

BAB II

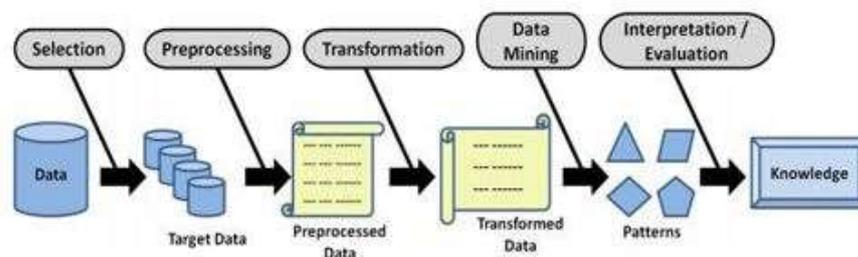
KAJIAN PUSTAKA

2.1 Knowledge Discovery in Database (KDD)

Ekstraksi informasi potensial, implisit, dan tidak diketahui dari sekumpulan data disebut sebagai *Knowledge Discovery in Databases* (KDD). (Mardalius, 2018) Hasil output yang diperoleh dapat dijadikan dasar pengambilan keputusan sebagai (knowledge base). (Gustientiedina et al., 2019a).

Menurut (Elisa, 2017) Sebuah teknik yang dikenal sebagai *Knowledge Discovery in Databases* (KDD) menggabungkan pengumpulan, pemanfaatan, dan persepsi informasi yang direkam dengan rekonsiliasi dan pengungkapan logis, terjemahan, dan persepsi contoh dalam berbagai kumpulan informasi untuk menemukan normalitas, contoh, atau koneksi dalam jumlah data yang besar.

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



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

Sumber: (Gustientiedina et al., 2019b)

1. Untuk memastikan tidak ada duplikasi atau kesamaan dalam penanganan data, digunakan 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.
3. Transformasi data sesuai dengan pengolahan data mining sesuai dengan format ekstensi.
4. Langkah utama dalam proses penggalian informasi baru dari data yang ada dikenal dengan data mining.
5. Evaluasi dan Interpretasi Mencari pola-pola menarik dalam basis pengetahuan yang telah diidentifikasi.
6. 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.

2.2.1 Pengelompokan Data mining

Menurut (Saikin & Kusrini, 2019), Dalam bukunya data *mining* dibagi menjadi beberapa kategori berdasarkan tugas yang dapat dilakukan yaitu:

1. Deskripsi Banyak peneliti hanya ingin mencoba berbagai cara untuk mendeskripsikan pola dan tren data.
2. Estimasi Laporan lengkap dengan nilai dan variabel target sebagai nilai estimasi sementara digunakan untuk membuat model. Klasifikasi dan estimasi hampir identik.
3. Prediksi Prediksi mirip dengan estimasi dan klasifikasi, dengan nilai hasil prediksi yang akan datang berbeda.
4. Klasifikasi memiliki variabel target kategorikal dalam klasifikasi. Misalnya, menentukan apakah aplikasi hipotek pelanggan memiliki kredit yang baik atau buruk.
5. Clustering adalah proses pengelompokan laporan, observasi, atau objek yang memiliki kesamaan fitur ke dalam tahapan. Kumpulan laporan yang tidak berbeda dengan laporan klaster lainnya tetapi memiliki kemiripan satu sama lain disebut dengan klaster. Misalnya, pengelompokan dalam penelitian bisnis digunakan untuk membedakan antara perilaku keuangan yang sah dan dipertanyakan dalam audit akuntansi.

Tugas asosiasi dalam data mining adalah menemukan atribut yang terjadi dalam satu tahun. Selain itu, ini disebut sebagai analisis keranjang belanja dalam dunia bisnis.

Data mining juga dapat dikelompokkan berdasarkan algoritmanya, berikut beberapa algoritma data mining:

1. Algoritma *K-Means Clustering*

K-Means adalah perhitungan yang digunakan dalam kumpulan diskrit yang mengisolasi informasi ke dalam berbagai kumpulan (Sadewo et al., 2017). Ide dasar algoritma k-means sangatlah sederhana yaitu meminimalkan Sum of Squared Error (SSE) antara objek informasi dengan berbagai k centroid.

2. Algoritma Apriori

Penelitian ini menggunakan algoritma apriori, salah satu jenis association rule data mining, untuk mencari kombinasi itemset yang memenuhi kriteria atau filter yang diinginkan dan memiliki nilai frekuensi tertentu (Azwanti, 2019)

3. Algoritma C.45

Algoritme untuk membuat pohon keputusan disebut C4.5 (pengambilan keputusan). Keuntungan mendasar dari menggunakan pohon keputusan adalah kapasitasnya untuk menyederhanakan proses pengambilan keputusan yang kompleks sehingga pembuat keputusan dapat lebih mudah melihat potensi solusi untuk masalah. (Azwanti, 2019).

2.2.2 Manfaat Data mining

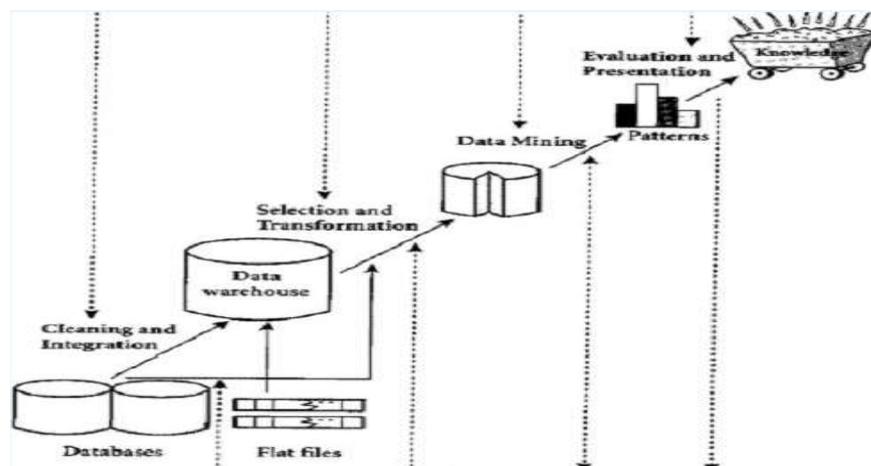
Menurut (Dr. Suyanto, S.T., 2017:3) kegunaan data mining ada 2 yaitu: deskriptif dan prediktif. Deskriptif adalah data mining yang digunakan untuk menemukan pola-pola yang dapat dimengerti oleh manusia yang menjelaskan karakteristik data. Prediktif adalah data mining untuk membentuk model ilmu pengetahuan yang akan digunakan untuk melakukan perkiraan atau prediksi.

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

1. Klasifikasi (classification) adalah menguraikan struktur yang diketahui untuk digunakan pada data-data baru.
2. Klasterisasi (clustering) adalah mengelompokkan data, yang belum diketahui jenis kelasnya, kedalam beberapa kelompok sesuai dengan kesamaannya.
3. Regresi (regression) adalah menemukan fungsi yang memodelkan data (kesalahan prediksi) seminimal mungkin.
4. Deteksi anomali (anomaly detection) adalah mengidentifikasi data yang tidak umum, perubahan yang sangat penting dan perlu penyelesaian lebih lanjut.
5. Pembelajaran aturan asosiasi (association rule learning) atau pemodelan perbandingan atau mencari relasi antar variabel.
6. Perangkuman (summarization) adalah menyediakan representasi data yang lebih sederhana yaitu visualisasi dan pembuatan laporan.

2.2.3 Tahapan Data mining

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



Gambar 2. 2 Gambaran Tahap- Tahap data mining

Sumber: [www: kompasiana.com](http://www.kompasiana.com)

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.
6. Presentasi Pengetahuan: Menggunakan pengetahuan untuk menyampaikan temuan data mining untuk mendapatkan data yang dicari pengguna.

2.3 Hasil Produksi

Hasil produksi merupakan keluaran dari suatu proses produksi yang dijalankan dalam sebuah perusahaan. Dalam penelitian ini hasil produksi yang dimaksud merupakan alat-alat elektronik yang diproduksi di PT. SIMATELEX Manufactory Batam.

Proses produksi barang kebutuhan dapur PT. Simatelex mempunyai ribuan tenaga kerja mulai dari proses awal yang berada di station1 hingga proses akhir station packing.

2.4 Metode Apriori

Dalam proses information mining ada beberapa teknik yang dapat dimanfaatkan. Ide kunci dari algoritma Apriori adalah membuat banyak lintasan atas basis data. Ini menggunakan pendekatan iteratif yang dikenal sebagai pencarian luas-pertama

(pencarian tingkat-bijaksana) melalui ruang pencarian, di mana k-itemsets digunakan untuk mengeksplorasi (k+1)-itemsets (Rao & Gupta, 2012). Penulis ulasan ini menggunakan metode perhitungan apriori. Salah satu perhitungan penting yang digunakan untuk mengubah basis informasi yang besar menjadi aturan afiliasi untuk melacak informasi adalah perhitungan apriori. Dalam basis informasi, kebutuhan untuk memenuhi kebutuhan dasar dan biaya bantuan ditentukan oleh struktur afiliasi. Persamaan berikut digunakan untuk menentukan nilai item:

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

Sementara nilai support antara 2 item diperoleh dengan menggunakan rumus:

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

Rumus 2. 1 Perhitungan Support

Frekuensi itemset menunjukkan frekuensi kemunculan itemset yang lebih besar dari nilai minimum yang ditentukan (). Misalnya, = 2 menunjukkan bahwa frekuensi kemunculannya lebih besar atau sama dengan 2 kali. FK menunjukkan kumpulan k-itemset yang sering. Hitung kepercayaan aturan asosiatif A-B untuk menemukan aturan asosiasi yang memenuhi persyaratan kepercayaan minimum setelah semua pola frekuensi dianalisis.

Menurut (Soma Darmawan, 2015), menetapkan bahwa algoritma apriori dapat diterapkan pada metode pemecahan masalah, yang menarik minat banyak akademisi untuk membuat algoritma yang efektif.

2.5 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.

2.6 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:**(Sriyanto et al., 2022),

Judul: Data Mining Penerapan Algoritma Apriori untuk Analisis Data Transaksi Sistem Inventory (Studi Kasus PT ABC President Indonesia)

Tahun: Vol. 3 No. 1 Juli 2022 ISSN 2715-2766

Untuk mencatat persediaan, PT. ABC President Indonesia tetap menggunakan sistem manual yang melibatkan penulisan data di atas kertas sebelum

memasukkannya ke dalam Microsoft Excel. Kesalahan pengumpulan dan pencatatan data, perhitungan dalam pengiriman barang, dan tingkat kesalahan yang lebih sering terjadi dalam pengiriman harian tidak dijamin. Akibatnya, pengiriman sering tertunda. Isu-isu ini panggilan untuk penggunaan data mining. Pendekatan tersebut memanfaatkan teknik algoritma apriori. Tujuan dari algoritma apriori adalah untuk mengidentifikasi produk yang paling diminati orang untuk dibeli. Algoritma apriori adalah aturan asosiasi dengan nilai dukungan dan nilai kepercayaan yang digunakan dalam penambangan data. Metode algoritma apriori akan digunakan dalam penelitian ini untuk menentukan penjualan desain aplikasi data mining yang paling berhasil. Menurut temuan penelitian, jika konsumen membeli produk NTHN450 maka mereka akan membeli produk NTHN330, dengan nilai support 47% dan nilai kepercayaan 78%, dan jika mereka membeli produk MTTK330, mereka akan membeli produk YOGURT, dengan nilai support 47% dan nilai kepercayaan 71%. NTHN450, NTHN330, MTTK330, dan YOGURT adalah produk yang paling diminati pelanggan di PT ABC President Indonesia.

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. Lakoe Dessert Pondok Kacang dapat mengembangkan strategi pemasaran untuk memasarkan produk lain dengan cara menganalisis keuntungan dari penjualan produk yang paling banyak terjual dan mengantisipasi kehabisan atau kosongnya stok atau bahan di beberapa titik. Dengan mengetahui produk yang paling banyak terjual dan pola pembelian barang yang dilakukan konsumen.

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

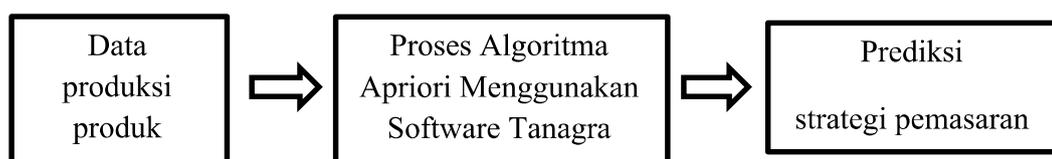
Tahun: Vol. 14, No. 1, April 2020 ISSN: 1978-8126

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.

2.7 Kerangka Pemikiran

Analisis data penting untuk melakukan pengolahan data hasil produksi. Pendekatan algoritma apriori adalah teknik yang sering digunakan untuk mengantisipasi penjualan atau menganalisis data secara akurat. Data produksi enam bulan terakhir menjadi bahan analisis. Untuk memberikan proyeksi yang tepat, aplikasi Tanagra akan mengolah data manufaktur. Untuk dijadikan sebagai data olahan yang akan dimasukkan ke dalam program Tanagra, data produksi akan disusun dan diolah dalam klasifikasi bulanan. Persentase tingkat akurasi tertinggi akan ditentukan dari hasil aplikasi Tanagra. Tingkat akurasi atau % akan ditetapkan selama proses pengolahan data dan digunakan sebagai tolok ukur. Akibatnya, akan menghasilkan output sebagai nilai yang diharapkan meningkat atau menurun di tahun mendatang. Kerangka penelitian ini dapat disimpulkan dari informasi latar belakang dan bagian teknik sebagai berikut:



Gambar 2. 4 Kerangka Berpikir
Sumber: Penelitian 2022