

**IMPLEMENTASI *DATA MINING*
PADA PENJUALAN TAS MENGGUNAKAN
ALGORITMA *APRIORI***

SKRIPSI



**Oleh:
Amelia Sitorus
150210079**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS PUTERA BATAM

TAHUN 2020

**IMPLEMENTASI *DATA MINING*
PADA PENJUALAN TAS MENGGUNAKAN
ALGORITMA *APRIORI***

SKRIPSI

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



**Oleh:
Amelia Sitorus
150210079**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS PUTERA BATAM

TAHUN 2020

PERNYATAAN

Yang Bertanda tangan di bawah saya :

Nama : Amelia Sitorus

NPM : 150210079

Fakultas : Teknik

Program Studi : Teknik Informatika

Menyatakan bahwa “SKRIPSI” yang saya buat dengan judul:

**IMPLEMENTASI DATA MINING PADA PENJUALAN TAS
MENGUNAKAN ALGORITMA APRIORI**

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

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

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 20 Februari 2020
Yang membuat pernyataan,

Amelia Sitorus
150210079

**IMPLEMENTASI *DATA MINING*
PADA PENJUALAN TAS MENGGUNAKAN
ALGORITMA *APRIORI***

SKRIPSI

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

Oleh:

Amelia Sitorus

150210079

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

Batam, 20 Februari 2020

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

ABSTRAK

Toko tas jc jolin *collection* merupakan salah satu toko tas di kota Batam yang memanfaatkan penjualan tas di toko maupun media sosial. Dengan sistem penjualan tersebut tentu meningkatkan tingkat penjualan tas yang kadang tinggi dan juga kadang rendah, untuk Toko Jc jolin *Colecction* sendiri penjualan tas bisa mencapai 348 pcs sampai 605 pcs tas perbulanya. Hal tersebut cukup sulit diketahui pemilik toko tas dengan pencatatan setiap penjualan tas masih bersifat manual sehingga penjualannya tidak maksimal dimana pemilik toko tidak mengetahui merek tas mana saja yang kehabisan *stock*. Untuk itu dilakukan penelitian ini guna menganalisa data transaksi penjualan tas tersebut dengan menggunakan teknik *data mining*. *Data mining* ialah teknik analisa pada data untuk mencari informasi yang tersembunyi pada beberapa data besar yang dicatat pada saat bisnis yang dijalankan. Algoritma yang dipakai algoritma *apriori* yaitu mendapatkan *accosiation rule* dari setiap kombinasi *item*. *Data sample* yang dipakai yakni 14 *item* dari 12 kali transaksi dalam satu tahun yaitu tahun 2018. Pada penelitian menggunakan nilai minimum *support* 50% dengan *minimum confidence* 80%. yang menghasilkan *frequent 1-itemset*, *frequent 2-itemset* dan *frequent 3-itemset* dengan total *final rule* yaitu 19 *rule* dengan aplikasi pengujian pada penelitian ini yakni aplikasi Tanagra 1.4. Sedangkan untuk *item* yang paling tinggi tingkat penjualannya ada Furla, LV, Prada, MV , *Dior*, *Jeep*, *Gucci* dan *Coach* dengan masing-masing *confidence* 80%-100%. Hasil penelitian ini baik pengolahan secara manual dan pengujian menggunakan aplikasi tangara sama, baik dari *accoasiation rule* maupun untuk *final rulanya*.

Kata Kunci: *Data mining*, Tas, Data Transaksi Penjualan Tas, Algoritma *Apriori*

ABSTRACT

Jc jolin bag store is one of the bag shops in Batam that utilizes bag sales at stores and social media. With this sales system, it certainly increases the level of sales of bags which are sometimes high and sometimes low, for Jc Jolin Colecction Stores, bag sales can reach 348 pcs to 605 pcs per month. It is quite difficult to know the bag shop owner by recording each bag sale is still manual so that sales are not optimal where the shop owner does not know which brand of bag is out of stock. For this reason, this research is conducted in order to analyze the transaction data of the sale of the bag using data mining techniques. Data mining is a data analysis technique to find information hidden in some big data that is recorded when a business is run. The algorithm used a priori algorithm is obtaining the acceleration rule of each combination of items. The sample data used are 14 items from 12 transactions in one year, 2018. In this study, using a minimum support value of 50% with a minimum confidence of 80%. which results in frequent 1-itemset, frequent 2-itemset and frequent 3-itemset with a total final rule of 19 rules with the application of examiners in this study namely the Tanagra 1.4 application. Whereas the highest selling items are Furla, LV, Prada, MV, Dior, Jeep, Gucci and Coach with 80% -100% confidence respectively. The results of this study are both manual processing and testing using the same tangara application, both from the accoasiation rule and for the final rule.

Keywords: Data mining, Bags, Data Sales Transaction Bags, Apriori Algorithms

KATA PENGANTAR

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

Penulis menyadari bahwa penulisan skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai piha. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terimakasih kepada:

1. Rektor Universitas Putera Batam.
2. Bapak Andi Maslan, S.T.,M.Si. Selaku Ketua Program Studi Teknik Informatika.
3. Ibu Yulia, S.Kom., M.Kom selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
4. Dosen dan Staff Universitas Putera Batam.
5. Kedua Orang Tua yang selalu mendoakan dan memberikan semangat dari segi materi maupun non materi kepada penulis.
6. Adek Wensi Sitorus dan Ermila Sitorus yang mendukung saya baik dari segi materi maupun non materi.
7. Kepada teman saya Wida Estaria Sitinjak yang menjadi teman begadang dalam penyusun skripsi masing-masing.
8. Kepada teman-teman seperjuangan Program Studi Teknik Informatika Universitas Putera Batam.

Semoga Tuhan membalas setiap kebaikan dan selalu mencurahkan berkat dan rahmat-NYA. Amin

Batam, 20 Februari 2020

Amelia Sitorus

DAFTAR ISI

PERNYATAAN	iii
ABSTRAK	v
ABSTRACT	Error! Bookmark not defined.
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	3
1.3 Pembatasan Masalah.....	4
1.4 Perumusan Masalah	4
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	5
1.6.1 Aspek Teoritis.....	6
1.6.3 Aspek Praktis	6
BAB II TINJAUAN PUSTAKA	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 Kegunaan <i>Data Mining</i>	11
2.2.3 Tahapan <i>Data Mining</i>	12
2.2.4 Pengelompokan <i>Data Mining</i>	13
2.2.5 Teknik Pembelajaran <i>Data Mining</i>	15
2.2.6 Penerapan <i>Data Mining</i>	16
2.3 Metode <i>Data mining</i>	19
2.3.1 Algoritma C.45	19
2.3.2 Algoritma <i>Nearest Neighbor</i>	19
2.3.3 Algoritma <i>Apriori</i>	20
2.3.4 <i>Fuzzy C-Means</i>	22
2.3.5 <i>Bayesian Clasification</i>	23
2.3.6 <i>Backpropagation</i>	24
2.4 <i>Software Pendukung</i>	24

2.5	Penelitian Terdahulu	25
2.6	Kerangka Pemikiran	29
2.7	Hipotesa	30
BAB III METODE PENELITIAN		31
3.1	Desain Penelitian	31
3.2	Teknik Pengumpulan Data	33
3.2.1	<i>Observasi</i> (Pengamatan)	33
3.2.2	Wawancara (<i>Interview</i>)	34
3.2.3	Dokumentasi	34
3.2.4	Studi <i>Literature</i>	34
3.3	Operasional Variabel	34
3.3.1	Jumlah Transaksi	35
3.3.2	Merek tas	36
3.4	Metode Analisis Dan Perancangan Sistem	36
3.4.1	Metode Analisis	36
3.4.2	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	Analisa Data	40
4.2	Hasil Pengujian	44
4.2.1	Pengujian Secara Manual	44
4.2.1	Pembentukan Pola Aturan <i>Asosiasi</i>	56
4.2	Pengujian Menggunakan Tanagra	65
BAB V SIMPULAN DAN SARAN		76
5.1	Simpulan	76
5.2	Saran	77
DAFTAR PUSTAKA		78
LAMPIRAN		
Lampiran 1. Pendukung Penelitian		
Lampiran 2. Daftar Riwayat Hidup		
Lampiran 3. Surat Keterangan Penelitian		

DAFTAR TABEL

Tabel 3. 1 Data Transaksi Penjualan Tas Di Toko Jc Jolin <i>Colecction</i>	35
Tabel 3.2 Jadwal Penelitian	39
Tabel 4.1 Daftar Merek Tas	40
Tabel 4.2 Data Set Penjualan Tas	41
Tabel 4. 3 Kode Transaksi	42
Tabel 4. 4 Hasil Integrasi Data	43
Tabel 4. 5 <i>Frequent 1-Itemset</i>	45
Tabel 4.6 Perhitungan <i>Frekuensi 1-itemset</i>	47
Tabel 4.7 Perhitungan <i>2-Itemset</i>	50
Tabel 4.8 Perhitungan <i>3-Itemset</i>	54
Tabel 4.9 Hasil Pembentukan Aturan <i>Asosiasi 2-Itemset</i>	58
Tabel 4.10 Hasil Pembentukan Aturan <i>Asosiasi 3-Itemset</i>	61
Tabel 4.11 <i>Final Rule</i>	63

DAFTAR GAMBAR

Gambar 2.1 Tahapan <i>Knowledge Discovery In Database</i> (KDD)	8
Gambar 2.2 Kerangka Pemikiran	30
Gambar 3.1 Desain Penelitian	31
Gambar 3.2 Perancangan Sistem	37
Gambar 3.3 Denah Lokasi	38
Gambar 4.1 Tampilan tabel tabular pada excel	65
Gambar 4.2 Tampilan awal Tanagra	66
Gambar 4.3 Import Data	66
Gambar 4.4 Pilih Data	67
Gambar 4.5 Import Data	68
Gambar 4.6 Data View	68
Gambar 4.7 Jendela Parameter <i>Define attribute statuses</i>	69
Gambar 4.8 Metode <i>Apriori</i>	70
Gambar 4.9 <i>Parameters</i>	71
Gambar 4.10 <i>Card Itemset</i>	72
Gambar 4.11 Langkah Melihat <i>Itemset</i>	73
Gambar 4.12 Hasil <i>2-Itemset</i>	73
Gambar 4.13 Hasil <i>3-Itemset</i>	74
Gambar 4.14 <i>Final Rule</i>	75

DAFTAR RUMUS

Rumus 2.1 Kedekatan antar 2 kasus	19
Rumus 2.2 Nilai <i>Support</i>	21
Rumus 2.3 Nilai <i>Support</i> dari 2 <i>item</i>	22
Rumus 2.4 Nilai <i>Confidence</i>	22
Rumus 2.5 Probabilitas <i>Bayes</i>	23
Rumus 4.1 Perhitungan Support 1- <i>Itemset</i>	45
Rumus 4.2 Perhitungan support 2- <i>itemset</i>	47
Rumus 4.3 Rumus mencari <i>confidence</i>	56

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan dan kemajuan teknologi dari tahun ketahun semakin pesat dan canggih. Teknologi ini terus diperbaharui dengan tujuan agar menghasilkan suatu teknologi yang dapat membantu menyelesaikan pekerjaan manusia agar lebih efektif dan efisien, termasuk pekerjaan dalam mengolah data. Data yang besar apabila dibiarkan begitu saja akan tidak berguna bahkan tidak bernilai sama sekali, oleh karenanya diperlukan teknologi yang mampu mengolah data yang besar sehingga mampu menghasilkan informasi yang berguna.

Aktivitas penjualan yang beroperasi setiap harinya menghasilkan data transaksi yang banyak sehingga menjadi tumpukan data. Salah satu transaksi penjualan itu yaitu di toko penjualan tas yang bernama toko tas Jc Jolin *Collection*. Tas merupakan suatu barang yang bisa termasuk ke dalam kebutuhan dasar baik itu bagi pelajar atau bahkan pekerja. Tas banyak dibeli karena karena kebutuhan dasar atau juga bisa hanya sebagai penunjang derajat sosial, para pembeli tas juga bisa beragam dari setiap kalangan masyarakat. Dengan beragam merek dan model adalah salah satu daya tarik dalam penjualan tas.

Hal ini tentu mampu menghasilkan peluang transaksi yang cukup besar dalam bisnis penjualan tas ini sehingga menghasilkan penumpukan data yang terus bertambah banyak, dimana toko tas Jc Jolin *Collection* dapat menjual tas sebanyak 10 sampai 50 tas perharinya dengan sistem penjualan di toko dan juga

penjualan secara *online* dan juga toko ini *welcome to reseller*. Di toko ini pengolahan data masih bersifat manual yaitu hanya memakai mesin kasir biasa tanpa ada *databasenya* kemudian mencatat setiap jumlah penjualan tas pada sebuah buku dan toko ini tidak hanya menjual tas saja tapi juga menjual sepatu dan ada juga jam yang mana semakin membuat penumpukan data laporan penjualan yang besar. Data transaksi tadi hanya dipakai sebagai dokumen atau sebagai bukti pembayaran saja, setelah itu data pada toko tas tersebut dibiarkan begitu saja sehingga hanya sekedar tumpukkan data transaksi yang tidak ada manfaatnya.

Ada beberapa masalah yang bisa muncul dalam bisnis penjualan tas yakni toko akan mengalami kesulitan dalam memperoleh informasi *update* dalam setiap tingkat penjualan pertahunnya bahkan perbulanya. Untuk Toko Jc jolin *Colecction* sendiri, penjualan tas bisa mencapai 348 pcs sampai 605 tas perbulanya. Tak jarang juga pembeli atau bahkan *consumen* terkadang kecewa karena merek tas yang dicari tidak ada. Hal ini tentu akan mempengaruhi besar perkembangan toko tersebut, dalam persaingan bisnis yang tinggi ini toko tersebut harus mampu merancang sebuah strategi bisnis. Dengan ketersediaan data penjualan yang besar harusnya bisa diolah dengan baik sehingga dapat digunakan untuk merancang sebuah strategi bisnis dalam meningkatkan penjualan baik di toko maupun bisnis lainnya (Badrul, Studi, & Informasi, 2016). Tapi dengan ketersediaan data yang besar saja ditambah belum adanya suatu sistem dan metode untuk mengolah data tersebut maka data akan sia-sia. Oleh karenanya diperlukan suatu metode yang

bisa mengolah data agar pengolahannya efisien dan akurat yaitu dengan mengimplementasikan penggunaan *data mining* dengan algoritma *apriori*.

Data mining merupakan salah satu cara eksploitasi atau penyeleksian data dengan memanfaatkan himpunan data dalam skala cukup besar dengan melewati rangkaian proses untuk memperoleh informasi yang bernilai dari data (Pandiangan, 2019). Algoritma *Apriori* yaitu suatu proses untuk menemukan suatu aturan *Apriori* yang memenuhi syarat minimum untuk *support* dan syarat minimum untuk *confidence*. Algoritma *Apriori* bertujuan untuk menemukan *frequent item sets* di sekumpulan data (Salamah & Ulinnuha, 2017). *Data mining* menggunakan algoritma *apriori* menghasilkan aturan *asosiatif* antar suatu kombinasi item untuk menghasilkan algoritma yang efisien dengan analisis pola frekuensi tinggi dari setiap transaksi penjualan tas sehingga akan dapat diketahui merek tas yang mana yang sering terjual dan toko tas dapat mengontrol persediaan setiap merek tas pada toko dan juga dapat dimanfaatkan untuk meningkatkan strategi bisnis. Dari uraian diatas peneliti tertarik untuk melakukan penelitian yang berjudul **“IMPLEMENTASI *DATA MINING* PADA PENJUALAN TAS MENGGUNAKAN ALGORITMA *APRIORI*”**.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka identifikasikan masalah dari penelitian ini adalah:

1. Minimnya pemanfaatan data transaksi pada sistem penjualan tas di toko Jc Jolin *Collection*.

2. Kurang efektifnya penjualan tas terhadap *konsumen* dikarenakan pengolahan data tidak diolah dengan baik.
3. Sulitnya mengetahui merek tas apa yang paling banyak terjual yang mana membuat toko Jc Jolin *Collection* tidak mengetahui merek tas apa yang kehabisan *stock*.

1.3 Pembatasan Masalah

Berdasarkan latar belakang dan identifikasi masalah di atas, maka dalam penelitian ini perlu adanya batasan masalah agar pengkajian masalah dalam penelitian ini dapat lebih fokus dan terarah. Adapun pembatasan masalah dalam penelitian ini sebagai berikut:

1. Penelitian ini menggunakan data transaksi penjualan tas ditoko Jc Jolin *Colecction*.
2. Data yang digunakan dalam penelitian ini adalah data transaksi penjualan tas tahun 2018.
3. Aplikasi yang digunakan untuk pengujian data adalah aplikasi *Tanagra* versi 1.4.

1.4 Perumusan Masalah

Berdasarkan batasan masalah yang telah penulis pilih, maka adapun perumusan masalah sebagai berikut:

1. Bagaimana menggunakan algoritma *apriori* untuk membentuk pola kombinasi *itemsets* yang dapat mempermudah toko Jc Jolin *Colecction* dalam mengolah data penjualan tas ?
2. Bagaimana cara keterkaitan *rule* transaksi merek tas pada toko Jc Jolin *Colecction* yang terjual ?
3. Bagaimana aplikasi *data mining* menggunakan algoritma *apriori* memberikan keputusan yang tepat terhadap penjualan tas ?

1.5 Tujuan Penelitian

Adapun tujuan penelitian berdasarkan rumusan masalah diatas, yaitu:

1. Untuk mengetahui bagaimana algoritma *apriori* dapat membentuk pola kombinasi *itemsets* yang dapat mempermudah toko Jc Jolin *Colecction* dalam mengolah data penjualan tas.
2. Untuk mengetahui keterkaitan *rule* terhadap transaksi tas yang terjual pada toko Jc Jolin *Colecction*.
3. Untuk mengetahui bagaimana aplikasi *data mining* menggunakan algoritma *apriori* memberikan keputusan yang tepat terhadap penjualan tas.

1.6 Manfaat Penelitian

Secara khusus, penelitian ini sangat berpeluang memberikan manfaat baik dari segi aspek teoritis maupun aspek praktis.

1.6.1 Aspek Teoritis

1. Menambah pengetahuan peneliti khususnya tentang *data mining* dengan algoritma *apriori*.
2. Sebagai referensi dalam mengolah data penjualan dengan *data mining*.
3. Sebagai referensi bagi pembaca dalam melakukan penelitiannya.

1.6.3 Aspek Praktis

1. Bagi toko Jc Jolin *Colecction* aplikasi ini sangat bermanfaat dalam mengolah data transaksi penjualan tas.
2. Untuk mempermudah mengetahui merek tas yang paling banyak terjual perperiodenya ditoko Jc Jolin *Colecction*.
3. Untuk mengetahui keterkaitan antar *rule* yang dihasilkan oleh algoritma *apriori* dengan menggunakan aplikasi Tanagra.

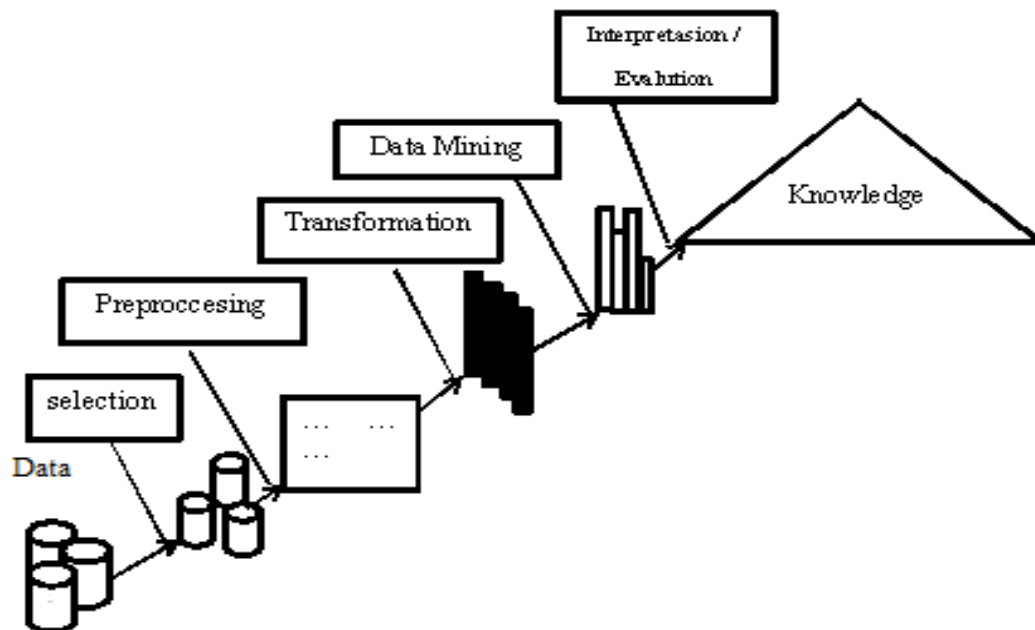
BAB II

TINJAUAN PUSTAKA

2.1 *Knowledge Discovery In Database (KDD)*

Knowledge Discovery In Database (KDD) merupakan proses pemisahan informasi tersembunyi, tersirat dan yang tidak diketahui dari sekumpulan data. Proses *knowledge discovery in database* melibatkan *output* dari proses *data mining* (bagian pemisahan dominan suatu pola data), lalu mengkonversikan *outputnya* menjadi informasi yang mudah dimengerti (Yulia & Azwanti, 2018)

Istilah *data mining* dan *knowledge discovery in database (KDD)* kerap dipakai secara bergantian dalam mengartikan cara penggalian informasi tersembunyi pada suatu basis data yang besar. Sesungguhnya sebutan kedua hal tersebut mempunyai konsep yang tidak sama, tetapi berkaitan satu dengan lainnya. *Data mining* merupakan salah satu tahap pada pada semua proses KDD . Proses KDD secara garis besar akan jelaskan seperti di bawah ini :



Gambar 2.1 Tahapan *Knowledge Discovery In Database* (KDD)
Sumber : (Retno Tri Vlandari, 2017, hal : 3)

1. *Data Selection*

Sebelum tahap pencarian informasi dalam KDD dimulai perlu dilakukan pemilihan ataupun penyeleksian data operasional dari data yang besar. Data hasil penyeleksian akan digunakan dalam proses *data mining*, disimpan dalam suatu berkas yang terpisah dari basis data yang operasional yang besar.

2. *Pre-processing/cleaning*

Proses *cleaning* pada data yang menjadi fokus KDD dilaksanakan terlebih dahulu lalu setelahnya baru proses *data mining* bisa dilakukan. Proses pembersihan tersebut meliputi menghapus data yang serupa ataupun ganda, memeriksa data yang tidak sesuai / konsekuen dan memperbaiki kesalahan pada data seperti kesalahan cetak (*tipografi*). Proses yang juga harus

dilaksanakan yaitu proses *enrichment* dimana yang berarti proses “memperkaya” data yang telah ada atau informasi lain sesuai dengan kebutuhan KDD. Contohnya data dan informasi eksternal.

3. *Transformation*

Coding ialah proses perubahan untuk data yang akan digunakan, yang akhirnya data yang dipakai konsekuen untuk proses *data mining*. Proses tersebut merupakan proses inovatif dan sangat tergantung pada pola serta jenis informasi yang diseleksi pada *database*.

4. *Data mining*

Proses *data mining* merupakan proses menemukan pola dan informasi yang menarik dari data yang dipilih bisa menggunakan teknik ataupun model tertentu. Teknik, metode dan algoritma dalam *data mining* sangat beragam. Tujuan dan proses KDD secara keseluruhan menjadi tolak ukur dalam pemilihan metode maupun algoritma yang akan digunakan secara tepat.

5. *Interpretation / Evaluation*

Tahap ini berhubungan dengan pemeriksaan apakah pola dan informasi yang dihasilkan berbeda dengan fakta atau hipotesis yang telah ada sebelumnya atau disebut juga tahap *interpretation* dan pada tahap ini harus mudah dimengerti.

6. Presentasi Pengetahuan (Knowledge)

Pada tahapan ini menggunakan metode Visualisasi

2.2 *Data Mining*

2.2.1 *Pengertian Data Mining*

Data mining yaitu teknik analisa pada data unntuk mencari informasi yang tersembunyi pada beberapa data besar yang dicatat pada saat bisnis perusahaan dijalankan. *Data mining* adalah teknik menemukan pola-pola yang tersembunyi (*hidden patern*) berupa pengetahuan (*knowledge*) yang mana sebelumnya tidak diketahui dari sekelompok data dimana data tersebut bisa terdapat dalam *database*, *data warehouse*, atau media penyimpanan informasi yang lain (Rahmawati & Merlina, 2018)

Para pakar juga mendefinisikan *data mining* yang dituangkan kedalam buku yang dikarangnya, yaitu: *Data mining* ialah suatu teknik yang digunakan dalam menelusuri nilai tambah berbentuk informasi yang tidak diketahui secara manual dari suatu basis data (Retno Tri Vulandari, 2017, hal : 2). *Data mining* adalah suatu sebutan yang dipakai untuk menjabarkan suatu penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang memakai teknik statistik, matematika, kecerdasan buatan, dan *machine learning* yang saling terkait dari beragam *database* besar (Kusrini dan Emha Taufiq Luthfi, 2009, hal : 3)

Data mining adalah metode analisa terarah pada cara penemuan pengetahuan pada basisdata atau *knowledge discovery in database* (KDD). Pengetahuan berupa data atau hubungan antar data valid (yang sebelumnya tidak diketahui) (Dr.Suyanto, 2017, hal : 1). Berlandaskan penjelasan *data mining* yang telah diuraikan sehingga bisa ditarik kesimpulan bahwasanya *data mining* itu ialah

suatu cara, teknik ataupun metode yang digunakan dalam mencari ataupun menggali data pada *database* besar dengan output informasi sebagai nilai tambah yang berguna bagi pemilik data.

2.2.2 Kegunaan *Data Mining*

Secara umum kegunaan *data mining* dibagi menjadi dua: deskriptif dan prediktif. Deskriptif berarti *data mining* dipakai untuk menemukan pola – pola yang mudah dimengerti manusia yang menjelaskan karakteristik data. Sedangkan prediktif berarti *data mining* dipakai untuk susunan model pengetahuan yang akan digunakan dalam prediksi (Dr.Suyanto,2017, hal : 3). Berdasarkan fungsionalitasnya, fungsi-fungsi *data mining* bisa dikelompokkan pada enam kelompok berikut ini :

- a) Klasifikasi (*classification*): menyamaratakan pola yang sudah diketahui untuk diaplikasikan pada data-data baru
- b) Klasterisasi (*clustering*): pengolongan data yang tidak diketahui nama kelasnya kedalam sejumlah kelompok tertentu sesuai dengan besar keserupaanya.
- c) Regresi (*regression*): fungsi yang ditemukan memodelkan data dengan kesalahan prediksi yang sangat kecil.
- d) Deteksi anomaly (*anomaly detection*): mengenali data yang tidak umum, seperti transformasi atau deviasi yang mungkin sangat penting dan perlu dilakukan investigasi selanjutnya.
- e) Pembelajaran aturan asosiasi (*association rule learning*) atau pemodelan

kebergantungan (*dependency modeling*): menemukan hubungan antar variabel.

- f) Perangkuman (*summarization*): menyediakan representasi data yang lebih simpel, berupa visualisasi dan pembuatan laporan.

2.2.3 Tahapan *Data Mining*

Data mining tidak dapat dipisahkan dari rangkaian *Knowledge in database* (KDD, hal ini karena *data mining* merupakan salah satu rangkaian dari *Knowledge in database* KDD). Serangkaian proses tersebut memiliki tahap sebagai berikut (Retno Tri Vulandari, 2017, hal : 2-3):

1. Pembersihan data (*data cleaning*)

Pembersihan diperlukan untuk membuang data yang tidak konsisten dan noise. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai sumber. data yang kita peroleh biasanya tidak hanya dari *database* yang sama tapi bisa dari beberapa *database*.

3. Transformasi data (*data transformation*)

Transformasi data yaitu data diubah menjadi bentuk yang sesuai untuk untuk di-*mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan.

4. Aplikasi Teknik *Data Mining*

Aplikasi teknik *data mining* yaitu proses ekstraksi pola dari data yang ada.

5. Evaluasi pola yang ditemukan (*Pattern Evaluation*)

Proses interpretasi pola bisa dipakai sebagai pengetahuan yang bisa dimanfaatkan sebagai pendukung pengambilan keputusan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

6. Presentasi pengetahuan (*Knowledge Presentasion*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.2.4 Pengelompokan *Data Mining*

Berdasarkan tugas yang dilakukan, *data mining* dibagi menjadi beberapa golongan, yaitu (Kusrini dan Emha Taufiq Luthfi, 2009, hal : 10-12):

1. Deskripsi

Adakalanya peneliti dan analis secara simpel hendak mencoba cara untuk menggambarkan pola dan kecenderungan yang ada dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih berorientasi pada numerik dari pada kearah kategori. Model estimasi yang dihasilkan dapat digunakan untuk kasus barunya.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan datang dimasa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Pada klasifikasi, terdapat target variabel kategori. Sebagai contoh, pengelompokan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Pengklusteran

Merupakan pengelompokan *record*, observasi, atau memperhatikan dan membentuk kelas objek – objek yang menyerupai. Kluster adalah gabungan

record yang mempunyai kemiripan satu dengan yang lainya dan memiliki ketidakmiripan dengan *record – record* pada kluster lain.

6. Asosiasi

Tugas asosiasi dalam *data mining* adalah mendapatkan atribut yang tampak dalam satu waktu. Analisa keranjang belanja merupakan sebutan dalam dunia bisnis pada umumnya.

2.2.5 Teknik Pembelajaran *Data Mining*

Teknik yang dipakai pada *data mining* akrab hubunganya dengan “penemuan” (*discovery*) dan “pembelajaran” (*learning*) yang dibagi ketiga metode utama dalam pembelajaran yaitu (Retno Tri Vlandari, 2017, hal : 7-8) :

a. *Supervised Learning*

Supervised Learning ialah teknik yang sering dipakai. Teknik ini serupa dengan “*programming by example*”. Teknik ini mengaitkan fase pelatihan yang mana data pelatihan historis yang karakteristik-karakteristiknya bisa dipetakan ke *output* yang sudah diketahui, mengolah ke algoritma *data mining*. Proses tersebut melatih algoritma dalam mengidentifikasi variabel-variabel dan nilai-nilai kunci yang kelak akan dipakai untuk dasar pada pembuatan dugaan-dugaan saat ada data baru.

b. *Unsupervised Learning*

Teknik pembelajaran ini berlawanan arah dengan *supervised learning*. Teknik ini bergantung pada penggunaan algoritma yang mendeteksi semua pola,

seperti *associations* dan *sequences*, yang muncul dari kriteria penting yang spesifik dalam data masukan. Pendekatan ini mengarah pada pembuatan banyak aturan (*rules*) yang mengkarakterisasikan penemuan *associations*, *clusters*, dan *segments*. Beberapa aturan tersebut akan dianalisa untuk mendapatkan sesuatu yang penting.

c. *Reinforcement Learning*

Teknik pembelajaran sedikit lebih minim dipakai dibandingkan teknik pembelajaran di atas, namun mempunyai pemakaian yang terus ditingkatkan dari waktu ke waktu dan memiliki kontrol adaptif. Teknik ini hampir sama dengan kehidupan nyata yakni seperti “*on-job-training*”, dimana seorang pekerja diberikan setumpuk tugas yang membutuhkan keputusan. Pada beberapa titik waktu kelak diberikan penilaian atas *performance* pekerja tersebut kemudian pekerja diminta mengevaluasi keputusan-keputusan yang telah dibuatnya sehubungan dengan hasil *performance* pekerja tersebut. *Reinforcement learning* sangat tepat penggunaannya pada masalah yang rumit yang bersangkutan dengan waktu.

2.2.6 Penerapan *Data Mining*

Penerapan *data mining* dapat dilakukan pada (Retno Tri Vlandari, 2017, hal : 5-7):

1. Analisa pasar dan manajemen

Transaksi kartu kredit, kupon diskon, keluhan pembeli, studi tentang gaya hidup publik ditambah kartu anggota club tertentu adalah sumber data yang sering dipakai. *Data mining* dapat menyelesaikan beberapa solusi, diantaranya yaitu:

a) Menembak target pasar

Model-model konsumen dan membuat kategori pada setiap pemesan sesuai dengan kategori yang diharapkan contohnya kedukaan yang serupa, tingkat penghasilan yang sama, etiket membeli maupun karakteristik yang lain dapat dikelompokkan (*clustering*) memanfaatkan *data mining*.

b) Melihat pola beli pemakai dari waktu ke waktu

Data mining bisa dipakai guna melihat sistem pembelian dari waktu ke waktu selanjutnya. Misalnya, pada saat seseorang *merid* bisa jadi dia selanjutnya memilih pindah dari *single account* ke *joint account*.

c) *Cross market analysis*

Data mining dapat dimanfaatkan untuk melihat hubungan antara penjualan satu produk dengan beberapa produk yang lain.

d) *Profil customer*

Data mining dapat melihat *profil customer* sehingga dapat diketahui kategori *customer* dengan pembelian produk tertentu yang sering dibeli.

e) Identifikasi kebutuhan *customer*

Dapat mengidentifikasi produk mana saja yang paling bagus untuk setiap kalangan *customer* dan faktor apa saja yang mampu menjadi daya tarik bagi pembeli.

f) Menilai *loyalitas customer*

g) Informasi *summary*

Dapat dimanfaatkan dalam menyusun laporan *summary* yang bersifat multi dimensi dan difasilitasi dengan informasi *statistic* lainnya.

2. Analisa perusahaan dan manajemen resiko

a) Perencanaan keuangan dan evaluasi aset

Data mining dapat dimanfaatkan dalam hal analisi dan prediksi *cash flow* juga untuk melakukan *contingent claim analysis* dalam mengevaluasi asset serta mampu menganalisis *trend*.

b) Perencanaan sumber daya

Dari sekumpulan ringkasan informasi juga pola pembelanjaan dan pemasukan dari masing-masing *resource*. Maka dapat digunakan untuk *resource planning*.

c) Persaingan

Memanfaatkan *data mining* dalam hal memantau para pesaing bisnis dengan melihat *market direction* mereka. *Data mining* juga mampu mengelompokkan *customer* dan dapat memberikan variasi harga untuk setiap grup.

3. Telekomunikasi

Data mining dapat menemukan jutaan transaksi yang masuk, dan mengetahui transaksi mana sajakah yang masih harus ditangani secara manual, sasaranya yakni menambah layanan otomatis.

2.3 Metode *Data mining*

2.3.1 Algoritma C.45

Algoritma C.45 yaitu algoritma yang dipakai dalam membentuk pohon keputusan. Pohon keputusan dibentuk oleh algoritma c.45 melalui proses berikut (Kusrini dan Emha Taufiq Luthfi, 2009, hal : 13-91) :

- a. memilih atribut sebagai akar
- b. lalu pilih cabang untuk setiap nilai
- c. membagi kasus dalam cabang
- d. ulangi untuk proses pada setiap cabang hingga semua kasus pada tiap cabang hingga memiliki kelas yang sama.

2.3.2 Algoritma *Nearest Neighbor*

Pendekatan dalam mencari masalah dengan menghitung kedekatan antar kasus baru dengan kasus lama, yaitu berpusat dengan kesamaan berat dari sejumlah fitur yang sudah ada. Adapun rumus untuk melakukan kedekatan antar dua kasus adalah sebagai berikut (Kusrini dan EmhaTaufiq Luthfi, 2009, hal:93-148):

$$\text{similarity}(T,S) = \frac{\sum_{i=1}^n f_i(S_i) * w_i}{W_I}$$

Rumus 2.1 Kedekatan antar
2 kasus

Keterangan :

T : Kasus Baru

S : Kasus yang ada pada penyimpanan

n : Jumlah atribut disetiap kasus

i : Atribut individu diantara 1 s.d. n

f : Fungsi *similarity* atribut i diantara kasus T dan juga kasus S

w : Bobot yang diperkenankan pada atribut ke-i

Biasanya untuk kedekatannya berada dinilai antara 0 s.d. 1. Nilai 0 memiliki arti kedua kasus mutlak tidak mempunyai kemiripan, kebalikannya untuk nilai 1 kasus mirip dengan mutlak.

2.3.3 Algoritma *Apriori*

Salah satu jenis aturan yang ada dalam asosiasi pada *data mining* ialah algoritma *apriori* . Selain *apriori*, yang terkandung dalam kelompok ini ialah metode *Generalized Rule Induction* dan juga *Algoritma Hash Based* (Kusrini dan EmhaTaufiq Luthfi, 2009, hal : 149-176)

Aturan yang mengemukakan asosiasi diantara beberapa atribut dikenal juga dengan sebutan *affinity analysis* atau *market basket analysis*. Analisis asosiasi ataupun *association rule mining* ialah metode *data mining* dalam menemukan aturan asosiatif diantara suatu kombinasi *item*. Untuk analisis asosiasi banyak dikenal dikarenakan aplikasinya dalam analisis isi keranjang belanja di pasar swalayan, analisis asosiasi jugadikenal dengan *market basket analysis*.

Analisis asosiasi kerap kali disebut menjadi salah satu teknik *data mining* yang jadi pijakan untuk beberapa teknik *data mining* yang lain. Secara eksklusif, satu Dari banyaknya tahapan analisis asosiasi yang menjadi daya tarik perhatian beberapa peneliti dalam menciptakan algoritma yang efisien yaitu menganalisis pola frekuensi tinggi (*frequent pattern mining*). Bernilai maupun tidak bernilainya suatu aturan asosiatif akan diketahui menggunakan dua parameter, yakni satu disebut *support* dan dua disebut dengan *confidence*. *Support* (nilai penunjang) yaitu persentase kombinasi *item* pada *database*, sedang untuk *confidence* (nilai kepastian) ialah kuatnya relasi antar-*item* pada aturan asosiasi.

Analisis asosiasi ialah suatu proses dalam menemukan seluruh aturan asosiasi yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap:

1. Analisis pola frekuensi tinggi

Berfungsi menemukan kombinasi *item* yang mencukupi syarat minimum dari nilai *support* yang diberikan pada *database*. Nilai *support* sebuah *item* dihasilkan menggunakan rumus di bawah :

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

Rumus 2.2 Nilai *Support*

2. Pembentukan Aturan Asosiasi

Setelah seluruh pola frekuensi tinggi diperoleh, kemudian mencari aturan asosiasi yang mencukupi syarat minimum untuk *confidence* menggunakan cara dengan menghitung *confidence* aturan asosiatif $A \rightarrow B$.

Sementara itu, nilai *support* dari 2-item didapat dengan rumus 2.3 berikut.

$$Support = (A,B) P(A \cap B)$$

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

Rumus 2.3 Nilai *Support* dari 2 item

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

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

Rumus 2.4 Nilai *Confidence*

2.3.4 Fuzzy C-Means

Analisis kluster yaitu cara membagi data pada suatu himpunan kedalam sebagian himpunan yang kecocokan datanya dalam suatu himpunan lebih banyak daripada kecocokan data tersebut dengan data pada himpunan lain. Kluster secara lazim yaitu wujud kelompok bagian dari suatu kelompok data dan metode clustering dapat dikategorisasikan berdasarkan kelompok bagian yang dihasilkan: apakah *fuzzy* atau *crisp (hard)*. Terdapat banyak algoritma yang dipakai pada

clustering. *Fuzzy C-Means* (FMC) merupakan salah satunya. *Fuzzy C-Means* (FMC) ialah algoritma *clustering* data dimana semua data dapat dijadikan bagian dari suatu kluster sambil level yang diartikan sama dengan derajat keanggotaan (Kusrini dan Emha Taufiq Luthfi, 2009, hal : 177-187).

2.3.5 Bayesian Clasification

Bayes classification yaitu pengelompokkan statistik yang dipakai dalam memperediksi perobabilitas keanggotaan suatu *class*. klasifikasi *bayes* berdasarkan pada teorema *bayes* dan terbukti mempunyai akurasi dan kecepatan tinggi dalam pengaplikasiannya pada *database* dengan data yang besar (Kusrini dan EmhaTaufiq Luthfi, 2009, hal : 189-198).

Dimana bentuk umum *teorema bayes*, dapat dilihat di bawah ini :

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

Rumus 2.5 Probabilitas *Bayes*

Dalam hal ini:

X = data dengan class yang belum diketahui

H = hipotesis data X merupakan suatu *class* spesifik

P(H|X) = probabilitas hipotesis H berdasar kondisi X (*posteriori probability*)

P(H) = probabilitas hipotesis H (*prior probability*)

P(X|H) = probalitas X berdasarkan kondisi pada hipotesis H

P(X) = probabilitas dari X

2.3.6 Backpropagation

Backpropagation adalah algoritma yang terdapat pada pembelajaran jaringan syaraf tiruan. Proses pembelajaran ini dilaksanakan sesuai bobot- bobot jaringan syaraf tiruan dengan arah mundur berdasarkan nilai eror pada pembelajaran (Kusrini dan Emha Taufiq Luthfi, 2009, hal : 199-208).

Langkah pembelajaran dalam metode *backpropagation* adalah sebagai berikut:

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

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

2.4 Software Pendukung

Software yaitu istilah khusus untuk data yang diformat, dan disimpan secara digital, termasuk program komputer, dokumentasinya, dan berbagai informasi yang bisa dibaca, dan ditulis oleh komputer. Ada beberapa *software* yang bisa

digunakan pada *data mining*, Pada penelitian ini *software* yang digunakan yaitu Tanagra versi 1.4. Tanagra adalah suatu *software* berbasis terbuka dimana setiap pengguna dapat mengakses sumber kode, dan menambahkan algoritma pengguna, selama pengguna sepakat dan sinkron terhadap lisensi pendistribusian *software* (Sikumbang, 2018).

Tanagra merupakan sebuah *software* berbasis *open source*. *Software* ini mudah dipakai selayaknya dengan aturan yang telah ada pada pengembangan *software* untuk menganalisis data *real*. Pada Tanagra tidak perlu *coding* khusus, dikarenakan semua *tools* dan *platform* sudah tersedia pada *software* tersebut.

2.5 Penelitian Terdahulu

Pada penelitian ini, peneliti menggunakan beberapa penelitian terdahulu sebagai referensi penelitian yang dikutip dari beberapa jurnal ilmiah *data mining*, yaitu:

1. Menurut penelitian (Syahdan & Sindar, 2018) dengan judul “**Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota**”. E-ISSN: 2621-3052. Penelitian ini membahas tentang minimarket indomaret sering merasa kewalahan dengan tingkat belanja konsumen yang tidak beraturan. Dengan transaksi yang cukup besar maka perlu dilakukan penentuan tata letak produk makanan dan minuman. Teknik *data mining* yang dipakai yaitu *asosiatif rule* dengan algoritma *apriori* yang bertujuan mencari kombinasi dari setiap pola frekuensi transaksi. Dengan begitu akan diketahui pola frekuensi tertinggi, kemudian dicari aturan

asosiasi yang memenuhi syarat minimum untuk *confidence* aturan asosiatif $A \rightarrow B$ minimal *confidence* = 25%, nilai *confidence* dari aturan $A \rightarrow B$ dengan memakai 25 produk. Dengan setiap proses menggunakan algoritma *apriori*, maka item yang memenuhi *Support x Confidence* terbesar dan minimum *confidence* $\geq 25\%$ sehingga aturan asosiasi yang berurut yaitu Kode Produk = A, F Nama Produk = *Potabee chip bbq 35g, Plattos snack s.pg 30 g Confidence* = 33(%)

2. Menurut penelitian (Rezkiani, 2016) dengan judul **“Implementasi data mining dengan algoritma *apriori* untuk menentukan merek sepatu yang diminati pada mahasiswa pascasarjana kelas 14.1A.01 STMK Nusa Mandiri Jakarta”**. ISBN: 978-602-72850-1-9. Penelitian ini membahas tentang cara menentukan merek sepatu yang diminati mahasiswa pascasarjana, yang mana bisa membantu strategi penjual sepatu agar penjualan terus meningkat dengan menggunakan teknik analisis keranjang pasar. Proses pencarian asosiasi atau hubungan antar item data ini di ambil dari suatu basis data relasional. Proses tersebut menggunakan algoritma *apriori*. Dengan algoritma ini akan membantu mengetahui merek sepatu yang paling diminati dengan memanfaatkan transaksi penjualan. Dengan nilai *support* dan *confidence* yaitu minimum *support* 20% dan minimum *confidence* 60%. Hasil yang diperoleh dari penelitian ini yaitu *Nike-Adidas* dengan *support* 25% dan *confidence* 80% dan *Converse-Adidas* dengan *support* 25% dan *confidence* 66,67%.

3. Menurut penelitian (Maulana & Fajrin, 2018) dengan judul **“Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma FP-Growth Pada Data Transaksi penjualan Spare Part Motor”**. ISSN: 2406-7857. CV Thaja Baru merupakan dealer yang memiliki transaksi penjualan yang besar dimana mendukung untuk mendirikan cabang baru. Dalam pemilihan lokasi cabang harus bisa menarik pola pembelian konsumen, karenanya dibutuhkan suatu analisa yang lebih dalam agar mendapatkan informasi bermanfaat dan dapat menghasilkan keuntungan. Penelitian ini menerapkan algoritma FP-Growth dalam aplikasi yang dapat mengetahui pola pembelian konsumen disetiap cabang yang memiliki karakteristik yang berbeda. Aplikasi yang dipakai dalam mengolah data pada penelitian ini yaitu *Rapidminer*. *Spare part* yang melewati minimum support dan minimum *confidence* kemudian sering terjual ialah *battery assy, screw valve adjusting, axle, cable clutch, oil seal* dan *gasket cylinder*.
4. Menurut penelitian (Rodiyansyah, 2015) dengan judul **“Algoritma apriori untuk analisa keranjang belanja pada data transaksi penjualan”**. ISSN: 2460-1861. Pada perusahaan perdagangan, sering ada transaksi dan menghasilkan kumpulan data. Tumpukan data transaksi inilah yang diolah sebagai informasi yang bermanfaat. Dengan melakukan *data mining* metode *apriori* pada data transaksi penjualan tersebut diharapkan dapat menghasilkan pengetahuan mengenai pola pembelian oleh konsumen. Pada akhirnya pola yang ditemukan ini dapat digunakan sebagai penjunjang

keputusan manajer dalam mengelola aktifitas perusahaannya. Dengan nilai *support* dan *confidence* yaitu minimum *support* 30% dan minimum *confidence* 10%.

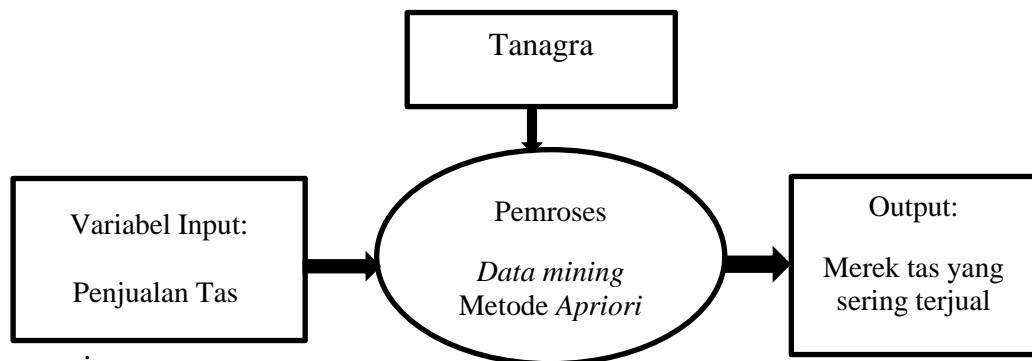
5. Menurut penelitian (Ardiansyah & Kardianawati, 2019) yang berjudul **“Implementasi Algoritma *Apriori* Untuk Menentukan *Cross Selling* Produk Pada Apotek RSUD Tugurejo Semarang”**. E-ISSN: 2528-0236. Perhitungan data transaksi yang tersimpan tiap harinya dan menyebabkan proses transaksi penjualan obat memerlukan waktu yang lama karena masih menggunakan cara yang sudah terkomputerisasi tetapi masih menggunakan cara dengan memilih daftar obat atau menentukan kode obat satu persatu berdasarkan obat yang akan dipesan. Untuk mempercepat proses penjualan tersebut dapat dilakukan dengan teknik *data mining* adalah *Association Rule* dengan memanfaatkan *cross selling* produk melalui perhitungan algoritma *Apriori*. agar proses penjualan lebih mudah dan lebih cepat sehingga dapat mempercepat proses tiap transaksi yang akan mengakibatkan berkurangnya panjang antrean para pembeli. Dengan implementasi algoritma *apriori* dengan acuan nilai minimum *support* 11% dan nilai minimum *confidence* 60% didapati pola transaksi pembelian obat dengan frekuensi tertinggi sebanyak 3 itemset yaitu jika membeli *abilify discmelt* 10mg (*aripiprazole*) *box* 10 tab dan *cath. Folley 2 way 16* (*rusch/urocare*) maka membeli *apialasit drop* dengan nilai *support* 18.6% dan nilai *confidence* 96.87% dan jika membeli *alprazolam 0.5 mg tab* dan *clonidin 0.15mg* maka membeli

nevirapine 200mg @60 (D) dengan nilai support 11.6% dan nilai confidence 90.62%.

6. Menurut penelitian (Mirmozaffari, Alinezhad, & Gilanpour, 2017) yang berjudul ***"Data Mining Apriori Algorithm for Heart Disease Prediction"***. ISSN: 2349-1469. *To find the unknown trends in heart disease, Apriori algorithm in association rule are applied to a unique dataset and their accuracy are compared in two different software. To further enhance accuracy and achieve more reliable variables, the dataset is purified by Discretization unsupervised filter. Finally, better performance software for Apriori algorithm with better accuracy is introduced. The system was implemented in WEKA and MATLAB software and prediction accuracy within Apriori algorithm in 3 steps, re compared. Despite "min support", the exact number of supports are introduced. For instance, in The Second WEKA Apriori rule, only "min support" (0.6) is introduced. However, in The Second MATLAB Apriori rule, despite "min support" (0.6), the exact number of support (0.622) is evaluated.*

2.6 Kerangka Pemikiran

Untuk dapat melakukan penelitian penjualan tas ini, diperlukan kerangka pemikiran yang bisa dipakai dalam menjelaskan bagaimana cara metode ini mampu memecahkan permasalahan yang ada. Kerangka pemikiran tersebut digambarkan dengan diagram dibawah ini:



Gambar 2.2 Kerangka Pemikiran
Sumber: Data penelitian, 2019

Pada gambar 2.2 menjelaskan bahwasanya di dalam kerangka pemikiran ini mempunyai variabel input, pemroses, dan terakhir variabel *output*. Untuk variabel input ialah transaksi penjualan tas di Toko Jc Jolin *Colecction*. Kemudian variabel input diproses menggunakan teknik *data mining* menggunakan metode *apriori*, lalu diuji memakai *Software Tanagra*, agar menemukan merek tas apa saja yang paling banyak terjual. Sedangkan untuk *output* akan mendapatkan hasil *item set* yang banyak terjual dari keterkaitan *rule* atau pola.

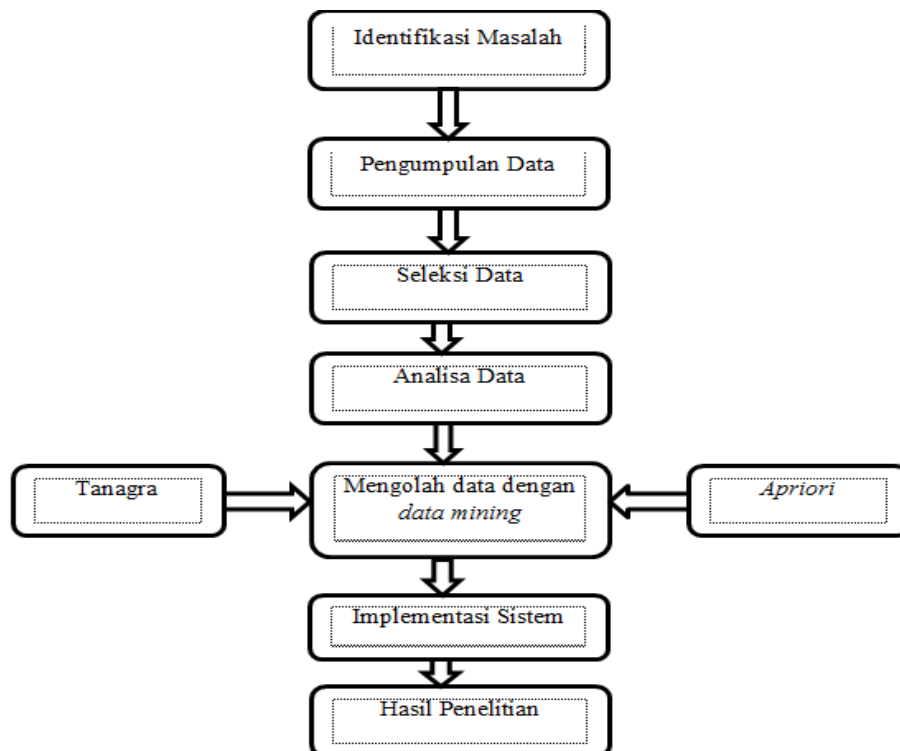
2.7 Hipotesa

Penelitian diharapkan mampu menghasilkan sebuah keputusan dengan menggunakan teknik *data mining* metode algoritma *apriori* dimana *rule* atau pola yang terbentuk dari transaksi penjualan tas.

BAB III METODE PENELITIAN

3.1 Desain Penelitian

Pada sebuah penelitian, desain penelitian sangat penting guna memperoleh informasi yang dibutuhkan dalam penyusunan dan penyeleksian masalah pada penelitian sehingga berjalan dengan baik dan sistematis.



Gambar 3.1 Desain Penelitian
Sumber : Data Penelitian, 2019

Tahap-tahap penelitian sebagai berikut:

1. Identifikasi Masalah

Langkah awal dalam melakukan sebuah penelitian, dimana mencari sebuah permasalahan yang bisa dijadikan untuk bahan penelitian. Pada penelitian ini permasalahan yang terjadi yaitu Toko tas Jc Jolin *Colection* sulit dalam mengolah data, dimana toko ini penyimpanan datanya masih bersifat manual yaitu mencatat setiap penjualan pada buku dan terjadilah penumpukan data yang dibiarkan begitu saja. Oleh karenanya digunakan *data mining* menggunakan metode *apriori* sehingga dapat mempermudah pengolahan data menjadi informasi yang berguna.

2. Pengumpulan Data

Adapun langkah selanjutnya setelah identifikasi masalah yakni pengumpulan data yang relevan dengan penelitian. Pada penelitian ini pengumpulan data dilakukan dengan cara observasi langsung ke Jc Jolin *Colection*, wawancara dengan ibu zhang angelina jolin dan dokumentasi yaitu foto bersama salah satu pengurus toko tersebut dengan memperoleh data penjualan tas yang akan dipakai sebagai sumber olahan data penelitian.

3. Seleksi Data

Setelah data diperoleh, maka peneliti akan melakukan seleksi data agar data tersebut dilakukan pembersihan terhadap variabel yang tidak terpakai.

4. Analisa Data

Analisa dilakukan menggunakan algoritma *apriori* yang bertujuan agar data lebih efektif dan efisien dalam pengolahannya dengan hasil yang lebih akurat.

5. Mengolah Data Dengan *Data Mining*

Setelah tahap analisa, data kemudian diolah dengan *data mining* menggunakan aplikasi tanagra sehingga memperoleh *rule assosiasi*.

6. Implementasi Sistem

Tahapan selanjutnya yakni implementasi sistem menggunakan aplikasi Tanagra untuk menguji kecocokan data olahan sesuai dengan hitungan manual sama dengan tanagra *rule* yang dihasilkan.

7. Hasil Penelitian

Output dari seluruh tahapan penelitian yakni hasil penelitian berupa *itemset* paling sering terjual. Dimana akan dibentuk *rule assosiasi* hasil dari data penjualan tas.

3.2 Teknik Pengumpulan Data

Suatu cara pengumpulan data yang diterima langsung dari narasumber yang bertautan. Teknik pengumpulan yang dipakai pada penelitian ini, yakni:

3.2.1 Observasi (Pengamatan)

Observasi yakni suatu metode pengumpulan data dengan mengamati kegiatan yang sedang terjadi. pada penelitian ini peneliti terjun langsung ke toko tas Jc Jolin *Collection* untuk melakukan pengamatan guna untuk mendapatkan data penjualan tas yang dibutuhkan. Data yang diperoleh yaitu transaksi penjualan tas tahun 2018.

3.2.2 Wawancara (*Interview*)

Wawancara yaitu percakapan secara langsung yang bertujuan untuk mendapatkan informasi. Setelah observasi peneliti langsung melanjutkan kebagian wawancara, disini peneliti memberikan pertanyaan dan narasumber memberikan jawaban yang bisa dipakai sebagai pendukung sumber data penelitian. Peneliti melaksanakan kegiatan wawancara dengan pemilik toko yaitu ibu zhang Angelina jolin selaku pemilik toko tas Jc Jolin *Colecction*.

3.2.3 Dokumentasi

Dokumentasi yakni berupa pelaporan ataupun gambar kegiatan yang didapatkan dari tempat penelitian. Pada penelitian ini, dokumentasi yang dilakukan yakni foto peneliti dengan salah satu pengurus toko tas Jc Jolin *Colecction* sebagai bukti pendukung tempat pengumpulan data pada penelitian ini.

3.2.4 Studi *Literature*

Untuk pendukung guna menghasilkan data olahan yang lebih akurat, peneliti menggunakan beberapa buku dan jurnal *data mining* sebagai referensi dalam penulisan penelitian ini.

3.3 Operasional Variabel

Operasional variabel adalah suatu petunjuk ataupun arah bagi peneliti dalam mengukur suatu variabel. Operasional variabel merupakan informasi ilmiah yang dapat membantu peneliti lain dengan variabel yang sama sehingga peneliti tersebut dapat menarik kesimpulan apakah menggunakan pengukuran yang sama

atau perlu pengukuran yang lain. Operasional variabel yang digunakan yaitu data penjualan tas di toko Jc Jolin *Colecction*. Variabel-variabel yang diambil yaitu:

3.3.1 Jumlah Transaksi

Jumlah transaksi yakni banyaknya transaksi penjualan tas dalam kurun waktu perbulan dalam satu tahun terakhir. Transaksi perbulanya mencapai ± 605 pcs tas yang terjual dengan 25 merek tas.

Tabel 3. 1 Data Transaksi Penjualan Tas Di Toko Jc Jolin *Colecction*

MEREK TAS	Tahun 2018											
	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Agust	Sept	Okt	Nov	Des
Jeep	26	19	20	54	58	14	19	45	15	23	27	40
Dior	43	55	58	12	22	19	20	19	45	50	17	54
Emori	9	16	15	11	0	2	2	0	0	11	18	9
M kors	11	62	30	19	44	49	19	17	19	18	22	38
M vania	16	25	40	29	58	10	44	14	22	26	47	19
VB	10	11	18	6	14	3	11	5	8	5	0	12
MJ	22	17	16	24	18	14	43	17	19	19	40	16
Lv	29	59	18	80	46	23	37	15	22	31	42	37
Moschino	0	6	0	20	11	19	5	19	17	16	16	15
Aigner	4	15	14	0	16	14	0	6	4	7	19	33
Fossil	16	19	20	17	0	12	3	14	0	3	0	0
Bonia	20	51	23	15	12	52	40	19	23	29	16	24
TKorea	15	11	21	19	8	0	4	14	12	0	18	18
Prada	28	52	57	19	24	17	54	41	15	43	38	39
Fila	2	24	18	19	18	25	14	18	19	17	18	22
Pedro	0	15	5	6	8	1	2	13	4	1	0	4
Polo	35	12	59	17	29	12	19	17	14	12	21	19
Polo Xidi	5	0	3	2	0	14	16	6	0	0	19	0
Gucci	39	56	17	20	46	29	42	17	15	51	15	24
Miu-Miu	21	19	23	21	18	18	16	47	19	30	16	19
Kipling	4	5	0	7	0	2	6	4	3	2	0	9
x bouluo	10	11	0	2	0	8	18	5	6	4	12	18
Coach	28	21	94	82	41	19	58	35	17	21	51	18
Fendi	8	5	0	0	0	0	10	8	6	9	0	0
Furla	75	19	20	17	39	22	37	18	24	12	43	26

Sumber: Data Olahan Peneliti , 2019

3.3.2 Merek tas

Merek tas yaitu merek tas yang ada ataupun jenis tas yang ada di toko Jc Jolin *Colecction*. Adapun merek tas yang ada pada toko ini yaitu Gucci, dior, michael kors, marc jacobs, fossil, coach, jeep, emori, lv, monna vania, victoris beckham, polo, pedro, miu-miu, fendi, furla, polo xidi, x bouluo, moschino, fila, kipling, tas korea, prada, aigner, bonia.

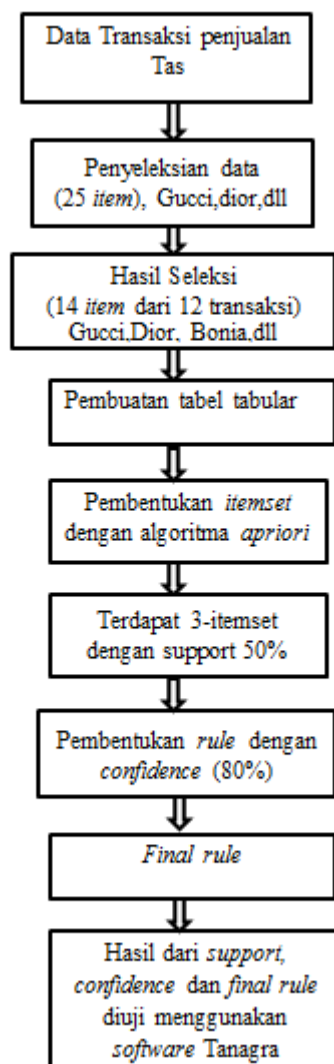
3.4 Metode Analisis Dan Perancangan Sistem

3.4.1 Metode Analisis

Metode analisis ialah satu hal yang amat penting dan wajib dilakukan pada penelitian. Hal tersebut dikarenakan bila suatu penelitian tidak melakukan metode analisis sehingga untuk keakuratan hasil dari suatu penelitian masih dipertanyakan. Hasil dari penelitian sangat dipengaruhi besar metode analisis. Untuk itu penelitian ini menggunakan *data mining* dengan metode *apriori*.

3.4.2 Perancangan Sistem

Adapun metode perancangan sistem yang digunakan dalam penelitian ini adalah sebagai berikut:



Gambar 3.2 Perancangan Sistem
Sumber: Data Penelitian, 2019

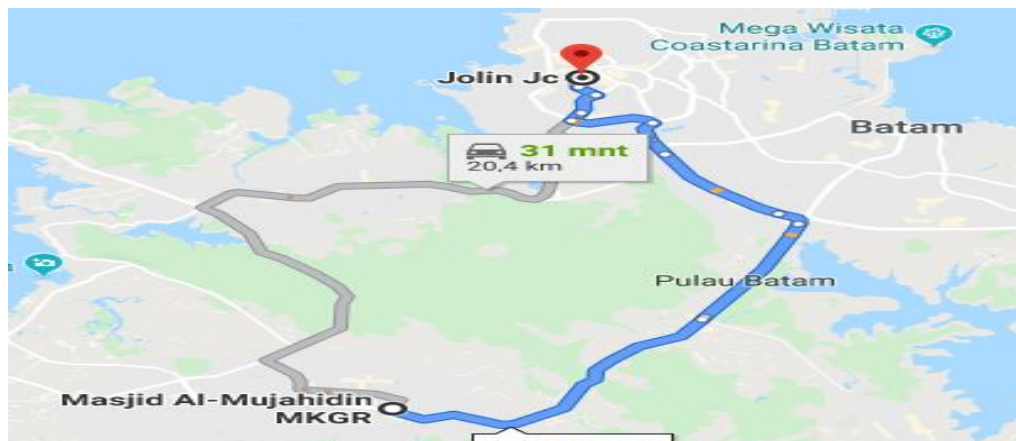
Dari gambar 3.2 dapat dilihat bahwa proses penelitian dimulai dari mengumpulkan data penjualan tas ditoko tas Jc Jolin *Colecction* dengan cara observasi dan wawancara. Kemudian data yang didapatkan peneliti diseleksi untuk mendapatkan sampel yang akan digunakan lalu data diolah memakai algoritma *apriori* dan hasil olahan data akan diperoleh dengan menggunakan aplikasi Tanagra.

3.5 Lokasi dan Jadwal Penelitian

Pada sebuah penelitian dengan tujuan dalam mengolah data penjualan tas di toko tas Jc Jolin *Colecction* dengan diterapkan *data mining* menggunakan algoritma *apriori* dibutuhkan lokasi dan waktu yang menjadi tempat penelitian. Adapun lokasi dan waktu yang dipakai pada penelitian ini sebagai berikut:

3.5.1 Lokasi Penelitian

Lokasi penelitian dilakukan di Toko Jc Jolin *Collection*, Alamat Komplek Pertokoan Nagoya 2000 No.08, Lubuk Baja, Batam.



Gambar 3.3 Denah Lokasi
Sumber: Data Penelitian, 2019

3.5.2 Jadwal Penelitian

Penelitian ini dilakukan dari bulan September 2019 hingga bulan Januari 2020. Adapun jadwal kegiatannya dapat dilihat di bawah ini.

Tabel 3.2 Jadwal Penelitian

NO	Kegiatan	Waktu Pelaksanaan Kegiatan																			
		September 2019				Oktober 2019				November 2019				Desember 2019				Januari 2020			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Survey Penelitian		■	■																	
2	Pengumpulan Data				■																
3	BAB 1						■	■													
4	BAB II								■												
5	BAB III										■	■	■								
6	BAB IV													■	■	■	■				
7	BAB V																	■	■	■	
8	Pengumpulan Skripsi																				■

Sumber : Data Penelitian, (2019)