

BAB II

LANDASAN TEORI

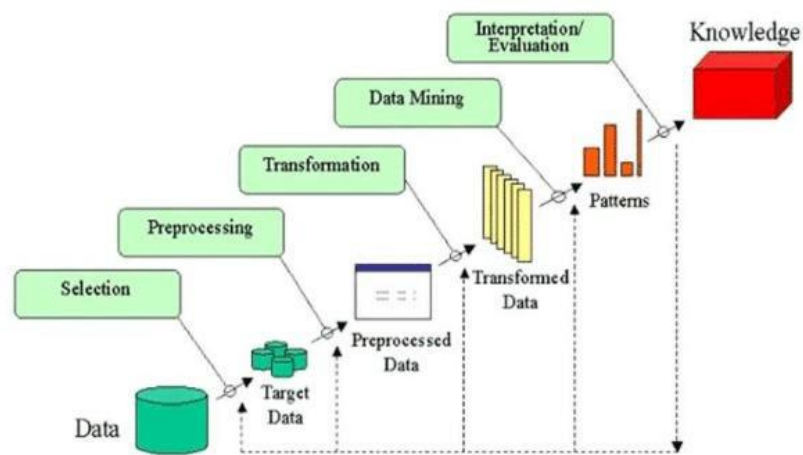
2.1 *Knowledge Discovery In Database*

Knowledge Discovery In Database (KDD) merupakan metode yang digunakan untuk mencari pengetahuan atau informasi yang belum diketahui dari sebuah database [9]. *Knowledge Discovery In Database (KDD)* merupakan keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. Teknologi dan Informasi adalah suatu hal yang lumrah bagi masyarakat. Pesatnya perkembangan teknologi dan informasi tentunya memberikan pengaruh besar bagi masyarakat dan juga pastinya bagi pengusaha atau pebisnis dalam kalangsungan bisnisnya, terutama yang berhubungan dengan data. Tidak sedikit segala sesuatu yang berkaitan dengan bisnis, teknologi dan informasi tentunya selalu menggunakan data sebagai sumber informasinya.

Data mining adalah kegiatan untuk mengidentifikasi beragam jenis data. Dari database yang tersimpan, data-data akan diidentifikasi baik dicari kemungkinan adanya pola ataupun lainnya yang dianggap berpotensi untuk menghasilkan ssesuatu yang bias dipakai oleh organisasi atau perusahaan yang memiliki data tersebut. *Knowledge Discovery In Database (KDD)* adalah salah satu metode yang bisa digunakan dalam melakukan data mining. *Knowledge Discovery In Database (KDD)*

merupakan nama lain dari *Data Mining* walaupun sesungguhnya kedua istilah tersebut memiliki konsep yang tidak sama, namun berkaitan satu sama lain, dan salah satu dari tahapan proses keseluruhan *Knowledge Discovery In Database (KDD)* merupakan *Data Mining* yang menjadi inti dari proses KDD.

Gambar 2.1 Ilustrasi mengenai proses KDD



Gambar 2.1 Ilustrasi Proses *Knowledge Discovery In Database*

KDD adalah sebuah proses untuk mencari dan mengidentifikasi *pattern* dalam sebuah *database*, pada sebuah *Knowledge Discovery In Database* atau KDD memiliki beberapa tahapan di antaranya:

1. Seleksi Data

Proses Seleksi Data dilakukan dengan memilih data yang relevan dengan tugas menganalisis dari database, menciptakan himpunan data target, atau memfokuskan pada contoh data dimana *discovery* akan dilakukan

dan hasil dari seleksi disimpan dalam suatu berkas terpisah dari *database* operasional.

2. Pemrosesan dan Pembersihan Data

Pada tahapan ini dilakukan menghilangkan *noise* dan data yang inkonsisten. Sebelum proses *Data Mining* dikerjakan, perlu dilakukan proses *cleaning* pada data yang akan menjadi fokus KDD dan proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data lain yang relevan untuk keperluan KDD.

3. Transformasi

Transformasi adalah proses transformasi pada data yang dipilih, sehingga data tersebut sesuai untuk proses *Data Mining*. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau *pattern* informasi yang akan dicari pada *database*.

4. *Data Mining*

Pada tahapan ini merupakan proses mencari *pattern* atau pola dan informasi dari sebuah *database* dengan menggunakan teknik atau metode. Pada proses *Data Mining* terdapat banyak teknik, metode atau algoritma yang dapat digunakan dan sangat bervariasi dan untuk menentukan pemilihan metode yang akan digunakan tergantung pada tujuan dan proses KDD secara keseluruhan.

5. Evaluasi

Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Proses ini merupakan proses penerjemahan pola-pola yang didapatkan dari *Data Mining*. Pola informasi yang dihasilkan dari proses *Data Mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

2.2 *Data Mining*

Data Mining merupakan sebuah inti dari proses KDD, meliputi dugaan algoritma yang mengeksplor data, membangun model dan menemukan pola yang belum diketahui [10]. KDD bersifat otomatis, dapat didefinisikan sebagai pengorganisasian proses untuk pengidentifikasian yang benar, berguna dan penemuan pola dari kumpulan data yang besar dan kompleks [11]. Data mining adalah suatu proses pengumpulan informasi dan data yang penting dalam jumlah yang besar. Dalam proses ini sering kali memanfaatkan beberapa metode, seperti matematika, statistika dan pemanfaatan teknologi. Pengertian data mining ini juga dikenal dengan istilah lain yakni *Knowledge Discovery In Database (KDD)* dan *Data Analysis*.

Data Mining merupakan salah satu teknik untuk menemukan, mencari, atau menggali informasi atau pengetahuan baru dari sekumpulan data yang sangat besar, dengan integrasi atau penggabungan dengan disiplin ilmu lain seperti statistika, kecerdasan buatan, serta *machine learning*, menjadikan *Data Mining*

sebagai salah satu alat bantu untuk menganalisa data yang kemudian menghasilkan informasi yang berguna [12] Dari penjelasan tersebut, *Data Mining* dapat diartikan sebagai sebuah proses dari sekumpulan data yang memiliki jumlah besar untuk mendapatkan data yang hasilnya tidak hanya sekedar informasi melainkan merupakan sebuah pengetahuan atau *knowledge* yang tersembunyi dari data tersebut [13]. *Knowledge* yang diperoleh dari hasil *mining* terhadap sekumpulan data tersebut dapat membantu dalam pengambilan sebuah keputusan [14].

2.3 Association Rule

Association rule merupakan suatu proses pada *Data Mining* untuk menentukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support (minsup)* dan *confidence (minconf)* pada sebuah *database* [15]. Kedua syarat yang digunakan untuk *interesting association rules* dibandingkan dengan batasan yang telah ditentukan dengan *minimum support* dan *minimum confidence*. *Association rule* merupakan prosedur yang akan berguna untuk mencari suatu relasi antar objek dalam suatu dataset yang telah ditetapkan. Untuk tahap asosiasi tersebut terdapat 2 tahapan yaitu: dalam mencari kombinasi yang sering terjadi dan untuk mendefinisikan kondisi.

Association rule adalah teknik *Data Mining* untuk menemukan aturan asosiasi antara suatu kombinasi barang [16] Fungsi *association rule* seringkali disebut dengan “*market basket analysis*”, yaitu Analisis dari kebiasaan membeli *customer* dengan mencari asosiasi dan korelasi antara item-item berbeda yang

diletakkan *customer* dalam keranjang belanjanya. Contoh aturan asosiasi dari analisis pembelian di suatu produk pada toko sembako adalah dapat diketahuinya berapa besar kemungkinan seorang konsumen membeli beras bersamaan dengan membeli gula. Dengan pengetahuan tersebut, pemilik toko dapat menyediakan stok lebih barang tertentu yang sering dibeli oleh konsumen. Secara teori, terdapat beberapa hal yang akan dipakai dalam mengukur apakah kumpulan item yang akan muncul bersamaan atau tidak muncul dan ada beberapa rule metode seleksi. Untuk support of an item set ini merupakan sebuah persentase dari setiap transaksi yang terjadi dan mengandung item set tersebut. Dan untuk confidence of an association rule ini merupakan perbandingan antara nilai support dari kumpulan item tersebut dan nilai support dari kumpulan yang mendahuluinya. Tetapi dari sini ada satu pokok permasalahannya yang tidak begitu kita perhatikan, yaitu adalah rule yang bagus belum tentu rule yang akan menarik.

Dalam menentukan nilai *minimum support* sebuah *item* dapat menggunakan rumus persamaan seperti di bawah ini :

1. $Support(A) = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}}$
2. $Support(A \cap B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}}$

Sedangkan untuk menentukan nilai *minimum confidence* sebuah *item* dapat menggunakan rumus persamaan seperti di bawah ini :

1. $Confidance(A \rightarrow B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Jumlah Transaksi Mengandung A}}$

2.4 *FP-Tree*

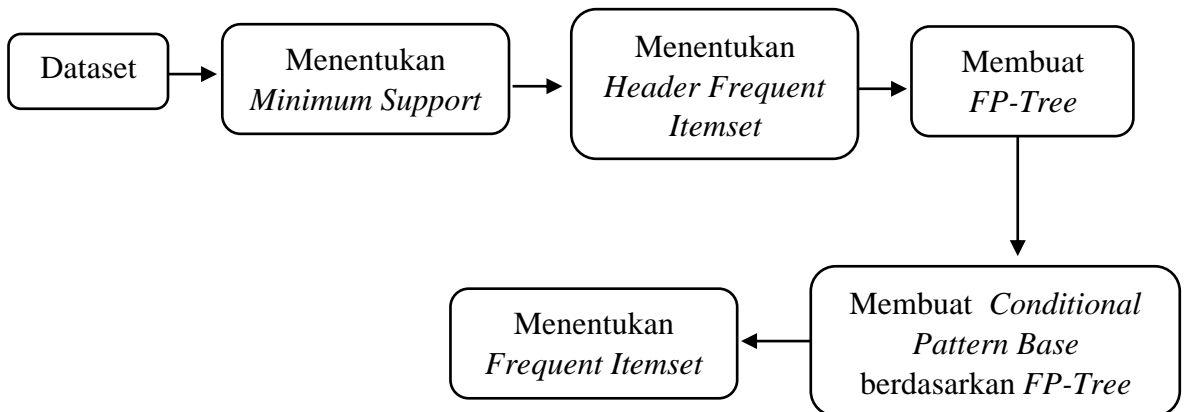
FP-Tree merupakan struktur penyimpanan data yang dimampatkan. *FP-Tree* dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam *FP-Tree* [17]. Karena dalam setiap pembelian atau transaksi yang dipetakan, mungkin ada transaksi yang mempunyai item yang sama, maka lintasannya kemungkinan juga untuk saling tumpang tindih. Semakin banyak data pembelian atau transaksi yang memiliki jenis(item) yang sama, maka pemanfaatan menggunakan struktur data *FP-Tree* semakin efektif. Kelebihan dari *FP-Tree* adalah Cuma membutuhkan dua kali pemindaian data transaksi yang terbukti sangat efisien. *FP-Tree (Frequent Pattern Tree)* digunakan bersamaan dengan algoritma *FP-Growth* untuk menentukan *frequent itemset* (data yang paling sering muncul) dari sebuah *dataset*.

2.5 **Algoritma *FP-Growth***

FP-Growth adalah algoritma alternatif yang dapat digunakan untuk menentukan itemset yang paling sering muncul dalam satu set data [7]. Algoritma *FP-Growth* merupakan salah satu cara alternatif untuk menemukan himpunan data yang paling sering muncul (*frequent itemset*) tanpa menggunakan generasi kandidat. Algoritma *FP-Growth* sangat efisien untuk menentukan *frequent pattern* baik dalam data yang besar maupun kecil, dibandingkan dengan algoritma apriori algoritma *FP-Growth* lebih cepat karena algoritma *FP-Growth* tidak perlu melakukan iterasi secara berulang seperti algoritma apriori yang membutuhkan waktu yang cukup lama. Dan algoritma apriori tidak dapat menangani data yang

besar, sehingga algoritma *FP-Growth* inilah yang dapat dijadikan sebagai solusi bagi permasalahan yang terjadi pada algoritma apriori.

Gambar 2.2 Tahapan-tahapan yang dilakukan dalam algoritma *FP-Growth* adalah sebagai berikut :



Gambar 2.2 Blog Diagram Algoritma *FP-Growth*

Contoh dari tahapan-tahapan yang dilakukan dalam algoritma *FP-Growth* misalnya pada data transaksi pembelian produk pada toko sembako untuk menentukan itemset yang paling sering muncul dalam satu set data. Untuk langkah awal mempermudah dalam pengolahan data maka data tersebut dikonversi ke dalam bentuk *spreadsheet* dengan format *.xls*.

Langkah pertama adalah menyusun data transaksi seperti pada tabel 2.1

TID	Jenis Produk
1	Gula, Kopi, Roti
2	Beras, Minyak, Gula, Garam
3	Gula, Teh, Mie Instan
4	Susu, Gula, Kopi
5	Minyak, Beras, Teh, Kopi

Tabel 2.1 Contoh Data Transaksi

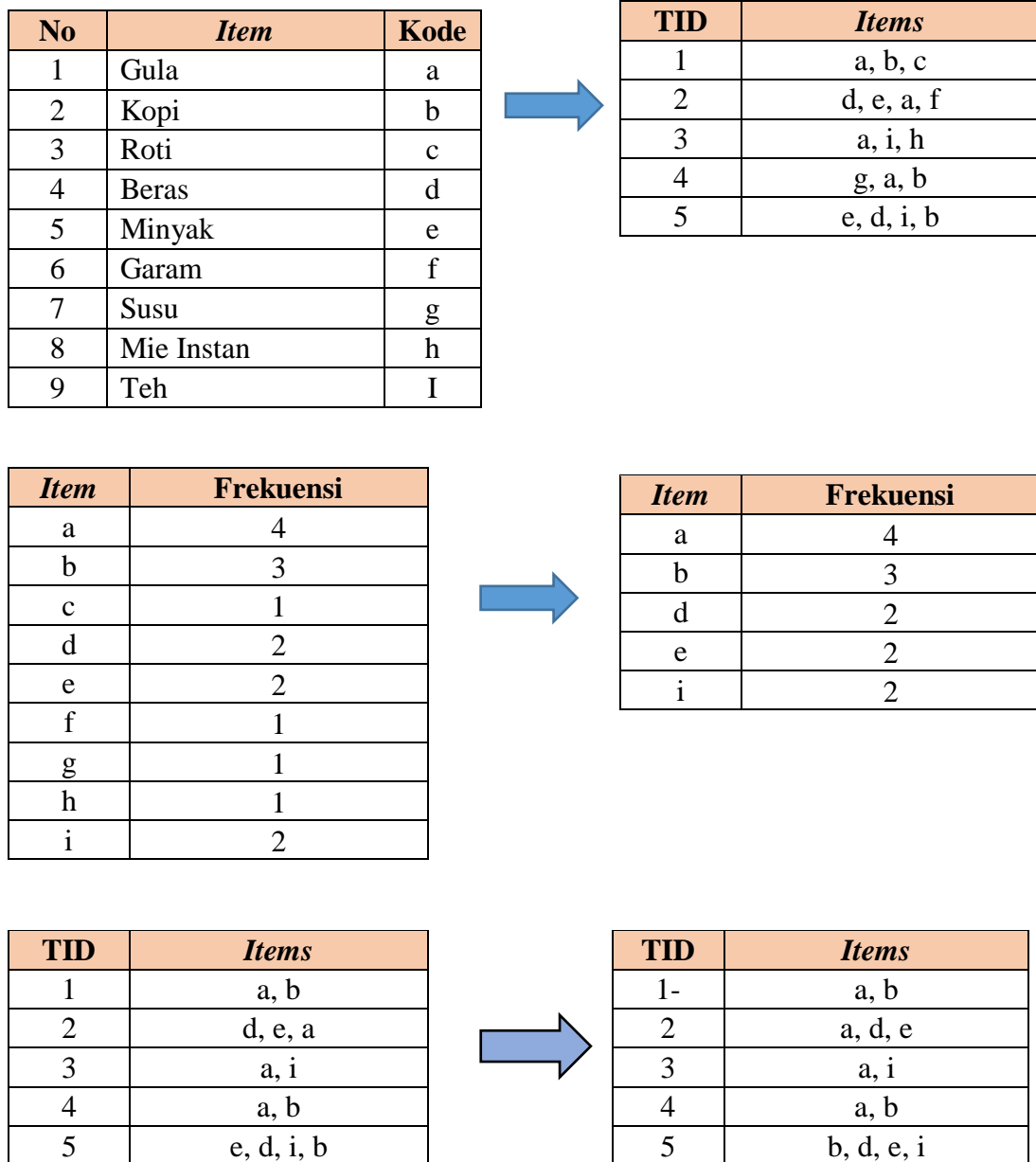
Data transaksi kemudian disusun ke dalam bentuk tabular data, maka data transaksi penjualan toko sembako dikonversi ke dalam bentuk *binary* dengan angka 0 dan 1. Di mana 1 adalah jika barang dibeli dan 0 jika barang tidak dibeli. Hasil proses konversi data transaksi penjualan dalam bentuk tabular data dapat dilihat seperti Tabel 2.2

TID	Gula	Kopi	Roti	Beras	Minyak	Garam	Susu	Mie Instan	Teh
1	1	1	1	0	0	0	0	0	0
2	1	0	0	1	1	1	0	0	0
3	1	0	0	0	0	0	0	1	1
4	1	1	0	0	0	0	1	0	0
5	0	1	0	1	1	0	0	0	1

Tabel 2.2 Contoh Tabular Data Transaksi

Setelah mengkonfersi data dalam bentuk *binary* kemudian memberikan kode kepada setiap *item* agar lebih mudah dalam melakukan proses data. Setelah dilakukan pengkodean selanjutnya adalah mempersiapkan data dengan kode, selanjutnya dibuatlah frekuensi kemunculan setiap *item* dari keseluruhan transaksi awal. Setelah frekuensi setiap *item* diperoleh, kemudian dibatasi dengan *support count*. Jika frekuensi *item* tidak kurang dari *support count* maka *item* tersebut akan dihapus dan tidak dipakai dalam proses *data mining*. Selanjutnya dilakukan *filter* sesuai dengan *support count*. Dan kemudian Tahap selanjutnya adalah memindai

tabel berdasarkan frekuensi tertinggi, seperti pada Tabel 2.3 Tahapan-tahapan dalam algoritma *FP-Growth* .



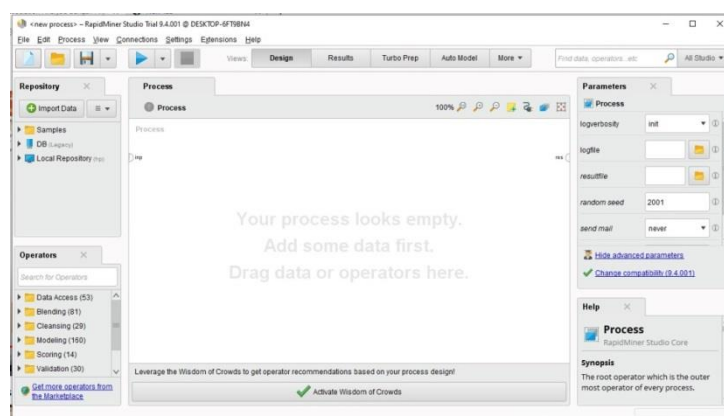
Tabel 2.3 Tahapan-tahapan dalam algoritma *FP-Growth*

2.6 Rapidminer

Rapid Miner merupakan perangkat lunak yang bersifat terbuka (open source). Rapid Miner adalah solusi untuk melakukan analisis terhadap *Data Mining*, text mining dan analisis pengelompokan [3]. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner adalah platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan bernama sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambahan teks, dan analisis prediktif. Hal ini digunakan untuk bisnis dan komersial, juga untuk penelitian, pendidikan, pelatihan, *rapid prototyping*, dan pengembangan aplikasi serta mendukung semua langkah dalam proses pembelajaran mesin termasuk persiapan data, hasil visualisasi, validasi model, dan optimasi.

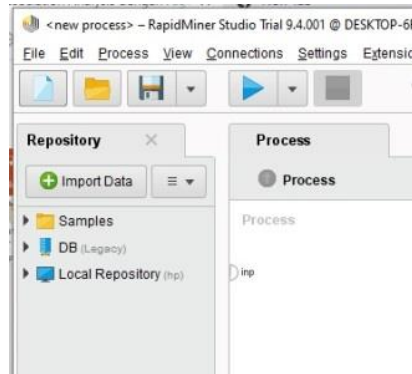
Adapaun menu-menu yang digunakan pada proses data mining dengan metode algoritma FP-Growth pada aplikasi Rapidminer adalah sebagai berikut :

a. Tampilan Awal Rapidminer



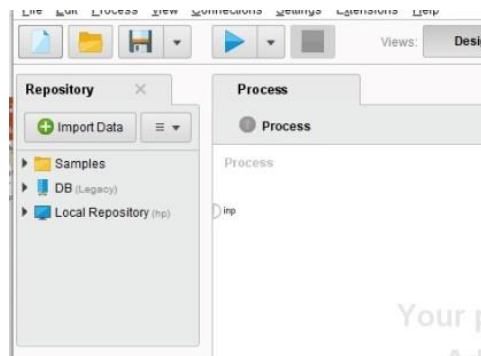
Gambar 2.3 Tampilan Default Rapidminer 9.4

- b. Menu *new process* digunakan untuk membuat proses baru seperti tampilan dibawah ini Gambar 2.4



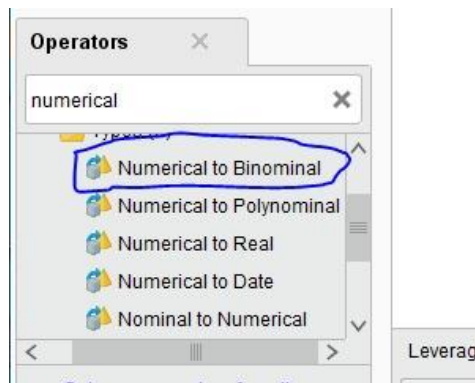
Gambar 2.4 Menu *new process*

- c. Menu *Import Data* digunakan untuk memasukkan data berupa file .xlsx seperti pada Gambar 2.5 dibawah ini



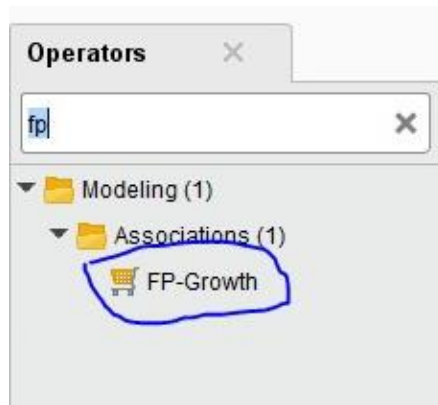
Gambar 2.5 Menu *Import Data*

- d. Selanjutnya ada 3 operator yang digunakan dalam memproses data menggunakan algoritma FP-Growth yang pertama adalah operator *Numerical to Binominal* digunakan untuk membaca data yang numeric ke binominal untuk tampilan menunya dapat dilihat pada Gambar 2.6 dibawah ini



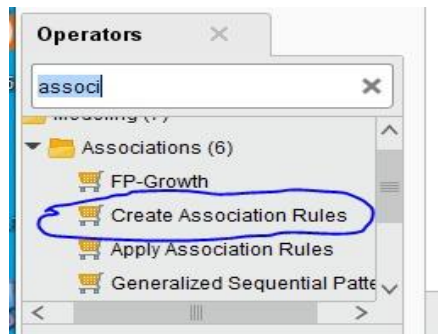
Gambar 2.6 Menu Operator *Numerica to Binominal*

- e. Yang kedua Menu Operator FP-Growth digunakan untuk memproses data dengan algoritma FP-Growth bisa dilihat pada Gambar 2.7 dibawah ini



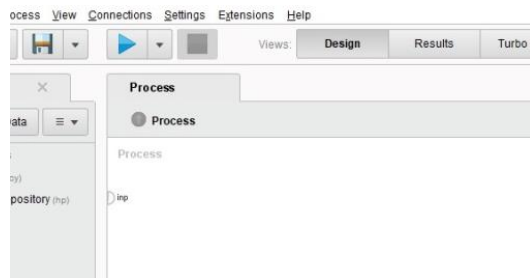
Gambar 2.7 Menu Operator *FP-Growth*

- f. Selanjutnya operator ketiga yaitu Operator *Association Rules* digunakan untuk mendapatkan pola-pola assosiasi untuk gambar menunya dapat dilihat dibawah ini



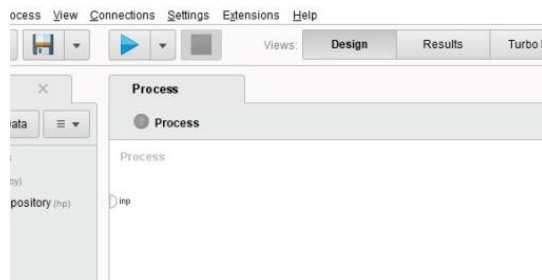
Gambar 2.8 Menu Operator *Create Association Rules*

- g. Menu Run atau Eksekusi digunakan untuk memproses / menjalankan data yang akan diolah terlihat pada Gambar 2.9 dibawah ini



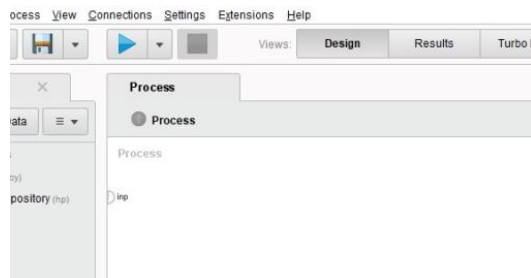
Gambar 2.9 Menu Run

- h. Menu Design digunakan untuk mengatur data dan operator-operator yang digunakan terlihat pada Gambar 2.10 dibawah ini



Gambar 2.10 Menu Design

- i. Selanjutnya Menu Result ini digunakan untuk menampilkan hasil dari proses data mining yang sudah set di menu design terlihat pada Gambar 2.11 dibawah ini



Gambar 2.11 Menu Result

2.7 Metodologi Penelitian

Dalam Kamus Besar Bahasa Indonesia (KBBI), metodologi adalah ilmu tentang metode; uraian tentang metode. Sedangkan pengertian penelitian dalam Kamus Besar Bahasa Indonesia (KBBI), penelitian adalah kegiatan pengumpulan, pengolahan, analisis, dan penyajian data yang dilakukan secara sistematis dan objektif untuk memecahkan suatu persoalan atau menguji suatu hipotesis untuk mengembangkan prinsip-prinsip umum. Metodologi penelitian merupakan cara untuk mengetahui hasil dari sebuah permasalahan yang spesifik dari suatu penelitian. Metodologi penelitian ini dilakukan dengan cara sistematis yang akan digunakan sebagai acuan dalam penelitian.

Manfaat metodologi penelitian adalah:

1. Memudahkan pekerjaan peneliti agar sampai pada tahap pengambilan keputusan atau kesimpulan-kesimpulan.

2. Untuk mengatasi berbagai keterbatasan yang ada, misalnya keterbatasan waktu, biaya, tenaga, etik, dan lain-lain.
3. Kesimpulan yang diambil oleh peneliti dapat dipercaya.
4. Kesimpulan yang diambil dapat digunakan untuk memecahkan permasalahan.