

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Knowledge Discovery In Database

*Knowledge Discovery in Database* (KDD) adalah proses menentukan pola-pola dan informasi yang berguna yang ada di dalam data yang berjumlah besar yang berpotensi bermanfaat. Terdapat lima tahapan *Knowledge Discovery in Database* antara lain:

1. *Data Selection*

*Data Selection* adalah proses pemilihan himpunan data pada sampel data dimana pengolahan akan dilakukan.

2. *Pre-Processing* dan *Data Cleaning*

*Pre-Processing* dan *Data Cleaning* adalah proses pembersihan data yang tidak relevan, data yang tidak konsisten, data yang terduplikasi, *noise* pada data dan memperbaiki kesalahan pada data.

3. *Transformation*

*Transformation* adalah proses mentransformasi atau menggabungkan data sebelum melakukan proses *mining*.

4. *Data Mining*

Proses mencari informasi dan pola-pola yang menarik dari data yang sudah ditransformasi dengan menggunakan teknik, metode dan algoritma tertentu.

#### 5. *Interpretation* dan *Evaluation*

Proses menerjemahkan dan menguji atau mengevaluasi pola-pola yang terdapat dari data yang diolah dengan teknik, metode dan algoritma yang sudah ditentukan. Hasil dari Evaluasi dipresentasikan dalam bentuk visualisasi. (Tamaela, Sedyono, & Setiawan, 2018)

### **2.2 Data Mining**

Data *Mining* adalah proses pengumpulan dan pengolahan data yang bertujuan untuk memperoleh informasi penting dari data tersebut. Data *Mining* digunakan untuk memenuhi beberapa tujuan tertentu, antara lain:

#### 1. Sebagai Sarana menjelaskan (*Explanatory*)

Data *Mining* dapat digunakan sebagai sarana menjelaskan hasil dari data yang sudah diolah.

#### 2. Sebagai Sarana Konfirmasi (*Confirmatory*)

Data *Mining* dapat digunakan sebagai sarana untuk memastikan sebuah pertanyaan atau mempertegas hipotesis.

#### 3. Sebagai Sarana Eksplorasi (*Exploratory*)

Data *Mining* dapat digunakan sebagai sarana untuk mencari pengetahuan lebih banyak atau pengalaman baru.

Data *Mining* juga biasa digunakan untuk mengelola data yang berjumlah banyak dan untuk memudahkan memberikan informasi yang akurat bagi

penggunanya. Karena itu, beberapa tahun belakangan ini data *mining* menjadi fokus yang menarik bagi industri untuk membantu dalam manajemen pengambilan keputusan perusahaan atau dalam perumusan strategi pemasaran dan penjualan perusahaan. (Dicoding, 2021)

### **2.2.1 Metode-Metode Data Mining**

Secara Umum, terdapat beberapa metode yang dilakukan untuk melakukan proses data *mining*, Antara lain : *Association*, *Classification*, *Clustering*, *Forecasting* dan *Regression*, Setiap metode mempunyai tujuan yang berbeda, yaitu:

#### 1. *Association*

*Association* bertujuan untuk mencari atribut yang muncul dalam satu waktu. *Association* juga disebut sebagai *Market Basket Analysis*.

#### 2. *Classification*

*Classification* bertujuan untuk memberikan target variabel kategori sehingga mendapatkan hasil analisa perbedaan dan persamaan karakteristik data.

#### 3. *Clustering*

*Clustering* bertujuan untuk mengelompokkan kelas ke dalam beberapa bagian yang sudah ditentukan. *Clustering* juga disebut sebagai Segmentasi

#### 4. *Forecasting*

*Forecasting* bertujuan untuk mengetahui presentase informasi pada periode yang akan datang. Metode *Forecasting* menggunakan data deratan angka dari periode sebelumnya.

## 5. Regression

*Regression* bertujuan untuk mencari dan menentukan nilai numerik pada suatu pola. Metode *regression* biasanya digunakan untuk mencari perbedaan suhu, prakiraan cuaca dan tekanan udara. (Center, 2021)

### 2.3 Association Rules

*Association Rules* yang juga disebut *market basket analysis* termasuk salah satu teknik *data mining* yang bertujuan mencari keterhubungan antara satu item dengan item lainnya dan menghasilkan *output* item aturan asosiatif antara suatu kombinasi item. *Association Rules* didefinisikan sebagai suatu proses untuk menemukan semua *frequent itemset* yang memenuhi syarat minimum *support* dan syarat minimum *confidence*. *Support* adalah ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item/itemset dari keseluruhan transaksi. (Junianto & Rachman, 2020) Untuk mendapatkan nilai *support* dari satu item dapat diperoleh dengan rumus sebagai berikut:

$$Support(x) = \frac{\text{Jumlah Transaksi yang Mengandung Item X}}{\text{Total Transaksi}}$$

Kemudian untuk memperoleh nilai *support* dari dari dua item dapat diperoleh dengan rumus sebagai berikut:

$$Support(x, y) = \frac{\text{Jumlah Transaksi yang Mengandung Item X dan Y}}{\text{Total Transaksi}}$$

*Confidence* yaitu probabilitas produk dibeli bersamaan dengan produk lain. Berikut ini adalah persamaan untuk nilai suatu *confidence*.

$$\text{Confidence } X \rightarrow Y = \frac{\text{Jumlah Transaksi yang Mengandung Item X dan Y}}{\text{Jumlah Transaksi yang mengandung X}}$$

Contoh jika diketahui  $X = \{\text{Roti, Selai}\}$  dan  $Y = \{\text{Mentega}\}$ , jumlah transaksi yang berisi  $X \cup Y = \{\text{Roti, Selai, Mentega}\}$  ada 5 dan total jumlah ada 100 maka  $\text{support}(X, Y) = 5/100 = 0,05$ . Sedangkan untuk mencari *confidence* adalah dengan membagi  $X \cup Y$  yaitu 5 dengan total jumlah transaksi X yaitu 10, maka  $\text{confidence } X \rightarrow Y = 5/10 = 0.5$ .

Penentuan minimum *support* sebagai ambang batas *support* dan minimum *confidence* sebagai ambang batas *confidence* dapat diatur sesuai dengan kebutuhan sehingga hasil *support* dan *confidence* yang akan muncul sesuai keinginan. (Junianto & Rachman, 2020) Sebagai contoh, jika menentukan minimum *support* 20% dan minimum *confidence* 30%, maka yang akan muncul adalah *support* yang memiliki nilai 20% ke atas dan *confidence* yang memiliki nilai 30% ke atas.

Hasil Penelitian yang digunakan penulis sebagai tinjauan pustaka berikutnya merupakan penelitian tentang Implementasi Data Mining Market Basket Analysis Menggunakan Association Rules pada Minimarket 212 Mart Batam (Rumahorbo & Arnomo, 2020), menyatakan bahwa penerapan metode market basket analysis dapat menganalisa kebiasaan belanja konsumen dan dapat digunakan untuk mengefisiensikan persediaan produk agar transaksi penjualan dapat berjalan lancar, selain itu dapat menentukan promosi produk agar penjualan meningkat.

## 2.4 Algoritma FP-Growth

Algoritma *FP-Growth* merupakan algoritma yang dapat digunakan dalam menentukan himpunan data yang sering muncul (*frequent itemset*) dalam sebuah kumpulan data. *FP-Growth* dimulai dengan menghitung item tunggal sesuai dengan jumlah kemunculan item yang ada di dalam *dataset*. Setelah proses perhitungan selesai maka akan dibuat struktur pohon yaitu *FP-Tree*. *FP-Tree* merupakan struktur penyimpanan data yang dimampatkan. *FP-Tree* dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam *FP-Tree*. Di dalam setiap data transaksi yang dipetakan terdapat transaksi dengan item yang sama, memungkinkan lintasannya untuk saling menimpa. Bertambah banyaknya data transaksi yang memiliki item yang sama akan membuat pemampatan pada struktur data *FP-Tree* akan semakin efektif. (Junianto & Rachman, 2020)

Adapun *FP-Tree* adalah sebuah pohon dengan definisi sebagai berikut :

1. *FP-Tree* dibentuk oleh sebuah akar yang diberi label null dan sekumpulan berupa pohon yang beranggotakan item -item tertentu
2. Setiap simpul dalam *FP-Tree* mengandung tiga informasi penting, yaitu label item yang meninformasikan jenis item yang direpresentasikan oleh simpul dan *support count* yang menginformasikan jumlah lintasan transaksi yang melalui simpul tersebut.

Setelah pembangunan *FP-Tree*, maka akan diterapkan algoritma *FP-Growth* yang berguna untuk mencari *frequent itemset* yang signifikan. Tahap pertama algoritma *FP-Growth* adalah Pembangkitan *Conditional Pattern*

*Base*. *Conditional Pattern Base* adalah *sub-database* yang berisi lintasan prefix dan pola akhiran. Setelah tahap *Conditional Pattern Base* selesai, dilakukan tahap *Conditional FP-Tree*. Di dalam tahap *Conditional FP-Tree*, *support count* dari setiap item akan dijumlahkan lalu item yang memiliki jumlah *support count* yang lebih besar sama dengan *minimum support* akan diambil untuk pembangkitan *frequent pattern*. Dalam pembangkitan *frequent pattern*, hasil dari *Conditional FP-Tree* dikombinasikan sehingga membentuk *frequent itemset* yang menampilkan *support count*. Jika bukan lintasan tunggal maka dilakukan pembangkitan *FP-Growth* secara rekursif. (Tajrin, Tamara, Wijaya, Josephin, & Kurniawan, 2021)

## **2.5 Rapidminer**

Rapidminer merupakan perangkat lunak yang berguna untuk melakukan analisis terhadap data *mining*, *text mining* dan analisis prediksi. Rapidminer juga menyediakan berbagai macam operator yang berguna untuk menjalankan berbagai macam algoritma, seperti: *Apriori*, *decision tree*, *FP-Growth*, dan sebagainya. Rapidminer juga menyediakan prosedur data *mining* seperti: data *preprocessing*, visualisasi, *modelling*, dan evaluasi. (Siregar & Puspabhuana, 2017) Pada penelitian ini Rapidminer digunakan untuk perhitungan algoritma *FP-Growth*.

## **2.6 Penelitian Terkait**

Hasil penelitian yang digunakan oleh penulis sebagai tinjauan pustaka merupakan penelitian tentang Penerapan Algoritma *FP-Growth* Untuk Menentukan Pola Pembelian Konsumen Pada AHASS Cibadak (Suhada,

Ratag, Gunawan, Wintana, & Hidayatulloh, 2020), Penelitian ini bertujuan untuk mendapatkan informasi berupa *frequent itemset* yang bermanfaat dari pengolahan data transaksi penjualan dengan menggunakan algoritma *FP-Growth*. Hasil penelitian ini menyatakan bahwa penelitian ini mampu mengetahui seberapa banyak barang yang dibeli oleh konsumen dan membantu dalam pemesanan barang atau produk.

Hasil penelitian yang digunakan penulis sebagai tinjauan pustaka berikutnya merupakan penelitian tentang Penerapan Algoritma *FP-Growth* dalam menentukan kecenderungan mahasiswa mengambil mata kuliah pilihan (Tahir & Sitompul, 2021), penelitian ini bertujuan untuk memanfaatkan data pemilihan matakuliah yang dimiliki menjadi informasi pola pemilihan matakuliah dengan menggunakan algoritma *FP-Growth*. Penelitian ini mendapatkan hasil bahwa algoritma *FP-Growth* mampu memproses data pemilihan matakuliah dan *association Rules* yang didapatkan dapat memberikan informasi matakuliah mana yang paling banyak diminati oleh mahasiswa.

Hasil penelitian yang digunakan penulis sebagai tinjauan pustaka berikutnya merupakan penelitian tentang Penerapan *FP-Growth* dalam penjualan perlengkapan ibadah umat muslim (Goni, Suratno, Nursyi'ah, & Gustian, 2020) penelitian ini bertujuan untuk mengetahui barang yang dibeli secara bersamaan sehingga pemilik toko mampu mengatur tata letak barang. Penelitian ini mendapatkan hasil bahwa algoritma *FP-Growth* menghasilkan *association rules* sehingga dapat diterapkan untuk mendukung strategi

promosi dengan melakukan pemberian diskon pada barang yang dibeli secara bersamaan dan pemilik toko dapat menempatkan barang-barang yang dibeli bersamaan saling berdekatan.

Hasil Penelitian yang digunakan penulis sebagai tinjauan pustaka berikutnya merupakan penelitian tentang Implementasi Data *Mining* Metode Asosiasi Algoritma *FP-Growth* pada Perusahaan Ritel (Wibowo & Jananto, 2020) penelitian ini bertujuan untuk mencari pola asosiasi barang yang sudah terjual untuk menentukan tata letak barang berdasarkan data transaksi penjualan. Penelitian ini menggunakan algoritma *FP-Growth* dan mendapatkan hasil bahwa pola asosiasi yang didapatkan memberikan informasi untuk barang keperluan sekolah dan barang keperluan kantor harus diletakkan berdekatan dengan alat tulis.

Hasil penelitian yang digunakan penulis sebagai tinjauan pustaka berikutnya merupakan penelitian tentang Analisa Data Profil Pelanggan Menggunakan Algoritma *FP-Growth* (Suryani, Defit, & Yunus, 2020) penelitian ini bertujuan untuk menentukan aturan asosiasi dari basis data pesanan dengan menggunakan algoritma *FP-Growth* sehingga menghasilkan strategi promosi untuk memperkenalkan perusahaan. Penelitian ini mendapatkan hasil bahwa penggunaan algoritma *FP-Growth* dapat meningkatkan keakuratan analisis pelanggan dan tingkat kepercayaan yang didapatkan dapat membantu dalam peningkatan penjualan.

Hasil penelitian yang digunakan penulis sebagai tinjauan pustaka berikutnya merupakan penelitian tentang Implementasi *FP-Growth* Dalam

Menganalisa Penjualan Obat (Hutabarat, Saripurna, & Azlan, 2020) penelitian ini bertujuan untuk mengimplementasikan algoritma *FP-Growth* ke dalam data transaksi penjualan obat sehingga menghasilkan informasi untuk strategi peningkatan penjualan obat. Penelitian ini mendapatkan hasil bahwa *Association rules* yang didapatkan dapat menjadi acuan untuk peningkatan penjualan obat.