

BAB II

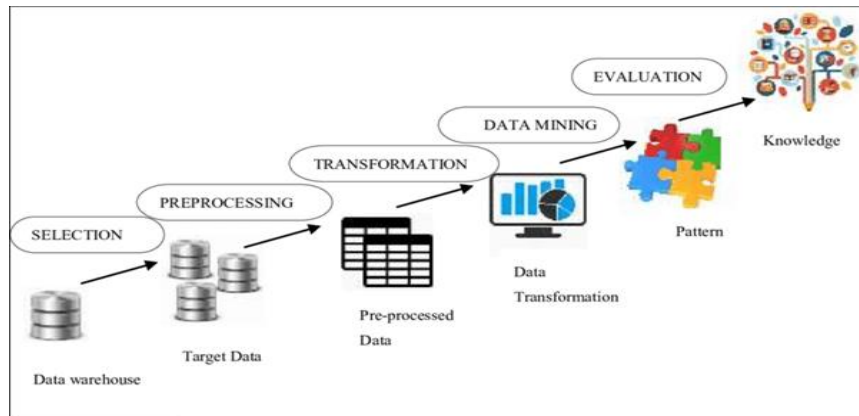
LANDASAN TEORI

2.1. *Knowledge Discovery In Database*

Knowledge Discovery in Database (KDD) merupakan proses yang proses yang melibatkan ekstraksi informasi yang berguna, sebelumnya tidak diketahui, dan informasi yang memiliki potensi berharga dari kumpulan data (Zega & Fauzi, 2023). *KDD* tidak hanya sekadar pengumpulan data, melainkan juga serangkaian langkah-langkah yang terdiri dari pembersihan data, pemilihan data, transformasi, serta penggunaan algoritma untuk menemukan pola-pola yang relevan dan berguna bagi pengambilan keputusan. Proses ini mencakup analisis data yang terstruktur, tidak terstruktur, maupun semi-terstruktur, dan berfokus pada pencarian informasi yang bisa digunakan untuk meningkatkan pemahaman atau untuk mendukung proses bisnis. *KDD* merupakan alur aktivitas untuk menemukan suatu pola dalam data. *KDD* digunakan dalam penemuan pola atau keterkaitan data dalam sekumpulan banyaknya data pada data historis (Syahriani, 2022).

Langkah dalam menganalisis *data mining* dengan menggunakan pendekatan *Knowledge Discovery in Database (KDD)* secara garis besar memiliki beberapa langkah seperti data selection, preprocessing, transformation, data mining dan evaluation (Cindy et al., 2025). *KDD* sendiri melibatkan berbagai disiplin ilmu, seperti statistik, pembelajaran mesin, basis data, dan pemrograman komputer. Tujuan utamanya adalah menghasilkan pengetahuan baru yang sebelumnya tidak diketahui, namun dapat dimanfaatkan secara praktis oleh

pengambil keputusan di sebuah organisasi atau perusahaan. Proses *KDD* dapat digambarkan sebagai rangkaian langkah-langkah yang terdiri dari:



Gambar 2.1. Tahapan Dalam KDD

(Sumber: Lathifah & Dana, 2024)

Pada setiap langkahnya, data diproses dan dikendalikan dengan teknik-teknik yang sesuai, hingga pada akhirnya menghasilkan pola atau pengetahuan yang relevan yang dapat diterapkan untuk memecahkan masalah yang ada. Seperti yang ditunjukkan pada proses sebagai berikut:

1. Seleksi Data (*Data Selection*)

Langkah pertama dalam proses *KDD* adalah memilih dan mengumpulkan data yang relevan dengan masalah yang ingin diselesaikan. Data yang dipilih biasanya berasal dari berbagai sumber, baik internal maupun eksternal, yang akan memberikan wawasan yang berguna.

2. Pembersihan Data (*Data Cleaning*)

Data yang dikumpulkan tidak selalu bersih dan siap pakai. Oleh karena itu, tahap pembersihan data sangat penting. Pada tahap ini, data yang tidak lengkap, duplikat, atau tidak relevan akan dihapus atau diperbaiki.

Pembersihan data membantu mengurangi ketidakakuratan yang dapat mempengaruhi hasil analisis

3. Transformasi Data (Data Transformation)

Setelah data dibersihkan, data sering kali perlu ditransformasi atau dikonversi ke dalam format yang lebih sesuai untuk analisis selanjutnya. Proses ini dapat mencakup normalisasi, penggabungan data dari berbagai sumber, atau penyusunan data dalam bentuk tabel atau format lain yang dapat dipahami oleh algoritma analitik.

4. Penambangan Data (Data Mining)

Penambangan data adalah inti dari proses *KDD*. Di sini, algoritma penambangan data diterapkan untuk menemukan pola atau pengetahuan baru yang tersembunyi dalam data. Salah satu metode penambangan data yang populer adalah **algoritma Apriori**, yang digunakan untuk mencari asosiasi antar item dalam dataset yang besar.

5. Evaluasi Pengetahuan (Knowledge Evaluation)

Setelah pola ditemukan, evaluasi dilakukan untuk menentukan seberapa baik pola tersebut mencerminkan hubungan yang relevan atau berguna dalam konteks yang lebih luas.

2.2. *Data Mining*

Data Mining adalah proses eksplorasi dan analisis data besar untuk menemukan pola, hubungan, atau informasi yang tersembunyi yang dapat digunakan untuk membuat keputusan atau memecahkan masalah. *Data mining* adalah proses yang menggunakan Teknik statistik, matematika, kecerdasan buatan,

dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (ASANA et al., 2022). Dalam konteks ini, data mining menggabungkan teknik-teknik dari berbagai disiplin ilmu seperti statistik, kecerdasan buatan, pembelajaran mesin, dan basis data untuk mengekstrak pengetahuan dari data yang ada. *Data mining* merupakan proses analisis data untuk menemukan suatu pola dari kumpulan data tersebut. *Data mining* mampu menganalisa dan mengolah sekumpulan data menjadi sebuah informasi yang dapat menentukan pola-pola yang dapat menunjang sebuah keputusan (Harahap et al., 2022).

Data Mining khususnya Algoritma *Apriori*, membantu mengelola data dengan menemukan pola asosiasi antara item menggunakan penghitungan support dan confidence, yang berguna untuk analisis pembelian (Parinduri et al., 2024). *Data mining* diartikan sebagai menambang data atau upaya untuk menggali informasi yang berharga dan berguna pada database yang sangat besar (Darmawan et al., 2022). Tujuannya adalah untuk mengidentifikasi informasi yang berguna yang dapat membantu dalam pengambilan keputusan yang lebih baik, peramalan tren di masa depan, atau peningkatan strategi bisnis. *Data mining* dapat diterapkan dalam berbagai bentuk, seperti klasifikasi (untuk mengkategorikan data ke dalam kelas tertentu), klusterisasi (untuk mengelompokkan data yang serupa), asosiasi (untuk menemukan hubungan antara item yang sering muncul bersama), dan regresi (untuk memprediksi nilai kontinu). Dengan menggunakan metode ini, organisasi dapat menggali wawasan yang tidak terlihat sebelumnya, sehingga dapat meningkatkan efisiensi operasional, merencanakan strategi bisnis yang lebih

baik, dan memberikan layanan yang lebih baik kepada pelanggan. *Data mining* sendiri memiliki kemampuan untuk mengolah data dari suatu basis data agar menghasilkan sebuah informasi baru yang bermanfaat (Amsury et al., 2023).

2.2.1. *Association Rules*

Association Rules adalah salah satu teknik dalam *data mining* yang digunakan untuk menemukan hubungan antar *item-item* dalam umlah dataset yang besarnya telah ditentukan (Faishal Zoelfiandi & Utomo Budiyo, 2022). Teknik ini bertujuan untuk mengidentifikasi *item-item* yang sering dibeli bersama dalam satu transaksi, sehingga bisa membantu pengambil keputusan dalam merancang strategi pemasaran, promosi, atau penataan produk yang lebih efektif. Analisis asosiasi dikenal sebagai sebutan *data mining*, banyak peneliti untuk menghasilkan algoritma yang efektif dan berkualitas adalah analisis pola frekuensi tinggi (*frequent pattern mining*).

Association Rules adalah proses untuk mendapatkan pola yang sering muncul, yaitu mengetahui seberapa sering suatu kombinasi *item* tampil dalam sebuah basis data. Asosiasi diproses berdasarkan mekanisme nilai support dan nilai confidence dari hubungan item tersebut (Sari & Khoiriah, 2022). Aturan asosiasi juga merupakan metode yang dapat digunakan, bahkan sering diterapkan ketika mencari pola yang muncul dalam berbagai transaksi penjualan di toko. Setiap algoritma *apriori* pasti menggunakan Aturan Asosiasi untuk mendukung proses analisis data, di mana setiap transaksi dapat terdiri dari beberapa *item* atau produk yang dijual oleh toko, menjadi acuan untuk menghitung *data mining* dengan algoritma *apriori*. Ini mungkin memerlukan pembacaan data transaksi

secara berulang-ulang dalam jumlah data transaksi yang besar untuk menemukan pola hubungan yang berbeda-beda (Tamba & Edbert, 2022). Analisis asosiasi dalam *data mining* mencakup serangkaian teknik untuk menemukan aturan terkait kombinasi beberapa *item*. Aturan asosiasi adalah metode dalam penambangan data untuk menemukan hubungan antara kombinasi elemen yang berbeda.

Salah satu langkah dalam analisis asosiasi yang menarik perhatian banyak peneliti untuk menciptakan algoritma yang efisien adalah seringnya muncul pola-pola selama proses *data mining*. Prinsip dasar dari asosiasi atau analisis hubungan menjelaskan apa yang terjadi saat ini dan kemungkinan yang akan datang. Misalnya, dalam penelitian transaksional di supermarket, ada pembeli yang membeli produk A dan juga ingin membeli produk B. Maka, perlu untuk menemukan kombinasi yang paling sering muncul dari sekumpulan *item*, sering disebut dengan himpunan *item* yang sering muncul. Aturan asosiasi sangat bermanfaat untuk mengidentifikasi hubungan penting antar *item* dalam *data mining* dan setiap transaksi yang ada. Untuk mendukung sistem dalam *data mining*, metode Algoritma *Apriori* adalah pilihan terbaik untuk mendapatkan nilai *support* yang diinginkan serta untuk menentukan nilai terbaik. Hubungan antara *itemset* dapat menunjukkan seberapa kuat atau lemah aturan dalam asosiasi yang sedang dianalisis saat ini. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) yang menarik perhatian banyak peneliti untuk menciptakan algoritma yang efektif. Metode dasar analisis asosiasi dipecah menjadi 2 tahap adalah (Prayuda et al., 2024).

1. Analisis pola sampel frekuensi paling tinggi

$$Supoport (A) = \frac{\text{Jumlah Transaksi yang Mengandung Item A}}{\text{Jumlah Total Transaksi}} * 100\%$$

Rumus 1. Item Set Nilai Support

$$Supoport (A, B) = \frac{\text{Jumlah Transaksi yang Mengandung Item A dan B}}{\text{Jumlah Total Transaksi}} * 100\%$$

Rumus 2. Dua Item Set Atau Lebih

2. Membentuk rule asosiatif

$$Confidance (A \rightarrow B) = \frac{\text{Jumlah Transaksi Mengandung Item A dan B}}{\text{Jumlah Transaksi Mengandung A}} * 100\%$$

Rumus3. Nilai Confidance

2.3. Algoritma Apriori

Algoritma *Apriori* adalah sebuah algoritma yang digunakan dalam data mining untuk menemukan aturan asosiasi (*association rules*) dalam sebuah basis data transaksi. Algoritma ini sangat populer dalam analisis pasar untuk menggali hubungan atau pola pembelian antara produk yang sering dibeli bersama. Algoritma apriori merupakan jenis aturan asosiasi pada metode *data mining* yang memakai *teknik association rules* untuk mencari *frequent itemset* (Al Faridzi & Prehanto, 2022). Tujuan utama dari algoritma *Apriori* adalah untuk menemukan *itemset* yang sering muncul (*frequent itemsets*) dalam transaksi dan kemudian menghasilkan aturan asosiasi yang menghubungkan *item-item* tersebut. Pada algoritma *apriori* menentukan kandidat yang mungkin muncul dengan cara memperhatikan minimum *support* dan minimum *confidence*. *Support* adalah nilai

pengunjung atau persentase kombinasi sebuah *item* dalam *database* (Pamungkas & Fatah, 2025).

Analisis asosiasi didefinisikan sebagai suatu proses untuk menemukan semua aturan asosiasi yang memenuhi syarat minimum untuk *support* (minimum support) dan syarat minimum untuk *confidence* (minimum confidence) (Sumantri, 2023). *Apriori* secara umum adalah mencari *frequent itemset* dari transaksi penjualan melalui beberapa tahap iterasi (Annisa Ilham Fathimah et al., 2022). Algoritma *Apriori* termasuk jenis aturan asosiasi pada *data mining* aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *market basket analysis* (Rahayu et al., 2024). Misalnya, dalam analisis pembelian barang di supermarket, algoritma *Apriori* dapat digunakan untuk mengetahui produk apa saja yang sering dibeli bersamaan oleh pelanggan, seperti roti dan mentega. Algoritma *Apriori* merupakan salah satu algoritma *data mining* yang efektif untuk mengidentifikasi pola asosiasi dalam data transaksi (Jessfry & Siddik, 2024).

2.3.1. Konsep Dasar Algoritma *Apriori*

Konsep algoritma *Apriori* yang berkaitan dengan dengan asosiasi:

1. *Itemset* adalah sekumpulan *item* atau produk yang muncul bersama dalam sebuah transaksi atau dataset.
2. *Frequent itemsets* adalah kombinasi *item* yang muncul lebih sering dari batasan *threshold* tertentu (misalnya, suatu persentase dari total transaksi). Misalnya, dalam konteks belanja, *frequent itemsets* adalah produk-produk yang sering dibeli bersama oleh konsumen.

3. *Support* adalah mengukur seberapa sering *item* atau kombinasi *item* muncul dalam dataset. Algoritma Apriori mencari *itemset* yang memiliki *support* lebih besar dari nilai minimum yang ditentukan.
4. *Confidence* adalah mengukur seberapa besar kemungkinan *item* B akan muncul jika *item* A sudah muncul. Ini mengukur kuatnya hubungan antar *item*.

2.3.2. Prinsip Kerja Algoritma *Apriori*

Prinsip utama dari algoritma *Apriori* adalah "sebuah *itemset* yang sering muncul (*frequent itemset*) pada suatu dataset, semua subset dari *itemset* tersebut juga harus sering muncul". Dengan prinsip ini, algoritma *Apriori* dapat mengurangi jumlah kombinasi *item* yang perlu diperiksa, sehingga proses pencarian aturan asosiasi menjadi lebih efisien.

Secara umum, algoritma *Apriori* beroperasi dengan langkah-langkah berikut:

1. Menemukan *Frequent Itemsets*

Pada langkah pertama, algoritma akan memindai seluruh transaksi untuk menemukan *itemset* yang sering muncul. Misalnya, jika dalam suatu dataset ada kombinasi produk yang sering dibeli bersamaan, maka kombinasi tersebut dianggap sebagai *frequent itemset*. Proses ini dimulai dengan *itemset* tunggal (*item* yang hanya berisi satu produk), kemudian *itemset* yang lebih besar (kombinasi dua atau lebih produk) diperiksa untuk menentukan apakah mereka memenuhi ambang batas *support* yang telah ditentukan.

2. Membangun Aturan Asosiasi

Setelah *itemset* yang sering ditemukan, algoritma kemudian menghasilkan aturan asosiasi. Aturan asosiasi adalah hubungan berbentuk "Jika A, maka B", yang mengindikasikan bahwa jika *item* A dibeli, maka kemungkinan besar *item* B juga akan dibeli. Aturan ini dihitung berdasarkan dua parameter utama: *support* dan *confidence*.

3. *Pruning* (Pemangkasan)

Proses pemangkasan adalah langkah penting dalam mengurangi jumlah kombinasi *itemset* yang harus diproses. Algoritma *Apriori* menggunakan sifat *anti-monotonicity*, yang berarti bahwa jika suatu *itemset* tidak memenuhi ambang batas *support*, maka kombinasi yang lebih besar yang mengandung *itemset* tersebut juga tidak akan memenuhi batas *support*. Dengan cara ini, banyak *itemset* yang tidak perlu diperiksa, sehingga mempercepat proses komputasi.

4. Menghasilkan Hasil

Setelah aturan asosiasi yang memenuhi kriteria tertentu ditemukan, hasilnya dapat digunakan untuk membuat keputusan bisnis yang lebih baik. Misalnya, supermarket dapat menggunakan aturan asosiasi untuk merancang promosi *bundling* produk atau menempatkan produk yang sering dibeli bersama di rak yang berdekatan.

2.4. **Prediksi Penjualan Bolen**

Prediksi penjualan bolen sangat penting bagi produsen atau perusahaan yang ingin memastikan ketersediaan produk yang tepat pada waktu yang tepat,

menghindari *overstocking* atau *understocking*, dan merencanakan produksi dengan lebih efisien.

Dengan mempertimbangkan faktor-faktor eksternal dan menggunakan metode yang tepat, seperti analisis teknik *data mining*, prediksi penjualan dapat dilakukan dengan lebih akurat. Ini memungkinkan Bolen Bunda Ibra untuk meningkatkan keuntungan, mengoptimalkan sumber daya, dan menyajikan produk yang lebih sesuai dengan keinginan konsumen.

2.5. *RapidMiner*

Aplikasi *RapidMiner* merupakan aplikasi yang digunakan untuk proses data mining. Didalam aplikasi ini terdapat banyak algoritma yang terkandung guna kebutuhan *data mining* (Romdon & Kholil, 2022). *RapidMiner* adalah sebuah *platform* perangkat lunak *open-source* yang digunakan untuk analisis data dan pembelajaran mesin (*machine learning*). *Platform* ini menawarkan alat yang kuat untuk *data mining*, analisis prediktif, dan pemrosesan data dalam berbagai aplikasi industri, penelitian, dan pengembangan. *RapidMiner* dirancang untuk memudahkan pengguna, baik yang berpengalaman maupun pemula, untuk mengolah, menganalisis, dan memvisualisasikan data dalam satu lingkungan yang terpadu. Aplikasi ini memungkinkan pengguna untuk memasukkan data mentah, termasuk database dan teks, yang kemudian dianalisis secara otomatis dan cerdas dalam skala besar (Nurhidayanti & Kurniawati, 2022).

RapidMiner memiliki antarmuka pengguna grafis (GUI) yang memungkinkan pengguna untuk mengembangkan alur kerja (*workflow*) analisis data dengan cara yang sangat visual. Pengguna dapat menarik dan melepaskan

komponen (seperti operator untuk pembersihan data, transformasi, model pembelajaran mesin, dll.) untuk membangun alur analisis mereka tanpa perlu menulis kode secara langsung. Ini menjadikan *RapidMiner* sangat ramah bagi pengguna yang tidak memiliki latar belakang pemrograman.

RapidMiner mendukung berbagai teknik analisis data, termasuk **data mining**, **pemrosesan bahasa alami (NLP)**, **pembelajaran mesin** (*machine learning*), **analisis prediktif**, dan **analisis teks**. Platform ini juga dapat digunakan untuk berbagai jenis proyek, mulai dari analisis pasar, prediksi tren penjualan, deteksi penipuan, analisis pelanggan, hingga peramalan data.

Design Perspective adalah tampilan utama di mana pengguna bekerja di *RapidMiner*. Ini digunakan sebagai tempat untuk merancang dan mengelola proses analisis. Tampilan ini menampilkan berbagai tampilan dengan fungsi masing-masing, yang mendukung proses analisis data. Sementara itu, *Result Perspective* adalah tampilan di mana hasil analisis ditampilkan dalam berbagai bentuk, seperti grafik, teks, pohon keputusan, tabel, tergantung pada data dan metode analisis yang digunakan (Lathifah & Dana, 2024). Pengguna dapat mengimpor data dari berbagai format file, seperti CSV, Excel, database SQL, serta layanan *cloud* atau *platform big data* seperti *Hadoop*. Dengan menggunakan *RapidMiner*, pengguna dapat mengolah dan membersihkan data mentah, mengekstrak fitur penting, dan melakukan analisis yang lebih dalam untuk mendapatkan wawasan yang berguna.

RapidMiner juga mendukung berbagai algoritma pembelajaran mesin, baik yang bersifat **terawasi** (*supervised*) seperti regresi dan klasifikasi, maupun yang

tidak terawasi (*unsupervised*) seperti klasterisasi dan asosiasi. *Platform* ini memungkinkan pengguna untuk membangun, melatih, dan menguji model pembelajaran mesin tanpa memerlukan pemahaman teknis yang mendalam tentang algoritma yang mendasarinya. Dengan demikian, *Rapid Miner* dapat digunakan untuk integrasi dan transformasi data (ETL) selain analisis data dengan hasil yang luar biasa, bersama dengan ratusan operator untuk pra-pemrosesan data. Selain itu, pelanggan dapat memilih dari berbagai format perangkat lunak (Fatah et al., 2024). Platform ini dikembangkan untuk mendukung berbagai kegiatan seperti bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran, dengan menyediakan lebih dari 100 solusi untuk pengelompokan, klasifikasi, dan analisis regresi (Purba et al., 2025).



Gambar 2.2. Logo *Rapidminer*

(Sumber: <https://www.prnewswire.com/news-releases/rapidminer-announces-record-revenue-for-q2-2017-300505241.html>)

2.6. Microsoft Excel

Microsoft Excel merupakan perangkat lunak spreadsheet populer, banyak digunakan dalam bisnis untuk berbagai tugas, termasuk penambangan data. Excel memiliki banyak alat bawaan yang dapat digunakan untuk penambangan data,

menjadikan penambangan data di excel sebagai cara yang berguna bagi bisnis untuk mengekstrak wawasan berharga dari data mereka. Alat analisis data excel dapat membantu bisnis melakukan berbagai Teknik penambangan data, seperti analisis regresi, pengelompokan, dan klasifikasi.

Selain itu, excel dapat terhubung ke sumber data eksternal, seperti basis data dan layanan web, sehingga memudahkan impor dan analisis kumpulan data besar. Excel juga mendukung alat visualisasi data seperti bagan dan grafik, yang dapat membantu bisnis menyajikan data mereka dengan lebih mudah dipahami dan mendalam.



Gambar 2.3. Logo Microsoft Excel

(Sumber: https://id.wikipedia.org/wiki/Berkas:Microsoft_Office_Excel_%282018%E2%80%93present%29.svg)

2.7. Penelitian Terdahulu

Bagian ini menyajikan penelitian terdahulu yang relevan sebagai dasar acuan dan pembandingan dalam pengembangan penelitian yang dilakukan.

Tabel 2.1 Penelitian Terdahulu

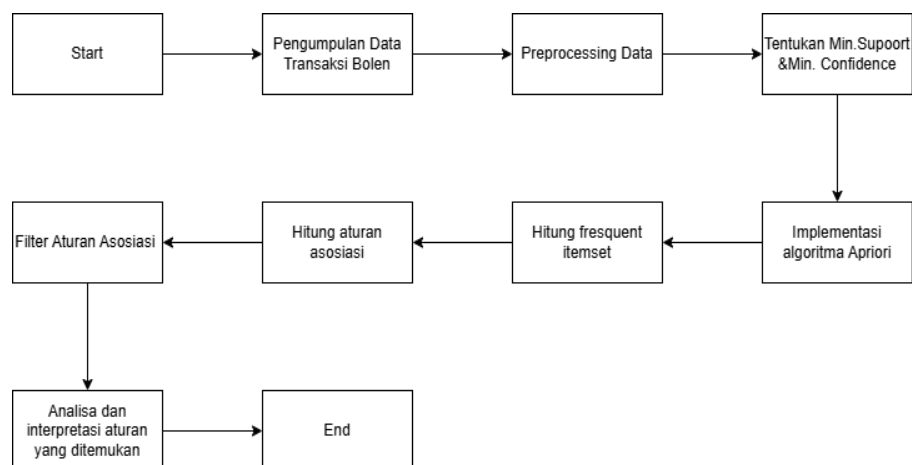
No	Judul	Nama Penulis	Tahun	Hasil
1.	Implementasi Algoritma	Dina Nurhidayanti,	2022	Algoritma Apriori

	Apriori Dalam Menemukan Association Rules Pada Persediaan Sparepart Motor	Ika Kurniawati		digunakan untuk menemukan 3 aturan asosiasi kuat pada penjualan sparepart untuk membantu pengaturan stok
2.	Penerapan Data Mining Menggunakan Algoritma Apriori Dalam Memprediksi Penjualan Kain Tenun Sumba	Elfrida Bulu, Gergorius K	2024	Menggunakan Algoritma Apriori untuk menemukan produk kain tenun yang paling banyak terjual, dengan sarung baju, kain tenun wewewa, dan kain tenun loura memiliki nilai support dan confidence tinggi
3.	Implementasi Data Mining dengan Metode Apriori Dalam Menentukan Pola Pemilihan Pemeriksaan Kimia	Denih Romdon, Iskhak Kholil	2022	Menggunakan Apriori untuk Menghasilkan 3 aturan asosiasi pemeriksaan kimia dengan support dan confidence tinggi.
4.	Implementasi Data Mining dalam Memprediksi Stok Herbal Menggunakan Algoritma	Fitriani Harahap, Nidia Enjelita Saragih, Ester Dayanti Paskaria Situmeang,	2022	Menganalisis data transaksi penjualan herbal, dengan hasil aturan asosiasi untuk produk yang

	Apriori	Elida Tuti, Erwin Ginting, Wirhan Fahrozi		sering dibeli bersama, seperti Madu Hutan Baarik 800gram dan Madu Hutan Baarik 470 gram, serta Habbatussauda cap Kurma Ajwa.
5.	Penerapan Data Mining Menggunakan Algoritma Apriori Dalam Membangun Sistem Persediaan Barang	Vincent Jessfry, Muhammad Siddik	2024	Penelitian ini menggunakan Menghasilkan 4 aturan asosiasi untuk mendukung pengelolaan stok barang lebih efisien. Apriori untuk

2.8. Kerangka Penelitian

Kerangka penelitian ini digunakan untuk memastikan proses penelitian berjalan secara terarah dan sistematis.



Gambar 2.4 Kerangka Kerja Penelitian