BAB II
TIJAUAN PUSTAKA

2.1 *Knowladge Discovery in Database* (KDD)

Dari beberapa referensi yang ada mengenai *knowladge discovery in database* (KDD), yaitu merupakan suatu informasi yang bersifat potensisal dalam membantu proses penerapan *data mining*.

Pada bagian bab ini akan menjelaskan tentang teori yang akan digunakan dalam penelitian ini juga metode yang digunakan dan *software* pendukung dalam pengolahan data untuk menyelesaikan suatu permasalahan yang telah disebutkan pada bab sebelumnya. Sehingga penelitian ini bisa membantu untuk memecahkan permasalahan yang terjadi pada Apotik Vitka Farma dan juga mengembangkan metode *data mining* dengan *algoritma apriori* serta mengembangkan ilmu pengetahuan bagi peneliti dan menambah wawasan bagi para pembaca.

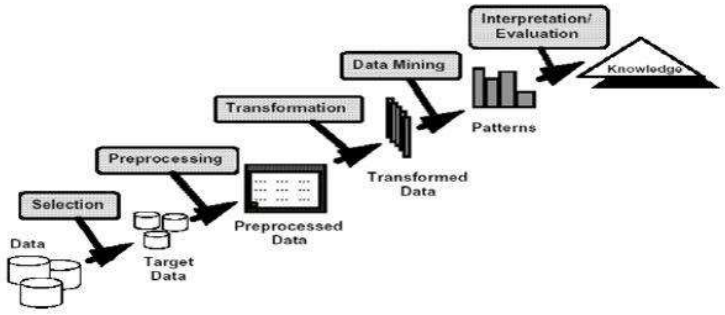
Data mining merupakan salah satu bentuk rangkaian dari *Knowledge Discovery in Database* (KDD). KDD berkaitan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari berbagai pola dalam sejumlah data. Dari rangkaian proses tersebut memiliki beberapa tahapan, yaitu (Vulandari, 2017, hal: 2-3).

1. Pembersih data (untuk membuang data yang tidak konsisten dan *noice*)
2. Integrasi data (gabungan data dari beberapa sumber)
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-mining)

4. Aplikasi teknik Data Mining, proses ekstraksi pola dari data yang sudah ada
5. Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan)
6. Persentasi pengetahuan (dengan teknik visualisasi)

Dari semua tahapan-tahapan tersebut merupakan bagian dari proses penemuan pengetahuan yang meliputi pemeriksaan informasi maupun pola yang telah didapatkan apakah berlawanan dengan fakta maupun hipotesa yang sudah ada sebelumnya. Adapun langkah terakhir dari KDD yaitu menjelaskan segala pengetahuan yang telah didapat dengan cara mempresentasikannya dalam bentuk yang mudah dipahami oleh pengguna.

Adapun proses atau tahapan-tahapan pada KDD dapat dilihat pada gambar 2.1 berikut (Nurajizah, 2019):



Gambar 2.1 Tahapan Knowladge Discovery in Database (KDD)
(Sumber: (Nurajizah, 2019)**)**

Ada pun penjelasan dari tahapan-tahapan KDD yaitu:

1. *Selection*

Penentuan data dari sebuah data operasional akan dilakukan sebelum tahap penggalian informasi di dalam KDD dimulai. Data hasil seleksi yang akan

dipergunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Prepocessing*

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses pembersihan data yang menjadi fokus dalam KDD. Proses pembersihan data meliputi proses membuang duplikasi data, memeriksa data yang inkonsistensi, dan memperbaiki kesalahan pada data, misalnya kesalahan tipografi.

3. Transformasi Data

Pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada tujuan yang ingin dicapai. Proses transformasi data yang telah dipilih, bertujuan agar data yang diperoleh sesuai untuk proses *data mining*.

4. Data mining

Proses *data mining* adalah proses mencari pola atau informasi yang menarik dalam data terpilih dengan menggunakan metode atau algoritma tertentu. Pemilihan metode atau algoritma yang tepat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Evaluation*

Proses penerjemahan pola yang dihasilkan dari *data mining*. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap *evaluation* adalah bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

2.2 *Data Mining*

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data (Vulandari, 2017, hal: 1-2). *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasikan dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Berdasarkan definisi–defenisi yang telah disampaikan dari beberapa sumber mengenai *data mining*, yaitu (Kusrini, 2009, hal: 3-4):

1. *Data mining* merupakan suatu kegiatan yang secara otomatis dalam proses pengolahan data yang sudah ada.
2. Data yang hendak di olah merupakan data yang nilainya besar.
3. Menyatukan suatu pola yang mungkin akan merberikan indikasi yang bermanfaat merupakan suatu tujuan dari *data mining*.

2.2.1 Manfaat Data Mining

Terlihat dari dua sudut pandang mengenai pemanfaatan *data mining*, yaitu sudut pandang komersial dan sudut pandang keilmuan (Vulandari, 2017, hal: 3-4).

1. Dari sudut pandang komersial, data mining guna untuk menangani meledaknya volume data, untuk bisa mendapatkan informasi yang merupakan asset untuk meningkatkan daya saing dalam institusi bisa menggunakan teknik komputasi.

Contohnya:

a. Bagaimana mengetahui hilangnya pelanggan karena pesaing
b. Bagaimana melihat konsumen atau item produk yang memiliki kesamaan karakteristik
c. Bagaimana cara mengidentifikasi produk-produk yang terjual bersamaan produk lain
d. Bagaimana cara memprediksi tingkat penjualan
e. Bagaimana menilai tingkat resiko dalam menentukan jumlah produksi suatu item
f. Bagaimana cara untuk memperkirakan tingkat bisnis dimasa mendatang

2. Dari sudut pandang keilmuan, penggunaan data mining untuk meng*capture*,

menganalisis serta menyimpan data yang bersifat real time dan sangat besar, seperti:

a. *Remot sensor* yang ditempatkan pada suatu satelit
b. *Telescope* yang digunakan untuk memindai langit
c. Simulasi saintifik yang membangkitakan data dalam ukuran *terrabytes*

2.2.2 Fungsi Data Mining

Terdapat beberapa fungsi umum yang diterapkan dalam data mining, yaitu (Vulandari, 2017, hal: 4-5):

1. *Association,* merupakan suatu prosess untuk menemukan aturan asosiasi dalam suatu waktu antara suatu kombinasi item

2. *Sequence,* proses dimana aturan asosiasi ditemukan dalam suatu waktu antara suatu kombinasi item dan diterapkan pada beberapa priode
3. *Clustering,* yaitu pengelompokan beberapa obyek/data ke dalam kelompok data sehingga setiap kelompok berisi data yang serupa
4. *Classification,* menemukan fungsi atau model yang menerangkan atau membedakan konsep atau kelas data, dengan tujuan agar bisa memprediksi kelas dari suatu objek yang labelnya tidak diketahui
5. *Regression,* merupakan suatu pemetaan data dalam penilaian prediksi
6. *Forecasting,* yaitu proses pengestimasian nilai prediksi berdasarkan pola-pola di dalam beberapa data.
7. *Solution,* merupakan proses pencarian akar masalah dan *problem solving* dari persoalan bisnis yang dihadapi atau sebagai informasi dalam pengambilan keputusan.

2.3 Metode *Data Mining*

Menurut (Kusrini, 2009)*, data mining* merupakan suatu istilah yang bisa digunakan menguraikan penemuan maupun pengetahuan didalam sebuah *database.*

Data mining terdapat beberapa konsep metode yang didasari oleh berbagai teori yang digunakan untuk melakukan pengolahan data pada suatu penelitian. Data *mining* memiliki beberapa metode yang bisa digunakan dalam menyelesaikan permasalahan dalam *data mining.*

2.3.1 Algoritma C4.5

Algoritma C4.5 merupakan suatu algoritma yang membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan (Kusrini, 2009).

2.3.2 Algoritma Nearest Neighbor

Nearest Neighbor adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada. Misalkan diinginkan untuk mencari solusi terhadap seorang pasien baru dengan menggunakan solusi dari pasien terdahulu (Kusrini, 2009, hal: 93).

2.3.3 Fuzzy C-Means

Analisis kluster atau *clustering* merupakan proses membagi data dalam suatu himpunan kedalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar dari pada kesamaan data tersebut dengan data dalam kelompok lain. Potensi *clustering* dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas

seperti klasifikasi, pengolahan gambar, dan pengenalan pola (Kusrini, 2009, hal: 177).

2.3.4 Bayesian Classification

Bayesian Classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan atau *class. Bayesian classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* (Kusrini, 2009, hal: 189).

2.3.5 Backpropagation

Backpropagation adalah salah satu algoritma pembelajaran dalam jaringan saraf tiruan. Proses pembelajaran dalam *backpropagation* dilakukan dengan penyesuaian bobot-bobot jaringan saraf tiruan dengan arah mundur berdasarkan nilai *error* dalam proses pembelajarannya. Jaringan saraf tiruan itu sendiri merupakan sebuah model yang mengadopsi cara kerja neuron secara biologi dengan fokus pada cara kerja saraf otak (Kusrini, 2009, hal: 199).

2.3.6 *Algoritma Apriori*

Dari beberapa metode *data mining* yang telah disampaikan diatas, maka salah satunya yang akan digunakan oleh peneliti dalam menyelesasaikan hasil dari pada penelitian ini adalah *algoritma apriori.*

Apriori merupakan algoritma yang dipakai pada saat mencari *frequent itemset* agar bisa memperoleh aturan asosiasi. Sesuai dengan namanya, algoritma ini menggunakan *prior knowledge* mengenai *frequent itemset properties* yang sudah diketahui terdahulu untuk memproses informasi yang akan datang. *Apriori* menggunakan pendekatan secara *iterative* seperti yang dikatakan sebagai *level-wish search* dimana k-*itemset* digunakan untuk mencari (k+1)-*itemset* (H. D. dkk Anggraeni, 2016).

Algoritma Apriori termasuk jenis aturan asosiasi pada *data mining.* Selain *apriori*, yang termasuk dalam golongan ini adalah metode *Generalized Rule Induction* dan *Algoritma Hash Based.* Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market basket analysis* (Kusrini, 2009, hal: 149).

Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menentukan aturan asosiatif antara suatu kombinasi *item.* Contoh aturan asosiatif dari analisis pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut, pemilik toko dapat mengatur penempatan barang untuk pemasaran dengan memakai kupon diskon untuk kombinasi barang

tertentu. Karena itu analisis asosiasi menjadi terkenal oleh aplikasinya untuk menganalisis isi keranjang belanja di pasar swalayan, analisis asosiasi juga sering disebut *market basket analysis.*

Analisis asosiasi dikenal juga sebagai salah satu teknik *data mining* yang menjadi dasar dari berbagai teknik *data mining* lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (*frequent pattern mining*). Penting atau tidaknya suatu aturan asosiatif dsapat diketahui dengan dua parameter, yaitu *support* dan *confidence. Support* (nilai penunjang) adalah persentase kombinasi *item* tersebut dalam database, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar-*item* dalam aturan asosiasi.

Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiasi yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap:

1. Analisis pola frekuensi tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam database. Nilai *support* sebuah *item* diperoleh dengan rumus berikut.

$$Support\ (A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ Transaksi}$$

Rumus 2.1 Nilai *Support* satu *item*

Sementara itu, nilai *support* dari 2 *item* diperoleh dari rumus 2 berikut.

$$Support\ (A,\ B) = P\ (A \cap B)$$

$$Support\ (A,B) = \frac{\sum Jumlah\ transaksi\ mengandung\ A\ dan\ \ B}{\sum Transaksi}$$

Rumus 2.2 Nilai *Support* dua *item*

2. Pembentukan Aturan Asosiasi

Setelah semua pola frekuensi tunggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif A → B. Nilai *confidence* dari aturan A → B diperoleh dari rumus berikut.

$$Confidence = P(A \mid B) = \frac{\sum Jumlah\ transaksi\ mengandung\ A\ dan\ \ B}{\sum Transaksi\ mengandung\ A}$$

Rumus 2.3 Nilai *confidence*

2.4 *Software* Pendukung

Software merupakan suatu aplikasi yang merupakan salah satu bagian dari sistem komputer yang apabila diperlukan oleh pengguna untuk membantu menyelesaikan suatu perintah atau tugas yang akan dikerjakan. Pada penelitian ini *software* yang digunakan oleh peneliti untuk membantu penyelasaian tugas akhir ini adalah WEKA.

WEKA merupakan sebuah paket tools machine learning yang praktis. *Waikato Environment for Knowledge Aanalysis* yang disingkat dengan WEKA, dibuat di Universitas Waikato, New Zealand untuk penelitian, pendidikan dan

berbagai aplikasi. Salah satu contoh penggunaan WEKA adalah dengan menerapkan sebuah metode pembelajaran ke dataset dan menganalisa hasilnya untuk memperoleh informasi tentang data, atau menerapkan beberapa metode dan membandingkan performansinya untuk dipilih (Vulandari, 2017, hal: 79-80).

Pengembangan WEKA mengikuti model releases Linux: digit kedua yang genap menunjukkan release yang stabil dan digit kedua yang ganjil menunjukkan release 'penembangan' (misalnya 3.0.x adalah release stabil, sedangkan 3.1.x adalah release yang sedang dikembangkan). WEKA memiliki beberapa versi diantaranya yaitu:

1. WEKA 3.0 : "versi buku" yang sesuai dengan deskripsi buku data mining
2. WEKA 3.2 : "versi GUI" yang menambahkan GUI dari CLI awal.
3. WEKA 3.3 : "versi pengembangan" dengan berbagai peningkatan.

2.5 Penelitian Terdahulu

Untuk pedoman dan sebagai bahan pertimbangan pada penelitian ini, untuk itu peneliti mencantumkan beberapa jurnal ilmiah *data mining* yang diambil sebagai referensi pada penelitian ini, yaitu:

1. Penelitian yang dilakukan oleh (Ulvah, 2018) dengan judul **"Implementasi Algoritma Apriori Aturan Keterkaitan Data Untuk Analisis Keranjang Belanja Sistem Persediaan Obat Pada Apotek Perdos Farma Makassar"** penelitian ini menjelaskan tentang meningkatkan sistem informasi persediaan obat pada apotek perdos farma makassar, dengan melakukan analisa

keranjang belanja menggunakan metode asosiasi dengan mengimplementasi *algoritma Apriori* didalamnya.

2. Penelitian yang dilakukan oleh (Nurajizah, 2019) dengan judul **"Analisa Transaksi Penjualan Obat Menggunakan Algoritma Apriori"** yang menjelaskan tentang peningkatan data transaksi yang cukup signifikan, tidak akan bernilai jika data tidak diolah menjadi sesuatu yang memiliki nilai lebih. Teknik data mining dapat menjadi salah satu cara untuk mengolah histori data transaksi penjualan yang bersumber dari sebuah dataset yang berukuran besar. Dalam penelitian ini, data mining yang digunakan adalah teknik asosiasi dengan menggunakan algoritma apriori menggunakan nilai minimum support sebesar 30% serta nilai minimum *confidence* sebesar 60%.
3. Penelitian yang dilakukan oleh (Surti Kanti & Richardus Eko Indrajit, 2017) dengan judul **"Implementasi Data Mining Penjualan Handphone Oppo Store SDC Tanggerang Dengan Algoritma Apriori"** penelitian ini menerangkan permasalahan yang terjadi pada Oppo store SDC yang belum memakai implementasi data mining, dimana barang yang sudah menjadi stock di showroom harus terjual semua tidak dapat dikembalikan ke pusat.

   Penelitian ini melakukan analisa data dengan menggunakan data mining dan metode algoritma appriori. Dengan metode tersebut dapat diketahui produk item yang paling banyak terjual, sehingga showroom oppo store SDC dapat menyusun strategi pemasaran untuk memasarkan produk dengan tipe lain dengan meneliti apa kelebihan produk yang paling banyak terjual tersebut

dengan produk lainnya dan dapat menambah persediaan stock di Oppo Store SDC.

4. Penelitian yang dilakukan oleh (Hapsari Dita Anggraeni, Ragil Saputra, & Beta Noranita, 2016) dengan judul **"Aplikasi *Data Mining* Analisis data Transaksi Penjualan Obat Menggunakan *Algoritma Apriori*"** yang menjelaskan tentang bahwa apotek merupakan suatu usaha industri yang menghasilkan data penjualan obat setiap hari, namun data penjualan tersebut belum dapat dimaksimalkan pemanfaatannya, karena data penjualan hanya disimnpan tanpa dilakulkan analisis lebih lanjut.
5. Penelitian yang dilakukan oleh (Indra Griha Tifik Isa & Dicky Jhoansyah, 2019) dengan judul **"Implementasi *Association Rules* Dalam Menentukan Posisi Gerobak *FoodCourt*"**, penelitian ini menerangkan tentang penggunaan metode dalam meningkatkan kegunaan data yang dihasilkan. Data yang akan di ekstraksi pada menelitian ini yaitu data penjualan *Foodcourt* yang menjadi sumber informasi yang berguna untuk manajemen yaitu rekomendasi tata letak gerobak yang menjadi strategi pemasaran dalam penjualan.
6. Penelitian yang dibuat oleh (Uci Beatulloh, Acep Irham Gufroni, & Rianto 2019) dengan judul **"Penerapan Metode *Association Rule* Mining Pada Data Transaksi Penjualan Produk Kartu Perdana Kuota Internet Menggunakan Algoritma Apriori"** yang menjelaskan tentang peningkatan penjualan produk yang lebih baik khususnya pada kartu perdana kuota internert, produka mana saja yang paling diminati di wilayah pemasaran dan

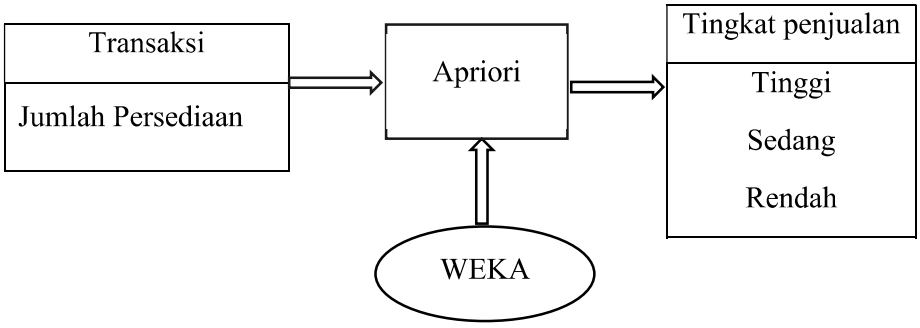
penjualan Priangan Timur meliputi *cluster* Ciamis, Garut dan Tasikmalaya. Dari ketiga wilayah tersebut produk yang paling diminati dipasaran *outlate* adalah produk dari operator kartu peradana kuota internet XL dengan Telkomsel dan produk Indosat dengan Telkomsel.

7. Penelitian yang dilakukan oleh (Irsyad Djamaludin, & Agus Nursikuwagus 2017) dengan judul **"Analisis Pola Pembelian Konsumen Pada Transaksi Penjualan Menggunakan Algoritma Apriori"** penelitian ini menjelaskan tentang untuk mendapatkan pola pembelian dan penjualan pada suatu barang yang dilakukan dengan analisis suatu pola pada algoritma apriori.
8. Penelitian yang dilakukan oleh (Agus Nursikuwagus, & Tono Hartono, 2016) dengan judul **"Implementasi Algoritma Apriori Untuk Analisis Penjualan Dengan Berbasis *Web"*** yang menjelaskan tentang bagaimana cara memberikan laporan penjualan yang pada penelitian ini di dapat dengan melakukan implementasi menggunakan algoritma apriori pada data transaksi untuk analis penjualan suatu produk.
9. Penelitian yang dilakukan oleh (Allfanisa Annurrullah Fajrin & Handoko, 2018) dengan judul penelitian yaitu **"Penerapan Data Mining Untuk Mengolah Tata Letak Buku Dengan Metode *Association Rule"*** yang menjelaskan tentang pengelolaan tata letak suatu buku agar mudah ditemukan oleh pengguna saat mencari suatu buku yang dibutuhkan.
10. *Research conducted by (Xioali Wang, Chen Song, Wei Xiong, and Xingxiao Lv, 2018) by title **"Evaluation of Floatation Working Condition Recognition Based on An Improved Apriori Algorithm"** that explains about*

A priori algorithm based on time series is proposed for flotation working conditions recognition. Based on the transaction database idea, a collection of time conditions data from working conditions is divided into a database of working conditions transactions with an auto-regressive model. Then it is used to find association rules that contain relationships between time series database of working conditions transactions to form a rule base for working conditions, which provides evaluation of the reliability of the results of recognition of online working conditions based on association rules.

2.6 Kerangka Pemikiran

Kerangka pemikiran adalah suatu gambaran bagaimana cara peneliti dalam memecahkan suatu permasalahan yang ada khususnya pada Apotik Vitka Farma Batam. Ada pun kerangka pemikirannya, yaitu:



Gambar 2.2 Kerangka Pemikiran
(Sumber: Data Olahan Peneliti, 2019**)**

Dari kerangka pemikiran tersebut terdapat variable *input* yang kemudian akan diolah atau diproses dengan metode apriori dan akan diuji dengan *software* pendukung yaitu WEKA, yang akhirnya bisa diketahui *output* hasil *item set* dalam menentukan tingkat penjualan obat pada Apotik Vitka Farma Batam.

2.7 Hipotesis

Dengan dilakukannya implementasi *algoritma apriori* pada Apotik Vitka Farma dari olahan data transaksi penjualan obat agar bisa memberikan hasil yang diharapkan terhadap tingkat penjualan obat pada apotek tersebut.