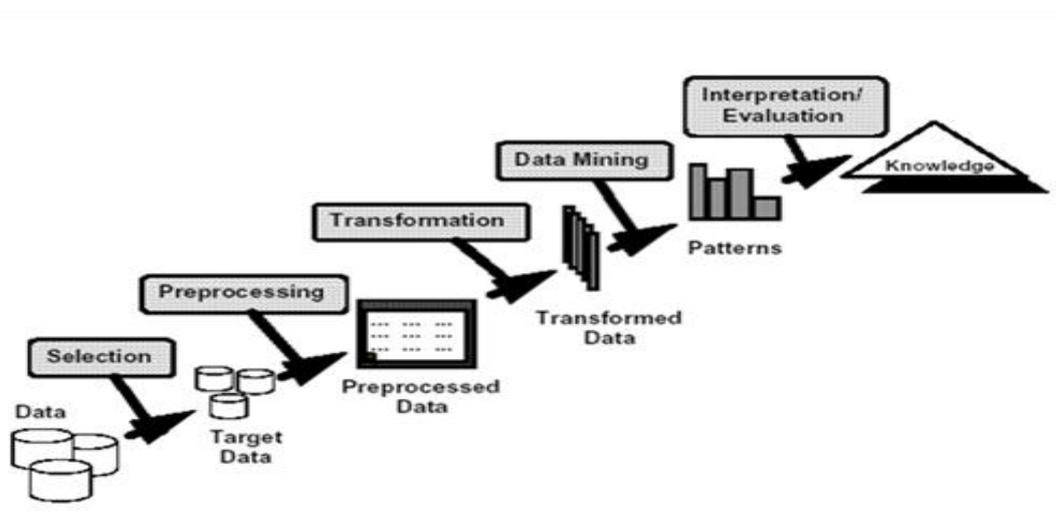


## BAB II

### LANDASAN TEORI

#### 2.1 *Knowledge Discovery In Database*

*Knowledge Discovery In Database (KDD)* merupakan metode yang digunakan untuk mencari pengetahuan atau informasi yang belum diketahui dari sebuah database. *Knowledge Discovery In Database (KDD)* merupakan nama lain dari *Data Mining* walaupun sesungguhnya kedua istilah tersebut memiliki konsep yang tidak sama, namun berkaitan satu sama lain, dan salah satu dari tahapan proses keseluruhan *Knowledge Discovery In Database (KDD)* merupakan *Data Mining* yang menjadi inti dari proses KDD .



**Gambar. 2.1** Proses KDD Data Mining

KDD adalah sebuah proses untuk mencari dan mengidentifikasi *pattern* dalam sebuah *database*, pada sebuah *Knowledge Discovery In Database* atau KDD memiliki beberapa tahapan di antaranya :

1. Seleksi Data

Proses Seleksi Data dilakukan dengan memilih data yang relevan dengan tugas menganalisis dari database, menciptakan himpunan data target, atau memfokuskan pada contoh data dimana *discovery* akan dilakukan dan hasil dari seleksi disimpan dalam suatu berkas terpisah dari *database* operasional.

Tabel 2.1. Seleksi Data Obat

| <b>NO.</b> | <b>NAMA</b>           | <b>NO.</b> | <b>NAMA</b>         |
|------------|-----------------------|------------|---------------------|
| 1          | AMBROXOL              | 26         | PANADOL EXTRA       |
| 2          | AMOXICILIN            | 27         | DECOLSIN            |
| 3          | AMPICILLIN            | 28         | MAMY POKO PANTS (S) |
| 4          | BODREX                | 29         | SGM 1+ 150G         |
| 5          | PARACETAMOL           | 30         | SANMOL SIRUP        |
| 6          | RIVANOL 300 ML        | 31         | SILEX SIRUP         |
| 7          | ALKOHOL 100 ML        | 32         | BODREXIN            |
| 8          | KASSA STERIL          | 33         | OBH COMBI ANAK      |
| 9          | KAPAS SELECTION       | 34         | OBH COMBI ATT 60ML  |
| 10         | HANSAPLAST ROLL BESAR | 35         | VIDORAN SMART       |
| 11         | HANSAPLAST ROLL KECIL | 36         | VIKS FORMULA 44     |
| 12         | BETADINE 60 ML        | 37         | ANTIMO              |
| 13         | PONSTAN               | 38         | KOMIX SACHET        |
| 14         | PROMAG                | 39         | TOLAK ANGIN SACHET  |
| 15         | MKP LANG 30 ML        | 40         | SAMCODIN            |
| 16         | PANADOL EXTRA         | 41         | SCANDEXON           |
| 17         | NORFLAMIN             | 42         | SALBUTAMOL 4 MG     |
| 18         | ASAMEFENAMAT          | 43         | CTM                 |
| 19         | CEFADROXIL 500 MG     | 44         | RANITIDINE          |
| 20         | SANMOL TABLET         | 45         | DEGIROL             |
| 21         | CDR                   | 46         | LANSOPRAZOLE        |

|    |               |    |                 |
|----|---------------|----|-----------------|
| 22 | REDOXON       | 47 | CHIL KID 400 GR |
| 23 | KADITIC       | 48 | ENTROSTOP       |
| 24 | ANASTAN FORTE | 49 | OMERIC 100 MG   |
| 25 | SANGOBION     | 50 | DICLOFENAC      |

## 2. Pemrosesan dan Pembersihan Data

Pada tahapan ini dilakukan menghilangkan *noise* dan data yang inkonsisten. Sebelum proses *Data Mining* dikerjakan, perlu dilakukan proses *cleaning* pada data yang akan menjadi fokus KDD dan proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data lain yang relevan untuk keperluan KDD.

Tabel 2.2. Data Sample Obat

| NO. | NAMA OBAT    |
|-----|--------------|
| 1   | Amoxicillin  |
| 2   | Nonflamin    |
| 3   | Asemefenamat |
| 4   | Cefadroxil   |
| 5   | Sanmol       |
| 6   | CDR          |
| 7   | Redoxon      |
| 8   | Katidic      |
| 9   | Anastan      |
| 10  | Sangobion    |

### 3. Transformasi

Transformasi adalah proses transformasi pada data yang dipilih, sehingga data tersebut sesuai untuk proses *Data Mining*. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau *pattern* informasi yang akan dicari pada *database*.

### 4. *Data Mining*

Pada tahapan ini merupakan proses mencari *pattern* atau pola dan informasi dari sebuah *database* dengan menggunakan teknik atau metode. Pada proses *Data Mining* terdapat banyak teknik, metode atau algoritma yang dapat digunakan dan sangat bervariasi dan untuk menentukan pemilihan metode yang akan digunakan tergantung pada tujuan dan proses KDD secara keseluruhan.

### 5. Evaluasi

Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Proses ini merupakan proses penerjemahan pola-pola yang didapatkan dari *Data Mining*. Pola informasi yang dihasilkan dari proses *Data Mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

## 2.2 *Data Mining*

*Data Mining* merupakan sebuah inti dari proses KDD [10], meliputi dugaan algoritma yang mengeksplor data, membangun model dan menemukan pola yang belum diketahui. KDD bersifat otomatis, dapat didefinisikan sebagai pengorganisasian proses untuk pengidentifikasian yang benar, berguna dan penemuan pola dari kumpulan data yang besar dan kompleks. *Data Mining* merupakan salah satu teknik untuk menemukan, mencari, atau menggali informasi atau pengetahuan baru dari sekumpulan data yang sangat besar, dengan integrasi atau penggabungan dengan disiplin ilmu lain seperti statistika, kecerdasan buatan, serta *machine learning*, menjadikan *Data Mining* sebagai salah satu alat bantu untuk menganalisa data yang kemudian menghasilkan informasi yang berguna.

Dari penjelasan tersebut, *Data Mining* dapat diartikan sebagai sebuah proses dari sekumpulan data yang memiliki jumlah besar untuk mendapatkan data yang hasilnya tidak hanya sekedar informasi melainkan merupakan sebuah pengetahuan atau *knowledge* yang

tersembunyi dari data tersebut. *Knowledge* yang diperoleh dari hasil *mining* terhadap sekumpulan data tersebut dapat membantu dalam pengambilan sebuah keputusan.

### 2.3 *Association Rule*

*Association rule* merupakan suatu proses pada *Data Mining* untuk menentukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support (minsup)* dan *confidence (minconf)* pada sebuah *database*. Kedua syarat yang digunakan untuk *interesting association rules* dibandingkan dengan batasan yang telah ditentukan dengan *minimum support* dan *minimum confidence*. *Association rule* adalah teknik *Data Mining* untuk menemukan aturan asosiasi antara suatu kombinasi barang. Fungsi *association rule* seringkali disebut dengan “*market basket analysis*”, yaitu Analisis dari kebiasaan membeli *customer* dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan *customer* dalam keranjang belanjanya.

Contoh aturan asosiasi dari analisis pembelian di suatu produk pada toko sembako adalah dapat diketahuinya berapa besar kemungkinan seorang konsumen membeli beras bersamaan dengan membeli gula. Dengan pengetahuan tersebut, pemilik toko dapat menyediakan stok lebih barang tertentu yang sering dibeli oleh konsumen.

Dalam menentukan nilai *minimum support* sebuah *item* dapat menggunakan rumus persamaan seperti di bawah ini :

1.  $Support(A) = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}}$
2.  $Support(A \cap B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}}$

Sedangkan untuk menentukan nilai *minimum confidence* sebuah *item* dapat menggunakan rumus persamaan seperti di bawah ini :

1.  $Confidence(A \rightarrow B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Jumlah Transaksi Mengandung A}}$

### 2.4 *FP-Tree*

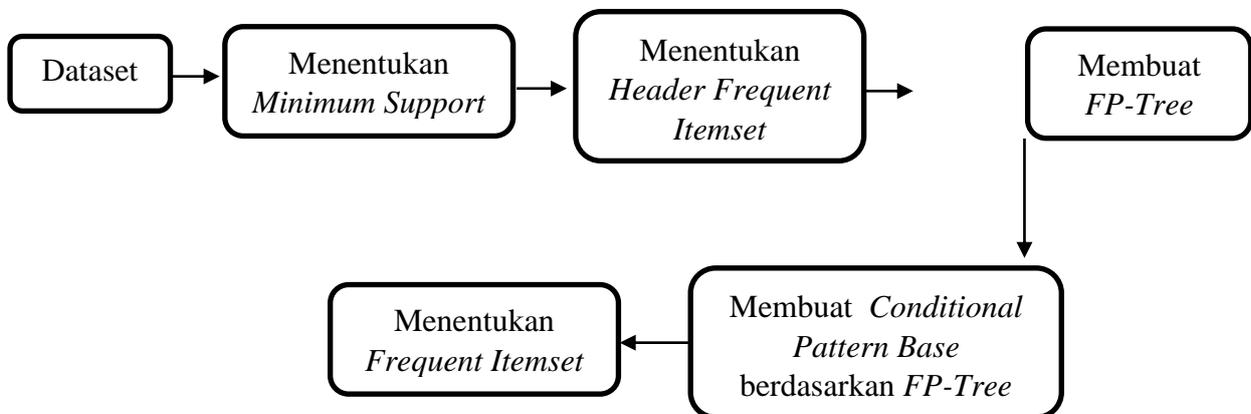
*FP-Tree* merupakan struktur penyimpanan data yang dimampatkan. *FP-Tree* dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam

*FP-Tree*. *FP-Tree* (*Frequent Pattern Tree*) digunakan bersamaan dengan algoritma *FP-Growth* untuk menentukan *frequent itemset* (data yang paling sering muncul) dari sebuah *dataset*.

## 2.5 Algoritma *FP-Growth*

*FP-Growth* adalah algoritma alternatif yang dapat digunakan untuk menentukan itemset yang paling sering muncul dalam satu set data. Algoritma *FP-Growth* merupakan salah satu cara alternatif untuk menemukan himpunan data yang paling sering muncul (*frequent itemset*) tanpa menggunakan generasi kandidat. Algoritma *FP-Growth* sangat efisien untuk menentukan *frequent pattern* baik dalam data yang besar maupun kecil, dibandingkan dengan algoritma apriori algoritma *FP-Growth* lebih cepat karena algoritma *FP-Growth* tidak perlu melakukan iterasi secara berulang seperti algoritma apriori yang membutuhkan waktu yang cukup lama. Dan algoritma apriori tidak dapat menangani data yang besar, sehingga algoritma *FP-Growth* inilah yang dapat dijadikan sebagai solusi bagi permasalahan yang terjadi pada algoritma apriori.

Gambar 2.1 Tahapan-tahapan yang dilakukan dalam algoritma *FP-Growth* adalah sebagai berikut :



**Gambar 2.2** Blog Diagram Algoritma *FP-Growth*

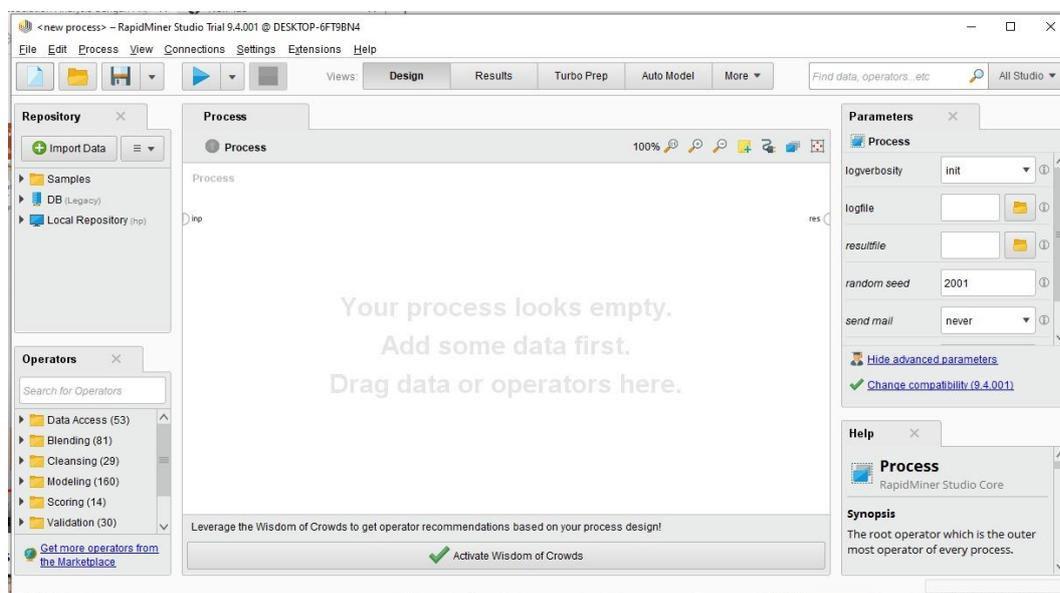
## 2.6 Metode K – Means

K – Means Clustering adalah metode untuk mengkategorikan atau mengelompokkan sekelompok objek sesuai dengan atribut yang sama atau karakteristik ke dalam sejumlah groups. Ini mendefinisikan sebuah cluster oleh massa yang mewakili mean dari cluster. Dalam data mining analisis cluster popular adalah K – Means. Ini adalah metode kuantisasi vector.

## 2.7 Rapidminer

Rapid Miner merupakan perangkat lunak yang bersifat terbuka (open source). Rapid Miner adalah solusi untuk melakukan analisis terhadap *Data Mining*, text mining dan analisis pengelompokan. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik.

Gambar 2.3 Rapidminer yang digunakan dalam menganalisis *Data Mining* adalah sebagai berikut :



**Gambar 2.3 Rapidminer**

## 2.8. Metodologi Penelitian

Metodologi penelitian merupakan cara untuk mengetahui hasil dari sebuah permasalahan yang spesifik dari suatu penelitian. Metodologi penelitian ini dilakukan dengan cara sistematis yang akan digunakan sebagai acuan dalam penelitian.

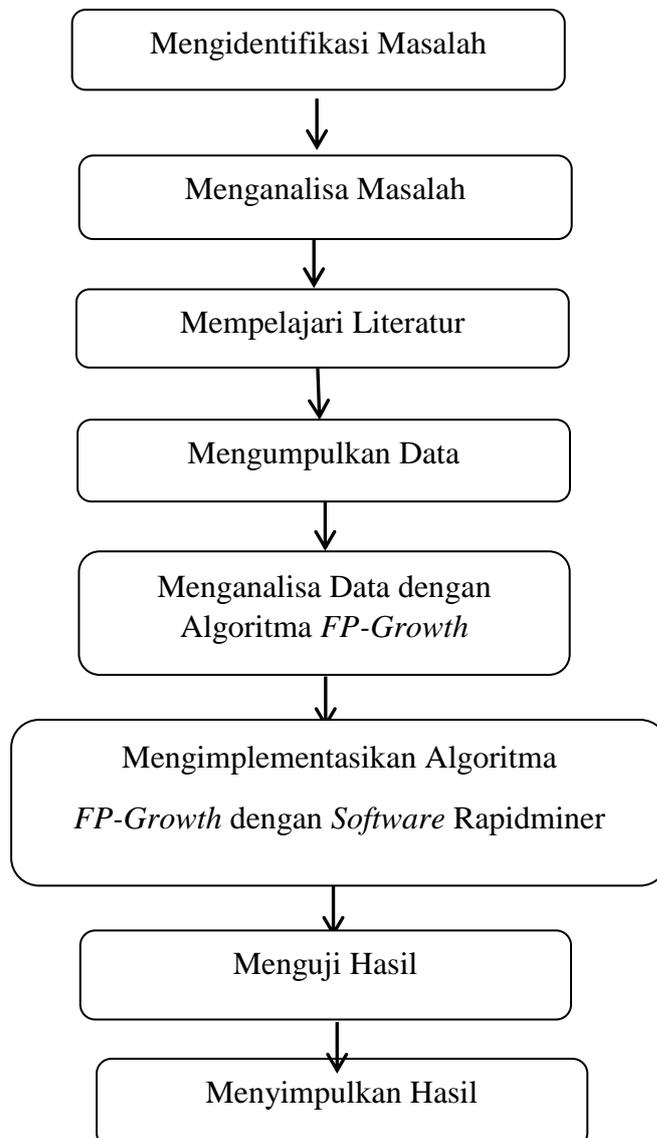
Manfaat metodologi penelitian adalah:

1. Memudahkan pekerjaan peneliti agar sampai pada tahap pengambilan keputusan atau kesimpulan-kesimpulan.
2. Untuk mengatasi berbagai keterbatasan yang ada, misalnya keterbatasan waktu, biaya, tenaga, etik, dan lain-lain.

3. Kesimpulan yang diambil oleh peneliti dapat terpercaya.
4. Kesimpulan yang diambil dapat digunakan untuk memecahkan permasalahan.

## 2.9. Kerangka Kerja

Kerangka kerja dalam penelitian merupakan kumpulan konsep penelitian yang tersusun secara sistematis supaya tujuan dari penelitian tercapai dengan baik. Kerangka penelitian ini dibentuk sebelum langkah penelitian dilakukan, kerangka kerja penelitian juga merupakan konsep suatu penelitian yang menghubungkan antara visualisasi satu variabel dengan variabel lainnya, sehingga penelitian yang dilakukan dapat diterima oleh semua pihak. Dengan adanya kerangka kerja diharapkan penelitian ini dapat diselesaikan dengan baik, gambaran kerangka kerja penelitian ini dapat lihat pada gambar 2.4



## Gambar 2.4 Kerangka Kerja Penelitian

Dari **Gambar 2.4** di atas terlihat jelas bahwa penelitian ini dilakukan secara bertahap dan sistematis, berikut ini adalah penjelasan dari gambar kerangka kerja di atas:

### 1. Mengidentifikasi Masalah

Tahap awal dari penelitian ini adalah mengidentifikasi masalah-masalah yang ada, permasalahan-permasalahan yang teridentifikasi inilah yang akan menjadi bahan acuan untuk mengerjakan tahap berikutnya.

### 2. Menganalisa Masalah

Adanya penelitian karena adanya permasalahan yang harus diselesaikan, artinya analisa masalah merupakan tahapan yang harus dilewati oleh setiap peneliti, pada tahapan ini peneliti melakukan proses analisa masalah pada data penjualan obat Apotik Dirgham. Dengan analisa ini sehingga pokok masalah yang diteliti akan tampak jelas (batasan, lingkup, latar belakang, dan signifikansinya) setelah dilakukan analisis terhadap pokok permasalahan yang bersangkutan.

### 3. Mempelajari Literatur

Dengan mempelajari literatur akan meningkatkan pemahaman peneliti mengenai teori-teori yang relevan, pemahaman akan teori-teori ini diharapkan peneliti dapat menjelaskan, membedakan, memprediksi fenomena-fenomena dan gejala-gejala yang berhubungan dengan masalah penelitian. Dalam studi literatur ini sebelum mengumpulkan data yang akan di proses, terlebih dahulu mengetahui pengetahuan-pengetahuan atau *knowledge* yang akan diterapkan dalam metode algoritma *FP-Growth* ini, literatur yang akan dipelajari ini bersumber dari jurnal-jurnal ilmiah yang di publikasikan di internet.

### 4. Mengumpulkan Data

Dalam pengumpulan data dilakukan observasi yaitu pengamatan secara langsung di tempat penelitian sehingga permasalahan yang ada dapat diketahui secara jelas. Kemudian dilakukan interview yang bertujuan untuk mendapatkan informasi atau data yang dibutuhkan

### 5. Menganalisa Data Menggunakan Algoritma *FP-Growth*

Pada tahap ini ini akan menampilkan bagaimana proses analisa *Data Mining* dengan algoritma *FP-Growth* dirancang berdasarkan data yang telah terkumpul. Dan

bagaimana mengembangkan proses analisa *Data Mining* dengan *Association Rule*, *FP-Tree* dan *FP-Growth* untuk mendapatkan hubungan antar tiap produk yang ada. Dengan menggunakan algoritma *FP-Growth* pada data transaksi penjualan untuk menentukan alternatif yang terbaik dari sekumpulan alternatif yang ada melalui suatu proses yang terstruktur dan analisa bagaimana *FP-Growth* untuk mendapatkan hasil sebagai tujuan yang akan dicapai kemudian dapat dijadikan oleh pihak apotik sebagai pengetahuan dalam meningkatkan penjualan.

6. Mengimplementasikan Algoritma *FP-Growth* dengan *Software* Rapidminer

Pada penelitian ini penulis mengimplementasikan berdasarkan hasil analisa data dengan algoritma *FP-Growth* dengan menggunakan alat bantu komputer dengan sistem operasi *windows* dan *software* Rapidminer 9.4.

7. Menguji Hasil

Pada tahapan ini akan dilakukan pengujian antara data yang dihitung secara manual dengan metode algoritma *FP-Growth* dengan menggunakan *software* Rapidminer 9.4.

8. Menyimpulkan Hasil

Setelah dilakukan implementasi algoritma *FP-Growth* pada Rapidminer 9.4, maka hasil yang didapat adalah produk-produk yang paling sering dibeli oleh konsumen dan kemudian dijadikan acuan untuk menstok barang-barang tersebut sehingga dapat meningkatkan penjualan.