

BAB II

LANDASAN TEORI

2.1 Teori Dasar

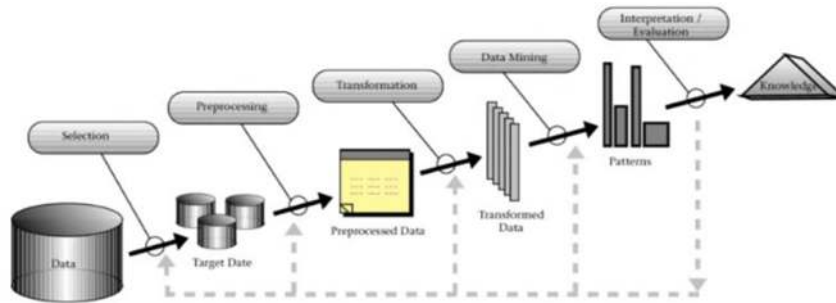
Peneliti akan menguraikan faktor-faktor yang mendukung penyelesaian topik ini. Ini termasuk pemahaman mengenai *Knowledge Discovery in Database* (KDD), Praktik *Data Mining*, penerapan algoritma *FP-Growth*, serta analisis dan penjelasan *software* pendukung seperti *RapidMiner*.

2.2 *Knowledge Discovery in Database*

KDD atau *Knowledge Discovery in Database* merupakan suatu proses di mana data historis digunakan untuk mengidentifikasi pola, keteraturan, dan hubungan yang mungkin tidak diketahui sebelumnya dalam catatan yang tersimpan dalam basis data besar, dengan tujuan mendapatkan wawasan yang bernilai (Bunda, 2020). Selama proses Saat mengumpulkan data, KDD menggunakan data historis untuk mengungkap pengetahuan, informasi, pola, keteraturan, dan hubungan dalam kumpulan data yang besar. Kegiatan KDD sendiri adalah langkah yang diambil untuk mendapatkan wawasan baru dari data yang ada (Fajrin & Handoko, 2018).

Penemuan pengetahuan dalam database (*Knowledge Discovery in Database/KDD*) mencakup ekstraksi potensi yang tersembunyi dan data yang tidak dikenal. Melalui langkah-langkah *Data Mining*, yang merupakan tahapan dalam KDD, informasi diubah menjadi hasil yang lebih akurat dan lebih sederhana dipahami (Dogan & Birant, 2021). Oleh karena itu, KDD dapat disimpulkan sebagai proses kunci untuk mengidentifikasi pola dalam data dan menghasilkan

bentuk informasi yang baru, bermanfaat, dan mudah dimengerti, yang didasarkan pada wawasan yang diperoleh melalui *Data Mining*.



Gambar 2.1 KDD

Langkah demi langkah dalam *Knowledge Discovery in Database* (KDD) adalah sebagai berikut:

1. **Selection:** Data yang relevan harus dipilih dari berbagai sumber yang tersedia.
2. **Preprocessing:** Data yang telah terpilih harus dibersihkan dan siap digunakan, termasuk mengatasi data yang hilang atau tidak valid.
3. **Transformation:** Proses *Data Mining* melibatkan penggunaan komputer untuk mengidentifikasi pola dan informasi penting dalam data yang ada.
4. **Data Mining:** Hasil dari *Data Mining* dianalisis untuk memeriksa relevansinya.
5. **Interpretation / Evaluation:** Hasil yang ditemukan diinterpretasikan dan sering kali diwujudkan dalam bentuk visual.
6. **Penerapan Pengetahuan:** Pengetahuan yang ditemukan digunakan untuk membuat keputusan yang lebih baik terhadap pola penjualan yang di telah ditemukan melalui perhitungan *FP-Tree*

Proses KDD juga memberikan kerangka kerja yang sistematis dan terstruktur untuk melibatkan langkah-langkah pemilihan data, preprocessing, penggalian data, evaluasi, interpretasi, dan penerapan pengetahuan. Ini membantu peneliti dalam menjalankan eksperimen dan analisis data dengan lebih terarah dan efisien

Selain itu, KDD memungkinkan para peneliti untuk menghasilkan bukti empiris yang kuat berdasarkan analisis data yang obyektif, yang sangat penting dalam penelitian ilmiah. Dengan menggunakan pendekatan KDD, peneliti dapat mendukung temuan mereka dengan fakta-fakta yang diperoleh dari data, yang meningkatkan validitas dan kepercayaan hasil penelitian mereka.

Dalam rangkaian penelitian ilmiah *Data Mining*, KDD berfungsi sebagai alat yang memadai untuk menggali pengetahuan, memvalidasi hipotesis, dan memberikan landasan yang kuat bagi pengambilan keputusan yang berbasis data. Dengan demikian, KDD menjadi penting dalam konteks penelitian ilmiah karena memungkinkan peneliti untuk menghadapi data dengan pendekatan yang ilmiah, sistematis, dan efektif untuk memunculkan wawasan yang berharga.

2.3 *Data Mining*

Data Mining merujuk pada proses mengungkap pola atau pengetahuan berharga dari sejumlah besar data. Praktek ini melibatkan penerapan teknik statistik, kecerdasan buatan, dan matematika komputasional untuk menganalisis kumpulan data dan mengidentifikasi pola yang tersembunyi di dalamnya. Sejarah *Data Mining* dapat ditelusuri kembali ke tahun 1930-an saat para statistisi pertama kali menggunakan teknik ini untuk menganalisis data sosial-ekonomi.

Data Mining merupakan proses yang sangat relevan dan berharga dalam analisis data penjualan barang (Antari et al., 2016). Dalam konteks ini, *Data Mining* dapat membantu mengidentifikasi pola pembelian yang penting, preferensi konsumen, dan tren penjualan yang dapat membimbing pengambilan keputusan yang lebih efektif. Dengan menggunakan teknik *Data Mining*, peneliti dapat menggali wawasan mendalam tentang perilaku pembelian konsumen, hubungan antara produk, dan faktor-faktor lain yang memengaruhi pola penjualan (AHMAD ADRI, 2021).

Metode *Data Mining*, peneliti dapat menganalisis pola pembelian bersama, mengungkap korelasi antara item produk, dan mengidentifikasi faktor-faktor yang memengaruhi keputusan pembelian (Kurnia et al., 2019). Ini membantu perusahaan mengoptimalkan strategi pemasaran, mengembangkan rekomendasi produk yang lebih sesuai, dan meningkatkan pengalaman pelanggan secara keseluruhan. Selain itu, *Data Mining* memungkinkan peneliti membuat prediksi penjualan berdasarkan pola penjualan sebelumnya, sehingga perusahaan dapat merencanakan stok barang dan strategi pemasaran secara lebih efisien.

Secara keseluruhan, dalam penelitian data penjualan barang, *Data Mining* berperan krusial dalam mengungkap pola dan informasi penting yang meningkatkan pemahaman tentang preferensi konsumen, membantu pengambilan keputusan, dan mendukung perusahaan dalam mencapai kinerja penjualan yang lebih optimal.

2.4 Metode *FP-Growth*

Saat menggunakan apa yang disebut teknik penambangan data, aturan asosiasi dilibatkan. Penggunaan teknik data mining melibatkan aturan asosiasi. Aturan asosiasi mengacu pada teknik penting dalam penambangan data yang bertujuan untuk menemukan hubungan dan korelasi antar objek dalam *database*. Metode ini mengandalkan analisis statistik dan kecerdasan buatan untuk mengidentifikasi interaksi kondisional antara sejumlah besar data masukan dan menghasilkan aturan keputusan dalam bentuk IF-THEN. (Ait-Mlouk et al., 2017).

Praktek *data mining* diprakarsai oleh Charu C. Aggarwal melalui analisis *database* transaksional. Aturan asosiasi didefinisikan sebagai persyaratan format umum $A \rightarrow B$ ($A, B \subset I$ dan $A \cap B = \emptyset$). Setiap aturan terdiri dari dua kumpulan elemen berbeda, dengan A sebagai item yang mendahului, dan B sebagai item yang mengikuti. Sebagai contoh, {Pengemudi} \rightarrow {Kendaraan}, menunjukkan korelasi yang signifikan antara kedua item {Pengemudi, Kendaraan}.

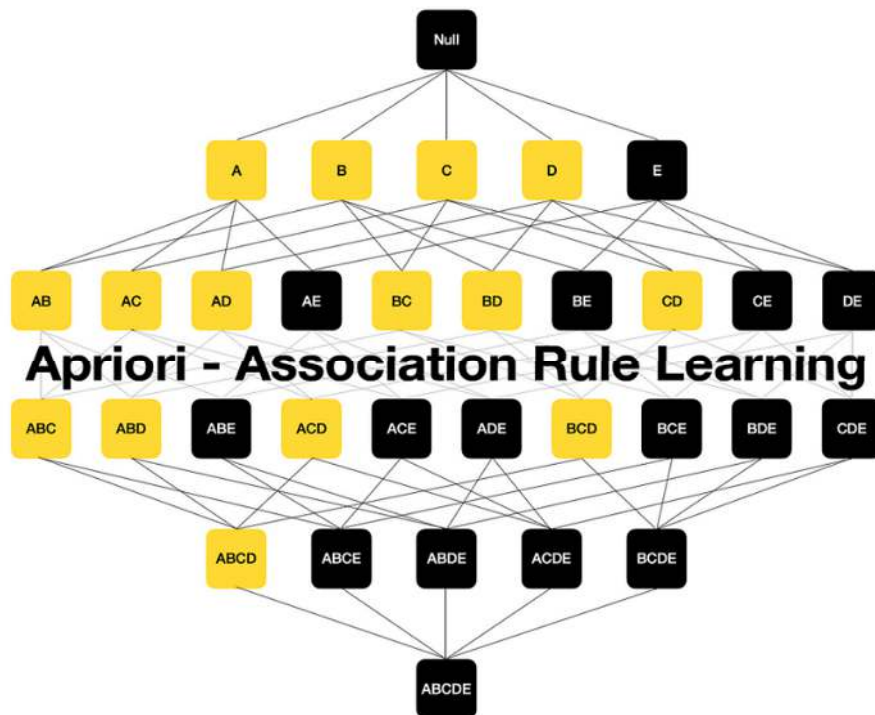
Seperti yang dijelaskan oleh (Antari et al., 2016), menciptakan aturan hubungan antar gabungan elemen merupakan sistem penting dalam *Data Mining*. Proses pemrosesan dan perhitungan data tertentu dilakukan untuk menentukan kepentingan tertentu (*actions of interest*) Tetapkan aturan asosiasi. Secara umum, aturan asosiasi memiliki dua metrik penting. Yaitu:

1. *Support* adalah keseluruhan transaksi yang mewakili kemungkinan pembelian beberapa produk secara bersamaan. Metrik ini mengukur apakah suatu item atau kelompok item layak diselidiki berdasarkan nilai *Confidence*-nya. Nilai *Support* yang tinggi berarti produk X dan produk Y cenderung dibeli secara bersamaan.

2. *Confidence*, atau tingkat kepercayaan, adalah kemungkinan bahwa suatu produk akan dibeli secara bersamaan dengan produk lain berdasarkan beberapa kejadian yang mungkin terjadi (contohnya, frekuensi penjualan item Y ketika konsumen membeli item X).

2.4.1 Aturan Asosiasi

Tujuan utamanya adalah menunjukkan hubungan asosiatif antara jenis produk yang dibeli oleh pelanggan, memungkinkan identifikasi pola dalam bentuk pembelian produk yang sering dilakukan oleh pelanggan. Dengan mengetahui jenis produk yang umum dibeli, informasi tersebut dapat digunakan sebagai landasan untuk menentukan produk mana yang sebaiknya dipromosikan kepada pelanggan.



Gambar 2. 2 Apriori

Konsep aturan asosiasi merupakan elemen penting dalam analisis data dan membantu mengidentifikasi hubungan antar elemen dalam kumpulan data. Dalam penelitian yang dilaksanakan oleh Agrawal, Imielinski, dan Swami (1993), Aturan asosiasi didefinisikan sebagai aturan yang menghubungkan kelompok elemen atau atribut dengan kelompok elemen atau atribut lain dalam kumpulan data. Penerapan aturan asosiasi merupakan hal yang umum dalam *Data Mining* untuk mengungkap korelasi antar elemen atau atribut yang sering terjadi bersamaan dalam suatu konteks.

Proses penemuan aturan asosiasi melibatkan dua langkah penting, yakni pembentukan kandidat aturan dan pengujian kandidat aturan. Langkah pertama adalah membuat aturan kandidat Dengan memilih elemen atau atribut yang sering muncul bersamaan dalam suatu *dataset*. Langkah kedua kemudian menguji aturan kandidat dengan menghitung tingkat *Support* dan *Confidence* untuk menentukan aturan mana yang memiliki nilai di atas ambang batas tertentu. tingkat dukungan menunjukkan seberapa sering aturan tersebut muncul dalam kumpulan data, dan tingkat *Confidence* mengukur seberapa andal aturan tersebut.

Penerapan aturan asosiasi dalam analisis data memberikan banyak manfaat di berbagai bidang. Misalnya, dalam analisis keranjang pasar, aturan asosiasi membantu toko dan bisnis mengidentifikasi produk yang sering dibeli konsumen secara bersamaan. Dengan memahami hubungan antar produk, Toko dapat meningkatkan strategi penjualan lintas produk dan meningkatkan pengalaman berbelanja pelanggan. Selain itu, aturan asosiasi digunakan dalam sistem

rekomendasi yang merekomendasikan produk yang sesuai berdasarkan perilaku dan preferensi konsumen.

2.5 Objek Penelitian

Gudang Computer adalah sebuah perusahaan yang berdiri pada bulan Oktober 2020 dan berlokasi di Jl. Duyung, Ruko *DCMALL* Block A 25, Kota Batam, Kepulauan Riau. Sebagai perusahaan yang bergerak dalam industri komputer, *Gudang Computer* menyediakan berbagai macam produk dan layanan terkait teknologi, seperti penjualan perangkat keras komputer, perangkat lunak, serta layanan perbaikan dan perawatan komputer. Meskipun telah beroperasi selama beberapa tahun, perusahaan ini menghadapi tantangan dalam mengoptimalkan penjualan dan stok barangnya.

Salah satu masalah utama yang dihadapi oleh *Gudang Computer* adalah kurangnya pemahaman tentang pola penjualan mereka. Hingga saat ini, perusahaan ini belum pernah melakukan perhitungan *Data Mining* terhadap data penjualan mereka. Akibatnya, mereka belum memiliki pemahaman yang mendalam tentang tren pembelian pelanggan, produk-produk yang paling diminati, atau pola pembelian bersamaan. Hal ini mengakibatkan pengelolaan stok barang yang kurang efisien dan potensi peningkatan penjualan yang belum dimaksimalkan.

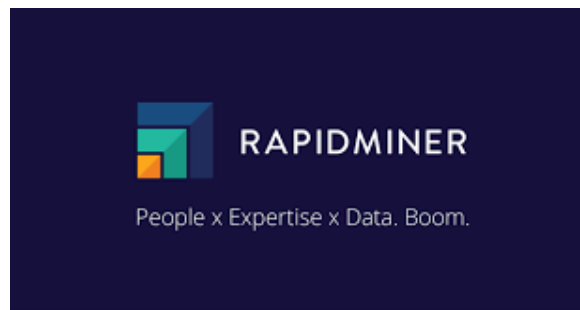


Gambar 2. 3 Gudang Computer

Oleh karena itu, penelitian ini akan menjadikan "*Gudang Computer*" sebagai objek penelitiannya. Dengan menerapkan algoritma *FP-Growth* dalam analisis data penjualan perusahaan ini, penelitian ini bertujuan untuk mengidentifikasi pola-pola pembelian yang sering terjadi, mengoptimalkan persediaan stok barang, dan membantu perusahaan dalam mengambil keputusan yang lebih cerdas. Dengan demikian, diharapkan *Gudang Computer* dapat meningkatkan efisiensi operasionalnya dan mencapai penjualan yang lebih maksimal dalam bisnisnya.

2.6 *Software RapidMiner*

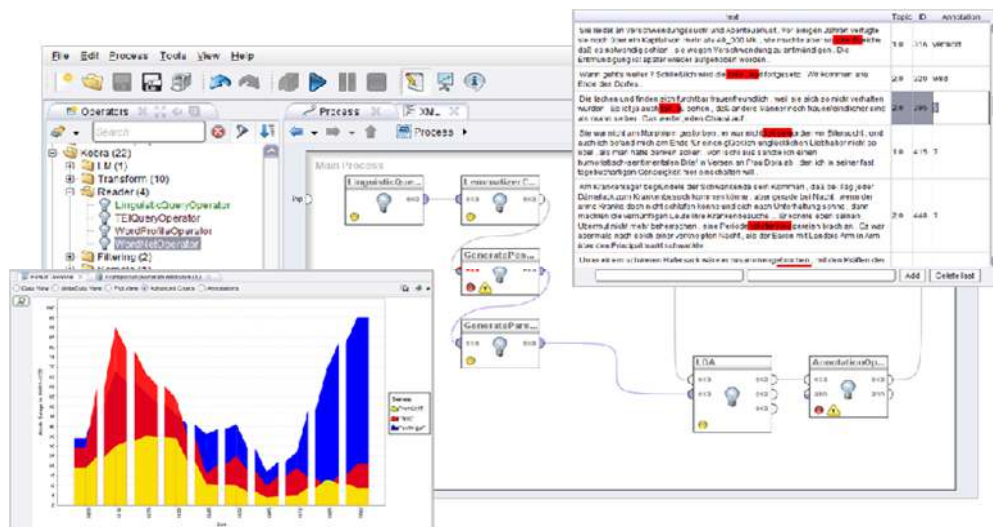
RapidMiner adalah perangkat lunak dengan kekuatan pemrosesan data yang kaya dan beragam fungsi dan operator unik. Software ini dapat digunakan mulai dari tahap awal pemrosesan dan pengeditan data hingga tahap pembelajaran mesin di mana data yang dikelola dapat dimodifikasi.



Gambar 2. 4 RapidMiner

Sebagai salah satu platform terkemuka dalam penambangan data dan analisis prediktif, *RapidMiner* menyediakan serangkaian fitur canggih untuk mengekstraksi wawasan dari data, membangun model prediktif, dan melakukan analisis statistik. Penggunaan *RapidMiner* menawarkan berbagai manfaat potensial untuk berbagai jenis organisasi.

RapidMiner juga memiliki kemampuan untuk menampilkan hasil analisis secara intuitif dan mudah dipahami. Platform ini menyediakan visualisasi data dan kemampuan pelaporan yang memungkinkan pengguna mengkomunikasikan wawasan penting secara efektif kepada pemangku kepentingan. Manfaat ini meningkatkan kemampuan pengambilan keputusan dan mempermudah mengkomunikasikan hasil analisis kepada tim dan pemangku kepentingan yang terlibat dalam proses pengambilan keputusan.



Gambar 2. 5 UI RapidMiner

2.7 Penelitian Terdahulu

Berdasarkan studi sebelumnya yang relevan dengan topik penelitian dan metode penelitian, terdapat beberapa penelitian terkait yang telah dilakukan sebelumnya. Beberapa dari penelitian tersebut meliputi topik yang serupa dengan judul penelitian saat ini:

1. (Hamid Mughal, 2018) “*Web Data Mining has evolved into a convenient and crucial platform for extracting valuable information. Users exhibit a greater preference for the World Wide Web when it comes to uploading and downloading data. With the escalating volume of data on the internet, the process of uncovering meaningful insights and patterns is becoming challenging and time-consuming.*”
2. (Butar & Elisa, 2022) “*Algoritma Apriori dapat Digunakan dalam sistem rekomendasi produk. Pengujian dengan menggunakan waktu perdagangan 20 hari, nilai support minimal 20%, dan nilai Confidence minimal 70% menunjukkan hasil yang memenuhi syarat tersebut. Ketika diuji menggunakan algoritma Apriori dan aplikasi yang dibuat, diperoleh hasil yang sesuai dengan nilai dukungan dan keyakinan yang diharapkan.*”
3. (Fajrin & Handoko, 2018) “*Metode penambangan data yang disebut aturan asosiasi. Istilah aturan asosiasi mengacu pada teknik penambangan data yang kuat untuk menemukan korelasi dan hubungan antar objek dalam database. Hal paling penting dalam teknik Data Mining adalah aturan untuk menemukan pola frekuensi tertinggi antara set itemsets. Data Mining juga memiliki tujuan untuk menemukan pengetahuan baru dari data yang telah diproses. Selama beberapa*

dekade terakhir, metode baru telah memiliki kemampuan untuk dikembangkan seputar pengumpulan dan pembangkitan data”

4. (AHMAD ADRI, 2021)“Dengan bantuan *Data Mining*, *Yati Cosmetics* dapat mengetahui tren penjualan kosmetik terpopuler di kalangan konsumen dan mengoptimalkan penawaran kosmetik yang tersedia. Analisis ini dapat bermanfaat bagi perbaikan sistem penjualan dan memberikan pemahaman yang lebih mendalam tentang preferensi pelanggan. Hasil analisis data juga memberikan aturan untuk membantu pengambilan keputusan, sehingga toko dapat meningkatkan efisiensi dan efektivitas bisnisnya.”
5. (Ashma Nurmeila et al., 2020) “Hasil penelitian menunjukkan *drum dc* dan *toner* adalah produk terlaris. Kesimpulan yang dihasilkan menunjukkan bahwa produk *toner* dan *drum dc* memberikan beberapa dukungan. (*support*) sebesar 50% dan tingkat kepercayaan (*Confidence*) sebesar 66,67%. Selain itu, produk *drum dc* dan *toner* juga memiliki tingkat dukungan sebesar 50% dan tingkat kepercayaan sebesar 85,71%.”
6. (Prasetya et al., 2021) “*This study delves into market basket analysis through Data Mining, specifically employing the Apriori algorithm, to unravel sales patterns at the O! Fish restaurant. The research introduces a web-based application designed to scrutinize consumer buying behaviors, enabling the formulation of promotional strategies tailored to the restaurant's needs. The outcome of the investigation manifests as a percentage derived from the total sales over the course of one year.*”

7. (Kumar & Mohbey, 2022)“*Utilizing Machine Learning and Data Mining in Manufacturing*” explores the advantages of incorporating machine learning techniques within the manufacturing industry. It outlines approaches to address specific challenges and suggests potential areas for future research. The vast amount of data from diverse sources, presented in different formats, qualifies as big data. Harnessing this extensive dataset allows for making informed decisions through diverse predictive analyses. The practice of making intelligent decisions across various scenarios using substantial datasets is commonly referred to as predictive analysis.”
8. (Kurnia et al., 2019) “Penerapan aturan asosiasi dalam analisis data memberikan manfaat yang banyak dalam berbagai bidang. Contohnya, Dalam analisis keranjang belanja, aturan asosiasi membantu toko dan bisnis meningkatkan strategi penjualan lintas produk dan meningkatkan pengalaman berbelanja pelanggan. Selain itu, aturan asosiasi juga digunakan dalam sistem rekomendasi untuk merekomendasikan produk yang sesuai berdasarkan perilaku dan preferensi konsumen.”
9. (Simanjuntak & Elisa, 2019) “Kartu perdana prabayar harganya terjangkau, sangat mudah didapat, dan biasanya tersedia dengan harga murah di banyak tempat. Penelitian ini bertujuan untuk memilih peta *mobile* yang optimal di Kota Batam dengan menggunakan *data mining* dan teknik pencocokan berbasis algoritma apriori. Penelitian ini dilakukan dengan menggunakan *software Tanagra 4.1*.”

10. (Simanjuntak et al., 2022) “*Data Mining* merupakan proses yang sangat relevan dan berharga dalam analisis data penjualan barang. Dalam konteks ini, *Data Mining* dapat membantu mengidentifikasi pola pembelian yang penting, preferensi konsumen, dan tren penjualan yang dapat membimbing pengambilan keputusan yang lebih efektif dan mengintegrasikan teknik dari berbagai disiplin ilmu, seperti statistika, teknologi pembelajaran, mesin basis data, pengenalan pola, jaringan saraf, pencarian informasi, analisis spasial data, dan berbagai aplikasi lainnya.”

2.8 Kerangka Pikiran

Kerangka pikiran merupakan sebuah gambaran yang menjelaskan penelitian dilakukan. Berikut ini merupakan kerangka pikiran yang telah dibuat:



Gambar 2. 6 Kerangka Pikiran

Pada langkah awal penelitian melibatkan pengumpulan data penjualan barang yang digunakan sebagai input proses. Data input ini terdiri dari informasi terkait produk atau barang, termasuk merek dan jenis produk, serta jumlah total transaksi yang telah dilakukan. Data ini kemudian disaring sebelum diproses oleh algoritma dengan memperhatikan nilai *Support* minimum dan *Confidence* minimal. Langkah selanjutnya adalah pengujian data menggunakan perangkat lunak *RapidMiner* untuk memastikan hasil yang diperoleh akurat. Proses ini bertujuan untuk mengidentifikasi produk yang paling sering dibeli oleh konsumen, sehingga informasi tersebut dapat berguna bagi toko yang ingin memahami pola penjualan.