

**PENERAPAN *DATA MINING* DALAM
MENENTUKAN STOCK BARANG PRODUK *ONLINE*
SHOP DENGAN METODE *K-MEANS CLUSTERING***

SKRIPSI



Oleh :

Wantri Junedi Manurung

150210169

**PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS PUTERA BATAM
2019**

**PENERAPAN *DATA MINING* DALAM
MENENTUKAN STOCK BARANG PRODUK *ONLINE*
SHOP DENGAN METODE *K-MEANS CLUSTERING***

SKRIPSI

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



Oleh :

Wantri Junedi Manurung

150210169

**PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS PUTERA BATAM
2019**

PERNYATAAN

Dengan ini saya menyatakan bahwa:

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

Batam, 09 Agustus 2019

Yang membuat pernyataan,

Wantri Junedi Manurung

150210169

**PENERAPAN DATA MINING DALAM MENENTUKAN
STOCK BARANG PRODUK *ONLINE* SHOPE DENGAN
METODE *K-MEANS CLUSTERING***

Oleh

Wantri Junedi Manurung

150210169

SKRIPSI

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

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

Batam, 09 Agustus 2019

Yulia, S.Kom., M.Kom

Pembimbing

ABSTRAK

Pengaturan stok yang dilakukan secara tidak akurat, mengakibatkan biaya simpan yang tinggi dan kurang ekonomis. Hal ini tentu akan merugikan semua pelaku bisnis atau usaha sama halnya *online shope*. Penelitian ini bertujuan untuk mengelompokkan produk yang dijual pada *online shope* menjadi lebih dari satu *cluster* untuk melihat produk mana yang paling diminati sehingga jumlah stok harus banyak, produk diminati untuk jumlah stok sedang, dan produk kurang diminati untuk jumlah stok sedikit. Metode yang digunakan dalam penelitian ini adalah *data mining k-means* merupakan salah satu metode dalam algoritma *clustering*, dimana *k-means* akan mencari partisi yang optimal dari data dengan memperkecil kriteria jumlah kesalahan kuadrat dengan prosedur iterasi yang optimal. Variabel yang digunakan adalah kode produk, jumlah transaksi, stok awal barang dan volume penjualan. Penelitian juga menggunakan *software Rapidminer* untuk pengolahan data dengan metode *K-Means*. Hasil akhir dari penelitian adalah berupa suatu program aplikasi yang dapat mengelompokkan produk menjadi kategori jumlah stok banyak, sedang dan sedikit berdasarkan transaksi penjualan.

Kata kunci: *data mining, k-means, stok barang, rapidminer.*

ABSTARCT

Stock management is done inaccurately, resulting in high save costs and less economical. This will certainly harm all business people or businesses as well as online shope. This study aims to classify products that are sold on the habibibe online shope to make more than one cluster to see which products are most desirable so that the number of stocks must be large, products desirable for moderate amounts of stock, and less desirable products for small amounts of stock. The method used in this research is k-means data mining is one of the methods in clustering agorithm, where k-means will find the optimal partition of data by minimizing the criteria for the number of quadratic errors with optimal iteration procedures. The variables used are product code, number of transactions, initial stock of goods and sales volume. The research also uses the Rapidminer software for data processing by the K-Means method. The final result of the research is in the form of an application program that can group products into categories of large, medium and small amount of stock based on sales transactions.

Keywords: *data mining, stock of goods, clustering, rapidminer.*

KATA PENGANTAR

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

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

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

1. Rektor Universitas Putera Batam.
2. Ketua Program Studi Teknik Informatika Universitas Putera Batam Bapak Andi Maslan, ST.,M.SI.
3. Ibu Yulia, S.Kom., M.Kom selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
4. Dosen dan Staff Universitas Putera Batam.
5. Ibu Nana Nurliana sebagai *Onwer online Shope* Habibie yang telah memberikan izin tempat penelitian.
6. Keluarga yang selalu memberikan doa dan motivasi yang baik

7. Rekan-rekan mahasiswa/i Universitas Putera Batam yang turut memberikan doa, dukungan, motivasi dan semangat dalam menyelesaikan skripsi.
8. Dan semua pihak yang telah membantu dalam penulisan skripsi ini
Semoga Tuhan Yang Maha Esa membalas serta kebaikan, kemurahan hati dan memberikan kemudahan dalam segala hal sera mencurahkan kasih karunia-Nya. Amin

Batam, 09 Agustus 2019

Wantri Junedi Manurung

DAFTAS ISI

PERNYATAAN.....	ii
ABSTRAK.....	iv
KATA PENGANTAR	vi
DAFTAS ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiii
DAFTAR RUMUS.....	xiv
BAB 1.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah	3
1.3 Pembatas Masalah.....	4
1.4 Perumusan Masalah.....	5
1.5 Tujuan Penelitian.....	5
1.6 Manfaat Penelitian.....	5
1.6.1 Secara Teoritis	5
1.6.2 Secara Praktis	6
BAB II.....	7
KAJIAN PUSTAKA.....	7
2.1 <i>Knowledge Discovey in Database (KDD)</i>	7
2.2 <i>Data Mining</i>	8
2.2.1 Pengertian <i>Data Mining</i>	8
2.2.2 Tahapan Pada <i>Data Mining</i>	12
2.2.1 Pengelompokann <i>Data Mining</i>	15
2.3 <i>K-MeanssClusteringg</i>	18
2.3.1 <i>K-Means</i>	18
2.3.2 <i>Clustering</i>	21
2.4 <i>Software</i> Pendukung.....	24
2.4.1 UML.....	24
2.4.2 <i>Rapidminerr</i>	29

2.4.3 HTML	30
2.4.4 PHP	31
2.4.5 MySQL	32
2.5 Penelitian Terdahulu	33
2.6 Kerangka Pemikiran	36
BAB III	38
METODE PENELITIAN	38
3.1 Desain Penelitian	38
3.2 Teknik Pengumpulan Data	41
3.2.1 Wawancara	41
3.2.2 Dokumentasi	41
3.2.3 Kajian Pustaka	42
3.3 Operasional variabel	42
3.4.1 Perancangan Data Mining	47
3.4.2 Perancangan Sistem dan Database	49
3.5 Lokasi dan Jadwal Penelitian	62
3.5.1 Lokasi Penelitian	62
3.5.2 Jadwal Penelitian	63
BAB IV	64
HASIL DAN PEMBAHASAN	64
4.1 Analisis Data	64
4.1.1 Seleksi Data	64
4.1.2 <i>Cleaning</i> Data	64
4.1.3 Perhitungan Algoritma	64
4.2 Hasil Pengujian	76
4.3 Halaman Antarmuka Sistem	79
4.4 Pengujian Unit	88
BAB V	93
5.1 KESIMPULAN	93
DAFTAR PUSTAKA	95
DARTAR RIWAYAT HIDUP	
SURAT IZIN PENELITIAN	

LAMPIRAN.....

DAFTAR GAMBAR

	Halaman
Gambar 2. 1 Tahapan Knowledge Discovery in Database.....	8
Gambar 2. 2 Tahapan pada Data Mining.....	12
Gambar 2. 3 Pengelompokan Data Mining	16
Gambar 2. 4 Kerangka Pemikiran	37
Gambar 3. 1 Desain Penelitian	39
Gambar 3. 2 Rancangan Data Mining	47
Gambar 3. 3 Use Case diagram sistem	49
Gambar 3. 4 Activity diagram Login.....	50
Gambar 3. 5 Activity diagram data barang.....	51
Gambar 3. 6 Activity diagram data transaksi	52
Gambar 3. 7 Activity diagram Data Pembeli	53
Gambar 3. 8 Activity diagram data mining Clustering	54
Gambar 3. 9 Activity diagram clustering result	55
Gambar 3. 10 Activity diagram logout.....	56
Gambar 3. 11 Sequence diagram Login	57
Gambar 3. 12 Diagram sequence data barang	57
Gambar 3. 13 Diagram sequence Data Transaksi.....	58
Gambar 3. 14 Diagram sequence data pembeli	59
Gambar 3. 15 Diagram sequence data mining clustering.....	59
Gambar 3. 16 Diagram sequence Clustering result	60
Gambar 3. 17 Activity diagram Logout.....	61
Gambar 3. 18 Class Diagram sistem	62
Gambar 4. 1 Plot Hasil Clustering Rapidminer.....	77
Gambar 4. 2 Hasil cluster model	77
Gambar 4. 3 Tampilan pengelompokan cluster C0	78
Gambar 4. 4 Hasil pengelompokan cluster C1	78
Gambar 4. 5 Hasil pengelompokan custer C2	79
Gambar 4. 6 Tampilan Halaman Login	80
Gambar 4. 7 Tampilan Keseluruhan Menu Sistem.....	80
Gambar 4. 8 Tampilan Halaman data barang.....	81
Gambar 4. 9 Tampilan Halaman Transaksi Barang	82
Gambar 4. 10 Tampilan Halaman Menambahkan Transaksi	82
Gambar 4. 11 Tampilan Halaman Pembeli	83

Gambar 4. 12	Add customer baru.....	83
Gambar 4. 13	Tampilan Halaman cluster.....	84
Gambar 4. 14	Tampilan Halaman cluster.....	84
Gambar 4. 15	Tampilan halaman menu cluster.....	85
Gambar 4. 16	Tampilan Halaman Tambah cluster C0.....	85
Gambar 4. 17	Tampilan halaman tambah cluster C1.....	86
Gambar 4. 18	Tampilan Halaman Tambah cluster C2.....	86
Gambar 4. 19	Tampilan clustering result.....	87
Gambar 4. 20	Tampilan clustering result.....	88

DAFTAR TABEL

	Halaman
Tabel 2. 1 Use case Diagram.....	25
Tabel 2. 2 Simbol Diagram Activity	26
Tabel 2. 3 Class Diagram	27
Tabel 2. 4 Sequence Diagram.....	28
Tabel 3. 1 Tabel Barang	42
Tabel 3. 2 Tabel Pembeli.....	44
Tabel 3. 3 Tabel Transaksi	45
Tabel 3. 4 Operasional Variabel.....	45
Tabel 3. 5 Jadwal Penelitian.....	63
Tabel 4. 1 Titik pusat cluster	65
Tabel 4. 2 Perhitungan Jarak data iterasi 1	68
Tabel 4. 3 Cluster baru iterasi ke-1	70
Tabel 4. 4 Hasil Perhitungan jarak dan pengelompokan data iterasi	70
Tabel 4. 5 Centroid baru iterasi ke-2.....	72
Tabel 4. 6 Hasil perhitungan jarak dan pengelompokan data iterasi ke-3.....	73
Tabel 4. 7 Hasil pengelompokan iterasi 1, 2 dan 3	74
Tabel 4. 8 Hasil pengelompokan produk jumlah banyak	75
Tabel 4. 9 Hasil pengelompokan produk jumlah menengah	76
Tabel 4. 10 Hasil pengelompokan produk jumlah sedikit.....	76
Tabel 4. 11 Pengujian Login	89
Tabel 4. 12 Pengujian data barang	89
Tabel 4. 13 Pengujian Halaman Transaksi	90
Tabel 4. 14 Pengujian data pembeli	90
Tabel 4. 15 Pengujian data mining clustering	91
Tabel 4. 16 Pengujian clustering result	92
Tabel 4. 17 Pengujian logout.....	92

DAFTAR RUMUS

	Halaman
Rumus 2. 1 Relokasi cluster 1	21
Rumus 2. 2 Relokasi cluster 0	21
Rumus 2. 3 Menghitung centroid cluster ke-1	23
Rumus 2. 4 Jarak masing-masing centroid.....	23
Rumus 4. 1 Euclidian Distance.....	65

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Teknologi internet saat ini berkembang sangat pesat terutama dalam dunia bisnis, hal ini dapat dilihat dengan kedatangan *electronic commerce (e-commerce)*. E-commerce dalam kegiatan perdagangan dan penjualan bisa dikenal dengan nama Toko *Online* atau *Online Shop* diterapkan oleh berbagai lembaga atau perusahaan.

Online Shop adalah proses pembelian transaksi barang atau jasa melalui media teknologi internet. *Online Shop* Habibie toko *online shop* yang ada di kota Batam. Toko *Online Shop* ini menjual barang antara lain : alat - alat rumah tangga, perabot rumah tangga, hiasan, *accessories, fashion* dari harga yang murah hingga barang - barang mahal. Distribusi *Online Shop* Habibie ini meliputi seluruh wilayah kota Batam. Konsumen melakukan pembelian melalui media sosial, setelah konsumen melakukan pemesanan, *Online Shop* Habibie memiliki kurir atau karyawan 15 orang untuk melakukan pengantaran ke alamat yang dituju diseluruh wilayah Batam. Sekarang ini *Online Shop* Habibie melakukan pencatatan transaksi secara manual.

Jumlah pemasok dari *customer* dan pelanggan yang fruktusi menimbulkan *stock* yang harus direadykan *Online Shop* Habibie menjadi kurang stabil. Selain dari pada itu produk yang beraneka ragam dan jenisnya banyak menjadikan

manajemen *stock* yang dilakukan menjadi tidak akurat. Sese kali supaya tidak terjadi kekurangan *stock* produk tertentu pada waktu jumlah permintaan *costumer* dan pelanggan dalam jumlah banyak jadi Toko *Online Habibie* melakukan pengorderan produk tertentu lebih banyak daripada sebelumnya. Strategi ini dilakukan *online Shop* Habibie untuk mengurangi *stock* barang dan *Online Shop* juga tidak lagi berulang-ulang melakukan pengorderan berulang-ulang ke produksi utama akan tetapi menimbulkan biaya simpan yang tinggi dan kurang ekonomis.

Permasalahan yang sering terjadi di *online shop* yang disebabkan oleh *Online Shop* yang mempengaruhi dalam menentukan persediaan minimum setiap barang yang harus di penuhi berdasarkan *costumer* dan pelanggan. *Online Shop* membutuhkan startegi dan perencanaan stok barang jauh lebih baik yang mampu menentukan persediaan barang atau produk mana yang mestinya di banyak, sedang, sedikit dan juga supaya *Online Shop* tidak perlu lagi meningkatkan atau menambah dalam mengisi stok produk tertentu dalam mengatasi permasalahan tersebut.

Cara statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengelola informasi yang berguna dan pengetahuan yang berkaitan dengan basis data yang besar dengan tahapan-tahapan yang dipakai disebut dengan *data mining*. Metode *K-Means* ialah metode pengelompokan yang simpel dan umum. Hal ini menyangkut *skill* pengelompokkan data dengan jumlah yang cukup banyak dengan rentang waktu perhitungan yang relatif singkat dan efisien dimiliki oleh *K-Means*, Akan tetapi

metode *K-Means* mempunyai kekurangan yang diakibatkan oleh asosiasi pusat mula *cluster* (Koko Handoko, 2018).

Pada penelitian sebelumnya, (Metisen, 2015), dalam melakukan pengetesan algoritma ini data yang digunakan adalah data barang diswalayan Fadhila Bengkulu yang diolah dan menghasilkan dua jenis data yakni data *low* pemasaran dan data *high* pemasaran. Dengan di pengelompokannya data ini, swalayan dapat mengetahui jenis barang yang banyak laku dan tidak banyak laku sehingga barang yang di gudang tidak menumpuk dengan menggunakan *data mining K-means clustering*.

Dari permasalahan di atas, mengingat betapa pentingnya untuk menentukan stock barang di toko *Online Shope*, sehingga menarik perhatian untuk membuat karya ilmiah yang berjudul: **“PENERAPAN DATA MINING DALAM MENENTUKAN STOCK BARANG PRODUK ONLINE SHOP DENGAN METODE K-MEANS CLUSTERING.”**

1.2 Identifikasi Masalah

Permasalahan timbul pada *scope* Toko *Online Shop* yang menggerakkan seseorang untuk melakukan observasi agar permasalahan yang dapat terselesaikan yang akan di jabarkan sebagai berikut:

1. Jumlah permintaan dari *costumer* dan pelanggan yang frukuasi menimbulkan *stock* yang harus disiapkan *Online Shop* Habibie menjadi kurang stabil.

2. Karena jenis produk yang di jual di toko *Online Shop* banyak menjadikan manajemen *stock* yang dilakukan menjadi tidak akurat.
3. Pemesanan produk tertentu lebih besar dari pada sebelumnya yang dapat mengurangi persediaan *stock* barang dan *Online Shop* juga tidak perlu berulang-ulang melakukan pengorderan ke produksi utama menimbulkan *cost* simpan yang tinggi dan kurang ekonomis.
4. Toko *Online Shop* Habibie mengalami kendala dalam menentukan *stock* minimum setiap produk yang harus dipenuhi melalui minat *costumer*.

1.3 Pembatas Masalah

Dengan penjelasan permasalahan di atas peneliti perlu melakukan pembatasan permasalahan supaya terselesaikan dengan terarah dan sesuai manfaat dan tujuan yang di harapkan. Berikut pembatasan permasalahan dalam karya ilmiah ini adalah:

1. Data yang diproses merupakan data pengelompokan *stock* barang jumlah banyak, sedang dan sedikit.
2. Penelitian ini dilakukan di Toko *Online Shop* Habibie.
3. Didalam karya ilmiah ini menerapkan *data mining* dengan metode *K-Means Clustering*.
4. Penelitian ini memakai *software RapidMiner* sebagai pendukung pengolahan data dalam menentukan *stock* barang produk *Online Shop*.
5. Perancangan dan pembuatan sistem menerapkan bahasa pemograman *PHP*, *MySQL*, dan *HTML* berbasis *web*.

1.4 Perumusan Masalah

Dengan penjelasan masalah diatas, supaya pembahasan dan penyelesaian permasalahan tidak simpang siur sehingga rumusan masalah tersebut adalah Bagaimana penerapan *data mining* dalam menentukan *stock* barang produk *Online Shop* dengan metode *K-Means Clustering* berbasis *web*?

1.5 Tujuan Penelitian

Dengan permasalahan yang tersedia dan upaya untuk mempermudah masyarakat awam dalam mengetahuinya , menjadikan tujuan dari karya ilmiah ini untuk menentukan *stock* barang produk *Online Shop* dengan memakai prinsip *data mining* metode *K-Means Clustering* berbasis *web*.

1.6 Manfaat Penelitian

Secara khusus, karya ilmiah ini sangat berpeluang memberikan manfaat baik dari aspek teoritis maupun aspek praktis.

1.6.1 Secara Teoritis

Berikut ini yang dirangkum manfaat teoritis dari karya ilmiah ini yaitu:

1. Pengelompokan produk atau barang ini dapat diterapkan dalam proses penjualan sehari-hari diseluruh toko *Online shop*.
2. Karya ilmiah ini dapat memperluas *knowlegde* khususnya *data mining* dengan metode *K-Means Clustering*.

1.6.2 Secara Praktis

Berikut ini dirangkum manfaat praktis karya ilmiah ini yaitu:

1. Untuk Toko *Online Shop* Habibie

Penelitian ini sangat berpeluang dapat membantu pihak toko *Online Shop* dapat menentukan stok barang yang tersedia.

2. Bagi Akademik

Kegunaan karya ilmiah ini bagi lembaga penelitian adalah untuk materi pembelajaran bagi mahasiswa dan sumbangan pemikiran tentang cara menentukan stok barang produk *online shop* dengan menerapkan *data mining* metode *K-Means Clustering*.

3. Bagi Peneliti

Bisa menambah pengetahuan sebagai bekal meningkatkan kemampuan dan pengetahuan dalam mengimplementasikan *data mining* metode *K-Means Clustering* dalam menentukan stok barang produk *Online Shop*. Lainnya juga penulis mendapat pengalaman baru dalam menyelesaikan tugas akhir yang bisa bermanfaat untuk tempat peneliti melakukan penelitian.

BAB II

KAJIAN PUSTAKA

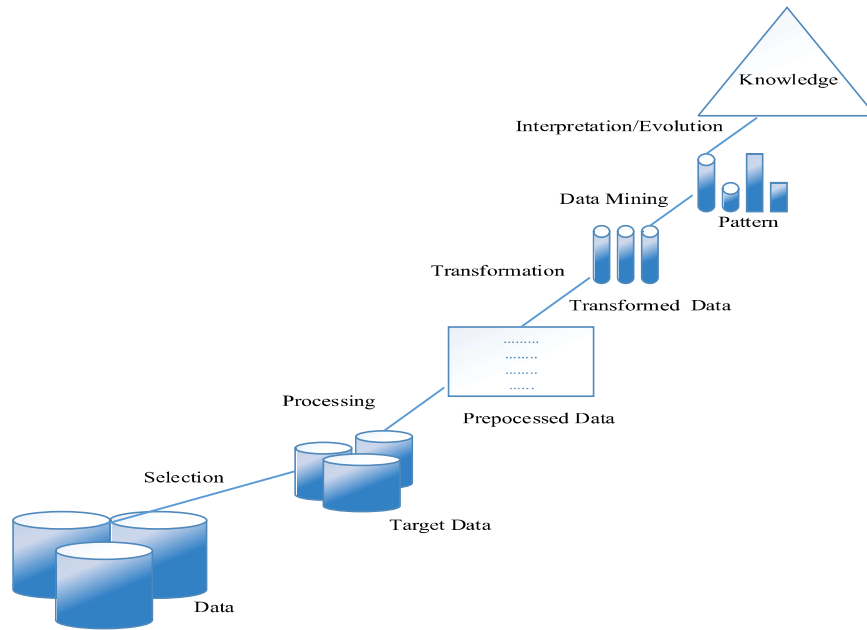
2.1 *Knowledge Discovey in Database (KDD)*

Sekarang ini, prinsip *data mining* semakin terkenal untuk alat dalam manajemen informasi karena jumlah informasi yang semakin tinggi. *Data mining* juga dikatakan sebagai *Knowledge Discovey in Database (KDD)* merupakan aktivitas yang terkait mengumpulkan, penggunaan data yang lalu untuk mendapatkan keteraturan, pola hubungan dalam mengatur data diperoleh besar. Lebih dari satu macam pendekatan yang tidak sama yang dikelompokan sebagai suatu cara pencarian informasi/ pengetahuan dalam *Knowledge Discovery in Database* KDD. Ada pendekatan kuantitatif, seperti pendekatan probabilistik seperti logika induktif, pencarian pola, dan analisis pohon keputusan (Koko Handoko, 2018).

Mula-mulanya ada enam elemen yang paling esensial dalam teknik pencarian informasi / pengetahuan dalam KDD yaitu:

1. Data yng berjumlah besar dikerjakan.
2. Yang berkaitan dengan volume data dibutuhkan efisiensi.
3. Mengedepankan kebenaran/keakuratan.
4. Bahasa tingkat tinggi diperlukan.
5. Memakai lebih dari satu bentuk dari *studi* otomatis.

6. Mendapatkan hasil yang memberentang.



Gambar 2. 1 Tahapan *Knowledge Discovery in Database*

Sumber : (Retno Tri Vulandari, 2017)

2.2 *Data Mining*

2.2.1 Pengertian *Data Mining*

Pada penelitian ini diuraikan tentang berbagai teori dalam penerapan *data mining*, kumpulan tahapan untuk mencari nilai tambah berupa informasi yang tidak di ketahui secara manual dari suatu database disebut dengan *data mining*. Informasi yang di hasilkan di peroleh melalui teknik mengestrak dan mengenali pola yang penting dan menarik dari data yang terdapat di database.

Tahapan yang memakai cara statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengelola informasi yang berguna

dan *knowledge* yang berkaitan dengan basis data yang besar dikenal dengan *data mining*. *Data Mining* umumnya mempunyai akar yang besar dari bidang seperti kecerdasan buatan, statistik dan database. Lebih dari satu cara yang sering disebut dalam literatur *data mining* antara lain: *clustering*, klasifikasi, asosiasi aturan data, jaringan saraf, dan algoritma genetika (Koko Handoko, 2018).

Menurut (Retno Tri Vulandari, 2017, pp. 1-2), proses yang disusun bagi menambah nilai tambah informasi yang selama ini tidak dipublikasikan secara manual dari suatu database yang didapat dengan teknik mengekstraksi dan menyusun pola yang penting atau menarik dari data yang diperlukan pada database disebut dengan *data mining*. *Data mining* adalah salah satu rangkaian dari *Knowledge Discovey in Database (KDD)* yang saling terkait dengan cara integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola sejumlah data. Serangkaian proses tahapan *data mining* tersebut adalah :

1. Pembersihan data (untuk melepaskan data yang tidak berubah-ubah dan *noise*).
2. Integrasi data (penggabungan data dari banyak sumber).
3. Transformasi data (data diubah menjadi bentuk yang selaras untuk *data mining*).
4. Aplikasi cara *data mining*, alur ekstraksi pola dari data yang tersedia.
5. Evaluasi pola yang ditemukan (alur interpretasi pola menjadi *knowledge* yang bisa dimanfaatkan untuk mengambil keputusan)

6. Presentasi pengetahuan .

Tahapan ini adalah proses pencarian *knowledge* yang membahas tentang pola atau informasi yang diperoleh melalui fakta atau dugaan sementara yang tersedia sebelumnya.

Menurut (Prasetyo, 2014) jika dilacak dari akar keilmuannya, *data mining* memiliki empat akar bidang ilmu yaitu :

1. Statistik

Bidang ini ialah akar tertua, tanpa statistik, mungkin *data mining* tidak ada. Memakai statistik *data mining* klasik dapat disimpulkan apa yang umum dikenal sebagai *exploratory data analysis* (EDA) yang bermanfaat untuk memfasilitasi *relasi* antarvariabel yang rumit/fitur kompilasi tidak ada informasi yang cukup yang dibawanya. Teknik eksplorasi data analisis *exploratory data analysis* klasik yang dipakai dalam *data mining* antara lain :

- a. Metode Komputasional : *statistic deskriptif* (distribusi, parameter statistik klasik (rata-rata, nilai tengah, rata-rata varian, dan lainnya), korelasi, tabel frekuensi, teknik eksplorasi *multivariat* (analisis *cluster*, analisis *factor*, analisis komponen utama dan klaifikasi, analisis karonik, analisis diskriminan, *classification tree*, analisis korespondensi) model linear/nonlinear lanjutan (regresi linear/nonlinear, *time series/forecasting*, dan sebagainya).

- b. Visualisasi data: mengarah pada representasi informasi dalam bentuk visual dan dapat dipandang sebagai satu yang paling berguna. Teknik visualisasi dikenal dengan histogram semua jenis (kolom, silinder, keruvut, piramida, lingkaran, batang, dan sebagainya), kotak, *scatter*, kontur, matriks, ikon dan sebagainya.

2. Kecerdasan Buatan atau *artificial intelligence*

Teori diciptakan melalui teknik heuristik berdasarkan kecerdasan buatan berkontribusi pada pemrosesan informasi berdasarkan pada model penalaran manusia, salah satu cabangnya dari pembelajaran mesin, yaitu disiplin ilmu yang utama yang direpresentasikan dalam bentuk pembangun *data mining*, menggunakan teknik dengan sistem komputer belajar dengan pelatihan.

3. Pengenalan Pola

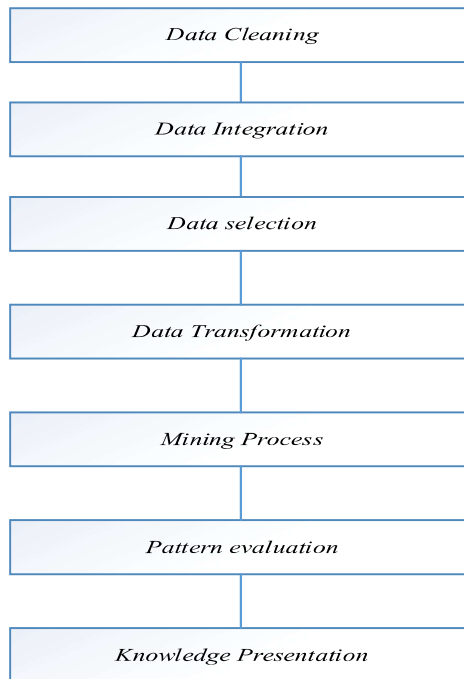
Data mining menjadi turunan pola tetapi hanya memproses data dari database. Data yang diambil dari data dasar untuk diolah bukan dalam bentuk relasi yang diperlukan dalam bentuk normal pertama, akan tetapi *data mining* memiliki ciri khas seperti pencarian pola asosiasi dan pola sekuensial.

4. Sistem Basis data

Memfasilitasi informasi terdiri dari data yang dicari menggunakan metode sebelumnya ialah akar pengetahuan dari *data mining*.

2.2.2 Tahapan Pada *Data Mining*

Berikut ini ialah tahapan *data mining* yang dapat dijabarkan :



Gambar 2. 2 Tahapan pada *Data Mining*

Sumber : (Data Mining Concepts and Techniques, 2017)

1. *Data Cleaning* (Pembersihan Data)

Proses menghapus data yang tidak relevan dalam suatu *dataset*. Data yang tidak relevan contohnya data yang gugur, data tidak *valid* dan data yang salah ketik akan dihapus supaya kualitas *data mining* menjadi lebih ringan karena jumlah yang akan dibutuhkan sedikit dan data menjadi lebih mudah disebut *Data cleaning* atau pembersihan data.

2. *Data Integration* (Integrasi Data)

Data integration atau Integrasi data adalah hasil perpaduan dari dua atau data dari banyak basis data yang diinputkan ke dalam basis data integrasi baru harus dilaksanakan dengan cermat karena kekeliruan pada proses ini bisa menyelesaikan dalam membuat tindakan berikutnya.

3. *Data Selection* (Seleksi Data)

Data Selection atau Seleksi data ialah tahapan pemilihan data yang dipakai karena sebagian data pada basis data akan dipakai dalam karya ilmiah. Hanya data yang selaras dengan standard peyelidikan yang akan diangkut dari basis data untuk diobservasi.

4. *Data Transformation* (Transformasi Data)

Data transformation atau transformasi data ialah bentuk penggantian data ke dalam format yang searah untuk ditindaklanjuti. Lebih dari satu metode *data mining* memerlukan jenis format data yang tidak sama. Seperti pada pengelompokan, input data harus dalam bentuk data kategorial supaya bisa dilakukan, menjadikan pada proses penggantian format data itu dilakukan.

5. *Mining Process* (Proses *Mining*)

Mining process atau proses *mining* ialah tahap yang paling didepan. Seluruh tahap sebelumnya ialah untuk mensupport tahap ini. Tahapan ini adalah tahapan memakai metode yang tersedia untuk mendapatkan *knowledge* dari data yang tersedia.

6. *Pattern Evaluation* (Evaluasi Pola)

Pattern Evaluation (Evaluasi Pola) ialah proses dalam mengenali pola yang tersedia melalui *result* dari proses *mining* yang dilakukan dan menarik kesimpulan berupa pengetahuan untuk menilai apakah hipotesis yang ada telah terpenuhi atau belum.

7. *Knowledge Presentation* (Presentasi Pengetahuan)

Knowledge presentation atau presentasi pengetahuan tingkatan visualisasi atau pengutaraan *result knowledge* yang telah diperoleh dalam penjabaran supaya diterima oleh seluruh manusia.

Menurut (Retno Tri Vulandari, 2017, p. 7), Teknik pembelajaran dalam *data mining* sangat berhubungan dengan penemuan (*discovery*) dan penataran (*learning*) yang terurai menjadi tiga metode utama penataran yaitu:

a. *Supervised Learning*

Supervised Learning ialah cara yang paling jamak dipakai yang *similar* dengan *programming by examples*. Cara ini mengaitkan bagian pelatihan di mana data pelatihan yang lalu yang keunikannya dipetakan ke *result* yang telah dipahami diolah dalam algoritma *data mining*. Proses ini melatih algoritma untuk mengevaluasi variabel-variabel dan kadar kunci yang akan dipakai sebagai kunci dalam membuat target kompilasi yang diperembahkan oleh data baru.

b. *Unsupervised Learning*

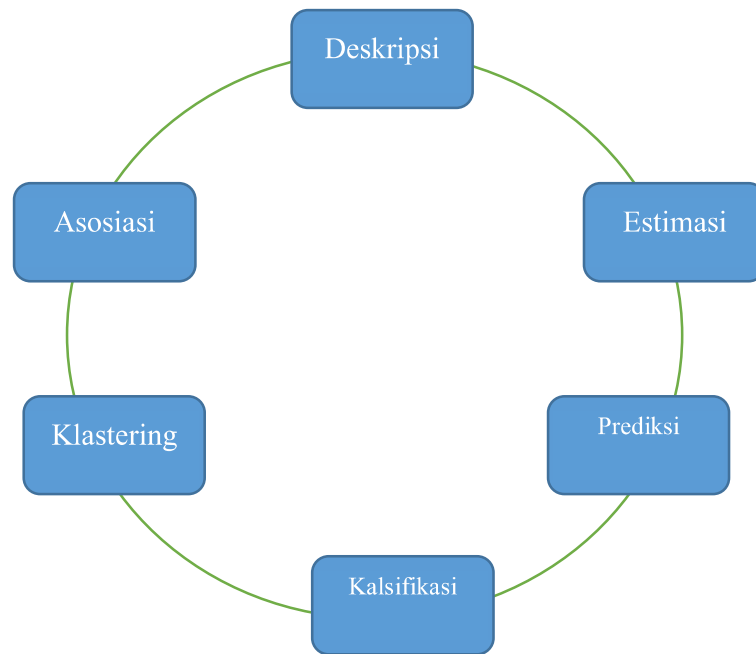
Cara pembelajaran ini kebalikan dari *supervised Learning* dengan tidak melibatkan fase pelatihan. Teknik ini membahas tentang pemakaian algoritma yang menghubungkan seluruh pola, contohnya asosiasi dan urutan, yang timbul pada standard penting yang khusus dalam data masukan. Tujuan ini mengkarakteristikan penemuan asosiasi, *cluster*, dan segmen. Norma-norma ini selanjutnya dianalisis untuk mendapatkan hal-hal yang penting.

c. *Reinforcement Learning*

Cara pembelajaran ini jarang dipakai namun mempunyai implementasi-implementasi yang terus dimaksimalkan dari waktu ke waktu dan mempunyai kontrol adaptif. *Reinforcement Learning* cocok dipakai untuk penyelesaian permasalahan yang sulit ketergantungan waktu.

2.2.1 Pengelompokan *Data Mining*

Data Mining dikelompokkan pada beberapa kelompok berlandaskan pekerjaan yang bisa dilaksanakan, yakni:



Gambar 2. 3 Pengelompokan *Data Mining*

Sumber : (Dr. Suyanto, 2017)

1. Deskripsi

Deskripsi dari bentuk kecondongan sering mempersembahkan peluang penjelasan untuk suatu pola atau kecondongan.

2. Estimasi

Tidak jauh berbeda atau sama besar dengan klasifikasi, kecuali variabel estimasi diperkirakan lebih ke numerik dan kategori. Model diciptakan memakai catatan *complete* yang menyajikan nilai variabel estimasi sebagai prediksi. Kemudian, pada tinjauan selanjutnya estimasi nilai dari variabel dibuat melalui kadar variabel prediksi.

3. Prediksi

Beda tipis dengan deskripsi dan perkiraan, kecuali dalam prediksi kadar *result* akan di masa depan. Ada lebih dari satu metode dalam deskripsi dan estimasi yang dapat dipakai untuk prediksi.

4. Klasifikasi

Dalam klasifikasi ini, estimasi variabel kategori. Contohnya, pengklasifikasian penghasilan bisa diterima dalam 3 golongan, yakni: penghasilan tinggi, penghasilan sedang dan penghasilan rendah.

5. Klastering

Pengklusteran yaitu klastering catatan, mengamati dan membentuk kelas objek yang mempunyai *similarity*. *Cluster* ialah himpunan catatan yang mempunyai *similarity* dengan *record* dengan *cluster* lain. Pengklusteran tidak sama dengan klasifikasi yakni tidak adanya variabel estimasi dalam pengklusteran. Pengklusteran tidak melakukan percobaan untuk klasifikasi, menargetkan, atau memperhitungkan nilai dari variabel estimasi. Melainkan, algoritma pengklusteran berusaha untuk melaksanakan pembagian terhadap diskusi data yang menuju kelompok yang memiliki *similarity (homogeneity)*, dimana *similarity* dalam suatu kelompok akan bernilai maksimal, sedangkan *similarity* dengan catatan dalam kelompok lain akan menjadi minimal.

6. Asosiasi

Fungsinya di *data mining* ialah mendapatkan atribut yang timbul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.3 *K-Means Clustering*

2.3.1 *K-Means*

Pada karya ilmiah ini, *data mining* yang digunakan ialah pengelompokan dengan memakai perpaduan antara *K-Means* dengan pengelompokan hierarkis. Dari algoritma *hierarchical clustering* dipakai untuk menentukan pusat *cluster*. Kemudian pusat klaster yang telah didapat klaster hirarkis itu dipakai untuk tahapan *clustering* data dengan memakai metode *K-means*. *K-Means* adalah algoritma *clustering* yang berkali-kali. Algoritma *K-Means* mematok nilai-nilai *cluster* (K) secara acak, untuk beberapa saat nilai tersebut menjadi pusat dari *cluster* atau bias disebut dengan *centroid*, mean atau "*means*". Kemudian menghitung jarak antar data yang ada pada setiap *centroid* memakai rumus *Euclidian* sampai ditemukan jarak yang paling dekat dari setiap data *centroid*. Golongkan setiap data berlandaskan kedekatannya dengan *centroid* (Retno Tri Vulandari, 2017, p. 54).

Menurut (Mudzakkir, 2018, p. 35), *K-Means* ialah metode pengelompokan data non hierarki yang memerlukan partisi data yang tersedia ke dalam bentuk satu atau lebih *cluster* atau group yang memerlukan data yang mempunyai fitur yang similar digolongkan ke dalam satu *cluster* yang sama dan data yang memiliki fitur yang tidak sama dengan kelompok yang lain. *K-Means* adalah metode pengelompokan berbasis jarak yang membagi data ke dalam jumlah kluster algoritma ini untuk atribut numerik. Algoritma *K-Means* terbilang pengelompokan partisi yang merenggangkan data ke k wilayah penggalan yang

terputus. Algoritma *K-Means* sangat terkenal karena kemudahan dan kemampuannya untuk mengcluster data yang besar dan data outlier dengan sangat cepat. Dalam algoritma *K-Means*, masing-masing data mestinya terbilang ke dalam *cluster* tertentu dan dapat diakses untuk setiap data yang termasuk *cluster* tertentu pada suatu proses tahapan, pada tahap selanjutnya pindah ke *cluster* lain.

Menurut Widodo (2013:9) dalam jurnal (Mudzakkir, 2018, p. 35), Mulanya pemakaian algoritma dalam melaksanakan proses *clustering* ketergantungan dari data yang tersedia dan konklusi yang hendak diraih. Maka dari itu dipakai algoritma *K-Means* yang didalamnya membuat patokan yakni:

- 1) Besaran *cluster* perlu dimasukkan.
- 2) Tipe atribut adalah angka

Algoritma *K-Means* ialah pola nonhierarki yang mengangkut alih sebagian dari komponen untuk diaktualkan sebagai pusat *cluster* awal. Pada kelompok ini pusat *cluster* ditentukan secara *random* dari sekumpulan populasi data. Berikutnya *K-Means* setiap elemen di dalam populasi data dan elemen cadangan ke salah satu pusat *cluster* yang telah ditentukan terpaut dari jarak minimum antar elemen dengan masing-masing *cluster*. *Cluster* pusat posisi akan dihitung kembali hingga semua komponen data dikelompokkan menjadi *cluster* setiap-pusat dan yang terakhir akan membentuk *Cluster* pusat yang baru. Algoritma *K-Means* mulanya melakukan duatahapan, yaitu tahapan pendeteksian lokasi pusat setiap kluster dan tahapan pencarian anggota dari setiap kluster. Cara kerja algoritma *K-Means*:

1. menentukan k sebagai besaran *cluster* yang hendak dibentuk.

2. membangkitkan k *centroid*(titik center *cluster*) mula secara acak.
3. menghitung itung jarak seluruh data ke tiap-tiap *centroid*.
4. *centroid* yang terdekat dipilih untuk setiap data.
5. data-data yang terletak pada *centroid* yang sama dengan cara menghitung nilai rata-rata untuk menentukan posisi nilai *centroid* yang baru.
6. jika posisi *centroid* baru dengan *centroid* masih berbeda, ulani proses ke-3.

Nilai k adalah parameter yang hendak diinputkan ketika memakai algoritma *K-Means*. Nilai k yang didasarkan pada informasi diketahui sebelumnya tentang sebenarnya berapa banyak *cluster* yang dibutuhkan untuk penerapannya, atau jenis *cluster* dicari dengan beberapa nilai k . Dalam *K-Means*, setiap *cluster* dari k diwakili titik tunggal dalam \mathbb{R}^d . Set representatif *cluster* dinyatakan $C = \{c_j | j=1, \dots, K$. Sejumlah K *representative* disebut juga sebagai *cluster means* atau *cluster centroid*. Untuk *set* data dalam X dikelompokkan berlandaskan *similarity*. Walaupun konsep yang diartikan untuk data-data yang berkumpul dalam satu *cluster* adalah data-data yang mirip, tetapi kuantitas yang digunakan untuk dilakukan pengukuran adalah ketidakmiripan (*dissimilarity*) untuk bergabung dalam satu *cluster* data-data ketidakmiripan yang kecil/dekat. Relokasi dalam sebuah *cluster* yang dinyatakan dalam nilai keanggotaan a yang bernilai 0 dan 1. Nilai 0 jika tidak menjadi anggota sebuah *cluster* dan nilai 1 jika menjadi anggota dalam sebuah *cluster*. Karena *K-Means* mengelompokkan secara tegas data hanya pada satu *cluster*, maka dari nilai a sebuah data dalam sebuah *cluster* hanya bernilai 1 sedangkan lainnya 0 dinyatakan dalam persamaan berikut:

$$a_{ij} = \int_0^1 \arg\{d(x_1, c_1)\} \quad \text{Rumus 2. 1 Relokasi cluster 1}$$

Jika diperhatikan langkahnya yang selalu memilih *cluster* yang dekat, maka sebenarnya *K-Means* berusaha untuk meminimalkan fungsi objek/fungsi biaya non-negatif dinyatakan dalam persamaan berikut :

$$J = \sum_{i=1}^N \sum_{c=1}^K a_{ic} d(x_i, c)$$

Rumus 2. 2 Relokasi cluster 0

2.3.2 Clustering

Menurut Widodo (2013:9) dalam jurnal (Mudzakkir, 2018, p. 35), metode yang dipakai untuk membagi kumpulan data menjadi lebih dari satu kelompok yang disetujui-disetujui yang telah ditentukan sebelumnya adalah *Clustering* atau klasifikasi. *Cluster* adalah kumpulan objek-objek data yang serupa satu dengan yang lain dalam *cluster* yang sama dan tidak sama terhadap objek-objek yang tidak sama *cluster*. Objek akan digolongkan ke dalam satu atau lebih kluster sehingga objek-objek yang berada di dalam satu kluster akan memiliki pertimbangan yang tinggi antara satu dengan yang lainnya. Objek – objek digolongkan berlandaskan prinsip maksiminasi objek pada kluster yang similar dan maksimalkan keberbedaan pada kluster yang tidak sama.

Pengelompokan pada suatu data pada suatu penggabungan untuk mengelompokkan suatu data himpunan yang atribut kelas tidak dijelaskan, berlandaskan kosepnya prinsip pengelompokan adalah untuk mengoptimalkan dan menggunakan kemiripan antar kelas. Misalnya, ada objek himpunan, langkah pertama dapat diklasterisasi menjadi lebih dari satu kelas himpunan selanjutnya

menjadi himpunan beraturan sehingga dapat diperoleh melalui klasifikasi tertentu. Analisis *clustering* ialah salah satu dari cara multivariat metode *interdependensi* (saling bergantung). Analisis *clustering* ialah suatu tahapan yang dipakai sebagai metode perpaduan pengamatan ke dalam golongan, sehingga:

- a. Karakteristik tertentu dimiliki oleh setiap himpunan homoge. Sehingga, pengamatan di satu kelompok yang sama similar dengan pengamatan di setiap kelompok.
- b. Setiap himpunan tidak sama dengan himpunan lain. Dengan begitu, pengamatan dalam himpunan satu tidak sama dari pengamatan dalam himpunan lain.

Pengelompokan data ialah bagian dari metode *data mining* yang mempunyai sifat tidak sejalan (tidak diawasi). Data ada dua jenis metode yang sering dipakai dalam pengelompokan data yakni pengelompokan data hierarkis dan pengelompokan data non-hierarkis. Algoritma *k-means* adalah bagian dari metode pengelompokan data non-hierarkis. (Nur, 2018, p. 2246).

Suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki pengumpulan paling dekat adalah *Hierarchical clustering*. Selanjutnya proses *dikontinue* ke objek wisata yang mempunyai kedekatan kedua. Demikian seterusnya sehingga *cluster* akan membuat pohon di mana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara keseluruhan, semua objek pada akhirnya hanya akan menata sebuah *cluster*. (Koko Handoko, 2018) . *K-Means Clustering*

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

1. Besaran *cluster* yang di bentuk dengan cara mematok nilai k. Untuk mematok banyaknya *cluster* k dilaksanakan lebih dari satu evaluasi seperti evaluasi teoritis dan konseptual yang mungkin dianjurkan untuk mematok berapa banyak *cluster*.
2. Bangun k *centroid* (titik *center cluster*) awal secara *random*. Penentuan *centroid* mula-mula dilaksanakan secara random dari objek-objek yang ada sebanyak k *cluster*, selanjutnya untuk memeperkirakan *centroid cluster* ke-i berikutnya, digunakan rumus sebagai berikut :

$$V = \sum_{i=1}^n X_i \quad ; i = 1, 2, 3, \dots n$$

Rumus 2. 3 Menghitung *centroid cluster* ke-1

3. Jarak setiap objek ke tiap-tiap *centroid* d menghitung jarak antara objek dengan *centroid* dengan memakai rumus *Euclidian Distance*.

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

Rumus 2. 4 Jarak masing-masing *centroid*

4. Tiap-tiap objek dialokasikan ke dalam *centroid* yang paling terdekat. Untuk melaksanakan pendistribusian objek ke dalam tiap-tiap *cluster* pada waktu iterasi secara umum bisa dilaksanakan dengan dua cara dengan hard *k-means*, yang dilakukan setiap objek yang diverifikasi sebagai anggota cluster dengan greenukur lain dapat dilaksanakan dengan *fuzzy C-Means*.

5. Iterasi dilakukan, selanjutnya menentukan posisi *centroid* baru dengan memakai persamaan.
6. Bila posisi *centroid* baru tidak sama ulangi langkah ke3.

Matriks *group assignment* pada iterasi sebelumnya dengan matrik *group assignment* pada iterasi yang sedang berjalan dibandingkan dengan dilakukan pengecekan *konvergensi*. Bila hasilnya sama maka algoritma *k-means cluster analysis* sudah *konvergen*, tetapi bila tidak sama maka belum *konvergen* sehingga perlu dilaksanakan iterasi selanjutnya.

2.4 Software Pendukung

2.4.1 UML

Menurut Windu Gata (2013) didalam jurnal (Urva, 2015), Hasil pemodelan pada *OOAD* terdokumentasikan dalam bentuk *Unified Modeling Language (UML)*. *UML* bahasa spesifikasi standar yang digunakan untuk mengarsipkan, mengkhususkan dan membuat perangkat lunak. Pengembangan dalam sistem yang berorientasi objek dan juga ialah *tool* untuk menunjang ekspansi sistem disebut dengan UML.

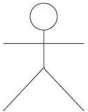





1. *Use case Diagram*

Use case diagram ialah pemodelan untuk kelakuan (*behavior*) sistem informasi yang akan dibuat. *Use case* mendefinisikan sebuah interaksi antara satu atau lebih aktor dengan sistem informasi yang akan dibuat. Digunakan untuk melihat fungsi apa saja yang ada di dalam sistem informasi

dan siapa saja yang berhak memakai fungsi-fungsi itu dapat disetujui.

Simbol-simbol yang dipakai dalam diagram *use case*, yaitu:

Tabel 2. 1 *Use case Diagram*




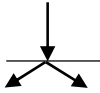



Gambar	Keterangan
	<p><i>Use case</i> yang ditunjukkan fungsionalitas yang diadakan sistem sebagai unit-unit yang bertukar pesan antar unit dengan aktor, biasanya diumumkan dengan memakai kata kerja di awal nama <i>use case</i>.</p>
	<p><i>Aktor</i> adalah <i>abstraction</i> dari orang atau sistem lain yang mengoperasikan fungsi dari sistem target. Untuk mengenali aktor, harus ditentukan pembagian tenaga kerja dan tugas-tugas yang terkait dengan peran pada konteks sistem target. Orang atau sistem dapat kelihatan dalam beberapa peran. Dibutuhkan sebagai aktor yang didukung dengan <i>use case</i>, tetapi tidak mempunyai kontrol terhadap <i>use case</i>.</p>
	<p>Asosiasi antara aktor dan <i>use case</i>, diilustrasikan dengan garis tanpa panah yang menandakan siapa atau apa yang meminta interaksi secara langsung dan bukannya menandakan aliran data.</p>
	<p>Asosiasi antara <i>actor</i> dan <i>use case</i> yang memakai panah terbuka untuk menandakan kala aktor berbagi dengan sistem.</p>
	<p><i>Included</i> di dalam <i>use case</i> lain atau pemanggilan <i>use case</i> oleh <i>use case</i> lainnya, contohnya pemanggilan sebuah fungsi program.</p>
	<p><i>Extend</i>, ialah ekspansi dari <i>use case</i> lain jika ketentuan terlaksana.</p>

Sumber: (Urva, 2015)

2. Diagram *Activity*

Activity Diagram menggambarkan *workflow* (aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis. Simbol-simbol yang digunakan dalam *activity diagram* yaitu:

Tabel 2. 2 Simbol *Diagram Activity*

Gambar	Keterangan
	<i>Start point</i> ditempatkan di pojok kiri atas dan adalah permulaan aktivitas.
	<i>End point</i> , aktivitas akhir.
	Aktivitas, mencerminkan suatu proses atau aktivitas bisnis.
	<i>Fork</i> (Percabangan), dipakai untuk memberitahu aktivitas yang dilaksanakan paralel atau untuk membagikan dua aktivitas paralel menjadi satu.
	<i>Join</i> (penggabungan) atau <i>rake</i> , dipakai untuk menggambarkan adanya dekomposisi.
	<i>Decision Points</i> , mencerminkan pilihan untuk pengambilan keputusan, <i>true</i> , <i>false</i> .
	<i>Swimlane</i> , pembagian <i>activity diagram</i> untuk membuktikan siapa dan melaksanakan apa.

Sumber: (Urva, 2015)

3. *Class Diagram*

Merupakan kaitan antar kelas dan penjelasan detail masing-masing kelas di dalam model desain dari suatu sistem, juga membahas patokan dan tanggung jawab entitas yang memastikan hubungan sistem. juga menunjukkan atribut-atribut dan operasi-operasi dari kelas dan batasan diagram kelas yang berkaitan dengan objek yang dihubungkan. Diagram kelas secara khas meliputi: Kelas (*Class*), Relasi, Asosiasi, Generalisasi dan Agregasi, Operasi Atribut (Operasi / Metode), Visibilitas, akses ke objek eksternal untuk setiap operasi atau atribut. Hubungan antar kelas memiliki informasi yang disebut dengan multiplisitas atau kardinaliti.

Tabel 2. 3 *Class Diagram*


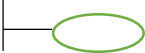




Multiplicity	Penjelasan
1	Satu dan hanya satu
0..*	Boleh tidak ada atau satu lebih
1..*	Satu atau lebih
0..1	Boleh tidak ada atau maksimal 1
n..n	Batasan antara. Contoh 2..4 mempunyai arti minimal 2 maksimal 4

Sumber: (Urva, 2015)

4. *Sequance Diagram*

Sequence diagram menggambarkan kelakuan objek pada *use case* dengan mendeskripsikan waktu hidup objek dan pesan yang diambil dan diterima antar objek. Simbol-simbol yang digunakan dalam diagram urutan yaitu:

Tabel 2. 4 *Sequence Diagram*

Gambar	Keterangan
	<i>EntityClass</i> , ialah serpihan dari sistem yang berisi himpunan kelas yang terdiri dari entitas-entitas yang membentuk sistem awal dan menjadi dasar untuk membentuk database.
	<i>Boundary Class</i> , beris ihimpunan yang menjadi <i>interface</i> atau interaksi antara satu atau lebih aktor dengan sistem, seperti tampilan <i>formentry</i> dan bentuk cetak.
	<i>Control class</i> , objek yang berisi logika aplikasi yang tidak bertanggung jawab kepada entitas, misalnya perhitungan dan aturan bisnis yang mengaitkan berbagai objek.
	<i>Message</i> , simbol mengirim pesan antar <i>class</i> .
	<i>Recursive</i> , Menyatakan pesan yang dikirim untuk aktivasi sendiri.
	<i>Activation</i> , aktivasi operasi dari objek, panjang kotak ini berbanding lurus dengan durasi aktivitas operasi.

	<i>Lifeline</i> , garis titik-titik yang terhubung dengan objek, sepanjang <i>lifeline</i> terdapat <i>activation</i>
--	---

Sumber: (Urva, 2015)

2.4.2 Rapidminer

Rapid Miner merupakan *software* yang dibuat oleh Dr. Markus Hofmann dari *Institute of Technology Blanchardstown* dan *Ralf Klinkenberg* dari *rapid-i.com* dengan tampilan *GUI (Graphical User Interface)* yang memungkinkan pengguna dalam memakai perangkat ini. Perangkat lunak ini bersifat *open source* dan dibuat dengan memanfaatkan program Java di bawah lisensi *GNU Public License* dan *Rapid Miner* dapat dijalankan di sistem operasi manapun. Dengan memakai *Rapid Miner*, tidak diperlukan kapabilitas koding khusus, karena semua fasilitas sudah disediakan. *Rapid Miner* dikhususkan untuk pengaplikasian *data mining*. Model yang ada juga cukup banyak dan lengkap, seperti Model *Bayesian*, *Pemodelan*, *Induksi Pohon*, *Jaringan Saraf Tiruan* dan lain-lain. Banyak metode yang disediakan oleh *Rapid Miner* mulai dari klasifikasi, klustering, asosiasi dan lain-lain. Jika tidak ada model atau model algoritma yang tidak ada di *Weka*, pemakai dapat menambahkan modul lain, karena *weka* hanya memakai sumber terbuka, sehingga dapat dipakai untuk mengembangkan *software* ini. (Haryati, 2015, p. 133)

Rapidminer ialah *software* untuk pengolahan data. Dengan memakai prinsip hari data algoritma *mining*, *RapidMiner* mengekstrak setengah dan setengah dari kumpulan data yang besar dengan memadukan metode statistik, basis data

kecerdasan buatan hari. menyediakan penggunaanya dalam melaksanakan perhitungan data yang sangat *RapidMiner* banyak memakai operator- operator. Operator ini berupaya untuk mengimplementasikan data. Data yang diambil dengan *node-node* jatuh operator kemudian kita hanya tinggal menghubungkannya beralih ke node hasil untuk hasil. Hasil yang diperlihatkan *Rapidminer* lengkap dapat dilihat secara visual hari grafik. Menjadikan *RapidMiner* adalah salah satu perangkat lunak pilihan untuk melaksanakan ekstrak data dengan metode-metode *data mining* (I, 2017, p. 60).

2.4.3 HTML

Konsep dasar HTML pemrograman yang lebih dikenal dengan sebutan *web scripting*. *Script* dikatakan karena perintah kode program tersebut akan diinterpretir dan tidak ada kompilasi untuk membuatnya dieksekusi. Berdasarkan *flyer* proses *interpreter poppy web scripting* dibagi menjadi dua kategori, yaitu yang sisi *klien* dan sisi *server*. Sisi *klien* dilakukan oleh *browser web* seperti *internet explorer, netscape, opera, firefox day*. Untuk contoh bahasa sisi *klien* adalah *HTML, CSS, Javascript, VBscript, XML day*. Sedangkan sisi *server* dilakukan oleh *server web* seperti *PWS (Personal Web Server)* untuk sistem operasi *windows 98*), *IIS* (untuk sistem operasi *windows 2000 / Windows XP*), *Apache, Tomcat, Xitami, hari ZOPE*. Untuk contoh bahasa sistiem server adalah *ASP (.Net), PHP, JSP, CFM, CGI / PL*.

Script web yang mendorong sisi *klien* akan menghasilkan halaman web yang statistik, format lebih fokus pada desain, informasi dan informasi yang

ditampilkan tidak dapat diperbarui seketika, karena tidak dapat melakukan permintaan interaktif dari pengguna dan proses tidak terjadi *output* apapun. Dapat dimengerti tidak mungkin untuk membuat aplikasi web yang dinamis dengan *scripting web* dan sisi klien ini, tetapi harus menggabungkan selatan dengan *web scripting* yang mengarah *serveride*. (Febio, 2011, p. 43)

2.4.4 PHP

PHP ialah skrip untuk skrip pemrograman sisi-*server web*, *skrip* yang membuat dokumen *HTML* secara langsung, dokumen maksimal *HTML* yang dihasilkan dari aplikasi bukan dokumen *HTML* yang dibuat memakai editor *tex* atau editor *HTML PHP / FI* ialah nama awal dari *PHP*. *PHP* adalah halaman rumah pribadi, *FI* adalah antarmuka bentuk. Pertama kali dibuat oleh *Rasmus Lerdoff*. *PHP*, mulai merupakan program yang dikhususkan untuk menerima input melalui formulir yang disetujui dalam *web browser*. Perangkat lunak ini disebarkan hari dilisensikan sebagai *software open source*. *PHP* secara resmi ialah kependekan dari *PHP Hypertext Preprocessor*, merupakan bahasa *script server-side* yang disisipkan pada *HTML*. Berikut adalah contoh yang umum digunakan untuk memaparkan tentang *PHP* sebagai *script* yang disisipkan dalam dokumen *HTML*:
`<html> <head> <title>Contoh</title> </head> <body> <?php echo "Tulisan ini dibuat dengan script PHP"; ?> </body> </html>`.

Kode/*script PHP* diapit dengan memakai tag awal dan tag akhir yang khusus (tag awal `<?php` atau `<?` dan tag akhir `?>`), yang mengharuskan

pemrogram untuk masuk dan keluar dari *mode script* PHP. Hal yang dapat dikerjakan *PHP* secara mendasar antara lain memperoleh data dari *form*, menghasilkan isi halaman *web* yang dinamik, dan menerima *cookies*, namun kapabilitas (*Feature*) *PHP* yang paling dipercayai dan bermakna adalah dukungan kepada banyak database. (Febio, 2011, pp. 43-44)

2.4.5 *MySQL*

Database secara sederhana, dapat kita sebut sebagai data gudang. Secara teori, basis data adalah himpunan data atau informasi yang kompleks, data-data tersebut disusun menjadi lebih dari satu golongan dengan tipe data yang disebut tabel / entitas), yang mana tiap-tiap datanya dapat saling berkaitan satu sama lain atau bisa dibuat sendiri, sehingga mudiriaks muders. *MySQL* merupakan basis data yang hanya menjalankan *crash Unix day Linux*.

Selama berjalannya waktu dan jumlah penggemar yang memakai database ini, *MySQL* merilis versi yang dapat *diinstal* jatuh semua *platform*, termasuk *Windows*. Hadiah *Lisens MySQL* adalah *freeware*. Kita dapat membeli dan membeli tanpa harus membayarnya. Meskipun kita menual produk yang disetujui perangkat lunak *MySQL*, kita tidak bisa menolak *hack*. Mungkin bagi kita yang baru dengan *MySQL* akan bingung dengan dua kata "*SQL*" hari "*MySQL*". Pertanyaan yang mungkin muncul adalah, apakah *SQL* yang sebenarnya, dan apa bedanya dengan *MySQL*? *SQL* merupakan kependekan kata kata "*Structured Query Language*". *SQL* ialah suatu permintaan bahasa yang terstruktur yang melekat pada satu database atau *SMBD* tertentu, sedangkan *MySQL* merupakan databasanya. Dengan kata lain, *MySQL* merupakan *SMBD*-nya dan *SQL* adalah

perintah atau bahasa yang melekat di dalam SMDB tersebut. Struktur dasar dari *express SQL* terdiri dari tiga klausa, yaitu: *select, from dan where*. (Febio, 2011, p. 44)

.2.5 Penelitian Terdahulu

Karya ilmiah yang terdahulu menyinggung pembahasan dan pembelajaran tentang tokoh-tokoh nasional yang mempergunakan media sebagai bahan pembelajaran selain buku, yang bisa menjadi pedoman atau patokan untuk pengembangan karya ilmiah sebelumnya. Ketersediaan penelitian terdahulu dilakukan sebelumnya, peneliti sangat terbantu karena sudah di paparkan dan implementasi *data mining* dengan metode *K-Means Clustering*.

1. (Mhd Gading Sadewo1, 2017),*ISSN* :2540-7600, "PENERAPAN *DATAMINING* PADA POPULASI DAGING AYAM RAS PEDAGING DI INDONESIA BERDASARKAN PROVINSI MENGGUNAKAN *K-MEANS CLUSTERING*" Karya ilmiah ini membicarakan tentang implementasi *data mining* pada ayam pedaging ras di Indonesia. Berlandaskan provinsi menggunakan *K- Means Clustering*. Pengumpulan data dikumpulkan melalui dokumen-dokumen penelitian tentang daging ayam yang diproduksi oleh Badan Pusat Statistik Nasional. Data yang dipakai dalam karya ilmiah ini adalah data dari tahun 2009-2016 yang terdiri dari 34 provinsi. Variabel yang digunakan (1) jumlah populasi dari tahun 2009-2016. Data diproses dengan melaksanakan pengelompokan

dalam 3 klaster yaitu klaster tingkat populasi tinggi, klaster tingkat populasi menengah dan rendah. Data *centroid* untuk *cluster* populasi tinggi 4711403141, Data *centroid* untuk *cluster* tingkat populasi 304240647, dan data *centroid* untuk *cluster* tingkat populasi rendah 554200. Diperoleh dari yang ada saat ini. sedang yaitu Sumatera Utara, Jawa Tengah, Jawa Timur, Banten, Kalimantan Selatan dan Kalimantan Timur, dan 27 provinsi lainnya termasuk tingkat partisipasi rendah. Keadaan ini dapat menjadi wejangan bagi pemerintah, provinsi yang menjadi perhatian pada populasi daging berlandaskan *cluster* yang telah dilakukan.

2. (Handoko, 2016). *ISSN 2476 – 8812*. “PENERAPAN *DATA MINING* DALAM MENINGKATKAN MUTU PEMBELAJARAN PADA INSTANSI PERGURUAN TINGGI MENGGUNAKAN *METODE K-MEANS CLUSTERING* (STUDI KASUS DI PROGRAM STUDI TKJ AKADEMI KOMUNITAS SOLOK SELATAN)” , Menerapkan *data mining* dengan memakai metode *clustering* untuk memperluas mutu pembelajaran pada instansi Perguruan Tinggi di program studi TKJ Akademik komunitas Solok Selatan berupa proses pengelompokan sejumlah data atau objek ke dalam *cluster* sehingga dalam setiap *cluster* tersebut akan berisi data yang semirip mungkin dan tidak sama dengan objek dalam *cluster* yang lainnya.
3. (Abadi, 2018), *ISSN 182-187* “*APPLICATION MODEL OF K-MENAS CLUSTERING: INSIGHT INTO PROMOTION STRATEGY OF VACATIONAL HIGH SCHOOL*” Analisis *data mining* dilakukan dengan

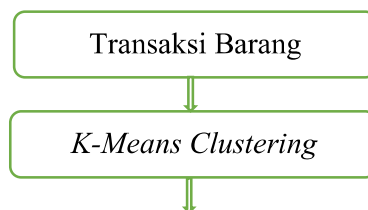
memakai metode *clustering K-Means*. Penggunaan metode ini, data yang diperoleh dapat dikelompokkan menjadi beberapa kelompok berdasarkan data, sehingga data yang mempunyai ciri khas yang sama dikelompokkan dalam satu *cluster* dan yang mempunyai ciri khas yang tidak sama dikelompokkan dalam kelompok yang berbeda yang memiliki ciri khas yang sama. Melalui pengelompokan data ini, diharapkan departemen pemasaran bisa melaksanakan pemasaran dengan kebijakan yang tepat untuk mendapatkan siswa baru.

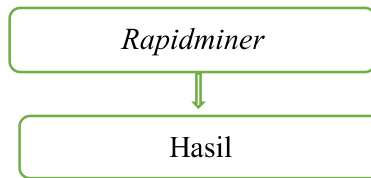
4. (Nurzahputra, 2017), *ISSN 17-24* “PENERAPAN ALGORITMA *K-MEANS* UNTUK *CLUSTERING* PENILAIAN DOSEN BERDASARKAN INDEKS KEPUASAN MAHASISWA.” Untuk melaksanakan penilaian kinerja dosen berdasarkan indeks kepuasan mahasiswa dapat mengaplikasikan metode *clustering K-Means*. Data didapat dari mahasiswa melalui angket sesuai aspek *Reliability, Responsiveness, Assurance*, dan *Empathy*. Data diproses untuk mendapatkan nilai dari setiap dosen. Data selanjutnya diproses memakai *Rapidminer* untuk dipatok nilai *centroid* dalam *cluster* baik dan *cluster* kurang dengan algoritma *K-Means*. *Centroid* data untuk *cluster* baik 17.099 dan *cluster* kurang 15.874. Sehingga didapat penilaian dosen berdasarkan indeks kepuasan mahasiswa dengan 5 dosen *cluster* baik dan 7 dosen *cluster* kurang.
5. (Metisen, ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA, 2015), *ISSN 1858 – 2680* “ANALISIS

CLUSTERING MENGGUNAKAN METODE *K-MEANS* DALAM PENGELOMPOKKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA. "Dalam implementasi ini, dipakai metode *clustering* dengan memakai algoritma *K-means*. Dari data yang diolah dengan sampel data yang diambil di Swalayan Fadhillah Bengkulu, maka diperoleh dua jenis kelompok data. yaitu data penjualan rendah dan data penjualan tinggi. Data ini disediakan oleh pihak yang berkepentingan Fadhillah dapat melihat jenis barang yang dijual dan tidak. menyimpan barang yang ada di gudang tidak membiak.

2.6 Kerangka Pemikiran

Kerangka berfikir yang baik akan memaparkan secara resmi tentang variabel antar yang akan diterbitkan. Oleh sebab itu, hendak dipaparkan relasi antara variabel independen dan dependen. Jika dalam karya ilmiah ada variabel moderator dan campur tangan, maka perlu juga jelaskan, Mengapa variabel harus ikut serta dalam karya ilmiah pertautan antar variabel tersebut, kemudian dirumuskan menjadi bentuk paradigma karya ilmiah. Oleh sebab itu pada setiap pertemuan paradigma karya ilmiah harus di dasarkan pada pertemuan berfikir. (Sugiyono, Metode Penelitian Kuantitatif dan Kualitatif dan R&D, 2014, p. 60).





Gambar 2. 4 Kerangka Pemikiran

Sumber: Data olahan Peneliti(2019)

Urutan kerangka berpikir diatas adalah:

1. Transaksi Barang : merupakan variabel peneliti yang akan diolah untuk penyelesaian permasalahan dengan cara memproses data yang sudah didapat peneliti dari narasumber atau tempat peniliti dengan cara wawancara maupun *observasi*.
2. *K-Means Clustering* : ialah metode *data mining* yang akan gunakan untuk mengolah data dari variabel yang sudah ditentukan.
3. *Rapidminer* : *tools* atau aplikasi yang akan digunakan peneliti untk penentuan stok barang yang peneliti harapkan dapat memberikan lebih baik karena aplikasi yang dihasilkan bisa menampilkan kategori produk dalam jangka waktu tertentu sesuai yang diinginkan.
4. Hasil : Hasil yang diperoleh dari pengolahan data dari ketiga variabelnya adalah produk yang paling digemari unttuk produk untuk jumlah stok banyak , produk untuk stok sedang dan produk kurang digemari untuk jumlah stok sedikit.

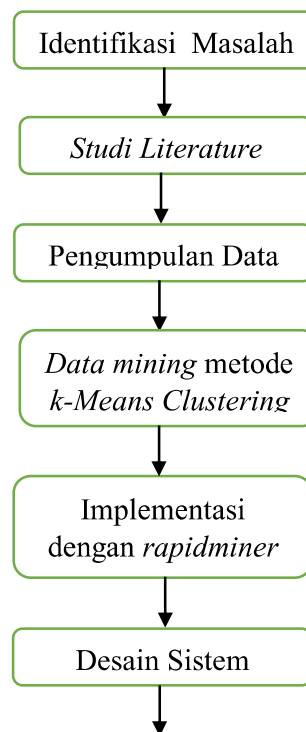
BAB III

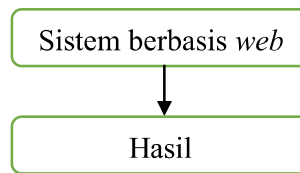
METODE PENELITIAN

3.1 Desain Penelitian

Menurut (Sugiyono, 2012, p. :2), untuk memperoleh data dengan tujuan dan kepentingan khusus diperlukan teknik ilmiah yang pada mulanya ialah metode karya ilmiah. Berlandaskan situasi tersebut terdapat empat kata kunci yang perlu di perhatikan yaitu, cara ilmiah, data, tujuan, dan kegunaan.

Pada karya ilmiah ini peneliti memakai desain penelitian dengan tahapan proses seperti yang terlihat pada gambar di bawah ini:





Gambar 3. 1 Desain Penelitian

Sumber: Data Penelitian (2019)

Mengenai uraian dari gambar 3.1 tahapan atau poses yang akan diselesaikan dalam karya ilmiah ini sebagai berikut :

1. Idenifikasi Masalah

Karya ilmiah ini dimulai dengan studi terlebih dahulu, yang bertujuan untuk memilah masalah yang terkait dengan topik penelitian, sehingga peneliti menerima apa yang benar-benar menjadi permasalahan yang akan pecahkan.

2. *Studi Literature*

Pada tahapan ini peneliti mempelajari buku, jurnal, internet dan pertanyaan yang berkaitan dengan subjek yang diteliti

3. Pengumpulan data

Melakukan wawancara dengan pihak pemilik (*owner*) *Online Shope* Habibie serta mengumpulkan data-data yang berhubungan dengan penelitian.

4. *Data mining* metode *K-means clustering*

Sesudah data-data sudah terkumpul, selanjutnya menganalisa data dengan cara *data mining k-means clustering* hingga akhirnya outputnya pengelompokan data

5. Implementasi dengan *Rapidminer*

Setelah data-data telah dianalisa yang di dpat hasilnya berupa pengelompokan produk, maka dilakukan implementasi dengan *rapidminer* untuk mencocokkan hasil pengelompokan barang yang sudah di hitung secara manual sama dengan pengujian sistem dengan *rapidminer*.

6. Desain sistem

Tahapan ini mendesain sistem penjualan, transaksi ditoko *Online Shope Habibie* dikarenakan selama ini masih melakukan perhitungan dan penjualan manual..

7. Sistem berbasis *web*

Setelah desain sistem sudah selesai maka implementasinya berupa sistem penjualan, transaksi dan *cluster* yang berbasis *web*.

8. Kesimpulan

Proses terakhir di karya ilmiah ini adalah merangkum hasil karya ilmiah yang berisi jawaban singkat terhadap rumusan masalah berlandaskan data yang tersedia. Pada saat penelitian ini, penelitian juga memberikan saran yang penting untuk memberikan masukan dalam penyelesaian permasalahan yang ada.

3.2 Teknik Pengumpulan Data

Teknik pengumpulan data ialah salah satu metode atau cara yang bisa di pergunakan dalam pengumpulan data. Pengumpulan data dalam karya ilmiah yang dimaksud yaitu dapat bahan, informasi, realitas dan informasi yang dapat dipercaya.

Jika dilihat dari segi cara atau teknik pengumpulan data, maka teknik pengumpulan data dapat dilakukan dengan wawancara (wawancara), angket (angket), observasi (observasi), dan ketiga kombinasi (Sugiyono, 2012, p. 137).

3.2.1 Wawancara

Menurut (Sugiyono, Metode Penelitian Kuantitatif Kualitatif dan R&D, 2012, p. 137), Wawancara dipakai sebagai metode pengumpulan data jika perlu melakukan *studi* pendahuluan untuk mendapatkan permasalahan yang perlu dipelajari, memahami masalah dari responden secara lebih rinci, dan jumlah responden kecil.

Peneliti sudah melakukan wawancara kepada pemilik Toko *Online Shope* Habibie untuk melaksanakan karya ilmiah serta pengumpulan data-data yang dibutuhkan dalam penelitian.

3.2.2 Dokumentasi

Adalah metode pengumpulan data dengan pengumpulan dengan mengumpulkan catatan-catatan atau dokumen-dokumen. Data yang diperoleh

berupa foto, *history* transaksi penjualan yang sudah dirangkum dalam bentuk data mentah sehingga menjadi sebuah dokumen yang ada dipakaidalam karya ilmiah.

3.2.3 Kajian Pustaka

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

3.3 Operasional variabel

Variabel operasional dibutuhkan untuk penentuan indikator dan jenis variabel yang terkait dengan karya ilmiah. Variabel operasional bertujuan untuk memastikan skala pengukuran tiap-tiap variabel, sehingga pengetesan dugaan sementara memakai *tool* bisa dilaksanakan dengan benar. Variabel operasional juga merupakan pemahaman variabel operasional, dalam praktiknya, secara signifikan dalam lingkup objek penelitian / objek yang diteliti. Berikut ini merupakan penjabaran dari operasional variabel yang peneliti akan olah.

Tabel 3. 1 Tabel Barang

No	Nama barang	Kode barang	Harga satuan(Rp)	stok awal barang
1	Gucci	P1	Rp. 80.000	50
2	Blender	P2	Rp500.000	150
3	Mixer	P3	Rp 400.000	100
4	Lemari Plastik	P4	Rp 250.000	125

5	Kursi	P5	Rp 50.000	100
6	Obat Herbal	P6	Rp 150.000	350
7	Keranjang Loundry	P7	Rp 50.000	400
8	Kasur	P8	Rp. 500.000	60
9	Dispenser	P9	Rp 300.00	320
10	Spray	P10	Rp 100.00	600
11	Selimut	P11	Rp 150.000	100
12	Speaker	P12	Rp 200.000	500
13	Mikropon	P13	Rp 160.000	80
14	Gantungan Kunci	P14	Rp 50.000	60
15	Handuk	P15	Rp 75.000	380
16	Pakaian dalam	P16	Rp 90.000	100
17	Topi LED	P17	Rp 60.000	600
18	Karpet	P18	Rp 600.000	40
19	Timbangan	P19	Rp 800.000	210
20	Piring	P20	Rp 85.000	160
21	Sendok	P21	Rp 30.000	100
22	Prasmanan	P22	Rp 200.000	450
23	Sapu	P23	Rp 30.000	15
24	Ember	P24	Rp 25.000	25
25	Box Continer	P25	Rp 300.000	50

26	Rak tas	P26	Rp 60.000	60
27	Gelas	P27	Rp 45.000	100
28	Teko	P28	Rp 65.000	35
29	Kacamata	P29	Rp 50.000	550
30	Boneka	P30	Rp 300.00	450

Sumber: Data penelitian (2019)

Uraian tabel diatas merupakan tabel barang yang yang didapatkan peneliti melalui teknik pengumpulan data di toko *Online shope* HABIBIE yang dijelaskan dengan nama barang, kode barang, harga satuan(Rp), stok awal barang selama satu bulan. Operasional yang dipakai dalam karya imiah ini adalah data hitoris diambil dari data- data pada bulan Januari 2019- Mei 2019. Akan tetapi peneliti melakukan pengambilan sample secara acak yaitu dibulan Mei 201

Tabel 3. 2 Tabel Pembeli

No	KodePembelian	Nama Pembeli	Alamat
1	PO1	Junedi	Sagulung Batam
2	PO2	Wandy	Nongsa batam
3	PO3	Frengky	Marina
4	PO4	Josua	Nusa batam

Sumber: Data Penelitian (2019)

Uraian tabel diatas merupakan tabel pembeli yang diakan digunakan peneliti dalam perancangan dan pembuatan sistem karena di toko *online Shope* Habibie

selama ini masih melaksanakan penentuan stok barang secara manual sehingga kurang saksama.

Tabel 3. 3 Tabel Transaksi

Kode Barang	Kode Pembeli	Jml	Harga satuan	Total Bayar	Stok awal Barang	Sisa Stok barang
P1	PO1	5	Rp 80.000	Rp 400.000	50	45
P15	PO2	3	Rp 75.000	Rp 150.000	400	397
P24	PO3	20	Rp 25.000	Rp 250.000	25	5
P8	PO4	10	Rp 500.000	Rp 5.000.000	80	60

Sumber: Data Penelitian (2019)

Uraian tabel diatas merupakan tabel transaksi pembelian barang yang akan dipakai penelaah dalam perancangan dan pembuatan sistem yang berkaitan dengan toko *Online Habibie* masih melakukan pemilihan persediaan barang secara manual.

Tabel 3. 4 Operasional Variabel

Kode Produk	Jumlah transaksi	Stok Awal Barang	Volume Penjualan
P1	8	50	37
P2	25	150	120
P3	20	100	85
P4	20	125	120
P5	8	100	50
P6	24	350	300
P7	18	400	300
P8	12	60	50
P9	35	320	300
P10	28	600	500
P11	8	100	50

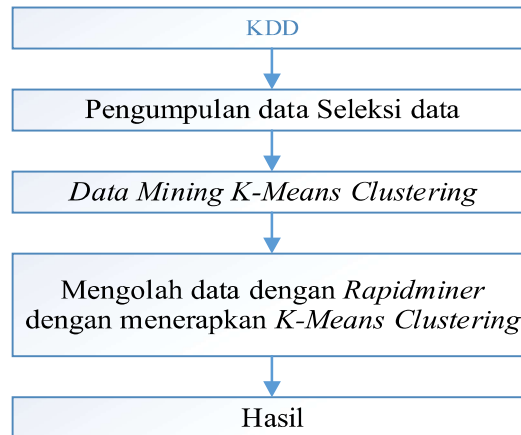
P12	30	500	476
P13	12	80	60
P14	12	60	50
P15	22	380	300
P16	14	100	85
P17	28	600	450
P18	4	40	30
P19	10	210	200
P20	13	160	150
P21	19	100	100
P22	24	450	370
P23	2	15	10
P24	7	25	20
P25	9	50	50
P26	15	60	50
P27	16	100	100
P28	10	35	30
P29	24	550	500
P30	28	450	420

Sumber: Data penelitian (2019)

Dari data yang sudah diperoleh dilakukan proses transformasi dengan cara mengubah kode produk yang terjual sebagai atribut lama dengan kode produk baru. Atribut yang dipakai dalam karya ilmiah ini adalah kode produk, jumlah transaksi, stok awal barang, volume penjualan. Untuk transaksi penjualan ini adalah seberapa banyak orang melakukan pembelian yang dapatkan dipenelitian ini dari toko *Online Shope* Habibie pada bulan Mei 2019 melalui teknik pengumpulan data. Sedangkan untuk stok awal tersebut adalah seberapa banyak barang atau produk yang distok didapatkan melalui teknik pengumpulan data. Volume penjualan merupakan seberapa banyak barang atau produk yang terjual selama Mei 2019 didapatkan melalui teknik pengumpulan data.

3.4 Metode Analisis dan Rancangan Sistem

3.4.1 Perancangan Data Mining



Gambar 3. 2 Rancangan *Data Mining*

Sumber: Olahan data peneliti (2019)

Pada uraian diatas penjelasan dari metode analisis dan rancangan adalah yaitu:

1. *KDD Goals*

Melalui observasi, *Online Shope* Habibie selama ini melakukan penentuan stok barang secara manual sehingga kurang akurat sehingga menimbulkan biaya simpan yang tinggi. Penentuan kategori stok yang tidak akurat menjadikan biaya penyimpanan tinggi, tidak efektif dan sering mengecewakan konsumen, sehingga dibutuhkan metode *K-Means* yang bisa meng-*cluster* produk menjadi kelompok jumlah stok banyak, sedang dan sedikit/kurang.

2. Pengumpulan dan Seleksi Data

Peneliti mengumpulkan data berupa penyediaan stok barang atau produk, transaksi penjualan dan volume penjualan berupa data historis penjualan di bulan Mei 2019

3. *Data mining dengan K-Means Clustering*

Data mining yang dipakai dalam karya ilmiah ini adalah *clustering* untuk pengelompokan produk untuk menentukan stok produk dengan jumlah besar untuk produk paling diminati, jumlah sedang untuk produk diminati jumlah sedikit untuk produk kurang diminati Tahapan *data mining* memilih algoritma *K-Means* yang mengcluster produk menjadi 3 kelompok yaitu grup pertama untuk stok banyak, kedua produk untuk sedang, dan kelompok tiga untuk produk sedikit.

4. Mengolah data dengan *Rapidminer* dengan menerapkan *K-Means Clustering*

Data yang sudah diolah oleh *K-Means Clustering* dimasukkan kedalam software *rapidminer* untuk membuktikan kesamaan hasil yang diolah secara manual dengan *software rapidminer*

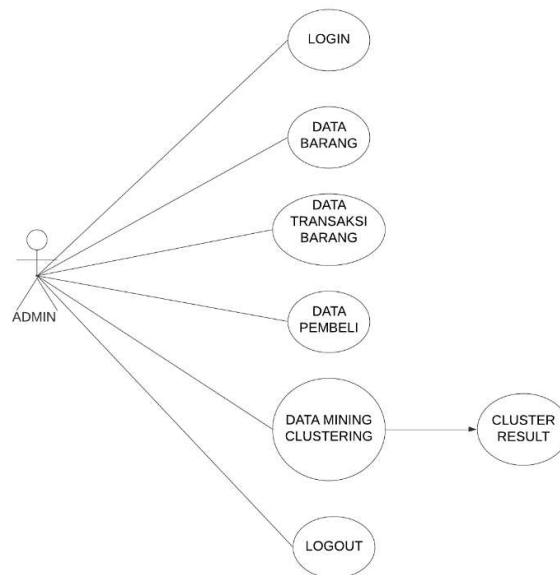
5. Hasil

Pada tahap ini diperoleh hasil, penerapan metode *K-Means clustering* produk dalam penentuan stok produk yang beragam dan banyak jenisnya untuk produk paling diminati, jumlah sedang dan jumlah sedikit untuk produk kurang diminat.

3.4.2 Perancangan Sistem dan Database

3.4.2.1 Use Case Diagram

Use case dibawah ini memaparkan satu aktor yang sedang menggunakan sistem yaitu *admin* sehingga pengerjaan *use case* lebih dikhususkan pada fungsionalitas yang ada pada system, melalui alur dan urutan. Didalam sistem ini terdapat 7 interaksi, diantaranya “*login*”, “*data barang*”, “*Data Transaksi Barang*”, “*Data Pembeli*”, “*Data mining Clustering*”, *cluster result* dan *logout*. Penjelasan *use case system* ini dijelaskan pada gambar dibawah ini:



Gambar 3.3 Use Case diagram sistem

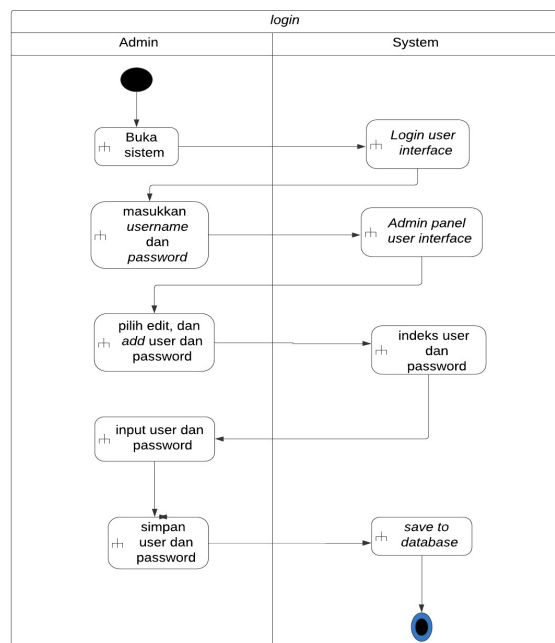
Sumber:Data Penelitian (2019)

3.4.2.2 Activity Diagram

Activity diagram mencerminkan kumpulan gerakan dari aktivitas, dipakai untuk memaparkan aktivitas yang dibentuk dalam satu operasi sehingga dapat juga untuk aktivitas lainnya. Dibawah ini adalah penjelasan dari *activity* diagram pada sistem ini.

1. *Login*

Activity diagram untuk *login, admin* dimulai dengan pemegang akses penuh terhadap sistem, *admin* melakukan *login*, memasukkan *user* dan *password*, *admin panel user interface*, pilih edit dan *add user login, indeks login, input user* dan *password*, simpan *user* dan *password*. Berikut adalah *activity* diagramnya:

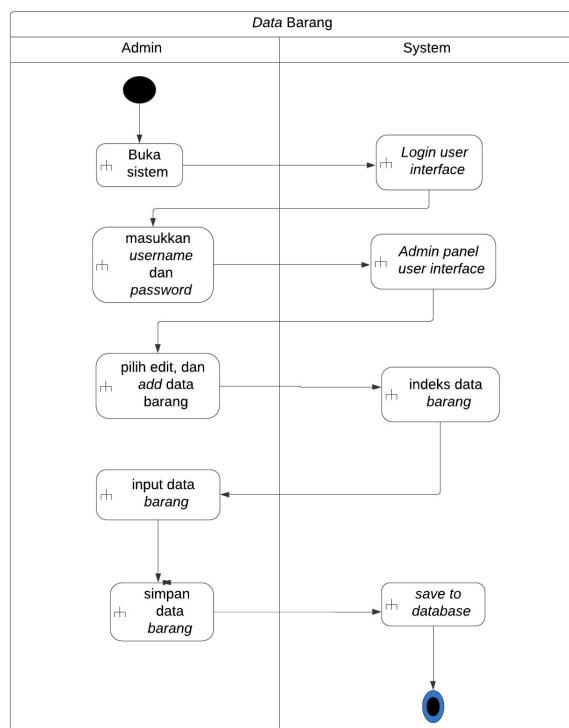


Gambar 3. 4 *Activity diagram Login*

Sumber: Data penelitian (2019)

2. Data Barang

Activity diagram untuk data barang, *admin* dimulai dengan *admin* sebagai pemegang akses penuh terhadap sistem. *Admin* buka sistem, melakukan *login*, memasukkan *user*, *password*, *admin panel user interface*, pilih edit dan *add* data barang, *indeks* data barang, *input* data barang, simpan data barang dan *save to database*. Berikut merupakan gambarannya:

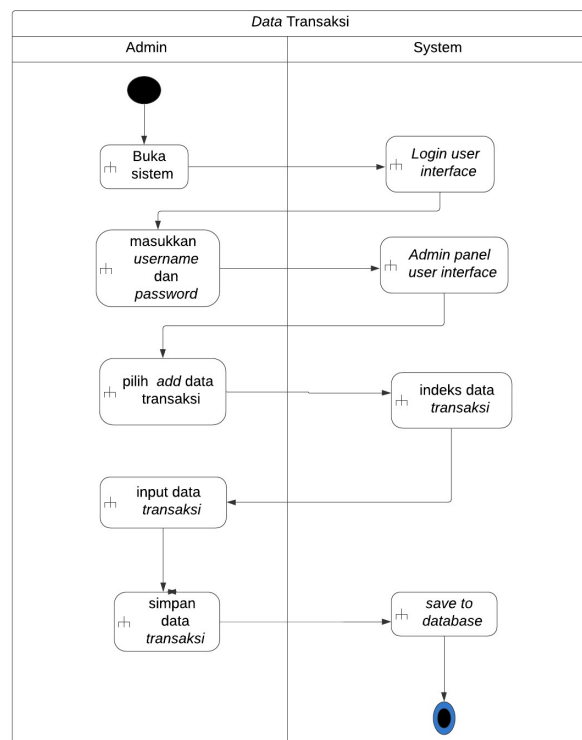


Gambar 3. 5 *Activity diagram* data barang

Sumber: Data Penelitian (2019)

3. Data Transaksi

Activity diagram untuk data transaksi, *admin* dimulai sebagai pemegang akses penuh terhadap sistem. *Admin* buka sistem, melakukan *login*, memasukkan *password user*, *admin panel user interfase*, pilih edit data transaksi, *indeks data transaksi*, *input data transaksi*, simpan data transaksi dan *save to database*. Berikut merupakan gambarannya

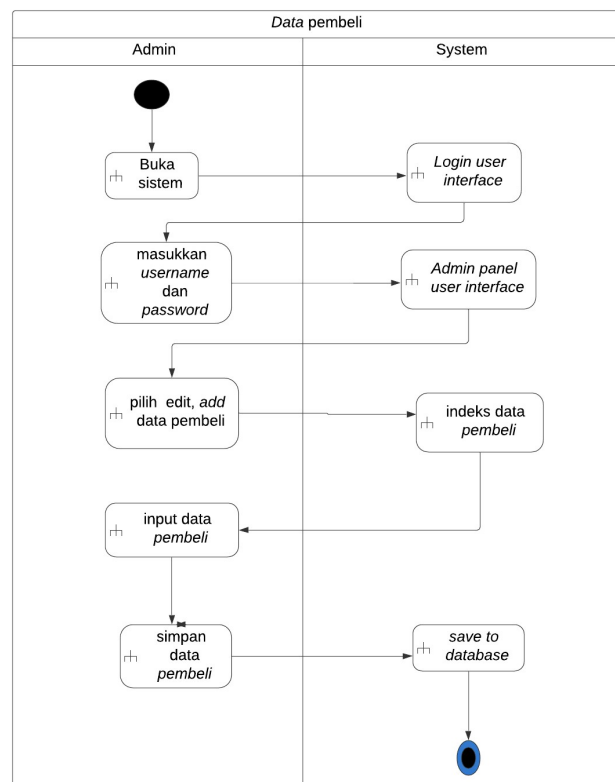


Gambar 3. 6 *Activity diagram* data transaksi

Sumber: Data Penelitian (2019)

4. Data Pembeli

Activity diagram untuk data pembeli, *admin* dimulai sebagai pemegang akses penuh terhadap sistem. *Admin* buka sistem, melakukan *login*, memasukkan *password user*, *admin panel user interface*, pilih edit, *add* data pembeli, *indeks* data pembeli, *input* data pembeli, simpan data pembeli, dan *save to database*. Berikut merupakan gambaran activitynya

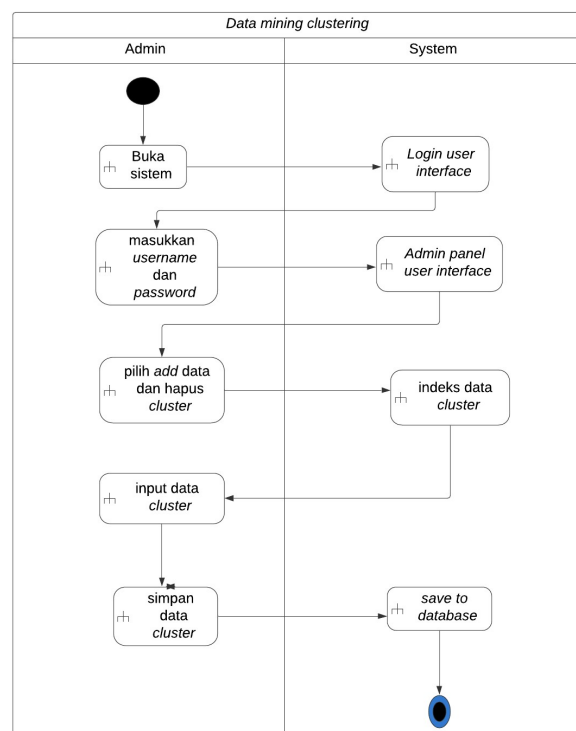


Gambar 3. 7 *Activity diagram* Data Pembeli

Sumber: Data Penelitian (2019)

5. *Data mining clustering*

Activity diagram untuk *data mining clustering*, *admin* dimulai dengan *admin* sebagai pemegang akses penuh terhadap sistem. *Admin* buka sistem, melakukan *login*, memasukkan *password user*, *admin panel user interface*, pilih *add* dan hapus *cluster* barang, *indeks data cluster*, *input data cluster*, simpan data *cluster*, dan *save to database*. Berikut merupakan gambaran activitynya

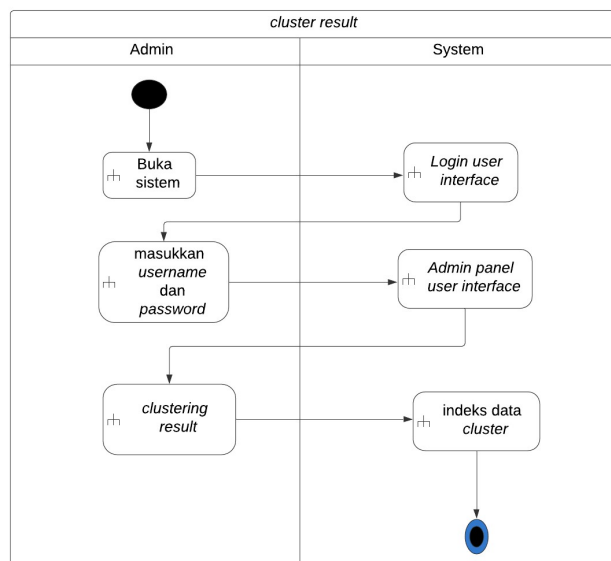


Gambar 3. 8 *Activity diagram data mining Clustering*

Sumber : Data Peneitian (2019)

6. Clustering Result

Activity diagram untuk *data mining clustering*, admin dimulai dengan *admin* sebagai pemegang akses penuh terhadap sistem. *Admin* buka sistem, melakukan *login*, memasukkan *password user*, *admin panel user interfase*, *clustering result*, *indeks data cluster*. Berikut merupakan gambaran activitynya

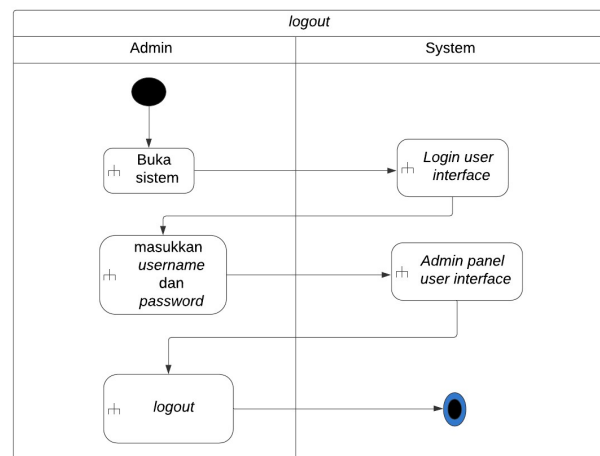


Gambar 3. 9 Activity diagram clustering result

Sumber: Data Penelitian (2019)

7. Logout

Activity diagram untuk *logout*, admin dimulai dengan pemegang akses penuh terhadap sistem, *Admin* buka sistem, melakukan *login*, memasukkan *password user*, *admin panel user interfase*, *logout*. Berikut merupakan gambaran activitynya.



Gambar 3. 10 Activity diagram logout

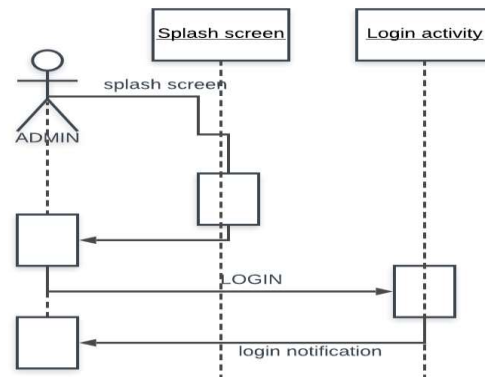
Sumber: Data penelitian (2019)

3.4.2.3 Sequence Diagram

Mencerminkan korelasi antara berpartisipasi objek dalam urutan waktu. Fungsinya untuk memperlihatkan deretan pesan yang dikirim antara objek juga korelasi antar objek yang terjadi pada titik tertentu dalam eksekusi sistem. Berikut penjelasan dari diagram urutan adalah

1. Login

Admin melakukan *login* terhadap sistem. Berikut penjelasan gambar *diagram sequence* dari *login* pada sistem

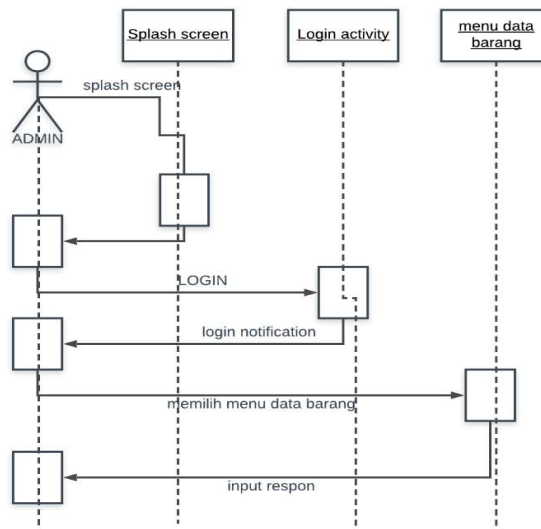


Gambar 3. 11 *Sequence diagram Login*

Sumber: Data Penelitian (2019)

2. Data Barang

Untuk dapat memilih menu data barang, admin harus memilih menu *login*, menu data barang. Berikut penjelasan gambar dari data barang pada sistem:

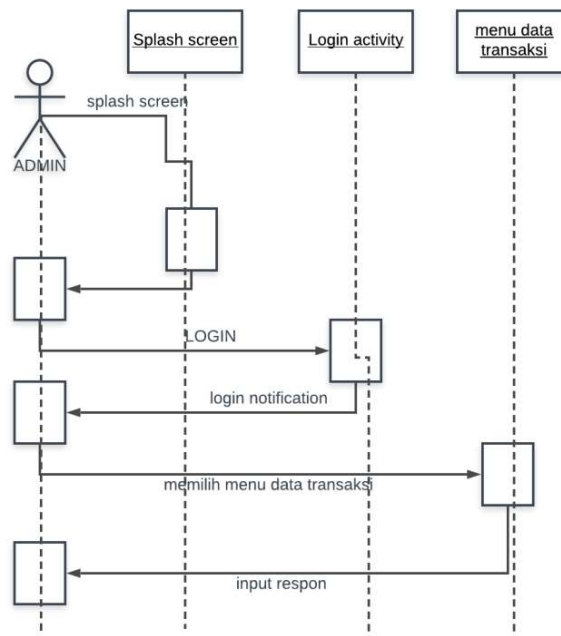


Gambar 3. 12 *Diagram sequence data barang*

Sumber: Data peneliian (2019)

3. Data Transaksi

Untuk dapat memilih menu edit data transaksi, admin harus memilih menu *login*, menu data transaksi. Berikut penjelasan gambar dari data transaksi pada sistem:

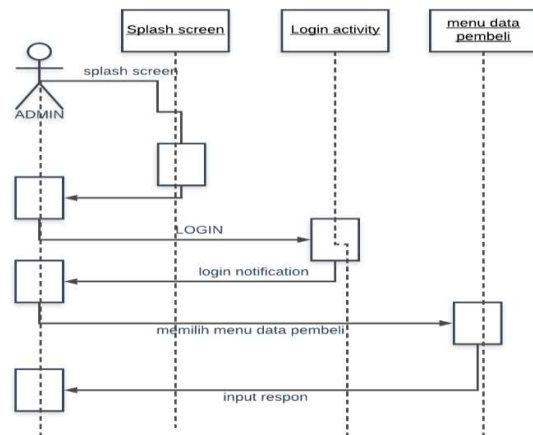


Gambar 3. 13 Diagram sequence Data Transaksi

Sumber: Data Penelitian (2019)

4. Data Pembeli

Untuk dapat memilih menu data pembeli, *admin* harus memilih menu *login*, menu data pembeli. Berikut penjelasan gambar dari data pembeli pada sistem:

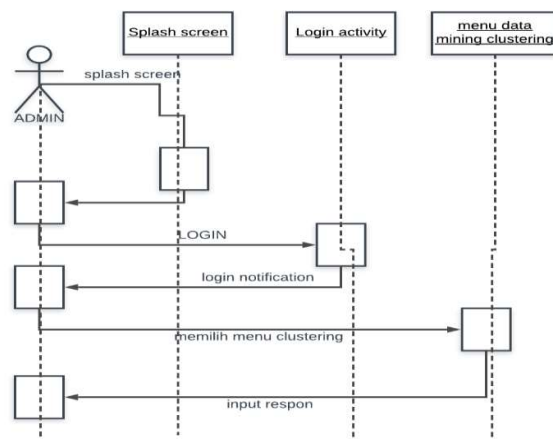


Gambar 3. 14 *Diagram sequence data pembeli*

Sumber: Data penelitian (2019)

5. *Data mining clustering*

Untuk dapat memilih menu *check cluster*, *admin* harus memilih *login*, menu *data mining clustering*. Berikut penjelasan gambar dari *data mining clustering* pada sistem:

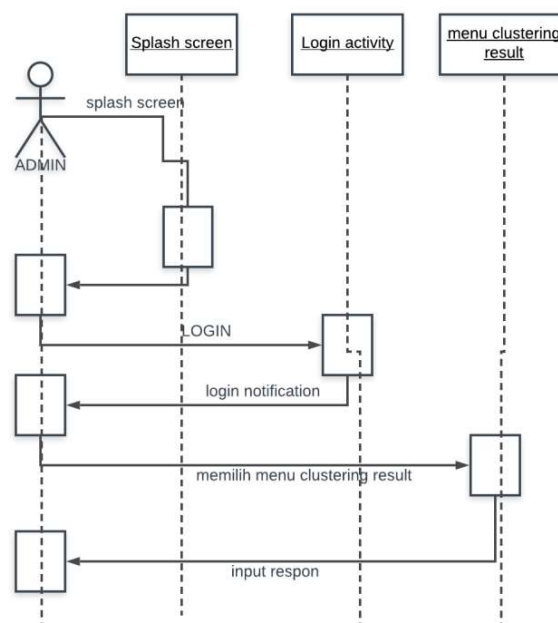


Gambar 3. 15 *Diagram sequence data mining clustering*

Sumber: Data penelitian (2019)

6. Clustering Result

Untuk dapat memilih menu *clustering result*, *admin* harus memilih *login*, menu *clustering resut*. Berikut penjelasan gambar dari *data mining clustering* pada sistem:

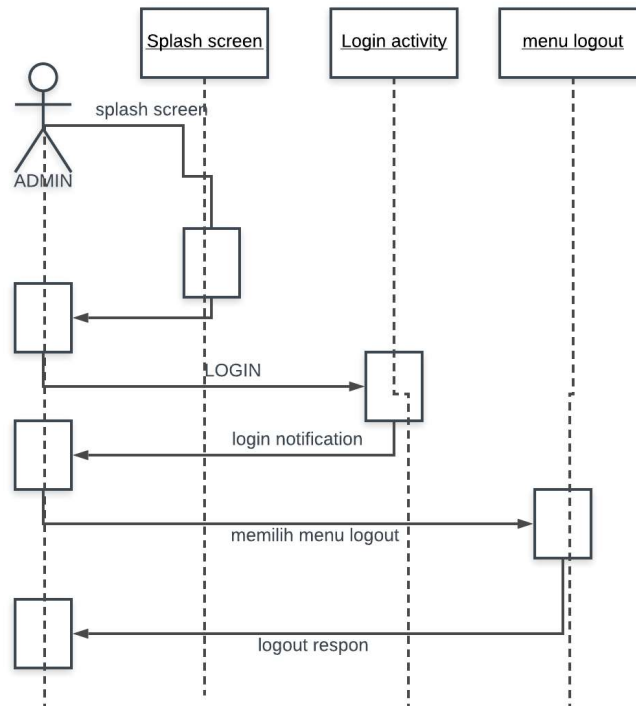


Gambar 3. 16 Diagram sequence Clustering result

Sumber: Data penelitian (2019)

7. Logout

Untuk dapat memilih menu *logout*, *admin* harus memilih *login*, menu *data logout*. Berikut penjelasan gambar dari *data mining clustering* pada sistem:

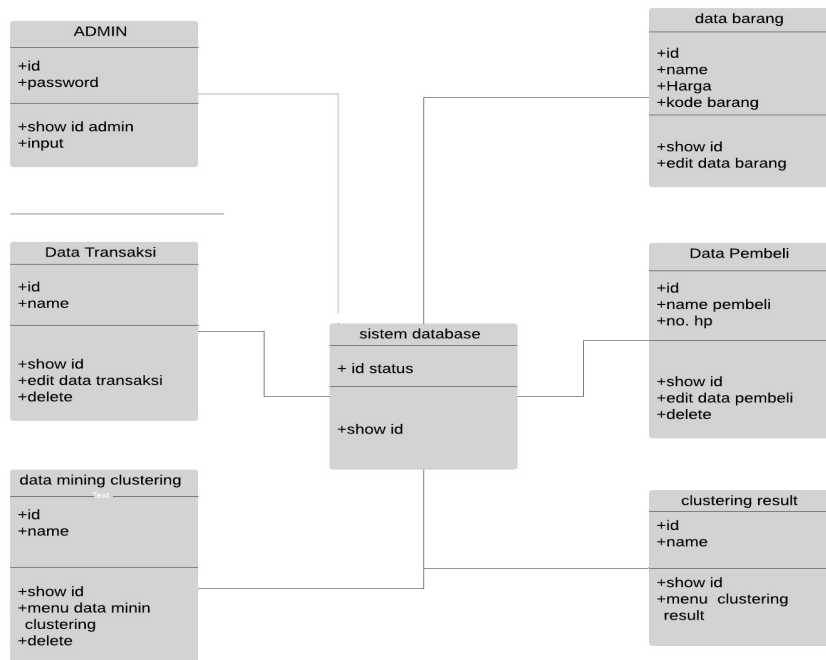


Gambar 3. 17 Activity diagram Logout

Sumber: Data penelitian (2019)

3.4.2.4 Class Diagram

Class diagram menjelaskan jenis– jenis objek dalam sistem dan berbagai macam kaitan statis yang terjadi. Pada sistem ini terdapat kelas utama, yaitu kelas admin, , kelas data barang, dan kelas transaksi, kelas data pembeli dan kelas *data mining clustering* dan *clustering result*. *Class diagram* pada aplikasi ini akan dijelaskan pada gambar dibawah ini.



Gambar 3. 18 *Class Diagram* sistem

Sumber: Data Penelitian (2019)

3.5 Lokasi dan Jadwal Penelitian

3.5.1 Lokasi Penelitian

Lokasi Penelitian yang dilakukan di Toko *Online Shope* Habibie yang beralamat di Ruko Buana Mas 2 No. 29 Kecamatan Sagulung, Kota Batam Kepulauan Riau.

3.5.2 Jadwal Penelitian

Penelitian ini dilakukan mulai dari Maret 2019 sampai dengan September 2019 yang dapat dilihat pada jadwal penelitian dibawah ini:

Tabel 3. 5 Jadwal Penelitian

No	Kegiatan	Mar-19				Apr-19				Mei-19				Jun-19				Jul-19			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Input Judul	■																			
2	Pengumpulan Data		■	■																	
3	BAB 1				■	■															
4	BAB 2						■	■													
5	BAB 3								■	■	■	■									
6	BAB 4											■	■	■	■	■					
7	BAB 5																■	■			
8	Pengumpulan Skripsi																			■	

Sumber: Data penelitian (2019)