

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Teori Umum

Dalam bagian ini, akan menjelaskan konsep penting dan terkait dari penelitian ini.

2.1.1. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah pengumpulan pengetahuan dari informasi yang ada. Basis data terdiri dari tabel yang ditautkan. Hasil dari pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai pengetahuan dalam pengambilan keputusan. (Mardi, 2017). Pada dasarnya menurut (Alfiyan et al., 2019) ada enam elemen yang dalam teknik pencarian informasi/pengetahuan dalam KDD yaitu:

1. Bekerja pada sejumlah data tertentu.
2. Harus ada dampak pada nilai data.
3. Akurasi adalah kunci penting.
4. Bahasa yang digunakan adalah bahasa tinggi.
5. Gunakan beberapa model dan teknik pembelajaran mesin.
6. Dapatkan hasil yang menarik.

Menurut (Adha et al., 2017) Proses KDD secara kasar dapat digambarkan sebagai berikut:

1. *Data Selection*

Proses memecahkan dan menyeleksi data gabungan yang bekerja sesuai dengan kriteria yang diinginkan oleh peneliti, seperti data konsumen.

2. *Pre-procesing/cleaning*

Hal pertama merupakan proses pembersihan. pembersihan merupakan proses yang meliputi pengecekan data yang tidak konsisten, menghilangkan duplikasi data, dan memperbaiki kesalahan pada data.

3. *Transformation*

Pada proses coding, data yang terpilih dimodifikasi agar sesuai dengan algoritma apriori. Urutan data dalam KDD 9 (*Data Discovery Database*) tergantung pada jenis dan pola data yang dicari dalam *database*.

4. *Data mining*

Data mining adalah langkah pencarian struktur dan data pada data terpilih dengan menggunakan metode yang berbeda. Dalam ekstraksi data, teknik dan metode sangat bervariasi. Pilihan strategi yang paling tepat tergantung pada sistem KDD secara keseluruhan.

5. *Interpretation/ Evaluation*

Informasi yang dihasilkan oleh data mining harus disajikan dengan cara yang mudah dipahami oleh semua pihak. Fase ini disebut fase definisi.

2.1.2. *Data Mining*

Data mining adalah proses menganalisis dan mengekstrak pengetahuan secara otomatis menggunakan satu atau lebih teknik pembelajaran komputer. (Eska, 2018). Sedangkan menurut Bramer dalam (Mardi, 2017) Data mining adalah proses yang menggunakan teknik statistik, kecerdasan buatan, matematika, dan pembelajaran mesin untuk mengekstrak dan menganalisis informasi dan pengetahuan berharga dari data besar. Dari pengertian data mining yang sudah dijabarkan, hal penting yang terkait dengan data mining menurut (Mardi, 2017):

1. Data mining adalah proses otomatis dari data yang ada
2. Data yang akan diproses merupakan data yang sangat besar
3. Tujuan dari data mining adalah untuk menemukan hubungan atau pola yang mungkin memberikan indikasi yang berguna

2.1.2.1. Pengelompokan *Data Mining*

Menurut Larose dalam jurnal (Alfiyan et al., 2019), Kecerdasan data dibagi menjadi beberapa kelompok berdasarkan fungsinya:

1. Deskripsi
Dalam proses menggambar, peneliti berusaha memperoleh informasi dari data yang tersedia yang menunjukkan gambaran tentang jenis dan gayanya.
2. Estimasi
Berdasarkan hasil variabel target, model prediktif dibuat menggunakan hasil record yang ada sebagai nilai referensi.
3. Prediksi
Tujuan dari prediksi ini adalah untuk memprediksi hasil di masa depan. Misalnya, prediksi adalah prediksi produk mana yang akan laris di pasar dalam beberapa bulan mendatang.
4. Klasifikasi
Klasifikasi adalah proses menggambarkan data dalam bentuk dan kemudian mengklasifikasikannya dengan cara yang telah ditentukan. Proses ini dilakukan untuk menentukan distribusi informasi berdasarkan kebutuhan masa depan.
5. Pengklasteran

Tujuan dari pengklasteran adalah untuk membentuk kelompok data, observasi, atau membuat kelompok objek dengan data yang sejenis. *Cluster* adalah kumpulan data yang memiliki kemiripan satu sama lain dan tidak ada kemiripan dengan data pada cluster lainnya.

6. Asosiasi

Pekerjaan sosial menemukan strategi yang bekerja sama pada waktu yang sama. Dalam bisnis, interaksi ini sering disebut sebagai analisis obrolan.

2.1.2.2. Tahapan – Tahapan *Data Mining*

Menurut Saleh dalam jurnal (Yanto, 2018) data mining dibagi menjadi beberapa tahapan yaitu:

1. Pembersihan data

Proses menghilangkan *noise* dan data yang tidak konsisten.

2. Integrasi data

Menggabungkan data dari berbagai *database* menjadi satu database baru.

3. Seleksi data

Data yang ada di database yang relevan dari hasil analisis yang akan diambil dari database.

4. Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format tertentu yang dapat diproses dalam *Data Mining*.

5. Proses *Mining*.

Proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi pola

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.

7. Presentasi pengetahuan

Penyajian pengetahuan dari metode yang digunakan untuk memperoleh pengetahuan.

2.1.2.3. Manfaat Data Mining

Menurut (Erwansyah, 2019) manfaat dari data mining dapat dilihat berdasarkan dua sudut pandang yang berbeda, antara lain sebagai berikut:

1. Sudut Pandang Komersial

Di area ini, data mining dapat digunakan bersama dengan teknologi komputer untuk memproses data dalam jumlah besar untuk mendapatkan informasi yang dapat digunakan untuk meningkatkan persaingan.

2. Sudut Pandang Keilmuan

Dari perspektif ilmiah, pengumpulan data dapat digunakan untuk menganalisis data mentah dan menyimpannya di lingkungan data besar.

2.2. Tinjauan Teori Khusus

2.2.1. Algoritma Apriori

Algoritma apriori adalah algoritma yang sudah lama dikenal banyak orang yang melakukan proses pencarian fenomena elemen dengan menggunakan teknik aturan kombinasi. Langkah-langkah untuk menentukan aturan yang ada dari pengumpulan data adalah berapa kali sesuatu ditemukan. Tabel periodik unsur adalah unsur atau kelompok unsur yang sering dikelompokkan bersama. Hasil kerjasama dapat dilihat dari hasil kerjasama dan nilai-nilai kepercayaan. Dukungan

mengukur seberapa banyak informasi dalam stok lebih dari cukup informasi dalam transaksi. Keyakinan atau kepercayaan adalah contoh keadaan yang menunjukkan hubungan antara satu hal dengan hal lainnya. (Putria, 2018).

Menurut (Santoso et al., 2016) terdapat dua proses utama yang dilakukan dalam algoritma apriori, yaitu:

1. *Join* (penggabungan)

Metode ini dimaksudkan untuk menghubungkan objek yang sudah ada dengan objek lain, bukan untuk membuat koneksi baru.

2. *Prune* (pemangkasan)

Metode ini merupakan metode penentuan nilai minimum yang direkomendasikan dari suatu campuran bahan.

2.2.2. Penjualan

Menurut Kotler dalam jurnal (Gusrizaldi & Komalasari, 2016) Pemasaran penjualan pada dasarnya adalah ilmu dan seni menggunakan model menjual untuk menarik perhatian orang lain agar membeli produk yang ditawarkan. (Gusrizaldi & Komalasari, 2016) ada beberapa faktor yang memberikan pengaruh terhadap penjualan sebagai berikut:

1. Kondisi Pasar

Pasar adalah tempat di mana barang-barang dijual antara penjual dan pembeli. Kondisi pasar, jenis pelanggan, pangsa pasar, daya beli, waktu dan syarat pembelian penting bagi pengecer untuk memaksimalkan keuntungan.

2. Kondisi dan Kemampuan Penjual

Jual beli melibatkan dua aspek, yaitu penjual pertama dan pembeli kedua. Sehingga penjual atau orang pertama akan dapat meyakinkan pembeli atau orang kedua untuk mencapai target penjualan yang diinginkan.

3. Modal

Penjual membutuhkan modal yang lebih tinggi sehingga dia dapat menunjukkan produk kepada pembeli melalui transaksi untuk meningkatkan biaya pengiriman, biaya, dan biaya lainnya.

4. Kondisi Organisasi

Di perusahaan besar, penjualan ditangani oleh manajer penjualan perusahaan atau profesional penjualan. Hal ini berbeda dengan perusahaan kecil atau usaha kecil dimana proses pemasarannya dilakukan oleh orang-orang yang terlibat dalam berbagai bidang perusahaan atau bisnis tersebut.

2.2.3. Association Rule Mining

Menurut (Kaur & Singh, 2013) dalam jurnal (Masnur, 2015) *Association Rule Mining* adalah teknik penambangan data yang memungkinkan menemukan aturan asosiasi dalam data layanan. Aturan penyortiran dirancang untuk menemukan produk umum dalam database. Misalnya, manajer toko mungkin ingin mengetahui bagaimana pelanggan membeli sepatu tenis saat berbelanja sepatu tenis. Pengetahuan ini dapat membantu manajer toko mengelola inventaris, mengatur ruang rak, dan menerapkan strategi promosi produk.

Dokumen merupakan salah satu variabel *boolean* 0 dan 1, dimana 0 berarti tidak ada dan 1 berarti ada. Setiap elemen dalam plot dapat mewakili vektor logis.

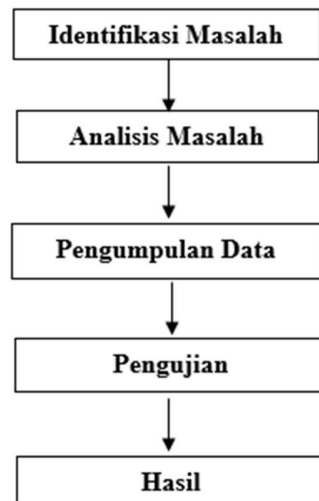
Vektor logis sering diperlukan untuk nilai yang mewakili objek terkait. Nilai-nilai ini dapat direpresentasikan sebagai aturan asosiasi.

2.2.4. Tanagra

Menurut (Rahmawati & Merlina, 2018) Tanagra adalah perangkat lunak gratis untuk tujuan pendidikan dan penelitian. Dalam penelitian ini, metode data mining berbeda analisis data. Dari pembelajaran statistik dan pembelajaran mesin hingga database. Menurut (Najib & Suryani, 2020) Tanagra memiliki 2 tujuan yaitu:

1. Memberi para peneliti dan siswa alat penambangan data sederhana untuk memenuhi standar pengembangan perangkat lunak industri saat ini (terutama dalam hal desain dan penggunaan grafis) dan membantu analisis data nyata atau sintetis.
2. Menyediakan struktur yang memungkinkan peneliti untuk dengan mudah menggambarkan pekerjaan mereka menggunakan metode pengumpulan data mereka.

2.2. Kerangka Pemikiran



Gambar 2. 1 Kerangka pemikiran

Sumber: (Penelitian)

Berikut penjelasan dari flowchart kerangka berfikir diatas:

1. Peneliti mengidentifikasi masalah yang terjadi
2. Setelah mengidentifikasi masalah peneliti menganalisis masalah. pada tahapan ini peneliti menentukan ruang lingkup masalah
3. Peneliti mengumpulkan data yang dibutuhkan untuk memecahkan masalah yang sudah di analisis. Peneliti harus melakukan kegiatan observasi langsung ke PT. Selatanindo Batam Mandiri sehingga bisa dilihat kesalahan yang terjadi. Membuat pengolahan data dari apa yang didapat dari data transaksi untuk mendapatkan berapa itemset yang muncul menggunakan pendekatan *association rule* dan proses pengujian data dilakukan dengan menggunakan *software* Tanagra.
4. Tanagra diuji secara manual dalam proses analisis manual dengan menghitung data menggunakan *baseline* yang sesuai dan meninjau data

yang dihitung secara manual untuk menentukan konsistensi data yang dihasilkan.

5. Hasilnya adalah tempat terakhir untuk menulis laporan yang menjelaskan temuan dan merekomendasikan penelitian lebih lanjut untuk meningkatkan pengetahuan ini.

2.3. Penelitian Terdahulu

Penelitian terdahulu dari penelitian ini adalah sebagai berikut:

Tabel 2. 1 Peneliti terdahulu

No	Penelitian	Judul	Hasil Penelitian
1	Alkadri Masnur 2015	Analisa Data Mining Menggunakan Market Basket Analysis Untuk Mengetahui Pola Beli Kosumen	Analisis korelasi kunci dan algoritma canggih ini dapat digunakan untuk membantu Toko Bangunan Gunung Sayo membuat angka yang memahami pola pembelian pelanggan dan apa yang paling sering mereka beli.
2	Adhithia Erfina, Melawati, Nunik Destria Arianti 2020	Penerapan Metode Data Mining Terhadap Data Transaksi Penjualan Menggunakan Algoritma Apriori (Studi kasus : Toko Fasentro Fancy)	Dengan penerapan algoritma apriori pada data transaksi penjualan pada tokoh, dari hasil penelitian ini yaitu untuk menentukan kombinasi itemset
3	Dini Silvi Purnia, Ai llah Wamilah 2017	Implementasi Data mining Pada Penjualan Kacamata Menggunakan Algoritma Apriori	Menggunakan algoritme apriori berdasarkan bagan di atas, merek kacamata terlaris menggunakan algoritme apriori untuk mengembangkan strategi pemasaran untuk menjual produk mereka bersama perusahaan lain dengan memvalidasi manfaat produk mereka.

Tabel 2.1 Lanjutan Penelitian Terdahulu

No	Penelitian	Judul	Hasil Penelitian
4	Erma Delima Sikumbang 2018	Penerapan Data Mining Penjualan Sepatu Menggunakan Algoritma Apriori	Data mining dan algoritma canggih sangat berguna untuk memahami hubungan antara frekuensi penjualan sepatu yang paling banyak diminta oleh konsumen, yang merupakan informasi yang sangat penting ketika memutuskan untuk membuat inventaris untuk jenis sepatu di pasar. Undang-undang anggaran yang penting membantu mengembangkan strategi penjualan sepatu.
5	Fitri Rahmawati, Nita Merlina 2018	Metode Data Mining Terhadap Data Penjualan Sparepart Mesin Fotocopy Menggunakan Algoritma Apriori	Setelah melakukan pengujian dengan cara perhitungan manual dan perhitungan menggunakan software Tanagra versi 4.1 dapat diketahui bahwa hasil yang di peroleh dari perhitungan manual tidak jauh berbeda dengan perhitungan menggunakan software Tanagra 1.4
6.	Ramadani Saputra dan Alexander J.P. Sibarani 2020	Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat	Penggunaan data mining dapat diimplementasikan dengan pre-algorithm pada sistem pengadaan untuk memprediksi pola koleksi dan aturan pengumpulan barang sebagai informasi penjualan dan menerapkan pre-algorithm dalam penjualan obat untuk mengetahui pola penjualan konsumen jika pada waktu pengolahan data kita dapat mengetahui obat mana yang paling sering dibeli.
7.	Arfhan Prasetyo, Numan Musyaffa, Ricki Sastra 2020	Implementasi Data Mining Untuk Analisis Data Penjualan Dengan Menggunakan Algoritma Apriori (Studi Kasus Dapoerin's)	Studi tentang algoritma apriori dapat membantu mengembangkan strategi pemasaran untuk berbelanja makanan ringan lainnya dengan pelajari keuntungan menjual makanan jajanan.

Sumber: (Penelitian)