

BAB 2

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Banyaknya penelitian (*research*) menghasilkan berbagai pengetahuan atau pandangan tentang penerapan metode *apriori*. Beberapa hasil penelitian 3 (tiga) tahun terakhir yang relevan dengan penelitian ini, akan dipaparkan sebagai bahan rujukan. Penelitian mengenai metode *apriori* ini bukan kali pertama dilakukan, sebelumnya sudah pernah dilakukan penelitian yang mengimplementasikan metode *apriori*, diantaranya adalah sebagai berikut :

2.1.1 Penerapan Data Mining Menggunakan Association Rules Untuk Mendukung Strategi Pemasaran Calon Mahasiswa Baru (Nurjoko, Abdi Darmawan, 2015)[7]

Dalam penelitian ini mereka menggunakan algoritma apriori dengan menggunakan data calon mahasiswa dan data mahasiswa untuk mengetahui informasi mengenai sumber media promosi yang digunakan calon mahasiswa baru. Hal ini dapat dimanfaatkan oleh pihak manajemen untuk menyusun strategi pemasaran dalam meningkatkan jumlah mahasiswa baru pada tahun berikutnya melalui pemilihan media promosi yang tepat dengan teknik data mining. Informasi yang didapat berupa nilai *support* dan *confidence* dengan aturan tingkat nilai maksimum *confidence* sebesar 81,1 % untuk sumber media promosi dari koran dan tingkat nilai maksimum *confidence* sebesar 77,8 % untuk sumber media promosi dari brosur. Sehingga pemilihan media promosi yang tepat untuk menjaring calon mahasiswa baru adalah melalui media koran dan media brosur namun juga tidak mengabaikan media promosi yang lainnya[7].

2.1.1 Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan (Kennedi Tampubolon, Hoga Saragih, dan Bobby Reza, 2013)[8]

Penelitian ini membahas mengenai pentingnya stok barang di apotek dan barang mana saja yang menjadi prioritas yang harus tersedia supaya tidak terjadi kekosongan barang. Jika hal ini dibiarkan stok barang akan berkurang dan berpengaruh pada loyalitas pelanggan dan pendapatan apotek. Pemasok alat-alat kesehatan menjadi faktor yang mendorong kelancaran penyaluran obat kepada pelanggan sehingga pelayanan obat dapat berjalan dengan lancar. Dengan menerapkan algoritma apriori dapat mempercepat proses pembentukan pola kombinasi itemset hasil penjualan alat-alat kesehatan di apotek dan hasil yang didapat adalah dengan nilai *support* dan *confidence* tertinggi yaitu *stick asam urat - stick gula* dan *stick kolesterol - stick gula*[8].

2.1.2 Penggunaan Market Basket Analisis Untuk Menentukan Pola Kompetensi Mahasiswa (Arief Jananto, 2012)[9]

Dalam penelitian ini mereka menggunakan nilai akademik mahasiswa kemudian dikelompokkan ke dalam suatu kompetensi kemudian mengambil rata-rata nya dan menghasilkan pencapaian kompetensi dengan *minimum support* sebesar 70% dan *minimum confidence* sebesar 75% yang dilakukan pada tiga kompetensi yaitu *Integration System (SI)*, *Information System (IS)*, dan *Network and Communication (NC)* yang memiliki arti bahwa 70% calon lulusan program studi sistem informasi angkatan dari tahun 2004 sampai 2007 memiliki kompetensi dalam bidang sistem integration, sistem informasi, dan *network and communication* dibandingkan dengan kompetensi yang lain[9].

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
1.	Nurjoko, Abdi Darmawan 2015.	Menyusun strategi pemasaran dalam meningkatkan jumlah mahasiswa baru melalui pemilihan media promosi yang tepat dengan teknik data mining.	<i>Association Rules</i> dengan algoritma <i>apriori</i> .	Nilai <i>support</i> dan <i>confidence</i> dengan aturan tingkat nilai maksimum <i>confidence</i> sebesar 81,1 % untuk sumber media promosi dari koran dan tingkat nilai maksimum <i>confidence</i> sebesar 77,8 % untuk sumber media promosi dari brosur.
2.	Kennedi Tampubolon, Hoga Saragih, Bobby Reza.	Pentingnya stok barang dan barang mana saja yang menjadi prioritas yang harus tersedia agar tidak terjadi kekosongan barang.	<i>Association Rules</i> dengan algoritma <i>apriori</i> .	Penentuan pola kombinasi itemset hasil penjualan alat-alat kesehatan di apotek dan menghasilkan nilai <i>support</i> dan <i>confidence</i>

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
				tertinggi yaitu <i>stick asam urat - stick gula</i> dan <i>stick kolesterol - stick gula</i> .
3.	Arief Jananto, 2012.	Pengelompokkan nilai akademik mahasiswa ke dalam suatu kompetensi dan mengambil rata-rata nya.	<i>Association Rules</i> dengan algoritma <i>apriori</i> .	Pencapaian kompetensi dengan <i>minimum support</i> sebesar 70% dan <i>minimum confidence</i> sebesar 75% yang dilakukan pada tiga kompetensi yaitu <i>Integration System (SI), Information System (IS),</i> dan <i>Network and Communication (NC)</i> yang

Nama				
No	Peneliti dan Tahun	Masalah	Metode	Hasil
				memiliki arti bahwa 70% calon lulusan program studi sistem informasi angkatan dari tahun 2004 sampai 2007 memiliki kompetensi dalam bidang <i>integration system, information system, dan network and communication</i> dibandingkan dengan kompetensi yang lain.

2.2 Universitas Dian Nuswantoro (UDINUS)

Universitas Dian Nuswantoro (UDINUS) adalah perguruan tinggi swasta yang ada di Semarang, Indonesia, berdiri sejak tahun 1990. Rektor saat ini adalah Dr. Ir. Edi Noersasongko M. Kom. UDINUS berada di pusat kota Semarang, Jawa Tengah, Indonesia. Terletak di bundaran Tugu Muda, di Jl. Imam Bonjol 207 dan

Jl. Nakula I no. 5-11 Semarang. UDINUS merupakan salah satu universitas swasta terbaik di Semarang, hal ini dibuktikan dengan prestasi para mahasiswanya yang meraih banyak gelar kejuaraan baik di dalam negeri maupun mancanegara.

Jumlah mahasiswa di UDINUS lebih dari 10.000 mahasiswa yang berasal dari beragangkota di Indonesia bahkan mancanegara. Namun jumlah mahasiswa yang diterima dengan jumlah mahasiswa yang lulus tiap tahunnya sering tidak sebanding atau seimbang. Hanya beberapa persen saja lulusan mahasiswa yang menyelesaikan studinya selama 4 tahun. Hal ini dapat mengakibatkan kekalahan UDINUS dalam persaingan perguruan tinggi swasta terutama di wilayah Jawa Tengah. Oleh karena itu, penentuan kriteria profil calon mahasiswa yang lulus tepat waktu menjadi faktor yang sangat penting agar dapat menyaring calon mahasiswa baru yang berpotensi lulus tepat waktu. Dan dapat membantu dalam pengambilan keputusan yang berkaitan dengan strategi pemasaran.

2.3 Kelulusan Mahasiswa

Mahasiswa merupakan sumber daya utama pada setiap perguruan tinggi baik negeri maupun swasta. Mahasiswa juga memiliki peranan yang penting dalam evaluasi keberhasilan penyelenggaraan program studi di perguruan tinggi sedangkan kelulusan merupakan tahap keberhasilan setelah menyelesaikan proses belajar mengajar di perguruan tinggi. Kelulusan mahasiswa sangat berpengaruh pada kualitas perguruan tinggi dan penilaian akreditasi.

Mahasiswa yang dinyatakan lulus harus memenuhi persyaratan tertentu yang ditetapkan oleh program studi dan keputusan rektor. Berikut merupakan syarat kelulusan mahasiswa sistem informasi di UDINUS :

1. Mahasiswa sudah menyelesaikan minimal 144 sks dalam waktu 8 semester dan maksimal 160 sks dalam waktu 14 semester.
2. Mahasiswa mendapat nilai minimal C untuk mata kuliah metodologi penelitian.
3. IPK minimal 2.00.
4. Skor toefl minimal maksimal 25 %.

5. Jumlah nilai D maksimal 3 mata kuliah dan tidak ada nilai E.
6. Telah mengambil seluruh mata kuliah wajib sesuai kurikulum program studi.

2.4 Penerimaan Mahasiswa Baru

Tingginya minat masyarakat atau siswa yang telah lulus SMA untuk melanjutkan studinya ke perguruan tinggi menjadi faktor pemicu bagi setiap perguruan tinggi untuk menyaring calon mahasiswa baru sebanyak mungkin. Namun seringkali masyarakat mengalami kesulitan dalam memilih perguruan tinggi. Dan perguruan tinggi swasta terkadang menjadi pilihan kedua dibanding dengan perguruan tinggi negeri. Oleh karena itu, perguruan tinggi swasta harus meningkatkan mutu dan kualitasnya agar menjadi pilihan utama bagi masyarakat. UDINUS saat ini telah menjadi perguruan tinggi swasta terbaik di Jawa Tengah. Terbukti dengan lulusan mahasiswa nya yang meraih gelar kejuaraan baik nasional maupun internasional. Pencapaian tertinggi UDINUS adalah mendapat akreditasi A untuk program studi Teknologi Informatika S-1, Sistem Informasi S-1, dan Desain Komunikasi Visual S-1. UDINUS juga memiliki banyak kerja sama dengan sejumlah perguruan tinggi di Indonesia bahkan hingga ke luar negeri. Dengan keuntungan tersebut diharapkan dapat meningkatkan jumlah mahasiswa baru di UDINUS.

Bagi calon mahasiswa baru terdapat 3 jalur pendaftaran mahasiswa baru di UDINUS antara lain :

1. Jalur PMDK

Jalur PMDK adalah jalur penerimaan mahasiswa baru yang dilakukan tanpa melalui ujian untuk program diploma (D-IV) dan program sarjana (S-1). Jalur PMDK ini hanya didasarkan pada nilai dan sertifikat prestasi akademik dan non-akademik.

2. Jalur Reguler

Jalur reguler adalah jalur penerimaan mahasiswa baru yang dilakukan melalui ujian potensi akademik yang diselenggarakan secara mandiri atau kolektif untuk program sarjana tingkat (S-1), program kejuruan (D-III) dan program diploma IV.

3. Program Penyetaraan

Program penyetaraan bagi yang telah lulus program vokasi atau pertukaran mahasiswa.

2.5 Data Mining

Data mining merupakan proses mencari hubungan atau pola yang menyaring data dengan jumlah yang besar, data disimpan dalam database, dan data *warehouse*. Data mining berhubungan dengan ilmu di bidang yang lain seperti data *warehousing*, *database system*, *machine learning*, *statistik*, *information retrieval*, dan komputasi tingkat tinggi [10]. Proses pengolahan data membutuhkan teknik statistik, *matemathics*, *machine learning*, dan kecerdasan buatan untuk dapat menggali informasi-informasi yang dapat dimanfaatkan.

Data mining berhubungan dengan pencarian suatu informasi yang tersembunyi dan penggunaan perangkat lunak untuk penemuan pola. Beberapa faktor yang mendukung perlunya dilakukan data mining adalah :

1. Data telah mencapai ukuran atau jumlah yang sangat besar.

Proses data mining akan menghasilkan informasi yang akan mendasari dalam pengambilan keputusan sehingga tingkat kebenaran informasi tersebut menjadi sangat signifikan, semakin banyaknya data yang digunakan maka hasilnya akan semakin akurat. Perkembangan data dalam jumlah atau ukuran mengalami peningkatan yang semakin cepat, sehingga ukuran basis data di perusahaan bisa mencapai kisaran *gigabyte* atau bahkan *terabyte*.

2. Telah dilakukan proses data *warehousing*.

Dalam data mining terdapat sumber data yang berisi data gabungan dari beberapa departemen, daerah operasi atau dari sumber lain yaitu data kependudukan untuk mencapai hasil yang memuaskan. Oleh karena itu, proses *warehousing* sangat diperlukan untuk menjaga konsistensi, memberikan perspektif yang lebih baik terhadap data dan menjaga integritas data.

3. Kemampuan komputasi yang semakin terjangkau.

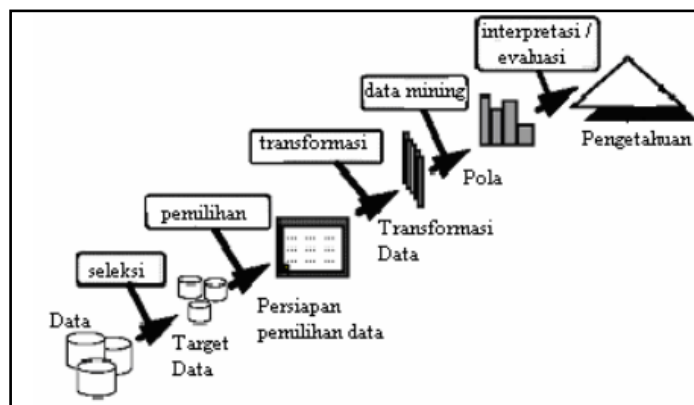
Proses komputasi membutuhkan sumber daya yang sangat besar. Penurunan harga yang terjadi pada perangkat keras komputer serta semakin meningkatnya kinerja pada perangkat komputer maupun teknologi pengolahan data.

4. Persaingan bisnis yang semakin ketat

Persaingan bisnis yang semakin membuat perusahaan harus memiliki inovasi untuk meningkatkan daya saingnya dipasar global.

2.5.1 Tahapan Proses Dalam Data Mining

Proses data mining memiliki beberapa tahapan yang berawal dari sumber data kemudian menghasilkan informasi yang didapat dari tahapan seperti pada gambar 2.1 berikut ini.



Gambar 2.1 Fase-fase dalam data mining

1. Data Selection

Pemilihan data dari sekumpulan data operasional perlu dilaksanakan sebelum tahap penggalian informasi dalam *Knowledge discovery in databases* (KDD) dilakukan. Proses pemilihan data akan menghasilkan data yang akan diproses dalam data mining, tersimpan dengan baik dalam suatu berkas, terpisah dari basis data operasional.

2. *Preprocessing / Cleaning*

Proses *cleaning* data perlu dilaksanakan sebelum proses data mining dilakukan. Proses *cleaning* berawal dari menghapus data yang sama, membuang data yang tidak konsisten dan memperbaiki data yang salah. Juga dilakukan proses enrichment yaitu proses “memperkaya” data melalui data atau informasi yang relevan dan diperlukan untuk KDD seperti data atau informasi eksternal.

3. *Transformation*

Koding merupakan proses transformasi data yang telah ditentukan sehingga data tersebut sesuai untuk proses data mining. Proses koding diperlukan untuk menemukan pola informasi dalam database.

4. Data Mining

Data mining adalah proses menemukan pola atau informasi yang tersembunyi dalam data menggunakan metode atau algoritma tertentu. Metode dalam data mining memiliki banyak variasi. Pemilihan metode yang sesuai berpengaruh pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation* atau *Evaluation*

Ketika data mining telah menghasilkan pola informasi sebaiknya ditampilkan dalam bentuk yang mudah dipahami oleh pihak yang bersangkutan. Tahap ini merupakan salah satu tahapan dari proses KDD atau *interpretation*. Tahapan ini berisi pemeriksaan pola atau informasi yang dihasilkan apabila tidak sesuai dengan fakta atau penelitian sebelumnya[11].

2.5.2 Peran Utama Dalam Data Mining

1. *Classification* (Klasifikasi)

Klasifikasi adalah proses menemukan model yang memiliki label/target/class dengan nilai nominal (kategorikal). Algoritma yang termasuk dalam klasifikasi yaitu *Naïve Bayes*, *C4.5*, *K-Nearest Neighbor*, *CART*, *ID3*, etc.

2. *Estimation* (Estimasi)

Algoritma estimasi hampir sama dengan algoritma klasifikasi bedanya adalah pada variabel label/target/class adalah bilangan numerik. Estimasi nilai dari variabel target berasal dari nilai atribut (variabel prediktor) Algoritma yang termasuk dalam estimasi yaitu *Neural Network, Linier Regression, Support Vector Machine*.

3. *Prediction* (Prediksi)

Prediksi hampir sama dengan estimasi yang memiliki target/class/label berupa bilangan numerik, perbedaannya adalah data pada prediksi menggunakan data *time series*. Algoritma yang termasuk prediksi/forecasting sama dengan algoritma dalam estimasi.

4. *Clustering* (Pengklasteran)

Teknik ini merupakan pengelompokan data tanpa didasarkan oleh class tertentu. Klastering berhubungan dengan sekumpulan data yang mempunyai kemiripan dengan data lain serta mempunyai perbedaan dengan data dari klaster lain. Klastering termasuk unsupervised learning karena tidak memiliki class/target/label. Algoritma yang termasuk dalam klastering yaitu K-Means dan Agglomerative Hierarchical Clustering (AHC)

5. *Assosiation* (Asosiasi)

Tujuan dari asosiasi adalah menemukan atribut atau kombinasi yang sering muncul secara bersamaan dalam satu transaksi. Algoritma ini biasa digunakan untuk *market basket analysis* artinya menemukan item atau produk mana yang dibeli secara bersamaan. Contoh algoritma yang termasuk *asociation rules* adalah *Apriori, FP-Growth, Hash Based Algorithm* dan *Generalized Rules Induction*[8].

2.5.3 Metode Data Mining

Dalam data mining terdapat metode yang terdiri 2 jenis yaitu :

1. *Supervised Learning*

Variabel pada *supervised learning* yang menjadi class/target/label ditentukan. Metode ini berisi proses belajar berdasarkan nilai dari variabel target yang diasosiasikan dengan nilai atribut (variabel predictor). Sebagian besar algoritma data mining seperti estimasi, prediksi, klasifikasi adalah *supervised learning*.

2. *Unsupervised Learning*

Variabel pada *unsupervised learning* yang menjadi class/target/label tidak ditentukan. Metode ini mencari pola dari semua variabel (atribut). Algoritma *clustering* merupakan algoritma *unsupervised learning*.

2.6 Association Rules

Association rules disebut juga frequent item set merupakan proses menemukan aturan asosiasi antara kombinasi item. *Association rules* adalah prosedur untuk mencari hubungan antar item dalam data set dan menampilkan dalam bentuk *association rules*[12]. Aturan asosiatif sangat memperhatikan *minimum support* dan *minimum confidence*. *Support* adalah nilai penunjang atau presentase kombinasi item yang ada pada basis data. *Confidence* adalah nilai kepastian atau kuatnya hubungan antar-*item* dalam sebuah apriori[13]. Sebagai contoh konsumen biasanya akan membeli kopi dan susu yang ditunjukkan sebagai berikut :

Kopi \rightarrow susu (*support* =2%, *confidence*=60%)

Contoh diatas dapat diartikan bahwa keseluruhan dari total transaksi konsumen tersebut sebanyak 2%. Sedangkan 60% transaksi yang menunjukkan bila konsumen membeli kopi dan pasti membeli susu sebesar 60%.

Aturan asosiasi dapat diartikan sebagai suatu proses untuk menemukan aturan asosiasi lain yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*).

Metodologi dasar analisis asosiasi dibagi ke dalam dua langkah :

1. Analisis pola frekuensi tinggi

Yang dilakukan pada langkah ini yaitu menemukan kombinasi item yang memenuhi syarat minimum *support* pada database. Rumus *support* adalah sebagai berikut :

$$Support (A) = (\text{jumlah transaksi A} / \text{Total transaksi}) \times 100\% \quad (2.1)$$

Sementara, nilai *support* dengan aturan $A \rightarrow B$ dapat diperoleh dengan rumus berikut.

$$Support (A, B) = P(A \wedge B)$$

$$Support (A, B) = \frac{\text{Total transaksi yang mengandung A dan B}}{\text{Total transaksi}} \times 100\% \quad (2.2)$$

2. Pembentukan aturan asosiasi

Untuk mencari aturan asosiasi *confidence* dapat dicari setelah semua aturan pola frekuensi munculnya sebuah item ditemukan. Rumus *confidence* adalah sebagai berikut :

Misalnya ditemukan aturan asosiatif $A \rightarrow B$ maka :

$$Confidence P(B|A) = \frac{\text{Total transaksi mengandung A dan B}}{\text{Total transaksi mengandung A}} \times 100\% \quad (2.3)$$

2.7 Algoritma Apriori

Apriori merupakan jenis algoritma yang termasuk dalam *association rules*. Algoritma *apriori* mengembangkan satu item menjadi dua item, tiga item dan seterusnya hingga ditemukan semua ukuran. Algoritma ini dapat diketahui dengan memperhatikan dua faktor penting yaitu minimum *support* atau *threshold* dan minimum *confidence*. *Threshold* dapat diartikan sebagai batas minimum dari suatu transaksi. Apabila jumlah transaksi tidak memenuhi *threshold* maka tidak diikutsertakan dalam perhitungan berikutnya. Untuk lebih memahami algoritma *apriori* terdapat contoh kasus pada tabel 2.2 dibawah ini.

Tabel 2.2 Tabel Transaksi

Transaksi	Item yang dibeli
1	teh, susu, gula
2	gula, teh, roti
3	gula, teh
4	roti, susu
5	roti, gula, susu
6	gula, teh
7	kopi, gula, susu
8	kopi, gula, susu
9	kopi, roti, susu
10	teh, gula, kopi

Tabel 2.2. adalah contoh transaksi dari database suatu supermarket. Tabel tersebut dapat direpresentasikan ke dalam database traksaksional seperti tabel 2.3 berikut.

Tabel 2.3 Perubahan data transaksi menjadi database transaksional

Transaksi	Item yang dibeli
1	Teh
1	Susu
1	Gula
2	Gula
2	Teh
2	Roti
3	Gula
3	Teh
4	Roti
4	Susu
5	Roti
5	Gula

5	Susu
6	Gula
6	Teh
7	Kopi
7	Gula
7	Susu
8	Kopi
8	Gula
8	Susu
9	Kopi
9	Roti
9	Susu
10	Teh
10	Gula
10	Kopi

Langkah selanjutnya membuat bentuk tabular untuk dilakukan perhitungan seperti tabel 2.4 dibawah ini.

Tabel 2.4 Tabulasi Data Transaksi

Transaksi	Teh	Gula	Kopi	Susu	Roti
1	1	1	0	1	0
2	1	1	0	0	1
3	1	1	0	0	0
4	0	0	0	1	1
5	0	1	0	1	1
6	1	1	0	0	0
7	0	1	1	1	0
8	0	1	1	1	0
9	0	0	1	1	1
10	1	1	1	0	0

Tabel berikut akan menunjukkan kombinasi *2-itemset* dari data transaksi pada tabel 2.2

Tabel 2.5 Kombinasi 2-itemset

Kombinasi	Jumlah
gula, teh	5
kopi, teh	1
susu, teh	1
roti, teh	1
kopi, gula	3
susu, gula	4
roti, gula	2
susu, kopi	3
roti, kopi	1
roti, susu	3

Dari tabel 2.5 diatas apabila ditentukan $\phi = 3$, maka:

$$F_2 = \{ \{ gula, teh \}, \{ kopi, gula \}, \{ susu, gula \}, \\ \{ susu, kopi \}, \{ roti, susu \} \}$$

Tabel 2.6 Kombinasi 3-itemset

Kombinasi	Jumlah
gula, teh, kopi	1
gula, teh, susu	1
susu, gula, kopi	2
susu, gula, roti	1
kopi, gula, roti	0
susu, kopi, roti	1

Hasil Kombinasi dari F_2 dapat digabungkan menjadi kombinasi 3-itemset. Kombinasi 3-itemset yang dapat dibentuk dari itemset F_2 dapat dilihat pada

tabel 2.6 diatas. Hasil kombinasi F3 yang terbentuk = {{susu, gula, kopi}}, karena hanya inilah kombinasi yang memiliki frekuensi kemunculan $\geq \phi$. Dari F3 yang telah ditemukan, maka persentase *confidence* yang dihasilkan dari aturan asosiasi dapat dilihat pada tabel 2.7 dibawah ini.

Tabel 2.7 Calon Aturan Asosiasi dari F3

Aturan	Confidence	
Jika membeli gula dan susu, maka akan membeli kopi	2/4	50%
Jika membeli gula dan kopi, maka akan membeli susu	2/3	67%
Jika membeli kopi dan susu, maka akan membeli gula	2/3	67%

Aturan yang dapat dihasilkan dari tabel 2.7 diatas yaitu aturan dengan 3 *antecedent* berikut ini.

Jika membeli gula dan kopi, maka akan membeli susu.

Jika membeli kopi dan susu, maka akan membeli gula.

Calon aturan asosiasi F2 dapat dilihat pada tabel berikut.

Tabel 2.8 Aturan Asosiasi

Aturan	Confidence	
Jika membeli teh, maka akan membeli gula	5/5	100%
Jika membeli gula, maka akan membeli the	5/8	62.50%
Jika membeli gula, maka akan membeli kopi	3/8	37.50%
Jika membeli kopi, maka akan membeli gula	3/4	75%
Jika membeli gula, maka akan membeli susu	4/8	50%

Jika membeli susu, maka akan membeli gula	4/6	67%
Jika membeli gula, maka akan membeli roti	2/8	25%
Jika membeli roti, maka akan membeli gula	2/4	50%
Jika membeli kopi, maka akan membeli susu	3/4	75%
Jika membeli susu, maka akan membeli kopi	3/6	50%
Jika membeli susu, maka akan membeli roti	3/6	50%
Jika membeli roti, maka akan membeli susu	3/4	75%

Aturan asosiasi final berurut yaitu *Support x Confidence* seperti dalam tabel 2.9 berikut.

Tabel 2.9 Aturan Asosiasi Final

Aturan	Support	Confidence	Support x Confidence
Jika membeli teh, maka akan membeli gula	50%	100%	50.0%
Jika membeli gula, maka akan membeli the	50%	62.50%	31.1%
Jika membeli susu, maka akan membeli gula	40%	67%	26.8%
Jika membeli kopi, maka akan membeli gula	30%	75%	22.5%
Jika membeli kopi, maka akan membeli susu	30%	75%	22.5%

Jika membeli roti, maka akan membeli susu	30%	75%	22.5%
Jika membeli gula dan kopi, maka akan membeli susu	20%	67%	13.4%
Jika membeli kopi dan susu, maka akan membeli gula	20%	67%	13.4%

Tabel 2.9 diatas menerangkan tentang persentase *support* dan *confidence* yang berasal dari kombinasi 2-itemsets dan kombinasi 3-itemsets. Hasil perkalian *support* dan *confidence* inilah yang menjadi hasil akhir dari algoritma *apriori* [14].

Algoritma *apriori* merupakan salah satu algoritma yang mudah dan sederhana untuk dieksekusi, maka dari itu algoritma ini sering sekali ditemukan dalam berbagai penelitian, selain sederhana algoritma ini juga mudah untuk dikembangkan di masa yang akan datang[15].

2.8 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.

Ketiga kriteria diatas harus seimbang dan sangat dibutuhkan karena belum tentu model yang akurat adalah handal, dan yang handal atau akurat belum tentu bermanfaat.

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. [16]

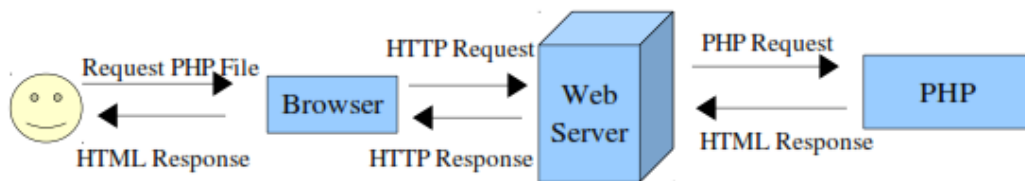
2.9 Hipertext Preprocessor (PHP)

PHP merupakan bahasa *script server-side* yang digunakan untuk membuat aplikasi web. *Server-side scripting* adalah kumpulan perintah atau sintaks yang akan dijalankan oleh server yang disertakan pada dokumen HTML sehingga keamanan dari halaman web terjamin.

PHP pada awalnya dikembangkan oleh Rasmus Lerdorf pada tahun 1995. Kemudian pada tahun 1998 fungsi-fungsi PHP mulai dikembangkan lagi oleh Zeev Suraski dan Andi Gutmants yang bernama PHP 3. Pada tahun 2002 performa PHP menjadi semakin meningkat menjadi versi 4 oleh Zend Engine dan yang terakhir menjadi PHP 5 dengan dukungan Zend Engine 2 [17].

2.9.1 Cara Kerja PHP

Cara kerja PHP yang akan dijelaskan adalah PHP sebagai bahasa pemrograman untuk mengembangkan aplikasi web. Karena selain digunakan untuk web programming PHP juga dapat digunakan untuk mengembangkan aplikasi berbasis desktop dan *Command Line Interface (CLI)*.



Gambar 2.2 Cara Kerja PHP

Dapat dilihat pada gambar 2.2 cara kerja PHP dapat dibagi menjadi enam proses yaitu :

1. User atau pengguna meminta halaman PHP.
2. Browser mengirim HTTP Request ke web server.
3. Web server mengirim permintaan file PHP tersebut ke PHP Processor.
4. PHP Processor akan memproses permintaan file PHP tersebut kemudian memberikan hasilnya ke web server.
5. Web server mengirimkan hasil tersebut ke browser dengan menambahkan HTTP header.
6. Browser memproses kiriman HTTP dan menampilkannya sebagai HTML kepada user atau pengguna[17].

2.9.2 Kelebihan PHP

PHP mempunyai banyak kelebihan diantaranya adalah :

1. Dapat menyesuaikan dengan beberapa database yang sering digunakan. Database yang didukung oleh PHP antara lain :
 - a. *MySQL*.
 - b. *Oracle*.
 - c. *PostgreSQL*.
 - d. *Sybase*.

2. PHP dapat digunakan dalam berbagai OS seperti *linux/unix, windows, macintosh*, maupun *solaris*.
3. PHP adalah *software open source* yang dapat di download tanpa harus mengeluarkan biaya.
4. Implementasi nya jauh lebih muda karena banyak developer yang membantu dalam pengembangan PHP.
5. Bahasa PHP adalah bahasa pemrograman yang paling mudah dan mempunyai banyak referensi.

2.10 Struktur Database

Database memiliki beberapa struktur yang terdiri dari :

1. Data

Kumpulan dari fakta mengenai objek tertentu yang diimplementasikan dalam bentuk seperti angka, gambar, huruf, suara, film, dll yang belum memiliki makna tertentu.
2. Informasi

Setelah data diolah maka menghasilkan informasi yang akurat dan sudah memiliki arti atau makna untuk mencapai tujuan tertentu.
3. Tabel

Adalah bagian paling dasar dalam database yang berhubungan dengan penyimpanan data.
4. *Field* (Kolom)

Adalah bagian dari tabel yang mengandung informasi yang lebih detail tentang sub judul tabel dalam sebuah item data. Berikut adalah ketentuan pembentukan *field* dalam tabel :

 - a. Item tersebut menarik atau unik.
 - b. Dapat disingkat.
 - c. Tidak diperbolehkan menggunakan spasi untuk pemisah *field* tetapi menggunakan tanda lambang seperti "_" Contoh: Kode produk menjadi Kdproduk, Kodeprdk, Kd_produk.

5. *Record*

Sekelompok informasi atau objek – objek tertentu.

2.10.1 Komponen Sistem Basis Data

1. *Hardware* (Perangkat Keras) yang terdiri dari :

- a. Komputer
- b. Media atau sarana komunikasi.
- c. Memori atau tempat penyimpanan sekunder online misalnya harddisk.
- d. Memori atau tempat penyimpanan sekunder offline misalnya removable disk untuk mem backup data.

2. *Operating System* (Sistem Operasi)

Sistem operasi adalah sistem yang mengontrol keseluruhan dari sumber daya yang ada pada komputer dan melakukan operasi-operasi dasar dalam computer. Misalnya : *MS-Windows 3.1, MS-DOS, MS-Windows 95/98/2000, MS-Windows NT, Novel-Netware, Unix* dan lain-lain.

3. *Database* (Basis Data)

Sistem basis data memiliki cakupan yang lebih luas dari basis data. Masing-masing basis data mempunyai sejumlah objek basis data. Selain menyimpan data, basis data juga berisi definisi struktur (baik untuk basis data maupun objek-objeknya secara detail).

4. *Database Management System* (Sistem Pengelolaan Basis Data)

Piranti yang termasuk dalam DBMS adalah *dBase III+, FoxBase dBase IV, Borland-Paradox* atau *Borland-Interbase, Ms Acces, MS-SQL-Server, Oracle, mySQL*.

5. *User* (Pengguna)

Ada 3 jenis user dalam sistem basis data antara lain :

- a. Orang yang memprogram aplikasi yaitu orang yang bertanggung jawab membuat program aplikasi menggunakan database.

- b. User akhir yaitu user yang melakukan interaksi dengan sistem basis data via online dengan menggunakan terminal atau *workstation*.
 - c. Administrator Basis Data yaitu orang yang mengatur strategi dan kebijakan data serta menyediakan kebutuhan dukungan teknik untuk menerapkan keputusan yang diambil.
6. Program Aplikasi.

Program Aplikasi biasanya bersifat optional sesuai kebutuhan user atau pengguna. DBMS yang digunakan lebih berperan untuk mengorganisasi data dalam database sementara untuk user atau pengguna akhir basis data disediakan program khusus untuk melakukan perubahan, pengisian, dan pengambilan data.

2.10.2 Language Database (Bahasa Basis Data)

Bahasa untuk berkomunikasi dengan sesama pengguna DBMS. Bahasa basis data terdiri dari sejumlah perintah yang digunakan untuk melaksanakan suatu pekerjaan tertentu. Bahasa basis data terdiri dari 3 jenis yaitu:

1. *DCL (Data Control Language)*.
2. *DDL (Data Definition Language)*.
3. *DML (Data Manipulation Language)*.