

BAB II LANDASAN TEORI

2.1 Tinjauan Pustaka

Beberapa penelitian terdahulu yang dijadikan acuan dalam penelitian ini adalah:

2.1.1 Tinjauan Terhadap Literatur 1

(Yakub et al., 2019) meneliti tentang Penerapan Data Mining Pengaturan Pola Tata Letak Barang Pada Toko Berkah Swalayan Untuk Strategi Penjualan Menggunakan Algoritma Apriori. Pengaturan tata letak barang dagangan di toko sangat mempengaruhi minat konsumen dalam berbelanja. Terbukti bahwa keputusan membeli barang diluar perencanaan diambil konsumen saat melihat barang yang disusun dengan baik didalam toko, sehingga dapat disimpulkan bahwa pengaturan tata letak barang sangat mempengaruhi volume penjualan. Akan tetapi dalam pengaturan tata letak barang pada Toko Berkah Swalayan masih mengalami kendala yaitu kurangnya pengetahuan pihak Toko Berkah Swalayan dalam pengaturan tata letak barang. Oleh karena itu dibutuhkan suatu sistem yang dapat membantu Toko Berkah Swalayan dalam menentukan pengaturan tata letak barang agar dapat meningkatkan strategi penjualan yaitu dengan membangun sebuah sistem Penerapan Data Mining Pengaturan Pola Tata Letak Barang Pada Toko Berkah Swalayan Untuk Strategi Penjualan Menggunakan Algoritma Apriori. dengan dibangunnya sistem ini diharapkan dapat membantu pihak Toko Berkah Swalayan dalam mengatasi permasalahan yang terjadi.

2.1.2 Tinjauan Terhadap Literatur 2

(Irfa'aturrochmah, 2018) meneliti tentang Penentuan Tata Letak Barang Dagangan Berdasarkan Data Transaksi Penjualan Harian Menggunakan Algoritma Apriori. Banyak transaksi yang terjadi setiap hari, mulai dari transaksi pembelian dan transaksi penjualan. Dengan banyaknya kegiatan transaksi yang terjadi, akan di sayangkan jika data tersebut tidak dimanfaatkan menjadi informasi penting untuk menunjang proses penjualan, salah satunya untuk menentukan tata letak barang dagangan seramart. Penelitian ini menerapkan pendekatan keranjang belanja atau Market Basket Analysis (MBA), algoritma Apriori untuk mengetahui pola barang yang dibeli konsumen secara bersamaan. Hasil analisis kemudian digunakan untuk menentukan tata letak barang berdasarkan perilaku belanja konsumen menggunakan alat bantu WEKA. Kemudian hasil dari analisis WEKA itulah yang akan dijadikan sebuah keputusan dalam penentuan tata letak barang di minimarket Seramart.

2.1.3 Tinjauan Terhadap Literatur 3

(Anggrawan, Mayadi and Satria, 2021) meneliti tentang Menentukan Akurasi Tata Letak Barang dengan Menggunakan Algoritma Apriori dan Algoritma FP-Growth. Pada penelitian ini peneliti melakukan analisis bertujuan untuk mencari kemiripan barang berdasarkan item pembelian dijadikan sebagai acuan dalam tata letak barang dan mengidentifikasi kesamaan barang yang dibeli ketika menambah stok barang. Untuk mengidentifikasi tujuan pada penelitian ini peneliti melakukan proses perhitungan menggunakan dua metode yaitu apriori dan FP-Growth dan melakukan pengujian dengan 2 pengujian yaitu pengujian hasil dan rasio adapun

hasil pengujian didapatkan FP-Growth menghasilkan rule yang lebih baik dibandingkan dengan algoritma apriori dengan total rule sebanyak 6, sedangkan algoritma apriori menghasilkan 4 rule, dan untuk pengujian dengan evaluasi hasil rule dari masing masing algoritma, algoritma FP-Growth memiliki hasil yang terbaik dengan lift ratio 1.27908.

2.1.4 Tinjauan Terhadap Literatur 4

(Azwanti, Elisa and Nopriadi, 2019) meneliti tentang Strategi Penentuan Tata Letak Barang dengan Teknik Asosiasi. Pada penelitian ini hanya membahas mengenai penjualan perlengkapan sekolah seperti tas, sepatu, kaos kaki dan lain sebagainya. Namun, diperlukan sebuah strategi penjualan guna mempertahankan bisnis dan pemasaran. Strategi yang baik pasti menghasilkan keuntungan dan produk ataupun barang akan terjual dengan cepat. Strategi tersebut adalah pengaturan tata letak barang. Terbukti bahwa keputusan membeli barang yang diluar perencanaan dipengaruhi oleh tata letak. Data mining menjawab masalah ini dengan menganalisa data yang besar tersebut kemudian membuat sebuah aturan, pola, ataupun model tertentu. Teknik asosiasi yang dapat digunakan adalah algoritma apriori yang dirancang untuk menemukan kombinasi item. Algoritma apriori berpedoman pada nilai min support dan min confidence. penelitian ini akan menganalisa pola pembelian peralatan sekolah pada objek penelitian Raffa Photocopy dengan Variabel yang diolah yakni Tas, Topi, Ikat Pinggang, Kacu, Topi Boni, Botol Minum, Kaos Kaki, Baret dan Dasi. Minimum support yang ditentukan yaitu 30% dan minimum confidence 75%. Hasilnya adalah Kaos Kaki – Topi

dengan minimum support 30% dan minimum confidence 80%, kemudian Dasi – Topi dengan minimum support 40% dan minimum confidence 100%. Dengan begitu, untuk tata letak yang harus dilakukan adalah topi harus berdekatan dengan kaos kaki dan dasi.

2.1.5 Tinjauan Terhadap Literatur 5

(Saputra, Wasiyanti and Nugraha, 2020) meneliti tentang Penerapan Algoritma Apriori Untuk Analisa Pola Penempatan Barang Berdasarkan Data Transaksi Penjualan. Tujuan penelitian ini menerapkan teknik association data mining yaitu algoritma apriori yang dapat menganalisa transaksi penjualan dari data historis. Pengetahuan baru diperoleh berdasarkan hasil perhitungan algoritma yang diimplementasikan menggunakan aplikasi RapidMiner. Informasi yang didapat dari pengolahan yaitu mengetahui kecenderungan pembeli dalam melakukan pembelian barang yang digunakan untuk memberikan rekomendasi pengadaan barang dan kebijakan tata letak barang berdasarkan intensitas penjualan dalam rangka menunjang aktifitas transaksi penjualan barang. Dari hasil analisa terhadap 31 transaksi penjualan yang menggunakan nilai minimum support sebesar 6% dan minimum confidence sebesar 60% didapatkan rekomendasi tata letak barang terhadap klasifikasi penempatan barang yaitu kantong plastik → tambang, rodeo → sterofom, plastik 9x25 → karet dan plastik 11x30 → sedotan.

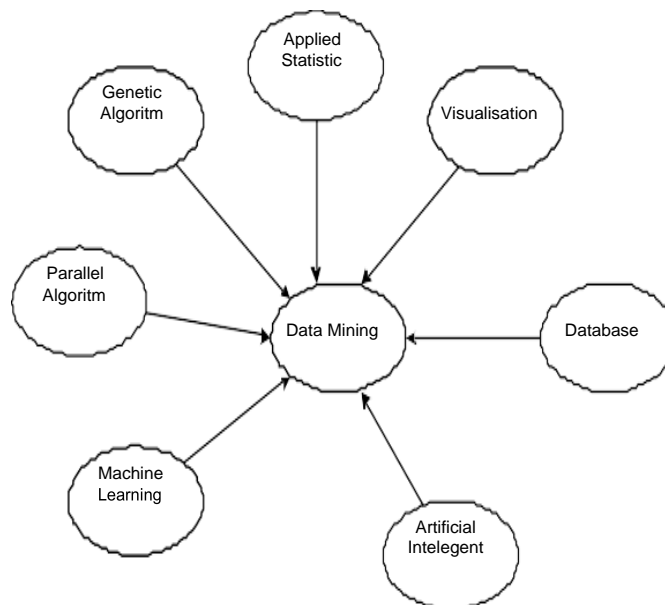
2.2 Data Mining

Menurut (Saputra, Wasiyanti and Nugraha, 2020) perkembangan *data mining* yang pesat tidak dapat lepas dari perkembangan teknologi informasi yang memungkinkan

data dalam jumlah yang besar terakumulasi. Tetapi pertumbuhan yang pesat dari akumulasi data telah menciptakan suatu kondisi yang disebut dengan “*rich of data but poor of information*” karena data yang terkumpul itu tidak dapat digunakan dalam suatu aplikasi yang berguna. Bahkan tidak jarang kumpulan data tersebut dibiarkan begitu saja sehingga tercipta “*data tombs*” (kuburan data).

Dalam jurnal ilmiah, *data mining* juga dikenal dengan nama KDD (*Knowledge Discovery in Database*). Namun pada tahun 1995, telah diadakan *International KDD Conference* di Montreal yang berhasil mendefinisikan bahwa KDD merupakan suatu proses dalam mengenali informasi atau suatu kebenaran baru dan benar-benar berguna serta mengenali pola yang dapat dimengerti dari data. Tujuan utama dari proses KDD adalah memprediksikan nilai-nilai yang berguna dari variabel-variabel yang ada atau menemukan pola-pola dari sebuah gugusan data yang dapat diinterpretasikan oleh manusia. Sesuai dengan tujuan tersebut, maka proses dalam mengenali informasi baru dan penemuan pola tersebut perlu diaplikasikan dengan *data mining*. Sehingga sebenarnya *data mining* merupakan suatu bagian yang tidak dapat dilepaskan dari proses KDD.

Perlu diketahui bahwa *data mining* merupakan salah satu bidang yang cukup banyak didukung oleh cabang ilmu lain di dalam teknologi informasi yaitu statistik, teknologi basis data, *machine learning*, sistem pakar, algoritma paralel, algoritma genetika, pengenalan pola, visualisasi data, dan lain-lain.



Gambar 2. 1 *Data mining* merupakan bidang multidisipliner

Ada beberapa faktor yang menjadi alasan utama mengapa menggunakan *data mining*:

1. Banyaknya data yang terkumpul sehingga memerlukan waktu yang sangat lama dan tenaga ahli yang cukup banyak untuk menganalisisnya.
2. Komputer menjadi salah satu pilihan utama karena kemampuannya dalam kecepatan, ketepatan, tidak pernah lelah dan mudah dioperasikan.
3. Tekanan dari kompetisi bisnis yang terus menguat sehingga menjadikan informasi menjadi sangat penting dan harus segera dimiliki.
4. Mampu menemukan suatu pola yang tidak terpikirkan sama sekali.

Menurut (Widiartha, 2019) *data mining* merupakan salah satu aktifitas dibidang perangkat lunak yang dapat memberikan ROI (*Return of Investment*) yang tinggi.

Hal yang perlu diperhatikan adalah bahwa *data mining* berbeda dengan *query tools*. *Query* dan *data mining* merupakan dua hal yang saling melengkapi. Keberadaan

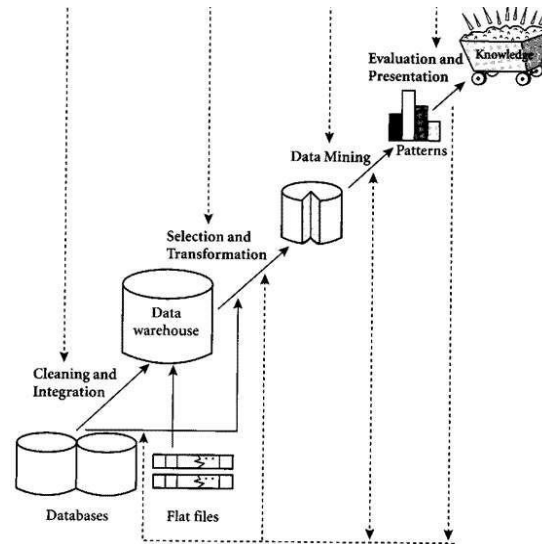
data mining bukan untuk menggantikan *query* tetapi menambahkan beberapa tambahan yang berarti. Jika menggunakan *query* sederhana maka informasi yang dapat diakses sekitar 80% dari data yang ada dalam basis data sedangkan 20% lagi akan menjadi informasi tersembunyi yang memerlukan teknik-teknik khusus dalam mengaksesnya.

2.2.1 Tahap-Tahap Data Mining

Karena *data mining* adalah suatu rangkaian proses maka dibagi menjadi beberapa tahap antara lain :

- a. Pembersihan data: untuk membuang data yang tidak konsisten dan noise.
- b. Integrasi data: untuk menggabungkan data dari beberapa sumber.
- c. Transformasi data : untuk mengubah data menjadi bentuk yang sesuai untuk *data mining*.
- d. Aplikasi teknik *data mining*.
- e. Evaluasi pola yang ditemukan : untuk menemukan informasi yang menarik ataupun bernilai.
- f. Presentasi pengetahuan dengan teknik visualisasi.

Tahap-tahap diatas dapat digambarkan sebagai berikut :



Gambar 2. 2 Tahap-tahap dalam *data mining*

2.2.2 Teknik Data Mining

Berdasarkan proses, yaitu :

1. *Supervised Learning*

Dalam *supervised learning* disyaratkan agar data analis telah mengidentifikasi atribut tujuan. Sebagai contoh, bila ada suatu pertanyaan tentang siapakah pelanggan yang baru-baru ini membeli mobil baru, untuk itu dapat dibuat target atribut 1 untuk “YA” dan 0 untuk “TIDAK”. Teknik-teknik yang termasuk dalam bagian ini antar lain *Clasification*, *Regressio*.

a. *Classification*

Classification adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika maka”, *decision tree* ataupun formula

matematis (Saputra, Wasiyanti and Nugraha, 2020) *Decision tree* merupakan salah satu metode *classification* yang paling populer karena mudah untuk diinterpretasikan oleh manusia. Setiap percabangan menyatakan kondisi yang harus dipenuhi dan tiap ujung pohon menyatakan kelas data atau atribut data. Dari decision tree tersebut, diketahui bahwa salah satu kelompok yang potensial membeli komputer adalah orang yang berusia dibawah atau sama dengan 30 dan juga merupakan seorang pelajar.

Algoritma *Decision tree* yang sering dipakai adalah ID3 dan C4.5, namun akhir-akhir ini sedang dikembangkan suatu algoritma yang dikenal dengan *RainForest*. Metode-metode classification yang lain adalah *Bayesian*, *Neural Network*, *Genetic Algorithm*, *Fuzzy*, *Case-based Reasoning* dan *K-Nearest Neighbor*.

b. *Regression*

Regression merupakan salah satu metode untuk menentukan hubungan variabel dengan variabel lainnya. metode *regression* yaitu regresi linier berganda, analisis regresi kolerasi daln lainnya

2. *Unsupervised Learning*

Berbeda dengan *supervised learning*, dalam *unsupervised learning* data analisis tidak perlu mengidentifikasi atribut target. Teknik-teknik *data mining* yang termasuk ke dalam bagian ini adalah *Clustering*, *Association Rule*, dan lain-lain. Berikut ini adalah gambaran tentang teknik *data mining* yang paling populer dari teknik-teknik *data mining* yang ada:

a. *Association Rule*

Merupakan teknik *data mining* untuk menemukan aturan asosiasi antara kombinasi *item*. Teknik ini pada umumnya diterapkan untuk basis data transaksi dimana transaksi terdiri dari sejumlah *item*.

Basis data transaksi tersebut dapat dipakai untuk menyelesaikan masalah pemilik pasar swalayan atau toko antara lain dalam mengatur tata letak barang, penyiapan stok barang, dan lain-lain. Dengan menemukan semua aturan asosiasi dan korelasi di antara *item* data dimana kehadiran salah satu dari sejumlah *item* data menunjukkan secara tidak langsung adanya kehadiran *item* data lainnya, maka masalah tersebut bisa diselesaikan. Contoh dari aturan asosiasi adalah bisa diketahui berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan informasi tersebut, seorang pemilik pasar swalayan atau toko dapat mengambil keputusan- keputusan strategis tentang pasar swalayan atau tokonya dalam mengatur penempatan barang atau merancang kampanye pemasaran dengan menggunakan kupon diskon untuk kombinasi barang. Adapun algoritma pada teknik ini antara lain *Apriori*, *FP-Growth*, *Closure/closed* dan lain-lain.

b. *Clustering*

Berbeda dengan *classification* dimana kelas data telah ditentukan sebelumnya, *clustering* melakukan pengelompokan data tanpa berdasarkan kelas tertentu. Bahkan *clustering* dapat dipakai untuk memberikan label pada kelas data yang belum diketahui itu. Prinsip *clustering* adalah memaksimalkan

kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas / *cluster*.

2.3 Market basket Analysis

Market basket analysis adalah analisis yang sering dipakai untuk menganalisa isi keranjang belanja konsumen dalam suatu pasar swalayan. Contoh penerapan dari aturan asosiatif adalah analisa pembelian produk pada sebuah toko alat tulis, pada analisa itu misalkan dapat diketahui berapa besar kemungkinan seorang pelanggan membeli pensil bersamaan dengan membeli penghapus. Penerapan aturan asosiasi dalam kasus tersebut dapat membantu pemilik toko untuk mengatur penempatan barang, mengatur persediaan atau membuat promosi pemasaran dengan menerapkan diskon untuk kombinasi barang tertentu (Choiriah and Akmal, 2019).

Fungsi *association rules* seringkali disebut dengan analisis keranjang belanja (Market Basket Analysis) yang digunakan untuk menemukan relasi atau korelasi di antara himpunan item-item. Analisis keranjang belanja adalah analisis dari kebiasaan membeli customer dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan customer dalam keranjang belanjanya.

Data keranjang belanja dalam bentuknya yang paling mentah adalah data transaksi pembelian oleh pelanggan, yang mengindikasikan hanya barang-barang yang dibeli bersamaan. Data ini menantang karena beberapa hal :

1. Jumlah datanya besar (kegiatan transaksi penjualan dengan data yang banyak)
2. Pembagiannya luas (setiap keranjang belanja hanya mengandung sebagian kecil dari benda-beda yang dijual)

3. Heterogenitas (orang-orang dengan selera yang berbeda cenderung sekelompok barang yang spesifik)

2.3.1 Tujuan Market Basket Analisis

Tujuan dari Market Basket Analisis adalah untuk menganalisa atas perilaku konsumen secara spesifik dari suatu golongan atau suatu kelompok tertentu. Selain itu tujuan dari market basket adalah untuk mengetahui produk-produk mana yang mungkin akan dibeli secara bersamaan. Analisis data transaksi dapat menghasilkan pola pembelian produk yang sering terjadi (Jumisah et al., 2017)

2.4 Association rules

Association rule adalah aturan asosiasi mengungkap item yang sering dikaitkan bersama”. Pada awalnya algoritma aturan asosiasi itu dikembangkan dalam konteks pasar analisis keranjang untuk mempelajari perilaku pembelian pelanggan yang dapat digunakan untuk pemasaran (Adiwihardja, Cahyati and Hilma, 2018).

Association rule berguna untuk menemukan hubungan penting antar item dalam setiap transaksi, hubungan tersebut dapat menandakan kuat tidaknya suatu aturan dalam asosiasi, Tujuan association rule adalah untuk menemukan keteraturan dalam data. Association rule dapat digunakan untuk mengidentifikasi item-item produk yang mungkin dibeli secara bersamaan dengan produk lain, atau dilihat secara bersamaan saat mencari informasi mengenai produk tertentu. Dalam pencarian *association rule*, diperlukan suatu variabel ukuran kepercayaan (*interestingness measure*) yang dapat ditentukan oleh user, untuk mengatur batasan sejauh mana dan sebanyak apa hasil output yang diinginkan oleh user.

Association rule adalah salah satu teknik *data mining* yang sudah mulai dipopulerkan oleh Rakesh Agrawal dan Ramakrishnan Srikant untuk analisis keranjang pasar pada tahun 1993 (Muchlis et al., 2021) Aturan asosiasi sering dipakai dalam penggalian data transaksi. Proses pencarian aturan asosiasi bertujuan untuk menemukan pola yang sering muncul, asosiasi antara suatu kombinasi *item* yaitu dengan membuat korelasi antara *item* data yang dikelompokkan ke dalam transaksi kemudian mengambil kesimpulan berdasarkan hubungan yang terbentuk dari beberapa *item* data. Hal-hal tersebut dapat merepresentasikan informasi penting yang ingin diketahui pada data yang ada.

Proses pencarian *frequent itemset* merupakan pra-syarat dan membutuhkan waktu sangat banyak, sehingga banyak algoritma dikembangkan untuk lebih mengefisienkan proses ini. Ada beberapa faktor yang mempengaruhi efisiensi pencarian *frequent itemset* diantaranya adalah bila basis data besar (jumlah transaksi yang banyak), *item* yang sangat banyak dan nilai *support* yang rendah.

Penting tidaknya suatu aturan asosiasi dapat diketahui dengan dua parameter, *support* yaitu prosentase kombinasi *item* dalam database dan *confidence* yaitu kuatnya hubungan antar *item* dalam aturan asosiasi (Saputra, Wasianti and Nugraha, 2020) Pencarian *association rule* bertujuan untuk menemukan semua aturan yang terdapat pada database dengan *minsup* (*minimum support*) dan *minconf* (*minimum confidence*) yang melebihi batas tertentu.

2.4.1 Support

Support dari suatu *association rule* adalah presentasi kombinasi item tersebut dalam database, dimana jika mempunyai item A dan item B maka support adalah proporsi dari transaksi dalam database yang mengandung A dan B. Support (dukungan) merupakan suatu ukuran yang menunjukkan seberapa besar dominasi suatu item atau itemset dari keseluruhan transaksi (Kusumo, Bijaksana, and Darmantoro, 2016).

2.4.2 Confidence

Confidence dari *association rule* adalah ukuran ketepatan suatu rule, yaitu presentasi transaksi dalam *database* yang mengandung A dan mengandung B. Dengan adanya *confidence* kita dapat mengukur kuatnya hubungan antar item dalam *association rule*. (Kusumo, Bijaksana, and Darmantoro, 2016).

$$S = \frac{\sum (Ta+Tc)}{\sum (T)} \dots\dots\dots (2.1)$$

Keterangan :

S = *Support*

$\sum (Ta+Tc)$ = Jumlah transaksi yang mengandung *antecedent* dan *consequent*.

$\sum (T)$ = Jumlah Transaksi

$$C = \frac{\sum (Ta+Tc)}{\sum (Ta)} \dots\dots\dots (2.2)$$

Keterangan :

C = *Confidence*

$\sum (Ta+Tc)$ = jumlah transaksi yang mengandung *antecedent* dan *consequent*.

$\sum (Ta)$ = jumlah transaksi yang mengandung *antecedent*

2.4.3 Analisa Pola Frekuensi Tinggi

Pada tahap ini, melakukan pencarian kombinasi item yang memenuhi syarat *minimum* dari nilai *support* dalam *database*. Nilai *support* item diperoleh dengan Persamaan 1.

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}} \dots\dots\dots(2.3)$$

Sedangkan Nilai *Support 2 Item* diperoleh dari persamaan 2.

$$\text{Support (A,B)} = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}} \dots\dots\dots(2.4)$$

2.4.4 Pembentukan Aturan Asosiatif

Pada tahap ini, setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat *minimum* untuk nilai *confidence* dengan menghitung nilai *confidence* dalam aturan asosiatif. Nilai *confidence* dari aturan diperoleh dari Persamaan 3.

$$\text{Confidence} = P(A|B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi Mengandung B}} \dots\dots\dots(2.5)$$

2.5 Apriori

Algoritma apriori digunakan untuk mencari *frequent itemset* yang memenuhi *minsup* kemudian mendapatkan *rule* yang memenuhi *minconf* dari *frequent itemset* tadi. Algoritma ini mengontrol berkembangnya kandidat *itemset* dari hasil *frequent itemset* dengan *support-based pruning* untuk menghilangkan *itemset* yang tidak

menarik dengan menetapkan *minsup*. Prinsip dari apriori ini adalah bila *itemset* digolongkan sebagai *frequent itemset*, yang memiliki *support* lebih dari yang ditetapkan sebelumnya, maka semua subsetnya juga termasuk golongan *frequent itemset* dan sebaliknya. Algoritma Apriori adalah algoritma analisis keranjang pasar yang digunakan untuk menghasilkan aturan asosiasi, dengan pola if-then. Pada algoritma Apriori menentukan kandidat yang mungkin muncul dengan cara memperhatikan minimum support dan minimum *confidence* (Jumisah et al., 2017).

Dalam metode ini terdapat salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien yaitu analisis pola frekuensi tinggi (*frequent pattern mining*). Suatu asosiasi dikatakan penting dengan memperhatikan dua tolak ukur yaitu: *support* dan *confidence*. *Support* yaitu persentase kombinasi item tersebut dalam database (nilai penunjang). *Confidence* yaitu kuatnya hubungan antar item dalam aturan asosiasi (nilai kepastian). Proses awal dalam algoritma Apriori adalah analisis pola frekuensi tinggi merupakan cara mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam basis data. *Frequent itemset* menunjukkan *itemset* yang memiliki frekuensi kemunculan lebih dari nilai minimum yang ditentukan. Tahap berikutnya adalah pembentukan aturan asosiasi, yaitu setelah semua pola frekuensi tinggi ditemukan, kemudian dicari aturan asosiasi yang memenuhi syarat *minimum confidence* dengan menghitung *confidence* $A \rightarrow B$. Untuk menentukan aturan asosiasi yang akan dipilih maka harus diurutkan berdasarkan *support x confidence*. Aturannya diambil sebanyak n aturan yang memiliki hasil terbesar. Semakin kecil minimum *support* dan *confidence* yang ditentukan, semakin

banyak pula *rules* yang dapat dihasilkan oleh aplikasi, dengan konsekuensi waktu proses akan lebih lama dibandingkan minimum *support* yang lebih besar.

Untuk membentuk kandidat itemset ada dua proses utama yang dilakukan algoritma apriori yaitu :

1. *Join Step* (Penggabungan) Pada proses ini setiap item dikombinasikan dengan item lainnya sampai tidak terbentuk kombinasi lagi.
2. *Prune Step* (Pemangkasan) Pada proses ini, hasil dari item yang dikombinasikan tadi kemudian dipangkas dengan menggunakan minimum support yang telah ditentukan oleh user.

2.6 RapidMiner Studio

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. Berbagai teknik deskriptif dan prediksi digunakan RapidMiner untuk memberikan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Terdapat kurang lebih 500 operator *data mining* yang dimiliki RapidMiner termasuk operator untuk *input*, *output*, *datapreprocessing* dan *visualisasi*. RapidMiner merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri (Afifi, Nastiti and Aini, 2020). RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi RapidMiner memiliki beberapa sifat sebagai berikut:

1. Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan di berbagai sistem operasi.
2. Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
3. Memiliki GUI, command line mode, dan Java API yang dapat dipanggil dari program lain.

Beberapa Fitur dari RapidMiner, antara lain:

1. Banyaknya algoritma *data mining*, seperti *decision tree* dan *self-organization map*.
2. Bentuk grafis yang canggih, seperti tumpukan diagram histogram, *tree chart* dan *3D Scatter plots*.
3. Banyaknya variasi *plugin*, seperti *text plugin* untuk melakukan analisis teks.
4. Menyediakan prosedur *data mining* dan *machine learning* termasuk: ETL (*extraction, transformation, loading*), *data preprocessing*, *visualisasi*, *modelling* dan evaluasi.
5. Proses *data mining* tersusun atas operator-operator yang *nestable*, dideskripsikan dengan XML, dan dibuat dengan GUI
6. Mengintegrasikan proyek data mining Weka dan statistika R.