

**PREDIKSI PENJUALAN TIKET MENGGUNAKAN  
METODE APRIORI PADA CV PATRIA MITRA  
SEJATI**

**SKRIPSI**



**Oleh:  
Ferry Dwi Nur Syaputro  
150210068**

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

**PREDIKSI PENJUALAN TIKET MENGGUNAKAN  
METODE APRIORI PADA CV PATRIA MITRA  
SEJATI**

**SKRIPSI**

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



**Oleh:  
Ferry Dwi Nur Syaputro  
150210068**

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

## **SURAT PERNYATAAN**

Yang bertanda tangan di bawah ini saya:

Nama : Ferry Dwi Nur Syaputro

NPM : 150210068

Fakultas : Teknik dan Komputer

Program Studi : Teknik Informatika

Menyatakan bahwa "skripsi" yang saya buat dengan judul:

### **PREDIKSI PENJUALAN TIKET MENGGUNAKAN METODE APRIORI PADA CV PATRIA MITRA SEJATI**

Adalah hasil karya sendiri dan bukan "duplikasi" darikarya orang lain. Sepengetahuan saya, didalam naskah skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah skripsi ini digugurkan dan skripsi yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan inisaya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 17 Februari 2020

**Ferry Dwi Nur Syaputro**  
150210068

**PREDIKSI PENJUALAN TIKET MENGGUNAKAN  
METODE APRIORI PADA CV PATRIA MITRA  
SEJATI**

**Oleh: Ferry Dwi Nur Syaputro  
150210068**

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

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

**Batam, 17 Februari 2020**

**Koko Handoko, S.Kom., M.Kom.  
Pembimbing**

## ABSTRAK

Dengan perkembangan teknologi dalam era yang sekarang ini, teknologi sangatlah berkembang dengan pesat dan penting apalagi jika dilihat dari sisi kehidupan, dan terutama dari segi perekonomian. Penelitian yang saat ini dilakukan guna untuk membantu atau mempermudah manajemen CV Patria Mitra Sejati dalam hal pengolahan data-data atau file-file penjualan tiket dengan penerapan *data mining* pada penjualan tiket dan untuk menentukan penjualan tiket pesawat yang paling sering terjual berdasarkan dari jenis maskapai yang tersedia di CV PMS. Dalam tahap penyeleksian data digunakan sebuah metode *apriori* dengan mengambil sebuah data dengan aturan asosiasi untuk menentukan sebuah keterkaitan dari suatu kombinasi item. Dimana perhitungan ini dilakukan dengan cara manual dan juga menggunakan *software* WEKA sehingga diperoleh data yang sama. Dengan memenuhi syarat minimal *Support* dan minimum *Confidence* yang telah ditentukan sebagai berikut tahun 2015 yang paling banyak terjual *Citilink* dan *Lion Air* dengan ketentuan *Support* 75% dan *Confidence* 90%, tahun 2016 yang paling banyak terjual *Lion Air* dengan ketentuan *Support* 50% dan *Confidence* 90%, tahun 2017 yang paling banyak terjual *Lion Air* dengan ketentuan *Support* 85% dan *Confidence* 90%, tahun 2018 yang paling banyak terjual *Lion Air*, *Citilink Air*, *Garuda Air*, dan *Sriwijaya Air* dengan ketentuan *Support* 85% dan *Confidence* 90%.

**Kata Kunci:** *Apriori*, *Data Mining*, Penjualan Tiket, WEKA

## **ABSTRACT**

*With the development of technology in the current era, technology is developing rapidly and is important especially when viewed from the side of life, and especially in terms of the economy. Research that is currently conducted in order to help or simplify the management of CV Patria Mitra Sejati in terms of managing data or ticket sales files by applying data mining to ticket sales and to determine the sale of airline tickets most often sold based on the type of airline available on the PMS CV. In the data selection stage, an apriori method is used by taking data with association rules to determine an association of a combination of items. Where this calculation is done manually and also using WEKA software so that the same data is obtained. By fulfilling the minimum Support and Confidence minimum requirements that have been determined as follows in 2015 the most sold Citilink and Lion Air with the terms of Support 75% and Confidence 90%, in 2016 the most sold Lion Air with the terms of Support 50% and Confidence 90%, in 2017 the most sold Lion Air with the provisions of 85% Support and 90% Confidence, in 2018 the most sold Lion Air, Citilink Air, Garuda Air, and Sriwijaya Air with the provisions of Support 85% and Confidence 90%.*

*Keywords: Apriori, Data Mining, Ticket Sales, WEKA*

## 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 kata 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. Bapak Koko Handoko, 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 Jainuri dan Ibu Ely Julaikah yang selalu mendoakan, mendukung baik dari segi moral maupun materil kepada penulis sehingga dapat menyelesaikan penulisan skripsi ini.

6. Kepada sahabat terbaik penulis Nike Winda, Fikardi, Sapar, Zabur, dan juga orang sepesial Ainul Mardiah yang telah menemani dan selalu memberikan semangat kepada penulis dalam memberi masukan.
7. Teman seperjuangan Program Studi Teknik Informatika Universitas Putera Batam.

Semoga Allah SWT membalas kebaikan dan selalu mencurahkan hidayah serta taufikNya, Aaamiin.

Batam, 17 Februari 2020

Ferry Dwi Nur Syaputro

## DAFTAR ISI

	Halaman
<b>HALAM SAMPUL</b>	
<b>HALAMAN JUDUL</b> .....	ii
<b>SURAT PERNYATAAN</b> .....	iii
<b>HALAMAN PENGESAHAN</b> .....	iv
<b>ABSTRAK</b> .....	v
<b>ABSTRACT</b> .....	vi
<b>KATA PENGANTAR</b> .....	vii
<b>DAFTAR ISI</b> .....	ix
<b>DAFTAR GAMBAR</b> .....	xi
<b>DAFTAR TABEL</b> .....	xii
<b>DAFTAR RUMUS</b> .....	xiii
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang .....	1
1.2 Identifikasi Masalah .....	3
1.3 Batasan Masalah .....	4
1.4 Rumusan Masalah .....	4
1.5 Tujuan Penelitian .....	5
1.6 Manfaat Penelitian .....	5
<b>BAB II TINJAUAN PUSTAKA</b>	
2.1 Teori .....	7
2.2 <i>Knowledge Discovery in Database (KDD)</i> .....	7
2.3 <i>Data Mining</i> .....	11
2.4 <i>Metode Data Mining</i> .....	19
2.5 <i>Software Pendukung</i> .....	23
2.6 Penelitian Terdahulu .....	24
2.7 Objek Penelitian .....	28
2.8 Kerangka Pemikiran .....	31
<b>BAB III METODE PENELITIAN</b>	

3.1	Arsitektur Sistem.....	32
3.2	Teknik Pengumpulan Data.....	34
3.3	Oprasional Variabel .....	35
3.4	Metode Analisis .....	37
3.5	Rancangan System .....	38
3.6	Lokasi Dan Jadwal Penelitian .....	39
<b>BAB IV HASIL DAN PEMBAHASAN</b>		
4.1	Analisa Data .....	41
4.2	Hasil Pengujian .....	45
<b>BAB V KESIMPULAN DAN SARAN</b>		
5.1	Kesimpulan .....	74
5.2	Saran.....	75
<b>LAMPIRAN</b>		
	Lampiran 1. Pendukung Penelitian .....	78
	Lampiran 2. Daftar Riwayat Hidup.....	81
	Lampiran 3. Surat Keterangan Penelitian .....	83

## DAFTAR GAMBAR

Halaman

Gambar 2. 1 KDD .....	9
Gambar 2. 2 Kerangka Pemikiran .....	31
Gambar 3. 1 Arsitektur Penelitian .....	32
Gambar 3. 2 Report Transaksi 2015-2018 .....	36
Gambar 3. 3 Perancangan system .....	38
Gambar 3. 4 Lokasi Penelitian .....	39
Gambar 4. 1 Pengolahan data .....	67
Gambar 4. 2 Pengolahan data .....	67
Gambar 4. 3 Penyimpanan data .....	68
Gambar 4. 4 Tampilan WEKA Tahap Awal .....	68
Gambar 4. 5 Proses WEKA ke-2 .....	69
Gambar 4. 6 Pengambilan Data .....	69
Gambar 4. 7 Pengambilan Data .....	70
Gambar 4. 8 Penyeleksian Data .....	70
Gambar 4. 9 Pengolahan Data .....	71
Gambar 4. 10 Hasil Pengolahan .....	72

## DAFTAR TABEL

Halaman

Table 3. 1 Jadwal Penelitian .....	40
Table 4. 1 Daftar <i>Items</i> .....	41
Table 4. 2 Data Awal Penjualan.....	42
Table 4. 3 Kode Transaksi .....	43
Table 4. 4 Integrasi data 2015 – 2018 .....	45
Table 4. 5 <i>Frequent</i> 1 .....	47
Table 4. 6 <i>Frequent C-1</i> .....	49
Table 4. 7 <i>Frequent</i> F-1 .....	51
Table 4. 8 <i>Frequent C-2</i> .....	53
Table 4. 9 <i>Frequent</i> F-2 .....	56
Table 4. 10 <i>frequent</i> F-2.....	57
Table 4. 11 Assosiasi F-2 .....	59
Table 4. 12 <i>Frequent</i> C-3 .....	61
Table 4. 13 <i>Frequent</i> F-3 .....	62
Table 4. 14 Assosiasi F-3 .....	63
Table 4. 15 Assosiasi Final .....	65

## DAFTAR RUMUS

	Halaman
Rumus 2. 1 Nilai Penunjang .....	21
Rumus 2. 2 Nilai Penunjang 2 item .....	21
Rumus 2. 3 Pembentukan Aturan <i>association</i> .....	21

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Seiring dengan berkembangnya teknologi informasi dan komunikasi pada era yang sering disebut milenial ini TIK juga semakin maju dan berkembang dan tidak dapat dihindari. Dimana perkembangan dan kemajuan teknologi informasi dan komunikasi berkembang dengan sangat cepat, seperti kebutuhan akan pengolahan data yang setiap hari semakin bertambah banyak dan jika dibiarkan maka data tersebut tidak akan berguna. CV Patria Mitra Sejati atau biasa disebut dengan PMS merupakan salah satu CV yang bergerak dibidang ticketing yang terletak dikota batam. CV Patria Mitra Sejati sehari-harinya melayani segala pemesanan tiket baik domestic maupun internasional. CV Patria Mitra Sejati memiliki beberapa akses penjualan tiket pada setiap maskapai didalamnya seperti Maskapai *Lion Air*, Maskapai *Citilink*, Maskapai *Garuda*, Maskapai *Trigana*, Maskapai *Express Air*, Maskapai *Sriwijaya Air*, Maskapai *Airasia*.

Peningkatan penjualan yang signifikan di CV Patria Mitra Sejati selalu menjadi pembahasan pokok yang sering dibahas pimpinan perusahaan untuk selalu meningkatkan target penjualan yang akan datang atau yang biasa disebut dengan peramalan (*foreccasting*). Misalnya, penjual dapat dengan mudah

memprediksi berapa banyak dan jumlah penjualan pada musim lebaran pada tahun yang akan mendatang. Dalam proses analisis guna untuk meningkatkan penjualan tiket belum adanya perhitung dengan menggunakan *data mining* (masih menggunakan cara manual) yang dapat mengakibatkan proses kinerja menjadi terhambat dan waktu akan terbuang secara Cuma-cuma. CV Patria Mitra Sejati memiliki data yang sangat banyak, sehingga memerlukan konsep pengolahan data yang lebih bagus guna untuk kemajuan dan perkembangan transaksi.

Dengan adanya permasalahan seperti itu, tentu saja dengan menggunakan pengolahan yang disebut *data mining* akan jauh lebih bagus dan akan mempermudah dalam memprediksi dan mengolah penjualan tiket, karena *data mining* dapat menampung data yang cukup banyak. *Data mining* merupakan bagian-bagian dari teknologi informasi dan komunikasi yang pantas untuk digunakan dalam hal meramal kejadian- kejadian yang akan terjadi kedepannya.

Suatu proses atau tata cara penambangan data didalam sebuah *database* yang terindex guna untuk menggali sebuah informasi-informasi yang terdapat didalam *data mining* itu sendiri (Ristianingrum & Sulastri, 2017). Dengan menggunakan method yang tepat dan benar *data mining* dapat dipergunakan untuk meramal kejadian yang akan datang. Metode association rule apriori dapat dipergunakan untuk menentukan pola data dan juga bisa dipergunakan untuk manajemen prediksi. Metode apriori merupakan sebuah *association rule* untuk menentukan suatu atribut yang telah diketahui sebelumnya untuk mengeksekusi atribut selanjutnya dengan memperhatikan *support* minimum sebuah kandidat

yang kemungkinan akan muncul, minimum *confidence* dengan tujuan untuk menentukan dan menemukan *frequent itemset* (Manalu, 2017).

Dengan mempergunakan teknik *data mining* dan menggunakan metode apriori data yang besar atau banyak akan dengan mudah untuk diolah untuk menghasilkan informasi-informasi yang penting. Tanpa mempergunakan teknik *data mining*, tentu saja akan sangat sulit untuk mendapatkan informasi-informasi penjualan. Kesulitan dalam mendapatkan informasi disebuah data yang besar dengan tidak menerapkan *data mining* tentu saja akan banyak membuang waktu secara Cuma-Cuma. Dalam era global ini efisiensi waktu sangat penting karena dapat dipergunakan untuk keperluan-keperluan yang lain. Sehingga kualitas dan kuantitas akan lebih baik.

## 1.2 Identifikasi Masalah

Berdasarkan dari uraian yang dipaparkan di atas dapat diidentifikasi masalah oleh penulis sebagai berikut:

1. Belum adanya perhitungan dengan menggunakan *data mining* pada CV Patria Mitra Sejati sehingga proses menjadi lebih lama dan sulit.
2. Belum adanya penerapan *data mining* sehingga dapat mengakibatkan pemborosan waktu.
3. Banyaknya jumlah data yang akan diolah dan akan sulit jika tidak menerapkan *data mining*.

### 1.3 Batasan Masalah

Ada beberapa hal yang menjadi pembatasan masalah agar penelitian ini tidak menyimpang dan dapat pada target. Batasan masalah di dalam penelitian ini adalah:

1. Penelitian ini dilakukan di CV Patria Mitra Sejati yang berlokasi di kota batam.
2. Data diperoleh dari histori transaksi atau *report* sebagai sumber data dari tahun 2015-2018.
3. Aplikasi yang akan dipergunakan dalam penelitian ini adalah WEKA.

### 1.4 Rumusan Masalah

Berdasarkan dari analisa penulis, dapat dirumuskan permasalahan yang akan diselesaikan peneliti sebagai berikut:

1. Bagaimana cara menganalisa data penjualan tiket dengan *data mining* menggunakan algoritma apriori?
2. Bagaimana cara mendapatkan persentasi penjualan tiket yang paling banyak terjual?
3. Bagaimana cara mengimplementasi *data mining* menggunakan *software* WEKA?

## 1.5 Tujuan Penelitian

Adapun tujuan penelitian berdasarkan perumusan masalah yang dipaparkan di atas adalah:

1. Untuk menganalisa data penjualan tiket di CV Patria Mitra Sejati yang bertujuan untuk membentuk pola dengan menggunakan algoritma apriori.
2. Untuk mendapatkan persentase penjualan tiket yang paling banyak terjual yang bertujuan mempermudah dalam memprediksi penjualan tiket.
3. Untuk mengimplementasikan *data mining* yang sebelumnya masih menggunakan perhitungan manual dan belum menggunakan *software* WEKA.

## 1.6 Manfaat Penelitian

Dengan tercapainya tujuan dari penelitian, diharapkan penelitian ini dapat bermanfaat bagi berbagai pihak. Manfaat penelitian ini adalah:

### 1.6.1 Manfaat Teoritis

1. Menambah pengetahuan tentang *data mining* dengan menggunakan algoritma *apriori*.
2. Dapat dipergunakan sebagai bahan referensi pengolahan data dengan *data mining*.

### **1.6.2 Manfaat Praktis**

1. Penelitian ini dapat dipergunakan sebagai acuan dalam mengolah data penjualan.
2. Memudahkan dalam hal mengetahui tiket yang paling banyak terjual dalam CV Patria Mitra Sejati.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teori**

Sebuah prinsip dasar suatu pembentukan dari sebuah ilmu pengetahuan. Dimana dasar dari teori bisa dikembangkan atau diciptakan dan menghasilkan sebuah pengetahuan yang baru yang lebih detail dan bisa memperkuat suatu pengetahuan tersebut. Pengekstrakan sebuah data untuk mendapatkan informasi dari data mining yang baru didapatkan dari sebuah bongkahan data yang besar membantu dalam mengambil keputusan. Memining data atau yang biasa disebut-sebut *Knowledge discovery* merupakan suatu teknik yang diciptakan guna untuk mendapatkan sebuah informasi dengan cara menelusuri dari data yang lalu atau transaksi yang sebelum-sebelumnya untuk membentuk suatu model, yang kemudian dipergunakan untuk mendapatkan pola suatu data.

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

Ditafsirkan sebagai informasi yang mungkin diekstraksi, implisit maka tidak dikenal dari set data yang berbeda. Metode knowledge discovery menghubungkan hasil dari metode data mining (metode yang cenderung mengekstraksi pola data), setelah itu hasilnya dimodifikasi secara tepat sehingga menjadi mudah untuk memahami informasi. KDD sendiri dapat di definisikan sebagai integritas metode non-sepele yang berfungsi untuk melacak dan mengenali pola (*pattern*) dalam

data, di mana pola yang diperoleh dalam bentuk pola baru dapat valid, dapat dipahami dan bermanfaat (Informatika et al., 2015).

Ada beberapa jenis pendekatan yang dikategorikan sebagai mencari informasi teknologi di KDD. Ada pendekatan kuantitatif, probabilistik seperti pendekatan logika induktif, mencari pola, dan menganalisa pohon keputusan. Pendekatan lain termasuk kelainan, analisis tren, algoritma genetika, JST (jaringan syaraf tiruan), dan kombinasi dari dua pendekatan atau lebih.

Singkatnya, ketika mencari informasi atau wawasan tentang KDD, ada enam faktor penting menurut (Tampubolon et al., 2013) yaitu:

1. Bekerja dengan data dalam jumlah besar.
2. Efisiensi diperlukan dalam volume data.
3. Penugasan atau prioritas akurasi.
4. Membutuhkan bahasa tingkat tinggi.
5. Menggunakan berbagai persyaratan automata untuk formulir pembelajaran.
6. Memberikan hasil yang baik.

Dalam tahapan KDD mencakup tiga proses utama menurut (Gunadi & Sensuse, 2012) sebagai berikut:

1. *Preprocessing*

*Preprocessing* terjadi pada data sebelum digunakanya algoritma *Data Mining*. Proses ini mencakup pembersihan data, seleksi, dan transformasi.

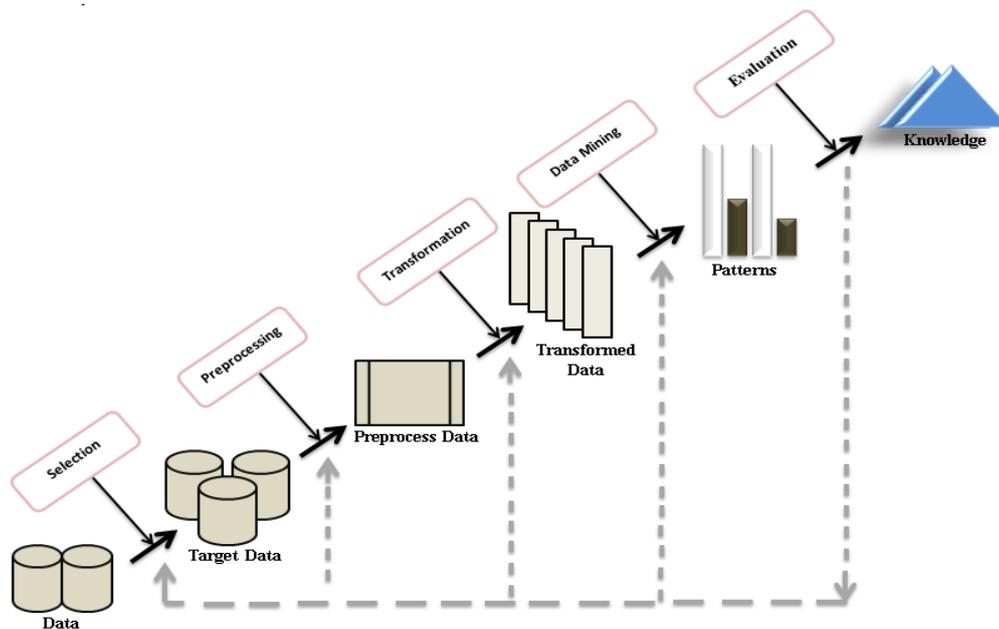
## 2. *Data Mining*

Merupakan proses yang paling penting dalam KDD penambangan data, yang menggunakan algoritma penambangan data untuk mengekstraksi informasi dari *database*.

## 3. *Post processing*

Merupakan hasil yang diperoleh dari proses *mining data* yang akan dievaluasi dalam post-processing.

Untuk menggambarkan proses penggalian informasi yang tersembunyi di dalam basis data yang besar sering kali digunakan secara bergantian. merupakan istilah dari data mining dan knowledge discovery in database (KDD). Padahal kedua istilah tersebut memiliki konsep berbeda tetapi saling terkait. Dan salah satu tahapan dalam seluruh proses KDD adalah data mining. Proses KDD dapat dijelaskan secara umum sebagai berikut (Informatika et al., 2015):



**Gambar 2. 1 KDD**  
(Sumber: Data Penelitian)

1. *Data selection*

Pemilihan (selection) data dari pengumpulan data operasional harus dilakukan sebelum fase penggalian informasi ke KDD. Data yang dipilih untuk digunakan dalam proses mining data disimpan dalam file terpisah dari database operasional.

2. *Pre-processing/ Cleaning*

Sebelum menambang data dapat dilakukan, data yang merupakan focus dari KDD perlu dibersihkan. Proses pembersihan melibatkan menghapus data ganda/ duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan, seperti kesalahan ketik. Pengayaan juga dilakukan, yaitu proses “memperkaya” data atau data yang ada dengan informasi lain tentang KDD, seperti data atau informasi eksternal.

3. *Transformation*

Proses mengubah data yang dipilih agar sesuai dengan proses penambangan data. Proses pengkodean di KDD adalah proses kreatif dan sangat tergantung pada jenis atau model informasi yang akan dicari dalam database.

4. *Data mining*

Data mining adalah suatu proses menemukan pola atau informasi menarik dalam data yang dipilih menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam menambang data sangat bervariasi. Pilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses KDD secara keseluruhan.

## 5. *Interpretasi/ Evaluasi*

Model informasi yang diperoleh dari proses penambangan data harus ditampilkan dengan cara yang mudah dipahami oleh yang bersangkutan. Fase ini adalah bagian dari proses KDD yang disebut interpretasi. Langkah ini juga termasuk memeriksa apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang sudah ada sebelumnya.

### 2.3 *Data Mining*

suatu metode yang menggunakan cara matematik, kecerdasan buatan, statistic, serta machine learning yang digunakan untuk mengidentifikasi sebuah informasi yang dapat bermanfaat dan pengetahuan didalam *database* ataupun data *warehouse*. Sebenarnya dalam *datamining* memiliki akar yang panjang di berbagai bidang seperti *machine learning*, *artificial intelligent*(kecerdasan buatan), *statistic* dan *basis data*. Serta beberapa metode yang sering dipergunakan dalam *data mining* adalah: *classification*, *association rule*, *clustering*, *neural network*, dan *algorithm genetic*(Handoko & Lesmana, 2018).

Merupakan kegiatan yang digunakan untuk menemukan susunan pemakaian dari *report data*, pengumpulan data, pola dan keterkaitan dalam data set yang besar. Penambangan data juga sering diartikan sebagai proses untuk menyelidiki atau menggali sebuah pengetahuan yang tidak diketahui dalam suatu kumpulan data secara manual. Dan tahapan untuk menemukan suatu pola atau informasi dalam satu set data dengan menggunakan teknik dan algoritma tertentu (Kamal, Iman Mustofa Hendro P, Tachbir Ilyas, 2017).

Hal-hal yang berkaitan dengan data mining, sebagaimana telah disampaikan menurut definisi-definisinya adalah:

1. Ekstraksi data adalah proses otomatis dari data yang telah ada.
2. Mengelola data dalam bentuk data yang sangat besar. Sehingga membuat hasil lebih dipercayai.
3. Tujuan dari data mining adalah untuk memperoleh tautan atau model yang dapat memberikan indikasi yang dapat bermanfaat.

Hasil dari ekstraksi data dapat digunakan untuk meningkatkan pengambilan keputusan dimasa depan(Tampubolon et al., 2013).

Data mining merupakan serangkaian proses yang akan menghasilkan pengetahuan baru yang diharapkan dapat berguna dan digunakan di masa depan.

Langkah-langkah dalam data mining dapat digambarkan sebagai berikut(Handoko & Lesmana, 2018):

1. Menghilangkan data dari noise dan yang tidak konsisten.
2. Menggabungkan data dari sumber-sumber yang berbeda.
3. Data dikonversi menjadi bentuk yang cocok untuk di ekstraksi.
4. Penerapan teknik data mining.
5. Mengevaluasi model yang ditemukan (untuk menemukan yang menarik/berharga).
6. Menyampaikan pengetahuan yang didapat dengan visual.

### 2.3.1 Manfaat Data Mining

Untuk manfaat dari data mining bisa dilihat dari dua perspektif, yaitu dari pandangan komersial dan keilmuan (Vulandari, 2017).

1. Jika dilihat dari sudut pandang komersial, eksploitasi data dapat digunakan dari sudut pandang komersial untuk menangani ledakan dari volume data, dengan dipergunakan teknik komputasi dapat diperuntukkan untuk menghasilkan sebuah informasi sebagai asset yang dapat meningkatkan daya saing suatu institusi.

Contohnya:

- a. Bagaimana cara melacak hilangnya pelanggan karena persaingan.
  - b. Bagaimana cara mengetahui *item* produk atau pengguna mana yang memiliki karakteristik yang sama.
  - 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. Jika dilihat dari sudut pandang keilmuan, data mining dapat diperuntukkan untuk menangkap, analisis, dan menyimpan data *real time* dan sangat besar, misalnya:
    - a. *Remote sensore* yang ditempatkan pada suatu satelit.
    - b. *Telescope* yang digunakan untuk memindai langit.

- c. Simulasi saintifik yang membangkitkan data dalam ukuran *terrabbytes*.

### 2.3.2 Fungsi dan Kegunaan *Data Mining*

Jika dilihat dari fungsi-fungsi *Data mining* sendiri yang secara umum diterapkan adalah (Vulandari, 2017):

1. *Association*, merupakan suatu proses untuk mendapatkan suatu aturan asosiasi antara kombinasi elemen sekaligus.
2. *Sequence*, suatu proses untuk memperoleh aturan untuk asosiasi antara kombinasi *item* sekaligus dan berlaku untuk lebih dari satu periode.
3. *Clustering*, suatu proses pengelompokan beberapa data/ obyek kedalam grup data sehingga setiap grup berisi data yang serupa.
4. *Classification*, proses untuk menemukan suatu model atau fungsi yang menjelaskan atau membedakan suatu konsep atau kelas data, yang bertujuan untuk dapat memprediksi kelas dari suatu obyek yang labelnya belum diketahui.
5. *Regression*, proses pemetaan data dalam suatu nilai prediksi.
6. *Forecasting*, merupakan suatu proses memperkirakan nilai pikiran berdasarkan model didalam set data.
7. *Solution*, proses menemukan akar penyebab dan menyelesaikan masalah-masalah bisnis, atau setidaknya sebagai informasi ketika membuat keputusan.

Secara garis besar, kegunaan *Data Mining* dibagi menjadi deskriptif dan prediktif. *Data mining* dipergunakan untuk mencari suatu pola-pola yang dapat

dimengerti dan dipahami oleh manusia yang menjabarkan karakteristik data disebut sebagai Deskriptif. Sedangkan *Data mining* yang digunakan untuk membangun suatu model pengetahuan yang akan diperuntukkan untuk peramalan atau prediksi disebut dengan prediktif (Suyanto, 2017). Berdasarkan fungsionalitasnya, pengelompokan *Data mining* dibagi menjadi enam kelompok sebagai berikut:

1. Klasifikasi (*classification*): men generalisasi struktur yang diketahui untuk di aplikasikan pada data-data baru. Seperti, klasifikasi penyakit kedalam sejumlah jenis, klasifikasi pada email ke dalam spam atau bukan.
2. Klasterisasi (*clusstering*): pengelompokan data, yang belum diketahui label kelasnya, ke dalam beberapa kelompok tertentu sesuai dengan kemiripannya.
3. Regresi (*RegressionI*): menemukan suatu fungsi yang memodelkan data dengan kesalahan prediksi seminimal mungkin.
4. Deteksi anomaly (*anomaly detection*): mengidentifikasi data yang tidak umum, bisa berupa *outlier* (pencilan), perubahan atau deviasi yang bisa sangat diperuntukkan dan perlu di analisis lebih lanjut.
5. Pembelajaran aturan asosiasi: pencarian relasi antar variable.
6. Perangkuman (*summarization*): sebagai penyampaian data yang lebih sederhana, dengan visualisasi atau laporan.

### **2.3.3 Pengelompokan *Data Mining***

Menurut (Gunadi & Sensuse, 2012) berdasarkan pengelompokan *data mining* dibagi berdasarkan tugas atau pekerjaan masing-masing, yaitu:

### 1. Deskripsi

Biasanya para peneliti dan juga analisis secara singkat mencari sebuah cara untuk memvisualkan dan juga keterkaitan sebuah pola didalam data. Pendeskripsian dari suatu pola lebih keterkaitan lebih sering membuat kemungkinan untuk penjelasan kecenderungan sebuah pola.

### 2. Estimasi

Prinsip dari kerja estimasi mirip dengan klasifikasi, terkecuali dari variabel sebuah target estimasi lebih cenderung kearah numeric dari pada kategori. Model didirikan dengan menggunakan sebuah baris *record* data lengkap yang memberikan nilai prediksi. Yang selanjutnya, peninjauan setelah estimasi nilai dari sebuah target variabel dibentuk berpatokan dari nilai prediksi variabel.

### 3. Prediksi

Untuk prinsip kerja prediksi mirip dengan cara kerja klasifikasi dan prinsip estimasi, terkecuali didalam nilai prediksi dari sebuah hasil bisa ada dalam masa yang akan datang.

### 4. Klasifikasi

Dalam klasifikasi, terdapat sebuah variabel target berkategori. Sebagai Contoh, pengelompokan dari pendapatan bisa di pisah menjadi tiga kreteria, yaitu tinggi pendapatan, sedang, dan rendah pendapatan.

### 5. Klaster

Merupakan sebuah pengelompokan sebuah *record*, analisis, atau dengan memperhatikan dan mendirikan sebuah kelas berupa objek-objek yang menyerupai. Kumpulan darisebuah *record* yang menyerupai dari satu dengan

yang lainnya dan tidak menyerupai *record* didalam klas yang lain merupakan definisi dari klaster.

Berbeda dengan klasifikasi, didalam sebuah *class* tidak memiliki variabel target. Dan dalam *clustering* tidak membentuk estimasi, klasifikasi, atau memperkirakan nilai sebuah variabel target. Melaikan, algoritma dari pengklasteran mencoba untuk membagi kesemua data yang bertujuan membentuk suatu kelompok-kelompok yang sejenis atau mirip, dimana kesamaan dari *record* didalam suatu kelompok dapat bernilai maksimal, akan tetapi kesamaan dengan data *record* disebuah kelompok yang lain akan mernilai minimal.

#### 6. Asosiasi

Memperoleh sebuah atribut keluar dalam waktu yang sama merupakan sebuah tugas dari asosiasi. Implementasi dari salah sau asosiasi ialah *market basket analysis* atau sebuah analisis keranjang belanja.

### 2.3.4 Teknik *Data Mining*

Teknik yang dipergunakan didalam data mining berkaitan dengan *discovery* dan *learning* yang dibagi dalam tiga metode pembelajaran utama adalah (Vulandari, 2017):

#### 1. *Supervised Learning*

Merupakan teknik yang paling sering dipergunakan. Teknik ini sama dengan "*programming by example*". Dalam teknik inimengikutkan fase pelatihan dimana karakter-karakter data pelatihan histori dipetakan kedalam sebuah hasil yang sudah diproses dengan *data mining*. Pada proses ini algoritma dilatih untuk

menandai variabel-variabel dan kunci nilai yang akhirnya akan dipergunakan sebagai dasar-dasar pikiran pada saat terdapat data baru.

## 2. *Unsupervised Learning*

Dalam pembelajaran teknik *Unsupervised Learning* tidak melibatkan teknik pelatihan seperti yang terdapat dalam fase *Supervised Learning*. Dimana pada teknik ini pola-pola seperti *sequences* dan *associations* tergantung pada pengguna algoritmanya, yang keluar dari definisi-definisi penting yang menyeluruh dalam data *input*. Dimana dalam pendekatan ini lebih mengarah pada banyaknya aturan pembuatan dengan karakteristik *cluster*, *association*, dan *segment*. Yang kemudian aturan tersebut di analisa untuk mendapatkan informasi yang penting.

## 3. *Reinforcement Learning*

Dimana teknik ini sangat jarang dipergunakan dibandingkan dengan kedua teknik yang lainnya, namun dalam teknik ini terdapat penerapan-penerapan yang selalu dioptimalkan secara terus menerus dan juga memiliki control adaptif. Dimana teknik ini sangat mirip dengan gambaran kehidupan nyata seperti “*on-job-training*”, dimana sekumpulan tugas diberikan kepada pekerja yang memerlukan keputusan-keputusan. Atas *performance* pekerja akan diberikan titik waktu kemudian para pekerja akan diminta untuk mengevaluasi tindakan-tindakan yang sudah dibuat yang berhubungan dengan hasil dari *performance* para pekerja tersebut. Dimana teknik ini sangat tepat jika dipergunakan untuk penyelesaian masalah-masalah yang bergantung dengan waktu yang sulit.

## 2.4 Metode *Data Mining*

Didalam pengolahan *data mining* data besar atau pun kecil, terdapat teknik atau sebuah cara yang dipergunakan guna untuk mempermudah dalam dalam pengolahan data dan biasanya di sebut dengan metode. Metode tersebut diantaranya adalah:

### 2.4.1 Algoritma *Apriori*

Metode *apriori* merupakan sebuah algoritma dari salah satu aturan asosiasi untuk mengekstraksi itemset dari dalam sebuah *database* raksasa yang bertujuan untuk menemukan informasi-informasi dengan menggunakan *data mining*. Dalam hal ini terdapat dua dasaran yang penting yaitu *support* dan keyakinan minimum. Dimana tahapan pertama, memastikan item lebih besar atau sama dengan *support*. Pada tahapan kedua, dimana kepercayaan terhadap kendala dipergunakan untuk membuat *rule* asosiasi.

Analisis asosiasi juga merupakan algoritma apriori dalam teknik *data mining*. Dimana dengan menggunakan teknik *association rule* dalam *data mining* untuk mencari suatu kombinasi itemset. Teknik ini merupakan tahapan analisis *association* yang banyak dipergunakan untuk menghasilkan algoritma yang efisien sehingga banyak menarik para peneliti dengan menggunakan analisis *frequesn pattern mining*. Suatu aturan asosiasi bisa diketahui dengan tolak ukur yaitu: nilai pasti (*confidence*) dan nilai penunjang (*support*). Nilai pasti merupakan kokohnya suatu ikatan antara item dengan aturan asosiasi. sedangkan

nilai penunjang didapatkan dalam persentase dari kombinasi item didalam sebuah database.

*Association* diartikan sebagai suatu proses yang diperuntukkan untuk mencari semua aturan-aturan asosiasi dengan memenuhi syarat minimum nilai penunjang dan juga minimum nilai kepastian. Sebuah algoritma apriori dilakukan proses utama dalam mendapatkan *frequent itemset* yaitu:

1. Penggabungan/ *join*

Suatu proses yang dilakukan dengan mengkombinasikan satu item dengan item yang lainnya sehingga tidak terjadi suatu kombinasi lagi.

2. Pemangkasan/ *prune*

Dimana suatu proses yang hasil dari sebuah item yang telah digabungkan kemudian dilakukan pemangkasan menggunakan minimum dari nilai penunjang (*support*).

Analisis *association* teknik yang dipergunakan untuk mendapatkan sebuah aturan asosiasi dari kombinasi item. Misalnya, dapat prediksi seberapa besar seseorang yang membeli roti dengan susu disuatu mini market. Sehingga pemilik minimarket dapat merancang dan mengolah tata cara penempatan barangnya atau memberi kupon diskon pemasaran untuk kombinasi barang tertentu (Putria, 2018).

Sebuah metode analisis dasar asosiasi dibagi menjadi dua tahapan yaitu:

1. Analisa pola frekuensi tinggi

Didalam tahapan ini mencari sebuah hubungan item yang memenuhi syarat minimal dari nilai penunjang dalam database. Rumus dari nilai penunjang diperoleh sebagai berikut:

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

**Rumus 2. 1** Nilai Penunjang

Sedangkan rumus nilai penunjang dari 2 item diperoleh sebagai 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. 2** Nilai Penunjang 2 item

## 2. Pembentukan aturan *association*

Jika pola dari frekuensi tinggi ditemukan semua, barulah mencari aturan-aturan asosiasi yang telah memenuhi syarat minimal *confidence* dengan menghitung  $A \rightarrow B$ . diperoleh dari rumus berikut:

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

**Rumus 2. 3** Pembentukan Aturan *association*

### 2.4.2 *Decision Tree*

Pohon pemecah masalah pengambil keputusan merupakan pemetaan tentang bagaimana cara-cara memecahkan sebuah masalah didalam masalah tersebut.

Pohon keputusan tersebut memperlihatkan penyebab-penyebab probabilitas/kemungkinan yang mungkin berpengaruh pada keputusan alternative tersebut, dengan disertai dengan prediksi dari hasil akhir yang didapatkan jika kita mengambil alternative tersebut (Vulandari, 2017).

### 2.4.3 Algoritma ID3

*Iterative Dichotomiser 3* atau biasa di sebut dengan istilah *Induction of Decision Tree* dimana algoritma ini bentuk yang bertujuan untuk memperoleh pohon keputusan yang unggul dan pada algoritma ini bisa mendapatkan hasil *decision tree* yang sederhana yang lebih baik dan konsisten dari sekumpulan data. Mekanisme pada ID3 ini dikerjakan secara *top-down* yaitu dengan memilih atribut manakah yang diperiksa tanpa *root* dari pohon keputusan (Vulandari, 2017).

Menurut dari (Suyanto, 2017) cara kerja dari algoritma ID3 ini terbilang lahap dalam melakukan pencarian data yang mengakibatkan belum secara keseluruhan optimal. Dimana algoritma ini diimplementasikan dengan menggunakan sebuah fungsi rekursi.

### 2.4.4 Nearest Neighbor

Dimana metode *Nearest Neighbour* ini merupakan metode klasifikasi yang sifatnya local sehingga tidak perlu menggunakan parameter yang rumit. Dan pada metode ini dapat dipergunakan dimana terdapat banyaknya data yang menyebabkan representasi model basis *rule*, kepadatan peluang atau model yang sulit dihasilkan. Pada model ini dikenal dengan model yang berbasis statistic yang

menyatakan aturan-aturan dengan menentukan tetangga yang terdekat(Suyanto, 2017).

## 2.5 *Software Pendukung*

*Software* atau biasa di sebut dengan perangkat lunak merupakan sebuah perintah program computer yang ketika dijalankan oleh pengguna akan memberikan fitur dan kinerja sebagaimana dimaksud oleh pengguna. Dengan kata lain, perangkat lunak yang berfungsi untuk memberi perintah kepada computer agar dapat berkinerja optimal sesuai dengan perintah atau arahan dari pengguna. Perangkat pembantu yang dipergunakan dalam penelitian ini adalah WEKA. Sebuah paket *tools mechine learning* yang sangat mudah dan praktis merupakan pengertian dari WEKA. WEKA atau yang biasa disebut dengan WAIKATO *Environment for Knowledge Analysis* yang diciptakan di Universitas Waikato, New Zealand yang difungsikan sebagai penelitian pendidikan dan juga berbagai aplikasi. WEKA juga dianggap telah mampu untuk menyelesaikan persoalan-persoalan *Data mining* di dunia nyata (Vulandari, 2017).

Penerapan WEKA banyak juga diterapkan pada beberapa tingkatan yang berbeda dan juga mudah dalam menggunakannya. WEKA juga memiliki implementasi pada algoritma-algoritma pembelajaran *state –of-the-art* yang dapat diterapkan kedalam *dataset* dari *command line*. Didalam WEKA juga banyak terdapat *tools* yang dapat dipergunakan untuk *preprocessing data*, klasterisasi, *classification*, visualisasi, dan *association rule learning*. Contoh metode pembelajaran WEKA yaitu dengan cara menerapkan sebuah metode kedalam

*dataset* dan menganalisis hasil pemrosesan hasil nilai *output* dipergunakan untuk mendapatkan informasi-informasi tentang data, atau dengan diterapkan metode dan mencocokkan hasil yang akan dipilih.

Seiring dengan pengembangan WEKA mengikuti dengan model yang di *releases* LINUX dengan digit kedua yang genap menandakan pengembangan yang stabil pada WEKA dan digit yang lain menandakan pengembangan. Berikut adalah beberapa versi dari WEKA, antara lain:

1. WEKA 3.0 : “very buku” dalam versi ini penjelasan sesuai dengan yang ada di dalam buku *Data mining*.
2. WEKA 3.2 : “versi GUI” penambahan pada versi ini adalah GUI dan CLI awal.
3. WEKA 3.3 : “versi pengembangan” versi yang telah ditambahkan peningkatan.

## 2.6 Penelitian Terdahulu

Dalam penelitian ini, peneliti mencantumkan beberapa bahan referensi dari penelitian terdahulu yang didapat dari jurnal *data mining* ilmiah, yaitu:

1. Sebuah penelitian dilakukan oleh (Handoko, 2016) yang bertemakan **“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)”**. Didalam sebuah penelitian dengan ISSN 2476-8812 membahas penerapan sebuah data mining dengan mempergunakan sebuah

metode klustering guna meningkatkan pembelajaran mutu dapa instansi perguruan tinggi deprogram studi TKJ akademik komunitas solok selatan. Dengan mempergunakan K Means sebagai algoritmanya dengan mengelompokan data-data atau obyek kedalam sebuah group dengan berisikan sebuah data-data yang sama persis dan berbeda dalam group yang lainnya. Dan tahapan pengujian dikerjakan dengan menggunakan sebuah aplikasi RapidMiner 5.3 sehingga dapat memperoleh sebuah group dalam meningkatkan mutu pembelajaran dengan mempergunakan sebuah table data mahasiswa yang sudah di transformasi sebagai sampel nya. Dengan pengujian 4 variabel dsebagai variabel pertamanya, yaitu jarak tempuh, IP, jumlah kehadiran, dan penghasilan.

2. Sebuah penelitian dilakukan oleh (Gunadi & Sensuse, 2012) yang bertemakan **“Penerapan Metode Data Mining Merket Basket Analysis Terhadap Data Penjualan ProdukBuku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (FP-Growth): Studi Kasus Percetakan PT. Gramedia”**. Dalam sebuah penelitian dengan ISSN 2085-725X membahas tentang ekperimen terhadap suatu data transaksi penjualan buku dipercetakan PT. Gramedia dengan mempergunakan metode *market basket analysis* dengan tujuan melihat suatu hubungan korelasi antara atribut penjualan. Apriori dan FP- growth adalah sebuah algoritma yang akan dipergunakan. Algoritma tersebut dipergunakan untuk membentuk suatu *frequent itemsets* yang akan digunakan sebagai bahan acuan dalam merumuskan aturan asosiasi yang diperoleh dengan model

*market basket analysis* dengan mempergunakan sebuah aplikasi WEKA 3.6 sebagai alat bantu dalam menentukan.

3. Sebuah penelitian dilakukan oleh (Kamal, Iman Mustofa Hendro P, Tachbir Ilyas, 2017) yang bertemakan “**Prediksi Penjualan Buku Menggunakan Data Mining Di Pt. Niaga Swadaya**”. Dalam sebuah penelitian dengan ISSN 2302-3805 membahas tentang sebuah permasalahan pendistribusian buku dengan permintaan dari konsumen kepada buku yang kurang menentu sehingga dapat menyebabkan tidak melakukan pemesanan dari toko penjualan kepada pihak penerbit, sehingga penerbit tidak bisa mendistribusikan sesuai dengan jadwal yang telah di setujui sebelumnya.
4. Sebuah penelitian dilakukan oleh (Tampubolon et al., 2013) yang bertemakan “**Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan**”. Dalam sebuah penelitian dengan ISSN 2339-210X membahas tentang pengimplementasian *data mining* dengan menggunakan apriori sebagai algoritmanya pada sistem persediaan alat-alat kesehatan di sebuah apotek dan jenis-jenis barang apa yang paling unggul dan harus memiliki stok lebih guna untuk mengantisipasi kekosongan/kelangkaan barang tersebut. Dikarenakan kurangnya stok bisa mempengaruhi tingkat pelayanan kepada konsumen dan juga pendapatan sebuah apotek tersebut. Dalam penelitian ini peneliti menerapkan fungsi dari sebuah *data mining* untuk mengelolah sebuah data dan juga untuk mendapatkan informasi yang dapat dipergunakan dalam peningkatan pelayanan dan penjualan.

5. Sebuah penelitian dilakukan oleh (Handoko & Lesmana, 2018) yang bertemakan "**Pengelompokan Data Mining Pada Jumlah Penumpang Di Bandara Hangnadam**". Dalam sebuah penelitian dengan ISSN 2337-8794 & E-ISSN 261-5292 membahas tentang sebuah pengelompokan sebuah data penumpang dengan menggunakan *data mining* dan data yang dipergunakan pada penelitian tersebut adalah sampel data dari rekaman data yang di ambil di bandara Hang Nadim yang berlokasi di Batam.
6. Sebuah penelitian dilakukan oleh (Putria, 2018) yang bertemakan "**Data Mining Penjualan Tiket Pesawat Menggunakan Algoritma Apriori Pada Terminal Tiket Batam Tour & Travel**". Dalam sebuah penelitian dengan ISSN 233-8794 & E-ISSN 2621-5292 membahas tentang pengimplementasian sebuah *data mining* dengan menggunakan apriori sebagai algoritmanya pada penjualan berbagai jenis tiket pesawat yang terdapat pada Terminal Tiket Batam Tour & Travel. Sehingga pemilik sebuah travel bisa mengatur sebuah strategi guna untuk meraup keuntungan dari penjualah tiket.
7. Sebuah penelitian dilakukan oleh (Kurniawan & Informasi, 2016) yang bertemakan "**Aplikasi Data mining untuk Memprediksi tingkat kelulusan Mahasiswa Menggunakan Algoritma Apriori di Ibi Darmajaya Bandar Lampung**". Dalam sebuah penelitian dengan ISSN membahas tentang prediksi dari tingkat kelulusan mahasiswa di Bandar lampung dengan menggunakan apriori sebagai algoritmanya. Dalam pembahasan ini peneliti mencari sebuah pola-pola atau asebuah aturan dari

sejumlah data yang besar. Data induk mahasiswa dan data kelulusan dimanfaatkan sebagai data olah, sehingga diharapkan bisa memberikan suatu informasi tingkat kelulusan melalui data induk dengan teknik *data mining*.

8. Sebuah penelitian dilakukan oleh (Informatika et al., 2015) yang bertemakan “**Data Mining Perkiraan Spanduk Dengan Algoritma Apriori ( Studi Kasus : Cv. Mentari Persada Medan)**”. Dalam sebuah penelitian dengan ISSN 2301-9425 membahas tentang menganalisa sebuah data pembuatan spanduk guna membentuk suatu pola kombinasi *itemsets* dengan algoritma apriori dan membentuk aturan-aturan dengan *association rules* dan juga cara menerapkan *data mining* dengan *tools* Tanagra 1.4.

## 2.7 Objek Penelitian

Objek penelitian adalah titik pusat perhatian didalam suatu penelitian, yang bertujuan untuk mengambil sebuah jawaban atau solusi didalam sebuah permasalahan dalam penelitian. Sebuah objek penelitian mendefinisikan tentang apa, siapa, dimana, kapan, penelitian tersebut dilaksanakan. Didalam penelitian ini yang menjadi obyek penelitian adalah Prediksi Penjualan Tiket CV Patria Mitra Sejati di Batam.

CV Patria Mitra Sejati pertama kali didirikan dipertengahan bulan pada tahun 2012, dimana pada saat louncing pertama kali kelengkapan hanya satu set perangkat komputer yang dirakit sendiri oleh pemilik usaha tersebut. Dan pada saat yang bersamaan dengan berdirinya CV PMS owner atau pemilik usaha

tersebut masih bekerja sebagai dosen. Sehingga usaha yang baru didirikan menjadi kurang diperhatikan karena setiap harinya masih melakukan aktifitas sebagai pengajar di sebuah Universitas.

Setelah usaha tersebut berjalan kurang lebih setengah tahun akhirnya memutuskan untuk berhenti mengajar dan mulai menekuni usahanya di bidang ticketing. Dan mulai berfikir untuk mengembangkan usahanya karena jika hanya mengandalkan satu usaha saja dirasa masih kurang dan kemudian pemilik usaha membuka tambahan usaha PPOB. Dimana dahulu system ticketing yang digunakan masih menggunakan *core system* yang mana proses *booking* dan *issued* masih terbilang cukup merepotkan dalam hal akses.

Seiring waktu yang terus berlalu akhirnya pemilik usaha mulai mencari tim IT untuk mengupdate system ticketing yang dikelola tersebut tepatnya pada tahun 2013. Dan program atau system baru telah selesai pada Januari 2014 yang mulai dipergunakan dalam *booking* dan *issued* tiket, dengan menggunakan system tersebut proses *booking* dan *issued* menjadi lebih simple dan mudah dalam penggunaannya. Karena dalam system tersebut mengadopsi system B2B (business to business) dimana pengguna bisa membuat sub atau orang lain yang ingin membuka usaha penjualan tiket maskapai bisa membuka usaha tentunya dengan biaya yang cukup terjangkau bagi orang kalangan menengah kebawah dan menengah ke atas.

Pada tahun 2013 Desember sub agent dari CV Patria Mitra Sejati sudah mencapai peningkatan yang cukup signifikan yaitu mencapai 700 agent yang berdomisili di kota Batam. Karena jika dirasa hanya di kota Batam saja system

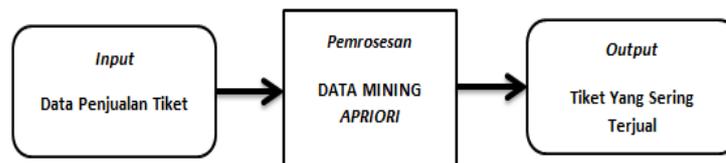
ticketing tersebut tidak bisa berkembang akhirnya mulai membuka cabang distributor di wilayah Kalimantan yang berlokasi di Balikpapan. Yang kemudian system tersebut dikembangkan lagi menjadi B2C (business to customer) dimana pembeli dapat melakukan akses tiket tanpa perlu datang ke travel agent.

Ditahun 2014 sub agent sudah mencapai di angka 3000 travel agent dimana sebagian pemakai atau agent adalah ibu rumah tangga yang ingin menambah penghasilan tambahan dan juga anak muda yang ingin mendapatkan penghasilan dengan cara yang simple, mudah, dan juga praktis. Sehingga menjadikan penjualan tiket pada CV PMS tersebut melambung tinggi dan mendapatkan penghargaan dari pihak maskapai sebagai top agent.

Dengan berjalannya usaha ticketing pada CV PMS tersebut tidak semua berjalan tanpa ada halangan yang merintang. Dikarenakan pendiri selalu memacu diri sendiri dan juga para karyawannya untuk selalu positif dalam menghadapi ombak-ombak yang menerjang. Dan sampai saat ini sub agent dari CV PMS tersebut sudah mencapai 10.000 agent dari sabang sampai merauke.

## **2.8 Kerangka Pemikiran**

Dalam membuat penelitian, peneliti membuat argument sementara ketika melakukan penelitian dari permasalahan yang diajukan. Kerangka berpikir dipergunakan untuk memberi penjelasan secara singkat terhadap suatu objek masalah. Ruanglingkup penelitian disajikan dalam diagram sebagai berikut:



**Gambar 2. 2** Kerangka Pemikiran

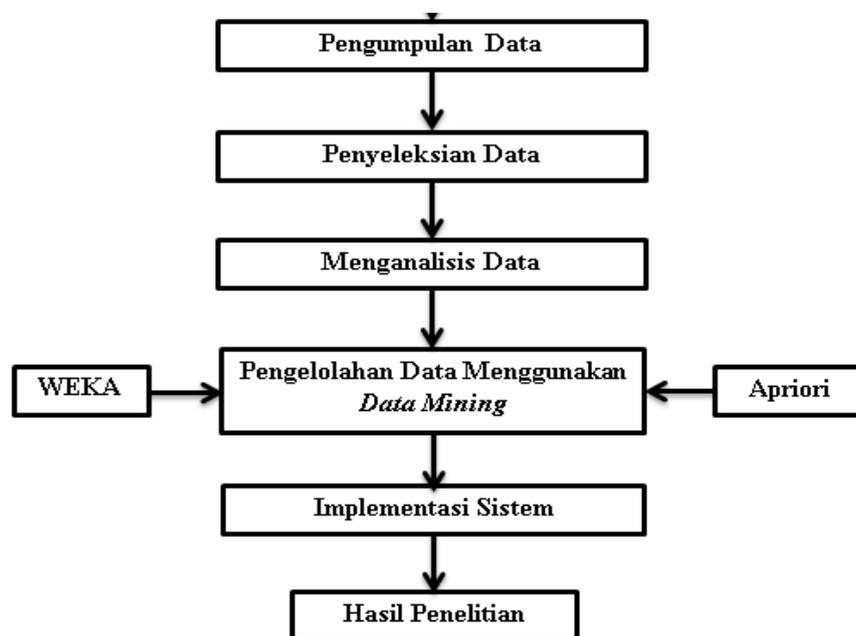
Penjelasan pada gambar menjelaskan bahwa dalam kerangka pemikiran tersebut memiliki beberapa tahapan dan variable berupa variable input, pemrosesan, dan output. Dimana variabel input terdiri dari masukan variabel penelitian yang digunakan sebagai acuan kedalam *Software* WEKA untuk diolah. Variabel yang digunakan untuk input adalah data penjualan tiket. Yang kemudian data diolah dengan menggunakan metode apriori pada tahapan pemrosesan dan diuji menggunakan WEKA. Dan tahapan output merupakan tahapan dimana data yang setelah diolah diaplikasi WEKA diharapkan sesuai dengan tujuan dari penelitian.

### BAB III

#### METODE PENELITIAN

##### 3.1 Arsitektur Sistem

Dalam suatu penelitian, peneliti membuat sebuah perencanaan penelitian ataupun sebuah perancangan, dengan tujuan untuk mempresentasikan keseluruhan dari alur kegiatan dari penelitian tentang *data mining* dari pertama memulai penelitian hingga sampai dengan akhir penelitian. Sebuah desain penelitian yang dapat dijelaskan berdasarkan gambar yang ada dibawah ini:



**Gambar 3. 1** Arsitektur Penelitian  
(Sumber: Data Penelitian 2019)

Keterangan berdasarkan dari tahapan penelitian berikut ini:

1. Mengidentifikasi masalah

Meneliti atau suatu kegiatan untuk mengeksplorasi atau mencari masalah yang potensial. Studi didalam penelitian ini terlalu banyaknya data yang tercecer yang tidak digunakan dengan benar di CV Patria Mitra Sejati yang menyebabkan hanya menjadi tumpukan data yang tidak berguna atau hanya menjadi tumpukan sampah. Oleh karena itu digunakanlah *data mining* dengan mempergunakan sebuah metode apriori supaya dapat mempermudah dalam pengolahan data.

2. Pengumpulan data

Setelah proses dari mengidentifikasi masalah telah selesai dikerjakan, langkah selanjutnya yang dikerjakan peneliti ialah mengumpulkan sebuah data-data yang terkait dengan mempergunakan metode pengamatan dengan terjun langsung ke lokasi dan melakukan wawancara, dengan tujuan untuk mendapatkan sebuah informasi-informasi dari data penjualan tiket yang digunakan untuk sumber penelitian.

3. Penyeleksian data

Setelah data berhasil di peroleh, langkah selanjutnya peneliti melakukan pemilahan data ataupun menyeleksi data. Dengan tujuan untuk dapat meminimalisir pengolahan data yang sama secara berulang.

4. Menganalisis data

Data yang telah berhasil diseleksi tadi di analisis ulang dengan tujuan untuk mengakuratkan sebuah data agar lebih akurat.

## 5. Pengolahan data

Ketika data-data selesai dianalisis data kemudian diolah dengan mempergunakan *data mining* untuk mendapatkan persentasi dati penjualan yang paling banyak terjual.

## 6. Implementasi sistem

Setelah data-data berhasil diolah langkahaselanjutnya yaitu menguji sebuah data tersebut guna untuk mendapatkan kesamaan data menggunakan software WEKA.

## 7. Hasil penelitian

Ketika semua proses diatas telah selesai, barulah didapatkan suatu hasil penelitian tersebut berupa *items* yang sering terjual. Yang menyebabkan terbentuknya persentasi dari tiket yang paling sering terjual.

### 3.2 Teknik Pengumpulan Data

Dimana suatuproses pengumpulan data merupakan sebuah metode yang dipergunakan dalam mendapatkan data-data yang dipergunakan dalam sebuah penelitian dari narasumber yang berkaitan. Ada pula teknik dari pengumpulan data yang dipergunakan dalam penelitian ini ialah:

#### 3.2.1 *Observasi* (Pengamatan)

Yang merupakan suatu metode mengumpulkan sebuah data ataupun informasi-informasi pada suatu obyek penelitian itu sendiri. Observasi dilakukan

peneliti secara langsung ke lokasi CV Patria Mitra Sejati guna pengambilan sebuah data yang akan di olah.

### **3.2.2 Wawancara**

Merupakan metode mengumpulkan data dengan melibatkan peneliti dengan narasumber. Dimana peneliti mengajukan pertanyaan yang berhubungan dengan penelitian kepada narasumbernya. Guna mendapatkan hasil atau informasi-informasi yang dipergunakan sebagai data sumber penelitian. Yang diwawancarai oleh peneliti adalah pendiri CV Patria Mitra Sejati beserta para staff nya.

### **3.2.3 Kajian Pustaka**

Guna mendalami pemahaman data sumber akurat, peneliti melakukan kajian pustaka yang mendukung dan berhubungan dengan penelitian ini. Dengan mengambil dari beberapa sumber-sumber ilmiah berupa buku, jurnal yang berkaitan dengan penelitian dari berbagai sumber.

## **3.3 Oprasional Variabel**

Oprasional variabel yang akan dipergunakan ialah data penjualan tiket pada CV Patria Mitra Sejati. Adapun variabel yang di angkat sebagai poin-poin dalam penjualan tiket sebagai berikut:

### **3.3.1 Report Transaksi**

*Report* Transaksi Merupakan Total dari banyak jumlah tiket yang terjual dalam batas waktu perbulan pada tahun 2015-2018. Dari data berikut jumlah

banyaknya transaksi dapat menentukan mana yang penjualah yang paling tinggi dan mana penjualan yang kecil.

No	Nama Items	Akumulasi Data 2018											
		Jan	Feb	Mart	Apr	Mei	Juni	Juli	Agus	Sept	Okt	Nov	Des
1	QG	838	594	591	610	922	746	1181	1102	1090	614	466	617
2	SJ	980	949	1031	1006	986	931	1113	857	784	707	933	477
3	JT	6814	6147	8519	7945	6979	5976	7732	6263	5693	6240	5008	5260
4	GA	807	862	1107	1052	1275	386	1285	1048	1028	1033	885	574
5	QZ	9	8	33	6	61	45	8	78	35	0	0	0
6	XN	11	8	29	17	45	14	40	62	12	58	30	17

No	Nama Items	Akumulasi Data 2016											
		Jan	Feb	Mart	Apr	Mei	Juni	Juli	Agus	Sept	Okt	Nov	Des
1	QG	308	651	441	264	674	443	408	477	94	64	83	1237
2	SJ	393	295	162	448	507	563	549	506	482	487	404	597
3	JT	4572	4122	4188	4979	4156	4102	3848	3721	3139	3250	4127	3464
4	GA	286	182	340	394	386	564	263	259	407	382	405	350
5	QZ	17	31	8	26	29	22	13	31	25	14	11	3
6	XN	19	16	5	13	16	10	4	16	26	12	14	8

No	Nama Items	Akumulasi Data 2015											
		Jan	Feb	Mart	Apr	Mei	Juni	Juli	Agus	Sept	Okt	Nov	Des
1	QG	1777	1540	2178	100	2415	2552	2271	2414	1508	1410	1538	1649
2	SJ	546	257	221	180	171	144	205	178	67	177	130	930
3	JT	3507	2737	2921	3207	4148	4473	3990	3988	3504	3849	4076	4528
4	GA	430	268	326	0	579	596	360	354	321	251	291	249
5	QZ	43	26	10	25	30	42	23	55	28	28	24	27
6	XN	10	2	7	7	7	14	16	6	11	7	1	18

No	Nama Items	Akumulasi Data 2017											
		Jan	Feb	Mart	Apr	Mei	Juni	Juli	Agus	Sept	Okt	Nov	Des
1	QG	551	417	1252	1276	1923	727	920	819	810	888	777	683
2	SJ	307	543	934	1262	2347	1073	1280	1286	1058	1057	1090	165
3	JT	3232	4199	4744	5796	11778	4856	5205	5694	7052	6505	7351	7056
4	GA	337	630	977	1194	1989	773	1037	947	772	1019	1166	1012
5	QZ	12	25	32	14	22	0	21	11	13	26	38	5
6	XN	6	40	43	88	77	40	19	21	17	18	45	21

**Gambar 3. 2** Report Transaksi 2015-2018  
(Sumber: Data Penelitian 2019)

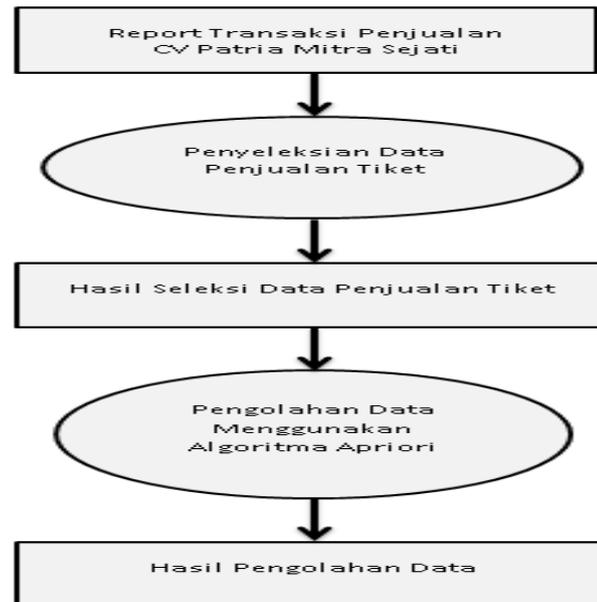
### **3.3.2 Nama Tiket Yang terjual**

Nama atau jenis tiket pesawat yang terjual di CV Patria Mitra Sejati dan maskapai apa saja yang paling banyak dibeli oleh pelanggan. Berikut nama-nama dari berbagai maskapai yang ada sebagai berikut adalah *Citilink Air*, *Lion Air*, *Sriwijaya Air*, *Air Asia*, *Garuda Air*, Express Air.

### **3.4 Metode Analisis**

Metode dalam sebuah analisis merupakan suatu hal yang penting guna untuk melakukan penelitian. Apabila suatu penelitian tidak mempunyai atau menggunakan suatu metode maka, keakurasian dari hasil sebuah penelitian masih mengambang atau diragukan. Suatu metode yang dipergunakan sangat berpengaruh hasil sebuah penelitian yang dilakukan tersebut. Sehingga penelitian ini mempergunakan *data mining* metode *apriori*.

### 3.5 Rancangan System



**Gambar 3. 3** Perancangan system  
(Sumber: Data Penelitian 2019)

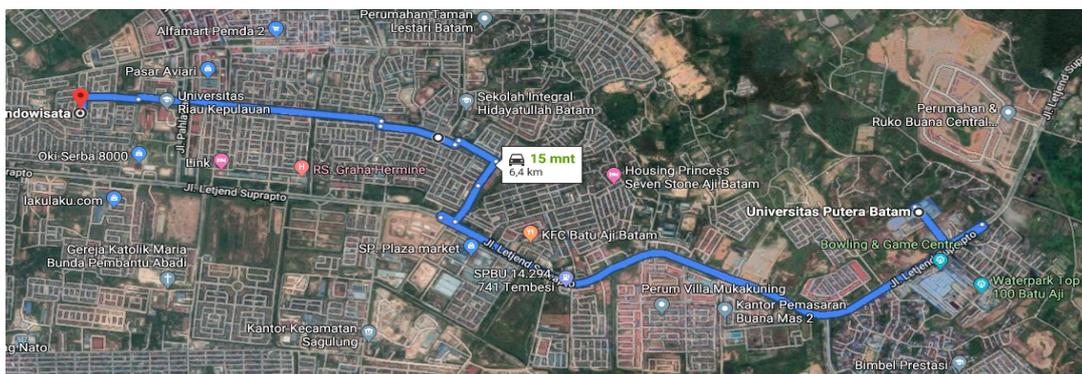
Dalam mengumpulkan data-data transaksi CV Patria Mitra Sejati peneliti mempergunakan sebuah metode yang sederhana yaitu dengan melakukan observasi dan wawancara. Ketika data-data sudah dikumpulkan kemudian dilakukan seleksi data dengan tujuan memilih dan memilah sebuah data yang akan dipergunakan untuk sampel dari penelitian dan yang tidak dijadikan sebagai sampel penelitian ini. Ketika data-data yang telah dijadikan sebagai sampel dari penelitian yang kemudian diolah dengan menggunakan algoritma *apriori* dengan memasukkan sampel-sampel yang telah didapatkan kedalam software WEKA untuk di olah yang kemudian akan didapatkan hasil dari pengolahan data tersebut.

### 3.6 Lokasi Dan Jadwal Penelitian

Dalam hal penelitian dengan tujuan untuk memproses atau mengolah sebuah informasi-informasi data penjualan tiket CV Patria Mitra Sejati dengan diterapkannya *data mining* dengan mempergunakan algoritma *apriori*, sehingga diperlukan sebuah tempat dan jadwal untuk penelitian. Dengan lokasi dan jadwal penelitian sebagai berikut:

#### 3.6.1 Lokasi Penelitian

Lokasi penelitian sangatlah penting dalam melakukan penelitian dalam rangka untuk mengolah data yang terdapat pada CV Patria Mitra Sejati dengan diterapkannya pengolahan data dengan menggunakan *Data Mining* dan *apriori* sebagai algoritmanya, lokasi penelitian ini dilaksanakan di CV Patria Mitra Sejati yang beralamat di Puskopkar Blok. B19 No. 7 Batu Aji, Batam.



**Gambar 3. 4** Lokasi Penelitian  
(Sumber: Data Penelitian 2019)

### 3.6.2 Jadwal Penelitian

**Table 3. 1** Jadwal Penelitian

No.	Kegiatan	Sept				Okto				Nov				Des				Jan				Feb			
		2019				2019				2019				2019				2020				2020			
		1	2	3	4	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)