

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teori**

Pada bab ini penulis akan membahas beberapa teori yang akan menjadi acuan penulis dalam melakukan penelitian dengan menjelaskan pengertian dan istilah yang berkaitan dengan topik penelitian.

##### **2.1.1 Teori Umum**

Dengan menjelaskan pengertian dan istilah yang terkait dengan judul penelitian, penulis membahas beberapa teori umum yang menjadi pedoman dasar dalam melakukan penelitian.

###### **2.1.1.1 KDD (*Knowledge Discovery in Database*)**

Knowledge Discovery (KDD) dalam database, Pada dasarnya bagaimana mendapatkan pengetahuan yang bermanfaat dari sekumpulan data (Elisa, 2018) KDD adalah suatu proses yang meliputi pengumpulan, penggunaan, data yang telah disimpan untuk mendapatkan keteraturan, pola dalam kumpulan data yang besar, metode darimana mendapatkan informasi yang terdapat dalam database, salah satu prosesnya adalah proses data mining.

Penambangan data dan KDD memiliki tujuan yang sama yaitu menggunakan data yang tersedia di database dan memanipulasi data tersebut untuk mendapatkan informasi baru dan berguna. Selain itu, masih banyak istilah lain yang memiliki arti yang mirip dengan penambangan data, seperti penemuan pengetahuan, ekstraksi pengetahuan, analisis skema/data, arkeologi data, dan penambangan basis data.

Banyak yang menganggap penambangan data sebagai sinonim untuk istilah lain yang umum digunakan "penemuan pengetahuan dari data" atau KDD, sementara yang lain menganggap penambangan data tidak lebih dari langkah kunci dalam proses penemuan pengetahuan. (Utomo & Mesran, 2020).

*Knowledge Discovery in Database* (KDD) memiliki tata cara tahapan-tahapan secara keseluruhan yang terdiri dari (Arifin et al., 2020):

1. *Data Selection*: Pemilihan data yang akan dipakai dalam proses data *mining* di mana sekumpulan data dikumpulkan sebelum tahap penyaringan informasi.
2. *Cleaning/ Pre-Processing*: Sebelum proses tahap data *mining* terlebih dahulu melalui proses pembersihan/*cleaning* pada data yang akan menjadi pokok dari KDD. Proses *cleaning* merupakan proses menghapus data-data yang ganda atau memeriksa data yang *inkonsisten*.
3. *Transformasi: Coding* merupakan perubahan data yang telah terpilih, sehingga data yang telah terpilih memenuhi syarat dalam proses data *mining*.
4. *Data Mining*: Bagian tahapan pencarian pola atau menemukan informasi yang dibutuhkan di dalam data yang digunakan, dengan menggunakan metode yang sesuai. Penggunaan metode dalam data *mining* mempunyai beberapa *variasi* sehingga menggunakan metode yang sesuai sangat mempengaruhi hasil akhirnya.
5. *Evaluasi/Interpretation*: informasi yang memiliki pola dari hasil proses data *mining* agar dapat disajikan dengan mudah dimengerti oleh pihak yang

membutuhkan informasi tersebut, tahapan KDD ini disebut dengan *interpretation*. Pada proses ini merupakan pemeriksaan pola atau informasi yang ditemukan bertolak belakang dengan yang terjadi di lapangan atau hipotesia yang ada sebelumnya.

#### **2.1.1.2 Data Mining**

Data *Mining* mungkin kita sudah sering mendengar kata ini data *mining* pengertian dari data *mining* ini adalah proses dari penggalian informasi dari dalam *database* berfungsi sebagai informasi yang berharga bagi penggunanya dengan menggunakan teknik *estimasi*, *clustering*, *deskripsi* dan banyak lagi (Lewis et al., 2021).

Data *mining* diartikan sebuah proses untuk menemukan pola yang terdapat dalam data yang dipilih, proses ini menggunakan *system* otomatis atau seringnya semiotomatis. Pola yang didapatkan harus memiliki fungsi yang berarti dan memberikan manfaat bagi pengguna informasi tersebut. Biasanya data yang dibutuhkan sangat besar (Santoso et al., 2016).

Menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT, data mining merupakan jenis analisis data yang memiliki kemampuan hebat untuk menemukan hubungan eksplisit dan meringkas data yang tidak diketahui dengan cara modern yang dapat dipahami pengguna. Data *mining* tidaklah hal yang baru sehingga salah satu kesulitan untuk menyimpulkan data *mining* ini dalam berbagai aspek (Mardi, 2019).

### 2.1.1.3 Pengelompokan Data Mining

Menurut (Alfiyan et al., 2019) data *mining* dibentuk menjadi beberapa kelompok berdasarkan yang ditentukan:

1. Deskripsi (*Description*)

Deskripsi adalah salah satu cara untuk menjabarkan bentuk dari kecenderungan yang terdapat dalam data yang digunakan, sering sekali menggunakan analisa sederhana sehingga menghasilkan kemungkinan pengertian untuk suatu pola

2. Estimasi (*Estimation*)

Estimasi mempunyai kesamaan dengan kategorikal, kecuali bahwa estimasi lebih numerik daripada kategorikal, dan membangun model dengan penggunaan data baris penuh, menghasilkan nilai variabel sebagai prediktor. Dalam ulasan selanjutnya, estimasi variabel target didasarkan pada nilai yang ditentukan dari variabel prediktor.

3. Prediksi (*Prediction*)

Prediksi sendiri memiliki kesamaan dengan klasifikasi dan estimasi, hanya saja nilai dalam prediksi merupakan hasil dari nilai yang dimasa mendatang. Beberapa bagian tahapan dalam klasifikasi dan estimasi dapat juga dipergunakan diprediksi sesuai penggunaan yang tepat.

4. Klasifikasi (*Classification*)

Klasifikasi dapat ditemukan target variabel katagori, contohnya dengan mengklasifikasikan pendapatan yang dapat di klasifikasikan menjadi tiga bagian, yaitu pendapatan besar, pendapatan sedang dan pendapatan kecil.

5. Pengklusteran (*Clustering*)

Pengelompokan atau pengelompokan bagian merekam, mengamati, atau berfokus pada dan membangun kelas objek yang serupa. Sebuah cluster sendiri dapat diartikan sebagai kumpulan record yang memiliki kemiripan satu sama lain dan memiliki record yang berbeda dalam cluster lainnya. Clustering tidak sama dengan klasifikasi pada tahap clustering, clustering tidak memiliki variabel target Clustering tidak mengklasifikasikan atau memperkirakan variabel target, tetapi algoritma clustering mengelompokkan semua data untuk membentuk kelompok yang serupa. Dimana pada data yang mempunyai kesamaan *record (homogen)* mempunyai nilai maksimal, sedangkan untuk data yang mempunyai kesamaan *record* dalam bagian lain akan mempunyai nilai minimum.

#### 6. Asosiasi

Peran asosiasi dalam proses penambangan data adalah untuk menentukan karakter mana yang muncul pada suatu waktu. Analisis keranjang merupakan bagian dari salah satu implementasi dari asosiasi atau analisis, contoh asosiasi adalah menentukan barang mana di supermarket yang akan dan tidak akan dibeli bersama oleh pelanggan.

##### **2.1.1.4 Association Rule Mining**

Teknologi penambangan aturan asosiasi adalah metode penambangan data yang memperoleh bentuk asosiasi yang sering terjadi dalam data, dan tahap analisis ini adalah pola frekuensi tinggi, Menggunakan pola frekuensi tinggi dalam analisis adalah proses menemukan portofolio item minimum yang diperlukan dengan nilai dukungan yang sesuai (Qoniah & Priandika, 2020). Metode ini menjadi populer

dikarena sering digunakan untuk menganalisis keranjang pasar. Aturan yang dihasilkan akan muncul dalam bentuk "jika maka" dalam transaksi. Dapat menjelaskan cara membeli barang A kemudian membeli barang B dalam satu kali transaksi (Santoso et al., 2016). Aturan dalam asosiasi sering dinyatakan dalam bentuk sebagai berikut (Fauzy et al., 2016):

1. *Support* Persentase transaksi yang berisi seluruh program. Jika Anda memiliki proyek A dan proyek B, maka dukungan adalah proporsi transaksi dalam database yang memiliki data A dan data B. Semakin tinggi support yang diterima maka dapat disimpulkan bahwa item tersebut laris manis di pasaran karena muncul dalam banyak transaksi, sedangkan semakin rendah support yang diterima maka item tersebut tidak laku.
2. *Confidence* merupakan Kepastian aturan komputasi, yaitu penyajian transaksi dalam database, memiliki A dan B. Adanya kepercayaan tersebut memungkinkan perhitungan kekuatan keterkaitan antar item dalam suatu asosiasi.

#### **2.1.1.5 Algoritma Apriori**

Algoritma apriori merupakan bagian dari kumpulan jumlah item tunggal (Febrianti dan Suryadi, 2018) tahapan untuk tahapan dalam menentukan *frequent-itemset* dengan melakukan literasi pada data. Dimana himpunan merupakan *itemset* dan diolah oleh sisten, sedangkan *frequent-itemset* menunjukkan *itemset* yang mempunyai kemunculan frekuensi yang memiliki nilai lebih minimum yang sudah ditentukan (Fauzy et al., 2016).

Pentingnya suatu asosiasi dapat dinilai dengan dua tolok ukur, dukungan dan kepercayaan. Nilai support didefinisikan sebagai persentase dari item-item tersebut yang tercampur dalam database, sedangkan nilai confidence adalah kekuatan hubungan antar item dalam aturan asosiasi. (Thoriq Muhammad & Nurhadiyono, 2014).

Menurut (Arifin et al., 2020) Algoritma apriori adalah bagian dari algoritma dasar yang dimanfaatkan sebagai saran untuk menentukan frequent itemsets untuk asosiasi Boolean. Asosiasi pada data *mining* merupakan algoritma apriori, aturan ini disebut analisis afinitas atau analisis keranjang pasar. Algoritma apriori, juga dikenal sebagai aturan asosiasi, adalah teknik penambangan data yang digunakan untuk menemukan aturan asosiasi dalam portofolio.

Metodologi dasar analisis asosiasi terbagi menjadi dua bagian yaitu (Kurniawan et al., 2014):

1. Analisis pola frekuensi tinggi

Identifikasi pola frekuensi tinggi selama fase analisis Portofolio proyek yang memenuhi persyaratan minimum Nilai dukungan untuk semua transaksi (Riszky & Sadikin, 2019). Tahap pertama dari algoritma sebelumnya adalah analisis pola frekuensi tinggi, yaitu dengan mendapatkan portofolio item dalam database yang mendukung nilai minimum yang diperlukan. Nilai *support* sebuah item diperoleh dengan rumus berikut:

$$Support (A) = \frac{Jumlah\ rekam\ medis\ mengandung\ keluhan\ A}{Total\ Rekam\ medis}$$

**Rumus 2.1** Menghitung Itemset-1

Sementara nilai support dua item menggunakan rumus sebagai berikut:

$$\text{Support}(A, B) = P(A \cap B)$$

$$\text{Support}(A, B) = \frac{\sum \text{Jumlah rekam medis mengandung } A \text{ dan } B}{\sum \text{Total rekam medis}}$$

### Rumus 2.2 Menghitung Itemset-2

Frequent itemset mewakili itemset yang terjadi lebih sering daripada nilai minimum yang ditentukan ( $\phi$ ). Misalnya, jika  $\phi = 2$ , semua itemset dengan frekuensi lebih besar atau sama dengan 2 kali disebut frequent itemset. Himpunan dari *frequent k-itemset* dilambangkan dengan  $F_k$  (Fauzy et al., 2016).

#### 2. Pembentukan Aturan Asosiasi

Setelah menemukan semua pola frekuensi tinggi, dengan menghitung kepercayaan aturan asosiasi  $A \rightarrow B$ , temukan aturan asosiasi yang memenuhi persyaratan kepercayaan minimum. Nilai *Confidence* dari aturan  $A \rightarrow B$  diperoleh rumus berikut (Kurniawan et al., 2014):

$$\text{Confidence } P(B|A) = \frac{\sum \text{Rekam medis mengandung } A \text{ dan } B}{\sum \text{Rekam medis mengandung } A}$$

### Rumus 2.3 Rumus Confidence

Untuk menentukan aturan asosiasi yang akan dipilih, jadi harus diurutkan berdasarkan  $\text{Support} \times \text{Confidence}$ . Aturan diambil sebanyak  $n$ - aturan yang memiliki hasil terbesar



## **2.1.2 Teori Khusus**

Penulis membahas beberapa teori khusus yang berkaitan dengan pelaksanaan penelitian yang terdiri dari pengertian dan istilah secara khusus yang berhubungan dengan penelitian penulis.

### **2.1.2.1 Klinik Gigi**

Praktek gigi adalah suatu fasilitas atau tempat yang memberikan pelayanan kesehatan gigi kepada seluruh masyarakat, meliputi pelayanan preventif, kuratif, dan rehabilitatif (Depkes RI, 1996). Dengan adanya klink ini memudahkan pasien langsung ketemu dokter yang menangani permasalahannya.

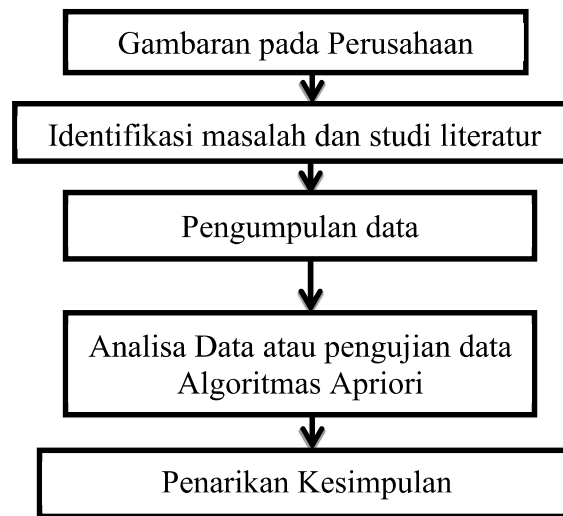
### **2.1.2.2 Administrasi Klinik Gigi**

Dengan adanya kegiatan ini, klinik gigi juga memiliki kegiatan administrasi, banyak klinik gigi yang sudah mencatat rekam medis dengan menggunakan *system* yang membantu dalam pendataan rekam medis dari pasien dan transaksi pembayaran. Dengan adanya data ini klinik gigi dapat menganalisa pola pasien yang datang untuk perawatan gigi.

### **2.1.2.3 Software Tanagra versi 1.4.50**

Untuk pengujian dalam tahap Analisa dan pengujian yang dilakukan secara manual, maka diperlukan Langkah pengujian dengan menggunakan *software Tanagra* menggunakan seluruh data yang didapatkan (Putra & Wadisman, 2018). Pengujian menghasilkan pengujian yang berasal dari dua kombinasi items, Rules terdiri dari *Actecedent* yang merupakan bagian dari kondisi dari pada rules sedangkan *consequent* merupakan pernyataan dari rules, untuk tingkat kekuatan rules dapat dilihat dari *lift* (Sikumbang, 2018).

## 2.2 Kerangka Pemikiran



**Gambar 2.1** Kerangka Pemikiran

**Sumber:** Penulis 2022

Penelitian ini bermula dari menentukan klinik yang akan diteliti, mengidentifikasi permasalahan yang ada pada penelitian dan menganalisa masalah yang terjadi dengan studi literatur. Dilanjutkan dengan mengumpulkan data yang peneliti butuhkan, kemudian dilakukan pengujian data atau analisa data. Setelah didapatkan hasil dilanjutkan ada penarikan kesimpulan dan pemberian saran berdasarkan hasil penelitian.

## 2.3 Penelitian Terdahulu

Penelitian sebelumnya dalam penelitian ini sebagai referensi adalah sebagai berikut:

1. Dari penelitian “Analisa Pola Data Penyakit Rumah Sakit Dengan Menerapkan Metode Association rule Menggunakan Algoritma

Apriori” yang dilakukan oleh (Kurniawan et al., 2014) disimpulkan hasil dari penelitian Melalui penelitian ini pola penyakit rumah sakit dengan menerapkan *Association rule* menggunakan Algoritma Apriori dapat membantu rumah sakit dalam menanyakan data riwayat medis pasien, membantu rumah sakit dalam pengambilan keputusan dan klasifikasi jenis penyakit.

2. Dari penelitian “Survei pada Penggunaan Teknik Data Mining pada Bidang Kesehatan di Indonesia” yang dilakukan oleh (Khomsah et al., 2017) disimpulkan hasil dari penelitian ini bahwa mengelolah rekam medis pada data pasien menghasilkan dampak yang berlimpah, terdapat beberapa jenis penyakit yang harus mendapatkan perhatian dari peneliti diantaranya DM, Jantung, Hepatitis, Kanker Payudara, TB, dan demam berdarah.
3. Dari penelitian “Penerapan Association rule Menggunakan Algoritma Apriori Pada Poliklinik Penyakit Dalam (Studi Kasus: Rumah Sakit Umum Daerah Bintan)” yang dilakukan oleh (Nola Ritha et al., 2021) Kesimpulan yang ditarik merupakan hasil dari penelitian ini. Pendekatan algoritma Apriori dapat digunakan untuk menentukan aturan asosiasi poliklinik. Aturan tersebut dapat digunakan oleh rumah sakit atau dinas kesehatan untuk menjangkau masyarakat dan menjadi keputusan dinas kesehatan dalam pencegahan penyakit.
4. Dari penelitian “Analisis Keterkaitan Penyakit Pasien Pada Puskesmas Menggunakan Metode Association Rule” yang dilakukan oleh

(Auliasari et al., 2016) disimpulkan hasil penelitian ini bahwa dari kedua kombinasi tersebut terdapat aturan nilai minimum *support* dengan nilai minimum *confidence* tertinggi dari data 77 yang telah diuji.

5. Dari penelitian "Penerapan *Association rule* dalam Bidang Kesehatan Pada Rumah Sakit X" yang dilakukan oleh (Karina Intan Indradewi, 2019) disimpulkan hasil penelitian ini bahwa Berdasarkan kesimpulan dari penelitian ini bawah penerapan *Association rule* dapat dimanfaatkan dalam dunia Kesehatan yang dapat membantun membentuk hubungan yang mempunyai kareteristik yang sama. Dengan menggunakan FP-Growth ditemukan itemset yang sering muncul dengan menggunakan prosedur Software Rapid Miner didapatkan 12 turan valid, aturan ini yang dapat dipergunakan sebagai informasi dalam mengdiagnosis pasien.
6. Dari penelitian "Implementasi Algoritma Apriori untuk Menentukan Pola Data Penyakit Pada Anak Usia Dini (Studi Kasus: RS. Estomihi) yang dilakukan oleh (Purba & Buulolo, 2020) disimpulkan hasil penelitian ini dengan metode ini dapat Melalui rekam medis rumah sakit dan data terkait penyakit pasien, kita dapat mengetahui pola penyakit pasien anak usia dini. Dengan menerapkan algoritma sebelumnya, ini memfasilitasi proses pembentukan pola komposisi di antara set item anak usia dini dengan dukungan dan kepercayaan diri yang tinggi.
7. Dari penelitian "Market Basket *Analysis* Pada Mini market Ayu dengan Algoritma Apriori" yang dilakukan oleh (Elisa, 2018) disimpulkan hasil

penelitian ini bahwa tahapan dalam penentuan analisis pola pembelian produk yang dilakukan menggunakan algoritma apriori dapat melihat konsumen kecenderungan dalam beli barang berdasarkan kombinasi 2 *itemset*. Sehingga melalui penelitian ini pemilik dapat mengatur tata letak barang dengan berdekatan untuk mempermudah pelanggan dalam menemukan barang tersebut.