

**DATA *MINING* UNTUK PENEMPATAN BARANG
BERDASARKAN FREKUENSI PERMINTAAN
DI PT LAUTAN LESTARI SHIPYARD**

SKRIPSI



**Oleh:
Nober Six Salvanius Mendrofa
180210121**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2022**

**DATA *MINING* UNTUK PENEMPATAN BARANG
BERDASARKAN FREKUENSI PERMINTAAN
DI PT LAUTAN LESTARI SHIPYARD**

SKRIPSI

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



**Oleh
Nober Six Salvanius Mendrofa
180210121**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2022**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Nober Six Salvanius Mendrofa
NPM : 180210121
Fakultas : Teknik dan Komputer
Program Studi : Teknik Informatika

Menyatakan bahwa “Skripsi” yang saya buat dengan judul:

**DATA MINING UNTUK PENEMPATAN BARANG BERDASARKAN
FREKUENSI PERMINTAAN DI PT LAUTAN LESTARI SHIPYARD**

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, di dalam naskah Skripsi ini tidak terdapat karya 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 Skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah Skripsi ini digugurkan dan gelar akademik 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, 21 Januari 2022



Nober Six Salvanius Mendrofa

NPM. 180210121

**DATA *MINING* UNTUK PENEMPATAN BARANG
BERDASARKAN FREKUENSI PERMINTAAN
DI PT LAUTAN LESTARI SHIPYARD**

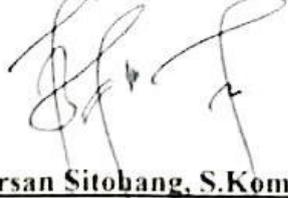
SKRIPSI

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

**Oleh:
Nober Six Salvanius Mendrofa
180210121**

**Telah disetujui oleh Pembimbing pada tanggal
seperti tertera dibawah ini**

Batam, 22 Januari 2022



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

ABSTRAK

Penataan barang di gudang yang baik merupakan salah satu hal penting dalam menunjang kegiatan suatu perusahaan. Pada PT Lautan Lestari Shipyard atau lebih dikenal dengan sebutan LLS, penempatan barang di gudangnya masih relatif tidak teratur dan terlihat berantakan sehingga menyebabkan kesulitan dalam pengambilan barang ketika diminta oleh pekerja di lapangan, hal ini dikarenakan penempatannya tidak berdasarkan frekuensi jumlah barang yang keluar. Untuk itu diperlukan suatu terobosan yang dapat mengatasi permasalahan tersebut baik menggunakan teknologi maupun kumpulan data-data. Penelitian ini bertujuan untuk menerapkan data mining dalam penyusunan *layout* barang dengan memanfaatkan arsip data pengeluaran barang yang tersedia di perusahaan yang telah disimpan. Arsip data yang telah terkumpul dapat memberikan pola pengambilan barang dan dari pola tersebut dapat diketahui barang dengan penggunaan yang banyak, sedang atau sedikit dengan menggunakan metode algoritma *k-means clustering*. Hasil perhitungan *Davies Bouldin Index* (DBI) menunjukkan jumlah *cluster* yang paling optimal adalah 5 *cluster* dengan nilai DBI 0,288, 5 *cluster* tersebut adalah *cluster* 0 sangat jarang digunakan, *cluster* 1 jarang digunakan, *cluster* 2 penggunaannya sedang, *cluster* 3 sering digunakan dan *cluster* 4 sangat sering digunakan. Barang yang masuk dalam kategori *cluster* 4 dan *cluster* 3 direkomendasikan untuk diletakkan di rak terdekat untuk memudahkan dalam pengambilan barang tersebut. Penelitian ini diharapkan dapat membantu mempermudah perusahaan dalam menata barang di gudang.

Kata Kunci: *Clustering, Data Mining, K-Means, Tata Letak*

ABSTRACT

Good arrangement of goods in the warehouse is one of the important things in supporting the activities of a company. At PT Lautan Lestari Shipyard or better known as LLS, the placement of goods in the warehouse is still relatively irregular and looks messy, causing difficulty in picking up goods when requested by workers in the field, this is because the placement is not based on the frequency of the number of goods that come out. For that we need a breakthrough that can overcome these problems using both technology and data collection. This study aims to apply data mining in the preparation of the layout of goods by utilizing the data archive of goods expenditures available in the company that has been stored. The data archives that have been collected can provide a pattern of retrieval of goods and from this pattern it can be seen that goods with a lot, moderate or little use are used by using the algorithm method k-means clustering. The results of the Davies Bouldin Index (DBI) calculation show that the most optimal number of clusters is 5 clusters with a DBI value of 0.288, these 5 clusters are cluster 0 which is very rarely used, cluster 1 is rarely used, cluster 2 is used moderately, cluster 3 is often used and cluster 4 is very rarely used. often used for. Items that fall into the cluster 4 and cluster 3 categories are recommended to be placed on the nearest shelf to make it easier to pick up these items. This research is expected to help make it easier for companies to organize goods in the warehouse.

Keywords: *Clustering; Data Mining; K-Means; Layout.*

KATA PENGANTAR

Segala Puji dan Syukur penulis panjatkan atas kehadiran Tuhan Yesus Kristus yang dengan senantiasa telah mencurahkan berkat dan karunia-Nya, sehingga penulis bisa menyusun dan menyelesaikan tugas akhir yang berjudul “Data Mining untuk Penempatan Barang berdasarkan Frekuensi Permintaan Barang dengan Metode *Clustering* di PT Lautan Lestari Shipyard”

Penulisan tugas akhir ini disusun guna memenuhi persyaratan dalam menyelesaikan program studi Strata Satu (S1) pada program Studi Teknik Informatika Universitas Putera Batam.

Penulis sadar bahwa tugas akhir ini tidak dapat terselesaikan dengan baik tanpa dukungan, bimbingan serta bantuan dari berbagai pihak. Untuk itu pada kesempatan ini dengan segala kerendahan hati, penulis mengucapkan terima kasih yang sebesar-besarnya kepada yang terhormat:

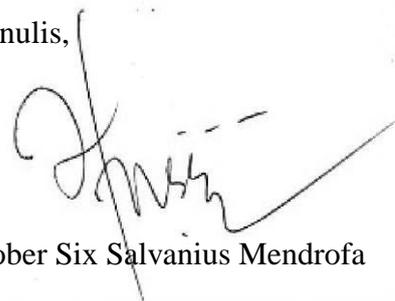
1. Ibu Dr. Nur Elfi Husda, S.Kom., M.SI. Selaku Rektor Universitas Putera Batam.
2. Dekan Fakultas Teknik dan Komputer, Bapak Welly Sugianto, S.T., M.M.
3. Ketua Program Studi Teknik Informatika, Bapak Andi Maslan, S.T., M.SI.
4. Kepada Bapak Sunarsan Sitohang, S.Kom., M.TI. Selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
5. Kepada Bapak Ellbert Hutabri, S.Kom., M.Kom. Selaku pembimbing Akademik pada Program Studi Teknik Informatika Universitas Putera Batam.

6. Para Dosen serta Staff di Universitas Putera Batam.
7. Kedua orang tua, abang dan adik yang selalu memberikan dukungan baik dari segi material maupun moril kepada penulis.
8. Seluruh keluarga besar penulis yang selalu memberi nasehat dan motivasi.
9. PT Lautan Lestari Shipyard yang telah memberikan izin kepada penulis untuk pengumpulan data penelitian.
10. Para teman-teman seperjuangan yang dengan setia mendampingi dan memberikan dukungan dan juga semangat.

Penulis juga menyadari keterbatasan pengalaman dan juga pengetahuan yang dimiliki penulis, sehingga dalam penulisan tugas akhir ini masih sangat jauh dari kata sempurna. Oleh sebab itu, penulis sangat mengharapkan saran dan kritik yang konstruktif dari berbagai pihak. Akhir kata semoga tugas akhir ini dapat diterima dan bermanfaat bagi para pembaca

Batam, 21 Januari 2022

Penulis,



Nober Six Salvanius Mendrofa

DAFTAR ISI

	Halaman
HALAMAN COVER	i
HALAMAN JUDUL	ii
SURAT PERNYATAAN ORISINALITAS	iv
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR RUMUS	xiii
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	4
1.4 Batasan Masalah	4
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	5
1.6.1 Manfaat Teoritis	5
1.6.2 Manfaat Praktis	6
BAB II TINJAUAN PUSTAKA	
2.1 <i>Knowledge Discovery in Database (KDD)</i>	7
2.2 <i>Data Mining</i>	9
2.3 <i>Metode Data Mining</i>	11
2.4 <i>Algoritma Data Mining</i>	16
2.4.1 <i>Algoritma Apriori</i>	16
2.4.2 <i>Algoritma Association Rule</i>	16
2.4.3 <i>Algoritma Rough Set</i>	17
2.4.4 <i>Algoritma C4.5</i>	17
2.4.5 <i>Algoritma Nearest Neighbor</i>	17
2.4.6 <i>Algoritma Naïve Bayes</i>	18
2.4.7 <i>Algoritma K-Means</i>	19
2.4.8 <i>Algoritma K-Nearest Neighbor (K-NN)</i>	20
2.4.9 <i>Algoritma K-Medoids</i>	20
2.5 <i>Davies Bouldin Index</i>	21
2.6 <i>Software Pendukung</i>	23
2.6.1 <i>Microsoft Excel 2019</i>	23
2.6.2 <i>Software RapidMiner</i>	24
2.7 <i>Penelitian Terdahulu</i>	25
2.8 <i>Kerangka Pemikiran</i>	30
BAB III METODOLOGI PENELITIAN	
3.1 <i>Desain Penelitian</i>	32

3.2	Teknik Pengumpulan Data	34
3.3	Operasional Variabel.....	35
3.4	Algoritma <i>K-Means Clustering</i>	35
3.5	Lokasi dan Jadwal Penelitian	37
3.5.1	Lokasi Penelitian	37
3.5.2	Jadwal Penelitian.....	38
BAB IV HASIL DAN PEMBAHASAN		
4.1	Analisis Data	39
4.1.1	Pengumpulan Data	39
4.1.2	Seleksi Data.....	40
4.1.3	Pembersihan Data.....	40
4.1.4	Transformasi Data	44
4.1.5	Implementasi	45
4.1.6	Perhitungan Algoritma <i>K-Means Clustering</i>	45
4.1.7	Perhitungan <i>Davies Bouldin Index (DBI)</i>	56
4.2	Hasil Pengujian	62
BAB V PENUTUP		
5.1	Kesimpulan.....	69
5.2	Saran.....	70
DAFTAR PUSTAKA		
LAMPIRAN		
Lampiran 1: <i>Davies Bouldin</i> untuk 2-10 cluster		
Lampiran 2: Foto Dokumentasi		
Lampiran 3: Form MWF dan Surat Jalan		
Lampiran 4: Daftar Riwayat Hidup		
Lampiran 5: Surat Permohonan Izin Penelitian		
Lampiran 6: Surat Balasan Izin Penelitian		

DAFTAR GAMBAR

Gambar 2.1. Tahapan Proses KDD	7
Gambar 2.2. <i>Flowchart</i> algoritma <i>K-Means Clustering</i>	12
Gambar 2.3. Cara Kerja Algoritma <i>K-Means</i>	19
Gambar 2.4. Logo <i>Microsoft Excel 2019</i>	23
Gambar 2.5. Logo <i>RapidMiner</i>	24
Gambar 2.6. Tahapan Kerangka Pemikiran	30
Gambar 4.1. Tampilan Data Mentah	39
Gambar 4.2. <i>Davies Bouldin 5 cluster</i>	62
Gambar 4.3. <i>Loading Screen software RapidMiner 9.10</i>	63
Gambar 4.4. Tampilan awal <i>software RapidMiner</i>	63
Gambar 4.5. <i>Import Data</i>	64
Gambar 4.6. <i>Select the cells to import</i>	64
Gambar 4.7. <i>Format your columns</i>	65
Gambar 4.8. <i>Where to store the data</i>	65
Gambar 4.9. Tampilan <i>Process</i>	66
Gambar 4.10 <i>Cluster dan jumlah item masing-masing cluster</i>	67
Gambar 4.11 <i>Grafik Tree Cluster Model</i>	68

DAFTAR TABEL

Tabel 3.1 Jadwal Penelitian	38
Tabel 4.1 Data setelah proses seleksi dan <i>cleaning</i>	41
Tabel 4.2 Lanjutan data setelah proses seleksi dan <i>cleaning</i>	42
Tabel 4.3 Lanjutan data setelah proses seleksi dan <i>cleaning</i>	43
Tabel 4.4 Lanjutan data setelah proses seleksi dan <i>cleaning</i>	44
Tabel 4.5 Titik Pusat <i>Cluster</i>	45
Tabel 4.6 Hasil Perhitungan Jarak <i>Euclidean</i> Iterasi 1	46
Tabel 4.7 Lanjutan Hasil Perhitungan Jarak <i>Euclidean</i> Iterasi 1	47
Tabel 4.8 Lanjutan Hasil Perhitungan Jarak <i>Euclidean</i> Iterasi 1	48
Tabel 4.9 Lanjutan Hasil Perhitungan Jarak <i>Euclidean</i> Iterasi 1	49
Tabel 4.10 Pusat <i>Cluster</i> Baru.....	51
Tabel 4.11 Hasil Akhir Perhitungan Jarak <i>Euclidean</i>	52
Tabel 4.12 Lanjutan Hasil Akhir Perhitungan Jarak <i>Euclidean</i>	53
Tabel 4.13 Lanjutan Hasil Akhir Perhitungan Jarak <i>Euclidean</i>	54
Tabel 4.14 Lanjutan Hasil Akhir Perhitungan Jarak <i>Euclidean</i>	55
Tabel 4.15 Pusat <i>centroid</i> iterasi ke-3	56
Tabel 4.16 Hasil Perhitungan SSW	58
Tabel 4.17 Hasil Perhitungan SSB	59
Tabel 4.18 Hasil Perhitungan <i>Ratio</i>	60
Tabel 4.19 Hasil Pengujian DBI.....	61
Tabel 4.20 Jumlah item per <i>cluster</i>	67

DAFTAR RUMUS

Rumus 2.1 Rumus <i>Euclidean</i>	20
Rumus 2.2 Rumus <i>Sum of Square Within-cluster</i>	22
Rumus 2.3 Rumus <i>Sum of Square Between-cluster</i>	22
Rumus 2.4 Rumus <i>Ratio</i>	23
Rumus 2.5 Rumus <i>Davies Bouldin Index</i>	23

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sejalan dengan perkembangan bisnis di masa modernitas serta kemajuannya dibidang teknologi data yang serba kilat ini membuat transformasi yang lumayan pesat, baik dalam bidang industri maupun jasa. Perihal tersebut pula mengakibatkan sesuatu pergantian besar pada tingkatan persaingan antara industri, sehingga para pelaku industri tersebut wajib senantiasa menghasilkan bermacam metode buat terus berkembang. Untuk mengimbangi persaingan khususnya di bidang bisnis dan meningkatkan pemasukan industri, para atasan perusahaan diharuskan untuk mampu mengambil tindakan yang akurat dalam menetapkan strategi yang akan dijalankan. Dalam hal tersebut perusahaan memerlukan data yang tepat dan perlu dianalisis lebih lanjut yang nantinya akan menjadi salah satu penentu strategi suatu perusahaan.

Ketersediaan data yang melimpah, kebutuhan pada informasi ataupun pengetahuan untuk dijadikan sebagai bahan pendukung penentuan keputusan agar membuat pemecahan permasalahan bisnis, serta sokongan infrastruktur di bidang teknologi data ialah awal mula dari munculnya data *mining* yang dirancangan sebagai pengganti pemecahan masalah yang nyata untuk para pimpinan dalam mengambil keputusan. Data *mining* juga adalah proses menambang pengetahuan dari kumpulan informasi dalam jumlah yang banyak. Data *mining* juga sering diistilahkan sebagai kegiatan mengekstraksi dan menganalisis informasi dengan jumlah besar yang bertujuan untuk menciptakan aturan dan pola yang bermakna,

selain itu data *mining* digunakan untuk mengekstrak informasi berharga dari *database* yang cukup besar, yang digunakan untuk memprediksi bentuk dan karakteristik bisnis dan membuat pola yang awalnya tidak dapat dikenali. Data *mining* bertujuan untuk menggali nilai tambah pada kumpulan informasi yang dapat berupa pengetahuan yang masih belum diketahui.

Menurut (Vulandari, 2017: 53) *clustering* merupakan salah satu metode data *mining* atau sering juga disebut dengan *segmentation*. Metode *clustering* digunakan untuk menentukan kelompok alami dari suatu kasus yang didasarkan pada sebuah kelompok atribut, kemudian mengelompokkan data yang memiliki persamaan atribut. *Clustering* juga merupakan metode yang *unsupervised*, karena seluruh atribut input diperlakukan sama.

Salah satu algoritma metode *clustering* yang masuk dalam kelompok *unsupervised learning* adalah algoritma *K-means*, algoritma ini digunakan untuk membagi data menjadi beberapa kelompok dengan sistem partisi. Algoritma ini menerima masukan berupa data tanpa label kelas. (Wanto et al., 2020: 3)

PT Lautan Lestari Shipyard merupakan perusahaan yang bergerak pada bidang perkapalan, membangun maupun memperbaiki kapal (*building and repair*). Melalui pengalaman selama ini, PT Lautan Lestari Shipyard selalu menjaga permintaan dari pelanggan mulai dari jadwal yang ditentukan untuk perbaikan kapal maupun kualitas dengan tujuan untuk memenuhi kepuasan pelanggan.

Dalam melakukan operasionalnya PT Lautan Lestari Shipyard membutuhkan peralatan maupun material penunjang dalam menggerakkan roda

perusahaan, baik yang dibeli secara lokal maupun yang diimport dari luar negeri. Barang-barang tersebut disimpan di 2 gudang yang saling berdekatan, penempatan barang yang baru masuk disusun di rak berdasarkan rak yang sudah kosong duluan dan barang-barang lama akan digeser atau ditumpuk disuatu tempat sehingga rak barang-barang tersebut nantinya akan diisi dengan barang baru. Penempatan yang seperti ini tentunya tidak berdasarkan dengan jumlah frekuensi pengambilan terbanyak dan terkesan tidak rapi, sehingga menyebabkan kesulitan dalam pengambilan barang ketika diminta oleh pekerja di lapangan.

Pengambilan alat dan bahan tersebut diatas akan dicatat dalam MWF (*Material Withdraw Form*) untuk pekerja internal perusahaan dan *form* Surat Jalan untuk pihak eksternal perusahaan, kemudian *form* tersebut direkap dengan menggunakan perangkat lunak *Microsoft Excel*. Jika data tersebut tidak dilakukan pengolahan yang tepat, data yang telah terkumpul akan sia-sia dan tidak bernilai guna. Arsip data yang telah terkumpul dapat memberikan pola pengambilan barang dan dari pola tersebut dapat diketahui barang dengan kegunaan yang banyak, sedang atau sedikit. Pola ini dapat berguna untuk mencari kaitan antara barang dan bisa digunakan untuk mengatur letak yang tepat dan penempatan barang yang harus berdekatan.

Berdasarkan uraian yang dijelaskan diatas, untuk itu penulis bermaksud mengangkat judul penelitian **“Data Mining untuk Penempatan Barang Berdasarkan Frekuensi Permintaan Barang di PT Lautan Lestari Shipyard”**.

1.2 Identifikasi Masalah

Berlandaskan pada latar belakang diatas, maka penulis menarik beberapa identifikasi masalah, yaitu:

1. Penempatan barang di rak masih belum punya konsep.
2. Kurangnya pengetahuan akan frekuensi barang dengan jumlah penggunaan yang banyak, sedang maupun sedikit.
3. Kurangnya pengetahuan dalam mengolah data yang ada.

1.3 Rumusan Masalah

Berlandaskan pada permasalahan yang telah diuraikan, penulis menarik rumusan permasalahan dalam penelitian ini:

1. Bagaimana analisis penempatan barang berdasarkan frekuensi permintaan dengan metode *clustering*?
2. Bagaimana penerapan metode *clustering* dalam penempatan barang berdasarkan frekuensi permintaan?

1.4 Batasan Masalah

Agar penelitian ini berfokus pada tujuan utama, maka penulis membuat beberapa batasan masalah yang akan diteliti, sebagai berikut:

1. Implementasi ini dilakukan berdasarkan pada data pengambilan barang di PT Lautan Lestari Shipyard periode Januari 2020 sampai dengan Desember 2020 dan Januari 2021 sampai dengan November 2021.

2. Data pengambilan barang yang di gunakan adalah data transaksi pengambilan barang untuk pekerjaan lapangan.
3. Pengujian data yang digunakan dalam penelitian ini adalah *software RapidMiner* dan menggunakan metode *clustering*.

1.5 Tujuan Penelitian

Berdasarkan gambaran rumusan masalah, maka tujuan dalam melakukan penelitian ini sebagai berikut:

1. Untuk menganalisis penempatan barang berdasarkan frekuensi permintaan dengan metode *clustering*.
2. Untuk menerapkan metode *clustering* dalam penempatan barang berdasarkan frekuensi permintaan.

1.6 Manfaat Penelitian

Manfaat Penelitian ini terbagi atas 2 (dua) manfaat antara lain manfaat teoritis dan manfaat praktis.

1.6.1 Manfaat Teoritis

1. Menerapkan ilmu yang telah didapatkan dalam perkuliahan di jurusan Teknik Informatika.
2. Salah satu bahan pertimbangan untuk mengetahui pemanfaatan data *mining* dalam pengambilan barang.

3. Tambahkan referensi bagi para pembaca yang akan melakukan penelitian serupa, terkhusus dibidang Teknik Informatika tentang data *mining* metode *clustering*.

1.6.2 Manfaat Praktis

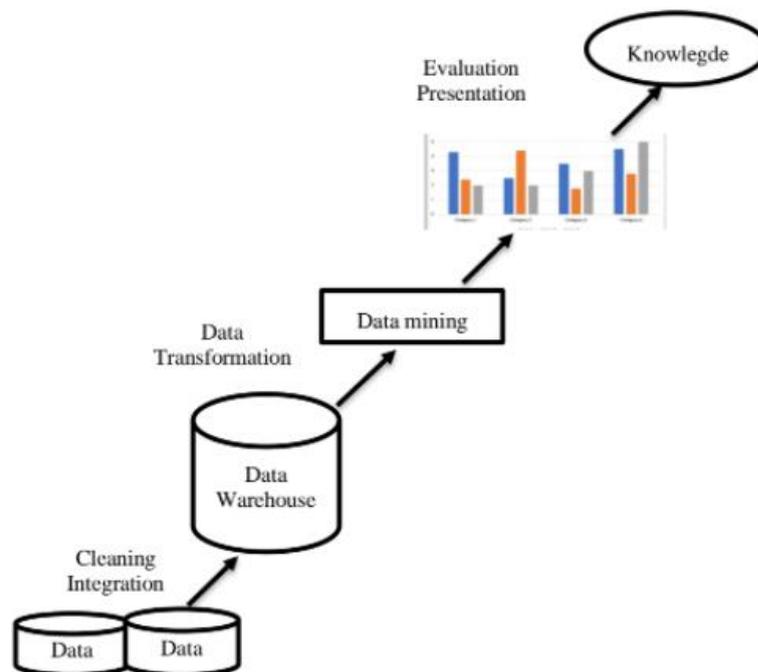
1. Memberikan informasi cara memanfaatkan arsip data pengeluaran barang, sehingga menjadi informasi yang bisa diolah menjadi data *mining* yang dapat menemukan solusi untuk penempatan barang.
2. Sebagai referensi untuk mendukung penentuan keputusan penempatan barang yang tepat.

BAB II

TINJAUAN PUSTAKA

2.1 Knowledge Discovery in Database (KDD)

Proses penemuan informasi dalam database dikenal sebagai *Knowledge Discovery in Database* (KDD). Seluruh proses pengorganisasian atau menemukan pola yang mendasari, pengetahuan, dan informasi dari pengumpulan data besar didefinisikan sebagai KDD. Metode KDD memberikan pengetahuan dan informasi praktis, baru, dapat dimengerti, dan penting (Buulolo, 2020: 2). Berikut ini adalah tahapan proses KDD.



Gambar 2.1 Tahapan Proses KDD

Sumber: (Peneliti, 2022)

1. *Data*

Sebelum memulai proses KDD data harus dipersiapkan terlebih dahulu. Data yang akan dipergunakan adalah data yang telah dipisahkan dengan data operasional.

2. *Selection*

Pemilihan data diperlukan untuk memilah data, karena data yang ada tidak semuanya dapat dipergunakan. Pemrosesan data ini, misalnya, menentukan kumpulan data target, menentukan variabel, memilih sampel data, dan menyimpan data dalam sebuah file.

3. *Pre-processing/cleaning*

Menghapus dan mengoreksi redundansi, inkonsistensi, dan kesalahan dari data yang dipilih dilakukan pada tahap ini. *Enrichment* adalah langkah dalam proses *Pre-processing/cleaning* yang melibatkan penambahan informasi tambahan yang relevan dan dapat diandalkan ke data yang telah dikumpulkan.

4. *Transformation*

Contoh dari tahap ini adalah transformasi dataset sehingga dapat digunakan dalam proses penambangan data. Ini adalah proses kreatif yang sangat bergantung pada pola pencarian database atau jenis informasi yang akan dicari.

5. *Data mining*

Tahapan ini merupakan tahapan utama dari proses KDD, dimana data yang terpilih akan dicari pola atau informasi yang menarik dengan

menggunakan metode dan teknik tertentu. Dalam data *mining*, ada berbagai pendekatan, strategi, dan algoritma untuk dipilih, tergantung pada tujuan atau informasi yang dicari.

6. *Interpretation/Evaluation*

Data *mining* menghasilkan pengetahuan dan informasi yang harus disajikan secara mudah dimengerti oleh mereka yang membutuhkannya. Pengetahuan dan informasi tersebut bisa dipresentasikan kedalam pohon keputusan, grafik atau dalam bentuk rule. Dalam proses KDD, tahap ini termasuk memverifikasi bahwa pola atau informasi yang didapatkan menurut fakta atau asumsi yang telah ditetapkan sebelumnya.

7. *Knowledge*

Tujuan akhir dari rangkaian proses KDD ialah Informasi berguna yang mampu dengan mudah dimengerti oleh beberapa pihak adalah tujuan. Pengetahuan atau informasi yang berguna perlu dipraktikkan sesuai dengan nilainya.

2.2 Data Mining

Data *mining* ialah suatu bidang ilmu yang dimanfaatkan dalam menangani permasalahan mendapatkan informasi dari database yang besar dengan menggabungkan teknik dari statistik, mesin pembelajaran, visualisasi data, pengenalan pola, dan juga database. Tujuan dari data *mining* adalah untuk mengekstrak informasi dengan teknik cerdas dari himpunan data yang selanjutnya

mengubah informasi menjadi terstruktur yang dapat dimengerti untuk penggunaan lebih lanjut (Indah Werdiningsih et al., 2020: 17).

Ada banyak cara untuk menambang data, tetapi salah satu yang paling umum adalah melalui penggunaan data *mining*. Data *mining* mencakup berbagai kegiatan, termasuk pengumpulan, ekstraksi, analisis, dan pelaporan data. *Knowledge discovery*, *knowledge extraction*, data / *pattern analysis*, *information harvesting*, dan istilah lainnya digunakan untuk menggambarkan penambangan data (Arhami & Nasir, 2020: 1).

Data *mining* juga menjadi proses logis untuk menemukan informasi yang berguna. Setelah ditemukan informasi dan pola dapat digunakan untuk alat pendukung dalam pengambilan keputusan dalam mengembangkan bisnis. Alat data *mining* dapat memberikan jawaban untuk berbagai pertanyaan yang terkait dengan bisnis dan terlalu sulit untuk diselesaikan. Data *mining* juga dapat digunakan untuk meramalkan tren masa depan, memungkinkan perusahaan untuk membuat pilihan yang lebih proaktif, efisien, dan dinamis.

Data yang telah dianalisis menggunakan metode data *mining* juga dapat mengungkapkan informasi yang sesuai dengan harapan. Misalnya pada bidang kesehatan, cukup banyak data yang dimiliki oleh rumah sakit, seperti data *medical record* dan *radiologi*, tetapi karena belum adanya standar koleksi data maka data-data tersebut sukar untuk diolah sehingga dengan kehadiran data *mining* maka diharapkan data-data yang dimiliki oleh pihak kesehatan dapat diolah sesuai dengan keperluan hingga menghasilkan informasi dan pengetahuan yang dapat dimanfaatkan oleh para pengambil kebijakan terutama pemerintah.

Menurut Siregar (2017: 3) Dengan "menggali" data, data *mining* mampu mengungkap informasi yang sebelumnya tidak diketahui dan memprediksi masa depan. Menggunakan pemodelan sebagai metode, tugas ini selesai. Sebuah "Jawaban" didefinisikan sebagai model yang dibangun dan kemudian digunakan dalam berbagai situasi di mana jawaban yang dicari.

Data *mining* biasanya dilakukan melalui tiga langkah (Arhami & Nasir, 2020: 4), yaitu:

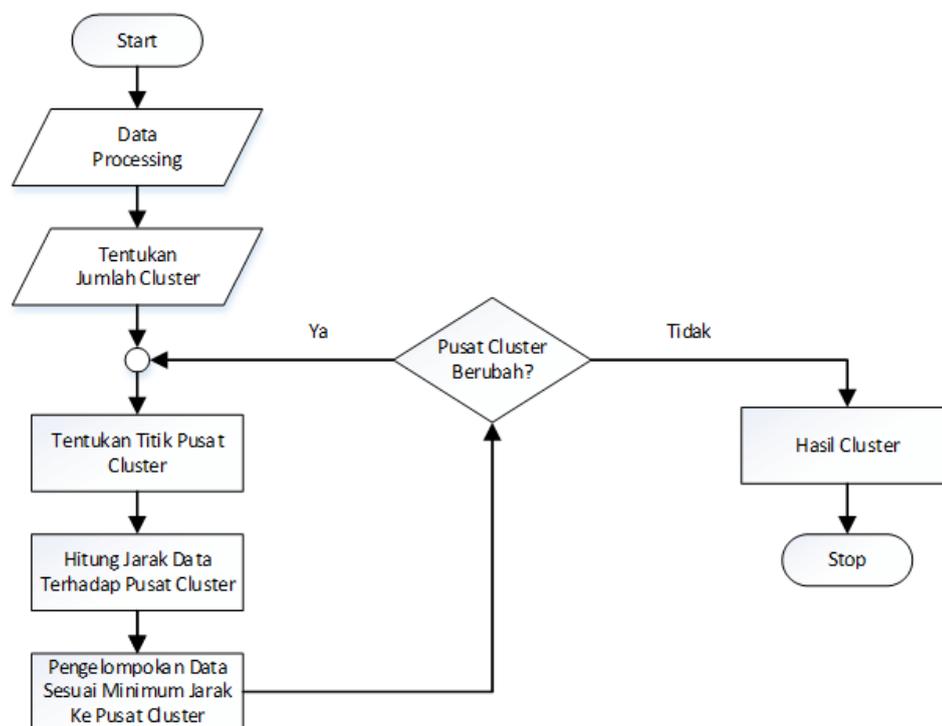
1. Eksplorasi: Hal yang harus dilakukan adalah mempersiapkan data dengan jumlah yang besar, selanjutnya data akan dibersihkan sesuai dengan yang diperlukan, serta menghilangkan data-data duplikat, alhasil data yang tertinggal hanya data yang benar-benar bisa digunakan.
2. Pemodelan atau identifikasi pola: Untuk membuat prediksi terbaik dan paling akurat, buatlah model statistik. Dikarenakan model yang berbeda diterapkan pada dataset yang serupa berulang kali serta hasilnya dibandingkan, prosedur ini bisa memakan waktu lama.
3. Penerapan: Model digunakan pada data pelatihan dan pengujian untuk memberikan perkiraan hasil yang diprediksi pada tahap akhir.

2.3 Metode Data *Mining*

Data *mining* dipisahkan menjadi beberapa kategori berdasarkan tugas-tugas yang dapat diselesaikan, seperti:

1. *Clustering* (Pengklasteran)

Pada tahun 1976, J.B. Mac Oueen adalah orang pertama yang memperkenalkan algoritma *K-Means*. Ketika mengelompokkan data berdasarkan karakteristik yang sama, algoritma ini ialah salah satu algoritma pengelompokan yang paling sering dipakai. Produk akhir yang dikenal sebagai *cluster* terbentuk sebagai hasil dari proses ini (Arhami & Nasir, 2020: 148). *Flowchart* algoritma *K-Means Clustering* dapat dilihat pada Gambar 2.2 berikut:



Gambar 2.2 *Flowchart* algoritma *K-Means Clustering*

Sumber: (Peneliti, 2022)

Clustering adalah pendekatan data *mining* yang berusaha untuk menemukan kelompok item dengan karakteristik serupa yang dapat diurutkan dari set lain dari objek, dengan objek dalam kelompok yang sama yang lebih homogen

daripada objek dalam kelompok yang terpisah (Jollyta et al., 2021: 1). Jumlah dan jenis data objek menentukan jumlah kelompok yang dapat diidentifikasi. Sekelompok data objek dibagi menjadi beberapa kelompok, kelompok-kelompok tersebut mempunyai ciri khas tertentu yang bisa dibedakan antara satu sama lain, tujuannya untuk melakukan interpretasi dan analisis selanjutnya sesuai dengan tujuan penelitian.

Dengan metode *clustering*, dapat kita identifikasikan area yang padat, menemukan hubungan yang menarik antar atribut data dan menemukan pola distribusi secara keseluruhan. Teknik *clustering* digunakan untuk mengelompokkan data dengan kriteria serupa ke wilayah atau grup yang sama, sambil memisahkan data dengan kriteria yang berbeda. Data akan dikelompokkan dalam satu kelompok atau cluster mempunyai tingkat kemiripan yang maksimum dan setiap kelompok mempunyai tingkat kemiripan yang minimum (Ginantra et al., 2021: 76).

Algoritma *Clustering* dibagi atas dua algoritma, yang pertama yaitu Hierarchical *Clustering*, yaitu suatu metode hirarkis untuk membuat kombinasi hirarkis dan diimplementasikan terhadap objek data, seperti *Hierarchical Agglomerative Clustering (HAC)*. *Non-hierarchical Clustering*, juga dikenal dengan algoritma pengelompokan partisi, memberikan jumlah n objek, dan k adalah banyaknya cluster yang dapat dibentuk. Algoritma partisi *clustering* memproses objek menjadi k grup sesuai dengan syarat optimasi tertentu, dimana setiap grup merupakan representasi dari sebuah cluster. Contoh algoritma partisi *clustering* termasuk *K-Means*.

Algoritma *K-means* akan dipakai guna pengelompokan dalam penelitian ini. Pengamatan dibagi menjadi kelompok k menggunakan metode *K-Means clustering* dalam statistik dan machine learning, dan setiap pengamatan ditempatkan dalam kelompok dengan rata-rata terdekat.

Pengelompokan data dengan nonhierarki (sekatan) memiliki beberapa metode salah satu metodenya adalah metode *K-Means* yang berupaya membagi data yang ada menjadi dua atau lebih kelompok. Metode ini memisahkan data ke dalam kategori, dengan setiap kategori termasuk data yang berbeda. Tujuan dari pengelompokan ini adalah untuk mengurangi jumlah spesies di setiap kategori sambil meningkatkan jumlah spesies di antara mereka.

2. *Classification* (Klasifikasi)

Klasifikasi adalah penggalan data yang menetapkan item dalam koleksi untuk menargetkan kelas tertentu. Klasifikasi dimulai dengan pengumpulan data dimana data tersebut terdapat kelas yang telah diketahui. Sebagai contoh, klasifikasi yang untuk mengidentifikasi peringkat kredit, berupa pemohon pinjaman sebagai risiko kredit rendah, sedang, atau tinggi. Data pemohon, seperti riwayat kredit pemohon, riwayat pekerjaan, kepemilikan atau sewa rumah pemohon, tahun tinggal pemohon, jumlah dan jenis investasi yang dilakukan oleh pemohon, dan sebagainya, dapat digunakan sebagai karakteristik. Target/kategori akan menjadi peringkat kredit, dan data untuk setiap aplikasi akan terjadi (Indah Werdiningsih et al., 2020: 19).

3. *Association* (Asosiasi)

Menurut Arhami & Nasir (2020: 160) analisis dengan kaidah asosiasi ini adalah untuk mendapatkan sekumpulan item atau itemset yang muncul secara bersamaan dalam transaksi-transaksi. Suatu itemset terdiri dari i buah item yang disebut dengan i -itemset. Persentase transaksi atau kombinasi yang terdiri dari satu item set disebut dengan *support itemset* dan *confidence* (nilai kepastian) ialah nisbah kekuatan hubungan diantara item dalam peraturan persatuan. Support dan confidence tersebut merupakan dua kriteria dasar dalam kaidah asosiasi.

4. Prediksi

Kegunaan dari metode prediktif adalah membangun model pengetahuan yang dapat digunakan untuk membentuk prediksi. Variabel dalam data dapat digunakan untuk menemukan pola yang berasal dari data. Pola ini dapat digunakan untuk meramalkan variabel yang tidak diketahui (Ginantra et al., 2021:79).

5. Estimasi

Estimasi sangat identik dengan metode klasifikasi; Perbedaannya adalah bahwa variabel target diasumsikan numerik daripada kategori.

6. Deskripsi

Deskripsi ialah teknik guna menggambarkan pola yang dimaksud. Penelitian analitis kadang-kadang bisa menjadi masalah mencoba memahami pola dan tren dalam data. Deskripsi pola dan tren kerap memberikan uraian yang mungkin pada pola atau tren (Tahyudin et al., 2021: 5).

2.4 Algoritma Data Mining

Berikut ini adalah beberapa algoritma yang biasa dikenakan dalam data *mining*, diantaranya:

2.4.1 Algoritma Apriori

Algoritma apriori, menurut Buulolo (2020: 11), telah banyak digunakan dalam data *mining* untuk mengidentifikasi pola data atau pola kejadian/frekuensi. Di pasar kecil, pola pembelian pelanggan biasanya ditemukan melalui penggunaan algoritma apriori berdasarkan data transaksi pembelian. Dalam algoritma apriori, istilah "nilai dukungan" mengacu pada persentase dari total data yang muncul di bagian data tertentu.

2.4.2 Algoritma Association Rule

Dalam data *mining*, Association Rule adalah metode untuk membentuk aturan asosiatif antara item dalam item kombinasi. Aturan hubungan ditulis dengan gaya "if then.... Atau jika... kemudian..." Proses Association Rule adalah kelanjutan dari metode sebelumnya, di mana aturan dibentuk dengan menggabungkan 2 item yang memenuhi persyaratan dukungan minimal, artinya yang akan dibuat *rule*-nya hanya kombinasi yang memenuhi *support* minimum.

Tujuan *association rule* adalah untuk mendapatkan hubungan/kekuatan hubungan antar item dalam sebuah kombinasi *itemsets*, contohnya adalah seberapa besar kemungkinan seseorang pembeli teh, juga membeli gula secara bersamaan dan sebaliknya. Kekuatan hubungan/peluang pada *association rule* antara minimal 2 (dua) *itemsets*, nilainya diukur dengan nilai *confidence* (Buulolo, 2020: 19).

2.4.3 Algoritma *Rough Set*

Metode *Rough Set* adalah pendekatan data *mining* yang menggunakan *Decision Systems* (tabel yang dilengkapi dengan *attribute* kriteria dan keputusan) untuk mengekstrak informasi/ pengetahuan dari kelompok data atau tabel. Output dari algoritma *rough set* dalam bentuk *rule* dan dapat digunakan untuk memperkirakan sesuatu yang akan terjadi berdasarkan data-data sebelumnya seperti memprediksi kelulusan mahasiswa berdasarkan nilai yang diperoleh per semester, memprediksi kenaikan harga emas berdasarkan harga emas sebelumnya dan lain sebagainya.

2.4.4 Algoritma C4.5

Algoritma C4.5 adalah alat kategorisasi data yang banyak digunakan dalam data *mining*. Algoritma C4.5 adalah algoritma untuk mengeksplorasi data yaitu untuk mencari hubungan variabel input atau *attribute* kriteria dengan variabel target atau *decision attribute* (atribut keputusan). Hasil dari eksplorasi data diubah dalam bentuk pohon keputusan (*decision tree*) sehingga mudah dipahami dan dimengerti oleh pengguna. Algoritma C4.5 mengubah sejumlah besar data menjadi pohon keputusan, yang kemudian ditafsirkan sebagai rule. Dalam data *mining* banyak algoritma pembentukan pohon keputusan seperti Algoritma *Cart*, Algoritma ID3, Algoritma C4.5 dan Algoritma 05.0.

2.4.5 Algoritma Nearest Neighbor

Sesuai dengan terjemahan ke bahasa Indonesia *nearest* = terdekat, dan *neighbor* = tetangga, berarti tetangga terdekat. Biasanya yang bertetangga selalu ada kemiripan antara rumah tangga yang satu dengan yang lainnya misalnya

model pagar rumah, bentuk rumah, merek TV, bentuk dapur rumah dan lain-lain. Oleh karena itu, algoritma *nearest neighbor* adalah penyelesaian kasus dengan menggunakan pendekatan atau perbandingan kasus lama dengan kasus baru dengan pendekatan atau perbandingan dengan menggunakan bobot nilai setiap kriteria kasus.

Contoh seorang hakim akan menjatuhkan hukuman kepada seorang terdakwa pencuri, selain berdasarkan undang-undang yang berlaku tetapi juga berdasarkan hukuman kepada terdakwa pencuri yang telah dijatuhkan hukuman sebelum-sebelumnya. Bank jika memberikan pinjaman kepada seseorang, akan menggunakan berbagai pertimbangan atau kriteria dan membandingkan dengan data-data nasabah sebelumnya.

2.4.6 Algoritma Naïve Bayes

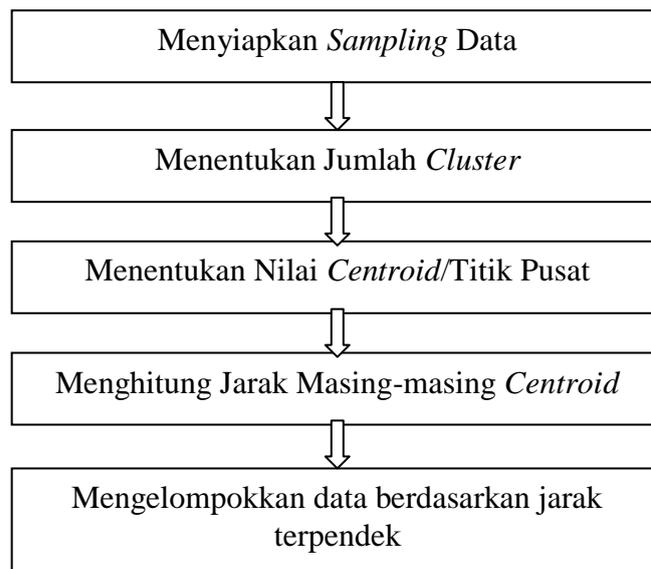
Algoritma Naive Bayes dikembangkan oleh ilmuwan Inggris Thomas Bayes dan memanfaatkan metode probabilitas dan statistik guna mengklasifikasikan data. Nama algoritma *Naive Bayes* mengacu pada *Teorema Bayes*, yang melakukan prediksi peluang masa mendatang bersumber dari pengalaman masa lalu.

Naïve Bayes Classifier ini memiliki asumsi yang sangat kuat bahwasanya setiap kondisi/peristiwa independen. Pengklasifikasi ini berkinerja mengagumkan dibandingkan dengan orang lain dari jenisnya. Untuk menentukan perkiraan parameter yang diperlukan untuk klasifikasi, metode ini menggunakan sejumlah kecil data pelatihan. Varians variabel dapat digunakan untuk menentukan

klasifikasinya, bukan seluruh matriks kovarians, karena diasumsikan bebas (Al-Khowarizmi, 2021: 219).

2.4.7 Algoritma *K-Means*

Algoritma *K-Means* menjadi salah satu teknik dalam data *mining* guna mengklusterkan (*Clustering*) data kedalam beberapa kelompok berdasarkan jarak, kriteria, kondisi atau karakteristik. Data dalam satu kelompok harus memiliki jarak terpendek, kriteria, kondisi atau karakteristik yang serupa atau mendekati sama antara satu dengan yang lain. Algoritma *K-Means* dapat mengelompokkan objek yang memiliki kemiripan. Berikut langkah-langkah dalam pengelompokan data dengan Algoritma *K-Means*:



Gambar 2.3 Cara Kerja Algoritma *K-Means*

Sumber: (Peneliti, 2022)

Untuk menetapkan jarak tiap-tiap *centroid* menurut jarak paling pendek, kriteria, kondisi atau karakteristik yang sama atau hampir sama, menggunakan model *Euclidean* yang dirumuskan berikut:

Rumus 2.1 Rumus *Euclidean*

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2}$$

Keterangan:

d_{ij} = Jarak dari data ke i ke pusat cluster j

x_{ki} = Data dari ke- i pada *attribute* data ke- k

x_{kj} = Data dari ke- j pada *attribute* data ke- k

2.4.8 Algoritma *K-Nearest Neighbor (K-NN)*

K-NN menjadi salah satu algoritma atau metode untuk klasifikasi yang biasa dikenakan dalam data *mining*. K-NN juga masuk dalam kategori regresi yang juga dapat digunakan untuk memprediksi seperti halnya regresi. K-NN dalam kelompok pembelajaran digolongkan ke dalam *supervised learning* yang berbasis pada jarak. K-NN juga dapat disebut sebagai *Casee Based Reasoning* yang merupakan suatu metodologi berbasis penalaran kasus-kasus berbasis pada data latih dengan suatu kasus disimpan, dilatih dan diakses untuk menyelesaikan permasalahan baru. Untuk mendapatkan nilai prediksi dari data baru maka dicari data-data yang sama atau berdekatan (Arhami & Nasir, 2020: 96).

2.4.9 Algoritma *K-Medoids*

Menurut Buulolo (2020: 72) algoritma *K-Medoids*, juga dikenal sebagai algoritma Partitioning Around Medoids (PAM), adalah pendekatan penambangan data untuk mengelompokkan (*clustering*) data ke dalam kelompok berdasarkan jarak, kriteria, kondisi, atau atribut. Algoritma *K-Medoids* dapat mengelompokkan objek yang mempunyai kemiripan. Perbedaannya yaitu algoritma *K-Medoids* ada perwakilan (*medoid*) untuk setiap cluster yang terbentuk, sedangkan algoritma *K-*

Means tidak mempunyai perwakilan karena dalam pembentukan pusat cluster (*centroid*) terbaru menggunakan nilai rata-rata (*mean*) dari cluster data yang terbentuk. Salah satu kelebihan algoritma *K-Medoids* adalah tidak sensitif terhadap *outlier* dan *noise* atau tidak mudah terjadi perubahan *cluster* terhadap perubahan nilai pusat *cluster* (*centroid*), dan sangat berbeda jauh dengan algoritma *K-Means* yang rentang terhadap perubahan *cluster* jika terjadi perubahan nilai pusat *cluster* (*centroid*).

2.5 Davies Bouldin Index

Pada tahun 1979 David L. Davies dan Donald W. Bouldin mengenalkan *Davies Bouldin Index* (DBI), sebuah metode untuk mengevaluasi efisiensi pengelompokan dan mendefinisikan kohesi sebagai ukuran kedekatan data dengan titik pusat cluster. Ada skema evaluasi berdasarkan jumlah dan kedekatan data hasil *clustering* saat menggunakan *Davies Bouldin Index* untuk *clustering* internal (Muningsih et al., 2021: 97)

Davies Bouldin Index menghitung jarak (*dissimilarity*) antar kluster. Algoritma pengelompokan dengan DBI rendah menghasilkan cluster dengan jarak intra-kluster rendah (kesamaan intra-kluster tinggi) dan jarak antar-kluster yang tinggi (kesamaan antar-kluster rendah). Semakin rendah nilai *Davies-Bouldin Index* suatu algoritma *machine learning* yang digunakan untuk klasterisasi, maka semakin tinggi kinerjanya. (Santoso & Azis, 2020: 53). Ini adalah tujuan evaluasi *cluster* untuk menentukan seberapa baik hasil pengelompokan kinerja. DBI adalah

metode yang digunakan untuk mengevaluasi hasil pengelompokan dan menentukan jumlah cluster yang optimal untuk digunakan.

Indeks Davies Bouldin dihitung dengan cara berikut (Manullang, et al. 2020: 1433):

1. *Sum of Square Within-cluster (SSW)*

Sum of Square Within-cluster (SSW) dapat digunakan untuk memperkirakan tingkat kohesi dalam cluster ke- i . Untuk mengukur kohesi, kita melihat berapa banyak titik data yang dekat dengan pusat *cluster* dari *dataset* yang dilacak. Persamaan yang digunakan untuk mendapatkan *Sum of Square Within-cluster* adalah sebagai berikut.

Rumus 2.2 Rumus *Sum of Square Within-cluster*

$$SSW = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$$

2. *Sum of Square Between-cluster (SSB)*

Hitung jarak antara cluster menggunakan perhitungan *Sum of Square Between-cluster (SSB)*. *Sum of Square Between-cluster* dapat dihitung menggunakan persamaan berikut.

Rumus 2.3 Rumus *Sum of Square Between-cluster*

$$SSB_{i,j} = d(c_i, c_j)$$

3. *Ratio (Rasio)*

Tujuannya adalah untuk menentukan nilai perbandingan antara cluster ke- i dan ke- j . Persamaan berikut digunakan untuk menghitung nilai rasio yang dimiliki oleh setiap *cluster*.

Rumus 2.4 Rumus *Ratio*

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

4. *Davies Bouldin Index*

Persamaan berikut digunakan untuk menemukan nilai *Davies-Bouldin Index* (DBI) menggunakan nilai yang diperoleh dari perhitungan rasio.

Rumus 2.5 Rumus *Davies Bouldin Index*

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$$

Banyaknya cluster yang disimpulkan dari persamaan, adalah k. Semakin kecil nilai DBI (non-negatif ≥ 0), semakin baik cluster yang dibentuk oleh pengelompokan memakai algoritma *clustering* (Butsianto & Saepudin, 2016 : 116).

2.6 Software Pendukung

Penelitian ini dilakukan dengan memanfaatkan software *Microsoft Excel* dan juga *Software RapidMiner* untuk melakukan pengujian data.

2.6.1 Microsoft Excel 2019

Gambar 2.4 Logo *Microsoft Excel 2019*

Sumber: (Peneliti, 2022)

Microsoft Excel adalah program *spreadsheet* (pengolah data) yang digunakan untuk mengotomatisasi manipulasi data menggunakan alat yang ada untuk memudahkan dalam melakukan perhitungan dasar, manipulasi data, pembuatan tabel, manajemen data grafis, dan lainnya. *Microsoft Excel* unggul dalam mengolah berbagai macam data, baik di bidang akuntansi, teknik, statistik, dan bidang lainnya yang membutuhkan perhitungan yang cepat dan akurat.

Pada lokasi penelitian ini pengolahan data menggunakan *Microsoft Excel*, data pengeluaran barang diketik secara manual dan direkap dalam satu *file*. Setelah itu akan disusun kedalam *stock card* berdasarkan item barang dengan menginput tanggal, jumlah item, *project*, dan nomor transaksi.

Software Microsoft Excel dalam penelitian ini akan digunakan sebagai pengolah data awal dengan cara melakukan pembersihan data seperti menghapus data duplikat, mengecek data yang kurang tepat dan lain sebagainya sehingga data siap diolah dengan menggunakan *software RapidMiner*.

2.6.2 *Software RapidMiner*



Gambar 2.5 Logo *RapidMiner*

Sumber: (Peneliti, 2022)

RapidMiner adalah program gratis dan *open-source*. Setiap kali berhubungan dengan data *mining*, *RapidMiner* adalah pilihan yang sangat baik. *RapidMiner* memanfaatkan bermacam-macam teknik deskriptif dan prediksi untuk membantu pengguna membuat keputusan terbaik.

Sejumlah fitur *RapidMiner*, meliputi:

1. Jumlah algoritma data *mining*, misalnya *decision tree* dan *self-organization map*.
2. Bentuk grafis yang canggih, misalnya tumpang tindih diagram histogram, *tree chart* dan *3D Scatter plots*.
3. Jumlah variasi *plugin*, misalnya *text plugin* untuk melakukan analisis teks.
4. Mencadangkan prosedur data *mining* dan machine learning meliputi: ETL (*extraction, transformation, loading*), data *preprocessing*, visualisasi, *modelling* dan evaluasi
5. Proses data *mining* terdiri dari operator-operator yang *nestable*, dideskripsikan memakai XML, dan diciptakan memakai GUI
6. Menggabungkan proyek data *mining Weka* dan statistika R.

2.7 Penelitian Terdahulu

Peneliti yang menggunakan data *mining* untuk menentukan tata letak menggunakan penelitian sebelumnya sebagai panduan. Jurnal berikut dikonsultasikan oleh para peneliti:

1. Penelitian oleh (Lestari, 2019) dari STMIK Budi Darma, Jurnal Riset Komputer (JURIKOM), ISSN: 2407-389X, Vol. 6 No. 2, dengan judul

penelitian “**Implementasi Data Mining Untuk Pengaturan Layout Swalayan Delimas Lestari Kencana dengan Menggunakan Algoritma K-Means Clustering**”. Tujuan penelitian ini adalah membuat model tata letak merek produk berdasarkan jenis produk menggunakan data dari tanda terima belanja. Data dibagi menjadi beberapa kelompok menggunakan algoritma pengelompokan *K-means*, yang membagi ke dalam kelompok data dengan ciri khas yang serupa secara bersamaan, dan data dengan karakteristik yang berbeda ke dalam kelompok lain. Sehingga dapat ditentukan berapa penempatan merek produk berdasarkan jenis produk di Supermarket Delima Lestari Kencana, tiga cluster dibuat menggunakan algoritma *K-Means: Cluster C1, Cluster C2, dan Cluster C3*.

2. Penelitian oleh (Syahra et al., 2020) dari STMIK Triguna Dharma, Journal of Technopreneurship and Information System (JTIS), E-ISSN: 2614-3089, Vol. 3 No.1, dengan judul “**Implementasi Data Mining Untuk Penyusunan Tata Letak Data Obat-Obatan Dengan Menggunakan Algoritma K Harmonic Means Pada Apotek Inti Fada Sidamanik**”. Sistem informasi manajemen data obat akan dibangun di Apotek Inti Fada Sidamanik untuk merampingkan serta menyederhanakan sistem informasi yang tidak terstruktur dengan baik. Hasil tes menunjukkan bahwa keanggotaan metode *clustering* metode KHM memiliki tingkat kesepakatan yang sangat bagus, dengan tingkat kesepakatan 92%.
3. Penelitian oleh (Fithri & Wardhana, 2021) dari Universitas Mercu Buana, Jurnal PILAR Nusa Mandiri, E-ISSN: 2527-6514, Vol. 17 No. 2, dengan

judul “*Cluster Analysis of Sales Transaction Data using K-Means Clustering at Toko Usaha Mandiri*”. Studi ini menggunakan algoritma pengelompokan *K-Means* untuk menganalisis data penjualan dan membuat kluster penjualan. Berupa klaster1, klaster2 dan klaster3 dengan persentase 62% (11 data), 8% (56 data) dan 30% (11 data), hasil pengelompokan data (25 data). Nilai 0,2 diperoleh dengan memvalidasi algoritma *K-Means Clustering* dengan *Davies Bouldin Index*. Manajemen inventaris dan strategi pemasaran tidak harus dibatasi oleh informasi pengelompokan penjualan.

4. Penelitian oleh (Lusiani & Liperda, 2020) dari Universitas Pertamina, Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi), E-ISSN: 2527-5941, dengan judul “**Perancangan Ulang Tata Letak Produk pada Gudang untuk Mengurangi Biaya Penanganan Material**”. Penelitian ini bertujuan untuk mengurangi biaya penanganan material dengan mendesain ulang tata letak produk gudang. Untuk penelitian ini, empat metode digunakan: *Class Based Storage*, *Dedicated Storage*, *Cube per Order Index*, dan Simulasi. Ada 27 skenario yang mungkin, masing-masing dengan tata letak yang berbeda, arrival, dan jumlah forklift. Tata letak dengan kebijakan Penyimpanan Khusus dan jumlah forklift sama dengan satu *forklift storage* dan satu *forklift delivery* dipilih berdasarkan perhitungan Manajemen Kinerja.
5. Penelitian oleh (Harpendi Bara et al., 2020) dari Institut Teknologi Nasional Malang, JATI (Jurnal Mahasiswa Teknik Informatika), E-ISSN:

2598-828X, Vol. 4 No. 2, dengan judul “**Pengelompokan Data Obat menggunakan Metode *K-Means Clustering* Pada UPT Puskesmas Kondoran Kec. Sangalla**”. Tujuan Penelitian: Mengurangi jumlah waktu yang diperlukan untuk memproses data untuk rencana pembelian obat di masa depan dengan mengembangkan sistem penambahan data yang dapat membantu mengelompokkan data obat ke dalam beberapa kelompok, dan dengan demikian mengurangi jumlah obat yang tidak dibutuhkan yang terakumulasi di gudang obat pusat kesehatan, di mana ia rentan terhadap kedaluwarsa. Sebanyak 183 data masuk ke dalam klaster 1 obat dengan penggunaan lambat dan 21 data dimasukkan ke dalam klaster 2 obat dengan penggunaan cepat, berdasarkan hasil metode pengujian pada 204 data.

6. Penelitian oleh (Hutapea & Fauzi, 2021) dari Universitas Putera Batam, Computer and Science Industrial Engineering (Comasie Journal), E-ISSN: 2715-6265, Vol. 04, No.06, dengan judul “**Data Mining Sistem Tata Letak Material di PT Batam Cyclelect**”. Ini adalah tujuan dari penelitian ini untuk memudahkan manajer toko sementara juga mengurangi jumlah waktu yang dibutuhkan untuk menemukan barang. Algoritma apriori memanfaatkan proses ini untuk mengurangi atau mempersempit ruang pencarian untuk frekuensi kandidat item yang ditetapkan. Perangkat lunak RapidMiner digunakan untuk memverifikasi keakuratan algoritma apriori. Setelah penelitian selesai, ditemukan bahwa memesan Marking Tape dan penanda menghasilkan nilai support 5% dan confidence 50%, sementara

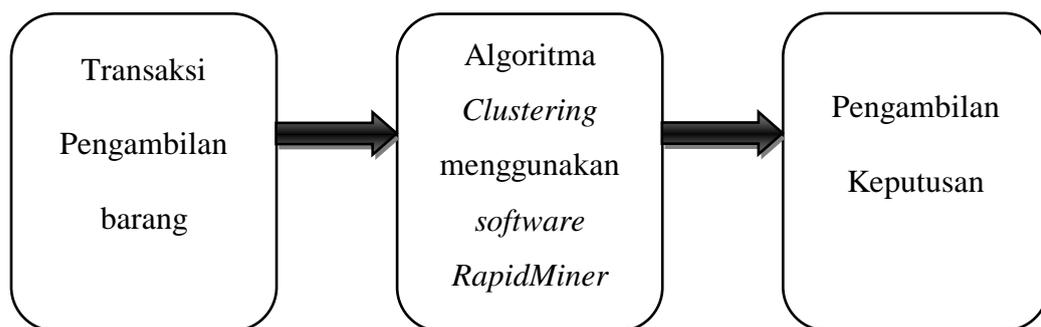
melakukan pemesanan PVC Insulation Tape menghasilkan nilai support 15% dan confidence 89%.

7. Penelitian oleh (Rajagukguk & Sitohang 2021) dari Universitas Putera Batam, Computer and Science Industrial Engineering (Comasie Journal), E-ISSN: 2715-6265, Vol. 5 No. 5, dengan judul **“Penerapan Data Mining dalam Pemilihan Laptop Menggunakan Metode Algoritma C4.5”**. Tujuan dari penelitian ini adalah untuk memanfaatkan perangkat lunak WEKA untuk mengkategorikan laptop berdasarkan harga, kapasitas RAM, kecepatan prosesor, ukuran, dan masa pakai baterai, yang semuanya dianalisis menggunakan algoritma C4.5. Dengan total 80 titik data, ditemukan tiga kriteria, yaitu harga, prosesor, dan daya tahan baterai, memiliki dampak signifikan pada pemilihan laptop. Dari 80 keping data, 53 memilih laptop yang tepat dan 27 memilih laptop yang salah.
8. Penelitian oleh (Fajrin & Handoko, 2018) dari Universitas Putera Batam, Jurnal Ilmiah Informatika (JIF), E-ISSN: 2615-1049, Vol. 06, No. 02, dengan judul **“Penerapan Data Mining Untuk Mengolah Tata Letak Buku Dengan Metode Association Rule”**. Tujuan dari penelitian ini adalah untuk menggunakan data transaksi pinjaman buku perpustakaan ABC untuk memecahkan tantangan pemrosesan data saat ini, mencari tahu buku mana yang sering dipinjam, dan membantu penempatan buku tergantung pada jenisnya. Pakai. Untuk membantu perpustakaan dalam meningkatkan kualitas layanan, tingkat pinjaman telah meningkat. Support minimum = 0,032, confidence minimum = 0,50, berdasarkan tes yang

dilakukan pada data dari 659 transaksi pinjaman buku dengan total 182 atribut. RapidMiner digunakan dalam uji ini.

2.8 Kerangka Pemikiran

Kerangka pemikiran ialah garis pemikiran yang masuk akal apabila membuat kesimpulan. Melalui pemanfaatan data *mining* algoritma *clustering* susunan letak barang di PT Lautan Lestari Shipyard dapat ditentukan susunannya, tahapan yang dibutuhkan untuk mendapatkan hasil rancangan dengan fungsi yang baik sebagai berikut:



Gambar 2.6 Tahapan Kerangka Pemikiran

Sumber: (Peneliti, 2022)

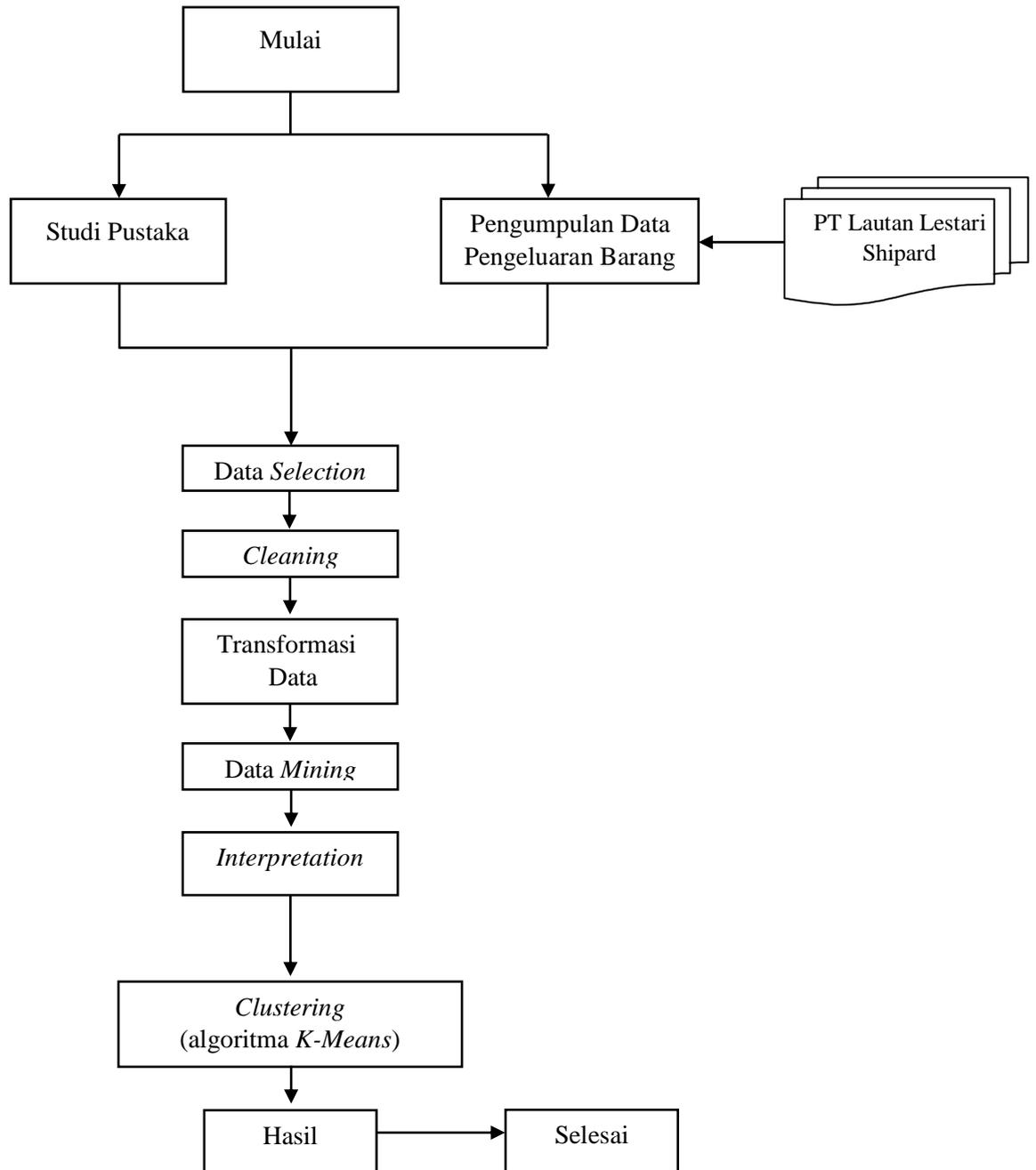
Seperti yang terlihat dari *flowchart*, langkah awal dari penelitian ini adalah:

1. Dihimpun lalu dilaksanakan *cleaning* data, *transformation* guna mengetahui pola pengambilan barang. Tahapan tersebut menghasilkan variabel masukan berwujud sejumlah pengambilan barang dan tipe item.

2. Tahap ini ialah proses pengkategorisasi data dengan memanfaatkan metode aturan *clustering*. Selanjutnya dilakukan pengujian dengan *software RapidMiner* berdasarkan data-data yang diperoleh.
3. Temuan penelitian ini adalah aturan untuk membuat keputusan dimana aturan yang digunakan sebagai panduan ketika memutuskan tata letak barang.

BAB III
METODOLOGI PENELITIAN

3.1 Desain Penelitian



Gambar 3.1 Tahapan Penelitian

Sumber: (Peneliti, 2022)

1. Studi Pustaka

Tinjau literatur terkait konsep data *mining* dengan menggunakan metode *clustering* dan algoritma *K-Means*. Literatur berasal dari data pengeluaran komoditas selama periode penelitian.

2. Pengumpulan Data

Untuk memahami data yang diperlukan, peneliti melakukan pengumpulan salinan data arsip pengeluaran barang dengan periode Januari-Desember 2020 dan Januari-November 2021.

3. Praproses data meliputi:

a. Data Selection

Untuk memilih kumpulan data (*dataset*) yang akan digunakan dalam penelitian ini dalam bentuk data pengeluaran barang, yang mencakup informasi tentang jenis barang dan nilainya.

b. *Cleaning*

Untuk merapikan data dengan cara yaitu menyempurnakan data, membuang data yang sama (duplikat).

c. Transformasi data

Guna membangun data supaya bisa di *cluster* atau dapat juga dengan cara menambah atribut baru.

d. Data *mining*

Menerapkan metode pengelompokan dan algoritma *K-Means* untuk memperoleh data pengeluaran produk, dan proses penambangan data akan dijabarkan menggunakan *software RapidMiner*.

3.2 Teknik Pengumpulan Data

Guna memperoleh data dari masalah penelitian perlu dilakukan pengumpulan data. Data yang telah dikumpulkan akan dilakukan pengolahan dengan data *mining* dan akan menghasilkan informasi yang bernilai guna. Ada beberapa metode pengumpulan data yang akan dikenakan, meliputi:

1. Observasi, Kegiatan yang dilakukan saat memasuki lokasi penelitian dengan mengamati langsung di lokasi penelitian. Peneliti dapat mengumpulkan data pengeluaran barang dan mengamati bagaimana barang ditempatkan di gudang PT Lautan Lestari Shipyard. Dalam penelitian ini yang akan diamati adalah penempatan barang di gudang, proses pengambilan barang hingga proses rekapan data pengambilan barang.
2. Dokumentasi, diperlukan sebagai alat bukti bahwa peneliti mendapatkan data langsung dari lokasi penelitian di PT Lautan Lestari Shipyard. Dokumentasi tersebut dapat dalam bentuk foto kegiatan penelitian dan juga foto rak-rak barang.

3. Studi Pustaka

Dalam studi ini, peneliti mencari informasi pengembangan teoritis yang relevan dengan subjek dengan melakukan tinjauan literatur yang terkait dengan topik penelitian dan prosedur penelitian, terutama jurnal dan buku yang relevan.

3.3 Operasional Variabel

Dalam sebuah penelitian, variabel operasional digunakan untuk memastikan atribut dan nilai yang terkait. Pembersihan dan penggalian variabel dari data yang dikumpulkan selama proses penelitian. Karakteristik berikut dimasukkan dalam analisis sebagai variabel independen:

1. Jenis item atau kategori digunakan untuk menentukan hal-hal yang memiliki pola frekuensi tinggi dan kombinasi item,
2. Tahun transaksi digunakan untuk mengetahui total keseluruhan pengambilan barang berdasarkan item yang akan diolah dengan metode *clustering*.

3.4 Algoritma *K-Means Clustering*

Awalnya, algoritma *K-Means* memanfaatkan beberapa komponen populasi sebagai pusat dari *cluster* awal. Pusat *cluster* dipilih dengan cara random dari satu set data kependudukan di tahap ini. Sebagai langkah selanjutnya, *K-Means* menjalankan tes pada data *swarm* untuk menentukan jarak minimum antara setiap komponen dan masing-masing pusat *cluster* yang telah ditentukan. Setiap kali komponen data baru perlu ditugaskan ke pusat *cluster* tertentu, posisi pusat *cluster* dihitung ulang (Darmi & Setiawan, 2017: 150).

Dalam bukunya (Wahyudi et al. 2020: 6) menegaskan Dengan menggunakan algoritma pengelompokan *K-means*, data dapat dibagi menjadi beberapa kelompok berdasarkan seberapa jauh data dalam kelompok-kelompok itu satu sama lain. Algoritma menggunakan fungsi untuk membandingkan set data

yang memiliki banyak kesamaan. Fungsi *Euclidean* digunakan untuk mengetahui seberapa jauh sesuatu. Selanjutnya, data diurutkan berdasarkan kedekatan dan disimpan.

Berikut merupakan beberapa langkah dalam pengelompokan data (Witanto et al., 2019: 704)

1. Tentukan banyaknya cluster yang hendak dibentuk.
2. Inisialisasi awal dan pusat cluster dilakukan dengan cara random
3. Menurut jarak antara objek, setiap titik data diposisikan di tengah cluster terdekat. Pada titik ini, metode *Euclidean* digunakan untuk menghitung jarak dengan menentukan kesamaan atau perbedaan data. Berikut rumusnya:

$$d(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2}$$

dimana:

“ $d(x, y)$ = ukuran ketidakmiripan

$x_i = (x_1, x_2, x_3, \dots, x_n)$ yaitu ukuran ketidakmiripan

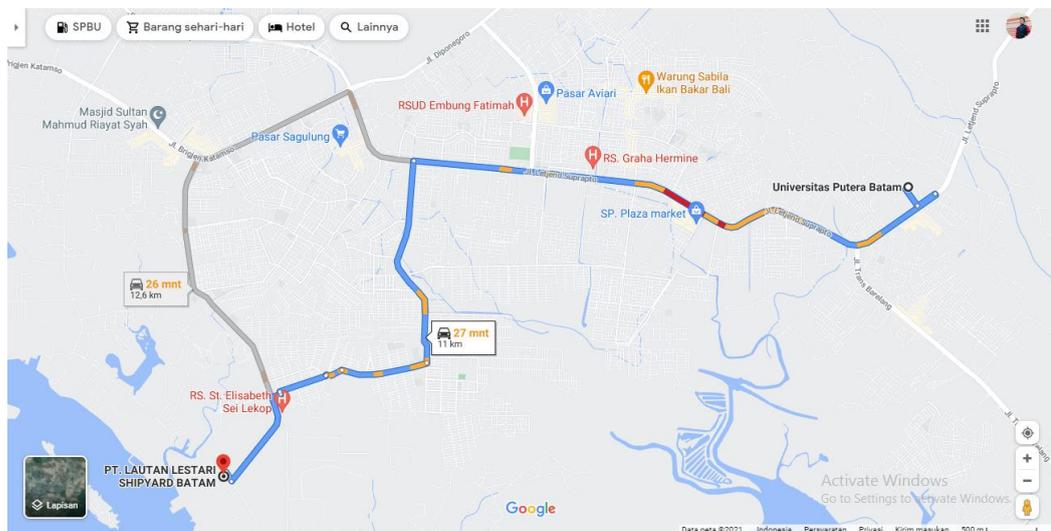
$y_i = (y_1, y_2, y_3, \dots, y_n)$ yaitu variabel pada titik pusat”

4. Rata-rata objek dalam cluster digunakan untuk menghitung pusat cluster. Selain itu, median juga dapat digunakan untuk membuat perhitungan.
5. Untuk menyelesaikan proses pengklasteran, harus menghitung ulang jarak antara setiap objek dan pusat cluster sampai cluster tetap tidak berubah.

3.5 Lokasi dan Jadwal Penelitian

3.5.1 Lokasi Penelitian

Penelitian ini dilakukan di PT Lautan Lestari Shipyards berlokasi di Sei Lekop Kecamatan Sagulung, dengan jarak tempuh sekitar 27 menit dari kampus UPB Tembesi.



Gambar 3.2. Peta Lokasi Penelitian

Sumber: (Peneliti, 2022)

3.5.2 Jadwal Penelitian

Penelitian ini membutuhkan waktu kurang lebih selama 3 bulan yaitu dilakukan pada bulan November 2021 hingga bulan Januari 2022. Berikut adalah jadwal yang peneliti lakukan.

Tabel 3.2 Jadwal Penelitian

Kegiatan	Nov. 2021				Des. 2021				Jan. 2022			
	1	2	3	4	1	2	3	4	1	2	3	4
Observasi												
Pengumpulan Data												
Pengumpulan Referensi												
Pengolahan dan Analisis Data												
Pengumpulan Laporan												

Sumber: (Peneliti, 2022)