

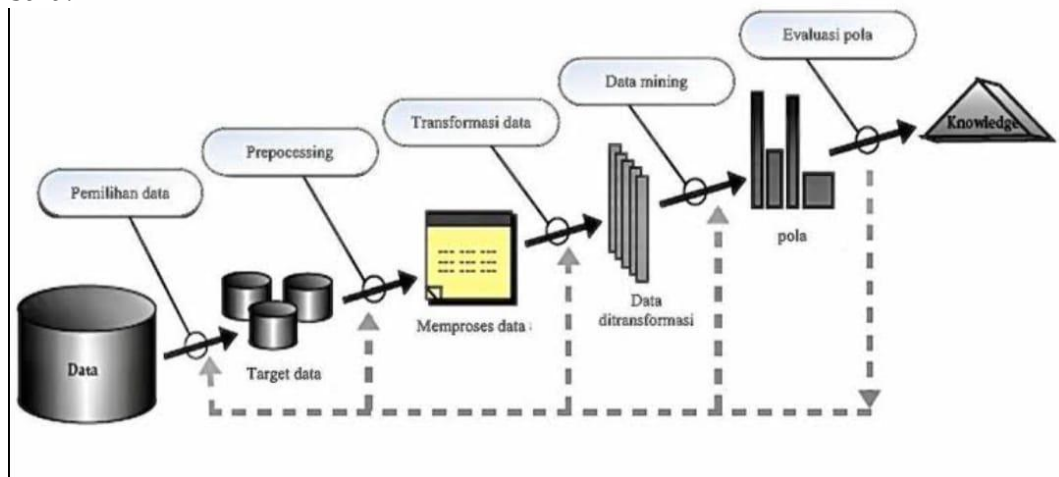
BAB II

LANDASAN TEORI

2.1 *Knowledge Discovery in Database (KDD)*

Knowledge Discovery in Databases (KDD) merupakan proses yang kompleks dan terstruktur untuk mengekstrak pengetahuan yang berguna dari kumpulan data besar. Proses ini melibatkan beberapa tahap, termasuk pembersihan data, integrasi, pemilihan, transformasi, penambahan data, serta evaluasi dan interpretasi hasil [1]. KDD sangat penting karena mampu mengungkap pola-pola yang sebelumnya tidak diketahui dan memberikan wawasan yang dapat diterapkan di berbagai bidang, seperti kesehatan, ilmu komputer, dan ilmu sosial [2]. Dalam konteks ilmu komputer, KDD telah berperan penting dalam pengembangan metodologi untuk menganalisis dataset yang kompleks, seperti yang digunakan dalam sistem deteksi intrusi. Misalnya, dataset KDD Cup '99 dan NSL-KDD telah banyak digunakan untuk benchmarking algoritma deteksi intrusi. Dataset KDD Cup '99, meskipun luas, telah dikritik karena redundansi dan kurangnya representasi dari lalu lintas jaringan yang nyata, yang mendorong pengembangan dataset NSL-KDD yang mengatasi masalah ini dengan menghilangkan catatan duplikat. Evolusi ini menunjukkan perlunya peningkatan kualitas dan relevansi data dalam aplikasi KDD. Lebih lanjut, teknik KDD juga telah diterapkan dalam berbagai domain, termasuk kesehatan, di mana mereka memfasilitasi analisis data pasien untuk diagnosis dan strategi perawatan yang lebih baik. Misalnya, model pembelajaran mesin yang dikembangkan melalui proses KDD telah digunakan untuk

menganalisis catatan medis dan memprediksi hasil pasien, menunjukkan utilitas praktis KDD dalam meningkatkan pengambilan keputusan di lingkungan klinis. Integrasi pengetahuan ahli dengan proses KDD juga memperkaya analisis, memungkinkan interpretasi yang lebih terinformasi dari data. Karakteristik iteratif dari KDD sangat penting untuk menyempurnakan proses ekstraksi pengetahuan. Setiap tahap KDD saling berhubungan, memastikan bahwa wawasan yang diperoleh tidak hanya valid tetapi juga dapat ditindaklanjuti. Pendekatan siklik ini memungkinkan peneliti dan praktisi untuk terus meningkatkan metodologi mereka dan beradaptasi dengan tantangan serta dataset baru.



Gambar: 2.1 Knowledge Discovery in Database (KDD)

Sumber: repository.unama.ac.id

Tahapan utama dalam KDD meliputi:

1. Seleksi Data: Mengidentifikasi dan memilih dataset yang relevan dari sumber data yang tersedia.
2. Praproses Data: Membersihkan data, menangani missing values, mendeteksi outliers, dan melakukan normalisasi.

3. Transformasi: Mengonversi data ke format yang sesuai untuk analisis, termasuk reduksi dimensi dan pemilihan fitur.
4. Data Mining: Menerapkan algoritma untuk mengekstraksi pola dan pengetahuan tersembunyi.
5. Interpretasi dan Evaluasi: Menilai signifikansi pola yang ditemukan dan mengkomunikasikan hasil.

2.2 Data Mining

Data mining, atau penambangan data, adalah proses analisis data yang bertujuan untuk menemukan pola, tren, dan informasi berharga dari kumpulan data yang besar. Menurut Muflikhah et al., data mining dapat didefinisikan sebagai proses otomatis yang menganalisis pengetahuan menggunakan berbagai teknik pembelajaran mesin (machine learning) untuk mengekstrak data yang menarik dari basis data yang besar. Proses ini melibatkan beberapa langkah, termasuk pemilihan data, pembersihan data, dan penerapan algoritma untuk mengekstrak informasi yang relevan [3]. Karim et al. menjelaskan bahwa data mining dapat diterapkan dalam berbagai konteks, termasuk dalam strategi dakwah di Kota Semarang, di mana pendekatan data mining digunakan untuk mengidentifikasi potensi dakwah dengan menerapkan algoritma fuzzy c-means[4]. Hal ini menunjukkan bahwa data mining tidak hanya berguna dalam konteks bisnis atau kesehatan, tetapi juga dalam konteks sosial dan keagamaan, di mana analisis data dapat membantu dalam pengambilan keputusan yang lebih baik. Riani et al. menambahkan bahwa data mining juga dapat digunakan untuk memprediksi penyakit, seperti dalam penelitian mereka yang menggunakan

metode Naive Bayes untuk memprediksi risiko penyakit jantung [5]. Dengan menggunakan berbagai indikator seperti usia, jenis kelamin, dan tekanan darah, mereka berhasil mencapai akurasi 86% dalam prediksi. Ini menunjukkan potensi data mining dalam bidang kesehatan untuk meningkatkan diagnosis dan pengobatan. Lebih lanjut, Bahri menekankan pentingnya penggunaan algoritma klasifikasi dalam data mining untuk memprediksi keterlambatan jam masuk kerja, yang dapat mempengaruhi kinerja perusahaan [6]. Dengan menerapkan teknik data mining, perusahaan dapat menemukan pola yang dapat membantu dalam meningkatkan disiplin dan produktivitas karyawan.

2.3 Sistem informasi Ritel

Sistem informasi ritel merupakan suatu sistem yang dirancang untuk mendukung operasi bisnis di sektor ritel, termasuk pengelolaan persediaan, penjualan, dan hubungan pelanggan. Menurut Krisnawati et al., sistem informasi ritel berfungsi untuk meningkatkan efisiensi operasional dan memberikan informasi yang akurat dan tepat waktu kepada manajemen untuk pengambilan keputusan yang lebih baik [7]. Dalam konteks ini, sistem informasi ritel tidak hanya berfokus pada pengolahan data transaksi, tetapi juga pada integrasi berbagai fungsi bisnis yang saling terkait, seperti pemasokan barang, manajemen persediaan, dan analisis penjualan. Salah satu aspek penting dari sistem informasi ritel adalah kemampuannya untuk menyediakan informasi yang cepat dan akurat. Misalnya, penelitian oleh Walingkas dan Saian menunjukkan bahwa penerapan sistem informasi pemasok barang dapat mempercepat proses pengadaan dan memastikan ketersediaan barang di toko [8]. Hal ini sangat penting dalam

industri ritel yang sangat bergantung pada kecepatan dan akurasi dalam pengelolaan persediaan untuk memenuhi permintaan pelanggan. Selain itu, sistem informasi akuntansi yang diterapkan dalam ritel juga berperan penting dalam mengelola keuangan dan persediaan. Krisnawati et al. menjelaskan bahwa sistem informasi akuntansi persediaan dapat membantu dalam merencanakan dan mengendalikan operasi bisnis, serta memberikan dasar yang kuat untuk pengambilan keputusan ekonomi.

Dengan adanya sistem informasi yang terintegrasi, perusahaan ritel dapat lebih mudah dalam mengawasi aliran barang dan keuangan, yang pada gilirannya dapat meningkatkan daya saing perusahaan di pasar. Penerapan teknologi informasi dalam sistem informasi ritel juga memungkinkan perusahaan untuk melakukan analisis data yang lebih mendalam. Dengan memanfaatkan data penjualan dan perilaku pelanggan, perusahaan dapat mengidentifikasi tren dan pola yang dapat digunakan untuk merancang strategi pemasaran yang lebih efektif [9]. Hal ini sejalan dengan perkembangan teknologi yang memungkinkan pengumpulan dan

analisis data dalam skala besar, sehingga perusahaan dapat lebih responsif terhadap perubahan kebutuhan pasar.

2.3.1 Konsep Dasar Sistem Informasi Ritel

Sistem informasi ritel adalah kombinasi dari perangkat keras, perangkat lunak, jaringan komunikasi, serta prosedur yang dirancang untuk:

1. Mengelola Transaksi Penjualan

Melakukan pencatatan transaksi secara real-time untuk menghindari kesalahan dan mempermudah pelaporan.

2. Mengoptimalkan Inventaris

Memastikan ketersediaan stok barang berdasarkan permintaan konsumen.

3. Menganalisis Data Pelanggan

Menggunakan data untuk memahami preferensi dan perilaku belanja konsumen.

2.1.2 Manfaat Sistem Informasi Ritel

1. Peningkatan Layanan Pelanggan

Dengan mengetahui pola belanja, toko dapat menyediakan barang yang lebih sesuai dengan kebutuhan pelanggan.

2. Efisiensi Operasional

Mengurangi biaya operasional melalui pengelolaan stok yang lebih baik dan otomatisasi proses.

3. Pengambilan Keputusan Berbasis Data

Data yang dikumpulkan memungkinkan toko membuat keputusan yang lebih terinformasi terkait promosi, tata letak produk, dan strategi penjualan.

2.1.3 Peran Sistem Informasi dalam Ritel Modern

1. Identifikasi Pola Pembelian Konsumen

Menggunakan data transaksi untuk mengetahui barang yang sering dibeli bersama.

2. Strategi Tata Letak Produk

Menempatkan produk yang sering dibeli bersamaan pada rak yang berdekatan untuk meningkatkan penjualan.

3. Strategi Promosi

Memberikan rekomendasi berbasis data untuk penawaran diskon atau bundling produk.

2.4 Algoritma Apriori

Algoritma Apriori adalah salah satu metode yang paling populer dalam data mining, khususnya dalam analisis asosiasi. Algoritma ini digunakan untuk menemukan pola pembelian produk yang sering terjadi dalam transaksi, yang dikenal sebagai market basket analysis. Prinsip dasar dari algoritma Apriori adalah mencari itemset yang sering muncul dalam data transaksi dan menghasilkan aturan asosiasi yang dapat digunakan untuk pengambilan keputusan bisnis [10]. Dalam penerapannya, algoritma Apriori bekerja dengan menggunakan dua parameter utama, yaitu support dan confidence. Support mengukur seberapa sering itemset muncul dalam dataset, sedangkan confidence mengukur seberapa sering itemset yang satu muncul bersamaan dengan itemset yang lain. Misalnya, jika kita menemukan bahwa 70% dari transaksi yang melibatkan produk A juga melibatkan produk B, maka kita dapat menyimpulkan

bahwa ada hubungan yang kuat antara kedua produk tersebut [11]. Penelitian oleh Fahrudin menunjukkan bahwa algoritma ini dapat mengidentifikasi barang-barang yang sering dibeli bersamaan, yang sangat berguna bagi pelaku bisnis untuk merancang strategi pemasaran dan penempatan produk. Salah satu contoh penerapan algoritma Apriori adalah dalam analisis penjualan di berbagai jenis toko, seperti swalayan dan kedai kopi. Dalam studi yang dilakukan di Cerry Mart, algoritma ini berhasil mengidentifikasi pola pembelian konsumen selama bulan tertentu, memberikan wawasan yang berharga untuk meningkatkan penjualan. Selain itu, penelitian di kedai kopi juga menunjukkan bagaimana algoritma ini dapat digunakan untuk menentukan kombinasi menu yang paling diminati oleh pelanggan, sehingga membantu dalam pengembangan strategi promosi. Kelebihan dari algoritma Apriori adalah kemampuannya untuk menangani dataset yang besar dan kompleks, serta menghasilkan aturan yang dapat diinterpretasikan dengan jelas. Namun, algoritma ini juga memiliki kelemahan, terutama dalam hal efisiensi, karena memerlukan beberapa kali pemindaian terhadap database untuk menemukan frequent itemsets [12]. Oleh karena itu, beberapa penelitian telah dilakukan untuk meningkatkan kinerja algoritma ini, seperti dengan menggunakan teknik partisi data untuk mempercepat proses pencarian.

1. Support:

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

2. Confidence:

$$Confidence(A \rightarrow B) = \frac{\text{Jumlah transaksi yang mengandung A dan B}}{\text{Jumlah transaksi yang mengandung A}}$$

Gambar 2.2 Rumus Support dan Confidence

2.5 Association Rules

Metode Association Rules dengan Algoritma Apriori adalah suatu prosedur untuk mencari hubungan antara item dalam suatu kumpulan data yang ditentukan. Algoritma Apriori digunakan untuk menentukan frequent itemset dan aturan asosiasi yang memenuhi syarat minimum support dan minimum confidence. Association Rules dapat digunakan untuk mengetahui kebiasaan atau perilaku pelanggan mengenai produk-produk yang sering dibeli secara bersamaan. Hal ini dapat membantu perusahaan dalam menentukan strategi promosi penjualan yang tepat. Dalam penerapan aturan asosiasi, dua metrik utama yang sering digunakan adalah support dan confidence. Support mengukur seberapa sering itemset muncul dalam dataset, sedangkan confidence mengukur seberapa sering itemset yang satu muncul bersamaan dengan itemset yang lain. Sebagai contoh, dalam penelitian oleh Gama et al., ditemukan aturan asosiasi yang menunjukkan bahwa jika produk A dibeli, maka produk B juga akan dibeli dengan tingkat kepercayaan tertentu.

Dengan demikian, aturan asosiasi tidak hanya membantu dalam memahami pola pembelian konsumen, tetapi juga memberikan wawasan berharga bagi pengusaha untuk merancang strategi pemasaran yang lebih efektif. Penelitian yang dilakukan oleh Fahrudin menegaskan pentingnya analisis asosiasi dalam konteks market basket analysis, yang dapat membantu bisnis dalam mengoptimalkan penempatan produk dan meningkatkan penjualan [13].

2.6 *RapidMiner*

RapidMiner adalah salah satu perangkat lunak yang banyak digunakan dalam bidang data mining dan analisis data. Perangkat ini dirancang untuk memudahkan pengguna dalam melakukan berbagai analisis data, termasuk klasifikasi, regresi, clustering, dan analisis asosiasi. *RapidMiner* menyediakan antarmuka grafis yang intuitif, sehingga memungkinkan pengguna, termasuk mereka yang tidak memiliki latar belakang teknis yang kuat, untuk melakukan analisis data secara efektif. Salah satu keunggulan *RapidMiner* adalah kemampuannya untuk menangani berbagai jenis data dan algoritma. Selain itu, *RapidMiner* juga digunakan dalam berbagai aplikasi, seperti analisis sentimen, pengelompokan data, dan prediksi permintaan, yang menunjukkan fleksibilitasnya dalam berbagai domain.

RapidMiner juga mendukung integrasi dengan berbagai sumber data, baik yang terstruktur maupun tidak terstruktur. Dalam konteks analisis data besar, *RapidMiner* mampu menangani volume data yang besar dan kompleks, yang merupakan tantangan utama dalam era big data saat ini. Misalnya, dalam penelitian *rapidMiner* digunakan untuk menganalisis tingkat pengangguran

dengan menggunakan metode K-Means clustering, yang menunjukkan kemampuannya dalam mengelompokkan data berdasarkan variabel yang relevan . Dalam hal evaluasi dan validasi model, RapidMiner menyediakan berbagai alat untuk mengukur akurasi dan kinerja model yang dibangun. Penelitian oleh Nurahman dan Aminah menunjukkan bahwa RapidMiner dapat digunakan untuk menguji model klasifikasi dengan hasil yang sangat akurat, mencapai 97,85% dalam pengujian algoritma C4.5 [14]. Ini menunjukkan bahwa RapidMiner tidak hanya berguna untuk membangun model, tetapi juga untuk mengevaluasi efektivitas model tersebut. Secara keseluruhan, RapidMiner merupakan alat yang sangat berguna dalam analisis data dan data mining, memberikan kemudahan bagi pengguna untuk melakukan analisis yang kompleks tanpa memerlukan pemrograman yang mendalam. Dengan berbagai fitur dan kemampuannya untuk beradaptasi dengan berbagai jenis data dan algoritma, RapidMiner menjadi pilihan yang populer di kalangan peneliti dan praktisi di bidang data science.



Gambar: 2.3 Rapidminer Studio

Sumber: RapidMiner Studio

2.7 Metodologi Penelitian

Dalam Kamus Besar Bahasa Indonesia (KBBI), metodologi adalah ilmu tentang metode; uraian tentang metode. Sedangkan pengertian penelitian dalam Kamus Besar Bahasa Indonesia (KBBI), penelitian adalah kegiatan pengumpulan, pengolahan, analisis, dan penyajian data yang dilakukan secara sistematis dan objektif untuk memecahkan suatu persoalan atau menguji suatu hipotesis untuk mengembangkan prinsip-prinsip umum. Metodologi penelitian merupakan cara untuk mengetahui hasil dari sebuah permasalahan yang spesifik dari suatu penelitian. Metodologi penelitian ini dilakukan dengan cara sistematis yang akan digunakan sebagai acuan dalam penelitian

2.7.1 Studi Pustaka

Studi pustaka dilakukan dengan menelusuri berbagai sumber literatur yang relevan, seperti jurnal ilmiah, buku referensi, dan penelitian terdahulu yang berkaitan dengan analisis pola pembelian konsumen, algoritma Apriori, serta penerapan data mining dalam sistem informasi ritel. Studi ini bertujuan untuk:

- 1 Memahami teori dasar mengenai data mining dan algoritma Apriori.
- 2 Mengetahui bagaimana algoritma Apriori telah diterapkan dalam berbagai penelitian serupa.
- 3 Menelaah keunggulan dan keterbatasan metode yang digunakan dalam penelitian sebelumnya.

2.7.2 Studi Wawancara

Wawancara dilakukan dengan pihak terkait, seperti manajer toko, kasir, dan pelanggan di Indomaret Simpang Marbau. Studi ini bertujuan untuk mendapatkan perspektif langsung mengenai:

- 1 Pola pembelian konsumen dari sudut pandang manajemen toko dan kasir.
- 2 Kendala yang dihadapi dalam mengelola transaksi dan stok barang.
- 3 Preferensi dan pengalaman pelanggan saat berbelanja di Indomaret Simpang Marbau.

2.8 Penelitian Terdahulu

Adapun Penelitian Terdahulu Dapat kita lihat pada tabel dibawah ini:

1	
Judul	Implementasi Market Basket Analysis dengan Algoritma Apriori pada Toko Elektronik
Nama Penulis	Kevin Brighton, Susanto Hariyanto
Tahun	2024
Hasil	Penelitian ini menunjukkan bahwa algoritma Apriori mampu meningkatkan efisiensi pengelolaan stok barang dan memberikan rekomendasi yang efektif untuk promosi di toko elektronik. Implementasi ini memungkinkan pengelolaan inventori menjadi lebih terstruktur, sehingga penempatan barang yang memiliki hubungan pembelian dapat dioptimalkan
2	
Judul	Analisis Pola Pembelian Konsumen di Toko Ritel DMART Menggunakan

	Algoritma Apriori Berbasis Website
Nama Penulis	Ahmad Najmi Fuadi, Mohamad Herdian Bhakti
Tahun	2023
Hasil	Penelitian ini berhasil menemukan pola pembelian konsumen yang dapat dimanfaatkan untuk mendukung strategi penataan tata letak produk di toko ritel. Selain itu, hasil pola ini juga digunakan untuk menyusun strategi promosi produk yang lebih terarah dan berbasis data aktual.
3	
Judul	Penerapan Algoritma Apriori pada Data Penjualan di Minimarket
Nama Penulis	Tim Penulis Jurnal INFOS
Tahun	2021
Hasil	Penelitian ini menyoroti bagaimana aturan asosiasi dari algoritma Apriori membantu menentukan kombinasi produk yang dapat dipromosikan dalam bentuk bundling. Kombinasi tersebut bertujuan untuk

	meningkatkan jumlah pembelian konsumen serta memperkuat strategi pemasaran minimarket.
4	
Judul	Prediksi Pola Penjualan di Supermarket Menggunakan Algoritma Apriori
Nama Penulis	W. Kuswayati
Tahun	2022
Hasil	Algoritma Apriori dalam penelitian ini mendukung optimalisasi tata letak produk, pengelolaan stok, serta strategi promosi. Penelitian ini juga memberikan wawasan mendalam tentang pola pembelian konsumen, yang dapat digunakan untuk meningkatkan efisiensi operasional dan kepuasan pelanggan.
5	
Judul	Penggunaan Algoritma Apriori untuk Analisis Pola Pembelian Konsumen di Pasar Tradisional
Nama Penulis	D. Ramadhan dan S. Lestari
Tahun	2022

Hasil	Penelitian ini menunjukkan bahwa algoritma Apriori dapat membantu mengidentifikasi pola pembelian konsumen di pasar tradisional. Pola ini digunakan untuk mendukung penempatan produk yang strategis di lokasi tertentu, sehingga memudahkan konsumen dalam menemukan barang yang sering dibeli bersamaan. Dengan demikian, pengelolaan tata letak pasar menjadi lebih efektif, dan peluang peningkatan penjualan pun dapat tercapai.
-------	--