

BAB 2

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Beberapa penelitian tentang implementasi market basket analysis. Beberapa hasil penelitian yang relevan dengan penelitian ini dalam 5 tahun terakhir akan dipaparkan dalam beberapa paragraph dibawah ini sebagai bahan rujukan.

Penelitian tentang market basket analysis yang pertama ini dilakukan oleh Rudy Ariyanto, Hendra Pradibta, dan Christine Puspa Sanjaya di tahun 2019 dengan judul “Penerapan Market Basket Analysis dengan Metode Algoritma Apriori untuk Menganalisis Pola Penjualan Batik Khas Banyuwangi di Sekar Kedaton”. Penelitian menyebutkan bahwa dalam bidang promosi kain batik yang dijual sangatlah kurang dikarenakan butik tersebut hanya melakukan penjualan secara offline. Karena banyaknya konsumen, butik Sekar Kedaton tidak dapat menentukan kain batik apa saja yang lebih diminati oleh pihak konsumen dan dibeli dengan bersamaan dalam satu transaksi sehingga kain batik yang tidak diminati konsumen mengalami penumpukan stok dan dapat merugikan butik dalam meningkatkan penjualan. Jumlah data produk dan transaksi yang cukup besar mengakibatkan butik tersebut mengalami kesulitan dalam manajemen data, yang dimana setiap waktu data yang masuk semakin bertambah ini menyebabkan data tersebut menjadi data yang tidak berguna. Oleh sebab itu diperlukan metode Market Basket dalam menentukan produk manakah yang nantinya akan dibeli dalam satu waktu sekaligus dengan melakukan analisa pada data transaksi konsumen. Dengan mengetahui daftar produk tersebut, maka sebuah sistem yang berbasis website ini dapat memudahkan pemilik butik dalam pengambilan keputusan untuk informasi mengenai pola belanja konsumen dan membantu dalam meningkatkan persediaan barang untuk penjualan selanjutnya [3].

Penelitian kedua tentang market basket analysis dilakukan oleh Septianti Ari Mindiawati Hasibuan dan Berto Nadeak di tahun 2018 yang berjudul “ Penerapan Data Mining Dengan Menggunakan Algoritma Ct-Pro Untuk Korelasi Obat. Penelitian ini menyatakan bahwa Kumpulan data transaksi sering kali hanya disimpan di dalam suatu database dan kurang diketahui pemanfaatannya. Data penjualan tersebut bisa diolah lagi sehingga didapatkan korelasi antara obat yang biasa dibeli bersamaan dalam satu transaksi. Selain itu dapat digunakan untuk peletakan obat yang sering dibeli secara berdekatan di dalam rak obat yang dijual pada Apotek Kimia Farma. Teknologi data mining menjadi jalan keluar bagi para manajer perusahaan untuk menentukan strategi pemasaran dan korelasi antara obat yang dibeli oleh customer sehingga bisa meningkatkan pelayanan pada konsumen. Dengan masalah yang ada pada Apotik Kimia Farma, maka peneliti mencoba memberikan suatu alternatif dengan menerapkan Algoritma CT-Pro. Algoritma market basket analysis digunakan untuk menghasilkan association rule, dengan pola “if-then” atau “jika-maka” [4].

Penelitian ketiga tentang marketbasket analysis dilakukan oleh Sholikin pada tahun 2019 yang berjudul “Algoritma Fp-growth Dalam Teknik Market Basket Analysis Sparepart Komputer Pada Jasa Servis Komputer (Studi Kasus : CV. Karya Computer Center)” Penelitian ini menyatakan bahwa Market Basket Analysis adalah suatu metode untuk menemukan pola belanja konsumen, misalnya menemukan produk B sering dibeli bersamaan dengan produk C dalam satu transaksi dan jumlah tertentu. Penelitian ini memiliki tujuan dalam pembuatan aplikasi dari data mining dengan menggunakan association rule atau aturan asosiasi dengan algoritma FP-Growth sebagai teknik analisis dari penjualan sparepart. Pengambilan data yang dilakukan dalam penelitian ini yaitu data transaksi penjualan selama satu bulan di CV. Karya Computer Center. FP-Growth adalah salah satu algoritma dari data mining yang digunakan dalam menentukan suatu himpunan data yang sering sekali muncul dalam suatu kumpulan data. Aturan

asosiasi menghasilkan suatu kombinasi dari beberapa barang yang dipasarkan dan sering dibeli oleh konsumen. Dengan hasil ini nantinya dapat membantu dalam merancang manajemen strategi pemasaran sparepart di CV. Karya Computer Center.

2.1.1 Ringkasan Penelitian Terkait

Berikut adalah ringkasan penelitian terkait mengenai metode yang akan digunakan pada penelitian ini :

Tabel 2.1 Ringkasan Penelitian Terkait

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
1.	Rudy Ariyanto, Hendra Pradibta, dan Christine Puspa Sanjaya, 2019	Butik Sekar Kedaton tidak dapat menentukan produk kain batik yang banyak diminati oleh konsumen dan dibeli bersamaan dalam sewaktu.	Metode Algoritma Apriori	Memudahkan pemilik butik dalam pengambilan keputusan untuk informasi mengenai pola belanja konsumen serta dapat membantu dalam meningkatkan persediaan barang untuk penjualan selanjutnya
2.	Septianti Ari Mindiawati Hasibuan dan Berto Nadeak 2018	Kumpulan data transaksi yang begitu besar sering kali hanya disimpan di dalam suatu	Metode Algoritma Ct-Pro	Membantu pihak apotik untuk mengetahui korelasi antara obat yang diminati oleh konsumen dalam

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
		database dan kurang digali pemanfaatannya		penyusunan peletakan obat.dengan memanfaatkan transaksi penjualan obat dengan pasien
3.	Sholikin, 2019	Data transaksi penjualan diinput dalam suatu sistem informasi dan data hasil inputan digunakan hanya untuk keperluan administrasi.	Metode Algoritma Fp-growth	Membantu untuk perancangan strategi pemasaran sparepart di CV. Karya Computer Center.

2.2 Landasan Teori

2.2.1 Pola Belanja Konsumen

Pola Belanja adalah bentuk terstruktur dari kegiatan belanja konsumen di dutik, swalayan, toko, pasar maupun tempat perbelanjaan lainya. Dari kegiatan belanja tersebut maka akan menghasilkan data transaksi belanja yang bisa diolah untuk menentukan pola belanja selanjutnya .Dari pola belanja konsumen inilah yang dapat diprediksi untuk pembuat keputusan strategi yang lebih efektif [5].

2.2.2 Data Mining

Data mining merupakan ilmu yang mempelajari sebuah metode dalam menemukan pola dari suatu data besar. Pengetahuan dapat menghasilkan berupa pola atau rumus [6].

Setiap hari manusia dapat menghasilkan banyak data yang besar baik jumlah dan ukurannya. Data-data tersebut diantaranya tentang data bisnis, kantor, penduduk dan sebagainya. Contoh dalam bidang bisnis busana ada banyak data mengenai data produk itu sendiri yang setiap hari akan bertambah banyak karena banyaknya pembeli di toko busana tersebut. Data itu akan terekap dalam computer dalam dataset yang nantinya akan diolah dengan metode dan algoritma yang akan menghasilkan pola belanja konsumen..

2.2.2.1 Tahapan Proses dalam Data Mining

Tahapan data mining yang merupakan Knowledge Discovery in Database (KDD) secara garis besar dijelaskan sebagai berikut [6]:

1. Data Cleaning

Proses memeriksa data yang tidak tetap, memperbaiki kesalahan data seperti penulisan yang salah, dan membuang duplikasi data. Pada dasarnya lebih baik data yang diperoleh dari database perusahaan ataupun hasil percobaan, mempunyai data yang kurang sempurna atau hilang, data yang tidak valid atau salah ketik. Selain itu, ada juga atribut data yang kurang relevan dengan hipotesis data mining yang dimiliki. Data cleaning akan berpengaruh dengan hasil informasi dari suatu teknik data mining karena data yang diolah akan berkurang jumlahnya dan kompleksitasnya.

2. Data Integration

Proses penambahan data atau informasi lain yang sudah ada dan relevan yaitu penggabungan data dari berbagai database ke dalam database baru yang dibutuhkan oleh KDD.

3. Data Selection

Memilih data yang relevan juga bisa dilakukan dengan Analisa dari data operasional. Dari data hasil pemilihan untuk data mining tersebut akan disimpan terpisah dalam database.

4. Data Transformation

Proses data transformasi kedalam bentuk format tertentu sehingga data dapat disesuaikan untuk proses data mining. Sebagai contoh metode seperti analisis klustering atau asosiasi hanya dapat menerima input data kategorikal saja.

5. Data Mining

Merupakan Proses pencarian pola atau informasi dalam data yang terpilih dengan metode atau teknik tertentu.

6. Pattern Evaluation

Merupakan identifikasi pola yang benar dari hasil data mining. Dari tahap ini hasil teknik data mining berupa pola ataupun evaluasi prediksi untuk menilai hipotesis ini tercapai atau tidak.

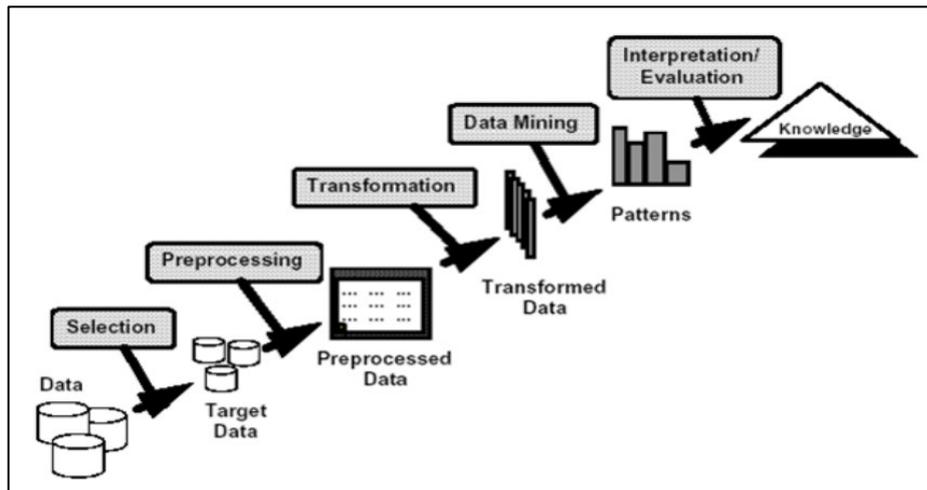
7. Knowledge Presentation

Tahap ini akan menampilkan suatu pola informasi yang dihasilkan dari proses data mining tersebut, visualisasi ini akan membantu menampilkan hasil dari data mining dalam bentuk yang sederhana.

Tahapan proses dari penggunaan data mining yang merupakan proses Knowledge Discovery in Database (KDD) dapat dijelaskan sebagai berikut [7]:

1. Memahami domain dari aplikasi dalam mengetahui dan menggali pengetahuan awal dan juga apa sasaran pengguna.
2. Membuat setiap target dataset yang terdiri dari pemilihan data dan fokus dalam sub-set data.
3. Transformasi data yaitu eliminasi , outliners, missing value serta pemilihan fitur dan reduksi dimensi.

4. Penggunaan algoritma data mining yang terdiri dari asosiasi, klasifikasi, sekuensial, klusterisasi, dan lainnya.
5. Interpretasi, evaluasi dan visualisasi dari awal pencarian pengetahuan dari suatu pola untuk mengetahui adakah sesuatu baru dan menarik yang dilakukan iterasi jika diperlukan.



Gambar 2.1 Proses KDD [8]

2.2.2.2 Pengelompokan Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan apa yang harus dilakukan, yaitu [9]:

1. Deskripsi

Peneliti dan analis secara sederhana terkadang ingin mencoba mencari cara untuk menggambar pola dan kecenderungan yang terdapat dalam suatu data. Sebagai contoh, dalam pilpres petugas pengumpul suara tidak dapat menemukan keterangan bahwa siapa yang kurang profesional akan sedikit didukung dalam pilpres. Deskripsi ini akan cenderung memberi penjelasan untuk suatu pola.

Tujuan dari deskripsi adalah untuk mengidentifikasi pola yang muncul secara berulang pada suatu data dan mengubah pola tersebut menjadi aturan dan kriteria yang dapat dengan mudah dimengerti oleh para ahli pada domain aplikasinya. Aturan yang dihasilkan harus

mudah dimengerti agar dapat dengan efektif meningkatkan tingkat pengetahuan (knowledge) pada sistem. Tugas deskriptif merupakan tugas data mining yang sering dibutuhkan pada teknik postprocessing untuk melakukan validasi dan menjelaskan hasil dari proses data mining. Postprocessing merupakan proses yang digunakan untuk memastikan hanya hasil yang valid dan berguna yang dapat digunakan oleh pihak yang berkepentingan.

2. Estimasi

Estimasi ini hampir sama seperti klasifikasi, kecuali variabel estimasi lebih ke numerik daripada kategori. Model ini menggunakan record lengkap juga menyediakan nilai dari variabel prediksi.

Estimasi juga hampir sama dengan prediksi, namun variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi

3. Prediksi

Prediksi hampir sama seperti klasifikasi dan estimasi, kecuali dalam prediksi nilai akan muncul di masa mendatang. Prediction atau fungsi prediksi merupakan salah satu fungsi data mining. Maksudnya yaitu dari proses nanti akan menemukan pola tertentu dari suatu data. Pola tersebut dapat diketahui dari variabel-variabel yang ada pada data. Pola yang didapat bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya.

Karena itulah fungsi satu ini dikatakan sebagai fungsi prediksi. Nantinya bisa digunakan untuk memprediksi variabel tertentu yang tidak ada dalam suatu data. Hal ini tentunya memudahkan dan menguntungkan bagi mereka pemilik kepentingan yang memerlukan prediksi akurat untuk membuat hal penting tersebut menjadi lebih baik.

Contoh prediksi dalam bisnis dan penelitian adalah:

- a. Prediksi harga gula dalam lima bulan mendatang.
- b. Prediksi persentase kenaikan kelas ditahun depan jika sampai saat ini siswa malas belajar.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel 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 bukan.
- b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- c. Mendiagnosa penyakit seorang pasien untuk mendapatkan termasuk kategori apa.

5. Pengklusteran

Pengklusteran merupakan pengelompokan pengamatan, record, membentuk dan memperhatikan kelas objek yang mempunyai kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan lainnya dan ketidakmiripan dengan record dalam kluster

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

6. Asosiasi

Tugas asosiasi dalam Data mining adalah menemukan atribut yang muncul dalam suatu 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. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli bersamaan.

2.2.3 Market Basket Analysis

Market basket analysis adalah Analisa data atas suatu kegiatan belanja konsumen disuatu tempat perbelanjaan manapun. Sumber data dari market basket analysis antara lain bersumber dari transaksi belanja di berbagai pusat perbelanjaan. Market basket analysis dimanfaatkan sebagai permulaan dalam pencarian pengetahuan dari sebuah transaksi ketika tidak diketahui pola secara spesifik apa yang akan cari. Kebutuhan dari market basket analysis bermula dari keakuratan dan manfaat yang nantinya dihasilkn dalam wujud sebuah aturan assosiasi (association rules). Association rules memiliki maksud mengenai keterkaitan antara pola-pola dari data dalam basis data.

Tujuan dari Market Basket Analysis yaitu untuk mengetahui keterkaitan atau hubungan antar produk yang kemungkinan dibeli oleh konsumen dalam satu transaksi. Dalam Market basket analysis membutuhkan data transaksi yang besar agar pola hubungan antar produk yang menjadi valid. Market basket analysis menghasilkan informasi mengenai produk mana saja yang biasanya dibeli dengan bersamaan. Contohnya jika seorang membeli sabun, maka kemungkinan akan membeli shampo. Informasi inilah yang akan diolah dari data mining dengan Market basket analysis. Produk yang biasa dibeli secara bersamaan akan diletakkan secara berdekatan agar pembeli mudah menemukan apa yang dicari. Dengan demikian para konsumen akan merasa puas dan penjualan juga akan meningkat.

Kelebihan dari proses *market basket analysis*:

1. Hasilnya jelas dan mudah dimengerti , contohnya : Jika sabun dan shampoo dibeli secara bersamaan , maka kemungkinan besar akan membeli pasta gigi sekaligus.
2. Market basket analysis berguna untuk mengolah data mining seperti pencarian pola belanja.

2.2.4 Algoritma Asosiasi

Algoritma asosiasi adalah bentuk algoritma data mining untuk memberikan informasi hubungan antar item data didatabase. Algoritma itu bisa dimanfaatkan dalam bisnis diantaranya dalam proses penjualan. Algoritma asosiasi bisa mempermudah dalam proses penjualan yaitu memberikan hubungan antar data penjualan yang dilakukan konsumen sehingga akan mendapat pola pembelian pelanggan. Pemilik perusahaan bisa memanfaatkan informasi untuk mengambil keputusan yang tepat. Ada beberapa algoritma asosiasi, namun disini akan dijelaskan sedikit mengenai algoritma asosiasi yaitu antara algoritma apriori, fp-growth dan Ct-Pro.

2.2.4.1 Algoritma Apriori

Algoritma apriori merupakan algoritma dasar yang dikemukakan oleh Agrawal & Srikant di tahun 1994 untuk menentukan frequent itemset dalam aturan asosiasi boolean. Algoritma apriori merupakan aturan asosiasi dalam data mining. Dalam menentukan aturan kombinasi item bisa menggunakan teknik data mining dengan menggunakan aturan asosiasi. Salah satu tahap analisis asosiasi pola frekuensi tinggi. Penting atau tidak suatu asosiasi bisa diketahui dari dua tolak ukur, yaitu support dan confidence. Nilai penunjang (support) yaitu persentase kombinasi item tersebut dalam database sedangkan nilai kepastian (confidence) adalah kuatnya hubungan antar-item dalam aturan asosiasi [10].

2.2.4.2 Algoritma Frequent Pattern Growth (FP-Growth)

Frequent Pattern Growth adalah salah satu algoritma yang bisa digunakan untuk menentukan himpunan data yang sering muncul dalam kumpulan data. Dalam Algoritma FP-Growth dapat menyimpan suatu informasi mengenai frequent itemset dalam bentuk struktur prefix-tree atau sering disebut FP-Tree [11].

Tidak melakukan candidate generation dalam proses pencarian frequent itemset, sehingga mengurangi scan database secara berulang dalam proses mining dan berlangsung cepat. Berikut ini adalah tahapan Proses Algoritma Fp-growth:

1. Penyiapan Dataset.
2. Pencarian itemset yang sering muncul.
3. Urutkan itemset berdasarkan jumlah terbesar hingga jumlah terkecil.
4. Buat fptree berdasarkan itemset yang sudah diurutkan.
5. Buat conditional pattern base.
6. Buat conditional fptree .
7. Buat frequent pattern.
8. Hitung nilai support dari frequent pattern yang sudah terbentuk.

$$\text{Support (A,B)} = \frac{\text{Transaksi Mengandung A dan B}}{\Sigma \text{Transaksi}} \times 100$$

9. Hitung nilai confidence dari frequent pattern yang sudah terbentuk.

$$\text{Confidence} = P(B|A) = \frac{\text{Transaksi mengandung A dan B}}{\Sigma \text{Transaksi mengandung A}} \times 100$$

10. Buat aturan asosiasi berdasarkan minimum support dan minimum confidence yang sudah ditentukan.

2.2.4.3 Algoritma Ct-Pro

Algoritma CT-Pro adalah aturan asosiasi pada data mining. Market basket analysis merupakan aturan asosiasi antara berbagai atribut. Teknik data mining dalam menemukan aturan asosiatif antara suatu kombinasi item biasanya menggunakan analisis asosiasi. Contoh aturan asosiatif dalam analisa belanja di pusat perbelanjaan adalah diketahuinya berapa besar kemungkinan membeli kopi sekaligus dengan gula. Dengan pengetahuan ini pemilik usaha dapat mengatur peletakan barang atau merancang promosi pemasaran dengan baik.

Langkah-langkah algoritma CT-PRO adalah sebagai berikut:

1. Menemukan item-item yang frequent
 - a. Data yang terkumpul harus diseleksi dahulu dan pilihlah data yang lengkap.
 - b. Kumpulan data tersebut, dilakukan transformasi data.
 - c. Setelah itu masing masing data dipilah berdasar minimum support yang telah ditentukan, dan akan mendapat Item Frequent Table.
 - d. Lalu masing-masing item dihitung frekuensi kemunculannya sehingga akan menghasilkan global item table.
2. Membuat CFP-Tree
 - a. Setelah ditemukan item-item yang frequent setelah itu lakukan pembangunan CFP-Tree. Frequent item yang ada diurutkan sesuai global item dari nilai yang terbesar ke terkecil.
 - b. Global item table diurutkan dari item terkecil lalu terbesar, karena algoritma CT-PRO bekerja dengan melakukan bottom-up mining.

- c. Masing-masing global item table di lakukan pencarian node yang berkaitan dengan item tersebut, setelah itu dipilah berdasarkan minimum supportnya maka diperoleh Local Item Table.
- d. Dari local item table dibuat CFP-Tree berdasarkan minimum support yang telah ditentukan.
- e. Kemudian item yang memenuhi minimal support dilakukan mining dengan rumus confidence.

2.2.5 Association rule

2.2.5.1 Pengertian Association rule

Association rule merupakan prosedur untuk mencari hubungan antara satu item dengan item yang lain. Association rule biasanya menggunakan “if” dan “then” misalnya “if B then C and D”, hal ini menunjukkan jika B maka C dan D. Untuk menentukan association rule maka harus menentukan support dan confidence untuk membatasi apakah rule tersebut menarik atau tidak [12].

Association rule digunakan untuk menemukan hubungan antar item dalam suatu transaksi, hubungan tersebut dapat mengetahui kuat atau tidaknya suatu aturan dalam asosiasi, Untuk menemukan keteraturan dalam suatu data adalah tujuan dari association rule. Association rule digunakan untuk mengetahui antar item produk yang kemungkinan dibeli secara bersamaan dengan produk lain dalam satu transaksi, dan mencari informasi mengenai produk tertentu. Untuk mencari association rule, maka perlu adanya variabel ukuran kepercayaan yang dapat ditentukan oleh user, untuk mengatur batasan sejauh mana dan sebanyak apa hasil output yang diinginkan oleh user.

2.2.5.2 Ukuran Kepercayaan Rule (Interestingness Measure)

Terdapat dua ukuran kepercayaan yang menunjukkan kepastian dan tingkat kegunaan suatu rule yang ditemukan yaitu [12]:

1. Support

Support merupakan tolak ukur dalam mengetahui berapa besar dominasi suatu item dari seluruh transaksi yang ada.

2. Confidence

Confidence merupakan suatu tolak ukur untuk mengetahui hubungan antar item secara conditional (semisal sesering apa item A dibeli jika orang membeli item B).

Untuk menemukan aturan asosiasi seperti yang diharapkan maka harus menemukan nilai dari support yang telah ditentukan. Support tersebut merupakan jumlah item pada setiap transaksi yang ada didalam database. Untuk dapat menemukan nilai support kita dapat mencari semua aturan yang jumlah support \geq minimum support. Dalam hal ini dapat digunakan sebagai cara untuk menemukan sebuah nilai confidence. Nilai confidence ditentukan dari nilai support suatu aturan dalam sebuah transaksi, apabila item pada setiap transaksi tidak sering muncul, maka kandidat yang tidak sesuai dengan nilai support \geq minimum support tersebut harus segera dipangkas tanpa harus menghitung confidencenya. Strategi biasa digunakan oleh banyak algoritma penggalian aturan asosiasi yaitu memecahkan masalah ke dalam dua pekerjaan utama.

3. Frequent Itemset Generation

Tujuannya yaitu untuk pencarian semua itemset yang sudah memenuhi batas minimum support. Itemset itu disebut itemset frequent atau yang sering muncul

4. Rules Generation

Mengekstrak aturan dengan confidence yang tinggi dari itemset frequent sehingga mendapatkan aturan asosiasi yang kuat (Strong rule) merupakan tujuan dari Rules Generation [13].

2.2.6 Algoritma Apriori

Bagian ini menjelaskan tentang algoritma Apriori sebagai metode yang digunakan dalam tugas akhir ini, yang meliputi definisi, langkah-langkah dan contoh kasus dengan menggunakan algoritma Apriori.

2.2.6.1 Pengertian Algoritma Apriori

Apriori merupakan algoritma yang menggunakan teknik association rule dalam pencarian frequent itemset. Algoritma Apriori dalam frequent itemset menggunakan knowledge yang telah diketahui sebelumnya, untuk mendapatkan proses informasi selanjutnya. Dalam algoritma Apriori untuk penentuan kandidat yang akan muncul dengan mengetahui nilai minimum support [14].

Algoritma apriori termasuk dalam *association rule* pada Data Mining. Metode selain algoritma apriori ini ada metode Hash Based dan Generalized Rule Induction. Market basket analysis adalah aturan dalam asosiasi antara beberapa atribut. Association rule mining atau analisis asosiasi merupakan teknik Data Mining yang digunakan dalam menemukan aturan asosiatif antara suatu kombinasi dari item. Metodologi dasar analisis asosiasi terbagi menjadi dua tahap:

1. Analisis pola frekuensi tinggi

Pada proses ini mencari kombinasi antar item yang memenuhi syarat minimum dari nilai support dalam database. Nilai support dari item akan diperoleh rumus sebagai berikut [14]:

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A}}{\text{Total transaksi}} \times 100$$

Nilai support dari 2 item diperoleh dengan menggunakan rumus:

$$\text{Support (A,B)} = \frac{\sum \text{Transaksi Mengandung A dan B}}{\sum \text{Transaksi}} \times 100$$

Itemset yang telah ditunjukkan pada frequent itemset memiliki frekuensi kemunculan lebih banyak nilai minimum yang ditentukan (ϕ). Misalkan $\phi = 2$, maka semua itemsets yang frekuensi kemunculannya lebih dari atau sama dengan 2 kali disebut frequent. Himpunan dari frequent k-itemset dilambangkan dengan F_k .

2. Pembentukan Aturan Asosiasi

Setelah diperoleh dan ditemukan semua pola dari frekuensi tinggi, kemudian mencari aturan asosiasi yang telah memenuhi syarat minimum untuk confidence dengan menghitung confidence dengan aturan asosiatif $A \rightarrow B$. Nilai Confidence dari aturan $A \rightarrow B$ diperoleh rumus berikut.

$$\text{Confidence} = P(B|A) = \frac{\text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi mengandung } A} \times 100$$

Dalam penentuan aturan asosiasi yang telah dipilih kemudian diurutkan berdasarkan $\text{Support} \times \text{Confidence}$. Lalu aturan diambil sebanyak n-aturan yang memiliki hasil terbesar.

Pembentukan aturan asosiasi adalah aturan dimulai ketika pola frekuensi tinggi telah ditemukan, kemudian dicari aturan asosiasi dengan acuan telah memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiasif $A \rightarrow B$. Nilai confidence dari aturan $A \rightarrow B$ diperoleh dengan rumus berikut [15]:

$$\text{Confidence} = P(B|A) = \frac{\text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi mengandung } A} \times 100$$

2.2.6.2 Proses Utama Algoritma Apriori

Pada proses ini bertujuan untuk meningkatkan efisiensi dalam mencari k-itemset, dapat digunakan dengan metode tambahan yang dinamakan Apriori Property. Metode ini dapat mengurangi lingkup pencarian sehingga dapat mempersingkat waktu.

Dalam algoritma Apriori terdapat dua proses utama, yaitu [12]:

1. Join (Penggabungan).

Setiap item pada proses ini dikombinasikan menggunakan item lainnya hingga tidak terbentuk lagi kombinasi. Untuk menemukan L_k , suatu set dari kandidat k -itemset dihasilkan dengan cara men-join L_{k-1} dengan dirinya sendiri. Set kandidat hasil join ini nanti akan dinotasikan sebagai C_k . Adapun aturan dari join ini adalah setiap kandidat yang dihasilkan tidak boleh mengandung kandidat yang kembar antara satu dengan yang lainnya.

2. Prune (Pemangkasan).

Pada tahapan ini, hasil dari item yang telah dikombinasikan kemudian dipangkas oleh user dengan menggunakan minimum support yang telah ditentukan. Semua $(k-1)$ -itemset yang tidak frequent tidak mungkin dapat menjadi subset dari frequent k itemset. Oleh karena itu, jika ada $(k-1)$ subset dari kandidat k -itemset yang tidak termasuk dalam L_{k-1} , maka kandidat tidak mungkin frequent juga dan oleh karena itu perlu dihapus dari C_k .

2.2.6.3 Formula Pencarian Nilai Support & Confidence

Analisa pola frekuensi tinggi dengan algoritma apriori adalah tahap dalam pencarian kombinasi item sehingga memenuhi syarat minimum pada nilai support sebuah item yang diperoleh dengan menggunakan rumus berikut [15]:

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A}}{\text{Total transaksi}} \times 100$$

Pencaarian dari nilai support dari 2 item diperoleh dengan menggunakan rumus :

$$\text{Support (A,B)} = P(A \cap B)$$

$$\text{Support (A,B)} = \frac{\text{Transaksi Mengandung A dan B}}{\Sigma \text{Transaksi}} \times 100$$

2.2.6.4 Langkah-langkah Algoritma Apriori

Adapun langkah-langkah atau cara kerja dari Algoritma Apriori secara ringkas sebagai berikut [16]:

1. Tentukan minimum support.
2. Iterasi 1 : hitung item-item dari support (transaksi yang mencakup seluruh item) dengan melakukan scan database pada 1-itemset, setelah 1-itemset diperoleh, dari 1-itemset apakah diatas minimum support, jika telah memenuhi minimum dari support, 1-itemset tersebut nantinya akan menjadi pola dari frequent tinggi.
3. Iterasi 2 : untuk memperoleh 2-itemset, harus dilakukan kombinasi dari k-itemset sebelumnya, kemudian dilakukan scan database lagi untuk mendapatkan hasil perhitungan item-item yang mencakup support. itemset yang telah memenuhi minimum dari support akan dipilih menjadi pola frequent tinggi dari kandidat.
4. Menetapkan nilai k-itemset pada support yang telah memenuhi minimum support dari k-itemset.
5. Melakukan proses untuk iterasi selanjutnya hingga tidak ada lagi k-itemset yang memenuhi minimum support.

2.2.6.5 Contoh Penyelesaian Algoritma Apriori

Untuk mempermudah penjelasan tentang Algoritma Apriori, akan dijelaskan tahapan-tahapannya menggunakan contoh soal sederhana. Berikut ini adalah itemset yang akan dibuat eksperimen:

Tabel 2.2 Transaksi yang dibeli

ID Transaksi	Barang yang Dibeli
T1	{Mango, Onion, Nintendo, Key-chain, Eggs, Yo-yo}
T2	{Doll, Onion, Nintendo, Key-chain, Eggs, Yo-yo}
T3	{Mango, Apple, Key-chain, Eggs}
T4	{Mango, Umbrella, Corn, Key-chain, Yo-yo}
T5	{Corn, Onion, Onion, Key-chain, Ice cream, Eggs}

1. Golden rule (threshold) menggunakan nilai prosentase 60% atau jumlah barang yang dibeli berjumlah paling sedikit 3. Untuk mempermudah nama-nama itemset, dilakukan penyingkatan dengan mengambil huruf depannya saja, sebagai contoh :

M = Mango; O = Onion; Dan sebagainya.

Tabel 2.3 Transaksi yang dibeli (2)

ID Transaksi	Barang yang Dibeli
T1	{M, O, N, K, E, Y}
T2	{D, O, N, K, E, Y}
T3	{M, A, K, E}
T4	{M, U, C, K, Y}
T5	{C, O, O, K, I, E}

2. Langkah ke-1 : Menghitung jumlah transaksi dalam setiap item. Hati-hati, dalam item O (Onion) dibeli dengan jumlah 4 biji, namun memiliki pembelian O dan hanya ada 3 transaksi.

Tabel 2.4 Banyaknya Transaksi per Item

Item	Banyaknya Item
M	3
O	3
N	2
K	5
E	4
Y	3
D	1
A	1
U	1
C	2

I	1
---	---

3. Langkah ke-2 : Mengacu pada golden rule yang telah dijelaskan dan disebutkan di atas, dilakukan penyaringan data pada Tabel 2.3, item yang memiliki transaksi minimal sebanyak 3 yang dipilih, Item yang jumlah transaksinya kurang dari 3, dibuang. Hasilnya dapat dilihat di Tabel 2.5

Tabel 2.5 Item yang Paling Sering Dibeli

Item	Banyaknya Transaksi
M	3
O	3
K	5
E	4
Y	3

4. Langkah ke-3 : Memasangkan item yang dimulai dari item pertama, yaitu MO, MK, ME, MY. Setelah itu dilanjutkan dengan item kedua. Misalnya OK, OE, OY. Perhatikan, OM tidak dibuat karena OM = MO (pasangan yang dibuat dari item pertama).

Tabel 2.6 Pasangan Item

Pasangan Item
MO
MK
ME
MY
OK
OE
OY
KE
KY
EY

5. Langkah ke-4 : Menghitung berapa kali dari pasangan item yang dibeli secara bersamaan. Contohnya pasangan MO dibeli secara bersamaan dalam itemset {M, O, N, K, E, Y}. Pasangan MK dibeli bersamaan sebanyak 3 kali yaitu di dalam {M,O,N,K,E,Y}, {M,A,K,E}, dan {M,U,C, K, Y}. Hasilnya dapat dilihat pada Tabel 2.7.

Tabel 2.7 Banyaknya Transaksi Pasangan Item

Pasangan Item	Banyaknya Transaksi
MO	1
MK	3
ME	2
MY	2
OK	3
OE	3
OY	2
KE	4
KY	3
EY	2

6. Langkah ke-5 : Menggunakan golden rule, menghilangkan semua pasangan item yang jumlah transaksinya kurang dari tiga. Hasilnya adalah pada Tabel 2.8.

Tabel 2.8 Transaksi Pasangan Item yang Terbanyak

Pasangan Item	Banyaknya Transaksi
MK	3
OK	3
OE	3
KE	4
KY	3

7. Langkah ke-6 : Membuat pasangan dari ketiga item dengan aturan dengan menggunakan pasangan item pada Tabel 2.8 yang memiliki huruf depan yang sama yaitu :

- OK dan OE, menjadi OKE
- KE dan KY, menjadi KEY

Setelah itu menghitung jumlah dari beberapa banyaknya transaksi pada pasangan tiga item berdasarkan Tabel 2.4. Hasilnya dapat dilihat pada Tabel 2.9.

Tabel 2.9 Banyak Transaksi 3 Pasang Item

Pasangan Item	Banyaknya Transaksi
OKE	3
KEY	2

Pada langkah ini, semisal ada tiga pasangan item ABC, ABD, ACD, ACE, BCD dan selanjutnya dibuatkan pasangan 4 item, carilah 2 huruf depan yang sama. Contoh :

- ABC dan ABD, menjadi ABCD
- ACD dan ACE, menjadi ACDE

Dilakukan secara menerus. Secara umum, mencari pasangan item yang huruf (item) terakhirnya berbeda.

8. Langkah ke-7 : Golden rule digunakan lagi untuk membuang pasangan dari tiga item yang jumlah transaksinya kurang dari 3. Hasil yang didapat yaitu OKE karena KEY hanya dua kali dibeli secara bersamaan.
9. Kesimpulan : Dari tiga item yang sering dibeli secara bersamaan adalah O, K, dan E.

Untuk menentukan confidence atau tingkat keyakinan dan frequent itemset (yaitu {O, K, E}) dapat digunakan dalam mencari

korelasi atau aturan-aturan asosiasi antar item yang terdapat pada frequent itemset tersebut. Caranya adalah sebagai berikut:

a. Itemset dibuat dalam himpunan bagiannya. Hasilnya akan seperti pada contoh di bawah :

- {O}
- {K}
- {E}
- {O, K}
- {K, E}
- {O, E}

b. Pencarian asosiasi dalam semua himpunan bagian yang telah dibuat, misal : {O} \rightarrow {K, E} artinya : jika O dibeli, bagaimana kemungkinan K dan E akan dibeli pada transaksi yang sama. O dibeli pada 3 transaksi dan di dalam 3 transaksi tersebut, K dan E juga dibeli. Maka keyakinannya adalah :

- {O} \rightarrow {K, E} : keyakinannya yaitu $3/3 \times 100\% = 100\%$.
- {K} \rightarrow {O, E} : keyakinannya yaitu $3/5 \times 100\% = 60\%$
- {E} \rightarrow {O, K} : keyakinannya yaitu $3/4 \times 100\% = 75\%$
- {K, E} \rightarrow {O} : keyakinannya yaitu $3/3 \times 100\% = 100\%$
- {O, E} \rightarrow {K} : keyakinannya yaitu $3/3 \times 100\% = 100\%$
- {O, K} \rightarrow {E} : keyakinannya yaitu $3/4 \times 100\% = 100\%$

2.2.6.6 Langkah-langkah dari proses Algoritma Apriori

Langkah-langkah algoritma Apriori untuk mendapatkan rules yang diinginkan oleh user, antara lain:

1. Scan database digunakan untuk memperoleh kandidat 1-itemset, yaitu C1 (Himpunan item yang terdiri dari 1 item) dan menghitung nilai supportnya. Bandingkan nilai support dengan minimum support yang sudah ditentukan, jika nilainya lebih besar atau sama dengan

- minimum support, maka itemset tersebut termasuk dalam large itemset yaitu L1 (Large itemset dengan 1 itemset)
2. Itemset yang tidak termasuk dalam large itemset tidak disertakan dalam iterasi selanjutnya (dilakukan pruning).
 3. Himpunan L1 yang dihasilkan dari iterasi pertama digunakan dalam iterasi selanjutnya. Kemudian L1 digunakan dalam proses join terhadap L1 itu sendiri dalam membentuk kandidat 2 itemset (C2). Proses selanjutnya membandingkan lagi support dari item-item C2 dengan minimum support, bila tidak ada pengurangan dari minimum support, maka itemset tersebut akan masuk dalam large itemset L2. Pada iterasi selanjutnya, hasil large itemset pada iterasi sebelumnya (Lk-1) akan dilakukan proses join terhadap dirinya sendiri untuk membentuk kandidat baru (Ck), dan large itemset baru (Lk). Setelahnya dilakukan proses pruning pada itemset yang tidak termasuk dalam Lk-.
 4. Dari seluruh large itemset yang memenuhi minimum support (frequent itemset) dibentuk association rule dan nilai confidencenya. Aturan-aturan yang nilai confidencenya lebih kecil dari minimum confidence, tidak termasuk dalam association rule yang dipakai.

2.2.7 Korelasi Lift

Proses mining Apriori ditandai dengan terbentuknya kekuatan hubungan kombinasi itemset dengan alat ukur asosiasi final. Namun untuk mengukur valid atau tidaknya asosiasi final tersebut maka dapat menggunakan lift ratio [17]. Lift ratio adalah alat ukur penting dalam aturan asosiasi. Fungsinya adalah mengukur ketepatan dan kecermatan suatu alat ukur (support dan confidence) agar dapat dipercaya sepenuhnya. Dalam penelitian ini lift ratio memastikan bahwa apakah penggunaan media A digunakan secara bersamaan dengan media B. Rumus perhitungan lift ratio dapat dirujuk pada peneliti [18]. Pada akhirnya sebuah kombinasi itemset dinyatakan valid dan kuat jika nilai lift ratio > 1 .

Nilai korelasi dapat diketahui dengan menggunakan rumus persamaan sebagai berikut:

$$Lift(A,B) = \frac{P(A \cup B)}{P(A)P(B)} \times 100$$

Lift (A,B) = Korelasi antara A dan B

$P(A \cup B)$ = Jumlah kemunculan antara A dan B dibagi dengan total transaksi

$P(A) P(B)$ = Jumlah kemunculan A dikali jumlah kemunculan B pada total transaksi.

Apabila dari perhitungan tersebut menghasilkan nilai dibawah 1 maka terdapat korelasi negatif. Untuk perhitungan yang menghasilkan nilai diatas 1 maka terdapat korelasi positif. Namun apabila menghasilkan nilai sama dengan 1 maka tidak ada korelasi antara X dan Y.

2.2.8 RapidMiner

RapidMiner merupakan software yang disediakan untuk pengolahan *data mining*. Pekerjaan RapidMiner dalam text mining adalah berhubungan dengan analisis teks dan mengekstrak pola-pola dari data set yang besar kemudian mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. Mendapatkan informasi bermutu tinggi dari teks yang diolah merupakan tujuan dari analisis.

Prosedur data mining dan machine learning disediakan oleh RapidMiner, di dalamnya termasuk ETL yaitu extraction, transformation, loading, data preprocessing, visualisasi, modelling dan evaluasi. Proses dalam data mining tersusun dari operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI. Penyajiannya dituliskan dalam bahasa pemrograman Java.

RapidMiner merupakan perangkat lunak *open source* atau bersifat terbuka. RapidMiner menjadi solusi dalam melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan

berbagai teknik prediksi dan deskriptif untuk memberikan wawasan kepada setia penggunanya sehingga dapat memberikan rekomendasi keputusan yang paling baik.

RapidMiner sebelumnya bernama YALE yaitu Yet Another Learning Environment, yang versi awalnya dikembangkan mulai tahun 2001 oleh Simon Fischer, Ingo Mierswa dan Ralf Klinkenberg di Artificial Intelligence Unit dari University of Dortmund. Pendistribusian dari RapidMiner dilakukan dan dilisensi oleh AGPL (GNU Affero General Public License) versi 3. Sampai saat ini telah ribuan aplikasi yang dikembangkan menggunakan RapidMiner di lebih dari 40 negara. RapidMiner tidak diragukan lagi sebagai software open source dalam pengolahan data mining karena sudah terkenal di dunia. RapidMiner menjadi peringkat pertama sebagai Software data mining pada polling oleh KDnuggets, sebuah portal data-mining pada 2010-2011.

2.2.8.1 Fitur dan Fungsi dari RapidMiner

Beberapa Fitur yang disediakan dari RapidMiner, antara lain:

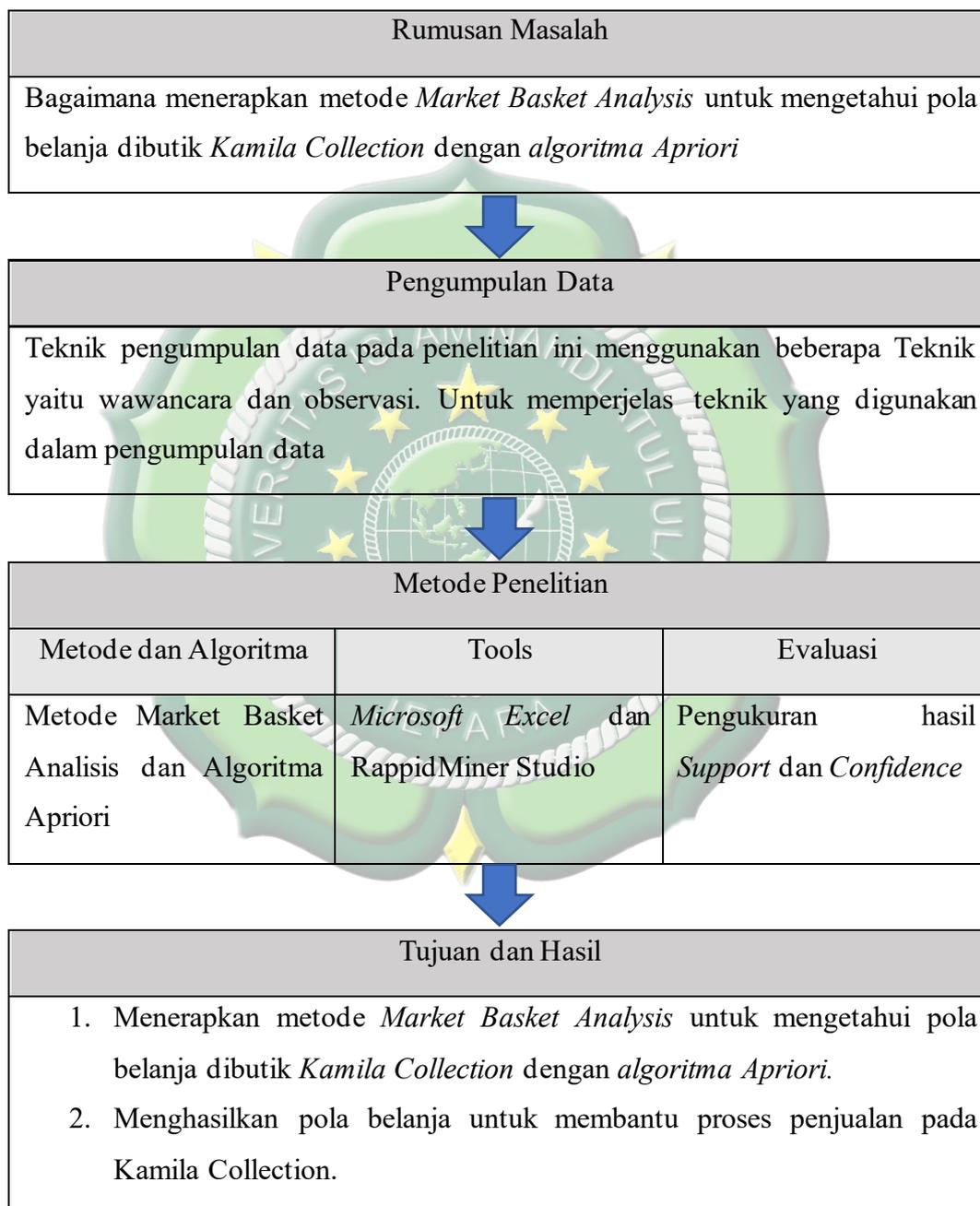
1. Banyak algoritma dari data mining, seperti self-organization map dan decision tree.
2. Penampilan dari grafis yang canggih, seperti tumpang tindih 3D Scatter plots, diagram histogram dan tree chart.
3. Variasi plugin yang banyak, seperti text plugin dalam melakukan analisis dari sebuah teks.
4. Penyediaan prosedur machine learning dan data mining termasuk: ETL atau transformation, extraction, loading, data preprocessing, visualisasi, modelling dan evaluasi
5. Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI

Fungsi-fungsi yang dapat digunakan melalui Explorer adalah sebagai berikut :

1. Preprocessing, merupakan panel yang digunakan untuk memilih data yang akan diproses. Data dapat berbentuk berkas ARFF, database, maupun data buatan yang dibangkitkan dengan pola tertentu. Pemilihan variabel atau fitur yang akan digunakan juga dapat dilakukan pada panel ini.
2. Classifier, merupakan panel untuk menggunakan algoritma klasifikasi atau regresi terhadap data yang telah dimasukkan di panel Preprocessing. Terdapat bermacam-macam algoritma klasifikasi dan regresi yang telah disediakan dan siap digunakan. Terdapat pula sejumlah metode evaluasi yang dapat digunakan untuk mengevaluasi hasil klasifikasi.
3. Cluster, merupakan panel yang memberikan akses terhadap berbagai metode clustering yang disediakan oleh Weka. Panel ini memiliki beberapa opsi pembagian data menjadi data latih dan data uji untuk keperluan pengujian hasil clustering.
4. Associate, menyediakan algoritma untuk association rule yang berguna untuk menggali relasi antar item pada data.
5. Select Attribute, menyediakan berbagai algoritma untuk memilih atribut atau variabel yang paling relevan untuk suatu permasalahan

2.3 Kerangka Pemikiran

Penyusunan skripsi ini memiliki kerangka pemikiran yang dibuat pedoman dalam menyusun skripsi, berikut ini adalah kerangka pemikirannya



Gambar 2.2 Kerangka Pemikiran