

BAB 2

TINJAUAN PUSTAKA

2.1 Tinjauan Terhadap Penelitian Sebelumnya

Dalam tinjauan pustaka, pertama-tama peneliti akan mengawali dengan mendeskripsikan penelitian terdahulu yang berkaitan dan relevan dengan penelitian yang akan dilakukan peneliti. Dengan demikian, peneliti memberi gambaran awal mengenai penelitian sebelumnya terkait permasalahan dalam penelitian sebelumnya ini.

Tabel 2.1 Penelitian Sebelumnya

Judul Penelitian	Nama Peneliti	Metode yang Digunakan	Hasil Penelitian	Perbedaan dengan Penelitian ini
Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk bagi Pelanggan	Ariefan a Ria Riszky dan Mujiono Sadikin	Data Mining, metode aturan asosiasi (association rule) dengan algoritma Apriori	Aturan asosiasi yang dapat digunakan sebagai acuan untuk merekomendasikan produk dengan membuat <i>bundling product</i>	Penelitian sebelumnya itu menerapkan metode aturan asosiasi dengan algoritma Apriori. Sedangkan penelitian ini bertujuan untuk membandingkan algoritma Apriori pada penelitian sebelumnya dengan algoritma lain

				seperti FP-Growth dan CT-Pro untuk diketahui algoritma mana yang paling optimal dari segi kekuatan aturan asosiasi untuk kasus perekomendasi produk ini.
--	--	--	--	--

2.2 Landasan Teori

Disini membahas mengenai teori-teori apa saja yang digunakan dalam membuat penelitian ini. Landasan teori yang diuraikan merupakan hasil studi literatur dari berbagai literatur, jurnal, paper, dan bacaan-bacaan lain yang ada kaitannya dengan judul penelitian.

2.2.1 Data Mining

Data mining adalah proses pencarian suatu informasi penting yang tersembunyi dengan menggali pola informasi dari data yang besar dengan menggunakan metode tertentu, dimana informasi tersebut digunakan untuk hal yang berguna dan mempermudah dalam pengambilan keputusan pengguna [6], [8]. Data mining juga sering digunakan untuk berbagai keperluan seperti untuk mengklasifikasi, mengklastering, mengestimasi, memprediksi, menemukan pola pada suatu data, dan lain-lain [5].

Data mining juga dikenal dengan istilah *pattern recognition* yang merupakan algoritma pengolahan data untuk menemukan suatu pola yang tersembunyi di dalam sebuah data yang besar. Teknik data mining ini dapat menghasilkan suatu pengetahuan baru dari hasil data yang telah diolah yang berfungsi untuk pengambilan suatu keputusan.

2.2.2 Metode Data Mining

Data mining memiliki beberapa metode yang biasa digunakan, yaitu :

1. Klasifikasi (Classification)

Klasifikasi adalah metode dengan cara menilai objek data dan memasukkannya kedalam kelas-kelas tertentu yang sudah disediakan untuk menemukan sebuah model dengan tujuan untuk dapat memperkirakan kelas dari suatu objek data yang baru.

2. Klastering (Clustering)

Klastering adalah metode dengan cara membagi data dalam satu himpunan menjadi beberapa kelompok (cluster), dimana data yang berada dalam satu kelompok memiliki kesamaan yang lebih besar dibandingkan dengan yang berbeda kelompok [9].

3. Aturan Asosiasi (Association Rules)

Aturan asosiasi adalah metode dengan cara menemukan aturan asosiatif antara kombinasi item dalam sebuah data [10]. Kombinasi-kombinasi item tersebut nanti akan diproses hingga mendapatkan kombinasi-kombinasi item yang sesuai dengan ketentuan yang nantinya akan digunakan sebagai aturan asosiatif.

2.2.3 Aturan Asosiasi (Association Rules)

Aturan asosiasi atau Association Rules adalah salah satu metode data mining untuk menemukan aturan asosiatif yang ada pada kombinasi item dalam sebuah data besar [11]. Aturan asosiasi ini menggunakan data latih untuk menghasilkan suatu pengetahuan dari data tersebut [12]. Pengetahuan tersebut berisi item-item mana saja yang sering muncul secara bersamaan dalam satu waktu. Aturan asosiasi ini menghasilkan suatu pengetahuan yang berbentuk “if...then...” atau “jika...maka...”.

Support (nilai penunjang) adalah persentase kombinasi itemset yang ada di dalam data [13]. Nilai support ini dilambangkan dengan $\text{Support}(A,B)$ “ $A \Rightarrow B$ ” yang dapat diartikan dengan “Jika A, maka B”. Berikut persamaan untuk menghitung nilai support [2]:

$$Support(A, B) = \frac{T_{A \cap B}}{T_{Total}}$$

Keterangan :

$T_{A \cap B}$ = Jumlah item yang mengandung A dan B

T_{Total} = Jumlah semua item

Confidence (nilai kepastian) adalah persentase kombinasi itemset yang menunjukkan kuatnya hubungan antara dua item atau lebih dalam aturan asosiasi [13]. Saat menghitung nilai confidence, dilakukan penukaran item, seperti kombinasi dua item, yaitu $A \rightarrow B$, lalu balikkan ke $B \rightarrow A$. Kombinasi 3 itemset yang sama yaitu $A, B \rightarrow C$, kemudian dibalik menjadi $A, C \rightarrow B$ dan $B, C \rightarrow A$. Setiap nilai support itemset nilainya bisa tetap sama, tapi bisa jadi memiliki nilai confidence yang berbeda. Berikut persamaan untuk menghitung nilai confidence [2]:

$$Confidence(A, B) = \frac{T_{A \cap B}}{T_A}$$

$$Confidence(A, B, C) = \frac{T_{A \cap B \cap C}}{T_{A \cap B}}$$

Keterangan :

T_A = Jumlah item yang mengandung A

$T_{A \cap B}$ = Jumlah item yang mengandung A dan B

$T_{A \cap B \cap C}$ = Jumlah item yang mengandung A, B, dan C

Nilai support adalah nilai pendukung dari kombinasi itemset sedangkan nilai confidence adalah nilai kepastian dari relasi antar item pada suatu kombinasi itemset [4].

2.2.4 Algoritma Apriori

Apriori adalah salah satu algoritma dari metode aturan asosiasi yang sangat terkenal dalam melakukan pencarian frequent itemset (itemset yang sering muncul). Algoritma Apriori digunakan untuk mencari frequent itemset yang memenuhi nilai minimum support, kemudian mendapatkan rule yang

memenuhi nilai minimum confidence dari frequent itemset yang sudah didapatkan [6].

Proses utama dari algoritma Apriori adalah sebagai berikut :

- Penggabungan (Join)
Proses ini adalah menggabungkan item dengan item lainnya dalam sebuah data sehingga tidak ada kombinasi lagi yang terbentuk [12].
- Pemangkasan (Prune)
Proses ini adalah memangkas hasil kombinasi item menggunakan nilai minimum Support yang telah ditentukan [12].

Algoritma Apriori memiliki kelebihan dan kekurangan dibandingkan algoritma lain. Kelebihannya adalah algoritma Apriori memiliki kemampuan untuk menangani data yang besar dengan baik [2]. Selain itu algoritma Apriori juga dapat menyederhanakan sebuah data. Untuk kekurangannya, algoritma Apriori memerlukan banyak waktu apabila memiliki banyak iterasi karena dalam setiap iterasi, algoritma apriori melakukan scan database. Oleh karena itu, algoritma Apriori memiliki kecepatan pemrosesan yang kurang dibandingkan dengan algoritma-algoritma lain.

2.2.5 Algoritma FP-Growth

Algoritma FP-Growth adalah salah satu algoritma dari metode aturan asosiasi yang dapat digunakan untuk menentukan frequent itemset (itemset yang sering muncul) dalam sebuah data. Algoritma FP-Growth juga merupakan algoritma yang berasal dari pengembangan algoritma Apriori, sehingga kekurangan dari algoritma Apriori diperbaiki oleh algoritma FP-Growth [14].

Kelebihan algoritma FP-Growth adalah bahwa hal itu dapat menghemat waktu dan ruang penyimpanan serta penggunaan sarana partisi untuk menghindari database skala besar. Pada algoritma Apriori diperlukan generate candidate untuk mendapatkan frequent itemsets. Akan tetapi, algoritma FP-Growth tidak memerlukan generate candidate karena FP- Growth menggunakan

konsep pembangunan tree untuk melakukan pencarian frequent itemsets. Dengan begitu kecepatan pemrosesan algoritma FP-Growth lebih cepat dibandingkan dengan algoritma Apriori [14], [15].

2.2.6 Lift Ratio

Lift Ratio adalah ukuran kekuatan aturan asosiasi yang diperoleh dengan membandingkan kepercayaan aturan dengan nilai kepercayaan patokan. Patokan kepercayaan itu sendiri adalah jumlah perbandingan semua item (bagian) berikutnya dengan jumlah total transaksi. Pada tahap ini, validitas aturan penelitian adalah apakah produk A dan produk B dibeli secara bersamaan. Jika rasio lift lebih besar dari 1, aturan tersebut dikatakan valid. Artinya aturan tersebut dapat dijadikan acuan untuk rekomendasi produk. Berikut persamaan untuk menghitung nilai lift ratio [2]:

$$Lift\ Ratio\ (A, B) = \frac{Confident\ (A, B)}{BC\ (A, B)}$$

$$BC\ (A, B) = \frac{N_c}{N}$$

Keterangan :

BC (A ,B) = Benchmark Confidence produk A dan B

N_c = Jumlah transaksi dengan item yang menjadi consequent

N = Jumlah transaksi keseluruhan

2.2.7 Tingkat Kekuatan Aturan Asosiasi

Kekuatan aturan asosiasi adalah salah satu tolak ukur untuk mengetahui bagus tidaknya algoritma dalam membuat aturan asosiasi yang valid. Untuk rentang kekuatan aturan asosiasi ini adalah 0 – 1, dimana aturan yang mendekati atau sama dengan 1 itu sangat baik. Berikut persamaan untuk menghitung nilai kekuatan aturan asosiasi [3] :

$$\text{Kekuatan Aturan Asosiasi} = \frac{\sum_{i=1}^n (Si \times Ci)}{n}$$

Keterangan :

n = Jumlah aturan asosiasi (Valid)

Si = Nilai support aturan asosiasi

Ci = Nilai confidence aturan asosiasi

2.2.8 Parameter Tuning

Parameter tuning merupakan suatu proses untuk menghasilkan parameter yang paling optimal dengan proses *tuning* yang nantinya akan digunakan dalam proses data mining. Proses *tuning* ini adalah metode mencari nilai parameter dengan cara mencoba berulang kali suatu nilai tertentu sebagai parameter sampai didapat sebuah parameter yang paling optimal yang akan digunakan pada proses utamanya [16].