

**IMPLEMENTASI *DATA MINING* PADA DATA  
PENJUALAN TOKO BANGUNAN  
TUJUH SAMUDRA DENGAN  
ALGORITMA *APRIORI***

**SKRIPSI**



**Oleh:  
Chiomen  
140210144**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
TAHUN 2018**

**IMPLEMENTASI *DATA MINING* PADA DATA  
PENJUALAN TOKO BANGUNAN  
TUJUH SAMUDRA DENGAN  
ALGORITMA *APRIORI***

**SKRIPSI**

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



**Oleh:  
Chiomen  
140210144**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
TAHUN 2018**

## **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, 05 September 2018

Yang membuat pernyataan,

Materai Rp 6.000,00

Chiomen  
140210144

**IMPLEMENTASI *DATA MINING* PADA DATA  
PENJUALAN TOKO BANGUNAN  
TUJUH SAMUDRA DENGAN  
ALGORITMA *APRIORI***

Oleh:  
Chiomen  
140210144

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

Telah disetujui oleh Pembimbing pada tanggal  
seperti tertera di bawah ini

**Batam, 05 September 2018**

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

## ABSTRAK

Data-data yang tersimpan dalam setiap transaksi sangat disayangkan jika tidak dimanfaatkan dengan baik untuk meningkatkan penghasilan dan strategi penjualan. *Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. *Apriori* adalah algoritma yang digunakan dalam melakukan pencarian *frequent itemset* untuk mendapatkan aturan asosiasi. Data-data dalam penelitian bersumber dari toko bangunan Tujuh Samudra. Data *sample* yang diambil adalah 15 *items* dari 14 transaksi pada bulan Januari 2017. Data tersebut kemudian diolah menggunakan algoritma *apriori* dan asosiasi, dimana harus menentukan *itemset-itemset* yang kemungkinan akan menjadi sebuah kombinasi yang sering dibeli oleh konsumen. Hasil dari penelitian ini menunjukkan frekuensi yang memenuhi nilai minimum *ssupport* dan minimum *confidence* adalah 3 jenis *item* saja yaitu kabel 2,5 100% kabel 1,5 100% dan isolasi 100% serta asosiasi final juga membuktikan hasil yang sama. Hasil penelitian ini dinyatakan benar karena telah diuji dengan menggunakan *software* pengujian yaitu Tanagra. *Item-item* inilah yang akan sering dicari oleh konsumen, oleh karena itu agar lebih mempermudah konsumen untuk membeli maka *item-item* tersebut bisa ditempatkan secara berdekatan. Selain sebagai penempatan *item*, pihak toko juga bisa menawarkan *item-item* yang berkemungkinan akan dibeli secara bersamaan serta hasil dari penelitian ini bisa sebagai pengatur untuk persediaan *item*.

**Kata kunci:** *Data Mining*, Algoritma *Apriori* dan Asosiasi, Toko Bangunan.

## **ABSTRACT**

*Databases that was saved from every transaction is very unfortunate if don't used for increase the market strategy and earnings. Data mining is process to search a pattern or interesting information from database which chosen by some method and technique. Technique, method, or algorithm in data mining have many variation. The precise election method or algorithm are hinged by the purpose and KDD processing. Apriori is an algorithm used for searching frequent itemset for getting association rule. Database in research is taken from Tujuh Samudra material shop, using 15 items from 14 transactions on January 2017 for the sample data. The data is processed using apriori algorithm and association, which is need to determine the itemset that possibility one combination and consumer will buy for it. The result of research is indicate fill the frequent minimum support and minimum confidence are Cable 2,5, Cable 1,5 and adhesive tape with 100% for each items. The result was evaluated with Tanagra and it's proved true because the result is the same. This items is always requested by consumer, set the items in one place for making the consumer more easier buy the items. Besides for placement, the seller could offered the items which is the consumer will buy at once and the result from this research could be arrangement for stock.*

**Keywords:** *Data Mining, Apriori Algorithm and Association, Material Shop.*

## 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 Data Penjualan Toko Bangunan Tujuh Samudra dengan Algoritma *A priori*”. 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. Rektor Universitas Putera Batam.
2. Dekan Fakultas Teknik dan Komputer Universitas Putera Batam.
3. Ketua Program Studi Teknik Informatika Universitas Putera Batam.
4. Bapak Koko Handoko, S.Kom., M.Kom. selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
5. Dosen dan staff Universitas Putera Batam.
6. Kedua orang tua yang memberikan semangat, dorongan dan selalu mendoakan penulis.
7. Kepada teman-teman seperjuangan penulis yang sudah memberikan masukan dan semangat dalam penyusunan skripsi ini.

Semoga Tuhan membalas kebaikan dan selalu mencurahkan hidayah serta taufikNya, Amin.

Batam, 04 Agustus 2018

Chiomen



## DAFTAR ISI

<b>HALAMAN SAMPUL DEPAN</b>	
<b>HALAMAN JUDUL</b> .....	ii
<b>HALAMAN 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> .....	xii
<b>DAFTAR RUMUS</b> .....	xiii
<b>DAFTAR LAMPIRAN</b> .....	xiv
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang Penelitian .....	1
1.2 Identifikasi Masalah .....	4
1.3 Pembatasan Masalah .....	4
1.4 Rumusan Masalah .....	5
1.5 Tujuan Penelitian .....	5
1.6 Manfaat Penelitian .....	6
1.6.1 Manfaat Teoritis .....	6
1.6.2 Manfaat Praktis .....	6
<b>BAB II KAJIAN PUSTAKA</b> .....	8
2.1. <i>Knowledge Discovery in Database (KDD)</i> .....	8
2.1.1 Proses <i>Knowledge Discovery in Database</i> .....	9
2.2. <i>Data Mining</i> .....	11
2.2.1 Pengelompokan <i>Data mining</i> .....	13
2.2.2 Arsitektur Sistem <i>Data Mining</i> .....	16
2.3 <i>Algoritma Association</i> .....	17
2.4 <i>Algoritma Apriori</i> .....	20
2.4.1 Tahapan <i>Algoritma Apriori</i> .....	22
2.5 Analisa Keranjang Belanja .....	24
2.6 <i>Software</i> Pendukung .....	24
2.7 Penelitian Terdahulu .....	25
2.8 Kerangka Pemikiran .....	28
<b>BAB III METODE PENELITIAN</b> .....	30
3.1 Desain Penelitian .....	30
3.2 Teknik Pengumpulan Data .....	32
3.3 Operasional Variabel .....	32
3.4 Metode Perancangan Sistem .....	35
3.5 Lokasi dan Jadwal Penelitian .....	36
3.5.1 Lokasi Penelitian .....	36
3.5.2 Jadwal Penelitian .....	37

<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>38</b>
4.1 Analisa Data .....	38
4.1.1 Pemilihan Data/Seleksi Data .....	38
4.1.2 Menentukan <i>Frequent Itemset</i> .....	38
4.1.3 Pengujian Menggunakan Tanagra .....	50
<b>BAB V SIMPULAN DAN SARAN.....</b>	<b>61</b>
5.1 Simpulan.....	61
5.2 Saran.....	61
<b>DAFTAR PUSTAKA.....</b>	<b>63</b>
<b>DAFTAR RIWAYAT HIDUP</b>	
<b>SURAT KETERANGAN PENELITIAN</b>	
<b>LAMPIRAN</b>	

## DAFTAR TABEL

<b>Tabel 3.1</b> Data Penjualan Toko Tujuh Samudra .....	33
<b>Tabel 3.2</b> Jadwal Penelitian .....	37
<b>Tabel 4.1</b> Tabel Tabular Data Transaksi .....	35
<b>Tabel 4.2</b> Representasi Data .....	40
<b>Tabel 4.3</b> Tabel Hasil <i>Support 1-itemset</i> .....	42
<b>Tabel 4.4</b> Hasil dari Kombinasi <i>2-itemset</i> .....	44
<b>Tabel 4.5</b> Hasil Kombinasi <i>Support 3-itemset</i> .....	47
<b>Tabel 4.6</b> Hasil Pembentukan Aturan Asosiasi pada <i>2-itemset</i> .....	48
<b>Tabel 4.7</b> Hasil Pembentukan Aturan Asosiasi pada <i>3-itemset</i> .....	49
<b>Tabel 4.8</b> Hasil Aturan Asosiasi Final .....	50

## DAFTAR GAMBAR

<b>Gambar 2.1</b> Proses <i>Knowledge Discovery in Database</i> .....	9
<b>Gambar 2.2</b> Arsitektur Sistem <i>Data Mining</i> .....	16
<b>Gambar 2.3</b> <i>Flowchart</i> Algoritma <i>Apriori</i> .....	22
<b>Gambar 2.4</b> Kerangka Pemikiran .....	28
<b>Gambar 3.1</b> Desain Penelitian .....	26
<b>Gambar 3.2</b> Perancangan Sistem.....	35
<b>Gambar 4.1</b> Atribut-atribut yang akan di <i>import</i> ke Tanagra .....	51
<b>Gambar 4.2</b> <i>Add-ins</i> ke Tanagra.....	52
<b>Gambar 4.3</b> <i>Execute</i> Tanagra pada <i>Excel</i> .....	52
<b>Gambar 4.4</b> Tampilan <i>File</i> pada Tanagra.....	53
<b>Gambar 4.5</b> Tampilan jendela <i>New</i> pada Tanagra.....	53
<b>Gambar 4.6</b> Tampilan pemilihan <i>File</i> pada Tanagra.....	54
<b>Gambar 4.7</b> Tampilan dari <i>File</i> yang telah dipilih .....	54
<b>Gambar 4.8</b> Tampilan <i>File Dataset</i> .....	55
<b>Gambar 4.9</b> Jendela Parameter <i>Define attribute statuses</i> .....	55
<b>Gambar 4.10</b> Tampilan <i>Dataset</i> yang telah diproses.....	56
<b>Gambar 4.11</b> Jendela <i>Frequent Itemsets</i> .....	57
<b>Gambar 4.12</b> Hasil dari Penentuan <i>Frequent Itemsets</i> .....	57
<b>Gambar 4.13</b> Jendela <i>Association rule parameter</i> .....	58
<b>Gambar 4.14</b> Hasil Pengujian <i>Rule Final</i> .....	59

## DAFTAR RUMUS

<b>Rumus 2.1</b> <i>Perhitungan Support 1-itemset</i> .....	19
<b>Rumus 2.2</b> <i>Perhitungan Support 2-itemset</i> .....	19
<b>Rumus 2.3</b> <i>Perhitungan Confidence 2-itemset</i> .....	20

## DAFTAR LAMPIRAN

**Lampiran 1.** Pembentukan Kombinasi 2-*itemset*

**Lampiran 2.** Tata letak dari hasil penelitian

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Penelitian**

Toko Bangunan merupakan sarana atau tempat yang menyediakan kebutuhan membangun dan merenovasi rumah mulai dari bahan bangunan hingga perlengkapan rumah tangga. Setiap harinya di toko bangunan tersebut terjadi beberapa transaksi penjualan. Data transaksi penjualan akan terus bertambah setiap harinya. Semakin banyak data transaksi yang tersimpan menyebabkan penyimpanan data yang sangat besar. Biasanya data-data transaksi penjualan tersebut hanya digunakan sebagai arsip dan tidak diketahui apa manfaat dari data tersebut untuk selanjutnya.

Dalam kumpulan data yang sangat besar tersebut, memiliki informasi-informasi yang tersembunyi. Kumpulan data tersebut dapat diolah untuk menghasilkan informasi-informasi yang bermanfaat yang bisa digunakan untuk pengambilan suatu keputusan dan untuk memperoleh pengetahuan. Pengolahan data tersebut bisa dilakukan dengan menggunakan teknik-teknik tertentu. Salah satu teknik yang digunakan dalam pengolahan data tersebut adalah dengan menggunakan metode *Algoritma Apriori*. *Data mining* adalah proses untuk mengekstraksi atau menggali pengetahuan dari data yang berjumlah besar. Data berskala besar memunculkan fenomena "*data rich but information poor*", dimana data yang berjumlah besar tidak diikuti dengan peningkatan informasi yang

didapat dari data tersebut. Fenomena itu sering muncul pada sistem penjualan yang dipakai di pasar swalayan, dimana data transaksi terus bertambah setiap hari tetapi hanya menghasilkan laporan transaksi harian yang monoton. Hal ini dikarenakan kurangnya analisa terhadap kumpulan data tersebut (Gama, Putra, & Bayupati, 2016).

*Apriori* adalah algoritma yang digunakan dalam melakukan pencarian *frequent itemset* untuk mendapatkan aturan asosiasi. Sesuai dengan namanya, algoritma ini menggunakan *prior knowledge* mengenai *frequent itemset properties* yang telah diketahui sebelumnya untuk memproses informasi selanjutnya. *Apriori* menggunakan pendekatan secara iterative yang disebut juga sebagai *level-wish search* dimana *k-itemset* digunakan untuk mencari *(k+1)-itemset* (Dita Anggraeni, Saputra, & Noranita, 2012).

Algoritma *Apriori* termasuk jenis aturan asosiasi pada *data mining*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market basket analysis*. Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan suatu kombinasi item. Salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (*frequent pattern mining*). Penting tidaknya suatu asosiasi dapat diketahui dengan dua tolak ukur, yaitu support dan confidence (Pane, 2013).

Toko bangunan Tujuh Samudra merupakan salah satu toko yang bergerak dibidang material bangunan. Toko ini setiap harinya harus memenuhi kebutuhan konsumen, tetapi terjadi beberapa permasalahan dalam proses transaksi dalam



kehidupan sehari-hari, salah satunya adalah penjual tidak bisa menawarkan *item-item* yang kemungkinan akan dibeli secara bersamaan oleh konsumen, sehingga pendapatan untuk toko tersebut tidak bisa maksimal. Untuk dapat mengatasi masalah tersebut, toko ini membutuhkan sumber informasi yang cukup banyak untuk dapat dianalisis lebih lanjut. Misalnya dengan memanfaatkan data transaksi penjualan yang telah tersimpan dalam *database*, pemilik toko dapat mengetahui kebiasaan pelanggan atau perilaku pelanggan mengenai apa saja bahan bangunan yang sering dibeli.

Banyaknya persaingan di dunia bisnis, khususnya dalam industri material bangunan, menuntut pemilik toko untuk menemukan suatu strategi yang dapat meningkatkan penjualan. Agar dapat mengetahui material bangunan apa saja yang dibeli oleh para konsumen, dapat dilakukan dengan teknik analisis yaitu analisis dari kebiasaan membeli konsumen. Pendeteksian mengenai material bangunan yang sering dibeli secara bersamaan dilakukan dengan menggunakan *association rule* (aturan asosiasi), yang mana proses pencarian asosiasi atau hubungan antar item data ini diambil dari suatu basis data relasional. Proses tersebut menggunakan algoritma *Apriori*, yang berfungsi untuk membentuk kandidat kombinasi *item* yang mungkin, lalu diuji apakah kombinasi tersebut memenuhi parameter *minimum support* dan *minimum confidence* yang merupakan nilai ambang yang diberikan oleh pengguna.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang diatas, peneliti mengidentifikasi beberapa masalah yang akan dijadikan bahan penelitian yaitu sebagai berikut:

1. Toko bangunan belum dapat memaksimalkan pemanfaatan data penjual, data penjualan hanya disimpan tanpa dilakukan analisis lebih lanjut. Jumlah data yang begitu besar justru bisa menjadi masalah bila organisasi tersebut tidak bisa memanfaatkannya.
2. Banyaknya persaingan di dunia bisnis, khususnya dalam toko material bangunan, menuntut pemilik toko untuk menemukan suatu strategi yang dapat meningkatkan penjualan.

## 1.3 Pembatasan Masalah

Adapun batasan masalah yang penulis dapat batasi dalam penelitian ini adalah sebagai berikut:

1. Data penjualan yang digunakan adalah data penjualan pada bulan Januari tahun 2017 yang bersumber dari Toko Tujuh Samudra yang lokasinya di Komplek Ruko Aku Tahu Blok A1 No.2, Sungai Panas, Batam.
2. Metode yang digunakan adalah metode asosiasi dengan algoritma *Apriori* untuk mencari pola hubungan pembelian pembeli.
3. Aplikasi yang digunakan untuk menguji hasil penelitian adalah Tanagra. Hasil dari penelitian hanya menganalisis pola hubungan pembeli tanpa membangun *software* aplikasi.

#### 1.4 Rumusan Masalah

Rumusan masalah yang dapat penulis ambil dari penilitan ini adalah sebagai berikut:

1. Bagaimana cara menerapkan algoritma *Apriori* pada toko bangunan Tujuh Samudra berdasarkan bahan bangunan yang paling banyak terjual?
2. Bagaimana caranya agar bisa meningkatkan perkembangan penjualan pada toko bangunan Tujuh Samudra dengan menggunakan algoritma *Apriori* dan asosiasi?

#### 1.5 Tujuan Penelitian

Tujuan penelitian yang dapat diambil dari penulisan ini adalah sebagai berikut:

1. Untuk mengetahui bagaimana cara menerapkan algoritma *Apriori* pada toko bangunan Tujuh Samudra berdasarkan bahan bangunan yang paling banyak terjual.
2. Supaya perkembangan penjualan pada toko bangunan Tujuh Samudra bisa meningkat.
3. Untuk mengetahui cara menggunakan algoritma *Apriori* pada *software* Tanagra.

## **1.6 Manfaat Penelitian**

Adapun manfaat dari penelitian ini yang bisa dimanfaatkan oleh pembaca atau peneliti adalah sebagai berikut:

### **1.6.1 Manfaat Teoritis**

Hasil dari penelitian ini penulis berharap bisa digunakan sebagai bahan referensi untuk peneliti lainnya, serta dapat digunakan untuk menambahkan pengetahuan baru atau sebagai bahan perbandingan sesuai dengan masalah yang akan diteliti.

### **1.6.2 Manfaat Praktis**

Selain manfaat teoritis, penelitian ini juga diharapkan dapat membawakan manfaat praktis seperti dibawah ini:

#### **1. Peneliti**

Hasil penelitian ini diharapkan bermanfaat dan memberikan kontribusi keilmuan pada penelitian bidang *data mining*.

#### **2. Toko Tujuh Samudra**

Hasil penelitian ini diharapkan dapat membantu pemilik toko bangunan Tujuh Samudara untuk memperoleh suatu pengetahuan baru dalam pengambilan keputusan penentuan pola hubungan pembeli dari data penjualan toko.

## **BAB II KAJIAN PUSTAKA**

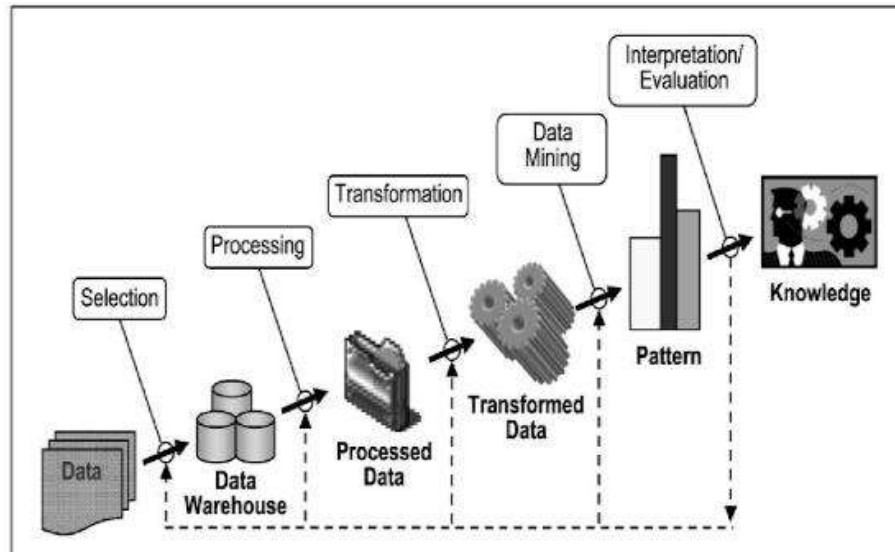
### **2.1. *Knowledge Discovery in Database (KDD)***

*Knowledge discovery in database (KDD)* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Istilah *data mining* dan *Knowledge discovery in database (KDD)* sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar (Choiriah, 2016).

*Knowledge discovery in database (KDD) was formalized in 1989, with reference to the general concept of being broad and high level in the pursuit of seeking knowledge from data. The term data mining was then coined; this high-level application technique is used to present and analyze data for decision-makers* (Pujari, 2001).Pernyataan tersebut dapat diterjemahkan, yaitu *Knowledge Discovery in Database (KDD)* telah disusun pada tahun 1989, yang mengacu kepada konsep umum yang lebih luas dan level tinggi pada pencarian pengetahuan dari data. Istilah *data mining* kemudian dikatakan sebagai teknik aplikasi level tinggi yang digunakan untuk memberikan dan menganalisa data sebagai pembuat keputusan.

### 2.1.1 Proses *Knowledge Discovery in Database*

Proses atau tahapan dari *Knowledge Discovery in Database* bisa dilihat pada gambar berikut ini (Norfiansyah, 2014):



**Gambar 2.1** Proses *Knowledge Discovery in Database*  
Sumber: (Norfiansyah, 2014)

Dari gambar tersebut dapat dijelaskan proses-proses yang terjadi yaitu, antara lain:

#### 1. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, kita memilih data-data seperti apa saja yang kita butuhkan untuk diproses lebih lanjut dan kemudian data disimpan dalam suatu berkas, terpisah dari basis data operasional sehingga memberikan kemudahan untuk penggunaan berikutnya.

## 2. *Pre-processing/cleaning*

Pada umumnya data yang diperoleh, baik dari *database* suatu perusahaan maupun eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil *data mining* nantinya. “*Garbage in garbage out*” (hanya sampah yang akan dihasilkan bila yang dimasukkan juga sampah) merupakan istilah yang sering dipakai untuk menggambarkan tahap ini. Pembersihan data juga mempengaruhi performansi dari sistem *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

## 3. *Transformation*

Beberapa teknik *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan klustering hanya bisa menerima *input-input* kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut *binning*.

Disini juga dilakukan pemilihan data yang diperlukan oleh teknik *data mining* yang dipakai. Transformasi dan pemilihan data ini juga menemukan kualitas dari hasil *data mining* nantinya karena ada beberapa karakteristik dari teknik-teknik *data mining* tertentu yang tergantung pada tahap ini.

#### 4. *Data mining*

*Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses KDD secara keseluruhan. Data yang digunakan untuk bisa menjadi sebuah model yang baik idealnya mencukupi sebagai data riset. Semakin banyak data dan semakin sedikit kesalahan (*error*) maka semakin bagus model yang akan dijadikan patokan.

#### 5. *Interpretation/Evaluation*

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

### **2.2. *Data Mining***

*Data mining* adalah salah satu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* (Kusrini & Luthfi, 2009).



*Data mining* merupakan analisis dari peninjauan kumpulan data untuk menentukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. *Data mining* merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar.” (Kusrini & Luthfi, 2009).

Dari definisi-definisi yang telah disampaikan, hal penting yang terkait dengan *data mining* adalah (Septiana & Dharmayanti, 2016):

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

Kemajuan luar biasa yg terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor, antara lain (Wijayanti, 2017):

1. Pertumbuhan data yang sangat cepat dalam pengumpulan data.
2. Penyimpanan data dalam data warehouse, sehingga seluruh perusahaan memiliki akses kedalam database yang baik.
3. Adanya peningkatan akses data melalui navigasi web dan intranet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.

5. Perkembangan teknologi perangkat lunak untuk data mining (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

### **2.2.1 Pengelompokan *Data mining***

Menurut Larose (2005) *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu sebagai berikut (Kusrini & Luthfi, 2009):

#### 1. Deskripsi

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.

#### 2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik daripada kearah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Contohnya yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pascasarjana dengan melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana.

#### 3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik

yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

#### 4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, pendapatan rendah.

#### 5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran.

Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

#### 6. Asosiasi

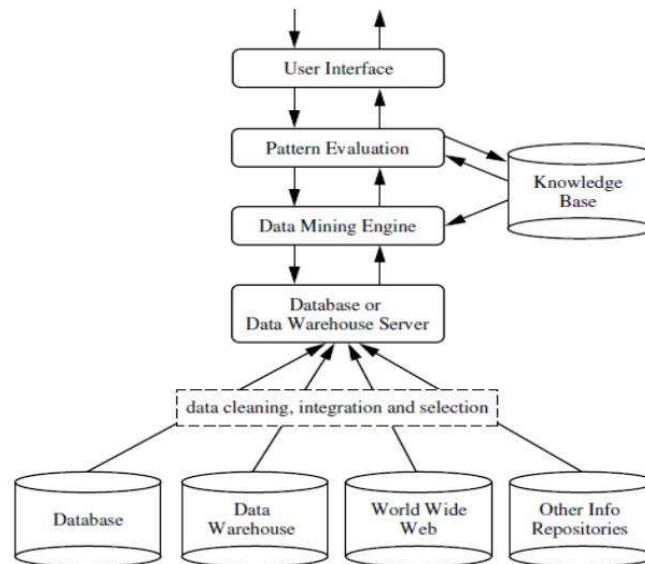
Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

*Market basket analysis* adalah suatu metode analisa atas perilaku konsumen secara spesifik dari suatu golongan/kelompok tertentu. *Market basket analysis* umumnya dimanfaatkan sebagai titik awal pencarian pengetahuan dari suatu transaksi data ketika kita tidak mengetahui pola spesifik apa yang kita cari. Teknik analisa keranjang pasar merupakan teknik yang mengadaptasi ilmu *data mining*. Teknik ini digunakan untuk merancang suatu strategi penjualan dan pemasaran barang melalui proses pencarian asosiasi atau hubungan antar *item* data dari suatu basis data relasional (Badrul, 2016).

Pencarian asosiasi berawal dari pengolahan data transaksi pembelian barang dari setiap pembeli, kemudian dicari hubungan antar barang-barang yang dibeli. Pencarian informasi ini hampir sama dengan mencari peluang kemunculan barang yang dibeli sesuai dengan kebiasaan berbelanja masyarakat dan jumlah transaksi yang ada (Badrul, 2016).

### 2.2.2 Arsitektur Sistem *Data Mining*

Gambar berikut merupakan bagian-bagian dari arsitektur sistem *data mining*, (Dita Anggraeni et al., 2012):



**Gambar 2.2** Arsitektur Sistem *Data Mining*  
Sumber: (Dita Anggraeni et al., 2012)

Penjelasan dari bagian-bagian arsitektur sistem *data mining* diatas adalah sebagai berikut:

1. Basis data, *data warehouse*, atau media penyimpanan lainnya

Media dalam hal ini dapat berupa basis data, *data warehouse*, *spreadsheets*, atau jenis-jenis penampungan informasi lainnya. Pembersihan data, integrasi data, dan seleksi data dilakukan pada bagian tersebut.

2. Server basis data/*data warehouse*

Server basis data/*data warehouse* bertanggung jawab dalam menyediakan data yang relevan berdasarkan permintaan pengguna *data mining*.

### 3. Basis pengetahuan

Pengetahuan yang digunakan dalam pencarian hubungan dari pola yang dihasilkan, seperti *concept hierarchies* digunakan untuk mengorganisasikan nilai atribut atau atribut-atribut ke dalam level abstraksi yang berbeda.

### 4. Mesin *data mining*

Mesin *data mining* merupakan bagian dari perangkat lunak yang menjalankan program berdasarkan algoritma yang ada.

### 5. Model evaluasi pola

Model evaluasi pola merupakan bagian dari perangkat lunak yang berfungsi untuk menemukan pola-pola yang terdapat dalam basis data yang diolah sehingga nantinya proses *data mining* dapat menemukan pengetahuan yang sesuai.

### 6. GUI

Bagian ini merupakan sarana antar pengguna dan sistem *data mining* untuk berkomunikasi, dimana pengguna dapat berinteraksi dengan sistem melalui *data mining query*, untuk menyediakan informasi yang dapat membantu dalam pencarian pengetahuan. Bagian ini mengizinkan pengguna untuk melakukan *browsing* pada basis data dan *data warehouse*, mengevaluasi pola tersebut dengan tampilan yang berbeda-beda.

## 2.3 Algoritma *Association*

Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi item. Analisis asosiasi dikenal juga sebagai salah satu teknik *data mining* yang menjadi dasar dari salah

satu teknik *data mining* lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien, yaitu analisis pola frekuensi tinggi (*frequent pattern mining*) (Fauzy, Rahmat Saleh, & Asror, 2016).

Algoritma asosiasi merupakan suatu bentuk algoritma dalam *data mining* yang memberikan informasi hubungan antar *item* data *didatabase*. Algoritma tersebut dapat dimanfaatkan secara luas dalam proses bisnis diantaranya dalam proses penjualan. Data mining algoritma asosiasi dapat membantu dalam proses penjualan dengan memberikan hubungan antar data penjualan yang dilakukan pelanggan sehingga akan didapat pola pembelian pelanggan. Pebisnis dapat memanfaatkan informasi tersebut untuk mengambil tindakan bisnis yang sesuai. Di dalam bidang usaha retail, metode *association rule mining* ini lebih dikenal dengan istilah analisa keranjang belanja atau *market basket analysis* (Badrul, 2016).

Dalam menentukan suatu aturan asosiasi, terdapat suatu ukuran ketertarikan (*interestingness measure*) yang didapatkan dari hasil pengolahan data dengan data perhitungan tertentu. Pada umumnya terdapat dua ukuran ketertarikan dalam aturan asosiasi, yaitu (Dita Anggraeni et al., 2012):

1. *Support*

*Support* adalah probabilitas konsumen membeli beberapa produk secara bersamaan dari jumlah seluruh transaksi. Ukuran ini menentukan apakah suatu *item/itemset* layak untuk dicari nilai *confidence*-nya (misal dari keseluruhan

transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan bahwa *item X* dan *Y* dibeli secara bersamaan.

## 2. *Confidence*

*Confidence* atau tingkat kepercayaan merupakan probabilitas kejadian beberapa produk yang dibeli bersamaan dimana salah satu produk sudah pasti dibeli (misal, seberapa sering *item Y* dibeli apabila konsumen membeli *item X*).

Metode dasar analisis asosiasi terbagi menjadi dua tahap, (Vulandari, 2017):

### 1. Analisa pola frekuensi tinggi

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

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

#### **Rumus 2.1** Perhitungan *Support 1-itemset*

Sumber: (Vulandari, 2017)

Sedangkan nilai dari *support* dari 2 *item* diperoleh dari rumus berikut :

$$\text{Support (A} \cap \text{B)} = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Total transaksi}}$$

#### **Rumus 2.2** Perhitungan *Support 2-itemset*

Sumber: (Vulandari, 2017)

### 2. Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung



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

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

**Rumus 2.3** Perhitungan *Confidence 2-itemset*

Sumber: (Vulandari, 2017)

## 2.4 Algoritma *Apriori*

Algoritma *Apriori* adalah suatu algoritma dasar yang diusulkan oleh Agrawal & Srikant pada tahun 1994 untuk menentukan *frequent itemset* untuk aturan asosiasi *boolean*. Algoritma *Apriori* termasuk jenis aturan asosiasi pada *data mining* (Sutradana & Wahyudi, 2017).

Algoritma *Apriori* menggunakan pengetahuan frekuensi atribut yang telah diketahui sebelumnya untuk memproses informasi selanjutnya. Pada algoritma *Apriori*, menentukan kandidat yang mungkin muncul dengan cara memperhatikan *minimum support* dan *minimum confidence*. Proses utama yang dilakukan dalam algoritma *Apriori* untuk mendapat *frequent itemset* sebagai berikut, (Widiati & Dewi, 2014):

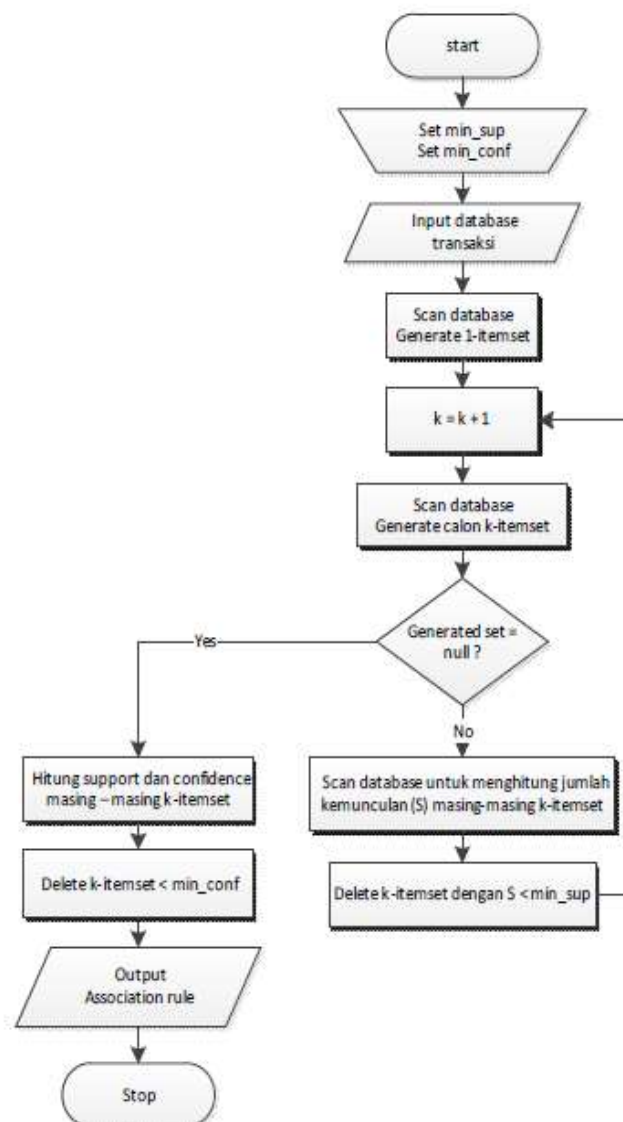
1. *Join* (Penggabungan). Proses ini dilakukan dengan cara pengkombinasian *item* dengan *item* yang lainnya hingga tidak dapat terbentuk kombinasi lagi.
2. *Prune* (Pemangkasan). Proses pemangkasan yaitu hasil dari *item* yang telah dikombinasikan kemudian dipangkas dengan menggunakan *minimum support* yang telah ditentukan oleh *user*. Pada iterasi ke- $k$  akan ditemukan semua *itemset*

yang memiliki  $k$  *item*, disebut dengan  $k$ -*itemset*. Tiap iterasi terdiri dari dua tahap, yaitu :

- a. Gunakan *frequent*  $(k-1)$  *itemset* untuk membangun kandidat *frequent*  $k$ -*itemset*.
- b. Gunakan *scan database* dan pencocokan pola untuk mengumpulkan hitungan untuk kandidat *itemset*.

### 2.4.1 Tahapan Algoritma *Apriori*

Algoritma *Apriori* dibagi menjadi beberapa tahap yang digambarkan melalui *flowchart* atau bagan alir untuk menemukan *frequent itemset* yang tersembunyi pada *database*, seperti berikut ini (Gama et al., 2016):



**Gambar 2.3** *Flowchart* Algoritma *Apriori*  
Sumber: (Gama et al., 2016)

Pada *flowchart* tersebut terdapat beberapa tahapan untuk menemukan *frequent itemset* yang tersembunyi pada database, yaitu:

1. *Apriori* menggunakan pendekatan iteratif dimana *k-itemset* digunakan untuk mengeksplorasi *(k+1)-itemset*.
2. Calon *(k+1)-itemset* dengan frekuensi yang jarang muncul atau dibawah *threshold (min\_support)* akan dipangkas dan tidak dipakai dalam menentukan aturan asosiasi.
3. Langkah pertama adalah *1-itemset* ditemukan dengan melakukan *scanning database* untuk mengakumulasi jumlah dari setiap kategori *item* dan kemunculannya pada setiap transaksi.
4. Selanjutnya *1-itemset* digunakan untuk menemukan *2-itemset*. Calon kandidat *2-itemset* ditemukan dengan cara saling memasangkan satu *item* dengan *item* yang lainnya sehingga didapatkan kombinasi yang memungkinkan untuk 2 buah *item*.
5. *2-itemset* kemudian dihitung nilai kemunculannya pada setiap transaksi. Nilai ambang (*min\_support*) ditentukan untuk memangkas calon kandidat yang dianggap tidak sering muncul atau tidak *frequent*.
6. *2-itemset* yang memenuhi nilai ambang akan dihitung nilai *support* dan *confidence*-nya. *2-itemset* yang memenuhi *min-support* dan *min\_confidence* akan dijadikan aturan asosiasi.
7. *2-itemset* digunakan untuk menemukan *3-itemset* dan seterusnya sampai tidak ada lagi *frequent (k+1)- itemset* yang bisa ditemukan.

8. Setelah semua aturan dari *frequent (k+1)-itemset* terbentuk, kemudian dihitung nilai *support* dan nilai *confidence*. Hasil perkalian dari nilai *support* dan *confidence* yang paling tinggi merupakan aturan asosiasi yang paling baik dari keseluruhan transaksi yang ada dalam.

## **2.5 Analisa Keranjang Belanja**

Menurut David Olson dan Yong Shi (2008), analisis keranjang belanja mengacu pada berbagai teknologi yang mempelajari komposisi keranjang belanja yang terdiri atas produk-produk yang dibeli pada satu kejadian belanja. Teknik ini telah diterapkan secara luas dalam berbagai operasi pasar swalayan. Data keranjang belanja dalam bentuknya yang paling mentah adalah daftar transaksi pembelian oleh pelanggan, yang mengindikasikan hanya barang yang dibeli bersamaan (Listriani, Setyaningrum, & Eka, 2016).

## **2.6 Software Pendukung**

Didalam *data mining* kita bisa menggunakan berbagai *software* untuk mendukung *data mining* dalam mengambil keputusan sesuai dengan tujuan dan proses dari data tersebut, oleh karena itu pada penelitian ini, peneliti akan menggunakan *software* Tanagra. Tanagra adalah *software data mining* bebas untuk tujuan akademik dan penelitian ini mengusulkan beberapa metode *data mining* dari analisis eksplorasi data, pembelajaran statistik, pembelajaran mesin dan daerah *database*. Tanagra adalah "proyek *open source*" karena setiap peneliti

dapat mengakses ke kode sumber, dan menambahkan algoritma sendiri, sejauh dia setuju dan sesuai dengan lisensi distribusi perangkat lunak (Badrul, 2016).

## 2.7 Penelitian Terdahulu

Berikut merupakan penelitian-penelitian terkait yang telah dibuat oleh peneliti sebelumnya, yaitu:

1. Penelitian yang dilakukan oleh (Sutradana & Wahyudi, 2017) yang berjudul ***PENERAPAN DATA MINING UNTUK ANALISIS PENGARUH LAMA STUDI MAHASISWA TEKNIK INFORMATIKA UIN SUNAN KALIJAGA YOGYAKARTA MENGGUNAKAN METODE APRIORI***, pada penelitian ini membahas tentang berapa lama mahasiswa menyelesaikan studinya dan untuk mengetahui apa penyebab mahasiswa tersebut bisa menyelesaikan studinya dalam waktu yang cepat atau lambat. Penelitian ini menentukan pola kombinasi *itemset* paling tinggi yaitu asal sekolah, pekerjaan orang tua, kelas dan kategori ipk lalu ditentukan aturan asosiasi untuk menghitung nilai *support* sebanyak 7% dan *confidence* sebanyak 40%.
2. Penelitian yang dilakukan oleh (Gama et al., 2016) yang berjudul ***IMPLEMENTASI ALGORITMA APRIORI UNTUK MENEMUKAN FREQUENT ITEMSET DALAM KERANJANG BELANJA***, pada penelitian ini membahas tentang *item-item* yang tersembunyi pada *database* transaksi atau keranjang belanja. Sampel data yang dipakai adalah 100 transaksi yang sudah melalui tahap *preprocessing* dengan membersihkan data yang rusak sehingga

hasil yang didapatkan adalah 2 *itemset* dengan nilai *support* 2,00% dan *confidence* 100,00% yang bermaksud kedua *itemset* tersebut banyak dibeli oleh pelanggan.

3. Penelitian yang dilakukan oleh (Fauzy et al., 2016) yang berjudul ***PENERAPAN METODE ASSOCIATION RULE MENGGUNAKAN ALGORITMA APRIORI PADA SIMULASI PREDIKSI HUJAN WILAYAH KOTA BANDUNG***, penelitian yang dibahas yaitu mengenai prakiraan cuaca saat ini dan esok harinya, dengan teknik *data mining* yang cepat namun akurat bisa menyelesaikan masalah ini. Variabel yang digunakan yaitu suhu, rata kelembapan, rata kecepatan angin, arah angin dan curah hujan serta *dataset* yang digunakan adalah data klimatologi. Hasil dari penelitian ini yaitu ternyata metode *association rule mining* dengan menggunakan algoritma *Apriori* dapat diterapkan pada sistem simulasi prediksi hujan.

4. Penelitian yang dilakukan oleh (Sitorus, 2016) yang berjudul ***PENERAPAN ALGORITMA APRIORI PADA DATA MINING UNTUK PENJUALAN TIKET BUS ( STUDI KASUS PT. ANTAR LINTAS SUMATERA )***, membahas tentang bagaimana pola penjualan tiket bus agar bisa menentukan tiket bus manakah yang paling sering dibeli oleh konsumen dengan menggunakan Algoritma *Apriori*. Penelitian ini membuktikan jika Algoritma *Apriori* bisa membantu mengembangkan strategi pemasaran serta dengan menggunakan *association rules* bisa menghasilkan pola kombinasi *itemsets* dari penjualan tiket bus.

5. Penelitian yang dilakukan oleh (Retnosari & Jananto, 2013) yang berjudul ***IMPLEMENTASI DATA MINING UNTUK MENEMUKAN HUBUNGAN***

**ANTARA KOTA KELAHIRAN MAHASISWA DENGAN TINGKAT KELULUSAN MAHASISWA PADA FAKULTAS TEKNOLOGI INFORMASI UNISBANK**, disini peneliti meneliti tentang bagaimana pengaruh kota kelahiran mahasiswa terhadap tingkat kelulusannya, dimana perhitungan dilakukan dengan tiga cara, yaitu secara manual, MS. Excel dengan *software* Tanagra dan *software* RapidMiner. Hasil dari penelitian ini yaitu, *software* RapidMiner dapat mengolah bilangan minimum *support* kurang dari 1,0 sampai tiga angka dibelakang koma, sedangkan Tanagra hanya dapat mengolah sampai dua angka dibelakang koma, hasil penelitian lainnya adalah bisa digunakan sebagai pertimbangan untuk pihak universitas melakukan promosi mahasiswa baru untuk kedepannya.

6. Penelitian yang dilakukan oleh (Putra & Wadisman, 2018) yang berjudul **IMPLEMENTATION OF DATA MINING FOR POTENTIAL CUSTOMER SELECTION USING K-MEANS ALGORITHM**, peneliti membahas tentang usaha suatu perusahaan yang memiliki banyak persaingan bisa mempertahankan pelanggannya, hal ini bisa dicegah dengan melakukan penelitian untuk mengetahui kelompok pelanggan yang berpotensi sehingga layak untuk perusahaan lindungi dengan cara memberikan pelayanan prima. Variabel yang digunakan yaitu data nama pelanggan, data jumlah transaksi dan data total transaksi sebanyak 20 *sample* data dalam bentuk format *excel*. Hasil dari penelitian ini adalah data dengan pusat *centroid* terbesar yang bisa termasuk ke dalam pelanggan yang paling berpotensi.

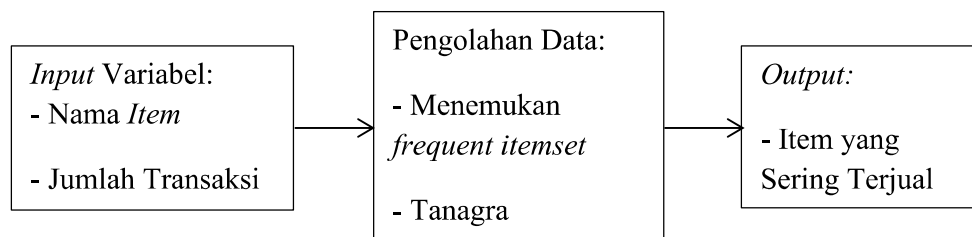
7. Penelitian yang dilakukan oleh (Pane, 2013) dengan judul **IMPLEMENTASI DATA MINING PADA PENJUALAN PRODUK**



***ELEKTRONIK DENGAN ALGORITMA APRIORI (STUDI KASUS : KREDITPLUS)***, penelitian ini membahas tentang penjualan produk elektronik yang paling banyak terjual dan dengan menggunakan algoritma *Apriori* apakah bisa membantu mengembangkan strategi pemasaran. Penelitian ini berhasil membuktikan jika merek Acer dan Toshiba adalah produk yang paling banyak terjual, yaitu dengan aturan asosiasi akhirnya jika membeli Acer, maka akan membeli Toshiba nilai *support* 50% dan *confidence* 66,67% dan jika membeli Toshiba, maka akan membeli Acer nilai *support* 50% dan nilai *confidence* 85,741%.

## 2.8 Kerangka Pemikiran

Kerangka pemikiran yang dapat penulis gambarkan untuk penelitian ini bisa kita lihat pada gambar berikut:



**Gambar 2.4** Kerangka Pemikiran  
Sumber: Data olahan peneliti, 2018

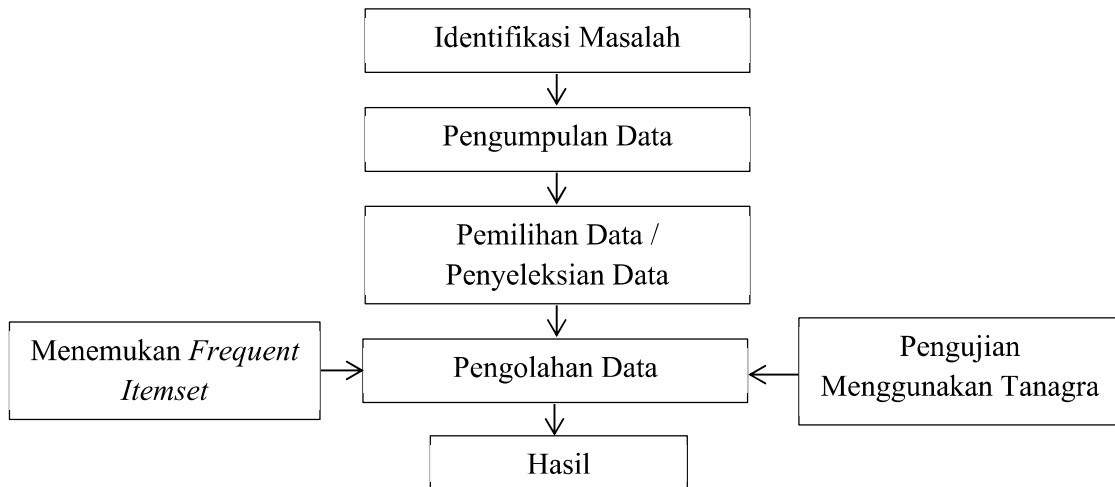
Pada gambar 2.2, dapat dijelaskan bahwa proses awal dari penelitian ini adalah adanya variabel yang telah dipilih berdasarkan nama *item* dan jumlah transaksi akan diseleksi terlebih dahulu, agar tidak ada tipografi. Selanjutnya data ini akan diproses dengan menemukan *frequent itemset* lalu diuji dengan Tanagra

untuk menemukan *frequent itemset* tertinggi dan untuk menentukan apakah hasil dari *software* pengujian sama dengan cara manual. Hasil akhir yaitu bisa mengetahui *item* yang sering terjual, pada hasil akhir inilah yang akan diterapkan pada toko tersebut yaitu dengan menata letak *item-item* yang telah dikombinasikan secara berdekatan serta sebagai pemilihan *stock* untuk kedepannya.

## BAB III METODE PENELITIAN

### 3.1 Desain Penelitian

Desain penelitian sangat dibutuhkan dalam suatu penelitian, hal ini dilakukan agar penelitian bisa berjalan dengan baik. Desain penelitian yang peneliti lakukan adalah sebagai berikut:



**Gambar 3.1 Desain Penelitian**  
Sumber: Data olahan penulis, 2018

Dari desain penelitian yang telah peneliti gambarkan diatas, dapat peneliti jelaskan sebagai berikut:

1. Identifikasi Masalah

Menjelaskan tentang apa saja yang menjadi permasalahan dalam penelitian yang dilakukan oleh peneliti.

## 2. Pengumpulan Data

Data yang dapat diambil oleh peneliti adalah data observasi langsung dari pihak toko bangunan Tujuh Samudra serta materi-materi studi kepustakaan didapatkan dari referensi jurnal nasional maupun internasional serta buku-buku yang berhubungan dengan *data mining* dan algoritma *Apriori* dan asosiasi.

## 3. Pemilihan Data / Penyeleksian Data

Data-data yang diperoleh akan dipilih atau diseleksi terlebih dahulu, proses ini dilakukan untuk menentukan data *sample* mana yang akan digunakan untuk menjadi bahan penelitian serta supaya data yang telah dipilih tidak ada yang duplikat, menghilangkan kesalahan pada data (tipografi) dan menghilangkan data yang kosong.

## 4. Pengolahan data

Data yang telah dipilih dan diseleksi akan diproses dengan cara algoritma *Apriori* yaitu secara manual dengan menemukan item *frequent-set*-nya lalu akan diuji kembali dengan menggunakan Tanagra.

## 5. Hasil penelitian

Setelah beberapa proses yang dilakukan pada penelitian tentu saja akan memberikan sebuah hasil yang peneliti cari. Hasil penelitian ini akan diuji menggunakan *software* Tanagra, untuk menentukan apakah hasil yang dicari sesuai dengan hasil pengujian.

### 3.2 Teknik Pengumpulan Data

Teknik pengumpulan data yang peneliti lakukan pada penelitian ini adalah sebagai berikut:

#### 1. Kajian Pustaka

Pada penelitian ini kajian pustaka yang dilakukan adalah dengan mengumpulkan materi-materi dari berbagai sumber seperti buku-buku mengenai *data mining*, serta jurnal-jurnal para senior yang telah diakui secara nasional maupun internasional.

#### 2. Observasi

Data-data yang peneliti gunakan dalam penelitian ini adalah data-data yang diperoleh dari pengamatan langsung ke lokasi yang akan diteliti, data inilah yang akan diproses untuk mencari nilai yang ingin diselesaikan.

### 3.3 Operasional Variabel

Variabel penelitian pada dasarnya adalah segala sesuatu yang berbentuk apa saja yang ditetapkan oleh peneliti untuk dipelajari sehingga diperoleh informasi dan kesimpulannya (Sudaryono, 2015).

Penjelasan dari variabel-variabel pada penelitian ini adalah, sebagai berikut:

1. Jumlah Transaksi, yaitu jumlah transaksi yang terjadi pada satu bulan tersebut. Jumlah transaksi per-*item* inilah yang sangat mempengaruhi penelitian ini, karena ia yang akan menentukan kombinasi-kombinasi *itemset* mana yang mungkin terjadi sebuah kombinasi.

2. Nama Item, yaitu nama-nama *item* yang sering dibeli oleh pembeli dan merupakan *item-item* yang akan dikombinasikan.

Data-data yang akan digunakan oleh penulis untuk diolah dalam penelitian ini bisa kita lihat pada tabel 3.1 seperti berikut:

**Tabel 3. 1** Data Penjualan Toko Tujuh Samudra

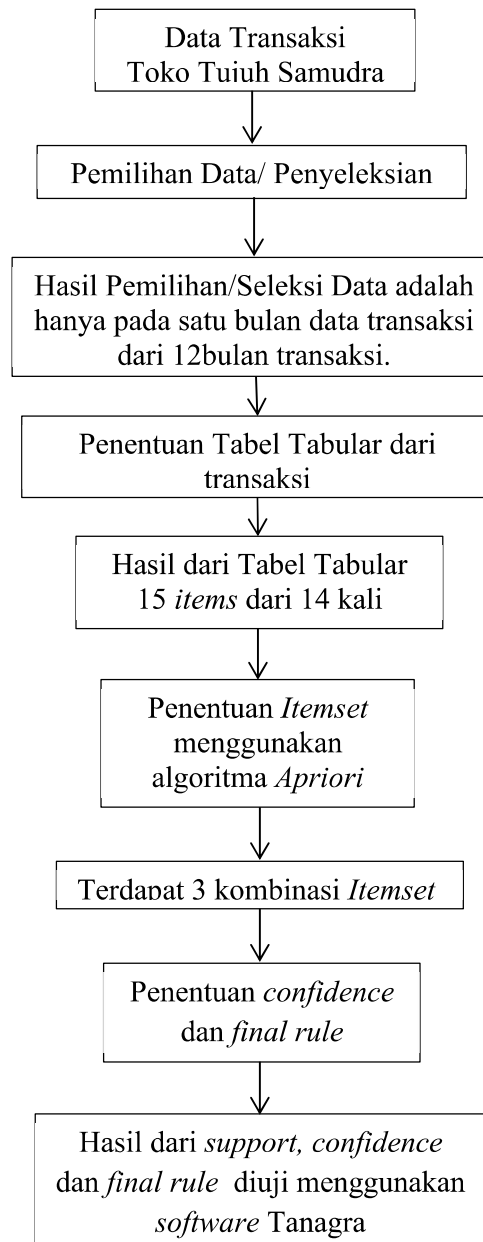
<b>Transaksi</b>	<b>Tanggal</b>	<b>Nama Item</b>
1	02/01/2017	Semen 40kg
		Bata Merah
		Pasir
2	06/01/2017	Seng Partisi
		Furing
		Stoping
3	08/01/2017	Kawat Beton
		Pasir
4	10/01/2017	Furing
		Stoping
		Ramset
5	12/01/2017	Kawat Beton
		Stoping
		Skrup
		Kabel2,5
		Kabel1,5
		Isolasi
6	15/01/2017	Ramset
		Kawat Beton
		Bata Merah
		Semen 40kg
7	16/01/2017	Ramset
		Furing
8	18/01/2017	Bata Merah
		Pasir
9	19/01/2017	Kabel Ties
		Kawat Beton
		Semen 40kg
10	24/01/2017	Kabel2,5
		Kabel1,5

<b>Transaksi</b>	<b>Tanggal</b>	<b>Nama Item</b>
		Kabel Ties
		Isolasi
11	25/01/2017	Kabel2,5
		Kabel1,5
12	26/01/2017	Stoping
		Skrup
		Furing
		BoxSaklar
13	27/01/2017	Isolasi
		Kabel2,5
		Kabel1,5
		BoxSaklar
14	31/01/2017	Kuas
		Semen 40kg

Sumber: Data olahan penulis, 2018

### 3.4 Metode Perancangan Sistem

Metode perancangan sistem yang terjadi pada penelitian ini adalah sebagai berikut:



**Gambar 3. 2** Perancangan Sistem  
Sumber: Data olahan penulis, 2018



Dari gambar 3.2 dapat diketahui bahwa proses dimulai dari pengumpulan data yang telah diseleksi dari pihak toko, data yang dikumpul yaitu hanya satu bulan transaksi dalam satu tahun. Kemudian data tersebut di format ke tabel tabular, hasil dari tabel tabular ini adalah terdapat 14 kali transaksi dalam satu bulan dan sebanyak 15 *item* yang terjual. Selanjutnya data-data ini akan di cari nilai *support*, *confidence* dan *final rule*-nya secara manual menggunakan algoritma *Apriori*. Setelah itu, akan diuji menggunakan Tanagra, sebagai pembuktian apakah hasil manual tersebut sesuai dengan hasil penguji dari *software* Tanagra.

### **3.5 Lokasi dan Jadwal Penelitian**

Pada penelitian ini lokasi dan jadwal yang dilakukan adalah sebagai berikut:

#### **3.5.1 Lokasi Penelitian**

Penelitian dilakukan di toko bangunan Tujuh Samudra yang berlokasi di Komplek Ruko Aku Tahu Blok A1 No.2, Sungai Panas, Batam.

### 3.5.2 Jadwal Penelitian

Jadwal penelitian ini dilakukan mulai dari bulan Maret hingga bulan Juli 2018, dengan rincian sebagai berikut:

**Tabel 3. 2** Jadwal Penelitian

No.	Kegiatan	Waktu Penelitian																						
		Maret 2018				April 2018				Mei 2018				Juni 2018				Juli 2018						
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4			
1	Bimbingan				■									■	■						■	■	■	■
2	Menentukan Judul Penelitian	■	■	■																				
3	Studi Kepustakaan/ Pengumpulan Materi				■	■	■	■	■	■	■	■	■											
4	Pengumpulan Data													■	■	■	■	■	■					
5	Pengolahan Data																			■	■	■	■	■
	Hasil Penelitian																							■

Sumber: Data olahan penulis, 2018