

BAB 2

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka Penelitian yang terkait dengan Algoritma Apriori

Tinjauan pustaka merupakan paparan penelitian lain yang berkaitan dengan penelitian yang akan dilakukan. Tinjauan pustaka yang dituliskan meliputi masalah, atribut atau variabel, metode penyelesaian, dan hasil. Penelitian dengan menggunakan algoritma apriori sudah banyak dilakukan untuk penelitian. Beberapa penelitian 5 tahun terakhir yang berhubungan dengan penelitian ini, maka penulis akan menggunakannya sebagai rujukan agar penelitian ini semakin baik. Berikut adalah beberapa penelitian nya :

2.1.1 Penerapan *Association Rule* Dengan Algoritma *Apriori* Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa Teknik S1 Fakultas Ilmu Komputer Universitas Dian Nuswantoro

Dalam penelitian ini permasalahan yang terjadi adalah jumlah mahasiswa yang mendaftar tiap tahun nya kebanyakan tidak lulus tepat waktu hal ini mengakibatkan banyak masalah antara lain turun nya akreditasi pada perguruan tinggi menurun, lulusan mahasiswa kurang berkualitas, serta dapat menurunkan kinerja program studi. Tujuan penelitian ini adalah mengetahui faktor apa sajakah yang membuat hal tersebut bisa terjadi. Oleh karena itu dalam penelitian ini penulis akan menerapkan Algoritma Apriori untuk menampilkan informasi tingkat kelulusan mahasiswa Fakultas Ilmu Komputer dengan aplikasi SPMF untuk mengetahui nilai support dan nilai confiden dari data mahasiswa yang telah diolah. [8]

Dari penelitian di atas maka dapat di simpulkan bahwa :

1. *Assosication rule mining* menggunakan algoritma apriori menghasilkan 8 rule menggunakan software SPMF dengan menentukan nilai minimum support 0.2% dan minimum confidance 50% dari keterkaitan antara data induk mahasiswa dan data kelulusan mahasiswa.

2. Pola data yang telah di temukan pada data induk mahasiswa yang memuat atribut jalur masuk dengan kategori reguler memiliki kecenderungan asosiasi yang kuat dengan memuat 6 rule dan atribut predikat prestasi dengan kategori lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,76 – 3,50 memuat 3 rule.

3. Kecenderungan pola yang terbentuk dari association rule mining data induk dan kelulusan mahasiswa dapat disimpulkan bahwa rata-rata mahasiswa dari jalur masuk reguler, PMDK, dan khusus memiliki tingkat kelulusan dengan lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,76 – 3,50.

2.1.2 Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus di Fakultas MIPA Universitas Diponegoro)

Di dalam penelitiannya membahas tentang algoritma apriori yang digunakan untuk menentukan tingkat kelulusan mahasiswa dengan menggunakan data induk serta data kelulusan mahasiswa yang sudah ada. Data yang digunakan adalah data yang berasal dari buku wisuda ke-115 di mana 48 dari 80 peserta yang telah di wisuda program reguler di Fakultas MIPA yang menempuh masa studi 8 semester lebih. Tingkat kelulusan di peroleh dari data lama studi dan IPK. [9]

Batas minimum support yang di berikan dalam penelitian ini adalah 1 hingga 3. Setelah proses yang di lakukan sistem akan menampilkan hasil berupa tabel yang menghubungkan kekuatan nilai support dan confidence dari masing-masing atribut dan threshold yang di gunakan. Maka dapat di simpulkan jika nilai confidence serta nilai support nya makin tinggi maka nilai dari hubungan antara atribut tersebut makin kuat.

2.1.3 Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat – Alat Kesehatan

Di dalam Apotek kelambir – 2 Medan permasalahan yang terjadi adalah sering kosong nya kebutuhan obat yang di butuhkan di dalam apotek tersebut , maka dari itu sangat penting adanya sistem persediaan barang pada apotek dan apotek juga harus tahu barang apa sajakah yang menjadi prioritas utama yang harus di stok untuk mengantisipasi kekosongan barang. Minimnya stok barang di apotek dapat mempengaruhi pelayanan terhadap konsumen dan pendapatan pada apotek. [6]

Berdasarkan data yang telah di proses dengan algoritma apriori dan di implementasikan pada aplikasi tanagra maka peneliti menarik kesimpulan bahwa: *Dataamining* dapat diterapkan dengan menggunakan database penjualan pada alat-alat kesehatan karena dapat diatemukan nya kecenderungan pola kombinasi *itemsets* sehingga dapat di jadikan nya sebagai acuan atau informasi yang berharga dalam pengambilan keputusan agar pihak apotek bisa mempersiapkan stok jenis barang apa sajakah yang diperlukan pada kemudian hari.

Dalam penerapan Algoritma Apriori pada teknik data mining hasil nya adalah sangat efisien dan mempercepat proses kecenderungan pola kombinasi itemset pada hasil penjualan alat kesehatan yang terdapat pada Apotek Kelambir – 2 Medan.

2.1.4 Penerapan Algoritma Apriori Dalam Menemukan Hubungan Data Awal Masuk Mahasiswa Dengan Prestasi Akademik (STUDI KASUS : STAI Miftahul Ulum Tanjungpinang)

Berdasarkan observasi yang dilakukan secara langsung di STAI Miftahul Ulum Tanjungpinang dapat diketahui bahwa jumlah kelulusan pada mahasiswa angkatan 2006,2007,2008 dan 2009 denganjumlah 1064 mahasiswa yang menempuh masa studi kurang dari 8 semester dan lebih dari 8 semester, di ketahui memiliki prestasi akademik yaitu IPK (indeks prestasi akademik) yang memuaskan, sangat memuaskan dan *cumlaude* (dengan pujian). [10] Dari hal

tersebut belum diketahui faktor apa sajakah yang mempengaruhi tinggi rendahnya lama studi dan IPK mahasiswa, oleh karena itu penelitian ini bertujuan untuk mencari hubungan antara data awal masuk mahasiswa dengan prestasi akademik yang diperoleh oleh mahasiswa. Algoritma apriori digunakan di karenakan algoritma ini dapat mencari nilai frekuensi yang tertinggi dengan cara mengkombinasikan item dengan item yang lainnya sehingga tidak terbentuk kombinasi lagi. Hasil yang di dapait dari proses algoritma apriori yaitu berupa nilai *support* dan *confidence*. Penerapan Algoritma Apriori ini dengan menggunakan bahasa pemrograman yaitu java neatbeans. Hasil yang di dapat adalah faktor yang mempengaruhi tinggi nya prestasi akademik berdasarkan data awal masuk mahasiswa yaitu dengan cara melihat nilai confidence tertinggi yang dimiliki oleh mahasiswa yang berasal dari program studi Pendidikan Agama Islam dengan anilai confidance 45,98% yang memiliki lama studi kurang dari 4 tahun dan IPK 3,50 – 4.00.

2.1.5 Implementasi Algoritma Apriori Dalam Menentukan Program Studi Dalam Menentukan Program Studi yang Diambil Mahasiswa

Tujuan dari penelitian ini adalah untuk menentukan konsentrasi program studi yang akan diambil oleh mahasiswa dengan menggunakan teknik data mining menggunakan metode Algoritma Apriori. Sistem ini membutuhkan data latih agar dapat membandingkan apakah mahasiswa dengan kriteria A cocok dengan program studi x. Prediksi kecocokan yang akan diperoleh berdasarkan hasil perbandingan mahasiswa berbeda yang memiliki kemiripan data demografi dengan mahasiswa A. Dengan digunakan nya Algoritma Apriori maka akan di dapatkan hasil yang berupa aturan – aturan yang terdiri dari kumpulan beberapa frequent item set dengan nilai confident yang tinggi. [7]

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
1.	Riko Adhi Saputro,	Jumlah mahasiswa yang masuk tidak sama dengan mahasiswa yang lulus	Penerapan association rule dengan algoritma apriori.	Association rule mining dengan menggunakan algoritma apriori berhasil di implementasikan dapat menemukan 8 rule dengan menggunakan software SPMF dengan ditentukan nilai minimum support 0.2% serta minimum confidence 50 %. Dari keterkaitannya antara dua data yaitu data induk mahasiswa dan data kelulusan mahasiswa.
2.	Nuqson Masykur Huda , 2010	Berdasarkan buku wisuda angkatan	Algoritma Apriori	Dengan cara memanfaatkan data induk mahasiswa dan data kelulusan

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
		115, 48 dari 80 peserta wisuda adalah mahasiswa yang menempuh masa studi lebih dari 8 semester dari yang di jadwalkan adalah 8 semester .		mahasiswa , maka akan di ketahui informasi tingkat kelulusan mahasiswa melalui teknik data mining .
3	Kennedi tampubolon ,et.al, 2013	Penting nya sistem persediaan barang yang terdapat di apotek dan barang apa sajakah yang menjadi prioritas utama yang harus di	Algoritma apriori dengan menetapkan support minimal 16 % dan nilai confidence nya adalah 70%.	Dapat menemukan pola kombinasi itemset sehingga dapat dijadikan nya sebagian informasi yang sangat berharga dalam mengambil keputusan untuk memepersiapkan stok jenis barang apa sajakah yang perlu di stok.

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
		stok untuk memenuhi kebutuhan pelanggan dan untuk mengantisipasi pasokan barang yang kosong.		
4	Heruandika Cahyono Pratama, et.al	Dari mahasiswa yang berjumlah 1064 banyak yang menempuh masa studi lebih dari 8 semester.	Penerapan Algoritma dengan menggunakan bahasa pemrograman java netbeans	Berupa nilai support dan nilai confidence.
5	Ahmad FikriFajri, 2016	Untuk masuk ke dalam program studi yang diinginkan adalah hal yang sangat	Algoritma apriori	Hasil berupa aturan yang merupakan beberapa kumpulan frequent item set dengan berupa nilai confidence yang tinggi.

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
		<p>di inginkan oleh mahasiswa tentu nya harus sesuai dengan kriteria. Jika tidak sesuai dengan kriteria maka akan timbul beberapa masalah</p>		

2.2 Landasan Teori

Untuk mendukung penelitian ini, berikut adalah teori – teori yang mendukung penelitian ini yang berasal dari berbagai sumber.

2.2.1 Konsep Data mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. [11] Dalam mencari informasi pada proses Knowledge Discovery in Database (KDD) mempunyai beberapa klasifikasi pendekatan . Beberapa pendekatan tersebut antara lain pendekatan kuantitatif dan probabilistik yang meliputi analisis pohon keputusan, logika induktif, pencarian

pola. Pendekatan selain kuantitatif dan probabilistik meliputi jaringan saraf tiruan, deviasi, algoritma genetik, dan lain sebagainya.

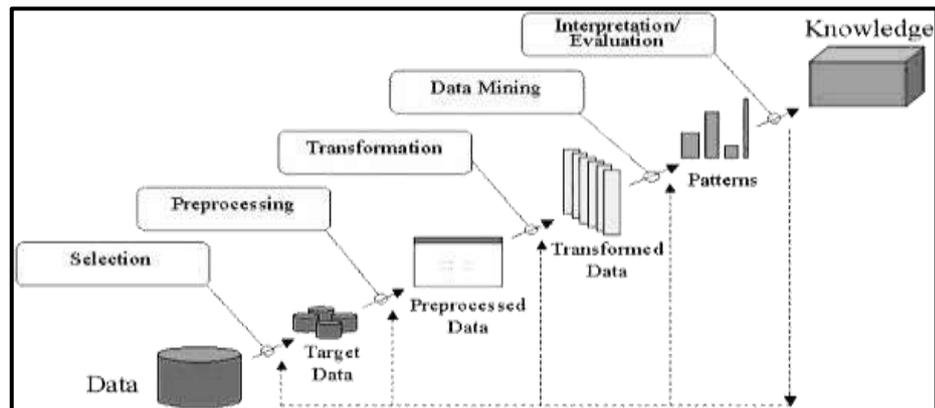
Ada Beberapa elemen dalam pencarian informasi pada Knowledge Discovery in Database (KDD) [6]:

1. Data yang dikerjakan harus dalam jumlah besar
2. Efisiensi dalam volume data harus di perhatikan
3. Keakuratan harus diutamakan
4. Bahasa yang digunakan adalah bahasa tingkat tinggi
5. Bentuk dari pembelajaran otomatis harus digunakan proses pencarian informasi
6. Hasil yang dikeluarkan harus semenarik mungkin

2.2.2 Definisi Data Mining

Data Mining berisi pencarian *trend* atau pola yang diinginkan dalam *database* yang besar untuk membantu pengambilan keputusan di waktu yang akan datang. Harapannya, perangkat *data mining* mampu mengenali pola – pola ini dalam data dengan jumlah yang sedikit. Ada beberapa hal yang penting tentang data mining yaitu :

1. Proses otomatis dari data yang tersedia
2. Data yang digunakan adalah data yang sangat besar
3. Keluaran yang diharapkan dari proses datamining adalah untuk memperoleh hubungan atau pola yang bisa digunakan.



Gambar 2.1 Proses Penemuan Pengetahuan

Proses *KDD* secara garis besar dapat dijelaskan sebagai berikut :

1. Penyeleksian Data

Dari data yang didapat merupakan data *raw* atau data mentah yang sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari data *raw*. Artinya dalam sebuah penelitian, data yang diambil harus sesuai dengan kriteria yang ada dan melalui proses seleksi yang cermat.

2. Pra pengolahan (*Preprocessing*)

Data yang masuk dari berbagai sumber (data source) banyak terdapat *noise data* atau kesalahan dan data yang tidak konsisten atau data tidak relevan. Data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid, dan atribut data yang tidak releva adengan hipotesa data mining yang dimiliki. Sehingga perlu dilakukan pembersihan data atau *data cleaning* sebelum memasuki data ditransformasikan. Atribut-atribut yang ada akan diidentifikasi entitas-entitas yang unik seperti atribut nama, jenis, nomor id, dan lainnya. *Preprocessing* perlu dilakukan secara cermat karena kesalahan pada data dapat menghasilkan hasil yang menyimpang pada pengambilan keputusan nantinya.

3. Transformasi data

Transformasi data (*transformation data*) merupakan penggabungan dengan format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum proses pengaplikasian. Data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval.

4. *Data Mining*

Proses *mining* ini dilakukan dengan menerapkan suatu metode atau perhitungan algoritma dari *data prepared*, untuk menemukan pola atau *patterns* pohon keputusan untuk mengetahui apakah hipotesa sudah tercapai.

5. Interpretasi dan Evaluasi

Dari pola yang ada, maka untuk memperoleh pengetahuan atau *knowledge*, tahap terakhir adalah mengidentifikasi pola atau *pattern* pohon keputusan yang diperoleh dan diterjemahkan ke dalam bentuk yang dapat dimengerti.

6. Presentasi pengetahuan (*knowledge presentation*)

Di dalam tahapan ini menampilkan dan menyajikan pengetahuan mengenai teknik yang digunakan untuk memperoleh pengetahuan yang diperoleh oleh pengguna. Ini adalah tahap terakhir dari proses *data mining* yang menjelaskan tentang bagaimana memformulasikan keputusan dari hasil analisa yang didapat.

2.2.3 Pengelompokan Data Mining

Berdasarkan tugasnya, data mining dikelompokan sebagai berikut :

1. Fungsi Deskripsi (*description*)

Banyak penelitian mencari data untuk mendapatkan pola yang dapat menggambarkan data tersebut . Kemungkinan penjelasan dari suatu pola didapat dari diskripsi pola itu sendiri

2. Fungsi Estimasi (*estimation*)

Estimasi dibangun dari beberapa record yang mempunyai nilai dari variable target yang dapat digunakan sebagai prediksi . Variable prediksi digunakan untuk peninjauan dari estimasi nilai dari variable target

3. Fungsi Prediksi (*prediction*)

Prediksi adalah hasil yang dapat digunakan untuk masa depan. Prediksi bisa menggunakan metode dan teknik yang ada pada klasifikasi dan estimasi

4. Fungsi Klasifikasi (*classification*)

Variable kategori digunakan dalam klasifikasi. Contohnya , Nilai dapat dipisahkan menjadi nilai tinggi, nilai sedang, nilai rendah.

5. Fungsi Pengklusteran (*cluster*)

Pengumpulan record, pengamatan, atau memperhatikan dan membentuk objek – objek yang mirip dapat dikatakan sebagai pengklusteran. Kluster sendiri dapat di definisikan sebagai recordh yang mempunyai kesamaan dan perbedaan dengan record data yang yang berada dalam kluster lain. Perbedaan pengklusteran dan klasifikasi ada pada variable target yang digunakan.

6. Fungsi Asosiasi (*association*)

Dalam proses menemukan atribut pada data mining merupakan tugas dari asosiasi. Pada dunia bisnis biasanya dikenal dengan analisis keranjang belanja.

2.2.4 Langkah – langkah Data Mining

Berikut adalah langkah – langkah data mining :

1. Mendefinisikan masalah yang ingin di teliti. Seperti contoh ingin mengetahui apakah seorang calon mahasiswa memiliki potensi masuk kedalam suatu jurusan tertentu atau tidak . pertanyaan tersebut kemudian dijawab oleh data mining, kemudian tentukan tieo tugas yang cocok dalam menjawab pertanyaan tersebut.

2. Mempersiapkan data yang dapat dijadikan sumber . Setelah masalah didapatkan kemudian mencari data yang dapat mendukung definisi dari masalah tersebut . menentukan jumlah data yang digunakan untuk mentraining data maining dengan algoritma data mining yang ada. Lalu sebagian data tersebut diberikan kedalam algortima data mining
3. Model *building* dan *validation*. Validasi dilakkan untuk memastikan data mining menghasilkan prediksi yang akurat. Pengujian dan validasi dari data mining dilakukan setelah proses training data selesai
4. *Deployment* . Proses deployment adalah proses dalam menentukan aplikasi data mining yang tepat.

2.2.5 Algoritma Apriori

Algoritma Apriori adalah jenis aturan asosiasi pada *data mining*. Selain algoritma apriori yang termasuk dalam jenis aturan asosiasi adalah *generalized rule induction* dan *algoritma hash based* [12]. Association Rule adalah salah satu teknik dari data mining yang digunakan untuk menemukan aturan asosiasi dari beberapa kombinasi item . Contoh Association Rule yaitu dalam toko perlengkapan olahraga dapat diketahui berapa besar kemungkinan seorang membeli sepatu bersamaan dengan membeli kaos kaki. Dari pengetahuan yang didapatkan pemilik toko perlengkapan olahraga dapat mengatur peletakan sepatu dengan kaos kaki olahraga atau promosi yang melibatkan kedua item tersebut. Data yang digunakan untuk mendapatkan sebuah data pengetahuan dalam aturan asosiasi adalah data latihan.

2.2.5.1 Tahap Algoritma Apriori

Analisis asosiasi merupakan dasar dari Teknik data mining lainnya . Analisis pola frekuensi tinggi (*frequent pattern mining*) adalah salah satu tahap Analisis asosiasi yang banyak digunakan oleh para peneliti untuk menghasilkan beberapa algoritma yang efisien.

2 tahap analisis asosiasi adalah :

1. Analisis pola nilai penunjang (*support*)

Dalam mencari kombinasi item yang memenuhi syarat minimum dari nilai pendukung yang ada di dalam database dilakukan pada tahap ini.

Rumus nilai support :

$$\text{Support}(A) = \frac{\text{jumlahTransaksiMengandung}A}{\text{NilaiTransaksi}} \quad (2.1)$$

Rumus nilai support 2 item :

$$\text{Support}(A, B) = \text{Support}(A \cap B) = \frac{\text{jumlahTransaksiMengandung}A \text{ dan } B}{\text{NilaiTransaksi}} \quad (2.2)$$

2. Perhitungan nilai kepastian (*confidence*)

Aturan asosiasi ditentukan setelah pola frekuensi tinggi ditemukan . Aturan asosiasi yang dicari adalah aturan asosiatif yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiasi “ jika A maka B “. Rumus mencari Nilai confidence dari aturan “ jika A maka B “ :

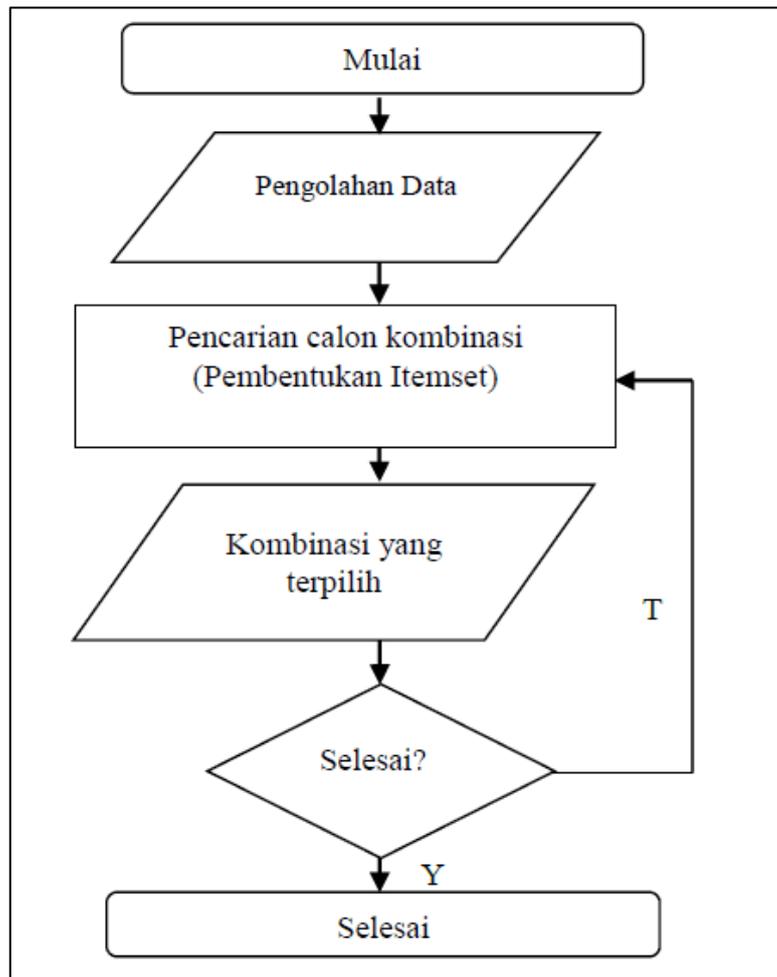
$$\text{confidence} = \frac{\text{jumlahTransaksiMengandung}A \text{ dan } B}{\text{jumlahTransaksi mengandung}A} \quad (2.3)$$

2.2.5.2 Langkah – Langkah Proses Aturan Asosiasi

Berikut adalah langkah – langkah dalam proses aturan asosiasi:

1. Mencari 1-itemset (himpunan item yang terdiri dari 1 item) dan menghitung nilai supportnya dengan cara men-scan database. Lalu bandingkan nilai support yang di dapatkan dengan nilai minimum support yang ditentukan terlebih dahulu . a itemset tersebut termasuk dalam large itemset jika nilainya lebih besar atau sama dengan minimum support.
2. Hanya Itemset termasuk dalam large itemset yang diikutkan dalam iterasi selanjutnya.

3. Hasil large itemset pada iterasi pertama (L_1) digunakan pada iterasi kedua sistem untuk membentuk kandidat itemset kedua (L_2). Hasil large itemset pada iterasi selanjutnya digunakan untuk iterasi selanjutnya menggunakan hasil large itemset pada iterasi sebelumnya (L_{k-1}) untuk membentuk kandidat itemset berikut (L_k). Sistem akan menggabungkan (join) L_{k-1} dengan L_{k-1} untuk mendapatkan L_k , seperti pada iterasi sebelumnya hanya itemset termasuk dalam large itemset yang diikutkan dalam iterasi selanjutnya (di prune).
4. Itemset hasil join kemudian dihitung supportnya
5. Proses join dan prune akan membentuk kandidat yang akan terus dilakukan hingga himpunan kandidat itemsetnya null, atau sampai kandidat sudah tidak ada yang bisa dibentuk lagi.
6. Kemudian, association rule yang memenuhi nilai support dan confidence yang telah ditentukan dibentuk dari frequent itemset tersebut
7. Nilai yang memiliki kemiripan dianggap sebagai satu data dalam association rule
8. Nilai minimum harus terpenuhi dalam pembentukan association rule
9. Cari himpunan bagian L yang tidak kosong untuk setiap large itemset L . Dihasilkan rule dengan bentuk $aB(L-a)$ jika supportnya (L) dan supportnya (a) lebih besar dari minimum support Untuk setiap himpunan bagian tersebut.



Gambar 2.2 Flowchart Algoritma Apriori

2.2.6 Evaluasi atau Validasi Data Mining

Metode data mining yang diusulkan harus dilakukan validasi atau evaluasi dengan metode pengukuran standard. Pengukuran model data mining dapat dibagi menjadi tiga kriteria yaitu :

1. Akurasi (Accuracy)

Ukuran seberapa baik hubungan antara model dengan hasil atribut yang berasal dari data yang telah tersedia.

2. Keandalan (Reliability)

Ukuran di mana model data mining yang diterapkan dengan dataset berbeda akan menghasilkan sebuah model data mining dan dapat diandalkan apabila menghasilkan pola umum yang sama terlepas dari data testing yang tersedia.

3. Kegunaan (Usefulness)

Berisi berbagai metrik yang dapat menilai seberapa baik model tersebut memberikan informasi yang bermanfaat.

2.2.6.1 Lift Ratio

Lift ratio adalah parameter penting setelah mendapatkan nilai support dan confidence dalam association rules yang digunakan untuk mengevaluasi sebuah aturan asosiasi. *Lift ratio* adalah nilai yang menunjukkan kevalidan proses transaksi dan memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B. *Lift Ratio* dapat dihitung dengan rumus sebagai berikut :

$$Lift\ ratio = \frac{Support(A \cap B)}{Support(A) * Support(B)} \quad (2.4)$$

Support (A ∩ B) = nilai support yang mengandung transaksi A dan transaksi B

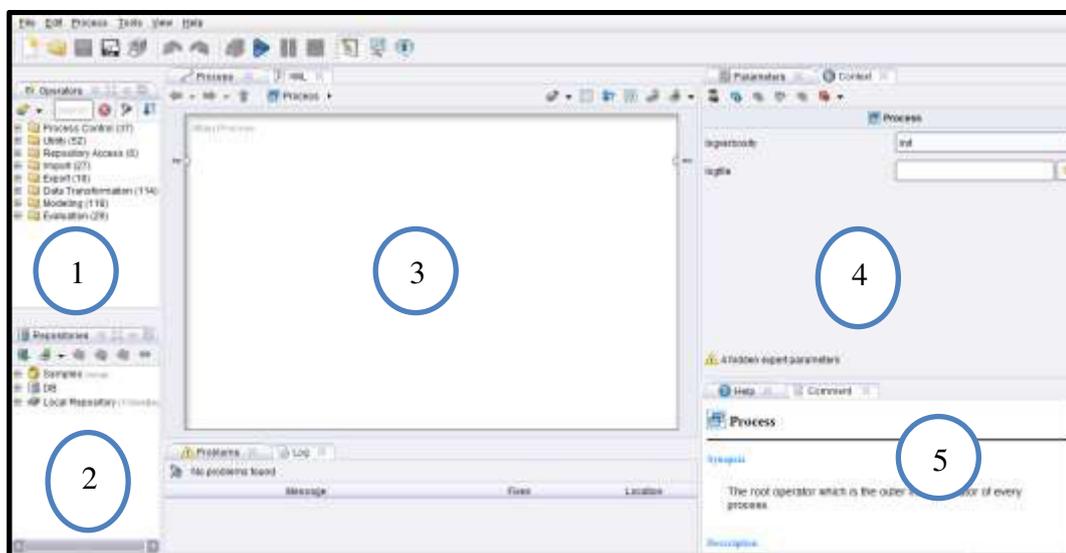
Support (A) = nilai support yang mengandung transaksi A

Support (B) = nilai support yang mengandung transaksi B

Jika nilai lift ratio > 1 maka artinya ada manfaat dari aturan tersebut. Semakin tinggi nilai lift ratio maka semakin besar kekuatan asosiasinya. Yang berarti bahwa produk A benar-benar dibeli bersamaan dengan produk B. [13]

2.2.7 RapidMiner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. [14] RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.



Gambar 2.3 Tampilan Lingkungan Kerja RapidMiner

Gambar 2.3 adalah tampilan dari lingkungan kerja RapidMiner. Berikut ini adalah penjelasannya :

1. Operator View

Operator View merupakan view yang paling penting pada perspective ini. Semua operator atau langkah kerja dari RapidMiner disajikan dalam bentuk kelompok hierarki di Operator View ini sehingga operator-operator tersebut dapat digunakan pada proses analisis. Hal ini akan memudahkan dalam mencari dan menggunakan operator yang sesuai dengan kebutuhan.

2. Repository View

Repository View merupakan komponen utama dalam Design Perspective selain Operator View. View ini dapat digunakan untuk mengelola dan menata proses Analisis menjadi proyek dan pada saat yang sama juga dapat digunakan sebagai sumber data dan yang berkaitan dengan meta data.

3. Process View

Process View menunjukkan langkah-langkah tertentu dalam proses analisis dan sebagai penghubung langkah-langkah tersebut. Di sini dapat menambahkan langkah baru dengan beberapa cara. Hubungan diantara langkah-langkah ini dapat dibuat dan dilepas kembali. Pada dasarnya bekerja dengan RapidMiner ialah mendefinisikan proses analisis, yaitu dengan menunjukkan serangkaian langkah kerja tertentu. Dalam RapidMiner, komponen proses ini dinamakan sebagai operator. Operator pada RapidMiner didefinisikan sebagai berikut :

- a. Deskripsi dari input yang diharapkan
- b. Deskripsi dari output yang disediakan
- c. Tindakan yang dilakukan oleh operator pada input, yang akhirnya mengarah dengan penyediaan output.
- d. Sejumlah parameter yang dapat mengontrol *action performed*.

4. Parameter View

Beberapa operator dalam RapidMiner membutuhkan satu atau lebih parameter agar dapat diindikasikan sebagai fungsionalitas yang benar. Namun terkadang parameter tidak mutlak dibutuhkan, meskipun eksekusi operator dapat dikendalikan dengan menunjukkan nilai parameter tertentu.

5. Help and Comment View

Setiap kali Anda memilih operator pada Operator View atau Process View, maka jendela bantuan dalam Help View akan menunjukkan penjelasan mengenai operator ini. Penjelasan yang ditampilkan dalam Help View meliputi:

- a. Sebuah penjelasan singkat mengenai fungsi operator dalam satu atau beberapa kalimat.
- b. Sebuah penjelasan rinci mengenai fungsi operator.
- c. Daftar semua parameter termasuk deskripsi singkat dari parameter, nilai default (jika tersedia), petunjuk apakah parameter ini adalah parameter ahli serta indikasi parameter dependensi.

Sedangkan Comment View merupakan area bagi Anda untuk menuliskan komentar pada langkah - langkah proses tertentu. Untuk membuat komentar, Anda hanya perlu memilih operator dan menulis teks di atasnya dalam bidang komentar. Kemudian komentar tersebut disimpan bersama-sama dengan definisi proses Anda. Komentar ini dapat berguna untuk melacak langkah-langkah tertentu dalam rancangan nantinya.