

BAB 2

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Gunadi dan Sensuse (2012) menerapkan dua algoritma yaitu algoritma *Apriori* dan algoritma *frequent pattern growth* (Fp-growth). Data yang digunakan merupakan data transaksi penjualan produk buku di Percetakan PT.Gramedia. Kedua algoritma ini digunakan untuk membentuk *frequent itemset* yang nantinya akan dijadikan sebagai acuan untuk merumuskan aturan-aturan asosiasi yang dihasilkan oleh model market basket *analysis* dengan menggunakan alat bantu perangkat lunak aplikasi Weka versi 3.6. Hasil yang diperoleh yaitu analisa yang dilakukan terhadap tingkat kekuatan aturan-aturan asosiasi menunjukkan bahwa aturan-aturan asosiasi yang dihasilkan oleh algoritma *apriori* memiliki tingkat kekuatan yang lebih tinggi dibandingkan dengan yang dihasilkan oleh algoritma FP-growth. Dengan tingkat akurasi dari algoritma *apriori* terhadap FP-growth sebesar 257,4543%.

Dewi Kartika Pane (2013) melakukan implementasi dengan menggunakan Algoritma Apriori, Penelitian ini menunjukkan bahwa sumber data penjualan Kreditplus untuk permintaan produk elektronik semakin meningkat. Data penjualan pada kreditplus selama ini tidak tersusun dengan baik, sehingga data tersebut hanya berfungsi sebagai arsip bagi perusahaan dan tidak dapat dimanfaatkan untuk

pengembangan strategi pemasaran. Penelitian ini bertujuan untuk menemukan *frequent itemsets* dijalankan pada sekumpulan data. Didefinisikan suatu proses untuk menemukan semua aturan apriori yang memenuhi untuk *support* dan syarat minimum untuk *confidence*.

Tampubolon, dkk (2013) melakukan implementasi dengan Algoritma Apriori pada data transaksi penjualan alat-alat kesehatan di Apotek Kelambir-2 Medan dari bulan Oktober sampai dengan bulan November 2013. Dalam penelitiannya Tampubolon dkk (2013) menggunakan program aplikasi Microsoft Excel 2007 dan Tanagra 1.4.48. Hasil yang diperoleh yaitu penerapan algoritma apriori pada teknik Data Mining sangat efisien dan dapat mempercepat proses pembentukan kecenderungan pola kombinasi itemset hasil penjualan alat-alat kesehatan di Apotek Kelambir-2 Medan.

Fadlina (2014) melakukan implementasi dengan menggunakan metode Algoritma Apriori, untuk menemukan pola-pola didalam data, dimana proses penemuan tersebut dilakukan secara otomatis atau semi otomatis dan pola-pola yang ditemukan harus bermanfaat sebagai ilmu pengetahuan yang baru dan informasi penting dari kejahatan jalanan (*Street Crime*) bagi kepolisian dan masyarakat dalam menanggulangi tingkat kejahatan jalanan dan sensitifitas kejahatan. Hasil dari perancangan *data mining* dengan algoritma apriori diperoleh informasi yang dibutuhkan oleh pihak kepolisian berupa presentase yang digunakan oleh bagian

reskrim untuk mengetahui kejahatan jalanan apa yang sering terjadi sehingga persoalan tingkat kejahatan jalanan dapat di minimalisasi.

Agung dkk (2015) melakukan implementasi dengan menggunakan metode Algoritma Apriori, untuk mengatur penempatan barang dimana aplikasi dibangun untuk menyajikan informasi yang berguna untuk meningkatkan pendapatan toko dengan acara mengatur penempatan barang yang saling berhubungan satu sama lain ketika konsumen membeli.

Berdasarkan penelitian-penelitian diatas, maka penelitian ini yaitu “Implementasi Data Mining Untuk Menganalisis Pola Peminjaman Buku Dengan Menggunakan Metode Algoritma Apriori Pada Perpustakaan STMIK AKAKOM Yogyakarta”.

Tabel 2.1 Perbandingan Penelitian

Nama (Tahun)	Judul	Isi
Goldie Gunadi Dana Indra Sensuse (2012)	Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori dan Frequent Pattern Growth (Studi Kasus : Percetakan PT.Gramedia)	Mengetahui tingkat kekuatan aturan-aturan asosiasi yang dihasilkan oleh algoritma <i>apriori</i> lebih besar dibandingkan dengan yang dihasilkan oleh algoritma <i>FP-growth</i> . Dengan tingkat akurasi dari algoritma apriori terhadap <i>FP-growth</i> sebesar 257,4543%.
Dewi Kartika Pane (2013)	Implementasi Data Mining Pada Penjualan Produk Elektronik Dengan Algoritma Apriori (Studi Kasus : Kreditplus)	Mengetahui produk elektronik terutama labtop untuk mencari <i>merk</i> labtop yang paling banyak dibeli pelanggan.

Tabel 2.2 Lanjutan Perbandingan Penelitian

Nama (Tahun)	Judul	Isi
Kennedi Tampubolon, Hoga Saragih, Bobby Reza (2013)	Implementasi Data Mining Algoritma Apriori pada Sistem Persediaan Alat-alat Kesehatan.	Ppenerapan Algoritma apriori pada teknik data mining sangat efisien dan dapat mempercepat proses pembentukan kecendrungan pola kombinasi itemset hasil penjualan alat-alat kesehatan.
Fadlina (2014)	Data Mining Untuk Analisa Tingkat Kejahatan Jalanan dengan Algoritma <i>Association Rule</i> Metode Apriori	Mengetahui kejahatan apa yang sering terjadi sehingga persoalan tingkat kejahatan jalanan dapat di minimalisasi.
Muhammad Thoriq Agung, Bowo Nurhadiyono (2015)	Penerapan Data Mining Pada Data Transaksi Penjualan Untuk Mengatur Penempatan Barang Menggunakan Algoritma Apriori	Pengaturan tata letak barang dilihat dari nilai <i>confidence</i> yang paling tinggi kemudian akan diletakan berdekatan, sedangkan dengan nilai <i>support</i> tertinggi akan diletakan ditempat yang paling mudah terlihat karena barang tersebut paling sering dibeli.
Maria Yolansia Sombo	Implementasi Data Mining Untuk Menganalisis Pola Peminjaman Buku Dengan Menggunakan Metode Algoritma Apriori Pada Perpustakaan STMIK AKAKOM Yogyakarta	Menghasilkan pola penataan buku untuk memberi kemudahan kepada semua pengunjung saat mencari buku.

2.2 Dasar Teori

Pada bagian ini penulis akan memaparkan beberapa tinjauan sebagai dasar teori dalam implementasi data mining untuk menganalisis pola peminjaman buku dengan menggunakan metode algoritma apriori pada perpustakaan STMIK AKAKOM Yogyakarta.

2.2.1 Data

Pengertian data menurut *Webster new World Dictionary*, data adalah *things known or assumed*, yang berarti bahwa data merupakan sesuatu yang diketahui atau dianggap. Data juga dapat didefinisikan sekumpulan informasi atau nilai yang diperoleh dari pengamatan (observasi) suatu objek, data dapat berupa angka dan dapat pula berupa lambang atau sifat. (syafizal helmi situmorang, 2010:1).

Menurut (AL-Bahra 2005:20), Data merupakan komponen dasar dari informasi yang akan diproses lebih lanjut untuk menghasilkan informasi. Sedangkan, menurut Longkutoy dalam bukunya “Pengenalan komputer”, Data adalah suatu istilah mejemuk yang berarti fakta yang mengandung arti yang digabungkan dengan kenyataan, simbol-simbol, gambar-gambar, angka-angka, huruf-huruf, atau simbol-simbol yang menunjukkan suatu ide, objek, kondisi, atau situasi dan lain-lain.

Dari beberapa definisi Dara dari para ahli dapat disimpulkan bahwa Data adalah suatu fakta yang bisa berupa simbol, gambar, angka, huruf yang dapat diproses lebih lanjut guna menghasilkan informasi.

2.2.2 Data Mining

Data mining merupakan proses iteratif dan interaktif untuk menemukan pola atau model baru yang sah (sempurna), bermanfaat dan dapat dimengerti dalam suatu *database* yang sangat besar (*massive database*).

Data mining berisis pencarian *trend* atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan teliti.

Data mining , sering juga disebut knowledge discovery in database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian sata historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan (Santosa, 2007).

Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola - pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasanyang kemudian dpat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya (Hermawati,2009).

Selain definisi di atas ada juga beberapa definisi maupun pengertian mengenai data mining seperti dibawah ini:

- a. Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang

berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data.

- b. Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar.
- c. Data mining ialah campuran dari statistic, kecerdasan buatan dan riset basis data yang masih berkembang.

Baskoro (2010) menyatakan bahwa adapun tujuan dari adanya data mining adalah:

- a. *Explanatory*, yaitu untuk menjelaskan beberapa kegiatan observasi atau suatu kondisi.
- b. *Confirmatory*, yaitu untuk mengkonfirmasi suatu hipotesis yang telah ada.
- c. *Exploratory*, yaitu untuk menganalisis data baru suatu relasi yang janggal.

Sebelum melakukan proses *data mining*, baiknya mengetahui terlebih dahulu apa yang bisa dilakukan oleh *data mining*, agar apa yang dilakukan nantinya memang sesuai dengan apa yang dibutuhkan serta menghasilkan nantinya memang sesuai dengan apa yang dibutuhkan serta menghasilkan sesuatu yang sebelumnya tidak diketahui dan bersifat serta bermanfaat bagi penggunaannya sendiri (Prasetyo Eko, 2012:25).

Pada dasarnya data mining mempunyai kegunaan serta tugas untuk mengspesifikasikan pola yang harus ditemukan dalam proses data mining. Secara umum tugas data mining dapat dibagi menjadi dua kategori yaitu:

a. Prediktif

Tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai dari atribut-atribut lainnya. Atribut yang diprediksi umumnya dikenal sebagai target atau variabel tak bebas, sedangkan atribut-atribut yang digunakan untuk membuat prediksi dikenal sebagai variabel bebas.

b. Deskriptif

Tujuan dari tugas deskriptif adalah menurunkan pola-pola (*korelasi, trend, cluster, trayektori*, dan *anomali*) yang meringkas hubungan yang pokok dalam data. Tugas data mining deskriptif sering disebut sebagai penyelidikan dan seringkali memerlukan teknik postprocessing untuk validasi dan penjelasan hasil.

Adapun pengelompokan Data Mining Menurut (Kusrini,2009:10), Pengelompokan data mining dijadikan sebagai teknik dari data mining berdasarkan tugas yang bisa dilakukan, yaitu:

a. Deskripsi

Para penulis biasanya mencoba menemukan cara untuk mendeskripsikan pola dan trend yang tersembunyi dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menentukan keterangan atau fakta bahwa siapa

yang tidak cukup profesional akan sedikit didukung dalam pemilihan Presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

b. Estimasi

Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih ke arah numerik dari pada kategori. Misalnya, akan dilakukan estimasi tekanan *systolic* dari pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Metode data mining yang terdapat di dalam estimasi yaitu (*Linear Regression, Neural Network, Support Vector*).

c. Prediksi

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi dimasa depan). Misalnya, ingin diketahui prediksi harga beras tiga bulan yang akan datang. Metode data mining yang terdapat di dalam prediksi yaitu (*Linear Regression, Neural Network, Support Vector*).

d. Klasifikasi

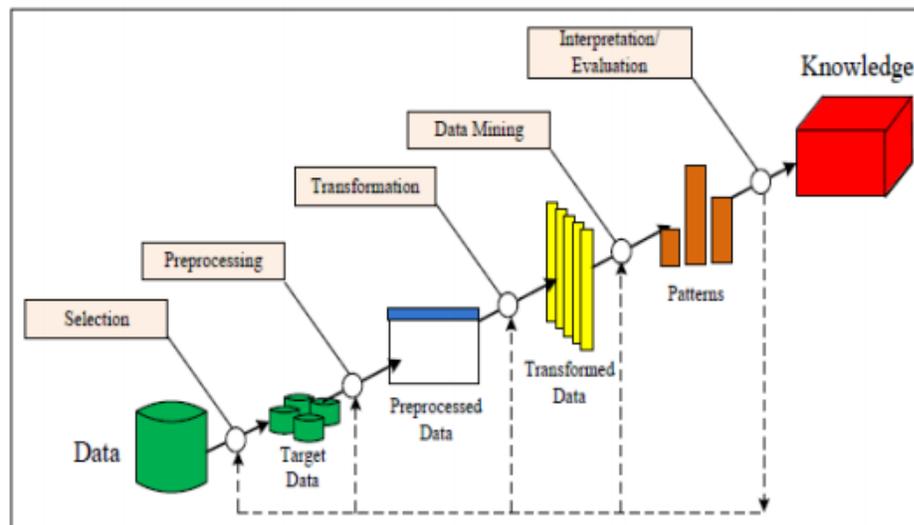
Dalam klasifikasi variabel, tujuan bersifat kategori. Misalnya, mengklasifikasikan pendapat dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Metode data mining yang terdapat di dalam klasifikasi yaitu (*Naive Bayes, K-Nearest, C4.5, ID3, CART, Linear Discriminant Analysis, Logistic Regression*).

e. Clustering

Clustering lebih kearah pengelompokan *record*, pengamatan, atau kasus dalam kelas yang memiliki kemiripan. Sebuah *cluster* adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lain dan memiliki ketidak miripan dengan *record-record* dalam *cluster* yang lain, misalnya untuk tujuan audit akuntansi akan dilakukan segmentasi perilaku financial dlam kategori dan mencurigakan. Metode data mining yang terdapat di dalam *clustering* yaitu (*K-Means*, *K-Medoids*, *Self-Organizing Map (SOM)*, *Fuzzy C-Means*).

2.2.3 Knowledge Discovery In Database (KDD)

Knowledge Discovery In Database (KDD) adalah proses menentukan informasi yang berguna serta pola-pola yang ada dalam data. Informasi ini terkandung dalam basis data yang berukuran besar yang sebelumnya tidak diketahui dan potensial bermanfaat. Data Mining merupakan salah satu langkah dari serangkaian proses *iterative* KDD (Kusrini,2009:7). Di bawah ini merupakan gambar Tahapan Dalam KDD :



Gambar 2.1 Tahapan Dalam KDD

Tahapan proses KDD terdiri dari:

a. *Data Selection*

Pada proses ini dilakukan pemilihan himpunan data, menciptakan himpunan data target, atau memfokuskan pada subset *variable* (sampel data) dimana penemuan (*discovery*) akan dilakukan. Hasil seleksi disimpan dalam suatu berkas yang terpisah dari basis data operasional.

b. *Pre-Processing* dan *Cleaning Data*

Pre-Processing dan *Cleaning Data* dilakukan dengan membuang data yang tidak konsisten dan noise, duplikasi data, memperbaiki kesalahan data, dan bisa diperkaya dengan data eksternal yang relevan.

c. *Transformastion*

Proses ini *mentransformasikan* atau menggabungkan data ke dalam yang lebih tepat untuk melakukan proses *mining* dengan cara melakukan peringkasan (*agregasi*).

d. Data Mining

Proses *Data Mining* yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik, metode atau algoritma tertentu sesuai dengan tujuan dari proses KDD secara keseluruhan.

e. *Interpretation* / Evaluasi

Proses untuk menerjemahkan pola-pola yang dihasilkan dari *Data Mining*. Mengevaluasi (menguji) apakah pola atau informasi yang ditemukan bersesuai atau bertentangan dengan fakta atau hipotesa sebelumnya. Pengetahuan yang diperoleh dari pola-pola yang terbentuk dipresentasikan dalam bentuk visualisasi.

2.2.4 Association Rules

Association rule merupakan salah satu metode yang bertujuan mencari pola yang sering muncul di antara banyak transaksi, dimana setiap transaksi terdiri dari beberapa *item* sehingga metode ini akan mendukung sistem rekomendasi melalui penemuan pola antar *item* dalam transaksi-transaksi yang terjadi. (Fadlina, 2014).

Dalam Kusrini dkk.2009, metodologi dasar analisis asosiasi terbagi menjadi 2 tahap:

1. Analisa Pola Frekuensi Tinggi dengan Algoritma Apriori

Tahap ini digunakan untuk mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam basis data. Nilai *support* sebuah *item* diperoleh dengan persamaan (2.1) (Kusrini dan Luthfi, 2009)

$$\mathbf{Support(A) = \frac{\text{jumlah transaksi mengandung A}}{\text{total transaksi}} \times 100\%} \quad (2.1)$$

Sedangkan nilai support dari 2 item diperoleh dari persamaan (2.2) dan persamaan (2.3) (Kusrini dan Luthfi, 2009)

$$\mathbf{Support (A,B) = P (A \cap B)} \quad (2.2)$$

$$\mathbf{Support (A, B) = \frac{\text{jumlah transaksi mengandung A dan B}}{\text{total transaksi}} \times 100\%} \quad (2.3)$$

Rumus diatas menjelaskan bahwa untuk menentukan nilai *support* pada satu item, jumlah transaksi yang mengandung *item* A (satu *item*) dibagi dengan jumlah total transaksi.

2. Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat *minimum* untuk *confidence* dengan menghitung confidence aturan asosiatif $A \rightarrow B$. Nilai confidence dari aturan $A \rightarrow B$ diperoleh dengan persamaan (2.4) (Kusrini dan Luthfi, 2009).

$$\mathbf{Confidence P(B|A) = \frac{\Sigma \text{Transaksi Mengandung A dan B}}{\Sigma \text{Transaksi Mengandung A}} \times 100\%} \quad (2.4)$$

2.2.5 Algoritma Apriori

Algoritma apriori termasuk jenis aturan asosiasi pada data mining. Selain apriori, yang termasuk pada golongan ini adalah metode *Generalized Rule Induction* dan Algoritma *Hash Based*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut affinity atau market basket analysis (Kusrini dkk,2009).

Algoritma apriori merupakan salah satu metode untuk menggali kaidah asosiasi yang paling sederhana dan paling terkenal untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah pola-pola item di dalam suatu database yang memiliki frekuensi atau support di atas ambang batas tertentu yang disebut dengan istilah minimum support. Pola frekuensi tinggi ini digunakan untuk menyusun aturan asosiatif dan juga beberapa teknik data mining lainnya.

Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item.

Menurut Han, dkk (2001), terdapat dua ukuran kepercayaan yang menunjukkan kepastian dan tingkat kegunaan suatu rule yang ditemukan yaitu:

1. *Support*

Support (dukungan) merupakan suatu ukuran yang menunjukkan seberapa besar dominasi suatu *item* atau *itemset* dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu *item/itemset* layak untuk dicari *confidence*-nya (misal, dari keseluruhan transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan bahwa *item* A dan B dibeli bersamaan).

2. *Confidence*

Confidence (tingkat kepercayaan) adalah suatu ukuran yang menunjukkan hubungan antar *item* secara conditional (misal, seberapa sering *item* B dibeli jika orang membeli *item* A).

Pada umumnya *association rule* yang didapatkan menarik apabila *rule* atau aturan tersebut memenuhi baik *minimum support* maupun *minimum confidence* yang telah ditentukan oleh *user*.

Algoritma apriori dibagi menjadi beberapa tahap yang disebut narasi atau pass (Devi dinda setiawan, 2009).

1. Pembentukan kandidat *itemset*

Kandidat *k-itemset* dibentuk dari kombinasi $(k-1)$ – *itemset* yang didapat dari iterasi sebelumnya. Satu cara dari algoritma apriori adalah pemangkasan kandidat *k-itemset* yang subjeknya berisi $k-1$ *item* tidak termasuk dalam pola frekuensi tinggi dengan panjang $k-1$.

2. Penghitungan *support* dari tiap kandidat *k-itemset*

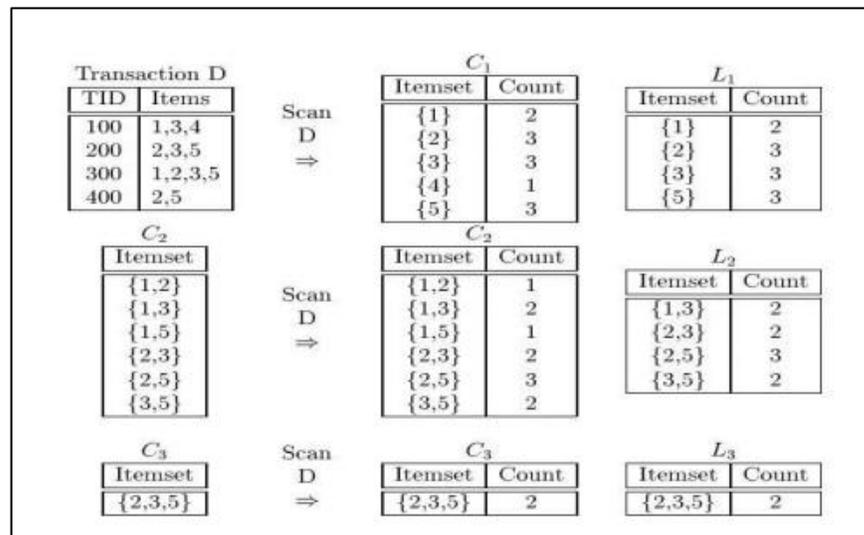
Support dari tiap kandidat *k-itemset* didapat dengan menscan database untuk menghitung jumlah transaksi yang memuat semua *item* didalam kandidat *k-itemset* tersebut. Ini adalah ciri dari algoritma apriori dimana diperlukan perhitungan dengan cara seluruh database sebanyak *k-itemset* terpanjang.

3. Tetapkan pola frekuensi tinggi

Pola frekuensi tinggi yang memuat k -item atau k -itemset ditetapkan dari kandidat k -itemset yang *support*-nya lebih besar dari minimum *support*.

4. Bila tidak didapat pola frekuensi tinggi maka seluruh proses dihentikan.

Berikut adalah contoh ilustrasi penerapan algoritma apriori seperti yang ditunjukkan gambar 2.2:



Sumber : [Silalahi,2016]

Gambar 2.2 Ilustrasi Penerapan Algoritma Apriori

2.2.6 RapidMiner

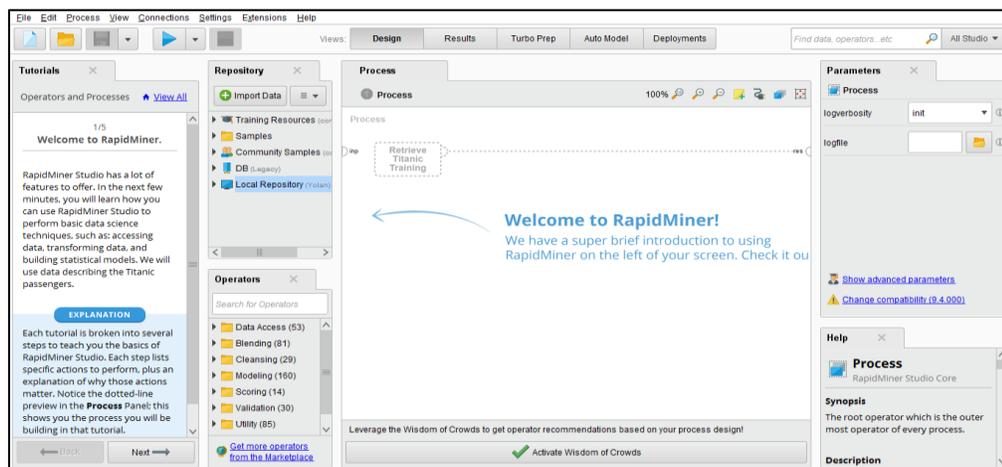
RapidMiner (YALE) adalah perangkat lunak *open source* untuk *knowledge discovery* dan data mining. RapidMiner memiliki kiran lebih 400 prosedur (operator) data mining, termasuk operator untuk *input*, *output*, *preprocessing* dan visualisasi.

Ribuan aplikasi data mining yang telah dikembangkan menggunakan RapidMiner banyak digunakan di dunia bisnis maupun penelitian.

Beberapa fitur dari RapidMiner:

1. Berlisensi gratis (open source).
2. Multiplatform karena diprogram dalam bahasa Java
3. Internal data berbasis XML sehingga memudahkan pertukaran data eksperimen.
4. Dilengkapi dengan *scripting language* untuk otomatisasi eksperimen
5. Memiliki GUI (*Graphical User Interface*), *command line mode (batch mode)*, dan Java API yang dapat dipanggil dari program lain.
6. Dapat dikembangkan dengan menambahkan plugin dan *ekstension*.
7. Fasilitas *plotting* untuk visualisasi data multidimensi dan model.

Berikut adalah tampilan aplikasi RapidMiner seperti yang ditunjukkan pada gambar 2.3:



Gambar 2.3 Tampilan Aplikasi RapidMiner

2.2.7 Perpustakaan

Perpustakaan menurut UU No.43 Tahun 2007, pasal 1, ayat 1 adalah institusi pengelola koleksi karya tulis, karya cetak, dan karya rekam secara profesional dengan sistem yang baku guna memenuhi kebutuhan pendidikan, penelitian, pelestarian, informasi, dan rekreasi bagi pemustaka.

Menurut Sulistyio Basuki perpustakaan adalah gedung atau ruangan yang berisi rak buku dan mengalami perubahan defenisi karena dipengaruhi oleh komponen koleksi dan perkembangannya muncul pengertian perpustakaan elektronik, multimedia, hibrida, perpustakaan maya (*virtual libarary*) dan perpustakaan tanpa tembok (*libraries without wall*).

Perpustakaan adalah fasilitas atau tempat yang menyediakan sarana bahan bacaan. Tujuan dari perpustakaan sendiri, khususnya perpustakaan perguruan tinggi adalah memberikan layanan informasi untuk kegiatan belajar, penelitian, dan pengabdian masyarakat dalam rangka melaksanakan Tri Dharma Perguruan Tinggi (Wiranto dkk,1997).

Dari pengertian diatas dapat disimpulkan bahwa perpustakaan adalah suatu institusi unit kerja yang menyimpan koleksi bahan pustaka secara sistematis dan mengelolanya dengan cara khusus sebagai sumber informasi dan dapat digunakan oleh pemakainya.