

BAB II

TINJAUAN PUSTAKA

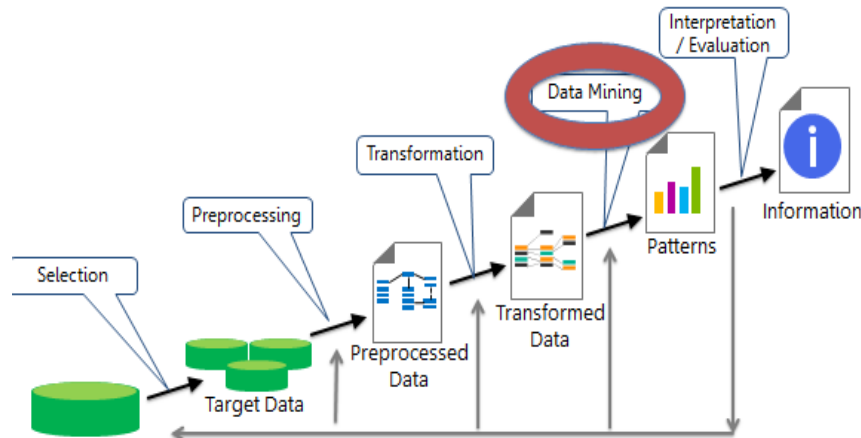
2.1 Teori dasar

Peneliti akan membahas elemen-elemen yang membantu menyelesaikan masalah ini, termasuk penemuan pengetahuan dalam database (KDD), *data mining*, algoritma *FP-Growth*, dan alat bantu seperti *RapidMiner*.

2.1.1 *Knowledge discovery in database (KDD)*

Menemukan keteraturan, pola, atau keterkaitan yang tersembunyi dan bernilai secara informasi dalam kumpulan data yang sangat besar merupakan bagian penting dari proses KDD, yang memerlukan analisis data historis. Ketika mengumpulkan informasi, KDD terutama berkonsultasi dengan basis data historis untuk menemukan pola tersembunyi dan hubungan sebab akibat dalam basis data dengan jumlah data yang sangat besar (Ikhsan et al., 2024). Mengekstrak kumpulan data yang tidak diketahui dan mungkin informasi implisit adalah tujuan dari penemuan pengetahuan dalam basis data (KDD) (Simanjuntak & Elisa, 2019). Penggabungan hasil data mining dengan proses penemuan pengetahuan memungkinkan transformasi data menjadi kesimpulan yang lebih tepat dan menyeluruh.

Dalam melakukan penerapan KDD perlunya pengetahuan dan penyusunan data yang akan diterapkan, pengetahuan yang dimaksud berupa tujuan dari analisa data yang ingin digunakan pada penelitian analisa dapat berupa pola-pola yang diberikan pada data transaksi penelitian sebagai peneliti yang melakukan penelitian berikut langkah-langkah yang dipertimbangkan untuk menggunakan KDD:



Gambar 2.1 Proses KDD

1. Seleksi, merujuk pada pemilihan data dari *dataset* operasional sebelum proses penggalian KDD difokuskan. Data yang dipilih untuk diteliti dalam *data mining* dialokasikan terpisah dari *database* utama seperti penjualan yang diambil berupa barang-barang tertentu pada periode penjualan Januari 2024 – Desember 2024.
2. Tahap pra-pemrosesan langkah awal dalam proses *data mining* yang terutama berfokus pada pembersihan data tahap ini mencakup berbagai bentuk kegiatan termasuk penghapusan data duplikat, verifikasi konsistensi data, koreksi ketidakakuratan data, dan prosedur pembersihan terkait lainnya.
3. Transformasi mengacu pada proses modifikasi kumpulan data yang terpilih untuk memenuhi persyaratan khusus dari proses *data mining*. Proses ini bergantung pada jenis informasi yang akan diekstraksi dan memerlukan penerapan algoritma tertentu.
4. *Data mining* proses sistematis untuk menemukan pola dan pengetahuan substantif diekstraksi dari data yang terkurasi melalui penerapan teknik komputasi dan metodologi tertentu dan langkah algoritma *data mining*

sistematis, sehingga diperlukan pemahaman komprehensif untuk penerapannya yang efektif.

5. Evaluasi dan interpretasi: Langkah terakhir dari KDD adalah menganalisis pola penggalian data. Hasil analisis ini harus disajikan dengan cara yang mudah dipahami oleh pembaca. Hasil analisis ini juga harus digunakan untuk memeriksa apakah pola-pola tersebut sesuai dengan fakta atau hipotesis yang ada.

2.1.2 Data mining

Pendekatan interaktif dan berulang yang berusaha menemukan model atau pola baru yang dapat membantu pengambilan keputusan di masa depan. Hal ini melibatkan pencarian tren atau pola-pola khusus yang diidentifikasi dalam basis data besar, dan mungkin mengintegrasikan *tools* pendukung keputusan tambahan. Analisis data yang dihasilkan dapat memberikan wawasan yang berguna untuk dipelajari dan dipahami secara lebih mendalam (Dogan & Birant, 2021).

Meskipun data yang dikumpulkan dari berbagai sumber seringkali terlalu besar untuk dianalisis secara manual, penggunaan teknologi komputer untuk menemukan informasi dan pengetahuan dalam jumlah data yang terus meningkat telah menjadi layak, murah, dan penting. Hasilnya, topik penelitian baru yang dikenal sebagai penemuan pengetahuan basis data, atau *data mining*, telah muncul. *Data mining* proses mengekstraksi informasi yang tidak sederhana, sebelumnya tidak diketahui, dan mungkin berguna dari sejumlah besar data. Ini menunjukkan bahwa penelitian ini melibatkan berbagai disiplin ilmu, seperti teori informasi, pembelajaran mesin, dan sistem basis data. Beberapa operasi penambangan data

yang paling umum adalah menambang aturan asosiasi, klasifikasi, pengelompokan, regresi, dan deteksi *outlier* (Khajuria et al., 2023).

Menurut (Vidiya & Testiana, 2023) penggunaan *data mining* bisa perluas wawasan bisnis dari pola data transaksi pembelian sehingga dalam konteks *cafe* dengan menerapkan teknik *data mining* pemilik bisnis bisa menjalankan operasional bisnis dengan optimal, meningkatkan kebahagiaan di antara para konsumen sambil belajar bagaimana membuat keputusan yang lebih tepat dengan menggunakan data yang ditambang dari aplikasi ini.

Penerapan *data mining* dalam konteks komersial bagi sebuah perusahaan *supplier marine* memiliki beberapa manfaat yang signifikan seperti, dengan *data mining* perusahaan dapat menganalisis perilaku pembelian konsumen perusahaan secara lebih mendalam, istilah *data mining* terdapat beberapa pandangan seperti penemuan informasi atau pengenalan pola. Ini memungkinkan perusahaan untuk mengidentifikasi preferensi, kebutuhan, dan tren pembelian, sehingga dapat menyesuaikan penawaran produk dan layanan perusahaan secara lebih akurat (Adri et al., 2021).

Menurut (Komariyah et al., 2023) terdapat banyak persamaan dari *data mining* seperti *knowledge discovery*, *knowledge extraction*, *data archaeology*, dan analisa data atau analisa pola, Salah satu metode, algoritma *fp-growth*, menggunakan *data mining* untuk menganalisis sebuah penelitian, namun istilah ini paling sering digunakan dalam kaitannya dengan operasi *data mining*.

Dalam penelitian (Saputra et al., 2023), salah satu bagian metode *data mining* yang menerapkan teknik algoritma *FP-Growth* dalam menemukan produk-produk

yang berasosiasi dengan *Market Basket Analysis*. Dalam hal pengambilan keputusan yang lebih cerdas, *data mining* adalah penyelamat. Bisnis dapat meningkatkan ketepatan dan berbasis data dari pilihan strategis mereka dengan menggunakan informasi yang lebih tepat dan relevan. Meningkatkan efisiensi operasional, memaksimalkan potensi keuntungan, dan meningkatkan daya saing perusahaan adalah tujuan yang dapat dicapai oleh bisnis pasokan maritim melalui penerapan strategis penggalian data. Mengingat (Kumar & Mohbey, 2022) untuk melakukan analisa *data mining* perlu dipertimbangkan sebuah proses yang sistematis agar hasil yang diberikan akurat sebagai berikut:

1. Tahap awal sebelum menggunakan perangkat lunak *data mining*. Analis harus memiliki pemahaman yang jelas tentang masalah yang dihadapi untuk dapat menciptakan solusi yang efektif. Penjelasan yang akurat mengenai masalah tersebut sangat penting dalam proses pengambilan keputusan.
2. Langkah penting dalam fase ini adalah mengidentifikasi dan menghilangkan data yang tidak akurat, melakukan transformasi data, serta memilih *subset dataset* yang relevan. Proses ini melibatkan langkah-langkah pertama dalam seleksi data, deskripsi data, dan visualisasi untuk memastikan kualitas data yang digunakan.
3. Pembangunan dan validasi model merupakan tahap di mana kinerja prediktif terbaik ditentukan dan jenis pemodelan yang tepat dipilih. Proses ini melibatkan pengujian dan evaluasi model untuk memastikan keakuratannya sebelum digunakan dalam pengambilan keputusan.

4. Implementasi adalah tahap terakhir dalam proses *data mining* di mana pencarian pola berurutan (*sequence mining*) dengan mencari sejumlah peristiwa yang umumnya terjadi secara bersamaan, jika diberikan sejumlah objek dan setiap objek dikaitkan pada waktu kejadian, maka prediksi bisa dibuat untuk mendapatkan pola yang memprediksi ketergantungan berurutan yang kuat dan model yang telah dibangun diimplementasikan ke dalam aplikasi yang sesuai. Selain itu, nilai-nilai seperti *support* (nilai penunjang), *confident* (nilai kepastian), dan *frequent item-set* ditentukan berdasarkan *association rule* untuk membantu dalam analisis dan pengambilan keputusan lebih lanjut.

2.1.3 Metode *fp-growth*

Algoritma *fp-growth* didasarkan pada struktur data *fp-tree*, yang terstruktur seperti pohon. *Fp-tree* dapat diekstraksi *frequent itemsets*-nya secara langsung dengan teknik yang digunakan dalam pekerjaan ini. Mengikuti pengembangan struktur pohon dari sekumpulan data transaksi, algoritma *fp-growth* akan digunakan untuk mengidentifikasi *set item* yang sering muncul. Proses ekstraksi melibatkan pembangunan struktur pohon data (*fp-tree*). (Nasyuha et al., 2021).

Pola-pola yang dipotong pada alat *Frequent Pattern Mining* yang sering digunakan dianggap kurang menarik, *data mining* telah menarik banyak perhatian komunitas penelitian basis data karena kegunaannya dalam banyak bidang. Untuk berbagai tujuan *data mining* seperti korelasi, asosiasi, periodisitas parsial, *set item* yang muncul, *set item* yang berurutan dan *set item* maksimum untuk *market basket analysis* buat memecahkan permasalahan ketentuan asosiasi *data mining* tujuan

utama dari temuan ketentuan asosiasi juga untuk memperhitungkan perilaku pelanggan. Penemuan pola umum yang efisien juga sekelompok *item* yang sering timbul dalam suatu gabungan data disebut tugas penting dalam *data mining* terutama dalam kumpulan data transaksi tujuan dalam penulisan ini sebagai pandangan efisiensi algoritma untuk menemukan pola yang sering dalam waktu komputasi serta permasalahan mengenai metode penerapan algoritma pada gabungan data yang berbeda.

Menurut (Dwiputra et al., 2023) dalam penerapan algoritma apriori dan *fp-growth* dalam *association rules data mining* hasil yang ditunjukkan bahwa algoritma *fp-growth* merupakan metode tercepat dalam pengujian dengan waktu 84,655 detik, meskipun begitu dalam algoritma apriori memiliki keunggulan tersendiri yaitu dalam pengujian metode yang dilakukan hanya memakai memori paling kecil pada 482,32 MiB, dengan kesimpulan apriori direkomendasikan untuk pengujian dengan memprioritaskan penggunaan memori dan kemudahan, sedangkan jika mempertimbangkan kecepatan eksekusi dan penggunaan data yang massif algoritma *fp-growth* merupakan pilihan yang bijak.

Dengan terdapatnya globalisasi perdagangan, usaha dagang mengalami kenaikan jumlah konsumen serta transaksi. alhasil, mereka wajib mengetahui resiko serta kemungkinan. Penambangan pola biasa bisa menolong dalam penemuan iklan, potongan harga, sistem letak ritel, penjualan yang luar biasa, manajemen penyimpanan, serta peramalan pasar semacam yang disajikan. Mengidentifikasi *association rule* pada kombinasi produk adalah teknik *data mining* yang penting. Proses ini melibatkan pengolahan data dan perhitungan tertentu untuk menentukan

aturan asosiasi yang relevan. Dua parameter variabel yang perlu dipertimbangkan serbagai penentu aturan adalah *Support* dan *Confidence*. *Support* mendeteksi probabilitas bahwa sejumlah *item* dibeli secara bersamaan, sedangkan *Confidence* mengukur seberapa sering produk C dibeli ketika *item* D telah dibeli. Kedua parameter ini digunakan sebagai batas ambang untuk menentukan keberhasilan aturan. Kedua parameter tersebut adalah:

1. *Support* sebagai ukuran nilai harapan pada pelanggan akan membeli sejumlah produk secara bersamaan berdasarkan jumlah transaksi yang ada. Nilai *support* menentukan apakah suatu *item* atau kelompok *item* layak untuk diteliti, dengan mempertimbangkan nilai *confidence*-nya. Jika *support* tinggi, itu menandakan bahwa *item* A dan B sering dibeli secara bersamaan.
2. *Confidence* suatu tingkat keyakinan terhadap sebuah produk akan dibeli jika produk lain telah dibeli, berdasarkan frekuensi kemungkinan kejadian tersebut. Sebagai contoh, *confidence* mengukur seberapa sering *item* A dibeli ketika *item* B telah dibeli oleh konsumen.

2.1.4 Penerapan rumus *fp-growth*

Pemanfaatan struktur *fp-tree* dapat mengurangi ketergantungan pada total transaksi yang massif hingga mengungkap informasi tentang *itemset* yang sering muncul maka penerapan *fp-growth* dalam *association rule mining* sangatlah efisien. Metode ini lebih cepat dibandingkan algoritma apriori tradisional, terutama saat mengelola *dataset* besar yang memiliki banyak pola yang sering terjadi.

$$\text{Confident } P(A \cap B) = \frac{\sum(A \cap B)}{\sum A} \times 100\%$$

Confident Transaksi A terhadap B

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

Support

$$\text{Support (A,B)} = \frac{\text{Jumlah Transaksi Yang Mengandung A dan B}}{\text{Total Transaksi}} \times 100\%$$

Support 2 Item

Rumus – rumus diatas akan digunakan untuk melakukan perhitungan minimal nilai dukungan (*support*) dan *confidence* yang akan diterapkan pada penelitian dengan menerapkan algoritma *FP-Growth*.

2.1.5 *Software* pendukung

Menurut (Lestari & Andriansyah, 2021) Dengan *RapidMiner*, mendapatkan tingkat akurasi yang tinggi berdasarkan algoritme yang dihitung sangatlah mudah. Tidak hanya itu, alat ini juga membantu hampir semua masalah dengan menghapus fitur-fitur yang tidak perlu dari *data input*, membuatnya lebih sederhana dan tidak mudah *overfitting*. Hal ini membuatnya lebih kuat daripada alat lain, terutama ketika menyangkut transformasi data kecil. Ditambah lagi, contohnya jauh lebih mudah dipahami karena sangat sederhana. Dengan fitur-fitur yang kuat, *RapidMiner* dapat digunakan untuk mengeksplorasi ilmu pengetahuan pada basis data, pembuatan model prediktif, dan analisis statistik diikuti dengan penerapan *machine learning* yang dapat mengakibatkan transformasi pada struktural data yang diproses. *RapidMiner* juga mampu menyajikan hasil analisis secara intuitif melalui fitur visualisasi data dan pembuatan laporan, *tools* ini memungkinkan penyajian

hasil secara efektif pemangku kepentingan, sehingga meningkatkan kompetensi pengambilan keputusan berbasis data yang lebih baik dan memudahkan penyebaran informasi temuan analisis kepada seluruh pihak terkait dalam pengambilan keputusan.

Penggunaan *RapidMiner* bisa diterapkan untuk aplikasi bisnis, komersial dan penelitian

2.2 Penelitian terdahulu

Tabel 2.1 Penelitian Terdahulu

No	Penulis (Tahun)	Judul Penelitian	Metode Penelitian	Hasil Penelitian
1	(Tahi et al., 2024) Scopus dan Sinta 1	<i>Using pattern mining determine fine climatic parameters for maize yield in Benin</i>	<i>Association Rules</i>	<i>The decision tree approach was the most used, although the information was later processed using the frequent pattern growth methodology to help find the rules. Medium minimum and maximum temperatures, rainfall, evapotranspiration, and humidity levels are associated with good maize yields in the Sudano-Guinean zone. Producing abundant maize in the Sudanian zone is possible with moderately high humidity and somewhat low minimum and maximum temperatures. The identified</i>

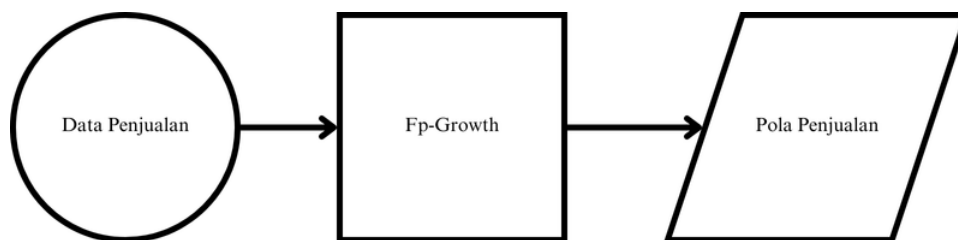
				<i>association rules demonstrated a reliable and effective method for maximizing maize yield.</i>
2	(Khajuria et al., 2023) Scopus dan Sinta 1	<i>Performance analysis of frequent pattern mining algorithm on different real-life dataset</i>	<i>Frequent-Pattern Growth</i>	<i>Four real-world datasets—the retail, accidents, chess, and mushrooms datasets—have been used to construct and evaluate these algorithms. The findings show that, when applied to the Retail dataset, the estDec+ method outperforms all others in terms of run time and memory use. When compared to other algorithms, Pre-Post performs better. Also, about the execution time and memory usage of Accident datasets.</i>
3	(Setiawan & Mulyanti, 2020) Sinta 2	<i>Market Basket Analysis dengan Algoritma Apriori pada Ecommerce Toko Busana Muslim Trendy</i>	Apriori	Perangkat lunak yang dikembangkan menggunakan metode market basket analysis (MBA) dengan algoritma apriori dapat membantu konsumen dalam melakukan rekomendasi produk berdasarkan pengujian aplikasi yang dilakukan melalui wawancara pengguna.

4	(Syahputri, 2020) Sinta 4	Penerapan <i>Data Mining</i> Asosiasi pada Pola Transaksi dengan Metode Apriori	Apriori	Secara keseluruhan, data sampel penjualan menghasilkan 52 aturan asosiasi dengan tingkat kepercayaan 50% dan tingkat dukungan 20%. Selain itu, hasil dari RapidMiner menunjukkan bahwa item dengan tingkat kepercayaan minimum di atas 50% adalah yang paling banyak diminati oleh pelanggan.
5	(Rajagukguk, 2020) Sinta 4	Implementasi <i>Association Rule Mining</i> Untuk Menentukan Pola Kombinasi Makanan Dengan Algoritma Apriori	Apriori	Proses yang telah ditentukan menghasilkan 25 aturan asosiasi, dengan minimum support 20% dan minimum confidence 70%, dengan menggabungkan ketiga itemset dan menghitung confidence. Aturan-aturan ini mengidentifikasi semua aturan asosiasi yang mungkin di antara item.
6	(Ikhsan et al., 2023) Nasional	Analisis Untuk Memprediksi Stok Aset di PT XYZ Menggunakan Algoritma Apriori	Apriori	Untuk mendapatkan nilai dukungan maksimum 7,9% dari kombinasi item Printer > Laptop dan Printer > PC, algoritma apriori dapat digunakan

				untuk meramalkan inventaris yang diperlukan.
7	(Sudarto & Handoko, 2023) Lokal	IMPLEMENTASI <i>DATA MINING</i> PADA PENGATURAN <i>DATA INVOICE</i> DISTRIBUTOR MENGUNAKAN ALGORITMA <i>FP GROWTH</i>	<i>Frequent- Pattern Growth</i>	Pola kombinasi terbaik yang didapatkan berupa Pembelian Layer Cake Pandan pasti akan membeli <i>Layer Cake Chocolate</i> dengan <i>Confidence</i> 100% dan nilai <i>support</i> tertinggi berupa setiap pembelian <i>Rota Prawn</i> 14 pasti akan membeli <i>Super Ring</i> 14 dengan <i>support</i> 32%.

2.3 Kerangka pemikiran

Kerangka pemikiran yang akan diterapkan pada penelitian yang menganalisa menggunakan algoritma *FP-Growth* sebagai berikut:



Gambar 2.2 Kerangka Pemikiran

Sumber: Data Peneliti, 2024

Data penjualan produk merupakan input utama untuk tahap pertama penelitian ini. Jumlah transaksi dan detail produk seperti nomor faktur dan deskripsi adalah bagian dari kumpulan data ini. Sebelum diproses lebih lanjut, data ini disaring untuk memastikan keakuratannya. Selanjutnya, data diproses menggunakan algoritma tertentu dengan mencapai nilai dukungan minimum.

Setelah proses ini selesai, data diuji memanfaatkan *tools* bantuan *RapidMiner* untuk menyajikan temuan yang akurat. Temuan pada proses ini adalah identifikasi produk yang memiliki frekuensi tinggi dibeli oleh pelanggan, memberikan wawasan yang berharga bagi perusahaan yang hendak memahami pola atau kecenderungan penjualan.

Mengumpulkan data penjualan barang sebagai *input* adalah tahap awal untuk mencapai tujuan ini. Detail produk termasuk merek, jenis, dan *volume* penjualan adalah bagian dari kumpulan data ini. Sebelum pemrosesan selanjutnya, data disaring untuk menjamin kebenarannya. Langkah selanjutnya adalah menentukan nilai dukungan minimal dengan menganalisis data dengan algoritma yang sesuai. Data kemudian dievaluasi menggunakan perangkat lunak *RapidMiner* untuk memastikan temuan yang dapat diandalkan ketika prosedur ini selesai. Toko dapat memanfaatkan hasil penelitian ini untuk meningkatkan inventaris dan strategi penjualan mereka dengan menentukan barang mana yang paling banyak dibeli pelanggan.