

**IMPLEMENTASI *DATA MINING* MENGGUNAKAN  
ALGORITMA *APRIORI* PADA PENJUALAN SUKU  
CADANG MOTOR**

**SKRIPSI**



**Oleh:  
Ainul Mardiah  
150210272**

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

**IMPLEMENTASI *DATA MINING* MENGGUNAKAN  
ALGORITMA *APRIORI* PADA PENJUALAN SUKU  
CADANG MOTOR**

**SKRIPSI**

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



**Oleh:  
Ainul Mardiah  
150210272**

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

## PERNYATAAN

Dengan ini saya menyatakan bahwa :

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana, dan/atau magister), baik di Universitas Putera Batam maupun di perguruan tinggi lain.
2. Skripsi ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan pembimbing.
3. Dalam skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian 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 diperguruan tinggi.

Batam, 08 Agustus 2019

Yang membuat pernyataan,

Ainul Mardiah

150210272

**IMPLEMENTASI *DATA MINING* MENGGUNAKAN  
ALGORITMA *APRIORI* PADA PENJUALAN SUKU  
CADANG MOTOR**

**Oleh:  
Ainul Mardiah  
150210272**

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

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

**Batam, 8 Agustus 2019**

**Yulia, S.Kom., M.Kom.  
Pembimbing**

## ABSTRAK

Penelitian ini dilaksanakan untuk mempermudah ataupun membantu para pemilik bengkel Candra Motor dalam mengelola data-data maupun arsip-arsip penjualan suku cadang motor dengan menerapkan *data mining* metode algoritma *apriori*. *Data mining* merupakan suatu operasi yang menggunakan teknik atau metode tertentu untuk mencari pola atau bentuk yang berbeda dalam sebuah data yang terpilih. Data-data penjualan selama setahun dengan jumlah 15 *item* diseleksi menggunakan metode algoritma *apriori*. Algoritma *apriori* adalah algoritma pengambilan data dengan aturan asosiatif (*association rule*) untuk menentukan hubungan asosiatif suatu kombinasi item. Pada algoritma *apriori* ditentukan *frequent itemset-1*, *frequent itemset-2*, dan *frequent itemset-3* sehingga diperoleh aturan asosiasi dari data-data yang sudah diseleksi sebelumnya. Untuk mendapatkan *frequent itemset* tersebut maka setiap data yang sudah diseleksi harus memenuhi syarat minimum *support* dan minimum *confidence*. Pada penelitian ini menggunakan minimum *support*  $\geq 7$  atau 0,583 dan minimum *confidence* 90%. Sehingga diperoleh beberapa aturan asosiasi, dimana perhitungan pencarian aturan asosiasi secara manual dan menggunakan *software* WEKA 3.9 diperoleh hasil yang sama. Dengan memenuhi syarat minimum *support* dan minimum *confidence*, maka suku cadang yang banyak terjual adalah ban dalam, oli Yamaha, dan oli MPX.

**Kata Kunci:** *Data Mining, Penjualan Suku Cadang Motor, Algoritma Apriori*

## **ABSTRACT**

*This research was carried out to simplify or assist Candra Motor workshop owners in managing data and archives of motorcycle parts sales by applying a data mining a priori algorithm method. Data mining is an operation that uses a particular technique or method to look for different patterns or shapes in a selected data. Sales data for a year with the number of 15 items selected using the a priori algorithm method. A priori algorithm is an algorithm for taking data with associative rules (association rule) to determine the associative relationship of an item combination. In a priori algorithm, it is determined frequent itemset-1, frequent itemset-2, and frequent itemset-3 so that the association rules can be obtained from previously selected data. To obtain the frequent itemset, each selected data must meet the minimum support and minimum confidence requirements. In this study using minimum support  $\geq 7$  or 0.583 and minimum confidence of 90%. So that some rules of association were obtained, where the calculation of the search for association rules manually and using WEKA software obtained the same results. By fulfilling the minimum support and minimum confidence requirements, the most sold spare parts are inner tube, Yamaha oil and MPX oil.*

*Keywords: Data Mining, Motor Parts Sales, Apriori Algorithms*

## **KATA PENGANTAR**

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

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

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

1. Rektor Universitas Putera Batam.
2. Ketua Program Studi Teknik Informatika Universitas Putera Batam Bapak Andi Maslan, S.T.,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. Kepada Orang Tua penulis Bapak Hanafi dan Ibu Rohaya yang selalu mendoakan, dan mendukung baik segi moral maupunmateril kepada penulis sehingga dapat menyelesaikan penulisan skripsi ini.
7. Kepada sahabat terbaik penulis Nike Winda, Ferry Dwi, Frandy, Fikardi dan Zabur yang telahmenemani dan selalu memberikan semangat kepada penulis dalam halmemberikan ide masukan.
8. Teman seperjuangan Program Studi Teknik Informatika Universitas Putera Batam.

Semoga Allah SWTmembalas kebaikan dan selalu mencurahkan hidayah serta taufikNya, Aamiin.

Batam,08 Agustus 2019

Ainul Mardiah

## DAFTAR ISI

HALAMAN SAMBUNG DEPAN	
HALAMAN JUDUL.....	ii
HALAMAN PERNYATAAN .....	iii
HALAMAN PENGESAHAN.....	iv
ABSTRAK .....	v
<i>ABSTRACT</i> .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xi
DAFTAR RUMUS .....	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Penelitian .....	1
1.2 Identifikasi Masalah .....	3
1.3 Pembatasan Masalah .....	3
1.4 Perumusan Masalah .....	4
1.5 Tujuan Penelitian .....	4
1.6 Manfaat Penelitian .....	5
1.6.1 Manfaat penelitian secara teoritis .....	5
1.6.2 Manfaat praktis.....	5
BAB II TINJAUAN PUSTAKA	
2.1 <i>Knowledge Discovery in Database (KDD)</i> .....	6
2.2 <i>Data Mining</i> .....	9
2.2.1 Manfaat <i>Data Mining</i> .....	11
2.2.2 Fungsi <i>Data Mining</i> .....	12
2.2.3 Pengelompokan <i>Data Mining</i> .....	13
2.2.4 Teknik Pembelajaran <i>Data Mining</i> .....	16
2.2.5 Penerapan <i>Data Mining</i> .....	17
2.3 Metode <i>Data Mining</i> .....	20
2.3.1 Klasifikasi <i>Bayes</i> .....	20
2.3.2 Metode <i>Decision Tree</i> .....	22
2.3.3 Algoritma <i>Nearest Neighbor</i> .....	23
2.3.4 <i>Fuzzy C-Means</i> .....	23
2.3.5 <i>Backpropagation</i> .....	26
2.3.6 Algoritma <i>Apriori</i> .....	29
2.4 <i>Software</i> Pendukung .....	30
2.5 Penelitian Terdahulu .....	32

2.6 Kerangka Pemikiran.....	35
2.7 Hipotesis.....	36

### BAB III METODE PENELITIAN

3.1 Desain Penelitian (Arsitektur Sistem).....	37
3.2 Teknik Pengumpulan Data.....	39
3.2.1 Pengamatan ( <i>Observasi</i> ).....	39
3.2.2 Wawancara ( <i>Interview</i> ).....	39
3.2.3 Kajian Pustaka.....	40
3.3 Operasional Variabel.....	40
3.3.1 Jumlah Transaksi.....	41
3.3.2 Nama Suku Cadang Motor.....	42
3.4 Metode Analisis dan Perancangan Sistem.....	42
3.4.1 Metode Analisis.....	42
3.4.2 Perancangan Sistem.....	43
3.5 Lokasi dan Jadwal Penelitian.....	44
3.5.1 Lokasi Penelitian.....	44
3.5.2 Jadwal Penelitian.....	45

### BAB IV HASIL DAN PEMBAHASAN

4.1 Analisa Data.....	46
4.2 Hasil Pengujian.....	51
4.2.1 Pengujian Secara Manual.....	51
4.2.2 Evaluasi (Tahap Pengujian).....	79
4.2.3 Pengujian Menggunakan WEKA.....	82

### BAB V SIMPULAN DAN SARAN

5.1 Simpulan.....	87
5.2 Saran.....	87

DAFTAR PUSTAKA  
 DAFTAR RIWAYAT HIDUP  
 SURAT KETERANGAN PENELITIAN  
 LAMPIRAN

## DAFTAR TABEL

	Halaman
<b>Tabel 3.1</b> Data Transaksi Suku Cadang Motor di Bengkel CandraMotor .....	41
<b>Tabel 3.2</b> Jadwal Penelitian .....	45
<b>Tabel 4.1</b> Daftar <i>Item</i> suku cadang motor .....	46
<b>Tabel 4.2</b> Data Set Awal Penjualan Suku Cadang Motor .....	47
<b>Tabel 4.3</b> Kode Transaksi.....	48
<b>Tabel 4.4</b> Hasil integrasi data transaksi (2018-2019).....	50
<b>Tabel 4.5</b> <i>Frequent 1-itemset</i> .....	52
<b>Tabel 4.6</b> <i>Frequent 2- Itemset</i> .....	57
<b>Tabel 4.7</b> Aturan Asosiasi F2 : 1 <i>antecedent</i> .....	67
<b>Tabel 4.8</b> <i>Frequent 3-Itemset</i> .....	71
<b>Tabel 4.9</b> Calon Aturan Asosiasi F3: 2 <i>antecedent</i> .....	74
<b>Tabel 4.10</b> Aturan Asosiasi Final .....	76
<b>Tabel 4.11</b> Hasil Evaluasi.....	81

## DAFTAR GAMBAR

	Halaman
<b>Gambar 2.1</b> Tahapan Knowledge Discovery in Database (KDD) .....	7
<b>Gambar 2.2</b> Kerangka Pemikiran .....	35
<b>Gambar 3.1</b> Desain Penelitian .....	37
<b>Gambar 3.2</b> Perancangan Sistem .....	43
<b>Gambar 3.3</b> Denah Lokasi .....	44
<b>Gambar 4.1</b> Pengolahan Data Di MS.EXCEL .....	82
<b>Gambar 4.2</b> <i>Type</i> Penyimpanan Data Di MS.EXCEL.....	82
<b>Gambar 4.3</b> Langkah Awal Pada Tampilan WEKA.....	83
<b>Gambar 4.4</b> Tahap Ke-2 Pengolahan Data Di WEKA .....	83
<b>Gambar 4.5</b> Pemilihan Data Yang Akan Diolah Pada WEKA.....	84
<b>Gambar 4.6</b> Membuka Data Yang akan Diolah .....	84
<b>Gambar 4.7</b> Memilih Metode Apriori .....	85
<b>Gambar 4.8</b> Memulai Pengolahan Data.....	85
<b>Gambar 4.9</b> Hasil Akhir Pengolahan Data Pada WEKA .....	86

## DAFTAR RUMUS

	Halaman
<b>Rumus 2.1</b> Probabilitas <i>Bayes</i> .....	20
<b>Rumus 2.2</b> <i>Teorema Bayes</i> .....	21
<b>Rumus 2.3</b> <i>Naive Bayes classifier</i> .....	21
<b>Rumus 2.4</b> Label Kelas <i>C<sub>i</sub></i> .....	22
<b>Rumus 2.5</b> Kedekatan antara 2 kasus .....	23
<b>Rumus 2.6</b> Algoritma FCM .....	25
<b>Rumus 2.7</b> Hitung <i>Cluster centre</i> .....	26
<b>Rumus 2.8</b> Hitung jarak .....	26
<b>Rumus 2.9</b> Matriks partisi .....	26
<b>Rumus 2.10</b> Penghitungan error dalam <i>output layer</i> .....	28
<b>Rumus 2.11</b> Error pada <i>hidden layer</i> .....	28
<b>Rumus 2.12</b> Modifikasi bobot jaringan .....	28
<b>Rumus 2.13</b> Nilai <i>Support</i> .....	30
<b>Rumus 2.14</b> Nilai <i>support</i> dari 2 <i>item</i> .....	30
<b>Rumus 2.15</b> Nilai <i>Confidence</i> .....	30

## **DAFTAR LAMPIRAN**

1. Lampiran Data Penelitian
2. Lampiran Wawancara dan Dokumentasi

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Penelitian**

Berbicara mengenai segala hal yang berhubungan dengan perkembangan teknologi itu tidak akan ada habis-habisnya sama sekali. Setiap tahunnya pasti akan mengalami perkembangan, dimana perkembangan dari teknologi tersebut memiliki tujuan agar tercipta suatu teknologi yang lebih baik lagi dan bisa membuat perubahan dalam mempermudah semua tugas termasuk dalam pengolahan data penjualan. Salah satu teknologi yang banyak digunakan oleh masyarakat untuk membantu mempermudah pekerjaan mereka dalam pengolahan data penjualan adalah *data mining* (Buulolo, 2013).

Setiap aktivitas penjualan memiliki banyak data yang harus diolah agar data tersebut menjadi bermanfaat dan tidak hanya menjadi tumpukan data saja. Salah satu tempat yang selalu melakukan banyak transaksi penjualan adalah bengkel motor. Bengkel motor merupakan tempat penyedia jasa *service* motor serta penjualan suku cadang motor. Bengkel bergerak dibidang dunia bisnis, dimana yang setiap harinya akan terjadi transaksi, baik transaksi *service* motor ataupun penjualan suku cadang motor. Karena setiap harinya terjadi transaksi penjualan, maka kemungkinan besar data transaksi yang diperoleh akan sangat banyak dan menumpuk. Data transaksi yang diperoleh sering kali hanya digunakan sebagai arsip ataupun bukti penjualan, selain itu data tersebut sudah tidak memiliki fungsi lagi bagi bengkel tersebut, maka data tersebut hanya akan menjadi tumpukan data

yang tidak berguna. Terkadang barang yang diminta oleh konsumen tidak tersedia karena permintaan barang yang terlalu banyak sehingga tidak sebanding dengan barang yang tersedia. Hal inilah yang membuat konsumen mengeluh dan pihak bengkel juga mengalami kehilangan pelanggannya.

Jika strategi penjualan suku cadang motor tidak direncanakan dengan baik dan benar, maka strategi penjualan suku cadang motor tersebut akan membutuhkan ataupun menghabiskan banyak waktu, biaya dan tenaga. Untuk itu diperlukan teknik *data mining* untuk mengelola data dalam jumlah besar. Dengan adanya *data mining*, maka data tersebut dapat memberi keuntungan bagi bengkel Candra Motor agar dapat meminimalisir waktu kerja pengguna karena dapat mengolah data dalam jumlah besar. Salah satunya adalah dengan mengimplementasikan penggunaan *data mining* dengan algoritma *apriori*.

*Data mining* adalah suatu operasi yang menggunakan teknik atau metode tertentu untuk mencari pola atau bentuk yang berbeda dalam sebuah data yang terpilih (Bulolo, 2013). Algoritma *apriori* adalah algoritma pengambilan data dengan aturan asosiatif (*association rule*) untuk menentukan hubungan asosiatif suatu kombinasi item. *Association rule* yang dimaksud dilakukan melalui mekanisme penghitungan *support* dan *confidence* dari suatu hubungan *item* (Yanto & Khoiriah, 2015). *Data mining* dengan algoritma *apriori* dapat menghasilkan *rule* atau aturan sehinggaketerkaitan antar data dari transaksi penjualan suku cadang motor dapat diketahui barang apa yang sering terjual dan pihak perusahaan dapat menyediakan stok barang bisa di kontrol serta waktu pelayanan terhadap konsumen menjadi lebih efektif.

Berdasarkan latar belakang permasalahan di atas, maka peneliti ingin mengangkat judul penelitian “**Implementasi *Data Mining* Menggunakan Algoritma *Apriori* Pada Penjualan Suku Cadang Motor**”.

## **1.2 Identifikasi Masalah**

Berdasarkan latar belakang masalah diatas, peeliti dapat mengidentifikasi masalah-masalah berdasarkan latar belakang masalah yang telah dijelaskan sebelumnya yaitu:

1. Kurang efektifnya pemanfaatan data transaksi pada sistem penjualan suku cadang motor di Bengkel Candra Motor.
2. Kurang efektifnya waktu pelayanan terhadap konsumen karena banyaknya data yang tidak dikelola dengan baik.
3. Sulitnya mengetahui produk (*item*) yang sering terjual sehingga bengkel Candra Motor tidak mengetahui barang apa saja yang sudah tidak memiliki stok lagi.

## **1.3 Pembatasan Masalah**

Agar tidak menyimpang dari tujuan utama penulis, maka penulis membatasi masalah-masalah dalam penelitian ini yaitu :

1. Penelitian ini menggunakan data transaksi penjualan suku cadang motor di Bengkel Candra Motor.
2. Data yang digunakan dalam penelitian ini adalah tahun 2018-2019 penjualan suku cadang motor tiap bulannya di bengkel Candra Motor.

3. Suku cadang yang didata yaitu ban dalam, kampas kramol, oli mestran. Kampas, kain jok, oli gerbok, batu arang stater, oli Castrol, oli mpx, kampas cakram, bola lampu, oli yamalub, busi, oli ultratec, aki air, sel sock, rantai, oli SGO, oli TOP1.
4. Aplikasi yang digunakan untuk pengujian data adalah aplikasi WEKA 3.9.

#### **1.4 Perumusan Masalah**

Adapun perumusan masalah dalam penelitian ini adalah:

1. Bagaimana menggunakan algoritma *apriori* untuk membentuk pola kombinasi *itemsets* yang dapat mempermudah bengkel Candra Motor dalam mengolah data penjualan suku cadang motor?
2. Bagaimana cara keterkaitan *rule* transaksi barang suku cadang motor yang terjual?
3. Bagaimana manfaat penerapan *data mining* dalam sistem penjualan suku cadang motor di bengkel Candra Motor?

#### **1.5 Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah :

1. untuk mengetahui bagaimana algoritma *apriori* dapat membentuk pola kombinasi *itemsets* yang dapat mempermudah bengkel Candra Motor dalam mengolah data penjualan suku cadang motor.
2. Untuk mengetahui keterkaitan *rule* terhadap transaksi barang yang terjual pada suku cadang motor Bengkel Candra Motor.

3. Untuk mengetahui manfaat penerapan *data mining* dalam sistem penjualan suku cadang motor di bengkel Candra Motor.

## **1.6 Manfaat Penelitian**

Manfaat penelitian dapat dikategorikan menjadi dua, yaitu manfaat teoritis dan manfaat praktis.

### **1.6.1 Manfaat penelitian secara teoritis**

1. Menambah wawasan yang sudah ada tentang *data mining* menggunakan algoritma *apriori* .
2. Sebagai referensi pengolahan data penjualan dengan *data mining* .

### **1.6.2 Manfaat praktis**

1. Manfaat aplikasi ini untuk bengkel Candra Motor adalah sebagai alat bantu untuk mengelola data sistem penjualan.
2. Untuk mempermudah mengetahui barang yang paling banyak terjual di bengkel Chandra Motor.



## BAB II

### TINJAUAN PUSTAKA

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

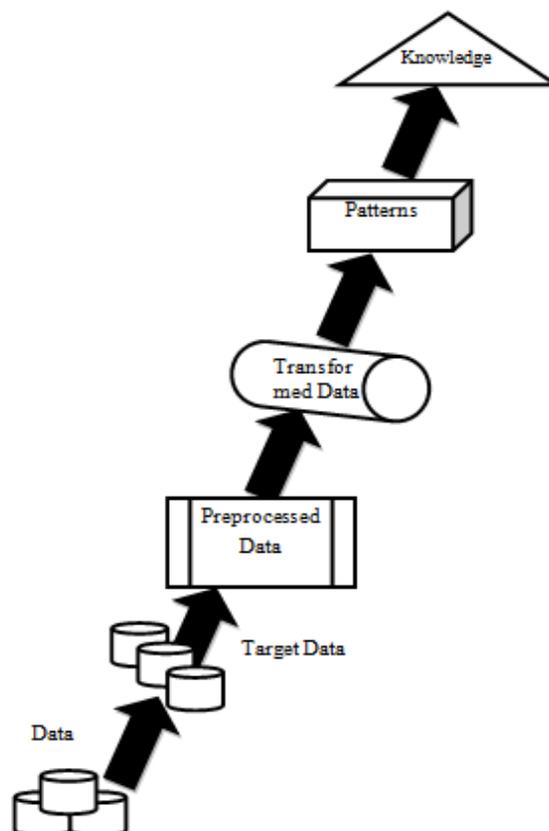
*Knowledge Discovery in Database* diartikan sebagai informasi potensial yang diekstraksi, implisit dan tidak diketahui dari beberapa kumpulan data. Suatu metode *knowledge discovery* mengaitkan hasil dari metode *data mining* (suatu metode yang kecenderungan mengekstrak pola suatu data), kemudian hasilnya diubah secara tepat sehingga menjadi suatu informasi yang mudah dipahami. KDD sendiri dapat didefinisikan sebagai keutuhan metode *non-trivial* yang berfungsi untuk menelusuri serta mengenali pola (*pattern*) didalam suatu data, dimana pola yang didapat berupa pola yang baru, sah, dapat dipahami serta dapat berguna (Annurullah & Maulana, 2018).

KDD berkaitan dengan suatu metode penemuan ilmiah serta integrasi, visualisasi dan interpretasi dari beberapa pola sejumlah data. Rangkaian metode tersebut mempunyai tahapan sebagai berikut (Retno Tri Vulandari, 2017, hal:2-3):

1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*).
2. Integrasi data (penggabungan data dari beberapa sumber).
3. Perubahan atau transformasi data (data diubah menjadi bentuk yang sesuai untuk di-*mining*).
4. Aplikasi teknik *data mining*, proses ekstraksi pola dari data yang ada.

5. Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan).
6. Presentasi pengetahuan (dengan teknik visualisasi).

Semua tahapan ini merupakan komponen dari metode penelitian pengetahuan yang melingkupi peninjauan apakah informasi maupun pola yang didapat bertolak belakang dengan hipotesis atau fakta yang sudah ada sebelumnya. Langkah terakhir KDD yaitu pengetahuan dipresentasikan dengan bentuk yang mudah dimengerti pengguna.



**Gambar 2.1** Tahapan Knowledge Discovery in Database (KDD)

(Sumber : Data Penelitian,2019)

Istilah *data mining* dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk memahami proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD akan dijelaskan secara garis besar seperti dibawah ini(Kusrini dan Emha Taufiq Luthfi, 2009, hal:6-7):

1. *Data Selection*

Penseleksian data dari beberapa sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil pemilihan yang akan dipakai untuk tahapan *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing/cleaning*

Sebelum proses *data mining* bisa dilakukan, perlu dilaksanakan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup beberapa hal, antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). Selain itu dilakukan juga proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. *Transformation*

*Coding* merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding*

dalam KDD adalah proses kreatif dan sangat tergantung pada jenis ataupun pola informasi yang akan dicari dalam basis data.

4. *Data mining*

*Data mining* merupakan proses mencari pola atau informasi menarik dalam data yang terpilih dengan memakai teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation / Evaluation*

Pola informasi yang didapatkan dari proses *data mining* perlu ditampilkan dalam bentuk yang dapat dimengerti oleh pihak yang berkepentingan. Tahap ini adalah bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya

## 2.2 *Data Mining*

*Data Mining* adalah metode yang digunakan untuk mengidentifikasi dan mengekstraksi informasi yang berguna dan pengetahuan yang tersusun dari berbagai *database* besar ataupun *data warehouse* yang menggunakan cara matematika, kecerdasan buatan, statistik, serta *machine learning* (Bahtiar et al., 2016).

Menurut definisi-definisi yang sudah disampaikan, hal penting yang berkaitan dengan *data mining* yaitu:

1. *Data mining* merupakan suatu metode otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan *data mining* adalah mendapatkan hubungan atau pola yang akan mungkin memberikan indikasi yang bermanfaat.

*Data Mining* merupakan metode pengekstraksian informasi dengan penggunaan algoritma dan cara yang mengaitkan bidang sistem manajemen *database*, ilmu *statistic*, dan mesin pembelajaran dari kumpulan beberapa data. *Data Mining* dipakai untuk mengekstraksi informasi penting dan berguna yang terdapat dalam *dataset* yang besar. Dengan menggunakan *data mining* maka akan ditemukan suatu informasi pengetahuan yang terdapat pada suatu kumpulan data yang sangat banyak (Yanto & Khoiriah, 2015).

Berdasarkan pemanfaatan *data mining* dilihat dari dua sudut pandang, yaitu sudut pandang komersial dan sudut pandang keilmuan. Pada sudut pandang komersial, *data mining* digunakan karena untuk menanggulangi volume data yang berlebihan, untuk menghasilkan informasi-informasi yang dibutuhkan yang merupakan asset yang dapat meningkatkan daya saing suatu institusi menggunakan teknik komputasi (Retno Tri Vlandari, 2017, hal:3-4).

Sedangkan dari segi sudut pandang keilmuan, *data mining* berfungsi untuk menganalisis, meng*capture* serta menyimpan data yang bersifat sangat besar dan *real time*, contohnya :

1. *Remote sensor* yang ditempatkan pada satelit luar angkasa.

2. *Telescope* yang digunakan untuk memindai langit yang digunakan oleh para ilmuan.
3. Simulasi saintifik yang membangkitkan data dalam ukuran besar atau ukuran *terabytes*.

### 2.2.1 Manfaat *Data Mining*

Pemanfaatan *data mining* dapat dilihat dari dua sudut pandang, yaitu sudut pandang komersial dan sudut pandang keilmuan (Retno Tri Vlandari, 2017).

1. Dari segi sudut pandang komersial, penggunaan *data mining* dapat dipakai untuk menangani meledaknya volume data, dengan menggunakan teknik komputasi dapat digunakan untuk menghasilkan informasi-informasi yang dibutuhkan yang merupakan asset yang dapat meningkatkan daya saing suatu institusi.

Contohnya:

- a. Bagaimana mengetahui hilangnya pelanggan karena pesaing.
- b. Bagaimana mengetahui *item* produk atau konsumen yang memiliki kesamaan karakteristik.
- c. Bagaimana mengidentifikasi produk-produk yang terjual bersamaan dengan produk lain.
- d. Bagaimana memprediksi tingkat penjualan.
- e. Bagaimana menilai tingkat resiko dalam menentukan jumlah produksi suatu *item*.
- f. Bagaimana memprediksi perilaku bisnis dimasa yang akan datang.

2. Dari sudut pandang keilmuan, *data mining* dapat digunakan untuk *mengcapture*, menganalisis serta menyimpan data yang bersifat *real time* dan sangat besar, misalnya:
  - a. *Remote sensor* yang ditempatkan pada suatu satelit.
  - b. *Telescope* yang digunakan untuk memindai langit.
  - c. Simulasi saintifik yang membangkitkan data dalam ukuran *terabytes*.

### 2.2.2 Fungsi Data Mining

Fungsi-fungsi *data mining* yang umum dan biasa diterapkan adalah (Retno Tri Vulandari, 2017, hal:4-5):

1. *Association*, adalah proses untuk mendapatkan aturan asosiasi antara suatu kombinasi *item* dalam suatu waktu.
2. *Sequence*, proses untuk mendapatkan aturan asosiasi antara suatu kombinasi *item* dalam suatu waktu dan diterapkan lebih dari satu periode.
3. *Clustering*, merupakan proses pengelompokan sejumlah data/obyek kedalam kelompok data sehingga setiap kelompok berisi data yang mirip.
4. *Classification*, proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.
5. *Regression*, merupakan proses pemetaan data dalam suatu nilai prediksi.
6. *Forecasting*, adalah proses pengestimasi nilai prediksi berdasarkan pola-pola didalam sekumpulan data.

7. *Solution*, merupakan proses penemuan akar masalah dan *problem solving* dari persoalan bisnis yang dihadapkan atau paling tidak sebagai informasi dalam pengambilan keputusan.

### 2.2.3 Pengelompokan *Data Mining*

*Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Kusrini dan Emha Taufiq Luthfi, 2009, hal:10-12):

1. *Description* (Deskripsi)

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikankemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. *Estimation*

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik dari pada kearah kategori. Model dibangun menggunakan *record* lengkap yang menyediakannilai dari variabel target sebagai prediksi.Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan

menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

### 3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang. Contoh prediksi bisnis dan penelitian adalah:

- a. Prediksi harga beras dalam tiga bulan yang akan datang.
- b. Prediksi persentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat puladigunakan (untuk keadaan yang tepat) untuk prediksi.

### 4. Klasifikasi

Dalam klasifikasi, terdapat target variable kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Contoh lain klasifikasi dalam bisnis dan penelitian adalah:

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau tidak.
- b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- c. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa.

## 5. Pengklusteran (*Clustering*)

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

- a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari satu suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku *financial* dalam baik dan mencurigakan.
- c. Melakukan pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.

## 6. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan *attribut* yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh asosiasi dalam bisnis dan penelitian adalah:

- a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran *upgrade* layanan yang diberikan.
- b. Menentukan barang dalam supermarket yang dibeli secara bersamaan dan yang dibeli secara bersamaan dan yang tidak pernah dibeli secara bersamaan.

#### **2.2.4 Teknik Pembelajaran *Data Mining***

Teknik yang digunakan dalam data mining erat kaitannya dengan “penemuan” (*discovery*) dan “pembelajaran” (*learning*) yang terbagi dalam tiga metode utama pembelajaran yaitu (Retno Tri Vlandari, 2017, hal:7-8) :

- a. *Supervised Learning*

*Supervised Learning* merupakan teknik yang paling banyak digunakan. Teknik ini sama dengan “*programming by example*”. Teknik ini melibatkan fase pelatihan dimana data pelatihan historis yang karakter-karakternya dipetakan ke hasil-hasil yang telah diketahui diolah dalam algoritma *data mining*. Proses ini melatih algoritma untuk mengenali variabel-variabel dan nilai-nilai kunci yang nantinya akan digunakan sebagai dasar dalam membuat perkiraan-perkiraan ketika diberikan data baru.

- b. *Unsupervised Learning*

Teknik pembelajaran ini tidak melibatkan fase pelatihan seperti yang terdapat pada *supervised learning*. Teknik ini bergantung pada penggunaan algoritma yang mendeteksi semua pola, seperti *associations* dan *sequences*, yang

muncul dari kriteria penting yang spesifik dalam data masukan. Pendekatan ini mengarah pada pembuatan banyak aturan (*rules*) yang mengkarakterisasikan penemuan *associations*, *clusters*, dan *segments*. Aturan-aturan ini kemudian dianalisis untuk menemukan hal-hal yang penting.

c. *Reinforcement Learning*

Teknik pembelajaran ini jarang digunakan dibandingkan dengan dua teknik lainnya, namun memiliki penerapan-penerapan yang terus dioptimalkan dari waktu ke waktu dan memiliki kontrol adaptif. Teknik ini sangat menyerupai kehidupan nyata yaitu seperti “*on-job-training*”, dimana seorang pekerja diberikan sekumpulan tugas yang membutuhkan keputusan-keputusan. Pada beberapa titik waktu kelak diberikan penilaian atas *performance* pekerja tersebut kemudian pekerja diminta mengevaluasi keputusan-keputusan yang telah dibuatnya sehubungan dengan hasil *performance* pekerja tersebut. *Reinforcement learning* sangat tepat digunakan untuk menyelesaikan masalah-masalah yang sulit bergantung pada waktu.

### **2.2.5 Penerapan *Data Mining***

*Data mining* dapat diterapkan pada (Retno Tri Vlandari, 2017, hal:5-7):

1. Analisa pasar dan manajemen

Sumber data yang digunakan seperti transaksi kartu kredit, kartu anggota *club* tertentu, kupon diskon, keluhan pembeli, ditambah dengan studi tentang gaya hidup publik.

Beberapa solusi yang dapat diselesaikan dengan *data mining* antara lain:

a. Menembak target pasar

*Data mining* dapat dilakukan pengelompokan (*clustering*) dari model-model pembeli dan melakukan klasifikasi terhadap setiap pembeli sesuai dengan karakteristik yang diinginkan seperti kedukaan yang sama, tingkat penghasilan yang sama, kebiasaan membeli dan karakteristik lainnya.

b. Melihat pola beli pemakai dari waktu ke waktu

*Data mining* dapat digunakan untuk melihat pola beli dari waktu ke waktu. Sebagai contoh, ketika seorang menikah bias saja dia kemudian memutuskan untuk pindah dari *single account* ke *joint account*.

c. *Cross market analysis*

Kita dapat memanfaatkan untuk melihat hubungan antara penjualan satu produk dengan produk lainnya.

d. *Profil customer*

*Data mining* dapat melihat *profil customer* sehingga dapat mengetahui kelompok customer tertentu suka membeli produk apa saja.

e. Identifikasi kebutuhan customer

Dapat mengidentifikasi produk apa saja yang terbaik untuk tiap kelompok customer dan factor apa saja yang dapat menarik konsumen.

f. Menilai *loyalitas customer*

g. Informasi *summary*

Dapat digunakan untuk membuat laporan *summary* yang bersifat multi dimensi dan dilengkapi dengan informasi *statistic* lainnya.

2. Analisa perusahaan dan manajemen resiko

a. Perencanaan keuangan dan evaluasi aset

*Data mining* dapat membantu melakukan analisi dan prediksi *cash flow* serta dapat melakukan *contingent claim analysis* untuk mengevaluasi aset. Selain itu dapat menggunakan untuk analisi trend.

b. Perencanaan sumber daya

Dengan melihat ringkasan informasi serta pola pembelanjaan dan pemasukan dari masing-masing *resource*. Maka dapat memanfaatkan untuk *resource planning*.

c. Persaingan

*Data mining* dapat membantu untuk memonitor pesaing-pesaing dengan melihat *market direction* mereka. *Data mining* juga dapat melakukan pengelompokan *customer* dan dapat memberikan variasi harga untuk masing-masing grup.

3. Telekomunikasi

*Data mining* melihat jutaan transaksi yang masuk, dan melihat transaksi mana sajakah yang masih harus ditangani secara manual, tujuannya adalah untuk menambah layanan otomatis.

## 2.3 Metode Data Mining

Didalam *data mining* terdapat beberapa metode yang dapat membantu dalam hal pengolahan data, diantaranya adalah :

### 2.3.1 Klasifikasi Bayes

Teori keputusan *Bayes* yaitu suatu pendekatan yang statistika fundamental dalam suatu data mining. Pendekatan tersebut dilandaskan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan probabilitas (Retno Tri Vulandari, 2017, hal:33-42). Metode *Bayes* ini ditemukan oleh Thomas Bayes pada abad ke-18. Dalam metode *Bayes*, probabilitas ataupun peluang bersyarat dinyatakan dengan rumus sebagai berikut :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

**Rumus 2.1** Probabilitas *Bayes*

Dimana keterangan rumus diatas yaitu X dilambangkan sebagai bukti, H sebagai hipotesis,  $P(H|X)$  sebagai probabilitas dimana hipotesis H benar untuk bukti X atau dengan kata lain  $P(H|X)$  merupakan probabilitas *posterior* H dengan syarat X,  $P(H|X)$  adalah probabilitas bahwa bukti X benar untuk hipotesis H atau probabilitas *posterior* X dengan syarat H,  $P(H)$  adalah probabilitas *prior* hipotesis H, dan  $P(X)$  adalah probabilitas *prior* X (Suyanto, 2017, hal:126-133).

*Naïve Bayesian Classifier* atau *Simple Bayesian classifier* dapat digunakan untuk melakukan estimasi  $P(H|X)$  untuk menentukan probabilitas bahwa tuple X berada didalam kelas C, dimana metode ini bekerja dengan lima langkah sebagai berikut :

1. Misalkan  $D$  adalah himpunan data latihan (*training set*) yang berisi sejumlah tuple beserta label kelasnya. Setiap tuple adalah berdimension  $n$  yang dinyatakan sebagai  $X = (X_1, X_2, \dots, X_n)$  yang didapat dari  $n$  atribut  $(A_1, A_2, \dots, A_n)$ .

2. Misalkan terdapat  $m$  kelas, yaitu  $C_1, C_2, \dots, C_m$ . Untuk sebuah tuple masukan  $X$ , *Naïve Bayes classifier* memprediksi bahwa tuple  $X$  termasuk ke dalam kelas  $C_i$  jika dan hanya jika

$$P(C_i|X) > P(C_j|X) \text{ untuk } \leq j \leq m, j \neq i$$

*Naïve Bayes classifier* bekerja dengan cara mencari nilai maksimal  $P(C_i|X)$  bernilai maksimal yang disebut *maximum posteriori hypothesis*. Dengan teorema Bayes,  $P(C_i|X)$  diperkirakan dengan rumus sebagai berikut :

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

**Rumus 2.2 Teorema Bayes**

3.  $P(X)$  bernilai sama untuk semua kelas (maksudnya, tuple  $X$  memiliki probabilitas yang sama untuk masuk ke dalam kelas manapun), maka hanya  $P(X|C_i) P(C_i)$  yang perlu dimaksimalkan.

4. Jika kita berhadapan dengan himpunan data yang memiliki sangat banyak atribut, kita dapat mereduksi atau menyaring kompleksitas perhitungan  $P(X|C_i)$  dengan asumsi naïf tentang independensi bersyarat kelas, yaitu: nilai-nilai atribut adalah saling independen. Maksudnya, antar atribut adalah saling bebas, tidak ada ketergantungan sama sekali. Dengan demikian, *Naïve Bayes classifier* merumuskan sebagai berikut:

$$P(C_i|X) = \prod_{k=1}^n P(X_k|C_i) = P(X_1|C_i) \times P(X_2|C_i) \times \dots \times P(X_n|C_i)$$

**Rumus 2.3 Naive Bayes classifier**

Menurut tuple-tuple pada himpunan data latih, kita dapat memperkirakan

$P(X_1 | C_i), P(X_2 | C_i), \dots, P(X_n | C_i)$  berdasarkan tipe atributnya.

5. Untuk memprediksi label kelas dari tuple  $X$ , kita harus menghitung probabilitas  $P(X|C_i)P(C_i)$  untuk setiap kelas  $C_i$ . Selanjutnya, kita hanya

perlu memaksimalkan probabilitas tersebut, yaitu mencari kelas  $C_i$  yang menghasilkan probabilitas  $P(X|C_i)P(C_i)$  maksimum sebagai kelas keputusan.

Secara matematis, tuple  $X$  diberi label kelas  $C_i$  jika dan hanya jika

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \text{ untuk } 1 \leq j \leq m, j \neq i$$

**Rumus 2.4** Label Kelas  $C_i$

$$P(X | C_i) P(C_i) > P(X | C_j) P(C_j)$$

### 2.3.2 Metode *Decision Tree*

*Decision tree* merupakan suatu metode klasifikasi praktis yang banyak digunakan serta sangat populer (Suyanto, 2017, hal:134). Pohon pengambilan keputusan dalam analisis pemecahan suatu masalah adalah pemetaan tentang cara-cara pemecahan masalah yang dapat ditemukan dari dalam masalah tersebut. Pohon tersebut juga menunjukkan penyebab kemungkinan/probabilitas yang akan berpengaruh pada alternatif-alternatif keputusan tersebut, beserta perkiraan hasil akhir yang akan didapatkan apabila kita mengambil alternatif keputusan tersebut (Retno Tri Vulandari, 2017, hal:15).

### 2.3.3 Algoritma *Nearest Neighbor*

*Nearest Neighbor* merupakan pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada (Kusrini dan Emha Taufiq Luthfi, 2009, hal:93-94).

Adapun rumus untuk melakukan perhitungan kedekatan antara dua kasus adalah sebagai berikut.

$$\text{Similarity}(T,S) = \frac{\sum_{i=1}^n f(T_i,S_i) * w_i}{w_i}$$

**Rumus 2.5** Kedekatan antara 2 kasus

Keterangan:

T : Kasus Baru

S : Kasus yang ada dalam penyimpanan

n : Jumlah atribut dalam setiap kasus

i : Atribut individu antara 1 s.d. n

f : Fungsi *similarity* atribut i antara kasus T dan kasus S

w : Bobot yang diberikan pada atribut ke-i

Kedekatan biasanya berada pada nilai antara 0 s.d. 1. Nilai 0 artinya kedua kasus mutlak tidak mirip, sebaliknya untuk nilai 1 kasus mirip dengan mutlak.

### 2.3.4 *Fuzzy C-Means*

Analisis kluster atau *clustering* adalah proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu

kelompok lebih besar daripada kesamaan data tersebut dengan data kelompok lain (Kusrini dan Emha Taufiq Luthfi, 2009, hal:177-181). Potensi *clustering* adalah dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klasifikasi, pengolahan gambar, dan pengenalan pola.

*Clustering* data diterapkan ke dalam data yang kuantitatif (numerik), kualitatif (kategorikal), atau kombinasi dari keduanya. Data dapat berupa hasil pengamatan dari suatu proses. Setiap pengamatan dapat memiliki  $n$  variabel pengukuran dan dikelompokkan dalam  $n$  dimensi vektor  $Z_k = [Z_{1k}, \dots, Z_{nk}]^T$ ,  $Z_k \in R^n$ . Sebuah himpunan dari  $N$  pengamatan dinotasikan dengan  $Z = \{Z_k \mid k = 1, 2, \dots, N\}$  dan direpresentasikan sebagai matriks  $n \times N$ .

$$Z = \begin{bmatrix} Z_{11} & \dots & Z_{1N} \\ \vdots & \vdots & \vdots \\ Z_{n1} & \dots & Z_{nN} \end{bmatrix}$$

Dalam terminology pengenalan pola, kolom dalam matriks disebut *patterns* atau *objects*. Baris disebut *features* atau *attribute*, dan  $Z$  disebut *pattern* atau *data matrix*. Arti kolom dan baris dalam  $Z$  bergantung pada konteks. Dalam diagnosis medik, sederhananya, kolom  $Z$  dapat direpresentasikan sebagai pasien, dan baris merupakan *symptoms* atau pemeriksaan pasien untuk pasien bersangkutan. *Cluster* secara umum merupakan wujud himpunan bagian dari suatu himpunan data dan metode *clustering* dapat diklasifikasikan berdasarkan himpunan bagian yang dihasilkan: apakah *fuzzy* atau *crisp (hard)*.

Metode *hard clustering* merupakan model yang berdasarkan pada teori himpunan klasik, yang mana suatu objek menjadi anggota atau tidak menjadi

anggota secara penuh ke dalam suatu kelompok. *Hard clustering* membagi data ke dalam sejumlah himpunan bagian secara eksklusif. Sebaliknya, metode *Fuzzy clustering* mengizinkan objek untuk menjadi bagian dari beberapa kelompok secara bersamaan dengan perbedaan level keanggotaan.

Terdapat banyak algoritma yang digunakan untuk *clustering*. Salah satunya *Fuzzy C-Means* (FCM). *Fuzzy C-Means Clustering* (FCM) adalah algoritma *clustering* data yang setiap datanya menjadi anggota dari suatu kluster dengan derajat didefinisikan dengan level keanggotaan.

Algoritma FCM didasarkan pada minimasi fungsi objektif yang diformulasikan dalam persamaan:

$$J_m = \sum_{i=1}^n \sum_{j=1}^c (u_{ij})^m \|Z_i - C_j\|^2$$

**Rumus 2.6** Algoritma FCM

Keterangan:

$u_{ij}$  merupakan level keanggotaan dari  $Z_i$  dalam cluster  $j$

$Z_i$  merupakan nilai data ke- $i$  dari  $d$ -dimensi data

$c_j$  merupakan nilai ke- $j$  dari  $d$ -dimensi *cluster center*

$M$  merupakan sembarang bilangan real lebih besar dari 1

Selanjutnya algoritma FCM dengan langkah sebagai berikut.

Langkah1 : Tentukan himpunan data  $Z$ . Tentukan jumlah *cluster* yang diharapkan  $1 < c < N$ , nilai pembobot  $m > 1$ , toleransi penghentian  $\epsilon > 0$ .

Langkah2 : Inisialisasi matriks partisi secara acak,  $U(0) \in M_{fc}$

ulangi untuk  $l = 3, 4, 5$

Langkah3 : Hitung *cluster center (means)*.

$$v_i^{(l)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(i-1)})^m z_k}{\sum_{k=1}^N (\mu_{ik}^{(i-1)})^m}, 1 \leq i \leq c$$

**Rumus 2.7** Hitung *Cluster centre*

Langkah4 : Hitung jarak.

$$D_{ikA}^2 = (z_k - v_i^{(l)})^T A (z_k - v_i^{(l)}), 1 \leq i \leq c, 1 \leq k \leq N$$

**Rumus 2.8** Hitung jarak

Langkah5 : Perbaharui matriks partisi.

Untuk  $1 \leq k \leq N$

Jika  $D_{ikA} > 0$  untuk semua  $i = 1, 2, \dots, c$

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^c (D_{ikA} / D_{jkA})^{2/(m-1)}}$$

**Rumus 2.9** Matriks partisi

Atau dengan kata lain:

$$\mu_{ik}^{(l)} = 0 \text{ jika } D_{ikA} > 0 \text{ dan } \mu_{ik}^{(l)} \in [0,1] \text{ dengan } \sum_{i=1}^c \mu_{ik}^{(l)} = 1$$

Ulangi sampai  $\|U^{(l)} - U^{(l-1)}\| < \varepsilon$ .

### 2.3.5 Backpropagation

*Backpropagation* adalah salah satu algoritma pembelajaran dalam jaringan saraf tiruan. Proses pembelajaran dalam *backpropagation* dilakukan penyesuaian bobot-bobot jaringan saraf tiruan dengan arah mundur berdasarkan nilai *error* dalam proses pembelajaran (Kusrini dan Emha Taufiq Luthfi, 2009, hal:199-202).

*Backpropagation* adalah metode pembelajaran jaringan saraf tiruan yang paling umum digunakan. *Backpropagation* bekerja melalui proses secara iterative dengan menggunakan sekumpulan contoh data (*data training*), membandingkan nilai prediksi dari jaringan dengan setiap contoh data. Dalam setiap proses, bobot relasi dalam jaringan dimodifikasi untuk meminimalkan nilai *Mean Squared Error* (MSE) antara nilai prediksi dari jaringan dengan nilai sesungguhnya. Modifikasi relasi jaringan saraf tersebut dilakukan dalam arah mundur, dari output layer hingga layer pertama dari *hidden layer* sehingga metode ini disebut *backpropagation*.

Langkah pembelajaran dalam metode *backpropagation* adalah sebagai berikut:

1. Inisialisasi bobot jaringan secara acak (biasanya antar -1.0 hingga 1.0).
2. Untuk setiap contoh data (*data training*), hitung keluaran berdasarkan bobot jaringan saat tersebut.
3. Selanjutnya lakukan proses perhitungan nilai error untuk setiap keluaran (*output*) dan *hidden node* (neuron) dalam jaringan. Bobot relasi jaringan dimodifikasi.
4. Ulangi langkah 2 hingga kondisi diinginkan tercapai.

Terdapat beberapa cara dalam modifikasi bobot-bobot jaringan saraf tiruan. Pertama, modifikasi dilakukan pada setiap akhir penghitungan setiap contoh kasus yang biasa disebut *case (online) updating*. Atau cara kedua, modifikasi bobot-bobot jaringan saraf tiruan dilakukan setelah semua contoh kasus dianalisis, Cara ini disebut *epoch (batch) updating*.

Untuk penghitungan error dalam *output layer* dilakukan dengan persamaan berikut.

$$Err_i = O_i (1-O_i)(T_i-O_i)$$

**Rumus 2.10** Penghitungan error dalam *output layer*

Dalam hal ini:

$O_i$  = Keluaran dari *output node* unit  $i$

$T_i$  = Nilai sesungguhnya dari *output node* dalam contoh kasus (*data training*)

Perhitungan *error* pada *hidden layer* dilakukan dengan persamaan berikut.

$$Err_i = O_i (1-O_i) \sum_j Err_j W_{ij}$$

**Rumus 2.11** Error pada *hidden layer*

Dalam hal ini:

$O_i$  = Keluaran dari *hidden node* unit  $I$  yang memiliki keluaran  $j$  dalam *layer*

$Err_j$  = Nilai *error* dalam *node* unit  $j$

$w_{ij}$  = Bobot antara kedua *node (neuron)* tersebut

Setelah nilai error pada tiap *node (neuron)* dihitung, lakukan modifikasi terhadap bobot jaringan dengan menggunakan persamaan berikut.

$$w_{ij} = w_{ij} + l \cdot Err_j \cdot O_i$$

**Rumus 2.12** Modifikasi bobot jaringan

Dalam hal ini:

$l$  = *Learning rate* dengan nilai antara 0 hingga 1;

Jika nilai  $l$  kecil, maka perubahan bobot akan sedikit dalam setiap iterasi; begitu pula sebaliknya. Nilai *leaning rate* biasanya berkurang selama proses pembelajaran.

### 2.3.6 Algoritma Apriori

Algoritma *apriori* salah satu jenis aturan asosiasi pada *data mining*. Selain *apriori*, yang termasuk pada golongan ini adalah metode *Generalized Rule Induction* dan *Algoritma Hash Based* (Kusrini dan Emha Taufiq Luthfi, 2009, hal:149).

Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market basket analysis*. Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk mendapatkan aturan asosiatif antara suatu kombinasi *item*. Karena analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisis isi keranjang belanja di pasar swalayan, analisis asosiasi juga sering disebut *market basket analysis*.

Analisis asosiasi dikenal juga sebagai salah satu teknik *data mining* yang menjadi dasar dari berbagai teknik *data mining* lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (*frequent pattern mining*). Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, yaitu *support* dan *confidence*. *Support* (nilai penunjang) adalah persentase kombinasi *item* tersebut dalam database, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar-*item* dalam aturan asosiasi.

Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiasi yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap:

### 1. Analisis pola frekuensi tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam database. Nilai *support* sebuah *item* diperoleh dengan rumus berikut.

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A}}{\text{Total Transaksi}}$$

**Rumus 2.13** Nilai *Support*

Sementara itu, nilai *support* dari 2 *item* diperoleh dari rumus 2 berikut.

$$\text{Support (A, B)} = P(A \cap B)$$

$$\text{Support (A, B)} = \frac{\Sigma \text{Transaksi mengandung A dan B}}{\Sigma \text{Transaksi}}$$

**Rumus 2.14** Nilai *support* dari 2 *item*

### 2. Pembentukan Aturan Asosiasi

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif  $A \rightarrow B$ .

Nilai *confidence* dari aturan  $A \rightarrow B$  diperoleh dari rumus berikut..

$$\text{Confidence} = P(B | A) = \frac{\Sigma \text{Transaksi mengandung A dan B}}{\Sigma \text{Transaksi mengandung A}}$$

**Rumus 2.15** Nilai *Confidence*

#### 2.4 Software Pendukung

*Software* merupakan bagian dari sistem komputer berupa program yang dapat menjalankan suatu perintah dan dapat dioperasikan oleh penggunanya melalui perangkat komputer. Pada penelitian ini *software* yang digunakan adalah

WEKA. WEKA merupakan sebuah paket *tools machine learning* mudah dan praktis. Waikato Environment for Knowledge Analysis atau sering disebut “WEKA”, yang dibuat di Universitas Waikato, New Zealand untuk penelitian, pendidikan dan berbagai aplikasi (Retno Tri Vlandari, 2017, hal:77).

Pada beberapa tingkatan yang berbeda, WEKA mudah digunakan dan diterapkan. Terdapat implementasi algoritma-algoritma pembelajaran *state-of-the-art* yang dapat dipakai pada dataset dari *command line*. WEKA mengandung tools untuk *pre-processing* data, klasifikasi, regresi, *clustering*, aturan asosiasi, dan visualisasi. Contoh penerapan WEKA yaitu dengan menerapkan suatu metode pembelajaran ke dataset dan menganalisa ataupun meneliti hasilnya untuk mendapatkan informasi tentang data tersebut, atau menerapkan beberapa metode dan membandingkan hasilnya untuk dipilih.

Pengembangan WEKA mengikuti model releases Linux: digit kedua yang genap menunjukkan release yang stabil dan digit kedua yang ganjil menunjukkan release ‘pengembangan’(contohnya 3.0.x adalah release stabil, sedangkan 3.1.x adalah release yang sedang berkembang). Dibawah ini beberapa versi awal dari WEKA, yaitu :

1. WEKA 3.0 : “versi buku” yang sesuai dengan penjelasan dari buku *data mining*.
2. WEKA 3.2 : “versi GUI” yang menambahkan GUI dari CLI awal.
3. WEKA 3.3 : “versi pengembangan” dengan berbagai peningkatan.

## 2.5 Penelitian Terdahulu

Sebagai bahan pertimbangan dalam penelitian ini, maka peneliti mencantumkan beberapa referensi penelitian yang diambil dari beberapa jurnal ilmiah *data mining*, yaitu:

1. Penelitian yang dilakukan oleh (Yulia & Azwanti, 2018) dengan judul **“Penerapan Algoritma C4.5 Untuk Memprediksi Besarnya Penggunaan Listrik Rumah Tangga di Kota Batam”**. Penelitian ini membahas tentang kebutuhan listrik yang telah menjadi kebutuhan primer, jika tidak ada listrik maka semua kegiatan yang ingin dilakukan akan terhambat. Kepadatan penduduk di Batam menyebabkan kebutuhan listrik semakin meningkat dan sangat tinggi. Jika penduduk tidak bijak dalam memakai listrik, hal ini akan menyebabkan menipisnya persediaan energy listrik di Kota Batam. Penelitian ini menerapkan penggunaan metode Algoritma C4.5 yang dapat menghasilkan model berupa pohon atau aturan yang mudah diimplementasikan, sehingga dapat membantu memprediksi penggunaan listrik pada rumah tangga di kota Batam.
2. Penelitian yang dilakukan oleh (Haryanto, Oslan, & Dwiyana, 2011) dengan judul **“Implementasi Analisis Keranjang Belanja Dengan Aturan Asosiasi Menggunakan Algoritma Apriori Pada Penjualan Suku Cadang Motor”** penelitian ini membahas bahwa penawaran suatu produk suku cadang motor kepada setiap konsumen yang biasanya dilakukan menggunakan pertimbangan tingkat produk yang banyak terjual ataupun tujuan tertentu. Salah satu pertimbangan pemasaran terhadap konsumen

yang bisa digunakan adalah dengan memperoleh pola saling keterkaitan antar barang yang satu dengan yang lainnya. Saling keterkaitan tersebut dapat dilihat dari barang yang paling sering terjual secara bersamaan. Maka pada penelitian ini akan dilakukan analisis dengan menggunakan metode *market basket analysis* terhadap data transaksi penjualan suku cadang motor agar dapat menentukan aturan penawaran barang tersebut dari pola transaksi penjualan barang tersebut.

3. Penelitian yang dilakukan oleh (Bahtiar et al., 2016) dengan judul **“Implementasi *Data Mining* Algoritme Apriori Pada Penjualan Suku Cadang Motor Delta Motor”** penelitian ini mengangkat permasalahan yang ada pada bengkel Delta Motor yang merupakan bengkel sepeda motor yang menyediakan layanan *service* serta pembelian berbagai macam suku cadang motor. Selama ini data-data penjualan yang ada di Delta Motor hanya dipakai sebagai rekapan ataupun arsip penjualan saja. Padahal data-data itu selain sebagai arsip penjualan, dapat-data tersebut juga digunakan pemilik untuk mengembangkan strategi pemasaran pada bengkelnya. Dengan teknologi *data mining* menggunakan *apriori* bisa menghasilkan sebuah rekomendasi yang dapat membantu dalam mengembangkan keputusan strategi penjualan.
4. Penelitian yang dilakukan oleh (Annurullah & Maulana, 2018) dengan judul **“Penerapan *Data Mining* Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma *FP-Growth* Pada Data Transaksi penjualan *Spare Part Motor*”** penelitian ini mengangkat permasalahan banyaknya data

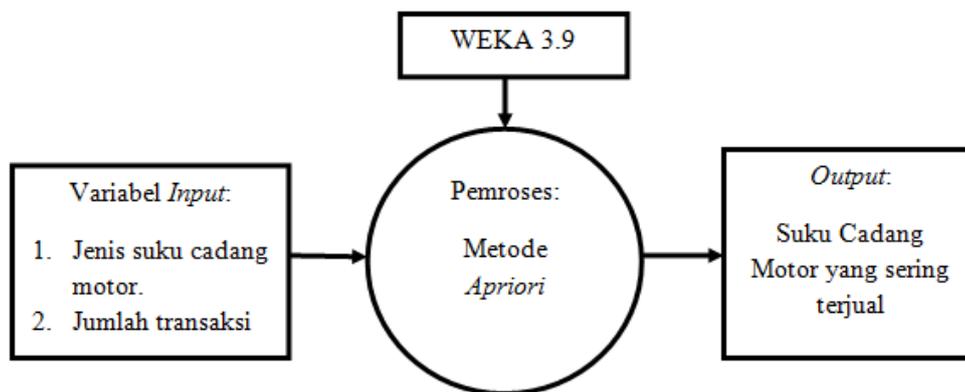
perusahaan yang semakin hari semakin bertambah, maka perusahaan tersebut membutuhkan proses penganalisis yang dapat dilakukan secara tepat dan akurat dengan aplikasi yang berbasis komputer. Penelitian ini memakai algoritma *FP-Growth* dalam aplikasi yang dibuat untuk mengetahui pola pembelian pada setiap cabang yang berbeda dengan karakteristik berbeda juga. Dari pola tersebut akan didapatkan kesimpulan dari informasi itu.

5. Penelitian yang dilakukan oleh (Yanto & Khoiriah, 2015) dengan judul **“Implementasi *Data Mining* Dengan Metode Algoritma Apriori Dalam Menentukan Pola Pembelian Obat”** penelitian ini membahas tentang persaingan yang di dunia bisnis, khususnya didalam industry Apotek, sehingga para pebisnis dituntut untuk mendapatkan strategi pemasaran yang dapat meningkatkan penjualan pada Apotek tersebut. Salah satu solusi untuk mengatasinya yaitu dengan tetap tersedianya stok jenis obat apa saja yang banyak dibutuhkan oleh konsumen. Untuk dapat mengetahui jenis obat apa saja yang paling sering dibeli oleh konsumen, penerapan algoritma apriori bisa membantu dalam membentuk kandidat kombinasi *item*. Jika sudah memenuhi parameter *support* dan *confidence* maka hasilnya bisa membantu dalam penentuan tata letak obat berdasarkan obat yang sering dibeli konsumen dan pola pembelian obat.
6. Penelitian yang dilakukan oleh (Zahra, Mazouri, Abounaima, & Zenkouar, 2019) dengan judul **“*Data mining combined to the multicriteria decision analysis for the improvement of road safety: case of France*”** *Data mining*

*“DM”, as a part of the general KDD “Knowledge Discovery in Databases” process is an attractive set of methods and techniques that has known a big popularity in recent years in various fields such as the fields of finance , education and healthcare.*

## 2.6 Kerangka Pemikiran

Untuk melakukan penelitian penjualan suku cadang motor ini, dibutuhkan kerangka pemikiran yan dapat digunakan untuk menjelaskan bagaimana cara metode ini dapat memecahkan permasalahan yang ada. Kerangka pemikiran tersebut digambarkan dengan diagram dibawah ini:



**Gambar 2.2**Kerangka Pemikiran

(Sumber: Data olahan peneliti, 2019)

Pada gambar 2.1 dapat dijelaskan bahwa didalam kerangka pemikiran ini memiliki variabel input, pemroses, serta variabel *output*. Dimana variabel input terdiri atas jenis suku cadang motor apa saja yang tersedia serta jumlah transaksi pada bengkel tersebut. Setelah itu variabel input diproses dengan metode apriori,

kemudian diuji dengan WEKA, agar menemukan barang apa saja yang banyak terjual. Dan pada *output* akan mendapatkan hasil *item set* yang banyak terjual dari keterkaitan *rule* atau pola.

## **2.7 Hipotesis**

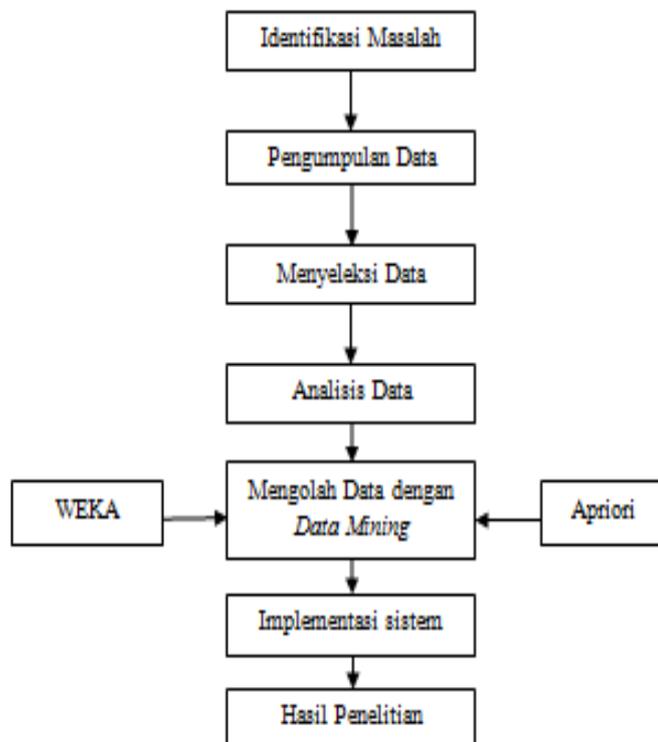
Dengan menggunakan data mining dengan algoritma apriori diharapkan dapat memberikan hasil yang tepat terhadap penjualan suku cadang motor bengkel Candra Motor dari transaksi pembelian barang.

### BAB III

#### METODE PENELITIAN

#### 3.1 Desain Penelitian (Arsitektur Sistem)

Dalam melakukan suatu penelitian, sangat diperlukan suatu perencanaan ataupun perancangan penelitian, agar penelitian tersebut dapat berjalan dengan baik dan sistematis. Suatu perencanaan ataupun perancangan penelitian tentang *data mining* ini dipresentasikan di dalam sebuah desain penelitian yang dijelaskan melalui gambar yang ada di bawah ini.



**Gambar 3.1** Desain Penelitian  
(Sumber : Data Penelitian, 2019)

Tahapan-tahapan penelitian sebagai berikut :

1. Identifikasi Masalah

Suatu kegiatan meneliti ataupun mencari sebuah permasalahan yang dapat diteliti. Di dalam penelitian ini, permasalahan yang terjadi banyaknya data yang tidak dapat diolah pada bengkel Candra Motor sehingga data tersebut menjadi tumpukan kertas sehingga tidak dipergunakan dengan benar. Untuk itu digunakan *data mining* metode apriori agar mempermudah dalam hal pengolahan data.

2. Pengumpulan Data

Setelah identifikasi masalah selesai dilakukan, maka selanjutnya peneliti mengumpulkan data yang berkaitan dengan menggunakan metode wawancara dan observasi secara langsung, agar mendapatkan data penjualan suku cadang motor yang akan dijadikan sumber data penelitian.

3. Menyeleksi Data

Setelah data didapatkan, kemudian peneliti melakukan pemilihan ataupun penyeleksian data. Hal ini dilakukan agar dapat menghindari terjadinya kesamaan data yang akan diolah secara berulang.

4. Analisis Data

Setelah data selesai diseleksi, kemudian data dianalisis kembali, agar data yang diperoleh lebih akurat.

5. Mengolah Data dengan *data mining*

Data-data yang sudah dianalisis kemudian diolah menggunakan *data mining* sehingga mendapatkan *rule assosiasi*.

## 6. Implementasi Sistem

Pengujian data yang dilakukan dengan aplikasi WEKA, untuk memperoleh kecocokan data yang diolah secara manual dengan sistem.

## 7. Hasil Penelitian

Dari semua proses yang sudah dilakukan maka diperoleh suatu hasil dari data penelitian tersebut berupa *item set* yang paling banyak terjual. Sehingga terbentuk *rule assosiasi* dari penjualan suku cadang motor.

### 3.2 Teknik Pengumpulan Data

Pengumpulan data merupakan suatu metode yang digunakan untuk memperoleh data yang dibutuhkan pada suatu penelitian dari narasumber yang bersangkutan. Adapun teknik pengumpulan data yang digunakan untuk penelitian ini adalah :

#### 3.2.1 Pengamatan (*Observasi*)

Pengamatan (*Observasi*) merupakan suatu kegiatan pengamatan ataupun pencatatan secara langsung. Peneliti melakukan observasi secara langsung ke bengkel Candra Motor yang bersangkutan agar memperoleh data suku cadang yang diperlukan.

#### 3.2.2 Wawancara (*Interview*)

Wawancara (*Interview*) merupakan kegiatan tanya-jawab antara peneliti dan narasumber, dimana peneliti akan mengajukan beberapa pertanyaan seputar

informasi yang bersangkutan dengan penelitian yang dilakukan agar memperoleh informasi yang dapat digunakan sebagai sumber data penelitian. Peneliti melakukan wawancara dengan seorang pemilik bengkel Candra Motor bernama ibu wati.

### **3.2.3 Kajian Pustaka**

Untuk memperoleh sumber data yang akurat, peneliti melakukan kajian pustaka pendukung penelitian ini berupa buku dan jurnal tentang *data mining* yang berasal dari berbagai sumber.

### **3.3 Operasional Variabel**

Menurut (Sugiyono, 2014) definisi operasional adalah penentuan data atau sifat yang akan dipelajari sehingga menjadi variabel yang dapat diukur. Definisi operasional menjelaskan cara tertentu yang digunakan untuk meneliti dan mengoperasikan konstruk, sehingga memungkinkan bagi peneliti yang lain untuk melakukan replikasi pengukuran dengan cara yang sama atau mengembangkan cara pengukuran konstruk yang lebih baik. Operasional variabel yang digunakan adalah data penjualan suku cadang motor di bengkel Candra Motor. Adapun variabel-variabel yang diambil sebagai sumber data adalah sebagai berikut :

### 3.3.1 Jumlah Transaksi

Jumlah transaksi yaitu banyaknya jumlah penjualan stok suku cadang motor dalam kurun waktu perbulan selama setahun terakhir. Dari data jumlah transaksi ini dapat menentukan *item set* yang paling banyak terjual.

**Tabel 3.1** Data Transaksi Suku Cadang Motor di Bengkel CandraMotor

Transaksi	Nama <i>Item</i>	Bulan Transaksi											
		Juni	Juli	Agust	Sept	Okt	Nov	Des	Jan	Feb	Mar	Apr	Mei
1	Ban Dalam	35	22	27	17	32	23	30	22	30	16	11	5
2	Kampas Kramol	3	9	6	4	8	3	3	5	8	6	1	0
3	Oli Mestran	12	5	8	6	11	9	8	4	8	6	1	2
4	Kampas	6	9	11	3	12	11	8	2	0	4	0	3
5	Kain Jok	7	5	8	8	7	3	0	0	1	2	1	0
6	Oli Gerbok	21	11	13	9	12	11	19	9	12	7	4	1
7	Batu Arang Stater	2	0	5	3	0	0	1	0	0	2	0	0
8	Oli Castrol	8	3	8	6	1	3	3	4	2	2	3	0
9	Oli MPX	33	28	26	24	34	30	32	22	24	22	16	8
10	Kampas Cakram	8	4	8	1	7	4	9	5	6	5	3	0
11	Bola Lampu	9	1	5	4	7	9	7	1	4	5	5	3
12	Oli Yamalub	38	25	31	20	46	31	37	23	37	28	12	2
13	Busi	16	9	4	5	17	10	12	3	9	11	12	1
14	Oli Ultratec	3	6	4	3	2	5	3	3	3	5	1	2
15	Aki Air	1	0	0	1	4	1	3	1	0	2	0	0
16	Sel Sock	0	2	0	1	0	0	0	1	2	1	0	1
17	Rantai	4	1	0	2	8	2	7	1	3	1	1	3
18	Oli SGO	2	7	4	2	8	4	1	4	5	5	0	0
19	Oli TOP1	2	1	1	1	0	1	2	0	1	1	1	0

(Sumber: Data Penelitian 2019)

### **3.3.2 Nama Suku Cadang Motor**

Nama suku cadang motor merupakan nama-nama produk ataupun barang yang tersedia di bengkel Candra Motor dan produk atau barang mana saja yang paling banyak terjual. Adapun nama-nama produk atau barang yang tersedia adalah oli mesin, oli samping, ban dalam, ban luar, kampas rem belakang, dll.

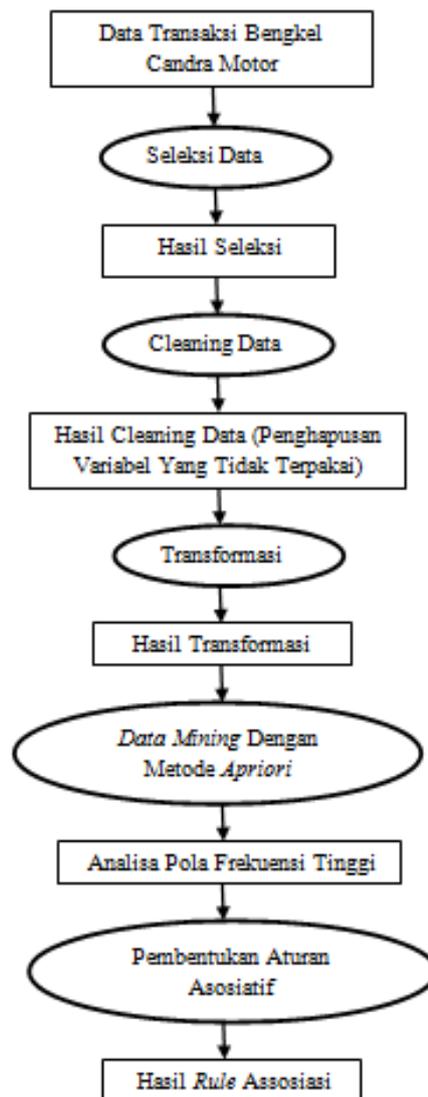
## **3.4 Metode Analisis dan Perancangan Sistem**

### **3.4.1 Metode Analisis**

Metode analisis merupakan suatu hal yang terpenting dalam melakukan sebuah penelitian. Jika metode sebuah penelitian tidak mempunyai metode analisis, maka keakuratan hasil dari sebuah penelitian tersebut masih diragukan. Metode analisis sangat mempengaruhi hasil dari sebuah penelitian tersebut. Sehingga pada penelitian ini menggunakan *data mining* dengan metode *apriori*.

### 3.4.2 Perancangan Sistem

Adapun metode perancangan sistem yang digunakan dapat dijelaskan pada gambar yang di bawah ini.



**Gambar 3.2** Perancangan Sistem

( Sumber : Data Penelitian, 2019)

Dilihat dari gambar 3.2 dapat dijelaskan bahwa semua proses penelitian dimulai dari mengumpulkan data penjualan pada bengkel Candra Motor dengan metode observasi dan wawancara. Setelah data didapatkan, selanjutnya peneliti melakukan seleksi data untuk memilih data yang layak untuk dijadikan sampel penelitian. Dari data penjualan perhari, perbulan, pertahun diperoleh hasil penyeleksian data yang akan dipakai. Data yang sudah diseleksi kemudian akan diolah menggunakan algoritma apriori dan kemudian akan diperoleh hasil pengolahan data menggunakan aplikasi WEKA.

### 3.5 Lokasi dan Jadwal Penelitian

Dalam melakukan sebuah penelitian yang bertujuan untuk mengolah data penjualan pada bengkel Candra Motor dengan menerapkan *data mining* algoritma *apriori*, maka diperlukan tempat dan jadwal penelitian. Lokasi dan jadwal penelitian dilaksanakan adalah sebagai berikut :

#### 3.5.1 Lokasi Penelitian

Penelitian ini dilakukan di bengkel Candra Motor Putri Hijau. Alamat Ruko Permata Rhabayu 2 blok.A No.17 Batu Aji, Batam.



**Gambar 3.3** Denah Lokasi  
(Sumber: Data Penelitian)

### 3.5.2 Jadwal Penelitian

Penelitian ini dilakukan dari bulan April 2019 sampai Agustus 2019.

Adapun jadwal penelitiannya dapat dilihat pada tabel dibawah ini:

**Tabel 3.2** Jadwal Penelitian

No.	Kegiatan	Waktu Pelaksanaan Penelitian																			
		April 2019				Mei 2019				Juni 2019				Juli 2019				Agustus 2019			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengumpulan data	■	■																		
2	BAB I			■	■																
3	BAB II					■	■														
4	BAB III							■	■	■	■										
5	BAB IV									■	■	■	■	■	■						
6	BAB V													■	■						
7	Pengumpulan Skripsi																	■	■	■	■

Sumber : Data Penelitian, (2019)