

BAB II

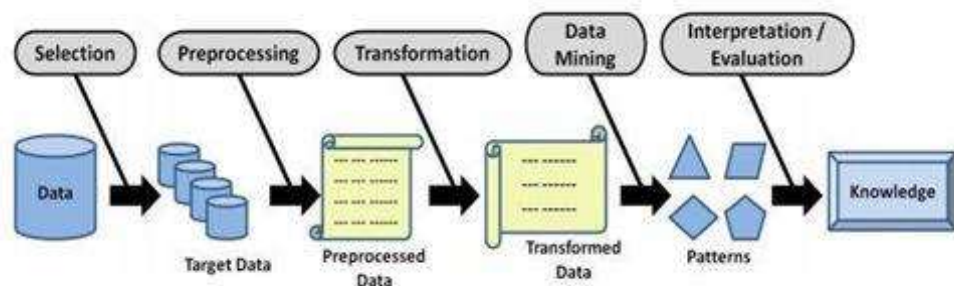
LANDASAN TEORI

2.1 *Knowledge Discovery in Database (KDD)*

Knowledge Discovery in Databases (KDD) atau Penemuan Pengetahuan dalam Basis Data adalah proses yang kompleks untuk mengidentifikasi, mengekstraksi, dan menerapkan pengetahuan baru yang berharga dari kumpulan data yang besar dan kompleks (Gustientiedina et al., 2019a) Proses KDD melibatkan berbagai tahapan yang saling terkait untuk memahami data, mengekstraksi pola, mengidentifikasi informasi penting, dan mengambil keputusan yang berdasarkan pada pengetahuan yang dihasilkan (Mardalius, 2018).

Menurut (Elisa, 2017) Sebuah teknik yang dikenal sebagai *Knowledge Discovery in Databases (KDD)* merupakan disiplin yang berfokus pada penggunaan teknik analisis data, statistik, dan pembelajaran mesin untuk menghasilkan pengetahuan baru dari data yang tidak terstruktur.

Secara lebih detail, proses *Knowledge Discovery in Database (KDD)* seperti gambar di bawah ini.



Sumber: (Gustientiedina et al., 2019b)

Gambar 2. 1 Proses Knowledge Discovery in Database (KDD)

1. Untuk memastikan tidak ada duplikasi atau kesamaan dalam penanganan data, digunakan seleksi seleksi untuk menyeleksi faktor-faktor yang akan diambil.
2. Dua tahap pemrosesan adalah sebagai berikut:

Integrasi data dilakukan secara kredit yang membedakan substansi baru guna membersihkan data yang tidak terpakai.

Transformasi Transformasi data sesuai dengan pengolahan data mining sesuai dengan format ekstensi.
3. Langkah utama dalam proses penggalian informasi baru dari data yang ada dikenal dengan data mining.
4. Evaluasi dan Interpretasi Mencari pola-pola menarik dalam basis pengetahuan yang telah diidentifikasi.
5. Pengetahuan Informasi yang baru diperoleh akan didiskusikan. Pada titik ini, informasi yang baru dihasilkan dapat menjangkau setiap individu untuk dijadikan sebagai sudut pandang pengambilan keputusan.

2.2 Data Mining

Penambangan data adalah proses menemukan contoh menarik yang sebelumnya tidak jelas dari data penting yang disimpan atau disimpan dalam kumpulan data atau tumpukan informasi yang sangat besar (Azwanti, 2019). Ukuran urutan robotisasi kasus ditentukan oleh desain informasi yang diperoleh dari indeks informasi, dan penambangan data juga merupakan hasil penggalian data dari kumpulan data lingkup besar.(Aras & Sardjono, 2016). Tujuan data mining adalah untuk mengekstraksi pengetahuan dan informasi yang dapat digunakan oleh

pekerja informasi yang dikenal dinamis dan berpikir kritis. (Waworuntu & Amin, 2018).

Menurut Handoko (2016) mengatakan bahwa data mining adalah proses yang menggunakan penalaran matematis, faktual, dan buatan manusia untuk menemukan dan mengekstraksi data yang berguna dan informasi terkait dari kumpulan data yang sangat besar. Penambangan data dapat digunakan dalam berbagai cara untuk mencapai tujuan tersebut (Ardiada et al., 2018). Tergantung pada hasil yang diharapkan, kegiatan data mining dapat dibagi menjadi beberapa kategori berikut:

1. **Prediksi** Prediksi atau disebut juga dengan prediksi adalah proses mengidentifikasi pola tertentu dalam data yang dapat diturunkan dari variabel dalam data tersebut dan digunakan untuk memprediksi variabel lain yang belum diketahui nilainya.
2. **Deskripsi** Deskripsi atau deskripsi memberikan pemahaman yang lebih mendalam tentang data yang diamati dan mengidentifikasi karakteristiknya.
3. **Klasifikasi**, juga dikenal sebagai klasifikasi, adalah proses pemrosesan data untuk menemukan fungsi atau model tertentu dari konsep data. Proses ini digunakan untuk membagi setiap data menjadi kelompok-kelompok yang berbeda.
4. **Fungsi asosiasi** adalah menemukan kombinasi atau aturan asosiatif dari data yang akan diolah untuk menemukan informasi tentang hubungan antar variabel.
5. Tujuan dari clustering lebih khusus untuk mengorganisasikan data ke dalam kelas-kelas dengan karakteristik yang serupa.

2.3 Manfaat Data mining

Menurut (Suyanto, 2017) Dua aplikasi utama dari data mining bersifat deskriptif dan prediktif. Tujuan dari metode deskriptif untuk data mining adalah menemukan pola dalam data saat ini yang dapat dipahami oleh orang-orang. Metode prediksi, di sisi lain, berfokus pada pengembangan model ilmiah yang digunakan untuk memperkirakan atau mengantisipasi kejadian atau nilai di masa depan berdasarkan data yang telah dikumpulkan di masa lalu.

Berdasarkan fungsinya, tugas-tugas data mining dibagi menjadi 6 kelompok yaitu:

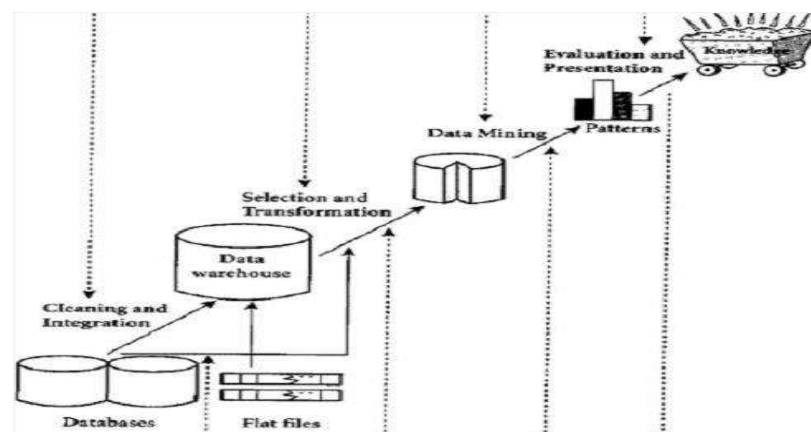
1. *Klasifikasi*: Mengklasifikasikan data ke dalam kategori atau kelompok tertentu berdasarkan pola dan karakteristiknya. Misalnya, mengklasifikasikan pelanggan sebagai pelanggan setia atau tidak setia berdasarkan riwayat pembelian mereka.
2. *Clustering* (Pengelompokan): Mengelompokkan data ke dalam kelompok-kelompok yang memiliki kesamaan karakteristik atau pola. Clustering membantu dalam mengidentifikasi segmen pelanggan yang berbeda berdasarkan preferensi atau perilaku yang serupa.
3. *Prediksi*: Memprediksi nilai atau kejadian masa depan berdasarkan pola dari data historis. Misalnya, memprediksi penjualan bulan depan berdasarkan tren penjualan dari bulan-bulan sebelumnya.
4. *Anomali Detection* (Deteksi Anomali): Mendeteksi data yang berbeda atau tidak biasa yang mungkin menunjukkan perilaku atau kejadian yang tidak normal, seperti mendeteksi transaksi penipuan dalam data keuangan.

5. Asosiasi: Menemukan hubungan dan asosiasi antara variabel atau item dalam data. Misalnya, mengidentifikasi pola pembelian bersamaan dari produk tertentu dalam daftar pembelian pelanggan.
6. Ringkasan (*Summarization*): Merangkum data menjadi bentuk yang lebih kompak dan informatif, sehingga memudahkan pemahaman dan analisis. Contohnya, membuat ringkasan statistik dari data atau menampilkan informasi kunci dari kumpulan data besar.

Setiap kelompok tugas ini memiliki peran penting dalam proses data mining dan membantu organisasi dalam menggali wawasan berharga dari data mereka untuk pengambilan keputusan yang lebih baik dan strategi bisnis yang lebih efektif.

2.4 Tahapan Data mining

Proses tahapan yang dilakukan merupakan salah satu rangkaian dalam proses data mining. Teknik integrasi, penemuan ilmiah, dan visualisasi pola data terjaln di sepanjang tahapan ini.



Gambar 2. 2 Gambaran Tahap- Tahap data mining

Sumber: www.kompasiana.com

Berikut ini adalah tahapan data mining:

1. Pembersihan data adalah proses membuang data atau kata-kata yang salah. Selain itu, lebih baik membuang data yang tidak relevan daripada mengurangi data yang akurat di masa mendatang.
2. Integritas data adalah proses menggabungkan data dari beberapa database menjadi satu.
3. Transformasi data (*Data Transformation*) Mengubah data dan menilai kegunaan temuan data mining. untuk membuatnya lebih sederhana untuk digunakan.
4. Implementasi metode untuk penambangan data Masalah satu bagian dalam penambangan data Metode untuk mengekstraksi pola dari data yang ada.
5. Pola yang akan dievaluasi untuk menentukan apakah hasil yang diamati benar-benar tercapai. meningkatkan proses data mining yang dihasilkan.
Presentasi Pengetahuan: Menggunakan pengetahuan untuk menyampaikan temuan data mining untuk mendapatkan data yang dicari pengguna.

2.5 Barang Elektronik

Barang elektronik merujuk pada perangkat yang menggunakan teknologi elektronik untuk berfungsi dan berinteraksi dengan pengguna. Istilah ini mencakup berbagai jenis peralatan dan perangkat yang menggunakan komponen elektronik, seperti sirkuit terpadu, transistor, dan perangkat semikonduktor lainnya. Barang elektronik dapat ditemukan dalam berbagai bentuk dan ukuran, mulai dari perangkat kecil seperti smartphone, tablet, dan kamera digital, hingga perangkat yang lebih besar seperti televisi, mesin pencuci, dan lemari es.

Keberadaan barang elektronik telah mengubah cara kita berkomunikasi, bekerja, dan berinteraksi dengan dunia sekitar. Mereka memungkinkan transfer data dan informasi yang cepat, pengolahan yang efisien, dan kontrol yang lebih mudah. Dalam masyarakat modern, barang elektronik telah menjadi bagian penting dari kehidupan sehari-hari, memfasilitasi kegiatan sehari-hari seperti berkomunikasi, mengakses informasi, hiburan, dan produktivitas. Selain itu, barang elektronik juga terus berkembang dengan inovasi dan teknologi baru yang terus muncul, menghadirkan perangkat yang lebih canggih dan fungsional bagi pengguna.

2.6 Metode Apriori

Metode Apriori adalah salah satu teknik dalam analisis data dan data mining yang digunakan untuk menemukan pola asosiasi atau hubungan yang relevan antara item dalam dataset transaksi atau daftar item. Metode Apriori digunakan secara khusus dalam analisis data transaksi, seperti data penjualan, data pembelian pelanggan, atau data transaksi lainnya (Rao & Gupta, 2012). Metode perhitungan apriori digunakan oleh penulis ulasan ini untuk mengubah data informasi yang besar menjadi aturan afiliasi yang membantu melacak informasi. Dalam proses ini, perhitungan apriori memainkan peran penting dalam mengidentifikasi hubungan dan asosiasi antara item dalam basis informasi. Penentuan kebutuhan dasar dan biaya bantuan dalam basis informasi juga ditentukan oleh struktur afiliasi. Untuk menghitung besaran nilai sebuah item, penulis menggunakan persamaan yang telah ditentukan sebelumnya.

$$\text{Support } (A) = \frac{\text{Jumlah Transaksi mengandung } A}{\text{Total Transaksi}} \times 100\%$$

Sementara nilai support antara 2 item diperoleh dengan menggunakan rumus:

$$Support (A) = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi}} \times 100$$

Rumus 2. 1 Perhitungan Support

Ketika lebih dari nilai minimum yang diberikan hadir, frekuensi dari itemset menampilkan frekuensi terjadinya lebih dari itu (). Misalnya, 2 Frekuensi kekambuhan dianggap sering jika lebih dari atau sama dengan dua kali. Huruf FK adalah singkatan dari frequent k-itemset set. Hitung kepercayaan aturan asosiatif A-B setelah semua pola frekuensi untuk menemukan aturan asosiasi yang memenuhi kepercayaan minimum.

Menurut (Soma Darmawan, 2015), Dalam kesimpulannya, algoritma Apriori terbukti menjadi metode yang efisien dan berguna dalam menyelesaikan berbagai permasalahan analisis data dan data mining. Algoritma ini telah menarik perhatian banyak peneliti karena kemampuannya untuk mengidentifikasi pola asosiasi dan hubungan antara item dalam basis data besar. Dengan menggunakan algoritma Apriori, peneliti dapat menemukan informasi yang relevan dan bermanfaat dari data transaksi atau daftar item. Keunggulan algoritma Apriori adalah kemampuannya untuk menghasilkan asosiasi dan aturan-aturan yang berarti dari data, yang dapat digunakan untuk pengambilan keputusan bisnis, analisis pasar, dan strategi pemasaran yang lebih efektif. Hal ini membuat algoritma Apriori tetap menjadi pilihan populer dalam penelitian dan analisis data. Contoh pada penelitian (Mardiah & Yulia, 2021) Data mining Algoritma apriori dapat digunakan untuk mengolah data hasil transaksi penjualan suku cadang motor pada bengkel Candra Motor, sehingga dapat menentukan rule asosiasi antar barang yang terjual sehingga

memudahkan pengguna jumlah barang yang terjual. Data mining algoritma apriori memberikan keputusan yang tepat pengolahan data secara manual maupun menggunakan software WEKA 3.9, maka suku cadang motor yang paling banyak terjual adalah ban dalam, oli Yamaha, dan oli MPX dengan nilai minimum support 90% dan minimum confidence 100%.

2.7 Software pendukung



Gambar 2. 3 Logo Tanagra

Sumber : Tanagra. com

Satu perangkat lunak penambangan data, Tanagra, menawarkan berbagai teknik penambangan data yang menggabungkan analisis data, pembelajaran statistik, pembelajaran mesin, dan basis data. Tanagra adalah perangkat lunak sumber terbuka yang dapat digunakan dan ditambahkan oleh siapa saja. Ini sering digunakan untuk alasan akademik dan penelitian.

Mahasiswa dan peneliti merupakan mayoritas pengguna Tanagra. Lebih mudah bagi peneliti dan siswa untuk menggunakan perangkat lunak Tanagra sesuai dengan praktik terbaik pengembangan perangkat lunak yang diterima untuk

memeriksa data nyata. Contoh pada penelitian (Saefudin et al., 2019) Aplikasi Data Mining di UD. Mumu Jaya Pandeglang menggunakan algoritma Apriori, UML, Borland Delphi 7, dan basis data MySQL. Membantu menentukan jenis ikan yang diminati konsumen dan pendataan transaksi serta stok ikan. Analisis data transaksi penjualan ikan berdasarkan pola asosiasi. Memberikan manfaat dalam mengoptimalkan stok ikan, layanan pelanggan, dan pengambilan keputusan yang cerdas.

2.8 Penelitian Terdahulu

1. Nama Pengarang: (Elisa, 2018), judul : Market Basket Analysis Pada Mini Market Ayu Dengan Algoritma Apriori.

Tahun: Vol. 2 No. 2 (2018) 472 – 478, ISSN : 2580-0760

Di Kota Batam, di minimarket Ayu. Karena kedekatannya dengan rumah penduduk, minimarket ini memiliki penjualan harian yang berdampak pada pertumbuhan data transaksi yang berkelanjutan dan penyimpanan data yang masif. Data dari transaksi penjualan hanya dijadikan arsip, tidak sebagaimana mestinya. Intinya, informasi yang dikumpulkan berubah menjadi kumpulan data yang sangat berguna. Algoritma Apriori adalah teknik yang digunakan dalam analisis keranjang pasar yang berusaha mengidentifikasi pola asosiasi dalam perilaku pembelian pelanggan dan produk yang sering mereka beli bersama. Menurut temuan penelitian, minyak dan susu memiliki tingkat dukungan dan kepercayaan tertinggi, dengan nilai dukungan masing-masing sebesar 42,85% dan 85,71%.

2. **Nama pengarang:** (Ependi & Akbar, 2021), **Judul: Implementasi Data Mining Pada Penjualan Produk Dengan Menggunakan Algoritma Apriori.**

Tahun: Vol. 3 No. 1 November 2021 ISSN 2685-2675

Implementasi Data Mining menggunakan algoritma Apriori pada CV. Berkat Motor membantu pemilik, Dedi Atmaja, untuk menganalisis data penjualan produk suku cadang dan jasa kendaraan. Dengan hasil analisis menggunakan R Studio, ditemukan 4 aturan asosiatif berdasarkan support (0.1) dan confidence (0.5) yang memberikan informasi tentang pola transaksi dan produk yang sering dibeli oleh pelanggan. Hal ini membantu pemilik dalam mengambil keputusan yang lebih tepat dalam mengontrol stok dan menentukan produk yang harus dibeli. Dengan adanya data mining, pemilik dapat menghindari spekulasi dan melakukan pembelian suku cadang secara lebih efisien dan berdasarkan data yang akurat.

3. **Nama pengarang:** (Purnia & Warnilah, 2017) , **Judul: Implementasi Data Mining Pada Penjualan Kacamata Menggunakan Algoritma Apriori**

Tahun: Vol.2 No.2, November 2017 ISSN: 2527-449X

Salah satu optik yang ada di Tasikmalaya adalah Optik Nasional. National Optics membeli dan menjual kacamata dari berbagai merek sebagai bagian dari operasinya. Data penjualan akan terus bertambah karena banyaknya transaksi penjualan setiap harinya. Data transaksi penjualan hanya akan menjadi sampah yang tidak berguna jika dibiarkan. Kapasitas untuk mengumpulkan dan mengolah data semakin berkembang seiring dengan

kemajuan teknologi. Penambangan data adalah praktik memanfaatkan pengetahuan dan informasi yang ada dalam kumpulan data.

- 4. Nama pengarang:** (Irene Ananda & Salamah, 2020), **Judul: Determination Of Sales Data Patterns Using The Association Rules Apriori Method**
Tahun: 06 Agustus 2020 eISSN : 2598-246X

Menemukan metode terbaik untuk pengoptimalan penjualan dalam persaingan dunia korporat sangatlah penting. Frekuensi pembelian konsumen merupakan faktor yang mempengaruhi kebutuhan analisis pasar. Oleh karena itu, diperlukan solusi untuk mengidentifikasi pola penjualan yang lebih efektif dan efisien dengan website. Untuk membuat asosiasi item dan aturan transaksi, data NS yang diperlukan diambil dari data transaksi penjualan untuk periode waktu tertentu. Ini akan memudahkan bisnis untuk meningkatkan omset penjualan selain memungkinkan mereka untuk mencari pola yang sering muncul di tengah banyak transaksi. Dengan memanfaatkan HTML untuk pembuatan halaman web, PHP untuk pembuatan website, dan MySQL untuk pengelolaan database, maka terciptalah aplikasi ini.

- 5. Nama pengarang:** (Alfianzah et al., 2020), **Judul: Implementation of Apriori Algorithm Data Mining for Increase Sales**

Tahun: Volume 5, Number 1, October 2020 e-ISSN : 2541-2019

Setiap perusahaan atau organisasi yang ingin bertahan harus memilih rencana bisnis yang sesuai. Sangat disayangkan jika statistik penjualan produk Lakoe Dessert Pondok Kacang tidak ditelaah lebih lanjut karena

pada akhirnya akan mengakibatkan penimbunan data. Untuk mengidentifikasi produk dengan penjualan tertinggi dan keterkaitan antara satu produk dengan produk lainnya, diperlukan salah satu algoritma data mining yaitu algoritma apriori. Dengan bantuan aplikasi Rapidminer 5 yang memiliki nilai support 2,4% dan nilai kepercayaan 50%, produk yang sering dibeli atau diminati pelanggan dapat diketahui. 209 transaksi yang merupakan data penjualan bulan Maret 2020 digunakan dalam penelitian ini. Menurut penelitian, orang paling sering membeli produk dengan label Strawberry Pudding dan Vanilla Pudding.

6. **Nama pengarang:** (Firmansyah & Merlina, 2020) , **Judul: Prediksi pola Penjualan Tiket Kapal PT.Pelni cabang Makassar Menggunakan Metode Algoritma Apriori ,Tahun: VOL. 5. NO. 2 FEBRUARI 2020 E-ISSN: 2527-4864**

Tujuan dari penelitian ini adalah untuk meramalkan penjualan tiket pesawat, yang sering dilakukan dengan menganalisis perilaku pembelian tiket secara historis. Algoritma Apriori digunakan dalam metodologi penelitian ini, dan penulis memilih Cabang Makassar sebagai studi kasus. Diketahui pola penjualan tiket kapal adalah jika membeli tiket KM Lambelu maka akan membeli juga tiket KM Bukit Siguntang, dengan nilai support 75% dan nilai confidence 90%. Algoritma Apriori menghasilkan aturan asosiasi antar item dari Januari 2018 hingga Desember 2018.

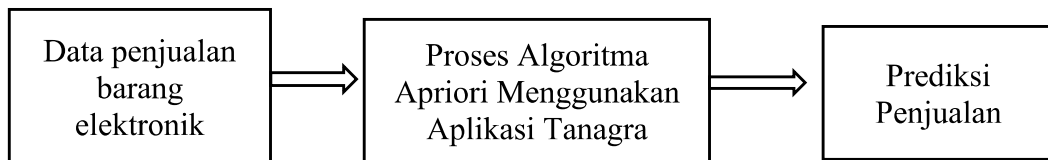
7. **Nama pengarang:** (Anas, 2020) **Judul:** **Implementasi Algoritma Apriori Untuk Menentukan Strategi Promosi STIE-Graha Karya Muara Bulian Tahun: Vol. 14, No. 1, April 2020 ISSN: 1978-8126**

Penerapan algoritma apriori merupakan tujuan utama dari penelitian ini untuk mendapatkan gambaran data mahasiswa baru di STIE-GK Muara Bulian. Untuk tahun ajaran 2018–19, penulis menganalisis data mahasiswa baru STIE–GK Muara Bulian menggunakan algoritma Apriori. Penambahan data diperlukan untuk mengumpulkan pengetahuan, dan metode apriori dipilih karena dapat mengevaluasi data yang muncul secara bersamaan, berulang, dan telah terakumulasi dalam waktu yang lama. Penulis menggunakan data dari 100 mahasiswa baru yang dipilih secara acak di STIE-GK Muara Bulian. Teknik Apriori digunakan penulis untuk menghitung data, yang kemudian diuji dengan menggunakan program data mining Weka. Dengan tingkat kepercayaan 13%, temuan penelitian menunjukkan bahwa variabel jurusan SMK dan TKJ menjadi kontributor utama siswa baru.

2.9 Kerangka Pemikiran

Metode algoritma apriori merupakan metode yang sering digunakan untuk memprediksi penjualan atau menganalisis data dengan tingkat akurasi tinggi. Data yang akan dianalisis adalah data penjualan barang elektronik. Data penjualan akan diproses menggunakan aplikasi Tanagra untuk mendapatkan prediksi yang akurat. Dari data penjualan yang didapatkan, data akan dikompilasi dan diproses dalam klasifikasi bulanan untuk digunakan sebagai data terproses yang akan dimasukkan

ke dalam aplikasi Tanagra. Dari hasil aplikasi Tanagra, akan didapatkan tingkat akurasi dengan persentase tertinggi. Dalam proses pengolahan data, tingkat akurasi atau persentase akan ditentukan yang akan digunakan sebagai patokan. Dengan demikian, akan dihasilkan output berupa nilai prediksi yang diprediksi akan mengalami peningkatan atau penurunan pada bulan atau tahun berikutnya. Dari deskripsi latar belakang dan metode yang digunakan, kerangka kerja penelitian ini adalah:



Gambar 2. 4 Kerangka pemikiran
Sumber: Data Penelitian 2023