

## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Pada halaman ini akan membahas beberapa sumber pustaka yang terhubung dengan kasus yang akan diteliti.

Penelitian yang dilakukan Tri Wahyuningsih (2017) dengan judul “Implementasi Data *Mining* Pada Data Transaksi Penjualan Untuk Mengetahui Pola Pembelian Menggunakan Algoritma Apriori”. Penelitian ini dilakukan dengan tujuan mengetahui pola pembelian konsumen untuk membantu pemilik swalayan dalam pengaturan *display* barang. Data yang digunakan adalah data transaksi penjualan dari toko Aswalayan dengan jumlah 474 data pada bulan Juli 2016.

Irsyad Djamaludin dan Agus Nursikuwagus (2017) menerapkan algoritma apriori untuk mendapatkan pola penjualan dan pembelian suatu produk pada perusahaan rumahan *Roseberry* yang bergerak dibidang industry makanan. Penentuan itemset yang digunakan adalah berjumlah 1,2, dan 3 item produk. Sedangkan untuk minimum *support* yang digunakan adalah 30% dari jumlah transaksi yang ada. Untuk asosiasi antar produk ditentukan dengan *confidence* sebesar 70%. Hasil dari penelitian ini adalah perangkat lunak analisis pola penjualan dan pembelian konsumen dengan Algoritma Apriori. Sedangkan keputusan yang diperoleh adalah asosiasi antar produk dengan menggunakan nilai *confidence*. Untuk final *Rule*

*Association* adalah produk bolu *cake* meses dan bolu *cake* keju dengan nilai *confidence* sebesar 84,62%.

Erma Delima Sikumbang (2018) menerapkan algoritma apriori untuk pada penjualan sepatu untuk mengetahui pola kombinasi item dan itemset frekuensi tinggi sehingga dapat dijadikan sebagai faktor pengambilan keputusan dalam memprediksi persediaan barang sepatu. Berdasarkan penelitian yang telah dilakukan dengan sumber data tahun 2016, hasil penelitian menunjukkan bahwa sepatu yang paling banyak diminati adalah *New Balance* (91,67%), *Adidas* (75%), *Geox* (50%), *Nike* (41,67%), dan *Palladium* (41,67%).

M. Afdal dan Muhammad Rosadi (2019) menerapkan algoritma apriori untuk analisis penempatan tata letak buku di perpustakaan Soeman Hs yang memiliki 60.583 judul dan 335.745 buku. Data yang digunakan peneliti adalah 11.550 transaksi peminjaman buku selama 3 tahun yang telah diproses menghasilkan 4 *rules* dengan kombinasi item terbesar adalah kategori buku agama dan ilmu sosial sering dipinjam secara bersamaan dengan nilai *support* 11,71% dan *confidence* 41,43%. Selain itu, kategori buku teknologi dan ilmu sosial sering dipinjam secara bersamaan dengan nilai *support* 13,8% dan *confidence* 40,75%.

Ramadani Saputra (2020) menerapkan algoritma apriori untuk menentukan hubungan antar item dalam satu dataset (sekumpulan data) yang telah ditentukan, dengan menggunakan teknik *association rule* dapat mencari kemungkinan kombinasi yang sering muncul dari suatu itemset.

Data yang digunakan peneliti adalah data Apotik Pusaka Arta dengan menggunakan nilai minimum *support* sebesar 20% dan minimum *confidence* 50%, aturan asosiasi yang didapatkan sebanyak 7 kombinasi.

Tabel tinjauan pustaka merupakan tabel yang dibuat untuk mendefinisikan penelitian yang sebelumnya hampir sama dilakukan dengan penelitian yang diajukan saat ini, adapun perbandingan yang menjadi tabel tinjauan pustaka penelitian yakni :

**Tabel 2.1 Tinjauan Pustaka**

No	Penulis	Objek	Metode	Hasil
1	Tri Wahyuningsih (2017)	Aswalaya	Algoritma Apriori	Menghasilkan hubungan barang yang sering dibeli sehingga bisa membantu pemilik dalam mengatur <i>display</i> barang.
2	Irsyad Djamaludin dan Agus Nursikuwagus (2017)	<i>Rosebarry</i>	Algoritma Apriori	Menghasilkan perangkat lunak analisis pola penjualan dan pembelian konsumen dengan

				keputusan yang diperoleh adalah asosiasi antar produk dengan menggunakan nilai <i>confidence</i> .
3	Erma Delima Sikumbang (2018)	Data Penjualan Sepatu	Algoritma Apriori	Mengetahui hubungan frekuensi sepatu yang paling diminati oleh konsumen.
4	M. Afdal dan Muhammad Rosadi (2019)	Perpustakaan Soeman HS	Algoritma Apriori	Mengetahui buku yang sering dipinjam secara bersamaan oleh pengunjung.
5	Ramadani Saputra dan Alexander J.P.Sibarani (2020)	Apotik Pusaka Arta	Algoritma Apriori	Menghasilkan aplikasi yang dapat mengetahui jenis obat yang dibeli konsumen secara bersamaan.

6	Muhammad Afrizal (2021)	Toko Harapan Kita	Algoritma Apriori	Menghasilkan perangkat lunak analisis pola penjualan dengan keputusan yang diperoleh adalah asosiasi antar produk dengan menggunakan nilai <i>support</i> dan <i>confidence</i> .
---	-------------------------------	-------------------------	----------------------	---

## 2.2 Dasar Teori

### 2.2.1 Data Mining

Data *mining* adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, data *warehouse*, atau penyimpanan informasi lainnya. Data *mining* berkaitan dengan bidang ilmu – ilmu lain, seperti *database* sistem, data *warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, data *mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, spatial data analisis, *image database*, *signal processing* (Han, 2006).

Karakteristik data *mining* sebagai berikut

- Data *mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- Data *mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
- Data *mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi (Davies, 2004).

Tahap-tahap data mining ada 7 yaitu :

1. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Pada umumnya 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 atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data *mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

## 2. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk data *mining* tidak hanya berasal dari satu *database* tetapi juga berasal dari beberapa *database* atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

## 3. Seleksi Data (*Data Selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

#### 4. Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data *mining*. Beberapa metode data *mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

#### 5. Proses *mining*

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

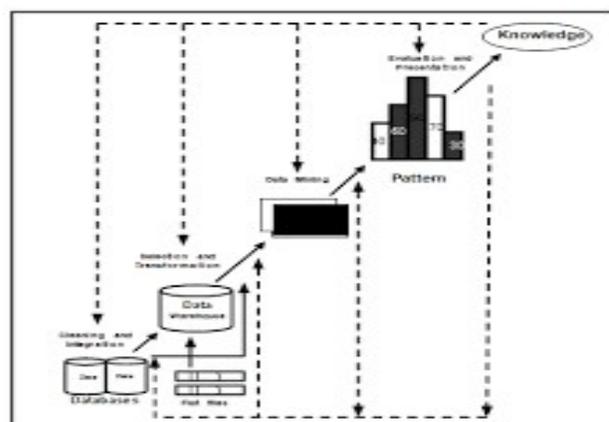
#### 6. Evaluasi pola (*pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data *mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data *mining*, mencoba metode data mining lain yang lebih sesuai,

atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

#### 7. Presentasi pengetahuan (*knowledge presentation*)

Presentasi pengetahuan (*knowledge presentation*), Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data *mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data *mining*. Karenanya presentasi hasil data *mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data *mining*. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data *mining* (Han, 2006).



**Gambar 2. 1 Ilustrasi Tahap Data Mining**

### 2.2.2 Algoritma Apriori

Algoritma apriori merupakan algoritma klasik pada data *mining*.

Algoritma apriori digunakan supaya computer dapat Algoritma apriori adalah algoritma yang paling terkenal untuk menemukan pola frekuensi tinggi. 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 subsetnya berisi  $k-1$  item tidak termasuk dalam pola frekuensi tinggi dengan panjang  $k-1$ .
2. Perhitungan *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 juga 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 kitemset ditetapkan dari kandidat  $k$ -itemset yang *support*nya lebih besar dari minimum *support*.
4. Bila tidak dapat pola frekuensi tinggi baru maka seluruh proses diberhentikan.

Algoritma apriori adalah satu algoritma dasar yang diusulkan oleh Agrawal dan Srikant pada tahun 1994 untuk menemukan *frequent*

*itemset* pada aturan asosiasi *Boolean*. Inde utama pada algoritma apriori adalah : pertama, mencari *frequent itemset* (himpunan item-item yang memenuhi minimum *support*) dari basis data transaksi. Kedua menghilangkan *itemset* dengan frekuensi yang rendah berdasarkan level minimum *support* yang telah ditentukan sebelumnya. Selanjutnya membangun aturan asosiasi dari *itemset* yang memenuhi nilai minimum *confidence* dalam basis data (Agrwal dan Srikant dalam jurnal Dewi ddk, 2016 : 122).

Algoritma apriori menggunakan *knowledge* mengenai *frequent itemset* yang telah diketahui sebelumnya, untuk memproses informasi selanjutnya. Pola frekuensi tinggi dalam pola-pola item dalam *database* yang memiliki frekuensi atau *support* di atas ambang batas tertentu yang disebut istilah minimum *support*. Pola frekuensi tinggi ini digunakan untuk menyusun aturan asosiasitif dan juga beberapa teknik data mining lainnya (Azwar Anas, 2016 : 632).

### **2.2.3 Teknik Asosiasi**

Pada penelitian ini teknik yang digunakan adalah teknik asosiasi, teknik asosiasi merupakan suatu metode penambangan data untuk mencari pola asosiasi yang sering muncul dalam data. Metode ini *popular* karena sering digunakan untuk menganalisis keranjang belanja. Aturan yang terbentuk akan berupa “jika maka” dalam suatu transaksi. Bisa digambarkan jika membeli barang A maka membeli barang B terjadi dalam suatu transaksi (Kusrini, 2009).

Algoritma yang digunakan dalam teknik asosiasi antara lain :

1. Apriori

Algoritma ini mencari *frequent itemset* dari *database* transaksi melalui beberapa tahap iterasi.

2. FP Growth

Algoritma ini berbeda dengan apriori, tidak perlu melakukan pencarian kandidat. Data direpresentasikan menggunakan *frequent Pattern Tree* melalui pendekatan *divide and conquer* untuk mendapat *frequent itemset*.

#### 2.2.4 Lift Ratio

Salah satu cara yang lebih baik untuk melihat kuat tidaknya aturan asosiasi adalah dengan menghitung *lift ratio*. Cara kerja metode ini adalah membagi *confidence* dengan *expected confidence*. *Confidence* dapat dihitung dengan rumus :

$$Confidence = \frac{\sum \text{Transaksi Mengandung Antesden dan Konsekuen}}{\sum \text{Transaksi Mengandung Antesden}}$$

Sedangkan nilai dari *expected confidence* dapat dihitung dengan rumus :

$$Expected Confidence = \frac{\sum \text{Transaksi Mengandung Konsekuen}}{\sum \text{Transaksi}}$$

*Lift ratio* dapat dihitung dengan cara membandingkan antara *confidence* untuk suatu aturan dibagi dengan *expected confidence*.

Berikut rumus dari *lift ratio* :

$$Lift Ratio = \frac{Confidence}{Expected Confidence}$$

Nilai *lift ratio* lebih besar dari 1 menunjukkan adanya manfaat dari aturan tersebut. Lebih tinggi nilai *lift ratio*, lebih besar kekuatan asosiasinya (Santosa, 2007). Jika nilai *lift ratio*  $< 1$  maka kemunculan A akan berkorelasi *negative* dengan kemunculan B, artinya kemunculan salah satu item mempengaruhi hal yang sebaliknya pada kemunculan item lainnya. Contoh dikorelasi *negative* adalah jika penjualan item A naik maka mempengaruhi jumlah penjualan B menjadi menurun. Jika didapatkan *lift ratio*  $> 1$  maka kemunculan A berkorelasi *positive* dengan kemunculan B, artinya kemunculan A ini berhubungan dengan kemunculan B. contoh korelasi *positive* adalah jika item A dibeli maka item B juga akan dibeli. Sedangkan jika *lift ratio* = 1 maka kemunculan A dan B *independent* dan tidak ada korelasi diantara kedua item tersebut. (Hanetla,2006).