

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Tinjauan Teori Dasar**

Pada bab ini peneliti menjelaskan mengenai teori-teori yang berkaitan dan akan digunakan pada penelitian ini.

##### **2.1.1. Tinjauan Teori Umum**

Teori umum merupakan teori pernyataan jika suatu pernyataan nilainya benar maka akan bernilai benar secara universal. Pernyataan ini berlaku di semua tempat, waktu, keadaan dan permasalahan. Pernyataan ini digunakan sesuai dengan penelitian yang akan dilakukan.

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

*Knowledge Discovery in Database (KDD)*, merupakan kegiatan yang mencakup pengumpulan, pemakaian, data yang telah tersimpan untuk mendapatkan keteraturan, pola dalam set data yang berukuran besar. Hasil dari data mining dimanfaatkan untuk memperbaiki penentuan keputusan (Putria, 2018).

Menurut (Ristekdikti et al., 2020) proses KDD merupakan proses yang dilakukan untuk mengambil informasi yang terdapat dalam database, dimana proses tersebut terdapat proses data mining.

Proses tahapan-tahapan data *mining* berdasarkan proses KDD secara garis besar sebagai berikut:

1. *Data selection*

Memisahkan dan memilih data dari sekumpulan data operasional berdasarkan kriteria yang diinginkan peneliti, seperti data berdasarkan yang ada pada konsumen.

2. *Procesing/cleaning*

Sebelum proses data mining dapat dilaksanakan, langkah yang dilakukan terlebih dahulu adalah proses *cleaning*. Proses *cleaning* yaitu proses yang meliputi pemeriksaan dan inkonsisten, menghilangkan duplikasi data, dan memperbaiki kesalahan pada data.

3. *Transformation*

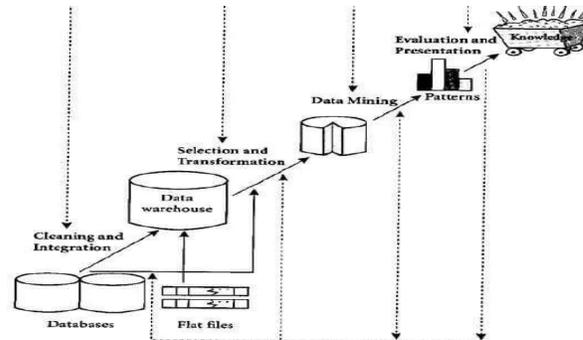
Pada proses coding dilakukan perubahan yang dipilih agar data sesuai dengan algoritma apriori. Proses *coding* yang ada dalam KDD (*Knowledge Discovery Database*) bergantung pada jenis dan pola informasi yang dicari didalam basis data.

4. *Data mining*

*Data mining* merupakan proses mencari pola dan informasi dalam data yang terpilih dengan menggunakan suatu metode tertentu. Dalam Data mining, teknik dan metode sangat bervariasi. Pemilihan teknik dan metode yang tepat sangat bergantung pada proses KDD (*Knowledge Discovery Database*) secara keseluruhan.

### 5. *Interpretation/ Evaluation*

Informasi yang dihasilkan dari data mining harus ditampilkan dalam bentuk yang mudah dipahami oleh pihak yang berhubungan. Tahap ini disebut juga tahap *interpretation*.



**Gambar 2. 2** Proses KDD

**Sumber:** (Putria, 2018)

#### 2.1.1.2. *Data Mining*

Data mining adalah kegiatan menambang pengetahuan dari data yang berjumlah besar, dengan cara mengekstrak data tersebut untuk menemukan pola tertentu dengan menganalisis untuk menghasilkan suatu informasi yang berguna (Astuti, 2019).

Menurut (Erwansyah, 2019) data mining merupakan proses untuk menggali nilai tambahan yang berupa informasi yang tidak diketahui oleh basis data dengan cara mengekstraksi pola yang ada didalam basis data.

Menurut (Budiyati et al., 2020) data mining digunakan untuk menguraikan penemuan dalam *database*. *Data mining* merupakan proses

yang menggunakan teknik statistik, matematika dan kecerdasan buatan dan *machine learning*.

Data mining merupakan proses untuk menemukan pola-pola yang ada dalam data. Proses ini otomatis atau semiotomatis. Dimana pola tersebut dapat memberi keuntungan secara ekonomi. Dan data yang digunakan adalah data yang berjumlah besar (Tumini, 2020)

Data mining sangat diperlukan terutama dalam mengelola data yang sangat besar untuk memudahkan aktifitas recording dari suatu transaksi dan proses data warehousing untuk dapat memberikan suatu informasi yang akurat bagi pengguna data mining (Takdirillah, 2020).

#### **2.1.1.3. *Pengelompokan Data Mining***

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

1. Deskripsi (*Description*)

Pada proses deskripsi Terkadang peneliti ingin menemukan data yang dapat memperlihatkan gambaran pola dan kecendrungan dari data yang ada. Sebagai contoh, masyarakat tidak bias mendapatkan fakta bahwa siapa pemerintah yang melakukan korupsi akan mendapatkan hukuman.

2. Estimasi (*Estimation*)

Estimasi hampir sama dengan klasifikasi, yang membedakan keduanya adalah variable target estimasi lebih kearah *numeric* dari pada kearah

kategori. Model dibangun *record* lengkap yang menyediakan nilai dari variable target sebagai nilai acuan.

### 3. Prediksi (*Prediction*)

Tujuan proses prediksi dilakukan adalah untuk memperkirakan suatu nilai dari sebuah hasil yang memiliki kemungkinan muncul pada waktu yang akan datang. Sebagai contoh prediksi adalah prediksi harga beras dalam dua bulan yang akan datang.

### 4. Klasifikasi (*Classification*)

Klasifikasi adalah proses mendapatkan model untuk menggambarkan data dan membaginya kedalam kelas-kelas yang telah ditetapkan. Proses ini dilakukan untuk mengetahui pembagian data-data sesuai dengan tingkat kepentingannya. Pada masa yang akan datang.

### 5. Pengklusteran (*Clustering*)

Pengklusteran merupakan membentuk kelompok data, pengamatan, atau untuk membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah sekumpulan data yang memiliki kemiripan satu dengan lainnya dan memiliki ketidakmiripan dengan dengan data dikluster lainnya.

### 6. Asosiasi

Dalam hal ini tugas asosiasi meliputi penemuan atribut yang keluar secara bersamaan. Dalam dunia perdagangan asosiasi dikenal dengan *Market Basket Analysis* atau analisa keranjang belanja. Contoh dari asosiasi adalah menemukan barang dalam supermarket yang dibeli

secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

#### **2.1.1.4. Tahapan-Tahapan Data Mining**

Menurut (Romadhon & Kodar, 2020) data *mining* dibagi menjadi beberapa tahapan yaitu:

1. Pembersihan Data

Suatu proses untuk membuang data yang tidak valid dan yang tidak sesuai untuk digunakan.

2. Integrasi Data

Proses menggabungkan data ke dalam satu *database* yang akan digunakan.

3. Seleksi Data

Data yang diambil harus sesuai dengan hasil analisis.

4. Transformasi Data

Data yang ada diubah menjadi bentuk yang tepat sesuai *data mining*.

5. Proses *Mining*.

Proses awal dalam metode pengkajian.

6. Evaluasi Pola

Proses untuk mengidentifikasi pola yang menarik untuk mewakili pengetahuan dari data tersembunyi.

7. Presentasi Pengetahuan.

Merupakan proses presentasi pengetahuan dan teknik visualisasi dapat membantu memberitahukan hasil data mining kepada pengguna.

#### 2.1.1.5. *Manfaat Data Mining*

Menurut (Erwansyah, 2019) manfaat data mining dilihat dari dua sudut yaitu:

##### 1. Sudut Pandang Komersial.

Dari sudut pandang komersial data *mining* bisa memanfaatkan untuk mengolah volume data yang berlebihan, dengan memakai teknik komputasi yang digunakan untuk memperoleh informasi dan hasil informasi tersebut merupakan sebuah untuk meningkatkan daya saing.

##### 2. Sudut Pandang Keilmuan.

Dari sudut pandang keilmuan, data *mining* digunakan untuk menganalisis data dan menyimpan data yang besar.

#### 2.1.1.6. *Algoritma Apriori*

Menurut (Arifin, 2020) Algoritma apriori merupakan suatu algoritma dasar yang dijadikan suatu usulan untuk menentukan *frequent itemsets* untuk asosiasi *Boolean*. Algoritma apriori termasuk asosiasi pada data mining. Aturan tersebut disebut *affinity analysis* atau *market basket analysis*. Algoritma apriori atau disebut juga dengan analisis asosiasi (*association rule mining*) merupakan teknik data mining untuk menemukan aturan asosiasi dalam suatu kombinasi item.

Ada beberapa tahap dalam mencari *association rule* dari suatu kumpulan data, tahap pertama adalah mencari *frequent itemset*

dahulu. *Frequent itemset* adalah kumpulan *item-item* atau barang-barang yang sering keluar secara bersamaan. Dalam mencari asosiasi dapat dilihat dari hasil *support dan confidence*. *Support* adalah sebuah ukuran dalam mengetahui berapa besar nilai kemunculan data *itemset* dari total keseluruhan data transaksi. *Confidence* atau tingkat kepercayaan adalah sebuah tolak ukur dalam menunjukkan keterkaitan dari satu *item* dengan item lainnya secara *conditional* (Putria, 2018).

Menurut (Zulfa et al., 2020) ada dua proses utama yang dilakukan dalam algoritma apriori, yaitu:

1. *Join* (penggabungan)

Proses ini bertujuan untuk menggabungkan *item* yang ada dengan *item* lainnya agar tidak muncul suatu kombinasi baru.

2. *Prune* (pemangkasan)

Proses yang menentukan nilai *minimum support* yang bertujuan untuk mengurangi jumlah *item* yang ada didalam kombinasi.

#### **2.1.1.7. Association Rule**

*Association rule* adalah suatu metode yang memiliki tujuan dalam memperoleh pola yang ada diberbagai transaksi. Banyaknya *item* yang dapat mendukung dari sebuah system rekomendasi melalui metode dalam menemukan pola antar *item* dalam transaksi. Contoh dari suatu aturan asosiasi dapat dilihat dari analisis pembelian produk disuatu swalayan yang dimana dapat diketahui dengan melihat seberapa besar

kemungkinan sebuah barang yang dibeli secara bersamaan dalam satu transaksi (Ikhwan et al., 2018)

*Association Rule* merupakan teknik yang bekerja dalam mencari suatu hubungan antar *item* yang didapatkan secara bersamaan. Tugas dari *association rule* adalah mencari aturan untuk mengukur hubungan lebih dari dua atribut (Vadilah et al., 2020).

Menurut (Arifin, 2020) *association rule* merupakan teknik *data mining* untuk mendapatkan sebuah aturan dari kombinasi *item*. Adapun alasan mengapa sebuah analisis asosiasi banyak diminati para peneliti karna dapat menghasilkan algoritma yang efisien melalui analisis pola frekuensi tinggi (*frequent pattern mining*).

*Association rule* atau asosiasi merupakan teknik data mining untuk mendapatkan aturan asosiasi yang terbentuk antara kombinasi *item*. Adapun metodologi dasar *association rule* sebagai berikut (Santoso et al., 2016)

#### 1. Analisis Pola Frekuensi Tinggi

Tujuan dari pola ini adalah untuk menemukan perpaduan dari *item* yang ada dalam basis data yang nilai *support-nya* memenuhi syarat *minimum*. Untuk mencari nilai *support item* dapat menggunakan rumus berikut :

$$\text{Support A} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}}$$

Untuk nilai *support* dari 2 item atau lebih diperoleh dengan rumus berikut:

$$\text{Support (A,B)} = \frac{\sum \text{Mengandung A dan B}}{\sum \text{Transaksi}}$$

**Rumus 2. 1** perhitungan support

**Sumber :** (Santoso et al., 2016)

## 2. Pembentukan Aturan Asosiasi

Setelah nilai *support* pada pola frekuensi tinggi didapatkan, dilanjutkan dengan mencari aturan asosiasi yang nilai *confidence* memenuhi syarat *minimum*. Nilai *confidence* dari aturan “jika A maka B” dapat menggunakan rumus berikut:

$$\text{Confidence P(B | A)} = \frac{\sum \text{Transaksi Mengandung A dan B}}{\sum \text{Transaksi Mengandung A}}$$

**Rumus 2. 2** pembentukan aturan asosiasi

**Sumber:** (Santoso et al., 2016)

### 2.1.2 Tinjauan Teori Khusus

Teori khusus merupakan teori yang memiliki keterkaitan dengan sejumlah fakta-fakta yang sifatnya particular.

#### 2.1.2.1. Tanagra

Aplikasi Tanagra adalah perangkat lunak bebas untuk tujuan akademik dan penelitian, yang mana penelitian ini melibatkan metode data mining yaitu analisis eksplorasi data, pembelajaran statistic, pembelajaran mesin, dan basis data (Rezki, 2020)

Menurut (Badrul, 2016) Tanagra merupakan software *open source* yang setiap orang bisa mengakses langsung pada kode *source* program dan melakukan pembahan algoritma yang sesuai izin dari perangkat lunak. Tanagra memiliki tujuan yaitu:

1. Tanagra bertujuan untuk memberikan kemudahan untuk peneliti dan mahasiswa dalam mengakses perangkat lunak *data mining*.
2. Tanagra bertujuan dalam memberikan usulan kepada peneliti agar dapat menambahkan metode data *mining* pada data mereka dan dijadikan sebagai tolak ukur kinerja mereka.
3. Tanagra ditunjukkan kepada pengembangan pemula, dimana dalam menyebarkan metodologi yang digunakan dalam membuat *software* dengan jenis yang sama. Mereka juga bisa menghasilkan keuntungan dalam mengakses secara gratis ke kode sumber ini, yang digunakan untuk mengetahui bagaimana perangkat lunak ini dibuat.

#### **2.1.2.2. Penjualan**

Penjualan adalah salah satu syarat berlangsungnya sebuah usaha, dengan Penjualan melakukan proses penjualan usaha tersebut akan mendapatkan sebuah keuntungan. Semakin tinggi proses jual beli yang terjadi maka keuntungan yang didapatkan akan berbanding lurus (Gusrizaldi & Komalasari, 2016).

Menurut (Gusrizaldi & Komalasari, 2016) pasar pokok penjualan merupakan suatu ilmu dan seni untuk memberikan pengaruh terhadap

pola penjualan untuk menarik perhatian orang agar membeli barang yang ditawarkan.

Menurut (Gusrizaldi & Komalasari, 2016) ada beberapa factor yang memberikan pengaruh terhadap penjualan yaitu:

1. Kondisi Pasar

Pasar adalah tempat terjadinya proses penjualan antara penjual dan pembeli. Faktor kondisi pasar yaitu jenis kelompok pasar pembeli, segmen pasar, daya beli, frekuensi pembelian, dan kebutuhannya sehingga penjual dapat menghasilkan pemasukan yang maksimal.

2. Kondisi dan Kemampuan Penjual

Transaksi jual beli melibatkan dua pihak yaitu pihak pertama sebagai penjual dan pihak kedua sebagai pembeli. Oleh karena itu, pihak pertama atau penjual harus bisa meyakinkan pembeli agar penjual bisa mencapai sasaran penjualan yang diinginkan.

3. Modal

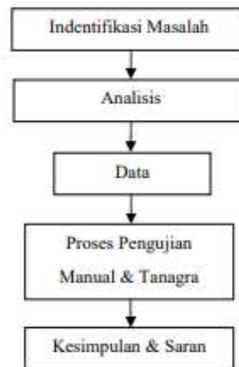
Untuk memulai sebuah usaha penjual memerlukan modal untuk membeli produk yang akan diperjualkan lagi melalui sarana usaha seperti biaya, transportasi, biaya mempromosikan barang, dan biaya lainnya.

4. Kondisi Organisasi Perusahaan

Pada perusahaan besar proses penjualan akan ditangani oleh bagian marketing perusahaan atau pihak yang ahli dibidang

penjualan. Lain halnya di perusahaan kecil atau usaha kecil dimana proses penjualan ditangani oleh pihak yang juga mengurus banyak hal dalam perusahaan atau usaha tersebut.

## 2.2 Kerangka Pemikiran



**Gambar 2. 3** kerangka pemikiran

**Sumber:** (Data Peneliti 2020)

Penjelasan dari kerangka pemikiran pada *flowcart* diatas:

1. Identifikasi masalah merupakan proses mendefinisikan problem atau masalah yang ada pada penelitian.
2. Analisis masalah, pada tahap ini peneliti menentukan ruang lingkup masalah dan menguraikan masalah menjadi sederhana.
3. Dalam mengumpulkan data yang dibutuhkan peneliti, peneliti harus melakukan proses observasi langsung di Toko Berkah Mart sehingga dapat diketahui masalah yang terjadi. Melakukan pengolahan data dari apa yang didapatkan dari data transaksi untuk mendapatkan seberapa besar kemunculan *itemset* dalam

memanfaatkan teknik *association rule* dan dilakukan proses pengujian data menggunakan *software* Tanagra.

4. Pengujian manual dan Tanagra, padatahap pengujian manual dilakukan dengan menghitung data menggunakan teknik *association rule* dan data hasil penghitungan manual dilakukan pengujian untuk mengetahui kesamaan antara kedua data yang telah didapatkan.
5. Kesimpulan dan saran, adalah poin akhir pada penulisan skripsi yang dimana menentukan hasil dari penelitian dan memberikan saran kepada penelitian selanjutnya supaya dapat mengembangkan ilmu ini.

### **2.3 Hipotesis Penelitian**

Hipotesis penelitian berdasarkan kerangka pemikiran diatas adalah sebagai berikut:

1. Diduga proses analisa *data mining* pada transaksi data penjualan dengan algoritma apriori untuk mengetahui pola pembelian konsumen.
2. Di duga nilai *support* dan *confidence* yang terbentuk sehinga dapat menentukan pola pembelian konsumen .

### **2.4 Penelitian Terdahulu**

1. Hasil penelitian yang dilakukan (Alfiyan et al., 2019) Dalam jurnal yang berjudul “Analisis Market Basket Dengan Algoritma Apriori Pada Transaksi Freshfood”. Dengan kesimpulan Algoritma apriori

digunakan dalam membentuk pola kombinasi *itemset*. Hasil dari teknik *association rule* yang menggambarkan hubungan antar *item* produk yang berada dalam kombinasi *itemset* yang sama.

2. Hasil penelitian yang dilakukan (Rezki, 2020) Dalam jurnal yang berjudul “Analisis Kinerja Guru Sekolah Dinas Pendidikan Dengan Menggunakan Metode Rought Set (Studi Kasus: Dinas Pendidikan Sumatra Utara)”. Dengan kesimpulan Metode Rought Set yang dirancang mampu menganalisa kinerja guru PNS di SMA Negeri 1 Medan dari data hasil analisa kinerja guru terakomodir dengan baik sehingga data ini dapat mudah dilacak demi kepentingan guru yang bersangkutan maupun untuk pihak Dinas Pendidikan.
3. Hasil penelitian yang dilakukan (Anggraini et al., 2020) dalam jurnal yang berjudul “Implementasi Algoritma Apriori Dalam Menentukan Penjualan Mobil Yang Paling Diminati Pada Honda Permata Serpong”. Dengan kesimpulan Algoritma Apriori yang diterapkan dapat mempercepat proses pembentukan kecendrungan pola kombinasi *itemset*, dengan hasil analisa penjualan pada Honda Permata Serpong dapat mengatur stok barang untuk pemasaran.
4. Hasil penelitian yang dilakukan (Ikhwan et al., 2018) dalam jurnal yang berjudul “*A Novelty of data Mining for Promoting Education Based on Fp-Growth Algoritma*”. Dengan kesimpulan Dalam mencari *frequent itemset* dapat diketahui dengan menyusun pohon keputusan dengan memanfaatkan algoritma *FP-Growth* dengan memperoleh

suatu asosiasi yang baru dari data siswa baru. Variable data sangat penting dalam menentukan akurasi dari suatu *itemset* yang saling berhubungan dengan menggunakan strategi promosi pendidikan.

5. Hasil penelitian yang dilakukan (Badrul, 2016) Dalam jurnal yang berjudul “Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa data penjualan”. Dengan kesimpulan Algoritma apriori digunakan dalam strategi marketing pada perusahaan atau institusi lainnya. Proses pelaksanaan strategi ini dilakukan dengan cepat karena menggunakan algoritma apriori yang diharapkan lebih efektif.
6. Hasil penelitian yang dilakukan (Budiyati et al., 2020) Dalam jurnal yang berjudul “Implementasi Metode Algoritma Apriori Untuk Penempatan Buku Pada Rak Perpustakaan STMIK Jakarta STI&K”. dengan kesimpulan algoritma apriori yang menghasilkan *association rule* untuk menghitung nilai *support* dan *confidence* dari data transaksi peminjaman buku dan hasil perhitungan dapat digunakan untuk penempatan buku pada rak buku diperpustakaan STMIK Jakarta STI&K.
7. Hasil penelitian yang dilakukan (Andini et al., 2020) dalam jurnal yang berjudul “Implementasi Algoritma K-Medoids untuk klasterisasi data penyakit pasien di RSUD Kota Bandung”. Dengan kesimpulan dari hasil *cluster* dapat informasi penyakit yang sering terjadi berdasarkan dari kamar dan jumlah pasien sehingga sehingga dapat membantu

pihak rumah sakit dalam mengambil Tindakan kebijakan dalam antisipasi pengobatan serta pencegahan penyakit.