

**IMPLEMENTASI *DATA MINING*
PADA PENJUALAN TIKET KAPAL MENGGUNAKAN
ALGORITMA *APRIORI***

SKRIPSI



Oleh:

Donda Lita Panjaitan

150210171

**FAKULTAS TEKNIK DAN KOMPUTER
PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS PUTERA BATAM
TAHUN 2019**

**IMPLEMENTASI DATA MINING
PADA PENJUALAN TIKET KAPAL MENGGUNAKAN
ALGORITMA APRIORI**

SKRIPSI

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



Oleh:

Donda Lita Panjaitan

150210171

**FAKULTAS TEKNIK DAN KOMPUTER
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 di perguruan tinggi.

Batam, 09 Agustus 2019
Yang membuat pernyataan,

Donda Lita Panjaiatan
150210171

**IMPLEMENTASI DATA MINING
PADA PENJUALAN TIKET KAPAL MENGGUNAKAN
ALGORITMA APRIORI**

**Oleh
Donda Lita Panjaitan
150210171**

SKRIPSI

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

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

Batam, 09 Agustus 2019

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

ABSTRAK

Pelabuhan domestik sekupang merupakan salah satu pelabuhan terbesar yang ada dikota batam, memiliki beberapa *rute* tujuan dan juga kapal ferry yang lumayan banyak. Banyaknya ferry membuat para *costumer* berlomba-lomba ingin menaiki ferry yang mana untuk mencapai tujuannya. Setiap hari banyak penumpang yang pergi sehingga kewalahan apalagi pada saat hari raya untuk menyiapkan dermaga tambahan atau cadangan. Banyaknya penjual tiket dan pelabuhan membuat penjualan tidak stabil ataupun musiman kadang meningkat ataupun turun. Oleh karena itu dibutuhkan strategi pemasaran untuk melakukan prediksi kedepan dengan melihat data hasil penjualan yang sebelumnya. Maka dilakukan penelitian untuk menganalisa dan memprediksi penjualan tiket kapal dengan tekni data mining. Data mining adalah proses mencari pola atau hubungan dari data yang diolah sehingga bermanfaat dalam skala besar. Adapaun metode yang dipakai adalah metode *apriori* yaitu metode yang memiliki aturan asosiasi atau hubungan antar item yang berguna untuk mempermudah para pemilik penjual tiket dalam mengelola data-data penjualan tiket. Aplikasi yang dipakai sebagai implementasi adalah Tanagra. Dalam penelitian ini terdapat 5 item set yang diperoleh dari auran asosiasi dengan minimum support 45% dan minimum *confidence* 75% serta 42 aturan asosiasi secara keseluruhan dan dengan 33 *final rule* yang didapat. Maka tiket kapal yang paling sering banyak terjual adalah, khromah, mega fajar, rahmat jaya, tenggiri dan caroline express.

Kata kunci: *data mining*, algoritma *apriori*, tiket kapal

ABSTRACT

The domestic port is one of the biggest ports in the city of Batam, has several destinations and also quite a number of ferries. The number of ferries makes customers to get on which ferry to reach the destination. Every day many passengers who just leave waiting on the holiday to prepare an additional dock or reserve. The number of ticket and port sellers makes sales unstable or sometimes goes up or down. Therefore we need a marketing strategy to make future predictions by looking at the previous sales data. Then a study was conducted to analyze and predict ship ticket sales with data mining techniques. Data mining is the process of finding patterns or relationships from processed data so that it is useful on a large scale. As for the method used is a priori method that is the method required by association or relationship between items that is useful to facilitate the ticket seller in managing ticket sales data. The application used as an implementation is Tanagra. In this study there are 5 set items obtained from association fees with a minimum support of 45% and a minimum of 75% trust and 42 overall association rules and with 33 final rules obtained. So the most frequently sold boat tect is, khromah, mega fajar, tenggiri, rahmat jaya and caroline express.

Keywords: data mining, a priori algorithm, ship ticket

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa yang telah melimpahkan berkat dan rahmat-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan judul “Implementasi *Data Mining* pada Penjualan tiket kapal menggunakan Algoritma *Apriori*”. Skripsi ini merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa dalam penulisan skripsi ini masih banyak kekurangan dan tidak sempurna, oleh karena itu kritik dan saran dari pembaca sangat penulis harapkan.

Dalam penyusunan skripsi ini penulis mendapatkan bantuan dari berbagai pihak, maka dalam kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Ibu Dr. Nur Elfi Husda S.Kom M.Kom selaku Rektor Universitas Putera Batam.
2. Bapak Andi Maslan S.T M.Si Ketua Program Studi teknik informatika
3. Bapak Sunarsan Sitohang S.Kom.,M.TI selaku pembimbing Skripsi dan pembimbing akademik pada Program Studi teknik informatika Universitas Putera Batam.
4. Dosen dan Staff Universitas Putera Batam
5. Kedua orang tua yang memberikan semangat, dorongan dan selalu mendoakan penulis.
6. Adek Ririn Panjaitan, Saut Ian Panjaitan, Nita Panjaitan , Frans Irwan yang mendukung saya baik secara materi dan nonmateri.
7. Amrin simanjuntak , Armeni simanjuntak yang selalu membantu dan mendukung secara materi ataupun non materi.
8. Kepada teman-teman seperjuangan penulis yang sudah memberikan masukan dan semangat dalam penyusunan skripsi ini.

Semoga Tuhan membalas kebaikan dan selalu mencurahkan rahmat dan berkatnya. Amin.

Batam, 09 Agustus 2019

Donda Lita Panjaitan

DAFTAR ISI

HALAMAN SAMPUL DEPAN	
HALAMAN JUDUL	ii
HALAMAN PERNYATAAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	Error! Bookmark not defined.
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	vii
DAFTAR TABEL	vii
DAFTAR GAMBAR	xi
DAFTAR RUMUS	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Batasan Masalah.....	5
1.4 Rumusan Masalah	5
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	6
BAB II LANDASAN TEORI	7
2.1 <i>Knowledge Discovery in Database (KDD)</i>	7
2.2 <i>Data mining</i>	10
2.2.1 Pengertian <i>Data mining</i>	10
2.2.2 Pengelompokan <i>Data mining</i>	11
2.2.3 Manfaat <i>Data mining</i>	13
2.2.4 Tahapan <i>Data mining</i>	14
2.3 Klasifikasi Data Mining	17
2.4 Algoritma apriori.....	19
2.5 Tiket	21
2.6 <i>Software</i> pendukung.....	22
2.7 Penelitian Terdahulu	24
2.8 Kerangka Pemikiran.....	28
BAB III METODE PENELITIAN	29

3.1 Desain Penelitian.....	29
3.2 Pengumpulan Data	32
3.2.1 Teknik Pengumpulan Data.....	32
3.3 Operasional Variabel.....	35
3.4 Metode Perancangan Sistem	36
3.5 Lokasi dan jadwal penelitian.....	38
3.5.1 Lokasi penelitian	38
3.5.2 Jadwal penelitian.....	38
BAB IV HASIL DAN PEMBAHASAN	40
4.1 Analisis Data	40
4.1.1 Data Set	40
4.1.2 Representasi Data Transaksi	42
4.1.3 Tabulasi Data Transaksi.....	43
4.1.4 Perhitungan Satu <i>Itemset</i>	44
4.1.5 Perhitungan 2 <i>itemset</i>	46
4.1.6 Perhitungan Tiga <i>Itemset</i>	54
4.1.7 Perhitungan Empat <i>Itemset</i>	58
4.1.8 Perhitungan Lima <i>Itemset</i>	61
4.1.9 Pembentukan Pola Aturan Asosiasi	63
4.2 Implementasi	68
4.2.1 <i>Import Data</i>	69
4.2.1 Proses Data.....	73
BAB V KESIMPULAN DAN SARAN	80
5.1 Kesimpulan	80
5.2 Saran.....	80
DAFTAR PUSTAKA	81
DAFTAR RIWAYAT HIDUP	82
SURAT KETERANGAN PENELITIAN.....	83
LAMPIRAN.....	84

DAFTAR TABEL

Tabel 3.1 Nama ferry	36
Tabel 3.2 Jadwal Penelitian.....	39
Tabel 4.1 Data Transaksi Penjualan Tiket.....	41
Tabel 4.2 Representasi Data Transaksi	42
Tabel 4.3 Tabel Tabular Transaksi.....	43
Tabel 4.4 Perhitungan Frekuensi Satu <i>Itemset</i>	46
Tabel 4.5 Hasil Perhitungan Satu <i>Itemset</i>	46
Tabel 4.6 Perhitungan 2 <i>itemset</i>	51
Tabel 4.7 Aturan Asosiasi 2 <i>Itemset</i>	53
Tabel 4.8 Perhitungan 3 <i>itemset</i>	56
Tabel 4.9 Aturan Asosiasi 3 <i>Itemset</i>	58
Tabel 4.10 Perhitungan 3 <i>Itemset</i>	60
Tabel 4.11 Aturan Asosiasi 4 <i>Itemset</i>	61
Tabel 4.12 Perhitungan 5 <i>Itemset</i>	62
Tabel 4.13 Aturan Asosiasi 5 <i>Itemset</i>	63
Tabel 4.14 Hasil Perhitungan 2,3,4 Dan 5 <i>Itemset</i>	63
Tabel 4.15 <i>final rule</i>	65

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Proses Tahapan KDD.....	8
Gambar 2.2 Gambaran Tahap- Tahap data mining	15
Gambar 2.3 Kerangka pemikiran.....	28
Gambar 3.1 Desain Penelitian	30
Gambar 3.2 Perancangan Sistem.....	36
Gambar 4.1 Tampilan halaman	69
Gambar 4.2 Langkah <i>import</i>	70
Gambar 4.3 Pilih Data	71
Gambar 4.4 Hasil <i>Import</i>	72
Gambar 4.5 Data <i>View</i>	73
Gambar 4.6 Proses Data	74
Gambar 4.7 Proses Asosiasi	74
Gambar 4.8 Metode <i>apriori</i>	75
Gambar 4.9 Parameter	76
Gambar 4.10 <i>Card Itemset</i>	77
Gambar 4.11 <i>Itemset</i>	78
Gambar 4.12 <i>Itemset 2</i>	79
Gambar 4.13 Hasil 3 <i>Itemset</i>	80
Gambar 4.14 Hasil 4 <i>itemset</i>	80
Gambar 4.15 Hasil 5 <i>item set</i>	81
Gambar 4.16 <i>Final Rule</i>	82

DAFTAR RUMUS

Rumus 2.1 1 <i>Item set</i>	20
Rumus 2.2 2 <i>Item set</i>	20
Rumus 2.3 <i>Confidence</i>	21

BAB I

PENDAHULUAN

1.1 Latar Belakang

Transportasi menjadi sarana atau akses tercepat yang digunakan manusia saat mau bepergian dari satu daerah ke daerah yang lain. Transportasi mempunyai peranan yang sangat penting dalam menentukan suatu pembangunan ataupun saat pemindahan. Transportasi dapat membantu berbagai kegiatan yaitu seperti kegiatan produksi, transaksi, distribusi, dan konsumsi. Ada 3 jenis transportasi yang sering digunakan oleh manusia yaitu: transportasi darat, transportasi udara dan transportasi air. Salah satu kota yang menggunakan ketiga jenis transportasi tersebut adalah kota Batam yaitu kota kecil yang berdekatan dengan negara Singapura dan Malaysia.

Batam merupakan salah satu kota dengan pertumbuhan ekonomi terpesat di Indonesia. Salah satu alat transportasi yang sangat diperlukan adalah transportasi laut. Letaknya yang strategis berada di jalur pelayaran internasional dan memiliki jarak yang sangat dekat dan berbatasan langsung dengan Singapura dan Malaysia. Sehingga sangat mudah bagi pengunjung dari negara luar untuk berkunjung ataupun dalam urusan bisnis. Salah satu transportasi yang digunakan adalah transportasi laut yaitu kapal untuk melakukan penyebarangan domestik maupun internasional. Ada

beberapa pelabuhan yang ada di Batam salah satunya pelabuhan internasional Sekupang yang memiliki banyak akses dalam domestik dan internasional.

Pelabuhan sekupang adalah pelabuhan domestik dan internasional yang berada dipantai barat pulau Batam, provinsi Kepulauan Riau. Pelabuhan ini menghubungkan kota Batam dengan pelabuhan-pelabuhan di kepulauan sebelah barat, seperti pelabuhan Moro, Tanjung Balai Karimun, selat panjang, Bengkalis, Tanjung Batu Karimun, Pelabuhan Dumai, Pulau Burung, serta kepulauan di sebelah selatan seperti pulau Bangka, Belitung dan juga dengan kota Palembang dan Jakarta. Pelabuhan sekupang dapat menjadi *route* tujuan untuk menjangkau keluar kota.

Banyak ditemukan pengunung data dari tahun ke tahun dan jika dihitung manual tidak akurat dan relevan. Jika dibiarkan saja, maka data-data tersebut akan menjadi sampah dan tidak berarti. Banyaknya kapal yang beroperasi setiap hari menunjukkan bahwa tingkat penumpang semakin banyak, oleh karena itu persediaan tiket dituntut untuk tetap ada namun jumlah penumpang yang tidak tentu jumlahnya membuat sulit untuk memprediksi persediaan tiket kapal.

Pelabuhan sekupang merupakan salah satu pelabuhan terbesar di kota batam dan memiliki *route* domestik dan internasional. Padatnya jadwal keberangkatan domestik dan laporan data keberangkatan dan data penumpang, membuat banyak data yang harus dilaporkan. Namun data tersebut masih bersifat manual dan laporan tersebut juga dilaporkan menggunakan manual. Dengan laporan manual pihak

pelabuhan kesulitan untuk mengetahui kapal mana yang paling diminati masyarakat dan kapal mana yang harus ditambahkan di hari-hari tertentu. Dengan banyaknya data tersebut maka digunakan metode *data mining* untuk menganalisis data laporan yang bisa membantu dengan cepat dengan akurasi yang tinggi.

Proses analisis data diperlukan sebuah teknik untuk mengetahui perbandingan tiket kapal laut dari batam ke daerah lain dan melakukan prediksi. Dengan menggunakan teknik *data mining* dapat melakukan prediksi tiket kapal yang berangkat dari kota Batam. *Data mining* adalah salah satu cara yang digunakan untuk mengolah data menjadi informasi yang sangat penting dari data. Data-data yang dihasilkan dari proses penjualan yaitu data transaksi penjualan tiket diolah menggunakan algoritma *apriori*. Menggunakan teknik tersebut, dapat diketahui seberapa besar tingkat penjualan tiket kapal di pelabuhan sekupang berdasarkan prediksi yang telah dilakukan. *Data mining* adalah proses untuk menggali informasi untuk menemukan hubungan pola yang berupa pengetahuan yang selama ini belum diketahui secara manual (Husin & Mulyaningsih, 2015). Menerapkan algoritma *apriori* pada penjualan tiket untuk penentuan penjualan tiket pesawat yang paling banyak terjual berdasarkan jenis maskapainya. Algoritma *apriori* digunakan untuk menggali informasi, fakta dan wawasan yang tidak biasa dan menambahkan pengetahuan itu sendiri (Narti Eka Putrie, 2018).

Ada beberapa metode dalam teknik *data mining*, dalam penelitian ini menggunakan metode algoritma *apriori*. Dalam memprediksi hujan pengolahan data cuaca merupakan hal yang penting. Data cuaca yang semakin hari semakin bertambah

menyebabkan penumpukan data sehingga pengolahan data tersebut perlu penanganan lebih lanjut (Fauzy & Asror, 2016). Data tersebut membentuk sebuah model asosiasi atau hubungan yang digunakan untuk melakukan prediksi dan digunakan untuk mengetahui tiket yang paling banyak terjual dipelabuhan sekupang. Dari uraian diatas peneliti tertarik untuk melakukan penelitian yang berjudul **“IMPLEMENTASI DATA MINING PADA PENJUALAN TIKET KAPAL MENGGUNAKAN ALGORITMA APRIORI”**.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan di atas maka indentifikasi masalah dari penelitian ini adalah :

1. Harga tiket kapal laut merupakan faktor pertimbangan konsumen dalam melakukan perjalanan
2. Persediaan tiket kapal yang habis membuat konsumen kadang kecewa terutama pada hari raya.
3. Sulitnya mengetahui kapal yang paling diminati masyarakat, karena laporannya masih bersifat manual.

1.3 Batasan Masalah

Masalah yang dibuat dalam penelitian sangat banyak, untuk itu penelitian membuat batasan masalah agar tidak menyimpang dari penelitian yang ingin dicapai, adapun batasan masalahnya sebagai berikut:

1. Peneliti hanya meneliti tentang tiket yang terjual dari pelabuhan sekupang ke antar daerah/domestik.
2. Data yang digunakan adalah data tahun 2019 yang diperoleh dari staff pelabuhan sekupang.
3. Data yang diambil yaitu tujuan Moro.
4. Penelitian ini menggunakan *Aplikasi Tanagra*.

1.4 Perumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan diatas maka dirumuskan permasalahannya sebagai berikut:

1. Bagaimana menganalisis data tiket kapal ke dalam aplikasi Tanagra sehingga menghasilkan prediksi yang akurat.
2. Bagaimana menerapkan algoritma *apriori* untuk mengetahui penjualan tiket kapal paling banyak terjual.

1.5 Tujuan Penelitian

Adapun tujuan penelitian yang dilakukan penulis adalah:

1. Untuk menganalisis tiket kapal kedalam aplikasi Tanagra sehingga menghasilkan prediksi yang tinggi.
2. Untuk mengetahui penerapan algoritma *apriori* penjualan tiket kapal paling banyak terjual di pelabuhan sekupang.

1.6 Manfaat Penelitian

Manfaat penelitian dapat dikategorikan menjadi 2 yaitu:

1.6.1 Aspek Teoritis

1. Menambah wawasan dan pengetahuan tentang penjualan tiket berdasarkan jumlah tiket yang terjual.
2. Sebagai sumber tambahan dalam membantu mahasiswa melakukan penelitian selanjutnya.

1.6.2 Aspek Praktis

1. Bagi pelabuhan sekupang
Membantu karyawan yang bekerja di pelabuhan dalam mengolah data yang jumlahnya besar.
2. Bagi kampus
Sebagai sumber tambahan dalam membantu mahasiswa melakukan penelitian selanjutnya.

3. Peneliti

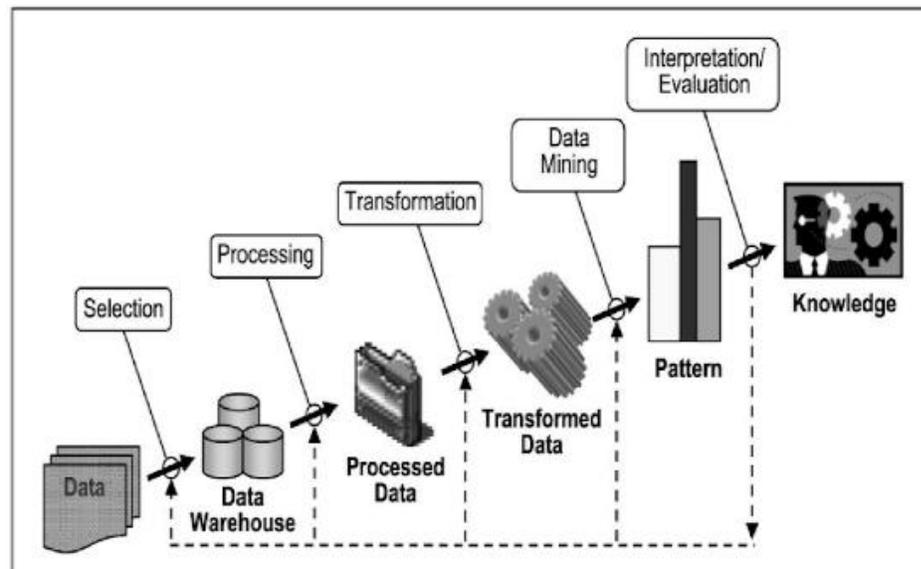
Untuk meningkatkan pengetahuan tentang *data mining* algoritma *apriori* serta pemenuhan terhadap salah satu syarat untuk memperoleh gelar S1 teknik informatika.

BAB II

LANDASAN TEORI

2.1 *Knowledge Discovery in Database (KDD)*

Menurut (Retno tri vulandari, s.si, 2017:2) *Data mining* adalah sebuah proses untuk mencari informasi yang masih banyak yang belum *real* dalam basis data yaitu proses *knowledge discovery in database* (KDD). KDD menghasilkan pengetahuan yang lebih baik dan lebih mudah untuk dimegerti. KDD mempunyai hubungan dengan teknik, integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari banyak data. KDD adalah rangkaian yang dilakukan untuk pengumpulan data, dalam pengolahan data akan digunakan pola atau hubungan yang tersusun teratur didalam database pada dalam dengan jumlah yang sangat besar. *Knowledge discovery in database* (KDD) pada intinya adalah proses menemukan pengetahuan yang bermanfaat pada pengumpulan data.



Gambar 2.1 Ilustrasi Proses Tahapan KDD

1. Seleksi Data (*Data Selection*)

Sekumpulan data yang dilakukan untuk mencari pengolahan informasi dalam KDD. Data hasil seleksi yang digunakan untuk *data mining*, menyimpan data, dan diolah dan akan dibedakan dari data yang tidak digunakan. Membuat data target, memilih data, fokus dalam satu variabel atau sampel untuk mencari data yang relevan.

2. Pemilihan data (*Preprocessing / Cleaning*)

Untuk membuat proses *data mining*, proses selanjutnya akan dilakukan pemilihan data yang digunakan pada data KDD. Dimana masih banyak data yang salah dan tidak relevan atau data yang sama (*double*), memeriksa data yang tidak valid atau *real* yang diperlukan pada KDD, seperti data atau informasi eksternal.

3. Transformasi (*Transformation*)

Yaitu proses transformasi atau perombakan data yang telah di pilih, dan data harus benar dengan fakta atau cara yang dilakukan *data mining*. Dalam proses transformasi diperlukan kreatif tergantung pada pola informasi yang diproses dari database. Fungsi transformasi adalah untuk menjelaskan atau menerangkan data yang relevan atau sudah siap untuk proses *data mining* berdasarkan goal yang ingin di capai.

4. *Data mining*

Proses *data mining* adalah proses untuk memilih dan menggali informasi yang sangat menarik untuk memilih data dengan menggunakan banyak cara atau teknik. Pada proses ini akan melakukan pemilihan algoritma untuk pencarian. Pemilihan hasil dari proses KDD adalah seperti *association, regression, forecasting, solution, clustering*. Pemilihan algoritma ini sangat bermanfaat pada goal untuk proses KDD secara menyeluruh.

5. Interpretasi / Evaluasi (*interpretation/evaluation*)

Didalam proses itu pola-pola yang dihasilkan dari *data mining* akan diamati. Pada proses ini informasi yang sudah didapat akan ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang ingin menggunakannya. Proses KDD akan memeriksa semuanya apakah informasi yang dicari berbeda dengan fakta atau hasil pengamatan yang dilakukan sebelumnya.

2.2 *Data mining*

2.2.1 *Pengertian Data mining*

Data mining merupakan data yang diperoleh berupa informasi yang akan diolah untuk informasi yang sangat bermanfaat dalam data base. *Data mining* adalah proses pengambilan informasi dari kumpulan data melalui algoritma dan teknik yang melibatkan ilmu *statistic*, *machine* dan *basis data* (Rodiyansyah, 1861). *Data mining* adalah rangkaian cara untuk mencari nilai tambah dari suatu data berupa pengetahuan yang selama ini tidak diketahui.

Para pakar juga memberikan pengertian tentang *data mining* yang dituangkan dalam buku yang ditulisnya yaitu : Menurut (Dr. Suyanto,S.T.,2017:1) *Data mining* adalah proses analisis pada pengetahuan dalam data base untuk menemukan ide-ide baru yang belum diketahui sebelumnya yang dikumpulkan dalam data yang sangat banyak dengan metode atau cara merupakan bagian dari kecerdasan buatan, mesin pembelajaran, statistika dan basis data. *Data mining* adalah suatu cara atau strategi untuk membuat nilai tambah yang selama ini belum diketahui secara manual dalam basis data (Retno Tri Vulandari, S.Si. 2017:1).

Data mining merupakan suatu hubungan untuk mencari pola dan kecenderungan untuk mengkoreksi data-data yang berjumlah banyak yang disimpan dalam penyimpanan database dengan menggunakan teknik matematika dan statistika (Kusrini & Emha Taufiq Luthfi, 2009:1). Berdasarkan pengertian *data mining* yang telah diuraikan maka dapat ditarik kesimpulan bahwa *data mining* itu merupakan suatu proses pencarian dan penggalian data yang selama ini belum diketahui dalam

data yang sangat banyak didalam database yang berguna untuk pemilik data kemas yang akan datang.

2.2.2 Pengelompokan *Data mining*

Menurut (Kusrini & Emha Taufiq, 2009:10) Dalam bukunya *Data mining* dibagi dalam beberapa kategori berdasarkan tugas yang dapat dilakukan yaitu:

1. Deskripsi

Banyak peneliti mencoba cara atau langkah dalam menafsirkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh petugas pemungutan suara mungkin tidak dapat menemukan keterangan data atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Model dibuat dalam bentuk laporan lengkap yang mempunyai nilai dan variable target sebagai nilai dugaan sementara. Kemudian peninjauan berikutnya estimasi nilai variabel akan dibuat menurut nilai prediksi. Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numeric daripada kearah kategori. Sebagai contoh yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pascasarjana dengan melihat nilai indeks prestasi mahasiswa pada saat mengikuti program sarjana.

3. Prediksi

Prediksi hampir sama dengan estimasi dan klasifikasi, bedanya dalam perkiraan nilai dari hasil akan ada dimasa yang akan datang. Sebagian metode yang digunakan dalam klasifikasi dan estimasi dapat juga digunakan untuk prediksi atau untuk keadaan yang tepat

Contoh : untuk memprediksi harga beras di bulan yang akan mendatang

4. Klasifikasi

Dalam Klasifikasi, ada target variable kategori. Sebagai contoh pembagian pendapatan dapat dipisahkan dalam 3 kategori yaitu pendapatan tinggi, pendapatn sedang, dan pendapatan rendah, contoh lainnya adalah memperkirakan apakah suatu pengajuan hiptek oleh nasabah merupakan suatu kredit yang baik dan buruk.

5. Pengklusteran

Merupakan pengklasifikasian laporan, observasi, atau memperhatikan dan membuat tahap objek-objek yang mempunyai kesamaan. Kluster adalah rangkaian laporan yang mempunyai kesamaan antara yang satu dengan yang lain dan tidak memiliki ketidaksamaan dengan laporan-laporan kluster lainnya. Pengklusteran berbeda dengan yang tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak melakukan klasifikasi, mengestimasi atau memperkirakan nilai dari target variabel. Misalnya dalam pengklusteran dalam bisnis penelitian yaitu untuk tujuan audit akuntansi yaitu

melakukan pemisahan terhadap perilaku finansial dalam baik dan mencurigakan

6. Asosiasi

Data mining bertugas untuk menemukan atribut yang muncul dalam satu tahun. Atau dalam dunia bisnis disebut juga analisis keranjang belanja. Contoh asosiasi dalam dunia bisnis dan penelitian adalah

- a. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

2.2.3 Manfaat *Data mining*

Menurut (Dr.Suyanto, S.T., 2017:3) kegunaan *data mining* ada 2 yaitu: deskriptif dan prediktif. Deskriptif adalah *data mining* yang dilakukan untuk menemukan cara- cara yang dapat dimengerti oleh manusia yang menerangkan karakteristik data. Prediktif adalah *data mining* untuk menghasilkan model ilmu pengetahuan yang akan digunakan untuk melakukan perkiraan atau prediksi.

Berdasarkan fungsinya, adapun tugas-tugas *data mining* dibagi menjadi 6 kelompok yaitu:

1. Klasifikasi (*classification*)

Klasifikasi adalah menguraikan cara yang diketahui untuk digunakan pada data-data baru. Misalkan, klasifikasi penyakit kedalam berbagai macam, klasifikasi email kedalam spam atau bukan.

2. Klasterisasi (*clustering*)

Klasterisasi adalah menggolongkan data, yang belum diketahui jenis kelasnya, kedalam beberapa kelompok sesuai dengan kesamaannya.

3. Regresi (*regression*)

Regresi adalah mencari fungsi yang memodelkan data (kesalahan prediksi) seminimal mungkin.

4. Deteksi anomali (*anomaly detection*)

Deteksi anomali adalah menemukan data yang tidak umum, bisa berupa *outlier* (pencilan), perbaikan atau deviasi yang sangat penting dan perlu penyelesaian selanjutnya.

5. Pembelajaran aturan asosiasi (*association rule learning*)

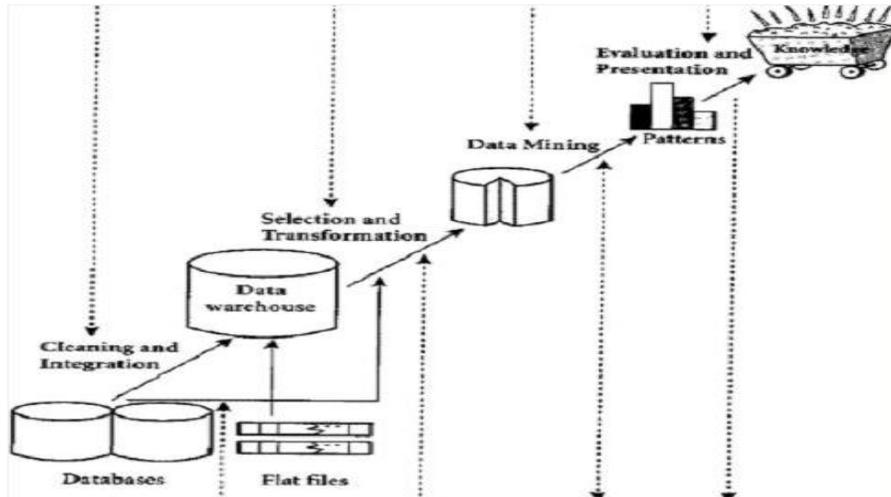
Adalah pemodelan perbandingan (*defency modeling*) atau menemukan relasi antar variabel.

6. Perangkuman (*summarization*)

Perangkuman adalah menyediakan representasi data yang lebih singkat yaitu visualisasi dan penyusunan laporan.

2.2.4 Tahapan *Data mining*

Salah satu susunan dalam proses *data mining* yaitu proses tahapan yang dilakukan. Tahap-tahap akan saling berhubungan dengan teknik integrasi dan penemuan ilmiah dan visualisasi dalam beberapa data.



Gambar 2.2 Gambaran Tahap- Tahap data mining

Tahap- tahap *data mining* adalah sebagai berikut:

1. Pembersihan data (*data cleaning*)

Secara umum, informasi yang diperoleh, baik dari database perusahaan dan hasil pengamatan, mempunyai konten yang belum sempurna seperti data yang hilang, data yang tidak valid atau juga hanya salah ketik. Ada juga atribut data yang tidak relevan dengan hipotesis *data mining* yang kita miliki. Data yang tidak bagus juga lebih baik dibuang karena dapat mengurangi kualitas atau keakuratan hasil *data mining* nantinya. Sampah dalam sampah (hanya sampah yang akan diproduksi jika sampah juga termasuk) adalah kata yang sering digunakan untuk menjelaskan tahap ini. Pemilihan data juga akan mempengaruhi kinerja sistem penambangan data karena data yang ditangani akan berkurang dalam jumlah dan kompleksitas.

2. Integritas Data (*Data Integritation*)

Integrasi data digunakan pada atribut yang mengidentifikasi entitas unik seperti nama, tipe produk, nomor pelanggan, dll. Integrasi data sangat penting dilakukan dengan hati-hati karena kesalahan dalam integrasi data dapat mengakibatkan hasil yang terdistorsi dan bahkan tindakan menyesatkan yang diambil kemudian. Misalnya, jika integrasi data berdasarkan jenis produk ternyata mengaitkan produk dari kategori yang berbeda, maka akan ada korelasi antara produk yang sebenarnya tidak ada. Dalam integrasi data ini juga perlu ditransformasikan dan dibersihkan data karena seringkali data dari dua database berbeda bukan cara penulisan yang sama atau bahkan data dalam satu database ternyata tidak ada di database lain.

3. Transformasi data (*Data Transformation*)

Teknik data mining memerlukan format data khusus sebelum dapat diterapkan. Misalnya sejumlah teknik standar seperti analisis asosiasi dan pengelompokan hanya dapat menerima input data kategorikal. Oleh karena itu data dalam bentuk angka numerik yang terus dibagi menjadi beberapa interval. Proses ini sering disebut *binning*. Di sini juga pemilihan data yang dibutuhkan oleh teknik penambangan data yang digunakan. Transformasi dan pemilihan data ini juga menentukan kualitas hasil penambangan data nantinya karena ada beberapa karakteristik teknik penambangan data tertentu yang tergantung pada tahap ini. Masalah satu bagian dalam *data mining*. Proses ekstraksi pola dari data yang sudah ada sebelumnya.

4. Evaluasi pola

Pada tahap ini, hasil dari teknik *data mining* dalam bentuk pola dan model prediksi yang berbeda dievaluasi untuk menilai apakah hipotesis yang ada telah tercapai. Jika ternyata hasil yang diperoleh tidak sesuai dengan hipotesis ada beberapa alternatif yang dapat diambil seperti: membuat umpan balik untuk meningkatkan proses penambangan data, mencoba teknik penambangan data lain yang lebih cocok, atau menerima hasil ini sebagai hasil tak terduga yang mungkin berguna.

5. Presentasi pengetahuan (*Knowledge Presentation*)

Tahap terakhir dari proses penambangan data adalah bagaimana merumuskan keputusan atau tindakan dari hasil analisis yang diperoleh. Ada saat-saat ini harus melibatkan orang-orang yang tidak mengerti penambangan data. Oleh karena itu penyajian hasil penambangan data dalam bentuk pengetahuan yang dapat dipahami oleh semua orang merupakan tahapan yang dibutuhkan dalam proses penambangan data. Dalam presentasi ini, visualisasi juga dapat membantu mengkomunikasikan hasil penambangan data.

2.3 Klasifikasi *Data Mining*

Menurut ((Kusrini & Emha Taufiq Luthfi, 2009:13) klasifikasi data mining terdiri dari:

1. Algoritma C4.5

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut –atribut yang ada. Algoritma C4.5 dapat menangani data numerik dan diskrit. Algoritma C4.5 digunakan ratio perolehan (*gain ratio*). Namun sebelum menghitung rasio hasil perlu dilakukan nilai informasi dalam antara bits dari suatu kumpulan objek yaitu dengan menggunakan konsep *entropy*.(Prismas, Jakarta, Asistiyasari, & Baidawi, 2017).

2. Algoritma *nearest neighbor*

Nearest neighbor adalah suatu strategi yang mencari masalah untuk menilai kedekatan antara masalah lama dengan masalah baru yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada. Misalnya yang diinginkan untuk mencari solusi terhadap seorang pasien baru dengan menggunakan solusi dari pasien terdahulu. Untuk mencari kasus mencari kasus pasien mana yang digunakan, maka dihitung kedekatan kasus pasien baru dengan semua kasus pasien lama. Kasus pasien lama dengan kedekatan terbesar yang akan diambil solusinya untuk digunakan pada kasus pasien baru.

3. *Fuzzy C-means*

Fuzzy C-means merupakan proses pembagian data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan data dalam satu kelompok besar daripada kesamaan data tersebut dengan data dalam kelompok lain. Potensi clustering adalah dapat digunakan untuk mengetahui struktur dalam data yang data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luasa seperti klasifikasi, pengelolaan *picture* dan pengenalan pola.

4. *Bayesian classification*

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memperkirakan probabilitas keanggotaan suatu kelas. *Bayesian classification* didasarkan pada teorema Bayes yang mempunyai kemampuan klasifikasi serupa dengan decision tree dan neural network. *Bayes classification* terbukti memiliki akurasi dan kecepatan yang sangat tinggi saat diaplikasikan kedalam database dengan data yang besar.

5. *Algoritma Association*

Algoritma aturan asosiasi akan menggunakan data latihan, sesuai dengan pengertian *data mining*, untuk menghasilkan pengetahuan. Pengetahuan apakah yang hendak dihasilkan dalam aturan asosiasi? Pengetahuan untuk mengetahui *item-item* belanja yang sering dibeli secara bersamaan dalam suatu waktu. Aturan asosiasi yang berbentuk “*if...then...*” atau “jika...maka...” merupakan pengetahuan yang dihasilkan dari fungsi aturan asosiasi.

2.4 **Algoritma *apriori***

Dalam proses *Data mining* ada beberapa metode yang dapat digunakan. Dalam penelitian ini, peneliti menggunakan metode algoritma *apriori*. Algoritma *apriori* termasuk juga dengan aturan asosiasi pada data mining. Aturan yang menyatakan asosiasi antara beberapa atribut yang sering disebut *affinity analysis* atau *market basket analysis*. Analisis asosiasi atau *association rule* data mining adalah

teknik data mining untuk menemukan aturan suatu kombinasi item set. Salah satu tahap yang banyak digunakan para peneliti untuk menghasilkan algoritma yang lebih singkat adalah analisis pola frekuensi tinggi (*frequent pattern mining*). Pentingnya suatu aturan asosiasi dapat diketahui dengan 2 yaitu : *support* dan *confidence*. *Support* (nilai penunjang) adalah presentase kombinasi item dalam database, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar item dalam aturan asosiasi(Pane, 2013).

Algoritma apriori adalah teknik data mining untuk menemukan aturan asosiasi antara suatu kombinasi item. *Algoritma apriori* sering juga dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik *data mining* lainnya.

Metode dasar analisis asosiasi terbagi menjadi 2 tahap yaitu :

a. Analisis pola frekuensi yang tinggi

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

$$Support (A) = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total Transaksi}} \times 100\% \quad \text{Rumus 2.1 1 Item set}$$

Sementara nilai support dari 2 item diperoleh dengan menggunakan rumus :

$$Support (A, B) = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi}} \times 100\% \quad \text{Rumus 2.2 2 Item set}$$

Frekuensi itemset menunjukkan *itemset* yang akan memiliki frekuensi kemunculan lebih dari nilai minimum yang ditentukan (ϕ). Misalkan (ϕ) = 2 maka semua *itemset* yang fekuensi kemunculannya lebih dari atau sama dengan 2 kali disebut *frequent*. Himpunan dari *frequent* k itemset dilambangkan dengan Fk.

b. Pembentukan aturan assosiatif

Setelah semua pola frekuensi tinggi ditemukan, kemudian dicari aturan *assosiatif* yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan *asosiatif* A- B.

Nilai *confidence* dari aturan A- B diperoleh dari rumus berikut:

$$Confidence = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi mengandung A}}$$

Rumus 2.3 Confidence

Untuk menemukan aturan asosiasi yang akan dipilih maka harus diurutkan berdasarkan *support x confidence*. Aturan diambil sebanyak n aturan nilai yang mempunyai nilai terbesar atau nilai yang paling tinggi.

2.5 Tiket

Tiket adalah suatu dokumen perjalanan yang dikeluarkan oleh suatu perusahaan yang berisi rute, tanggal, harga, data penumpang yang digunakan untuk melakukan perjalanan sesuai dengan tujuan penumpang. Tiket merupakan suatu kontrak pengangkutan antara perusahaan penerbangan dan penumpang,serta tanda

terima dari perusahaan penerbangan kepada sejumlah penumpang atas sejumlah uang yang dibayarkan. Pengertian tiket menurut UU RI No. 1 Tahun 2009, Tiket adalah dokumen yang berbentuk cetak, melalui proses elektronik, atau bentuk lainnya, yang merupakan salah satu bukti adanya perjanjian angkutan antara penumpang.

Seperti halnya tiket di pelabuhan domestik sekupang merupakan faktor utama untuk para penumpang melakukan perjalanan sampai ketujuan. Untuk melakukan perjalanan melalui kapal laut maka penumpang harus mempunyai tiket yang dibeli. Dengan tiket yang sudah ada maka penumpang bisa mengetahui informasi kapan keberangkatan kapal. Tiket merupakan suatu bukti atau indentitas penumpang untuk melakukan *boarding pass*.

2.6 Software pendukung

Tanagra adalah salah satu *software* pada *data mining* yang disediakan banyak metode *data mining* mulai dari perpaduan analisis data, pembelajaran statistik, dan pembelajaran mesin dan database. Tanagra adalah *software* pendukung berbasis *open source* dimana semua orang dapat mengakses *source* kodenya dan menambahkan algoritmanya, biasanya banyak digunakan untuk kepentingan akademik dan riset.

Tanagra adalah perangkat lunak *Data mining* gratis untuk tujuan akademik dan penelitian ini mengusulkan beberapa metode *data mining* dari analisis eksplorasi data, pembelajaran statistik, pembelajaran mesin, dan area basis data. Tanagra adalah "proyek sumber terbuka" dimana setiap peneliti dapat mengakses kode sumber, dan

menambahkan algoritma itu sendiri, sejauh ia setuju dan mematuhi lisensi distribusi perangkat lunak.

Tujuan utama dari proyek Tanagra adalah untuk memberikan para peneliti dan siswa perangkat lunak data mining yang mudah digunakan, sesuai dengan norma-norma yang ada dari pengembangan perangkat lunak dalam domain ini (terutama dalam desain GUI dan cara-cara menggunakannya), dan membuatnya mungkin untuk menganalisis data nyata atau sintetis.

Tujuan kedua Tanagra adalah mengusulkan kepada peneliti arsitektur yang memungkinkan mereka untuk dengan mudah menambahkan metode *data mining* mereka sendiri, untuk membandingkan kinerja mereka. Tanagra bertindak lebih sebagai platform eksperimental untuk membiarkan mereka pergi ke esensi pekerjaan mereka, membelanjakan mereka untuk berurusan dengan bagian yang menyenangkan dalam mengelola data manajemen (Badrul, Studi, & Informasi, 2016).

Tujuan ketiga arahan pengembang pemula, terdiri dalam menyebarkan metodologi yang mungkin untuk membangun perangkat lunak semacam ini. Mereka harus memanfaatkan akses gratis ke kode sumber, untuk melihat bagaimana perangkat lunak semacam ini dibangun, masalah yang harus dihindari, langkah-langkah utama proyek, dan alat dan pustaka kode yang digunakan untuk itu. Dengan cara ini, Tanagra dapat dianggap sebagai alat untuk mempelajari teknik pemrograman.

Sebagian besar yang menggunakan aplikasi Tanagra ini adalah mahasiswa dan peneliti. Dengan menggunakan *software* Tanagra memberikan peneliti dan

mahasiswa lebih mudah digunakan sesuai dengan aturan yang sudah ada di pengembangan *software* untuk menganalisis data yang *real*. Dengan menggunakan Tanagra tidak menggunakan koding atau program khusus, karena semua tools dan platform sudah tersedia di dalam *software* tersebut.

2.7 Penelitian Terdahulu

Dalam penelitian ini, peneliti menggunakan 5 jurnal dari penelitian tentang penjualan menggunakan data mining sebagai sumber referensi

1. Penelitian Mohammad Badrul yang berjudul “*Algoritma asosiasi dengan algoritma apriori untuk analisis data penjualan*” setiap perusahaan atau organisasi itu ingin mempertahankan dan menentukan pola yang benar dalam strategi bisnis. Data penjualan produk yang dibuat perusahaan pada akhirnya akan menghasilkan pengunungan data. Jadi sangat disayangkan jika tidak dianalisis kembali. Produk yang ditawarkan beragam dengan macam produk dan kadang-kadang merek orang untuk membeli produk, untuk mengetahui produk dengan penjualan yang tinggi dan keterkaitan produk yang dibutuhkan. Dalam algoritma apriori diimformasikan dengan aplikasi Tanagra, produk yang muncul secara bersamaan bisa diketahui. Satu tahap analisis asosiasi yang menarik perhatian banyak peneliti menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (pola sering pertambahan). Penting atau tidaknya sebuah asosiasi bisa diidentifikasi oleh dua tolok ukur, yaitu: dukungan dan kepercayaan diri. Dukungan (nilai

dukungan) adalah persentase kombinasi item ini dalam database, sedangkan kepercayaan (nilai kepastian) adalah hubungan yang kuat antara item dalam aturan asosiasi. Algoritma Priori dapat membantu pengembangan strategi pemasaran.

2. Menurut penelitian Khairul Umami yang berjudul “***Analisis Data mining dalam penjualan sparepart mobil dengan menggunakan metode algoritma apriori (Studi kasus: Di PT .IDK 1 MEDAN)***” PT. IDK 1 adalah salah satu Kantor cabang dealer mobil honda yang menjual berbagai jenis varian honda matic atau suku cadang mobil dan motor manual. Setiap penjualan atau barang yang dijual akan dilakukan dengan memasukkan basis data yang terhubung langsung ke kantor pusat. Tetapi PT. IDK 1 tidak tahu beberapa item yang sering dibeli secara bersamaan. Ketika stok suku cadang yang jumlahnya rendah, kantor hanya meminta mereka untuk mengirim stok suku cadang dari kantor pusat tanpa mengetahui bahwa suku cadang lainnya jika suku cadang dibeli maka suku cadang lainnya juga dibeli. Itu dianggap sulit ketika menata ulang barang karena banyaknya jenis onderdil. Teknik *data mining* telah banyak digunakan untuk menyelesaikan masalah yang ada dengan penerapan algoritma satu A-Priori untuk mendapatkan informasi tentang hubungan antara produk dari suatu transaksi basis data. Data transaksi penjualan suku cadang mobil honda di PT. IDK dapat diproses ulang menggunakan aplikasi *data mining* sehingga aturan asosiasi adalah tautan yang kuat antara penjualan *itemset itemset* sehingga bisa memberikan

rekomendasi dan memfasilitasi restocking item dalam pengaturan atau penempatan barang terkait dengan saling ketergantungan yang kuat.

3. Menurut penelitian Wirdah choriah yang berjudul “**Penggunaan *algorithm* *apriori data mining* untuk mengetahui tingkat kesetiaan konsumen (*Brand loyalty*) terhadap merek kendaraan bermotor (studi kasus dealer Honda rumbia)**” Algoritma yang biasa digunakan dalam proses mencari *itemset* sering (data yang paling sering muncul) adalah Apriori. Tetapi algoritma Apriori memiliki kelemahan yang telah memakan waktu lama dalam proses mencari *itemset* yang sering. Dengan memanfaatkan data transaksi konsumen terkait dengan pola loyalitas konsumen terhadap merek kendaraan bermotor Yamaha maka pola hubungan mereka melalui teknik *data mining*, aturan asosiasi. Kategori profesional, jenis kelamin konsumen dan merek kendaraan bermotor pada parameter yang diukur oleh tingkat minat konsumen pada kendaraan merek saat ini. Algoritma yang digunakan adalah algoritma apriori, informasi yang ditampilkan dalam bentuk nilai dukungan dan kepercayaan dari masing-masing kategori.
4. Menurut penelitian andre braga, Filipe portela, Antoni Abelha, alvaro silva yang berjudul “ ***Data mining to predict the use of vasopressors in intensive medicine patients***” Peran yang dimainkan oleh proses pengambilan keputusan dalam Pengobatan Intensif adalah sangat penting karena kondisi kesehatan yang buruk dari pasien yang pergi ke Unit Perawatan Intensif (ICU) dan kebutuhan akan keputusan yang cepat dan akurat. Karena itu setiap keputusan

sangat penting, karena itu dapat membantu menyelamatkan nyawa yang terancam punah. Keputusan harus selalu diambil pada pasien bunga terbaik setelah menganalisis semua data yang tersedia. Di mata para intensivists, selamanya semakin banyak data yang tersedia mengenai pasien, membuatnya semakin banyak sulit bagi mereka untuk mengambil keputusan berdasarkan banyak informasi. Penelitian ini didasarkan pada bukti klinis dan memberikan hasil yang sangat bergunadengan sensitivitas sekitar 90%. Model-model ini akan mengurangi kebutuhan obat vasopressor oleh membantu intensivists untuk bertindak dan mengambil keputusan yang akurat sebelum vasopressor dibutuhkan pasien. Ini akan meningkatkan kondisi pasien karena ketika saatnya tiba Diperkirakan kebutuhan vasopressor akan tidak ada lagi karena perawatan dini yang diberikan oleh intensivist. Keputusan dapat misalnya mengubah rencana terapi. Secara keseluruhan, proses pengambilan keputusan menjadi lebih dapat diandalkan dan efektif serta kualitas perawatan diberikan kepada pasien lebih baik.

5. Menurut Dr. Maruf Pasha, Meherwar Fatima, Abdul Manan Dogar dan Furrakh Shahzad yang berjudul "*Perfomance comparison of data mining algorithms for the predictive accuary of credit card defaulters*" Penggunaan kartu kredit untuk transfer saldo aman adalah kebutuhan waktu. Kegiatan penipuan juga muncul karena pertumbuhan yang cepat transaksi. Motif dari penelitian ini adalah untuk membandingkan akurasi prediksi pembayaran standar pelanggan menggunakan teknik penambangan data yang berbeda.

Akurasi dapat diprediksi di bentuk yang lebih ringkas dari sekedar menggambarkan hasil biner klasifikasi "Credible" atau "Tidak Credible" dalam hal risiko pengelolaan. Biasanya, "defaulters" peluang aktual default adalah gaib. Enam teknik penambangan data (FLDA, Naïve Bayes, J48, Regresi Logistik, MLP, dan IBK) diterapkan pada dataset. Hasil penelitian ini menunjukkan bahwa jaringan syaraf berkinerja terbaik untuk memprediksi standar kartu kredit klien dan menunjukkan akurasi tertinggi.

2.8 Kerangka Pemikiran

Untuk melakukan prediksi terhadap penjualan tiket kapal laut maka perlu dilakukan analisis data. Metode algoritma *apriori* adalah metode yang sering digunakan untuk melakukan prediksi penjualan ataupun analisis data dengan keakuratan yang tinggi. Data yang akan dianalisis adalah data penjualan kapal laut yang domestik. Data penjualan tersebut akan diolah menggunakan aplikasi Tanagra untuk mendapatkan prediksi yang akurat. Dari uraian latar belakang dan metode yang dipakai maka kerangka pemikiran dari penelitian ini adalah:



Gambar 2.3 Kerangka pemikiran

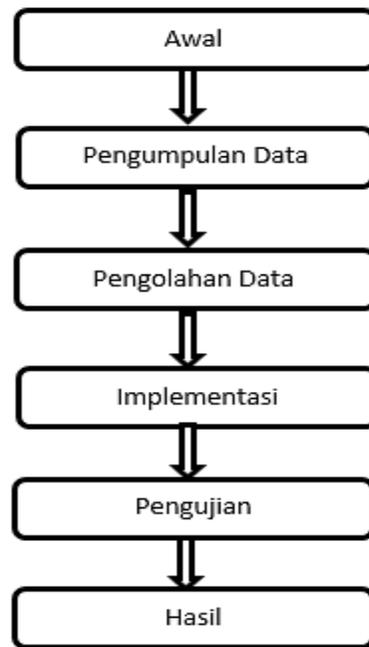
BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Desain penelitian adalah tahapan yang paling penting atau paling utama didalam proses penelitian. Desain penelitian adalah tahapan dimana prosedur atau cara untuk mendapatkan informasi yang diperlukan untuk menyusun dan menyelesaikan masalah dalam penelitian. Tahap penelitian harus mampu menghasilkan gagasan dalam proses penelitian secara mudah. Dalam unsur desain penelitian terdapat tahapan yang diawali mencari ide, tujuan, membuat konsep penelitian, permasalahan, perumusan, sumber informasi yang menjadi sumber pustaka.

Desain penelitian yang sangat berhubungan membuat penelitian lebih mudah memahami seperti hubungan antar variabel, pengumpulan data, analisis data , sehingga dengan adanya desain penelitian yang bagus maka peneliti dan pihak yang terlibat serta orang yang membaca mempunyai gambaran yang jelas dan mudah di mengerti tentang hubungan semua variable yang ada dalam penelitian yang dilakukan oleh seorang peneliti dalam melakukan sebuah penelitian.



Gambar 3.1 Desain Penelitian

Adapun desain yang dilakukan dalam penelitian ini sesuai dengan desain penelitian yang telah dibuat di atas adalah:

1. Awal

Langkah awal yang dilakukan dari penelitian ini adalah mencari dan mengerti masalah yang akan diteliti untuk dijadikan sebagai penelitian. Menentukan ruang lingkup masalah, latar belakang masalah, dan mempelajari beberapa teori yang berhubungan dengan masalah yang diteliti dalam bagiman untuk mencari solusi dari hasil masalah. Untuk mendapatkan tujuan yang diinginkan maka perlu dimengerti beberapa metode yang digunakan dalam penelitian.

2. Pengumpulan data

Metode yang akan digunakan dalam penelitian ini adalah wawancara, observasi dan dokumentasi. Wawancara digunakan untuk bertemu langsung dengan Narasumber untuk mengetahui secara langsung informasi – informasi yang penting dalam penelitian. Observasi digunakan untuk mengamati tempat untuk dijadikan sebagai tempat penelitian yang dilakukan dalam penelitian. Dokumentasi digunakan untuk mencari data mengenai hal- hal dari variabel dalam penelitian.

3. Pengolahan data

Pada tahap pengolahan data terlebih dahulu kita melakukan identifikasi masalah, kemudian mendeskripsikan masalah – masalah tersebut untuk mendapatkan hasil sebagai solusinya. Selanjutnya dilakukan analisi masalah dengan tujuan agar penulis mengetahui dan mendapat gambaran yang jelas untuk penyelesaian dan hasil yang didapatkan dalam penelitian.

4. Implementasi

Pada tahap implementasi ini peneliti dapat dengan mudah menjelaskan hasil analisis yang akan dilakukan, maka penulis menggunakan sebuah aplikasi data mining yang telah ada yaitu aplikasi Tanagra. Hasil tersebut akan diterapkan oleh pelabuhan sekupang, Batam.

5. Pengujian

Pada tahap ini data yang telah dikumpulkan akan diujikan lewat aplikasi Tanagra.

6. Hasil

Sebagai akhir dari penelitian ini akan didapatkan hasil dari data mining yang telah diolah berupa informasi tentang akurasi dari prediksi.

3.2 Pengumpulan Data

3.2.1 Teknik Pengumpulan Data

Menurut (Prof. Dr. Sugiono, 2014; 224) teknik pengumpulan data dilakukan dalam berbagai setting, berbagai sumber dan berbagai Cara. Dari settingannya, data yang dikumpulkan pada setting alamiah. Dari sumber datanya, pengumpulan data dilakukan dengan menggunakan sumber primer dan sumber sekunder. Sumber primer adalah sumber data yang diperoleh langsung dari sumber yang memberikan data kepada pengumpul data, sumber sekunder adalah sumber data yang diperoleh tidak langsung dari yang memberikan data kepada pengumpul data bisa dengan lewat perantara atau lewat dokumen. Dari segi cara, maka teknik pengumpulan data dilakukan dengan observasi, *interview* (wawancara), kuesioner (angket), dokumentasi dan gabungan.

1. Wawancara

Menurut (prof. Dr. Sugiyono, 2014; 231) wawancara adalah teknik pengumpulan data yang dilakukan untuk mendapatkan informasi langsung dari narasumber untuk mengetahui langsung permasalahan yang harus diteliti supaya lebih akurat dan mengetahui responden yang lebih mendalam. Dalam wawancara ini sudah mempersiapkan berupa pertanyaan – pertanyaan yang

akan di tanyakan langsung para narasumber. Dalam penelitian ini, peneliti menggunakan 3 teknik wawancara yaitu;

c. Wawancara terstruktur (*Structured interview*)

Wawancara terstruktur digunakan sebagai teknik pengumpulan data, jika peneliti dan pengumpul data sudah mengetahui pasti informasi yang akan diperoleh. Dalam wawancara ini peneliti harus sudah menyiapkan instrument penelitian berdasarkan pertanyaan- pertanyaan tertulis yang jawabannya pun harus ditulis. Dengan wawancara terstruktur ini setiap responden diberi pertanyaan yang sama dan pengumpul data mencatat.

d. Wawancara semiterstruktur (*semistructure interview*)

Dalam wawancara ini dapat dilakukan pelaksanaan bebas dibandingkan dengan wawancara terstruktur. Tujuan wawancara ini untuk menemukan permasalahan secara terbuka, dimana pihak nara sumber diajak wawancara diminta pendapat dan ide-idenya.

e. Wawancara tidak berstruktur (*unstructured interview*)

Wawancara tidak terstruktur adalah wawancara yang bebas dimana peneliti tidak menggunakan pedoman wawancara yang telah disusun secara lengkap untuk pengumpulan data. Wawancara tidak berstruktur ini sering digunakan dalam penelitian pendahuluan, dalam penelitian pendahuluan ini peneliti berusaha untuk mendapatkan informasi awal tentang permasalahan yang ada pada objek. Dalam wawancara tidak berstruktur peneliti belum mengetahui secara pasti data yang ingin diperoleh, sehingga peneliti lebih banyak

mendengarkan apa yang diceritakan oleh narasumber. Berdasarkan analisis terhadap jawaban setiap narasumber, maka peneliti mengajukan berbagai pertanyaan berikutnya yang lebih terarah pada tujuan.

2. Observasi

Menurut (Dr.sugyono, 2015; 226) observasi adalah suatu teknik dan pengumpulan data yang digunakan untuk mengumpulkan data berdasarkan data yaitu fakta mengenai dunia nyata dengan cara mengamati objek yang ingin diteliti. Dalam observasi ini digunakan dengan observasi partisipatif, observasi terus terang, dan observasi tak berstruktur.

a. Observasi partisipatif

Dalam observasi ini, peneliti terlibat dengan kegiatan orang yang sedang diamati atau yang digunakan sebagai sumber data peneliti. Sambil melakukan pengamatan, peneliti ikut merasakan apa yang dikerjakan oleh sumber data. Dengan observasi partisipatif maka data yang diperoleh akan lebih lengkap dan jelas.

b. Observasi terus terang

Dalam observasi ini peneliti dalam melakukan pengumpulan data dan menyatakan terus terang kepada sumber data, bahwa sedang melakukan penelitian.

c. Observasi tak berstruktur

Observasi tak berstruktur adalah observasi yang tidak dipersiapkan secara lengkap apa yang ingin diobservasi. Hal ini peneliti tidak ikut serta dalam

proses kegiatan yang sedang berlangsung, peneliti hanya berperan mengamati kegiatan.

3. Dokumentasi

Menurut (dr. sudaryono, 2015; 90) dokumentasi digunakan untuk memperoleh data langsung dari tempat penelitian berupa foto, rekaman suara atau video, laporan kegiatan dan data yang valid dengan penelitian. Dengan menggunakan dokumentasi hasil penelitian akan semakin relevan jika didukung dengan adanya bukti seperti foto atau rekaman.

3.3 Operasional Variabel

Operasional variabel adalah suatu cara untuk membantu alat komunikasi antar penelitian secara spesifik untuk digunakan arah atau petunjuk tentang sebuah variabel. Dalam penelitian ini mempunyai variabel yaitu:

1. Data penjualan

Data penjualan adalah data berupa jumlah tiket yang terjual yang terjadi dalam satu bulan yang dijadikan sebagai data. Jumlah penjualan tiket inilah yang mendasari penelitian ini untuk melakukan prediksi penjualan di bulan yang akan datang.

2. Nama Tiket Ferry

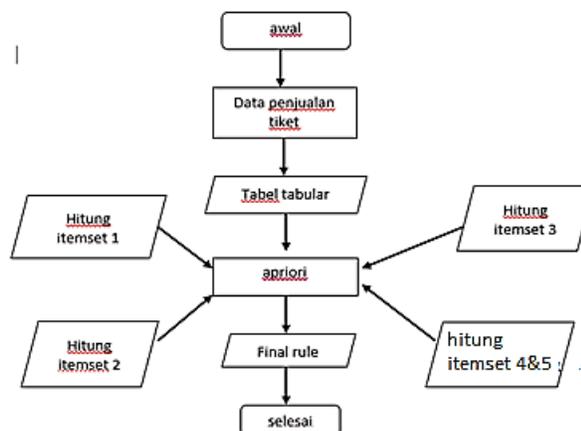
Nama- nama tiket ferry disini adalah nama ferry yang berlayar setiap harinya yang terjual dan dibeli oleh penumpang dan merupakan kapal yang dipilih oleh para penumpang. Berikut daftar ferry tersebut:

Tabel 3.1 Nama ferry

Rahmat Jaya
Mega Fajar
Tenggiri
Bintang Moro
Carolin Express
Budijasa
Batam Jet
Miko Natalia
Kharomah
Kurnia
Tobindo Express
Era Mandiri

3.4 Metode Perancangan Sistem

Adapun metode perancangan yang dibuat dalam penelitian adalah menggunakan metode apriori. Yang digunakan untuk menjelaskan langkah-langkah dalam penyelesaian secara rinci.

**Gambar 3.2** Perancangan Sistem

Sumber: Data Penelitian (2019)

1. Awal

Awal merupakan langkah awal yang digunakan dalam penyelesaian pengolahan data yang sedang diteliti.

2. Data penjualan tiket

Data penjualan tiket adalah data yang transaksi tiket yang diterima sebagai data olahan.

3. Tabel tabular

Data tabular adalah data deskriptif yang menyatakan nilai dari data grafis yang diterangkan. Data ini biasanya berbentuk tabel terdiri dari kolom dan baris. Kolom menyatakan jenis data (*field*), sedangkan baris adalah detail datanya (*record*).

4. *Apriori*

Apriori adalah metode yang dipakai dalam penyusunan penelitian ini sampai menghasilkan prediksi.

5. Hitung *itemset* 1,2,3 dan 4,5

Hitung item set adalah hitung item-item dari *support* (transaksi yang memuat seluruh item) dengan men-scan database untuk *1-itemset*, setelah *1-itemset* didapatkan, dari *1-itemset* apakah diatas minimum *support*, apabila telah memenuhi minimum *support*, *1-itemset* tersebut akan menjadi pola *frequent* tinggi, untuk mendapatkan *2,3,4,5 itemset*, harus dilakukan kombinasi dari *k-itemset* sebelumnya, kemudian *scan database* lagi untuk hitung item-item yang memuat *support*. *Itemset* yang memenuhi minimum *support* akan dipilih

sebagai pola frequent tinggi dari kandidat. Tetapkan nilai *k-itemset* dari *support* yang telah memenuhi minimum *support* dari *k-itemset*.

6. *final rule*

Final rule adalah penentuan confidence dan rule akhir setelah didapatkan kombinasi setiap *itemset*.

7. Selesai

Selesai adalah akhir dari pengolahan data.

3.5 Lokasi dan jadwal penelitian

3.5.1 Lokasi penelitian

Lokasi penelitian ini dilakukan di salah satu pelabuhan dikota batam, tepatnya pelabuhan domestik sekupang yang terletak disekupang.

3.5.2 Jadwal penelitian

Penelitian ini dilakukan Maret 2019 sampai Agustus 2019. Kegiatan yang dilakukan dapat dilihat pada table 3.1 berikut:

Tabel 3.2 Jadwal Penelitian

	kegiatan	Maret			April				Mei				Juni				Juli				
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Input Judul	■	■	■																	
2	BAB 1				■	■	■														
3	BAB 2							■	■	■											
4	BAB 3										■	■	■	■							
5	BAB 4												■	■	■	■	■	■			
6	BAB 5																		■	■	
7	Penyelesaian																				■

Sumber : Data Penelitian (2019)