

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teori Dasar**

Adalah upaya dalam menemukan suatu data penelitian yang berhubungan dengan penelitian . Berikut di bawah ini beberapa teori yang dipakai:

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

KDD adalah proses dalam mengumpulkan data dan mengolah data dengan tujuan mendapatkan data informasi yang penting. Sistem ini akan dilaksanakan dengan menggunakan perangkat lunak yaitu kecerdasan buatan. Pada umum nya manfaat KDD terbagi atas dua bagian yaitu deskriptif dan prediktif tetapi selain itu fungsi juga lain seperti *klasifikasi, asosiasi, forecasting, clustering, and sequencing* (Ulfha & Amin, 2020).

*KDD* merupakan seluruh dari proses untuk memecahkan masalah identifikasi paradigma didalam sebuah data, dimana paradigma tersebut ditemukan. KDD berhubungan dengan metode integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data dan juga adalah suatu cara untuk dapat menemukan informasi data yang menjadi solusi penyelesaian masalah data yang besar yang berasal dari database atau data setkedalam volume data yang lebih kecil (Aulia et al., 2021).

### 2.2.1 Tahapan KDD

Berikut beberapa tahapan dari KDD:

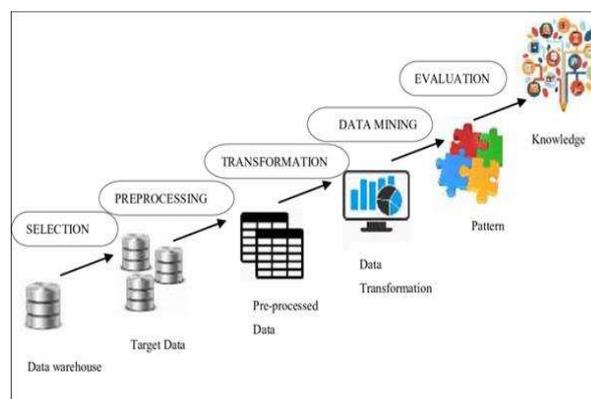
1. Data Seleksi
  - a. Merupakan suatu cara untuk mendapatkan data yang diinginkan atau sesuai dengan kebutuhan dalam pengolahan data yaitu berupa target dan memfokuskan pada variabel tersebut.
  - b. Pemilihan seleksi merupakan dari sekumpulan data operasional di haruskan untuk melakukan pencarian informasi dan mendapatkan hasil seleksi yang dipergunakan pada proses data mining.
2. Proses Data/pembersihan
  - a. Proses, pembersihan data, penghapusan *noise* akan dilakukan terlebih dahulu sebagai dasar operasional.
  - b. Proses *cleaning* meliputi menghapus data yang duplikasi, mengamati data yang tidak stabil serta memperbaiki kecacatan yang ada.
3. *Transformation*
  - a. Proses ini merupakan proses pelacakan karakteristik agar dapat dipresentasikan bersamaan dengan pencapaian yang akan dituju.
  - b. Adalah proses dari pemilihan data yang sinkron untuk proses data *mining*  
Proses ini sering juga disebut dengan proses kreatif dan sangat bergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data Mining*

- a. Pada langkah ini melakukan pilih tanggung jawab pada data mining, dengan pencapaian pada proses KDD misalnya pada klasifikasi data dan regresi.
- b. Melakukan pilih algoritma data mining untuk proses pencarian (*searching*).
- c. Proses yang mencari dan menemukan pradigma, banyak metode dan teknik. Pemilihan metode yang tepat adalah upaya proses pemilihan penyelesaian yang baik pada KDD.

5. *Interpretation/Evaluation*

- a. Penerjemahan pradigma yang berasal dari hasil data mining.
- b. Pradigma atau pola informasi telah ditemukan akan dmenampilkan dalam bentuk yang mudah dipahami.
- c. Adalah langkah bagian KDD yang terlibat atas pengawasan apakah pradigma telah didapatkan berlawanan dengan fakta.



**Gambar 2. 1** tahapan KDD  
 Sumber: (Ulfa & Amin, 2020)

### 2.2.2 Data Mining

*Data mining* merupakan kaidah asosiasi untuk dapat menganalisis hubungan transaksi penjualan dan menyediakan data dengan benar dan mudah di pahami sehingga bisa dijadikan dasar pengambilan keputusan. Proses analisis tersebut bertujuan untuk merancang strategi peningkatan penjualan pada apotek tersebut. Agar dapat mengetahui obat apa saja yang sering di beli oleh para konsumen maka analisis yang dilakukan berdasarkan kebiasaan konsumen dan mendeteksi menggunakan *association rule* (aturan asosiasi) yang mana proses asosiasi atau hubungan antar item data diambil dari suatu basis data rasional dan proses tersebut menggunakan *agoritma apriori* (Fajrin & Handoko, 2018).

Data mining adalah proses yang menggunakan mesin pembelajaran (*machine learning*) untuk menganalisis pengetahuan secara otomatis. Proses ini mencakup mencari pola atau tren dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Perangkat tertentu yang memiliki kemampuan untuk melakukan analisis data yang bermanfaat dan mendalam dapat mengidentifikasi pola-pola ini (Saputri et al., 2021). Pada KDD memiliki tahapan sebagai berikut:

1. *Deskriptif*

Fungsi pertama KDD adalah deskriptif, yang berfungsi untuk memahami data sehingga pengolahan dapat diteliti lebih lanjut. Tujuan dari proses deskriptif adalah untuk menemukan pola dan karakteristik dalam data, yang sebelumnya tidak dapat dilihat. Selain itu, fungsi deskriptif ini dapat digunakan sebagai bentuk yang semula tidak dapat melihat karakteristik dalam data.

## 2. *Prediktif*

Fungsi ini sangat terkait dengan proses yang digunakan untuk menemukan pola tertentu untuk data yang akan digunakan. Setelah menemukan pola untuk beberapa variabel dalam data, pola tersebut dapat digunakan untuk memperkirakan nilai variabel lain yang nilainya masih belum diketahui.

## 3. *Clasifikasi*

Dengan menggunakan fitur klasifikasi ini, beberapa pengertian tentang karakteristik kelompok data dapat dikumpulkan.

## 4. *Asosiasi*

Fungsi asosiasi ini merupakan fungsi penemuan pengetahuan dalam *database* yang dapat diproses untuk melakukan hubungan pada setiap data yang ada atau mengidentifikasi hubungan. Data yang dimaksud adalah data yang dikumpulkan sebelumnya dan saat ini.

## 5. *Forecasting*

*Forecasting* dan peramalan dapat dilakukan dengan mengumpulkan informasi yang pastinya besar. Fungsi dari forecasting ini adalah untuk mendapatkan gambaran tentang nilai data di masa mendatang.

## 6. *Clustering*

Fungsi *clustering* biasanya digunakan untuk mengidentifikasi kelompok dan produk dengan fitur khusus.

## 7. *Sequencing*

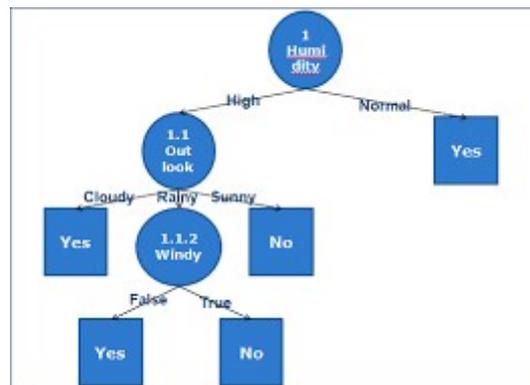
*Sequencing* memungkinkan identifikasi yang berbeda dalam jangka waktu tertentu. Ini dapat mencakup data pelanggan saat melakukan repeat berkas atau produk yang dibeli pelanggan berulang kali.

## 2.3 Algoritma Data Mining

Ada beberapa *algoritma* pada data mining yang sering dipergunakan dalam penyelesaian masalah pada data mining yaitu:

### 2.3.1 Algoritma C4.5

Algoritma C4.5 merupakan *algoritma* dari penambangan data yang digunakan untuk membangkitkan *classifier* berupa *decision tree* data yang telah diklasifikasi sebelumnya. *Classifier* merupakan alat penambangan data yang mengambil data yang perlu diklasifikasi dan mencoba memprediksi kelas data baru karena *decision tree* 4.5 sederhana maka untuk ditafsirkan dan dijelaskan cepat dan populer dari penambangan algoritma lainnya. *Algoritma* C4.5 digunakan untuk membentuk pohon keputusan yang dapat mengeksplorasi data, menemukan hubungan yang tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target.

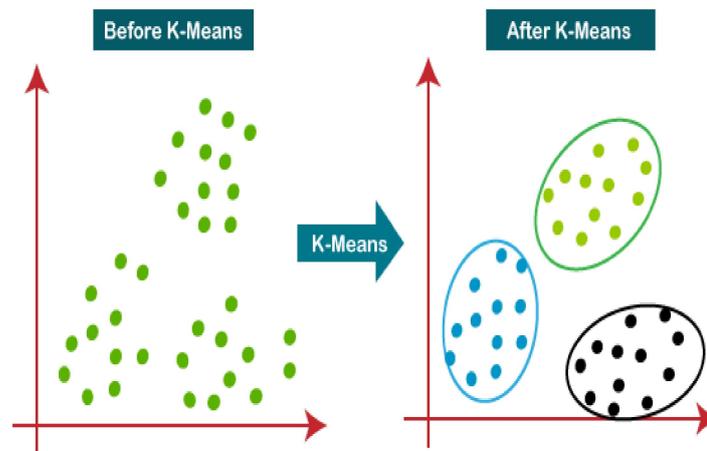


**Gambar 2. 2** Algoritma C4.5  
**Sumber:** (Ulfha & Amin, 2020)

### 2.3.2 Algoritma Kmeans

Salah satu algoritma pengajaran mesin yang paling umum digunakan untuk memecahkan masalah pengelompokan *dataset* yang belum dilabel ke dalam berbagai kluster. Pada clustering Kmeans, simbol K menunjukkan jumlah cluster yang digunakan. *Cluster* adalah titik data yang telah dikumpulkan bersama karena memiliki tipe data yang sama. Oleh karena itu, Kmeans didefinisikan sebagai algoritma iteratif yang membagi kumpulan data (*dataset*) yang tidak berlabel ke kluster yang berbeda sehingga setiap kelompok data hanya memiliki satu kelompok yang memiliki karakteristik yang sebanding. *Algoritma kmeans* melakukan dua tugas utama yaitu:

1. Penentuan nilai yang baik guna menjadi titik pusat K atau *centroid* dengan proses iteratif (perulangan)
2. Memastikan setiap titik data ke pusat K terdekat. Kemudian titik tersebut akan dibuatkan kedalam *cluster*.

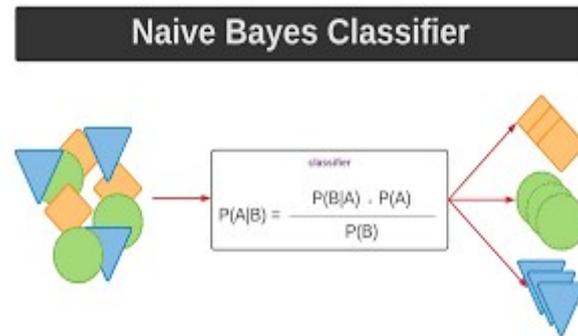


**Gambar 2. 3** *K-means*  
**Sumber:** (Saputri et al., 2021)

### 2.3.3 Algoritma *Naive Bayes*

Menurut (Surahman et al., 2023) Metode *Naive Bayes*, adalah prosedur statistik untuk menghasilkan produk probabilistik yang kuat, digunakan untuk menentukan kemungkinan atau preferensi seseorang dengan menggabungkan dan data sampling frekuensi dari kumpulan data yang tersedia. Algoritma yang menggunakan prinsip Bayes dapat meringkas setiap atribut secara mandiri atau tanpa mempengaruhi secara signifikan konsekuensi yang diberikan oleh nilai pada variabel kelas. Ketika digunakan untuk database dengan jumlah big data yang besar.

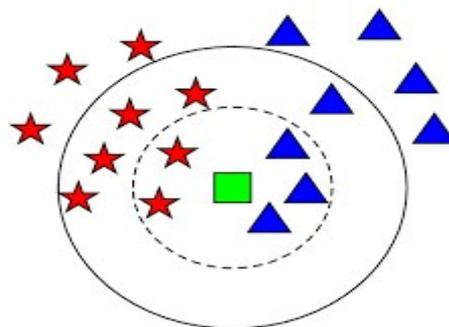
Teorema Bayes, yang menjadi dasar *Naive Bayes Classifier*, memberikan kemampuan klasifikasi yang seragam dengan *decision tree* serta *neural network*. Selain itu, *Naive Bayes Classifier* yang telah dicoba dan benar bekerja dengan sangat baik dan cepat saat digunakan dengan *database* yang berisi banyak data. Selain itu, *Naive Bayes Classifier* yang telah dicoba dan benar bekerja dengan sangat baik dan cepat saat digunakan dengan *database* yang berisi banyak data (Putri et al., 2021).



**Gambar 2. 4** Algoritma *Naive bayes*  
(Putri et al., 2021).

#### 2.3.4 Algoritma *K-Neares Neighbor*

Algoritma ini sering juga disebut dengan algoritma KNN. KNN merupakan metode yang sederhana namun memiliki tingkat akurasi yang tinggi yang bertujuan untuk mengklasifikasi objek baru dari jarak yang paling dekat *query instance ke training sample*. Algoritma KNN membuat suatu klasifikasi dari data baru yang belum di ketahui classnya dengan cara pilih sejumlah data yang terdekat (Yolanda & Fahmi, 2021).



**Gambar 2. 5** *KNN*  
Sumber:(Yolanda & Fahmi, 2021).

### 2.3.4 Algoritma Apriori

Algoritma apriori digunakan untuk membentuk kombinasi item yang mungkin dan kemudian menguji apakah kombinasi tersebut memenuhi parameter minimum dukungan dan minimum kepercayaan, yang merupakan nilai ambang yang diberikan oleh pengguna. Ini menemukan pola seperti produk yang akan sering dibeli bersamaan atau produk yang cenderung muncul bersama dalam sebuah transaksi (Saputri et al., 2021).

Jenis aturan asosiasi pada data mining disebut algoritma apriori. Selain apriori, jenis algoritma ini termasuk algoritma berbasis hash dan metode induksi umum. Aturan yang menunjukkan asosiasi antara beberapa atribut biasanya disebut analisis asosiasi atau analisis pasar keranjang. Algoritma apriori dibagi menjadi beberapa tahap yang disebut narasi. Tahapannya adalah sebagai berikut:

1. Buat daftar kandidat *itemset*. Kandidat  $k$ -itemset berasal dari kombinasi  $(k-1)$  *itemset* yang diperoleh dari iterasi sebelumnya. Dalam algoritma apriori, *subset* kandidat  $k$ -itemset yang berisi item  $k-1$  yang tidak termasuk dalam pola frekuensi tinggi dengan panjang  $k-1$  dipangkas.
2. Penghitungan *support* untuk setiap kandidat  $k$ -itemset. *Scan database* digunakan untuk menghitung jumlah transaksi yang memuat semua item kandidat  $k$ -itemset tersebut. Ini adalah ciri dari algoritma apriori, di mana penghitungan dilakukan dengan seluruh *database* sepanjang  $k$ -itemset terpanjang.
3. Tetapkan pola frekuensi tinggi. Pola ini memuat  $k$  item atau  $k$ -itemset yang memiliki *support* lebih besar dari minimum *support*.

4. Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan.

### 2.3.5 Association Rule

Metode data mining yang dikenal sebagai "aturan asosiasi" digunakan untuk mengidentifikasi pola asosiasi atau asosiasi item. Dengan menggunakan aturan asosiasi ini, misalnya, dapat menentukan seberapa besar kemungkinan seorang pelanggan akan membeli satu produk bersamaan dengan produk lainnya di suatu toko obat (Luhur et al., 2020).

1. Analisis pola frekuensi tinggi

Nilai *support* (Nilai Penunjang) merupakan presentase dari record-record yang mengandung kombinasi.

$$\text{Support } A = \frac{\sum (\text{Transaksi Mengandung } (A)) \times 100}{(\text{total transaksi } (A))};$$

2. Pembentukan Aturan Asosiasi

Akurasi dari suatu association rule sering disebut dengan *confidence* (nilai kepastian) adalah hubungan antara item dalam aturan asosiatif.

$$\text{Confidence } (A \Rightarrow B) = \frac{\text{Support } (A \text{ dan } B) \times 100}{\text{Support } A};$$

## **2.4 Objek Penelitian**

### **2.4.1 Pengertian Apotek**

Apotek sendiri merupakan tempat praktik tenaga profesional yang bekerja di farmasi, membantu meningkatkan kesehatan masyarakat. Menurut Undang-Undang Nomor 1 nomor 23 tahun 1992 tentang Kesehatan, pekerjaan kefarmasian mencakup pembuatan termasuk pengawasan kualitas sediaan farmasi, penanganan, pengadaan, penyimpanan, dan distribusi obat, pengolahan obat, dan pelayanan obat atas resep dokter. Apotek, menurut Permenkes RI nomor 1332/mankes/sk/x/2022, yang diubah dari Permenkes nomor 922/menkes/per/x/1993 mengenai ketentuan dan tata cara pemberian ijin apotek, adalah tempat di mana pekerjaan farmasi dilakukan dan obat diberikan kepada masyarakat.

### **2.4.2 Tugas dan fungsi apotek**

Menurut Peraturan Pemerintah No 51 tahun 2009 tentang tugas apotek sebagai berikut:

1. Tempat pengabdian profesi apoteker yang telah mengucapkan sumpah jabatan apoteker
2. Sarana yang digunakan untuk melakukan pekerjaan kefarmasian
3. Sarana yang digunakan untuk memproduksi dan distribusi sediaan antara lain obat, bahan baku obat, obat tradisional, kosmetik
4. Sarana pembuatan dan pengendalian mutu sediaan farmasi, pengamanan, pengadaan, penyimpanan dan pendistribusian obat.

### 2.4.3 Kualitas Pelayanan

Menurut (Luhur et al., 2020) menyatakan bahwa organisasi pelayanan kesehatan di Indonesia harus memprioritaskan penyediaan layanan kesehatan berkualitas tinggi yang berfokus pada kepuasan pasien atau konsumen. Untuk mengetahui jenis pelayanan yang dapat didefinisikan sebagai upaya untuk memenuhi kebutuhan orang lain, mutu pelayanan kesehatan harus mengacu pada keoptimalan pelayanan kesehatan 10 dalam mencapai kepuasan pasien dan konsumen. Pelayanan prima, juga disebut sebagai "pelayanan prima", adalah pelayanan terbaik yang diberikan kepada pelanggan, baik pelanggan internal maupun eksternal, berdasarkan prosedur pelayanan. Beberapa ahli telah menetapkan beberapa definisi kualitas pelayanan, menyatakan bahwa kualitas memiliki beberapa aspek (Aulia et al., 2021):

1. Harga murah, dianggap lebih rendah daripada harga yang berlaku di pasar atau di tempat lain.
2. Kelengkapan persediaan obat dan alat kesehatan, semua obat dan alat kesehatan yang tersedia dilengkapi dengan baik untuk memenuhi kebutuhan konsumen.
3. Fasilitas Sarana yang disediakan untuk memudahkan pelaksanaan fungsi dan memfasilitasi pelayanan yang baik dan nyaman bagi pelanggan.
4. Keramahtamahan Pelayanan, Kebaikan hati dan menarik budi bahasanya, manis tutur kata dan sikapnya, keakraban (dalam bergaul) dan menyenangkan dalam berinteraksi dengan pelanggan.

5. Waktu yang dibutuhkan untuk menyediakan obat dan resep kepada pelanggan, dan kecepatan penanganan obat dan resep.

## **2.5 Software Pendukung**

Merupakan *Software* yang digunakan oleh peneliti untuk mendukung penelitian yang dilakukan sehingga menemukan hasil yang ingin dicapai. Berikut di bawah ini merupakan *Software* yang digunakan:

### **2.5.1 RapidMinner**

RapidMinner yang sebelumnya dikenal sebagai YALE (*Yet Another Learning Environment*), dikembangkan oleh RalfKlinkenberg pada tahun 2001 dan merupakan salah satu alat yang digunakan untuk data mining. Ditulis dengan bahasa Java, itu dapat bekerja di semua sistem operasi. Selain itu, *RapidMiner* menggunakan GUI (*Graphic User Interface*) untuk membuat pipeline analitis. GUI membuat file XML yang menunjukkan proses analitis yang diinginkan pengguna untuk diterapkan pada data, yang kemudian dibaca oleh *RapidMiner* untuk memulai analisis secara otomatis. (Efori Buulolo, 2017).

*RapidMinner* merupakan perangkat lunak *open source* atau terbuka untuk *knowledge discovery* dan data mining. *RapidMinner* memiliki kurang lebih dari 400 operator *data mining*, termasuk operator masukan, proses, keluaran dan visualisasi. Berikut beberapa fitur dari *RapidMinner*, antar lain:

1. Bersifat *open source*
2. Multiplatform karena di program dalam bahasa java

3. Internal dan berbasis XML sehingga memudahkan dalam eksperimen
4. Dilengkapi dengan scripting language untuk otomatisasi eksperimen
5. Memiliki GUI (*Graphical User Interface*)
6. Dapat dikembangkan dengan menambahkan plugin
7. Fasilitas plotting untuk visualisasi data multidimensi dan model.



**Gambar 2. 6** Logo *RapidMiner*  
**Sumber:** (Aulia et al., 2021)

### **2.5.2** *Microsoft excel*

Tidak mengherankan bahwa *Microsoft Excel* adalah aplikasi yang banyak digunakan untuk keperluan administrasi perusahaan karena merupakan aplikasi yang memungkinkan pengolahan data otomatis dalam berbagai bentuk, seperti rumus, perhitungan dasar, pengolahan data, pembuatan tabel, dan pembuatan grafik. Karena aplikasi ini sangat lengkap dan penting untuk proses pengolahan data perusahaan, Excel juga memiliki formula yang memudahkan perhitungan dalam pengolahan data. Aplikasi ini juga sangat penting untuk membuat grafik khusus yang dibutuhkan perusahaan (Pratama and Arumi 2022).



**Gambar 2. 7** *Microsoft excel*  
**Sumber:** (Pratama and Arumi 2022).

## 2.6 Penelitian Terdahulu

Merupakan penelitian yang telah dilakukan oleh peneliti sebelumnya dan dijadikan sebagai referensi tambahan yang relevan pada pengembangan penelitian selanjutnya. Berikut beberapa penelitian yang dijadikan referensi pada penelitian ini:

1. Jurnal penelitian (Aulia et al., 2021) dengan judul “Implementasi Data Mining Menggunakan *Algoritma Apriori* Untuk Menentukan Pola Pembelian Obat di Rumah Sakit” menarik kesimpulan bahwa pengimplementasian dengan menggunakan algoritma apriori sebagai perhitungan dengan data yang sudah ada maka ditentukan bahwa obat yang paling sering terjual adalah sereplus Infusion Set Adult, IV Caath Pro 20 G, Nacl 0, Remedi Dispo 3.
2. Jurnal penelitian (Ulfha & Amin, 2020b) dengan judul “Implementasi Data Mining Untuk Mengetahui Pola Pembelian Obat Menggunakan *Algoritma Apriori*” menarik kesimpulan bahwa data mining sangat berguna untuk

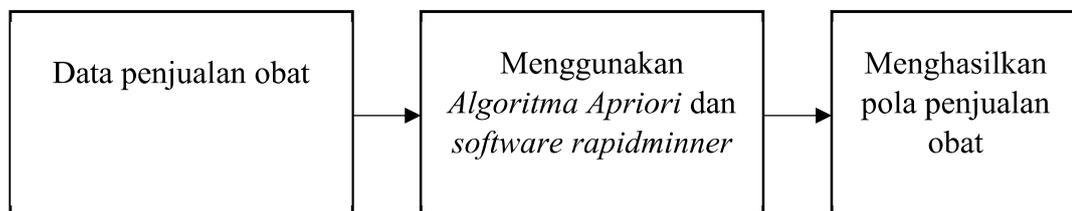
mengetahui hubungan pola frekuensi penjualan merek obat yang paling sering dibeli konsumen.

3. Jurnal penelitian (Anggraeni, Hapsari Dita, Saputra, Ragisl, Noranita, 2005) dengan judul “Aplikasi *Data Mining* Analisis Data Transaksi Penjualan Obat Menggunakan *Algoritma Apriori*” menarik kesimpulan bahwa aplikasi *data mining* menggunakan aturan asosiasi dengan *algoritma apriori* menyajikan informasi hubungan pembelian obat dengan *support* dan *confidence* tertinggi pada bulan januari.
4. Jurnal penelitian (Efori Buulolo, 2017) dengan judul “Implementasi *Algoritma Apriori* Pada Sistem Persediaan Obat” menarik kesimpulan bahwa dengan menggunakan algoritma apriori dan pengujian dengan aplikasi tanagra menghasilkan pola kombinasi itemset dan rule sebagai ilmu pengetahuan dan informasi penting dari data penjualan (data obat keluar)
5. Jurnal penelitian (Yanto & Khoiriah, n.d.) dengan judul “Implementasi Data Mining Dengan Metode *Algoritma Apriori* Dalam Menentukan Pola Pembelian Obat” menarik kesimpulan bahwa dengan metode tersebut dapat menentukan pola pola pembelian dapat dilakukan dengan melihat hasil dari kecendrungan konsumen membeli obat berdasarkan kombinasi 2 item set.
6. Jurnal penelitian (Luhur et al., 2020) dengan judul “Implemetasi *Data Mining* Menggunakan *Algoritma Apriori* Untuk Meningkatkan Pola Penjualan Obat” menarik kesimpulan bahwa penerapan *algoritma apriori* pada penjualan obat untuk melihat pola penjualan konsumen dimana pada saat pengolahan data dapat diketahui obat yang sering dibeli.

7. Jurnal penelitian (Fajrin & Handoko, 2018) dengan judul “Penerapan Data Mining Untuk Mengolah Tata Letak Buku Dengan Metode *Association Rule*” menarik kesimpulan bahwa *RapidMiner* salah satu *Software* data yang mudah digunakan dalam pengimplementasian data mining sehingga menemukan aturan *frequenci item set*.

## 2.7 Kerangka Penelitian

Merupakan upaya peneliti dalam menyusun pemikiran yang berbentuk kerangka dimulai dari input, proses hingga akhir yaitu output penelitian. Berikut dibawah kerangka penelitian:



**Gambar 2. 8** Kerangka Pemikiran