

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Wahyuningsih (2017) menerapkan algoritma *apriori* pada data transaksi penjualan di Aswalayan dengan data transaksi berjumlah 474 pada bulan Juli 2016. Selain itu penelitian yang dilakukan Wahyuningsih (2017) menghasilkan sebuah aplikasi.

Aplikasi yang dibuat oleh Wahyuningsih (2017) berguna untuk mengetahui pola pembelian konsumen untuk membantu pemilik swalayan dalam pengaturan *display* barang.

Penerapan algoritma juga dilakukan oleh Anas (2017) untuk mengetahui pola peminjaman buku di perpustakaan SMPN 24 Padang. Data yang digunakan terdiri dari atribut buku yang akan dipinjam, dalam hal ini terdapat 9 jenis buku sebagai sampel dan maksimal peminjamaan buku dalam 1 transaksi adalah 3. Untuk menguji kebenaran dari hasil pengolahan data, Anas (2017) menggunakan salah satu *software* aplikasi *Orange*.

Adapun hasil yang diperoleh dari penelitian yang dilakukan oleh Anas (2017) adalah sistem yang dibangun dapat membantu dalam menentukan pola pengunjung perpustakaan yang umumnya masih dilakukan secara manual selain itu *rule-rule* yang dihasilkan dapat digunakan sebagai referensi dalam pengadaan buku-buku koleksi di perpustakaan SMPN 24 Padang.

Nursikuagus dan Hartono (2016) menerapkan algoritma *apriori* pada data transaksi penjualan *adventure work* dengan 10 transaksi penjualan

sebagai sampel. Dalam penelitiannya Nursikuagus, dkk (2016) membuat aplikasi berbasis *web* dengan menggunakan PHP (*Personal Home Page*) sebagai bahasa pemrograman dan MySQL sebagai sistem manajemen basis data.

Adapun hasil yang diperoleh dari penelitian yang dilakukan oleh Nursikuagus dan Hartono (2016) yaitu aplikasi yang dibuat berhasil menunjukkan produk apa saja yang paling banyak terjual selain itu dapat membantu untuk mengadakan stok produk yang banyak disukai pembeli, dan menambah persediaan produk.

Tampubolon, dkk (2013) menerapkan 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.

Adapun hasil yang diperoleh dari penelitian yang dilakukan oleh Tampubolon, dkk (2013) 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.

Sementara itu, Gunadi dan Sensuse (2012) menerapkan dua algoritma yaitu algoritma *Apriori* dan algoritma *frequent pattern growth* (FPgrowth). Data yang digunakan merupakan data transaksi penjualan produk buku di Percetakan PT.Gramedia. Kedua algoritma ini digunakan untuk membentuk

frequent itemsets 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.

Adapun hasil yang diperoleh dari penelitian yang dilakukan oleh Gunadi dan Sensuse (2012) 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%.

Dalam tinjauan pustaka pada tabel 2.1 terdapat lima refrensi dan satu referensi dari penulis sebagai berikut:

Tabel 2.1 Perbandingan Metode Penelitian

Paramete Penulis	Objek	Metode	Hasil
Agus Nursikiagus, Tono Hartono (2016)	Produk Adventure work	<i>Algoritma Apriori</i>	Berhasil menunjukan produk apa saja yang sering dibeli.
Azwar Anas (2017)	Buku	<i>Algoritma Apriori</i>	Sistem yang dibangun dapat membantu dalam menentukan dalam menentukan pola pengunjung perpustakaan yang umumnya masih dilakukan secara manual.
Goldie Gunadi,	Buku	- <i>Algoritma Apriori</i>	Tingkat kekuatan aturan-aturan asosiasi yang dihasilkan oleh

Dana Indra Sensuse (2012)		- <i>Frequent Pattern Growth</i>	algoritma <i>apriori</i> lebih besar dibandingkan dengan yang dihasilkan oleh algoritma <i>FP-growth</i> . Dengan tingkat akurasi dari algoritma <i>apriori</i> terhadap <i>FP-growth</i> sebesar 257,4543%.
Kennedi Tampubolon, Hoga Saragih, Bobby Reza (2013)	Alat-alat Kesehatan	<i>Algoritma Apriori</i>	Penerapan Algoritma <i>apriori</i> pada teknik <i>data mining</i> sangat efisien dan dapat mempercepat proses pembentukan kecenderungan pola kombinasi itemset hasil penjualan alat-alat kesehatan.
Tri Wahyuningsih (2017)	Transaksi Penjualan Barang di Toko Aswalayan	<i>Algoritma Apriori</i>	Menghasilkan hubungan barang yang sering dibeli sehingga bisa membantu pemilik dalam <i>display</i> barang.
Robi Yanto, Riri Khoriah (2015)	Obat	<i>Algoritma Apriori</i>	Menghasilkan nilai yang sama sesuai dengan nilai <i>minimum support</i> dan <i>minimum confidence</i> yang telah ditentukan.

Tugas akhir yang akan diajukan menggunakan objek obat dengan metode Algoritma Apriori untuk menghasilkan sistem yang mampu membentuk pola kombinasi itemset penjualan obat.

2.1. Dasar Teori

2.2.1. Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*)

adalah proses pembentukan definisi – definisi konsep umum yang dilakukan dengan cara mengobservasi contoh – contoh spesifik dari konsep – konsep yang akan dipelajari (Hermawati, 2009).

Data mining, sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menentukan 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 trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola – pola inidikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan teliti, yang mungkin saja menggunakan perangkat pendukung keputusan lainnya (Hermawati, 2009).

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *data mining* adalah suatu teknik mengenali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui.

2.2.2. Association Rule

Association rule merupakan sebuah ekspresi implikasi $X \rightarrow Y$

berbentuk , dimana X dan Y merupakan disjoint *itemset* $(X \cap Y) = \emptyset$.

Dalam *association rule*, dapat menghitung *support* dan *confidence*.

Confidence menyatakan seberapa sering item – item dalam Y muncul dalam transaksi yang berisi X. Nilai *support* dan *confidence* dapat dinyatakan dalam rumus (2.1) dan (2.2) (Hermawati, 2013).

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (2.1)$$

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (2.2)$$

2.2.3. 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 disebut *affinity analysis* atau *market basket analysis* (Kursini dan Luthfi, 2009).

Algoritma *apriori* merupakan salah satu metode untuk mengenali 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*. Polafrekuensi 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.

2. *Confidence*

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

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

Dalam Kusri dan Luthfi, 2009. Metodologi dasar analisis asosiasi terbagi menjadi 2 tahap :

1. Analisis pola frekuensi tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat *minimum* dari nilai *support* dalam *database*. Nilai *support* sebuah *item* diperoleh dengan persamaan (2.3) (Kusri dan Luthfi, 2009)

$$\mathbf{Support (A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ transaksi} \times 100\%} \quad (2.3)$$

Sedangkan nilai *support* dari 2 *item* diperoleh dari persamaan (2.4) dan persamaan (2.5) (Kusri dan Luthfi, 2009).

$$\mathbf{Support (A,B) = P (A \cap B)} \quad (2.4)$$

$$\mathbf{Support (A \cap B) = \frac{\sum Transaksi\ mengandung\ A\ dan\ B}{\sum Transaksi} \times 100\%} \quad (2.5)$$

2. Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat *minimum* untuk *confidence* dengan menghitung *confidence* aturan asosiatif $A \rightarrow B$. Nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dengan persamaan (2.6) (Kusrini dan Luthfi. 2009).

$$\mathbf{Confidence} = P(B|A) = \frac{\text{Jumlah transaksi mengandung } A \text{ dan } B}{\text{Jumlah transaksi mengandung } A} \times 100\% \quad (2.6)$$

Algoritma *apriori* menurut Kusrini dan Luthfi. (2009) sebagai berikut:

1. Tentukan nilai minimum *frekuent itemset* yang menunjukkan *itemset* memiliki frekuensi kemuncukn lebih dari nilai minimum yang telah ditentukan (Φ) untuk mendapatkan kandidat 1 *itemset*.
2. Setelah mendapatkan hasil dari iterasi ke 1, maka akan dilanjutkan untuk mencari kandidat 2 *itemset* dan selanjutnya. Proses akan terhenti ketika tidak menentukan yang sesuai dengan nilai minimum yang telah ditentukan.
3. Dari kandidat *itemset* yang telah diperoleh kemudian dihitung nilai *support* dan *confidence*-nya, dengan syarat harus memenuhi nilai minimum *support* dan *confidence* yang telah ditentukan.
4. Diperoleh *rules* yang dapat digunakan sebagai informasi oleh pengguna.

2.2.4. Database

Menurut Sutarman (2012), *Database* sekumpulan *file* yang saling berhubungan dan terorganisasi atau kumpulan *record-record* yang menyimpan data dan hubungan diantaranya.

2.2.4. PHP (*Hypertext Preprocessor*)

PHP merupakan suatu bahasa *scripting* khususnya digunakan untuk *web development*. Karena sifatnya yang *server side scripting*, maka untuk menjalankan PHP harus menggunakan *web server*. PHP dapat melakukan pengambilan data, pengumpulan data dari *database*, menerima dan mengirim *cookie*. (Hidayatullah, 2014).

2.2.5. HTML (*Hypertext Markup Language*)

Shelly dan Velmaart (2011) menyatakan bahwa HTML (*Hypertext Markup Language*) adalah bahasa format khusus yang *programmer* gunakan untuk memformat dokumen untuk ditampilkan di *web*. Menurut Chaffey (2009) HTML adalah halaman standar *web* presentasi dengan menggunakan format untuk menentukan pesan dan tata letak halaman *web*. Berdasarkan pendapat para ahli di atas, penulis menyimpulkan bahwa HTML adalah bahasa pemrograman yang digunakan memformat dokumen untuk ditampilkan di *web*.

2.2.6. MySQL

MySQL merupakan database *multiuser* yang menggunakan bahasa *Structure Query Language* (SQL). MySQL dalam operasi *client-server*, melibatkan *server-daemon* MySQL di sisi *server* dan berbagai macam program, serta *library* yang berjalan disisi *client*. MySQL mampu menangani data yang cukup besar (Wirawan, 2009).