

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Agus Nursikuwagus dan Tono Hartono (2016) membahas tentang algoritma apriori untuk analisis penjualan dengan berbasis web. Hasil dari penelitian ini yaitu analisa penjualan yang berbasis web menggunakan algoritma apriori.

Maharani (2017) membahas tentang implementasi *data mining* untuk pengaturan *layout* minimarket dengan menerapkan *association rule*. Hasil dari penelitian ini membahas tentang rule kombinasi produk dengan penyusunan produk.

Dewi Listriani, Anif Hanifa Setyaningrum dan Fenty Eka M.A (2016) membahas tentang penerapan metode asosiasi menggunakan algoritma apriori pada aplikasi analisa pola belanja konsumen. Hasil dari penelitian ini untuk pencarian kombinasi item terbanyak berdasarkan data transaksi dan kemudian membentuk pola asosiasi dari kombinasi item tersebut.

Tri Wahyuningsih (2017) membahas tentang implementasi *data mining* pada transaksi penjualan untuk mengetahui pola pembelian menggunakan algoritma apriori. Hasil dari penelitian ini diperoleh hubungan barang yang sering dibeli sehingga bisa membantu pemilik dalam mengatur display barang.

Riris Unna Elvia (2015) membahas tentang toko *online* Riris dengan menggunakan metode apriori untuk pemilihan jenis bunga sesuai keinginan *customer*. Hasil dari penelitian ini untuk kombinasi produk dengan menghitung 3 *item support*.

Tabel 2.1 Tinjauan Pustaka

No.	Nama Pengarang	Judul	Objek	Hasil
1	Agus Nursikuwagus dan Tono Hartono (2016)	Implementasi Algoritma Apriori Untuk Analisis Penjualan Dengan Berbasis Web	Teori	Analisa Penjualan yang Berbasis Web dengan Menggunakan Metode Algoritma Apriori
2	Maharani (2017)	Implementasi Data Mining Untuk Pengaturan Layout Minimarket Dengan Menerapkan Association Rule	Minimarket	Mengetahui Rule Kombinasi Produk Untuk Penyusunan Layout
3	Dewi Listriani, Anif Hanifa Setyaningrum, Fenty Eka M.A (2016)	Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Analisa Pola Belanja Konsumen	Buku	Mencari kombinasi item terbanyak berdasarkan data transaksi dan kemudian membentuk pola asosiasi dari kombinasi item tersebut
4	Tri Wahyuningsih (2017)	Implementasi Data Mining Pada Transaksi Penjualan untuk Mengetahui Pola Pembelian	Swalayan	Menghasilkan hubungan barang yang sering dibeli sehingga bisa membantu pemilik

No.	Nama Pengarang	Judul	Objek	Hasil
		Menggunakan Algoritma Apriori		dalam mengatur display barang
5	Riris Unna Elvia (2015)	Toko Online Riris dengan Menggunakan Metode Apriori Untuk Pemilihan Jenis Bunga Sesuai Keinginan Customer	Toko Bunga Online	Mencari kombinasi produk dengan menghitung 3item support
6	Usulan Noki Sugiyono (2019)	Implementasi Data Mining untuk Penetapan Strategi Penjualan Menggunakan <i>Association Rule</i> di Digizone Clothing Yogyakarta	Distro Digizone Clothing	Penetapan strategi penjualan menggunakan metode <i>Association Rule</i>

2.2. Dasar teori

2.2.1. Data Mining

Data Mining adalah proses yang memperkerjakan lebih teknik pembelajaran komputer *machine learning* untuk menganalisis dan mengekstraksi pengetahuan atau *knowledge* secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi atau (*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, knowledge discovery in database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data ukuran besar ini memperbaiki 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 berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya atau (Hermawati, 2009).

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

2.2.2. Metode dalam *Data Mining*

Menurut Prasetyo (2014), *Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yakni sebagai berikut.

a. Model Prediksi (*Prediction Modelling*)

Pekerjaan ini berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada

himpunan baru yang didapat. Ada dua jenis model prediksi, yaitu klasifikasi dan regresi. klasifikasi digunakan untuk variabel target diskrit sedangkan regresi digunakan untuk variabel target kontinyu (Prasetyo, 2014).

b. Analisis Cluster (*Cluster Analysis*)

Analisis kelompok melakukan pengelompokan data kedalam sejumlah kelompok berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada. data-data yang masuk dalam batas kesamaan dengan kelompoknya akan bergabung dalam kelompok tersebut, dan akan terpisah dalam kelompok yang berbeda jika keluar dari batas kesamaan kelompok tersebut (Prasetyo, 2014).

c. Analisis Asosiasi (*Association Analysis*)

Analisis asosiasi digunakan untuk menemukan pola yang menggambarkan kekuatan hubungan fitur dalam data. pola yang ditemukan biasanya mempresentasikan bentuk aturan implikasi atau subset fitur. Tujuannya adalah untuk menemukan pola yang menarik dengan cara yang efisien (Prasetyo, 2014).

d. Deteksi Anomali (*Anomaly Detection*)

Pekerjaan deteksi anomali berkaitan dengan pengamatan sebuah data dari sejumlah data yang secara signifikan mempunyai karakteristik yang berbeda dari sisa data yang lain. data-data yang karakteristiknya menyimpang atau berbeda dari data yang lain disebut sebagai outlier.

algoritma deteksi anomali yang baik harus mempunyai laju deteksi yang tinggi dan laju kesalahan yang rendah (Prasetyo, 2014).

2.2.3. Tahap – Tahap dalam *Data mining*

Dalam *data mining* terdapat beberapa tahapan, sebagai berikut (Kusrini, 2009) :

a. Pembersihan data (*Data Cleaning*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh dari basis data suatu perusahaan memiliki isian isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa *data mining* yang dimiliki.

b. Integrasi data (*Data Integration*)

Integrasi data merupakan penggabungan data dari berbagai basis data ke dalam suatu basis data baru. Data yang diperlukan untuk *data mining* tidak hanya berasal dari satu basis data juga berasal dari beberapa basis data. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya nya.

c. Seleksi Data (*Data Selection*)

Data yang ada pada basis data seringkali tidak semuanya di pakai, oleh karena itu hanya data yang sesuai untuk proses analisis yang akan diambil dari basis data. Sebagai contoh, sebuah kasus yang meneliti faktor

kecenderungan orang membeli dalam kasus analisis keranjang belanja tidak perlu mengambil nama pelanggan cukup dengan ID pelanggan.

d. Transformasi Data (*Data Transformation*)

Data diubah atau digabung kedalam format yang sesuai untuk diproses dalam *data mining*. Beberapa data membutuhkan yang diaplikasikan

e. Proses *Mining*

Proses *mining* ini merupakan proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

f. Evaluasi Pola (*Pattern Evaluation*)

Evaluasi pola bertujuan untuk menemukan pola-pola menarik ke dalam basis pengetahuan yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesis yang ada memang tercapai.

g. Presentasi Pengetahuan (*Knowledge Presentation*)

Presentasi pengetahuan merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna

2.2.4. 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 analisis atau *market basket analysis*. (Kusrini, 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* diatas ambang batas tertentu yang disebut dengan istilah *minimum support*. Pola frekuensi ini di gunakan menyusun dan lain 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 :

a. *Support*

Support (dukungan) merupakan suatu ukuran yang menunjukkan seberapa besar dominasi suatu *item* atau *itemset* dari keseluruhan transaksi.

b. *Confidence*

Confidence (tingkat kepercayaan) suatu ukuran yang menunjukkan hubungan antara item secara kondisional (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 Kusrini (2009), metodologi dasar analisis asosiasi terbagi menjadi dua 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 mengikuti persamaan an (2.1) (Kusrini, 2009).

$$Support (A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ transaksi} \times 100\% \quad (2.1)$$

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

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

$$Support (A \cap B) = \frac{\sum Transaksi\ mengandung\ A\ dan\ B}{\sum Transaksi} \times 100\% \quad (2.3)$$

2. Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, berulah dicari aturan asosiatif yang memenuhi syarat minimum untuk *confidence* dengan

menghitung *confidence* atau aturan asosiatif a sampai B. Nilai *confidence* dari aturan a sampai B diperoleh dengan persamaan (2.4)(Kusrini, 2009).

$$Confidence=P(B|A)=\frac{Jumlah\ transaksi\ mengandung\ A\ dan\ B}{Jumlah\ transaksi\ mengandung\ A} \times 100\% \quad (2.4)$$

Algoritma apriori menurut Kusrini (2009) sebagai berikut :

1. Tentukan nilai *minimum frequent itemset* yang menunjukkan item set memiliki frekuensi kemunculan 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 menemukan 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.