Kelulusan Mahasiswa Untuk Menganalisa Kualitas Lulusan”**, Semakin meningkatnya jumlah mahasiswa yang diluluskan setiap tahunnya menyebabkan banyaknya data mahasiswa yang perlu diolah sehingga menyebabkan kesulitan dalam pengelompokan data tersebut. Pada penelitian ini menerapkan Data Mining dengan menggunakan metode *Clustering* untuk mengelompokkan kualitas lulusan mahasiswa Fakultas Ilmu Komputer Universitas Dehasen Bengkulu berdasarkan IPK dan Program Studi. Algoritma yang digunakan yaitu *K-Means Clustering*, dimana data dikelompokkan berdasarkan karakteristik yang sama akan dimasukkan ke dalam kelompok yang sama dan set data yang dimasukkan

ke dalam kelompok tidak tumpang tindih. Informasi yang ditampilkan berupa kelompok – kelompok lulusan mahasiswa yang mendominasi Program Studi, sehingga diketahui kelompok yang memiliki kualitas lulusan terbaik. Hasil penelitian ini akan membantu pihak Universitas dalam menganalisa kualitas mahasiswa yang diluluskan dan program studi yang paling berpotensi diminati. *Software* yang digunakan untuk membantu pengelompokan ini adalah *Rapid Miner*.

3. (Siregar, 2018) jurnal dengan ISSN 2622-1659 yang berjudul **“Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode *K-Means* (Studi Kasus Di Toko Adi Bangunan)”**. Dalam persaingan dunia bisnis saat ini, kita dituntut untuk senantiasa mengembangkan bisnis agar selalu bertahan dalam persaingan. Untuk mencapai hal tersebut, ada beberapa hal yang bisa dilakukan yaitu dengan meningkatkan kualitas produk, penambahan jenis produk, dan pengurangan biaya operasional perusahaan dengan cara menggunakan analisis data perusahaan. Data mining adalah sebuah teknologi yang mengotomatisasi proses untuk menemukan pola menarik dan sensitif dari kumpulan-kumpulan data yang besar. Ini memungkinkan pemahaman manusia tentang menemukan pola dan skalabilitas teknik. Toko Adi Bangunan merupakan sebuah toko yang bergerak dalam bidang penjualan bahan-bahan bangunan dan peralatan rumah tangga yang memiliki sistem seperti pada swalayan yaitu pembeli mengambil sendiri barang yang akan dibeli. Data-data penjualan, pembelian barang maupun pengeluaran tidak

terduga tidak tersusun dengan baik, sehingga data tersebut hanya berfungsi sebagai arsip bagi toko dan tidak dapat dimanfaatkan untuk pengembangan strategi pemasaran. Pada penelitian ini, data mining diterapkan menggunakan model proses *K-Means* yang menyediakan proses standar penggunaan *data mining* pada berbagai bidang digunakan dalam klasifikasi karena hasil metode ini mudah dipahami dan diinterpretasikan.

4. (Mardalius, 2018) jurnal dengan ISSN 2407-1811 yang berjudul **“Pemanfaatan *Rapid Miner Studio 8.2* Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma *K-Means*”**. Toko Rafadel Acc menjual berbagai jenis aksesoris yang tersedia yang dijual di toko tersebut. Dari berbagai jenis aksesoris yang dijual tentu tidak semuanya yang laku terjual dan juga ada yang kurang laku serta ada juga yang tidak pernah terjual sama sekali. Dengan adanya masalah ini maka kita perlu melakukan perhitungan untuk menentukan atau mengelompokkan mana kategori aksesoris yang laku, kurang laku dan tidak laku terjual, dalam proses pengelompokan maka akan digunakan sebuah metode pengelompokan menggunakan Algoritma *K-Means Clustering* sebagai metode perhitungan secara manual dan dalam implementasinya maka digunakan sebuah *software Data Mining* menggunakan *RapidMiner Studio* versi 8.2. Dengan adanya aplikasi *RapidMiner Studio* ini pemilik toko dapat melihat hasil pengelompokan aksesoris mana yang paling laku, laku dan kurang laku. Maka, bila terdapat produk yang tidak laku, pemilik toko dapat mencari alternative

lain agar aksesoris yang tidak laku dapat menjadi laku. Metode yang digunakan dalam pengumpulan data adalah observasi dan wawancara kepada pemilik toko Rafadel Acc.

5. (Handoko & Lemana, 2018) jurnal dengan ISSN 2621-8794 yang berjudul **“Pengelompokan *Data Mining* Pada Jumlah Penumpang di Bandara Hang Nadim”**. *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 ata 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.*
6. (Purba & Dkk, 2018) jurnal dengan doi: 10.1088/1742-6596/1007/1/012049 yang berjudul **“ *The effect of mining data k-means*”**

*clustering toward students profile model drop out potential". The high of student success and the low of student failure can reflect the quality of a college. One of the factors of fail students was drop out. To solve the problem, so mining data with K-means Clustering was applied. K-Means Clustering method would be implemented to clustering the drop out students potentially. Firstly the the result data would be clustering to get the information of all students condition. Based on the model taken was found that students who potentially drop out because of the unexcit ing students in learning, unsupported parents, diffident students and less of students behavior time. The result of process of K-Means Clustering could known that students who more potentially drop out were in Cluster 1 caused Credit Total System, Quality Total, and the lowest Grade Point Average (GPA) compared between cluster 2 and 3.*

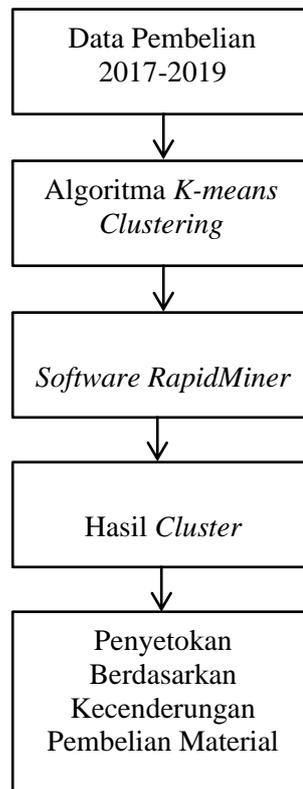
7. (Pernia & Dkk, 2018) jurnal dengan doi: 10.1088/1757-899X/364/1/012045 yang berjudul ***"A data mining approach for indoor air assessment, an alternative tool for cultural heritage conservation"***. *The exposure of cultural heritage to the environment has a significant impact on its degradation process and degradation rate. Consequently, managing the indoor air quality is vital to minimize further damage to historical artefacts and works of art. Despite its potential impact, the traditional assessment of the indoor air quality still represents a challenge for most collection guardians. This approach typically relays on the comparison of measured environmental parameters and corresponding*

*acceptable values. However, determining the acceptable values and relative importance of the different environmental parameters turns out to be quite complex since it depends on the material types present in the collection and their preservation state. Furthermore, the significant amount of data generated during the measurements hampers the application of traditional methods of analysis. Considering all these, we propose the use of data mining as an alternative method for the indoor air quality assessment in cultural heritage studies. Data mining can provide knowledge from vast volumes of heterogeneous data, through high-speed processing, detection, and analysis. Here we present its application to identify dynamics and patterns affecting the indoor air quality in a realistic case. Using data from a measuring campaign held at a late Gothic church in Belgium, we show that inappropriate periods can be identified without using standards. In addition, different types of periods can be identified by studying the relation between multiple parameters. For that we use the k-means clustering method, interpreting the results with both visual and statistical tools.*

## **2.7. Kerangka Pemikiran**

Kerangka pemikiran ialah gambaran alur berpikir yang dibangun dari dasar teori dan referensi-referensi yang mengarahkan peneliti sampai pada dugaan sementara dari pemecahan masalah yang telah dirumuskan. Bagian ini berisi bukti-bukti empiris, teori-teori, dan pemikiran logis dari peneliti. Kerangka pemikiran akan lebih secara operasional apabila dilengkapi dengan *flow chart* atau

diagram alir yang menggambarkan rangkaian alur berpikir peneliti. Deskripsi kerangka pemikiran ialah landasan bagi perumusan hipotesis. (Toto & Dkk, 2015)



**Gambar 2. 10** Kerangka Pemikiran  
**Sumber:** Gambar Peneliti (2020)

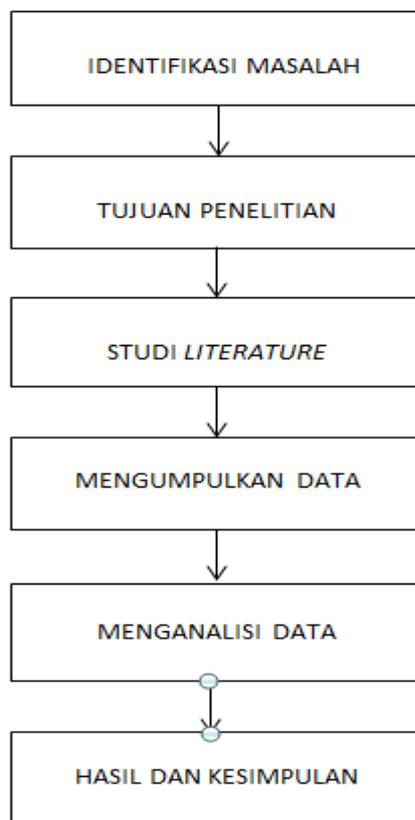
Berdasarkan kerangka pemikiran diatas, data pembelian yang telah dikumpulkan dari tahun 2017 hingga 2019 serta telah diseleksi dan ditransformasi kemudian diolah menggunakan metode algoritma *k-means clustering*. Setelah itu, hasil data dari *k-means clustering* diimplementasikan kedalam *software* atau aplikasi *RapidMiner* untuk menemukan *cluster* pada data, setelah ditemukan *cluster* dapat ditentukan kecenderungan pembelian material yang akan digunakan untuk memprediksi penyetakan material pada proyek sejenisnya dimasa yang akan datang.

## BAB III

### METODE PENELITIAN

#### 3.1. Desain Penelitian

Penelitian *clustering* ini memiliki beberapa tahapan, tahapannya sebagai berikut :



**Gambar 3. 1** Desain Penelitian  
**Sumber:** Gambar Peneliti (2020)

Berikut ini merupakan beberapa penjelasan dari desain penelitian yang terdapat pada gambar 3.1, yaitu:

1. Identifikasi Masalah

Memaparkan permasalahan-permasalahan di PT. Starindo Acetech Perkasa berdasarkan latar belakang masalah.

2. Menentukan Tujuan

Menentukan tujuan yang akan dicapai dari penelitian di PT. Starindo Acetech Perkasa yang didasari latar belakang masalah

3. Studi *Literatur*

Dalam memperkuat pemaparan dalam penelitian, peneliti melakukan pencarian informasi yang mendukung dan membantu kelancaran proses penelitian, mengumpulkan dan mempelajari dari berbagai sumber buku, jurnal, maupun website resmi yang memiliki hubungan langsung dengan objek yang diteliti.

4. Mengumpulkan Data

Peneliti mengumpulkan data-data di PT. Starindo Acetech Perkasa dengan menggunakan metode wawancara, observasi, studi pustaka, dan dokumentasi dalam bentuk berupa catatan dan foto di PT. Starindo Acetech Perkasa.

5. Menganalisis Data

Data yang telah diperoleh akan dilakukan ke tahapan analisis, data yang di analisis berupa data pembelian material tahun 2017 (Juli) hingga 2019 (Oktober). Analisis data tersebut menggunakan aplikasi *Excel* dan *RapidMiner*.

## 6. Hasil dan Kesimpulan

Tahap akhir dari judul yang diangkat dalam penelitian ini dijelaskan dalam bentuk hasil pengelompokan data dan kesimpulan.

### 3.2. Pengumpulan Data

Dalam penelitian ini untuk memperoleh data-data yang sesuai dan dapat dipertanggungjawabkan, maka diperlukan teknik dalam mengumpulkan data-data. Terdapat beberapa teknik yang peneliti gunakan dalam mengumpulkan data yaitu, antara lain:

#### 1. Penelitian lapangan

Penelitian lapangan yang dilakukan dalam penelitian ini sebagai berikut:

##### a. Wawancara

Metode pengumpulan data ini dengan dilakukan tanya jawab secara langsung kepada narasumber. Dalam hal ini, narasumber yang peneliti wawancara adalah Ibu Liany, sebagai *Manajer Accounting* PT. Starindo Acetech Perkasa.

##### b. Observasi

Metode observasi ini dilakukan pengamatan terhadap objek yang akan diteliti. Dalam hal ini, objek yang peneliti observasi meliputi data-data material yang disimpan, waktu tersedianya material dan tempat penyetakan material.

#### 2. Penelitian Perpustakaan

##### a. Studi Pustaka dan Dokumentasi

Metode pengumpulan data dengan dilakukan mengumpulkan data dan mempelajari masalah yang berhubungan dengan objek yang diteliti. Dalam hal ini, peneliti berpatokan atau bersumber pada buku, jurnal, ataupun contoh kasus yang berhubungan dengan penelitian disertai dengan dokumentasi berupa foto dan catatan dengan narasumber.

### 3.3. Operasional Data

PT. Starindo Acetech Perkasa merupakan perusahaan yang menjalankan bidang *general contractor*, *supplier* dan *maintenance*. Melalui hasil wawancara dengan Ibu Liany selaku *Manager Accounting* di PT. Starindo Acetech Perkasa, penelitian ini mengambil 3 variabel yaitu variabel besi, pvc dan ppr pada data pembelian material di PT. Starindo Acetech Perkasa berdasarkan data pembelian untuk proyek *MV Doulos Phos Hotel*, berdasarkan variabel tersebut sebagai acuan dalam pengambilan data pembelian sebagai berikut:

**Tabel 3. 1** Operasional Data Pembelian Material

No	Pengelompokan Material
1	<i>Quantity</i> pembelian

**Sumber:** Data Peneliti (2020)

*Quantity* pembelian dijadikan sebagai acuan data dalam algoritma *k-means*, dilihat dari seberapa banyak kali pembelian jenis barang besi, pvc dan ppr dalam kurun waktu perbulan.

### 3.4. Metode Perancangan Sistem

Metode yang digunakan dalam penelitian ini dalam perancangan data yaitu sebagai berikut:

### 1. *Data Selection*

Sumber data diperoleh dari data pembelian Juli 2017- Oktober 2019 pada PT. Starindo Acetech Perkasa, data terlampir berupa *statement* yang terdiri dari *invoice-invoice* yang terdiri dari gabungan beberapa proyek selain proyek *MV Doulos Phos Hotel*. Data tersebut juga terdiri dari identitas-identitas seperti nomor *invoice*, nomor *delivery order*, nomor *purchase order*, tanggal, *term of payment*, *salesman*, *currency*, nama barang, *quantity* barang, harga, diskon, jumlah bersih dan sebagainya. Dalam data tersebut diseleksi berdasarkan data yang dapat digunakan dan data yang tidak dapat digunakan, pengambilan data berupa data material yang memiliki identitas yang jelas dan data material yang memiliki hubungan dengan proyek *MV Doulos Phos Hotel* dan identitas yang di ambil berupa nama barang, *quantity* barang dan harga barang.

**Tabel 3. 2** Data Rekapitulasi Pembelian Material *MV Doulos Phos Hotel*

Nama Barang	Qty	Harga Peritem	Harga Total	Tanggal
cutting torch	2pc	250000/pc	500000	24 juli 2017
cutting nozzle pnme	2pc	30000/pc	60000	24 juli 2017
tee giv 1 1/2"	10pc	28000/pc	280000	25-Aug-17
45 elbow rucika 1 1/4" d	60pc	2200/pc	132000	25-Aug-17
clamp seng 1"	26pc	5000/pc	130000	30-Sep-17
clamp seng 1/2"	100pc	4000/pc	400000	30-Sep-17
rucika tee 1 1/4" x 3/4" aw	1pc	9000/pc	9000	13-Oct-17
rucika tee 2" x 3/4" aw	1pc	13000/pc	13000	13-Oct-17
rucika elbow 4" rd 45'	2pc	25000/pc	50000	21-Nov-17
gaston berent 18" bt-1164	1pc	155000/pc	155000	21-Nov-17
pipa pn 10 1/2"	125btg	63900/btg	3767100	13-Dec-17
pipa pn 10 3/4"	30btg	88900/btg	4632750	13-Dec-17

**Sumber:** Data Peneliti (2020)





**Tabel 3.2 Lanjutan**

sort flange a150n #150 x 4" aff	4pc	160000/pc	640000	22-Jan-18
PLFF flange ss400 pn16 4" aff	24pc	140000/pc	3360000	22-Jan-18
equal elbow 90' 1" x 1" ppr	30pc	15900/pc	276660	1-Feb-18
equal elbow 90' 1/2" x 1/2" ppr	60pc	7900/pc	274920	1-Feb-18
pipa trilliun (5.8m) aw 4"	10btg	441000/btg	3748500	12-Mar-18
pipa trilliun (5.8m) aw 6"	5btg	980000/btg	4165000	12-Mar-18
socket giv 1 1/4"	30pc	13000/pc	390000	18-Apr-18
union giv 1 1/2"	6pc	45000/pc	270000	20-Apr-18
d/d 200 x 200mm	55pc	13440/pc	739200	12-May-18
bar grill 650 x 450mm	1pc	49560/pc	49560	12-May-18
3 way tee trilliun 20mm	60pc	3300/pc	198000	22-Jun-18
joining coupling trilliun 20mm	100pc	500/pc	50000	22-Jun-18
303 stop valve 1 1/2"	1pc	838900/pc	486562	6-Jul-18
303 stop valve 1/2"	8pc	210900/pc	978576	6-Jul-18
rucika tee 3/4" raw	1pc	4000/pc	4000	24-Aug-18
rucika dop mati 3/4" aw	1pc	2000/pc	2000	24-Aug-18
klem selang s/s 3/8"	4pc	2500/pc	10000	24-Sep-18
klem selang s/s 3/4"	2pc	3000/pc	6000	24-Sep-18
baut stainless m12 x 40	4pc	9000/pc	36000	21-Oct-18
washer stainless m12 100's	8pc	1500/pc	12000	21-Oct-18
rucika elbow 1/2" aw	6pc	3000/pc	18000	11-Nov-18
klem selang s/s 5/8"	8pc	2500/pc	20000	11-Nov-18
equal staright union 1/2"	10pc	5900/pc	34220	3-Dec-18
male staright union 1/2"	10pc	43900/pc	254620	3-Dec-18
cable link 006	6pc	2000/pc	12000	8-Jan-19
cable link 016	4pc	4000/pc	16000	8-Jan-19
baut besi 3/8" x 1" - 13"	0.60kg	35000/kg	21000	7-Mar-19
mata obeng voss 65mm	2pc	10000/pc	20000	15-Mar-19
rucika tee 3/4" aw	8pc	5000/pc	40000	6-Apr-19
rucika tee 1" x 3/4"	8pc	7000/pc	56000	6-Apr-19
trangking pvc 2" x 2"	4btg	20000/btg	80000	31-May-19
baut kembang m10x 60	10pc	2500/pc	25000	30-May-19
ac rxq20aym	1pc	9600000/pc	9600000	4-Oct-19
anti-corrision for rxq20aym	1pc	1100000/pc	1100000	4-Oct-19

**Sumber:** Data Peneliti (2020)

## 2. *Pre-processing / Cleaning*

Setelah data yang diperoleh terseleksi, selanjutnya dibutuhkan tahap *cleaning* agar lebih fokus pada poin yang dituju. Proses tahap ini melakukan pemeriksaan data yang konsisten, membuang data-data yang duplikasi dan memperbaiki kesalahan pada data.

**Tabel 3. 3** Data Setelah *Cleaning*

No	Tanggal	Besi	Pvc	Ppr
1	Juli 2017	9	0	0
2	Agustus 2017	49	44	42
3	September 2018	24	11	28
4	Oktober 2017	102	77	16
5	Nopember 2017	91	54	1
6	Desember 2017	56	29	34
7	Januari 2018	33	13	32
8	Februari 2018	48	35	17
9	Maret 2018	83	80	11
10	April 2018	40	36	11
11	Mei 2018	11	21	16
12	Juni 2018	4	14	8
13	Juli 2018	8	15	8
14	Agustus 2018	23	43	7
15	September 2018	34	22	0
16	Oktober 2018	12	33	5
17	Nopember 2018	29	36	0
18	Desember 2018	1	1	3
19	Januari 2019	23	32	0
20	Maret 2019	33	27	0
21	April 2019	48	36	0
22	Mei 2019	10	10	0
23	Oktober 2019	2	0	0

**Sumber:** Data Peneliti (2020)

Pada tabel diatas merupakan data pembelian berdasarkan *quantity* barang besi, pvc dan ppr perbulannya, di *cleansing* sesuai kebutuhan yang diperlukan dalam algoritma.

### 3. *Transformation*

Setelah data-data dicelaning sesuai variabel yang dibutuhkan untuk tahapan *data mining*, data tersebut akan diinput ke dalam aplikasi untuk dilakukan tahapan lebih lanjut, yaitu *data mining*.

### 4. *Data Mining*

Setelah data diinput ke dalam aplikasi akan dilakukan proses *data mining* dengan aplikasi *RapidMiner* menggunakan metode *k-means*.

### 5. *Interpretation / Evaluation*

Tahapan akhir atau evaluasi ini, melalui proses *data mining* di aplikasi *RapidMiner* dengan menggunakan metode *k-means* menghasilkan pengelompokan data yang ditampilkan dalam tabel berikut:

#### a. *cluster 0*

Berdasarkan perhitungan manual iterasi dihentikan pada iterasi ke-2, maka hasil yang dikelompokkan ke dalam *cluster 0* yaitu:

**Tabel 3. 4** Hasil Pengelompokan *Cluster 1*

No	Bulan	Cluster	Besi	Pcv	Ppr
1	Juli 2017	<i>Cluster 0</i>	9	0	0
2	September 2017		24	11	28
3	Mei 2018		11	21	16
4	Juni 2018		4	14	8
5	Juli 2018		8	15	8
6	Agustus 2018		23	43	7

**Sumber:** Hasil Peneliti (2020)

**Tabel 3.4** Lanjutan

7	September 2018	<i>Cluster 1</i>	34	22	0
8	Oktober 2018		12	33	5
9	Nopember 2018		29	36	0
10	Desember 2018		1	1	3
11	Januari 2019		23	32	0
12	Maret 2019		33	27	0
13	Mei 2019		10	10	0
14	Oktober 2019		2	0	0

**Sumber:** Hasil Peneliti (2020)

b. *Cluster 1*

Berdasarkan perhitungan manual iterasi dihentikan pada iterasi ke-2, maka hasil yang dikelompokkan ke dalam *cluster 1* yaitu:

**Tabel 3.5** Hasil Pengelompokan *Cluster 1*

No	Bulan	Cluster	Besi	Pcv	Ppr
1	Agustus 2017	<i>Cluster 1</i>	49	44	42
2	Desember 2017		56	29	34
3	Januari 2018		33	13	32
4	Februari 2018		48	35	17
5	April 2018		40	36	11
6	April 2019		48	36	0

**Sumber:** Hasil Peneliti (2020)

b. *Cluster 2*

Berdasarkan perhitungan manual iterasi dihentikan pada iterasi ke-2, maka hasil yang dikelompokkan ke dalam *cluster 2* yaitu:

**Tabel 3.6** Hasil Pengelompokan *Cluster 2*

No	Bulan	Cluster	Besi	Pvc	Ppr
1	Oktober 2017	<i>Cluster 2</i>	102	77	16
2	Nopember 2017		91	54	1
3	Maret 2018		83	80	11

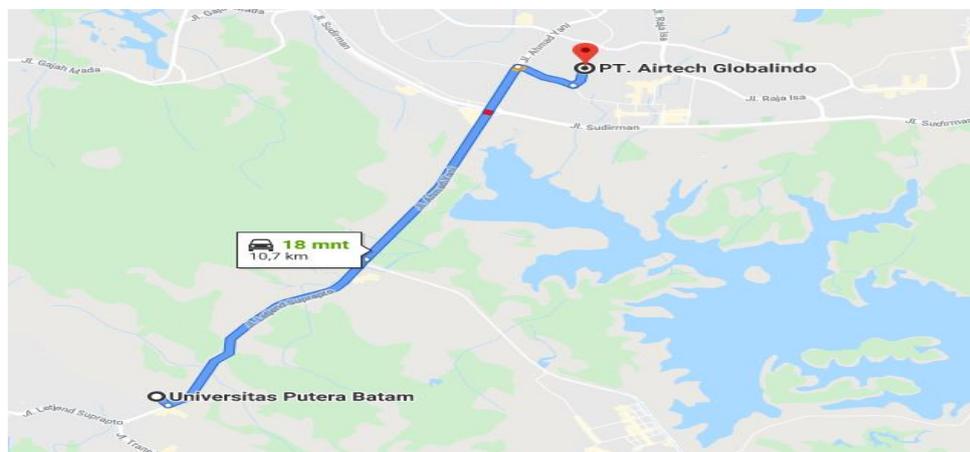
**Sumber:** Hasil Peneliti (2020)

Berdasarkan data pengelompokan yang diperoleh maka diketahui bahwa kelompok-kelompok jenis material tiap bulannya pada proyek *MV Doulos Phos Hotel* pada periode Juli 2017 – Oktober 2019.

### 3.5. Lokasi Dan Jadwal Penelitian

#### 3.5.1. Lokasi Penelitian

Penelitian ini dilaksanakan di PT. Starindo Acetech Perkasa, PT. Starindo Acetech Perkasa merupakan perusahaan yang bergerak dibidang *maintenance*, *general contractor* dan *supplier*, yang terletak di Komplek Mega Cipta Sejati Factory I, No. 01-05, Batam Center, Batam – Indonesia.



**Gambar 3. 2** Lokasi Penelitian

Pada gambar 3.2 PT. Starindo Acetech Perkasa berada di cakupan PT. Airtech Globalindo untuk itu jarak tempuh dari lokasi penelitian ke kampus Universitas Putera Batam (Tembesi) memerlukan waktu sekitar 18 menit.

### 3.5.2. Jadwal Penelitian

Pada penelitian ini terdapat waktu yang digunakan selama penelitian berlangsung sebagai berikut:

**Tabel 3. 7** Jadwal Penelitian

No	Jadwal Penelitian	Waktu Penelitian													
		Sep 2019		Okto 2019			Nop 2019		Des 2019		Jan 2020				Feb 2020
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Pengajuan Judul Penelitian														
2	Penyusunan BAB I														
3	Penyusunan BAB II														
4	Penyusunan BAB III														
5	Penyusunan BAB IV														
6	Penyusunan BAB V														
7	Pengumpulan Skripsi														

**Sumber:** Jadwal Peneliti (2020)

Berdasarkan tabel 3.2 merupakan kurun waktu yang digunakan dalam melakukan penelitian pengelompokan material proyek *MV Doulos Phos Hotel* menggunakan metode *k-means clustering*.