

**IMPLEMENTASI *DATA MINING* PENJUALAN BUAH
DENGAN ALGORITMA APRIORI**

SKRIPSI

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



**Oleh:
Miftahul Janah
150210036**

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

IMPLEMENTASI *DATA MINING* PENJUALAN BUAH DENGAN ALGORITMA APRIORI

**Oleh:
Miftahul Janah
150210036**

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

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

Batam, 08 Februari 2019

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

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, 08 February 2019

Miftahul Janah

(150210178)

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 buah Awi. Data *sample* yang diambil adalah 15 *items* dari 14 transaksi pada bulan Desember 2018. 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 sirsak 100% anggur 100% dan Mangga 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 Buah Awi

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 awi fruit shop, using 15 items from 14 transactions on for the december 2018 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 soursop, grape and mango 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 Awi*

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 Buah Awi dengan Algoritma *Apriori*”. Skripsi ini merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

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

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

1. 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, Aamiin.

Batam, 08 Februari 2019

Penulis

DAFTAR ISI

Halaman

HALAMAN SAMPUL DEPAN	
HALAMAN JUDUL	
HALAMAN PERNYATAAN	
HALAMAN PENGESAHAN	
ABSTRAK	i
ABSTRACT	ii
KATA PENGANTAR	iv
DAFTAR ISI	v
DAFTAR GAMBAR	viii
DAFTAR RUMUS	xi
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian	1
1.2 Identifikasi Masalah	3
1.3 Pembatasan Masalah	3
1.4 Rumusan Masalah	4
1.5 Tujuan Penelitian	4
1.6 Manfaat Penelitian	5
1.6.1 Manfaat Teoritis	5
1.6.2 Manfaat Praktis	5
BAB II KAJIAN PUSTAKA	6
2.1 <i>Knowledge Discovery in Database (KDD)</i>	6
2.1.1 Proses <i>Knowledge Discovery in Database</i>	7
2.2 <i>Data Mining</i>	8
2.2.1 Pengelompokan <i>Data mining</i>	11
2.2.2 Arsitektur Sistem <i>Data Mining</i>	14
2.3 Algoritma <i>Association</i>	16
2.4 Algoritma <i>Apriori</i>	17
2.5 Analisa Keranjang Belanja.....	18
2.6 <i>Software</i> Pendukung.....	18
2.7 Penelitian Terdahulu	19
2.8 Kerangka Pemikiran.....	21
BAB III METODE PENELITIAN	23
3.1 Desain Penelitian.....	23
3.2 Teknik Pengumpulan Data.....	25
3.3 Operasional Variabel.....	25
3.4 Metode Perancangan Sistem	29
3.5 Lokasi dan Jadwal Penelitian	30
3.5.1 Lokasi Penelitian	30
3.5.2 Jadwal Penelitian.....	30

BAB IV HASIL DAN PEMBAHASAN	32
4.1 Analisa Data	32
4.1.1 Pemilihan Data/Seleksi Data.....	32
4.1.2 Menentukan <i>Frequent Itemset</i>	32
4.1.3 Pengujian Menggunakan Tanagra.....	41
BAB V SIMPULAN DAN SARAN.....	52
5.1 Simpulan	52
5.2 Saran.....	52
DAFTAR PUSTAKA.....	54
DAFTAR RIWAYAT HIDUP	
SURAT KETERANGAN PENELITIAN	
LAMPIRAN	

DAFTAR TABEL

	Halaman
Tabel 3.1 Data Penjualan Toko Buah Awi.....	26
Tabel 3.2 Jadwal Penelitian.....	31
Tabel 4.1 Tabel Tabular Data Transaksi	33
Tabel 4.2 Representasi Data	34
Tabel 4.3 Tabel Hasil <i>Support 1-itemset</i>	36
Tabel 4.4 Hasil dari Kombinasi <i>2-itemset</i>	38
Tabel 4.5 Hasil Pembentukan Aturan Asosiasi pada <i>2-itemset</i>	41

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Arsitektur Sistem <i>Data Mining</i>	14
Gambar 2.2 Kerangka Pemikiran	22
Gambar 3.1 Desain Penelitian	23
Gambar 3.2 Perancangan Sistem.....	29
Gambar 4.1 Atribut-atribut yang akan di <i>import</i> ke Tanagra.....	42
Gambar 4.2 <i>Add-ins</i> ke Tanagra.....	43
Gambar 4.3 <i>Execute</i> Tanagra pada <i>Excel</i>	43
Gambar 4.4 Tampilan <i>File</i> pada Tanagra.....	44
Gambar 4.5 Tampilan jendela <i>New</i> pada Tanagra	44
Gambar 4.6 Tampilan pemilihan <i>File</i> pada Tanagra.....	45
Gambar 4.7 Tampilan dari <i>File</i> yang telah dipilih	46
Gambar 4.8 Tampilan <i>File Dataset</i>	46
Gambar 4.9 Jendela Parameter <i>Define attribute statuses</i>	47
Gambar 4.10 Tampilan <i>Dataset</i> yang telah diproses.....	48
Gambar 4.11 Jendela <i>Frequent Itemsets</i>	48
Gambar 4.12 Hasil dari Penentuan <i>Frequent Itemsets</i>	49
Gambar 4.13 Jendela <i>Association rule parameter</i>	50
Gambar 4.14 Hasil Pengujian <i>Rule Final</i>	50

DAFTAR RUMUS

	Halaman
Rumus 4. 1 <i>Perhitungan Support 1-itemset</i>	34
Rumus 4. 2 <i>Perhitungan Support 2-itemset</i>	36
Rumus 4. 3 <i>Perhitungan Confidence 2-itemset</i>	40

DAFTAR LAMPIRAN

Lampiran 1. Pembentukan Kombinasi *2-itemset*

Lampiran 2. Fotocopy Surat Izin Penelitian

Lampiran 3. Data Penjualan Toko Buah Awi dalam Per/Minggu

Lampiran 4. Dokumentasi

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Pesatnya perubahan teknologi informasi telah membawa perkembangan pola pikir pertimbangan dan bentuk analisa teknologi informasi yang saat ini menjadi sebuah alat untuk membantu kita merampungkan persoalan yang membutuhkan penyelesaian. Diantara beberapa persoalan yang ada yaitu tentang bidang pemesanan barang. Ketersediaan informasi transaksi pelanggan mendorong pengembangan teknik yang secara otomatis. Database penjualan menyimpan sejumlah *record* transaksi penjualan yang sangat besar. Banyaknya persaingan dalam dunia bisnis, menuntut para pengembang untuk menemukan suatu strategi yang dapat meningkatkan penjualan.

Data mining memiliki sifat *automacally* secara interaktif dalam menentukan pola informasi-informasi saat ini dan atau memperkirakan tingkah (*trend*) di masa yang akan datang. Terdapat beberapa metode dalam *data mining*, salah satunya adalah metode asosiasi (*association*). Metode asosiasi akan meng-asosiasikan data dengan menggunakan aturan *apriori* (pengalaman) yang memenuhi syarat minimum *support* (nilai penunjang), yaitu kombinasi tiap item dalam *database* dan syarat minimum *confidence* (Jaya, 2018).

Algoritma *apriori* masuk dalam kategori kaidah asosiasi pada *data mining*, algoritma *apriori* untuk menemukan *frequent item sets* dijalankan pada sekumpulan data. Analisis *apriori* didefenisikan suatu proses untuk menemukan

Semua aturan apriori yang memenuhi syarat minimum untuk *support* dan syarat minimum untuk *confidence*. Penggunaan algoritma apriori yang telah dilakukan oleh peneliti sebelumnya, Hal ini menjadi dasar untuk membuat aplikasi *data mining* untuk analisis penjualan dengan algoritma apriori berbasis *web* dengan menggunakan data contoh (sampel data) dari *Microsoft adventure work* (Nursikuwagus et al., 2016).

Toko buah Awi yang bergerak dalam penjualan buah memanfaatkan cara-cara ilmiah dalam teknik pengorderan barang. Dilihat dari hasil penelitian sebelumnya mendorong peneliti untuk melakukan penelitian dari permasalahan yang terjadi di Toko buah Awi bahwa algoritma apriori sangat efisien dalam membangun pembentukan kombinasi item yang dapat dikategorikan berdasarkan nilai parameternya yang nantinya akan menghasilkan perhitungan yang dapat membantu untuk menentukan strategi penjualan.

Penelitian ini di latar belakang dari hasil pengamatan, saat ini dalam melakukan transaksi penjualan toko buah Awi belum menggunakan komputer dan mencatat data transaksi yang ada, namun data tersebut hanya berfungsi sebagai arsip penjualan saja sehingga belum digunakan sebagai data untuk memprediksi hasil penjualan buah yang akan datang yang diminati konsumen. Hal ini bertujuan menyediakan data produk yang banyak dibeli oleh pelanggan agar usaha yang dirintis terus berjalan, sehingga *data mining* sangat cocok untuk kasus seperti ini.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan diatas maka identifikasi masalah dalam penelitian ini adalah :

1. Sulitnya mengetahui produk (item) yang sering terjual sehingga Toko Buah Awi mengetahui buah apa saja yang sudah tidak memiliki stok lagi.
2. Minimnya pemanfaatan data transaksi pada sistem penjualan buah di Toko Buah Awi.
3. Kesulitan dalam mengolah informasi penjualan buah di Toko Buah Awi, sehingga kurang efektifnya pengelolaan data penjualan buah.

1.3 Batasan Masalah

Mengingat masalah yang dijelaskan dalam penelitian ini sangatlah luas, agar tidak menyimpang dari tujuan utama penulis, maka penulis membatasi masalah-masalah yang akan ditulis, diantaranya adalah sebagai berikut:

1. Penelitian data mining ini menggunakan data transaksi penjualan buah di Toko Buah Awi dalam kurun waktu 5 tahun terakhir.
2. Aplikasi data mining yang digunakan untuk pengujian data penelitiannya adalah aplikasi Tanagra.
3. Metode yang digunakan pada penelitian ini adalah algoritma apriori.

1.4 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas maka rumusan dalam penelitian ini adalah:

1. Bagaimana menerapkan algoritma apriori untuk mengetahui pola kombinasi *itemsets* yang dapat mempermudah pasar dalam mengolah data penjualan buah?
2. Bagaimana cara mengetahui jenis buah yang terjual paling banyak dan stock buah yang masih tersedia?
3. Bagaimana manfaat *data mining* dalam sistem penjualan buah bagi penjual dan pembeli?

1.5 Tujuan Penelitian

Sebagaimana identifikasi permasalahan yang terjadi, maka tujuan yang akan dicapai antara lain:

1. Untuk mengetahui pola kombinasi *itemsets* penjualan buah sehingga dapat mempermudah dalam pengelolaan penjualan buah.
2. Untuk mengetahui jenis buah yang paling banyak terjual, sehingga digunakan sebagai bahan pertimbangan stock buah untuk masa mendatang.
3. Untuk mengetahui manfaat penerapan *data mining* dalam sistem penjualan buah di Toko Buah Awi, sehingga lebih efektifnya data penjualan buah.

1.6 Manfaat Penelitian

Manfaat penelitian dapat dikategorikan menjadi 2, yaitu manfaat teoritis dan manfaat praktis.

1.6.1 Manfaat Teoritis

1. Menambah wawasan yang sudah ada tentang data mining menggunakan algoritma apriori.
2. Sebagai salah satu referensi untuk mengetahui pemanfaatan data sistem penjualan.

1.6.2 Manfaat Praktis

1. Manfaat aplikasi ini untuk pasar adalah sebagai alat bantu untuk mengelola data sistem penjualan.
2. Diharapkan hasil penelitian ini berguna bagi Universitas Putera Batam sebagai bahan bacaan dan referensi bagi mahasiswa lainnya.

BAB II

TINJAUAN PUSTAKA

2.1 *Knowledge discovery in database (KDD)*

Knowledge discovery in database adalah suatu rangkaian proses yang digunakan oleh suatu lembaga untuk mendeskripsikan, memperbaharui, menerangkan, dan menyebarluaskan informasi untuk digunakan kembali, diketahui, dan dipelajari di dalam organisasi. Kegiatan ini biasanya terkait dengan objektif organisasi dan ditujukan untuk mencapai suatu hasil tertentu seperti wawasan bersama, pengembangan kinerja, keberhasilan daya saing, atau tingkat pembaharuan yang lebih baik (Sistem, Fakultas, Komputer, & Sriwijaya, 2016).

Knowledge Discovery in Database (KDD) adalah proses yang melingkupi pengambilan, penggunaan database untuk memperoleh kesamaan, rule atau keterkaitan dalam set data yang bermedium besar. Di dalam jurnal yang berjudul “Pembentukan Cluster dalam *Knowledge in Database* dengan Algoritma *K-Means*“. *Knowledge Discovery in Database (KDD)* didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data.

Istilah *data mining* dan *knowledge discovery in database (KDD)* sering kali digunakan secara bergantian untuk memahami proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining* (Studi, Informasi, Ilmu, Universitas, & Kuning, 2016).

2.1.1 Proses *Knowledge Discovery in Database*

Proses *knowledge discovery in databases (KDD)* secara garis besar dapat dijelaskan sebagai berikut (Pujiono, Budiono, & Fahmi, 2014) :

1. *Data selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing/cleaning*

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi focus KDD. Proses cleaning mencakup antara lain membuang penggandaan data, memeriksa data yang *inkonsisten*, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses "memperkaya" data yang sudah ada dengan data atau informasi lain yang relevan yang diperlukan untuk KDD, seperti data atau informasi eksternal.

3. Integrasi data

Integrasi data merupakan penggabungan data dari berbagai database kedalam database baru.

4. *Transformation*

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

merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

5. *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 bergantung pada tujuan dan proses KDD secara keseluruhan.

6. *Interpretation/Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dipahami 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 hipotesis yang ada sebelumnya.

7. Presentasi Pengetahuan (*knowledge*)

Merupakan visualisasi dan penyajian pengetahuan mengenai teknik yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat.

2.2 Data Mining

Data mining adalah kegiatan mengolah dengan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mendistribusikan dan

mendeskripsikan informasi yang berguna dan wawasan yang berhubungan dari berbagai data yang disimpan. Pada saat kita dihadapkan kepada sekumpulan data dari suatu objek atau kejadian, kita perlu mengolahnya untuk mendapatkan manfaat dari data itu. Kita perlu mengenali polanya sehingga kita akan menemukan kecenderungan dari data tersebut (Lunak, 2017).

Kemudian kita dapat juga melakukan perkiraan atas apa yang akan terjadi pada jangka waktu kedepannya berdasar data masa sebelumnya berkaitan dengan data jumlah penjualan tersebut. Jadi pengenalan pola adalah suatu disiplin ilmu yang mempelajari bagaimana kita mengelompokkan obyek ke berbagai kelas dan bagaimana dari data bisa kita temukan kecenderungannya. Yang pertama mengacu pada kasus klasifikasi dan yang kedua mengacu pada regresi. *Data mining* juga mengacu pada langkah- langkah menentukan variable atau fitur yang penting untuk dipakai dalam klasifikasi dan regresi. *Data mining* memegang peranan penting dalam bidang industry, keuangan, cuaca, ilmu dan teknologi.

Berikut ini adalah beberapa contoh yang memperlihatkan masalah-masalah dalam data mining (Lunak, 2017) :

1. Memprediksi harga suatu saham dalam beberapa bulan kedepan berdasarkan performansi perusahaan dan data-data ekonomi.
2. Memprediksi apakah seorang pasien yang diopname akan mendapat serangan jantung berikutnya berdasarkan catatan kesehatan sebelumnya dan pola makanannya.
3. Memprediksi permintaan semen dalam beberapa tahun mendatang berdasarkan data permintaan semen di tahun-tahun sebelumnya.

4. Memprediksi apakah akan terjadi tornado berdasarkan informasi dari sebuah radar tentang kondisi angin dan kondisi atmosfer yang lain.
5. Identifikasi apakah sudah terjadi penipuan terhadap kartu kredit dengan melihat catatan transaksi yang tersimpan dalam database perusahaan kartu kredit (Lunak, 2017).

Data mining menetapkan *set size* dan *set size frequency* pada data transaksi sehingga mengetahui ukuran dan frekuensi transaksi yang terjadi yang berisikan menggunakan algoritma apriori klasik yang sudah dikembangkan sebelumnya dan belum menggunakan teknik optimasi untuk memperoleh aturan asosiasi yang lebih efisien. *Data mining* dapat dibagi menjadi beberapa kelompok yaitu deskripsi, estimasi, prediksi, klasifikasi, klustering dan asosiasi (Jaya, 2018).

Data mining adalah suatu proses menemukan hubungan pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika, *Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pujiono et al., 2014).

Karakteristik *data mining* sebagai berikut (Pujiono et al., 2014):

1. *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. *Data mining* biasa menggunakan data yang sangat besar, biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
3. *Data mining* berguna untuk membuat keputusan yang kritis.

2.2.1. Pengelompokan *Data Mining*

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Tampubolon et al., 2013) :

1. *Description* (Deskripsi)

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

2. *Estimation*

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

3. Prediksi.

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

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

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

4. Klasifikasi

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

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

5. Pengklusteran (*Clustering*)

Pengklusteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang

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

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

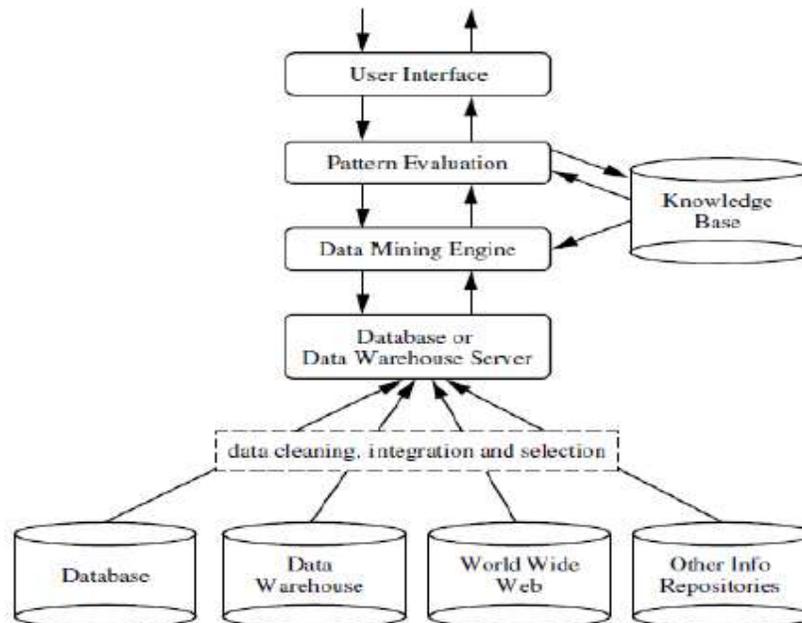
6. Asosiasi

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

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

2.2.2 Arsitektur Sistem *Data Mining*

Gambar berikut merupakan bagian-bagian dari arsitektur sistem *data mining*,



Gambar 2.1 Arsitektur Sistem *Data Mining*

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 menentukan 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 memungkinkan pengguna untuk melakukan browsing pada basis data dan data warehouse. Menevaluasi pola tersebut dengan tampilan yang berbeda-beda.

2.3 Algoritma Association

Algoritma aturan asosiasi akan menggunakan data latihan, sesuai dengan definisi *data mining*, untuk menghasilkan pengetahuan. Pengetahuan seperti apa akan dihasilkan dalam aturan asosiasi? Pengetahuan untuk mengetahui *item-item* belanja yang sering dibeli secara bersamaan dalam suatu waktu. Aturan asosiasi yang berbentuk “*if...then...*” atau “jika...maka...” merupakan pengetahuan yang dihasilkan dari fungsi aturan asosiasi (Studi et al., 2016).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap :

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:

$Support(A) = \frac{\text{aksi TotalTransndungA saksiMengaJumlahTran}}{\text{aksi TotalTransndungA saksiMengaJumlahTran}}$ Sedangkan nilai dari *support* 2 item diperoleh dari rumus berikut :

$Support(A,B) = P(A \cap B) = \frac{\text{aksi TotalTransndungAdanB saksiMengaJumlahTran}}{\text{aksi TotalTransndungAdanB saksiMengaJumlahTran}}$

2. Pembentukan aturan assosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan assosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan assosiatif A Nilai *confidence* dari aturan A diperoleh dari rumus berikut :

$Confidence = P(B|A) = \frac{\text{ndungA saksiMengaJumlahTran}}{\text{ndungA saksiMengaJumlahTran}}$

2.4 Algoritma Apriori

Algoritma apriori masuk dalam kategori kaidah asosiasi pada *data mining*, *Algoritma apriori* adalah *algoritma* yang paling terkenal untuk menemukan pola frekuensi tinggi. *algoritma apriori* dibagi menjadi beberapa tahap yang disebut narasi atau pass (Studi et al., 2016):

1. Pembentukan kandidat *itemset*, kandidat *k-itemset* dibentuk dari kombinasi ($k-1$)-*itemset* yang didapat dari iterasi sebelumnya. Satu cara dari *algoritma apriori* adalah adanya pemangkasan kandidat *k-itemset* yang *subset*-nya yang berisi $k-1$ item tidak termasuk dalam pola frekuensi tinggi dengan panjang $k-1$.
2. Penghitungan *support* dari tiap kandidat *k-itemset*. *Support* dari tiap kandidat *k-itemset* didapat dengan menscan *database* untuk menghitung jumlah transaksi yang memuat semua *item* didalam kandidat *k-itemset* tersebut. Ini adalah juga ciri dari *algoritma apriori* dimana diperlukan penghitungan dengan cara seluruh *database* sebanyak *k-itemset* terpanjang.
3. Tetapkan pola frekuensi tinggi. Pola frekuensi tinggi yang memuat k *item* atau *k-itemset* ditetapkan dari kandidat *k-itemset* yang *support*-nya lebih besar dari minimum *support*.
4. Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan. Bila tidak, maka k ditambah satu dan kembali bagian 1.

2.5 Analisa Keranjang Belanja

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, & A, 2016).

2.6 Software Pendukung

Tanagra adalah *software Data Mining* terbuka bagi akademik dan penelitian ini mengajukan beberapa metode *data mining* dari analisis pengembangan 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, Studi, & Informasi, 2016).

Tujuan utama dari proyek Tanagra adalah memberikan peneliti dan mahasiswa yang mudah untuk menggunakan perangkat lunak *data mining*, sesuai dengan norma-norma yang hadir dari pengembangan perangkat lunak dalam domain ini (terutama dalam desain GUI dan cara untuk menggunakannya), dan memungkinkan untuk menganalisis baik data yang nyata atau sintetis.

Tujuan kedua Tanagra adalah untuk mengusulkan kepada peneliti arsitektur yang memungkinkan mereka untuk dengan mudah menambahkan metode penambangan data mereka sendiri, untuk membandingkan kinerja mereka. Tanagra bertindak lebih sebagai platform eksperimental untuk membiarkan mereka pergi ke penting dari pekerjaan mereka, pengeluaran mereka untuk berurusan dengan bagian menyenangkan dalam *programmation* semacam ini alat pengelolaan data.

Tujuan ketiga dan terakhir, arah pengembang pemula, terdiri dalam menyebarkan metodologi yang mungkin untuk membangun perangkat lunak semacam ini. Mereka harus mengambil keuntungan dari akses gratis ke kode sumber, untuk melihat bagaimana perangkat lunak semacam ini dibangun, masalah untuk menghindari, langkah-langkah utama dari proyek ini, dan alat-alat dan perpustakaan kode yang digunakan untuk. Dengan cara ini, Tanagra dapat dianggap sebagai alat pedagogis untuk belajar teknik pemrograman.

2.7 Penelitian Terdahulu

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

1. Penelitian yang dilakukan oleh (Prisilla, Farmadi, & Candra, 2014) yang berjudul **IMPLEMENTASI METODE ALGORITMA APRIORI PADA SISTEM PENDUKUNG KEPUTUSAN**, pada penelitian ini membahas tentang algoritma apriori dapat diperolehnya beberapa rule dapat menentukan sistem pendukung keputusan yang dapat digunakan sebagai acuan. Hasil dari penelitian

ini adalah berupa keputusan bahwa barang yang sering dibeli konsumen kemungkinan berkisar antara rok dan blus, sehingga untuk kedepannya pada proses order barang pihak manajemen toko dapat menggunakan keputusan dari sistem yang telah dibangun.

2. Penelitian yang dilakukan oleh (Jaya, 2018) yang berjudul **ANALISIS PENJUALAN PRODUK RETAIL DENGAN METODE DATA MINING ASOSIASI**, pada penelitian ini membahas tentang penyediaan barang memiliki sifat yang saling berhubungan atau membentuk aturan asosiasi sehingga dapat membantu dalam analisis penjualan produk retail yang efektif. Hasil dari penelitian ini adalah pembentukan aturan apriori menunjukkan keterkaitan antar produk yang sering dibeli secara bersamaan antara lain tolak angin sachet, farm house sosis sapi, dll.

3. Penelitian yang dilakukan oleh (Nursikuwagus et al., 2016) yang berjudul **IMPLEMENTASI ALGORITMA UNTUK ANALISIS PENJUALAN DENGAN BERBASIS WEB**, pada penelitian ini membahas tentang aplikasi berdasarkan algoritma apriori dalam penjualan dengan berbasis web yang dapat membantu dalam memahaminya. Hasil penelitian ini memberikan paparan mengenai hasil dari implementasi dari algoritma apriori dengan bahasa pemrograman berbasis web. Beberapa hasil menunjukkan setiap proses yang dilakukan oleh tahapan algoritma apriori.

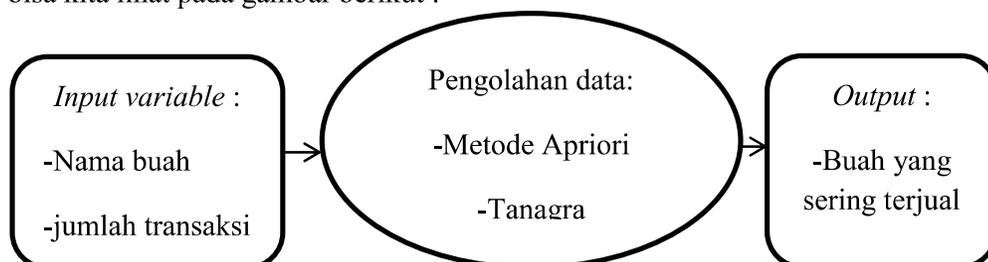
4. Penelitian yang dilakukan oleh (Apriori, 2016) yang berjudul **IMPLEMENTASI DATA MINING UNTUK PENENTUAN POSISI BARANG PADA RAK MENGGUNAKAN METODE APRIORI PADA PT**

MIDI UTAMA INDONESIA, pada penelitian ini membahas membantu pihak alfamidi untuk menyusun penempatan produk yang dijual berdasarkan penentuan posisi barang pad rak dengan menggunakan data mining dalam pengolahannya. Hasil penelitian ini adalah perhitungan dengan algoritma apriori yang berupa kombinasi itemsets atau keterkaitan barang dapat digunakan untuk melakukan penataan barang dalam rak atau etalase.

5. Penelitian yang dilakukan oleh (Handoko, 2016) yang berjudul **PENERAPAN DATA MINING DALAM MENINGKATKAN MUTU PEMBELAJARAN PADA INSTANSI PERGURUAN TINGGI MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI PROGRAM STUDI TKJ AKADEMI KOMUNITAS SOLOK SELATAN)**, pada penelitian ini membahas tentang penerapan *data mining* dengan menggunakan metode clustering k-means untuk meningkatkan mutu pembelajaran pada instansi perguruan tinggi studi kasus program studi TKJ akademi komunitas solok selatan. Hasil penelitian ini adalah penerapan *data mining* dengan metode clustering k-means sehingga menghasilkan cluster-cluster dalam meningkatkan mutu pembelajaran.

2.8 Kerangka Pemikiran

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



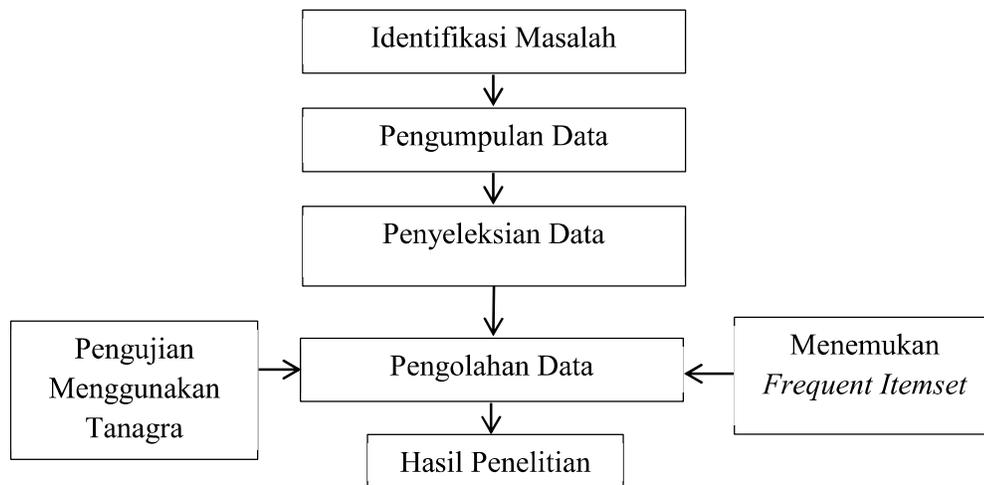
Gambar 2. 2 Kerangka Pemikiran

Pada gambar 2.2, dapat dijelaskan bahwa proses awal dari penelitian ini adalah adanya variabel yang telah dipilih berdasarkan nama buah dan jumlah transaksi akan diseleksi terlebih dahulu, agar tidak ada tipografi. Selanjutnya data ini akan diproses dengan metode Apriori lalu diuji dengan Tanagra untuk menemukan buah yang paling banyak terjual dan untuk menentukan apakah hasil dari *software* pengujian sama dengan cara manual. Hasil akhir yaitu bisa mengetahui buah yang sering terjual, pada hasil akhir inilah yang akan diterapkan pada toko tersebut sehingga membantu dalam pemilihan *stock*.

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

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

Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Data yang dapat diambil oleh peneliti adalah data observasi langsung dari pihak toko, wawancara secara langsung kepada kepada pihak toko buah Aviari 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. 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 sama, menghilangkan kesalahan pada data dan menghilangkan data yang kosong.

4. Pengolahan data

Data diinput kedalam bentuk yang lebih berarti berupa informasi, Kemudian data yang telah dipilih dan diseleksi akan diproses dengan cara algoritma *Apriori* yaitu secara manual dengan menemukan item *frequenset*-nya lalu akan diuji kembali dengan menggunakan *software* Tanagra.

5. Hasil penelitian

Setelah beberapa proses yang dilakukan pada penelitian tentu saja akan memberikan sebuah hasil yang penelitian yang dicari. Hasil penelitian ini akan diuji menggunakan *software* Tanagra dengan metode *Apriori*, 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(Kunang et al., 2013):

1. Pengamatan (Observasi)

Melakukan pengamatan langsung ke toko buah untuk mendapatkan data yang dibutuhkan.

2. Wawancara (*Interview*)

Mengadakan wawancara dengan pihak-pihak yang berkaitan langsung dengan permasalahan yang sedang dibahas pada tugas akhir ini untuk memperoleh gambaran dan penjelasan secara mendasar.

3. Studi Pustaka

Penulis mengumpulkan berbagai referensi dan literatur pendukung penelitian berupa buku, jurnal dan artikel yang berasal dari berbagai sumber yang erat kaitannya dengan objek permasalahan.

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.

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-buah inilah yang sangat mempengaruhi penelitian ini,

karena ia yang akan menentukan kombinasi-kombinasi *itemset* mana yang paling banyak terjual.

2. Nama buah, yaitu nama-nama buah yang sering dibeli oleh pembeli dan merupakan *item-item* sebagai pemilihan stock.

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 Buah Awi

Transaksi	Tanggal	Nama Item
1	02/12/2018	Jeruk
		Semangka
		Pisang
2	06/12/2018	Kelengkeng
		Apel
		Pepaya
3	08/12/2018	Pear
		Pisang
4	10/12/2018	Apel
		Papaya
		Duku
5	12/12/2018	Pear
		Papaya
		Alpukat
		Sirsak

Transaksi	Tanggal	Nama Item
		Anggur
		Mangga
6	15/12/2018	Duku
		Pear
		Semangka
		Jeruk
7	16/12/2018	Duku
		Apel
8	18/12/2018	Semangka
		Pisang
9	19/12/2018	Kiwi
		Pear
		jeruk
10	24/12/2018	Sirsak
		Anggur
		Kiwi
		Mangga
11	25/12/2018	Sirsak
		Anggur
12	26/12/2018	Pepaya
		Alpukat

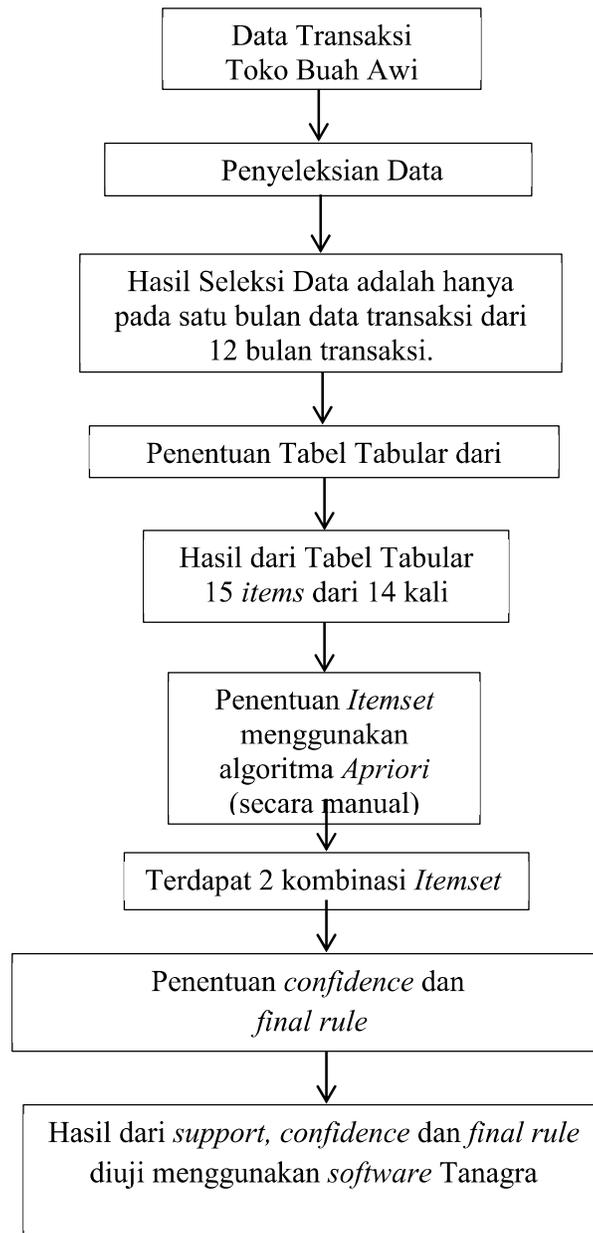
Transaksi	Tanggal	Nama Item
		Apel
		Nanas
13	27/12/2018	Mangga
		Sirsak
		Anggur
		Nanas
14	31/12/2018	Jambu air
		jeruk

Tabel 3.1

Sumber: Data olahan penulis, 2018

3.4 Metode Perancangan Sistem

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



Gambar 3. 2 Metode Perancangan Sistem

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 Buah SP Plaza yang berlokasi di Toko buah Awi blok D No 2 Batu Aji, Batam.

3.5.2 Jadwal Penelitian

Jadwal penelitian ini dilakukan mulai dari bulan September 2018 hingga bulan Januari 2019, dengan rincian sebagai berikut:

Tabel 3.2 Jadwal Penelitian

No.	Kegiatan	Waktu Penelitian																							
		September 2018				Oktober 2018				November 2018				Desember 2018				Januari 2019							
		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