

BAB II

TINJAUAN PUSTAKA

2.1. Data Mining

Menurut (Marisa, Maukar, & Akhriza, 2021) *data mining* adalah suatu proses penambangan data yang menghasilkan sebuah *output* (keluaran) yang berupa pengetahuan dengan tujuan untuk menemukan pola atau hubungan dalam data jumlah besar yang tidak disadari keberadaannya.

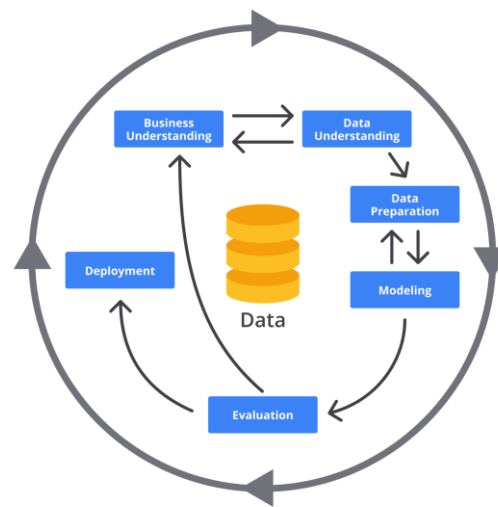


Gambar 2. 1 Siklus *data mining*
(sumber: E-book *Data mining konsep dan penerapannya*, 2021)

Di sebuah siklus *data mining* terdiri dari 3 fase utama yaitu *input*, proses, dan *output*. Pada hal ini, *input* berkaitan dengan adanya data yang akan dikelola, kemudian pada bagian proses terdapat beberapa metode yang akan digunakan seperti algoritma apriori, *FP- growth*, dan K-Means. Menurut (Ajeng & Andreas, 2024) bagian siklus *output* adalah bagaimana hasil data yang sudah di olah melalui proses akan membantu mengidentifikasi pola pembelian dan memberikan wawasan berharga untuk pengambilan keputusan strategis. Proses umum dalam *data mining* sering kali mengikuti model CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang terdiri dari enam fase:

1. *Business Understanding*: Memahami tujuan bisnis dan kebutuhan.
2. *Data Understanding*: Mengumpulkan dan mengeksplorasi data.

3. *Data Preparation*: Menyiapkan data untuk analisis.
4. *Modelling*: Menerapkan teknik pemodelan untuk data.
5. *Evaluation*: Mengevaluasi model untuk memastikan bahwa ia memenuhi tujuan bisnis.
6. *Deployment*: Mengimplementasikan model dalam lingkungan nyata.



Gambar 2. 2 Siklus CRISP-DM
(Sumber: Dicoding, 2024)

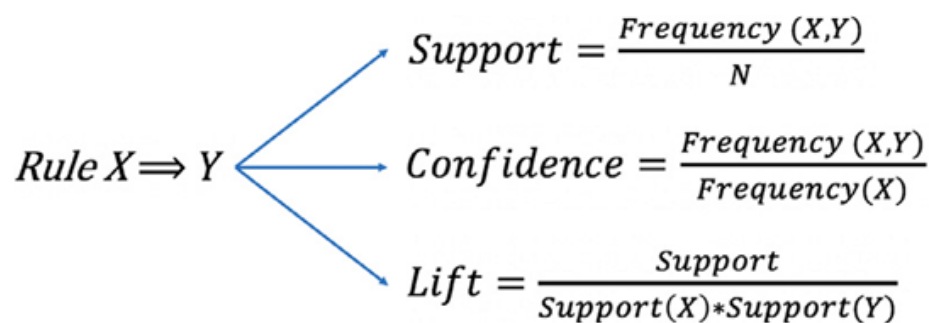
Menurut Penelitian (Galit & Otto R, 2021) *data mining* sangat penting dalam pengambilan keputusan berbasis data di berbagai bidang, termasuk bisnis, kesehatan, dan ilmu sosial. Dalam konteks penelitian ini, *data mining* akan digunakan untuk menganalisis pola transaksi musiman yang terjadi dalam data penjualan. (Galit & Otto R, 2021)

2.2. Association Rule

Association rule adalah teknik *data mining* yang digunakan untuk menemukan hubungan atau pola asosiasi antara item dalam kumpulan data yang besar (Artsitella, Apriliani, & Ashari, 2021). Dalam bisnis ritel, AR digunakan

untuk mengidentifikasi produk yang sering dibeli bersamaan. Konsep utama dalam teori ini meliputi *itemset*, *support*, *confidence*, dan *lift*, yang digunakan untuk mengidentifikasi pola pembelian yang signifikan. (Ajeng & Andreas, 2024) menekankan bahwa teori ini sangat berguna dalam analisis pola pembelian untuk mengidentifikasi produk yang sering dibeli bersamaan. Dalam konteks penelitian ini, teori asosiasi akan digunakan untuk menemukan pola pembelian yang terkait dengan transaksi musiman, sehingga dapat memberikan wawasan tentang produk yang memiliki hubungan kuat dalam periode tertentu. Penelitian oleh (Iskandar, Anggiane, Zulvia, & Nissya, 2024) juga menunjukkan bahwa algoritma Apriori, yang merupakan bagian dari teori asosiasi, efektif dalam meningkatkan penjualan melalui analisis pola pembelian.

Association rule biasanya digunakan dalam menentukan hubungan sebab akibat, dengan bentuk LHS ke RHS dengan interpretasi bahwa jika item dalam LHS (*Left Hand Side*) dibeli, maka item RHS (*Right Hand Side*) juga dibeli hal ini berdasarkan prinsip "jika-maka" (*if-then*), yang menunjukkan pola kemunculan item dalam transaksi. Aturan asosiasi biasanya ditulis dalam bentuk $X \rightarrow Y$, yang berarti jika seseorang membeli item X, mereka juga cenderung membeli item Y.



$$\begin{aligned}
 \text{Rule } X \Rightarrow Y & \begin{cases} \text{Support} = \frac{\text{Frequency}(X,Y)}{N} \\ \text{Confidence} = \frac{\text{Frequency}(X,Y)}{\text{Frequency}(X)} \\ \text{Lift} = \frac{\text{Support}}{\text{Support}(X) * \text{Support}(Y)} \end{cases}
 \end{aligned}$$

Gambar 2. 3 Rumus Association Rule
(Sumber: Sandaruwan Herath, 2024)

Berikut adalah langkah utama yang ada pada AR adalah:

1. Menemukan *Frequent Itemset*: Mengidentifikasi *itemset* yang sering muncul dalam transaksi.
2. Menghasilkan Aturan Asosiasi: Membuat aturan asosiasi dari *frequent itemset* yang memenuhi ambang batas kepercayaan (*confidence*) minimum. Sedangkan untuk parameter dalam *association rule* ada tiga parameter utama dalam mengevaluasi kekuatan aturan asosiasi yaitu:
 1. *Support*: Menunjukkan seberapa sering *itemset* muncul dalam data transaksi.
 2. *Confidence*: Mengukur tingkat kepastian bahwa aturan tersebut benar, yaitu rasio antara transaksi yang mengandung X dan Y dibandingkan dengan transaksi yang hanya mengandung X.
 3. *Lift*: Mengukur seberapa kuat hubungan antara item X dan Y dengan memperhitungkan kemungkinan kebetulan.

2.3. Algoritma Apriori

Menurut (Zulfa, Rayuwatib, & Koko, 2020) Metode apriori merupakan salah satu metode klasik dari *data mining* yang digunakan agar program dapat mempelajari aturan dari asosiasi untuk mendapatkan pola hubungan antara beberapa item dalam satu dataset. Metode ini juga banyak digunakan untuk data transaksi atau bahasa lainnya *market basket* karena dengan adanya metode ini pemilik toko dapat mengetahui pola dari data transaksi dan pembelian seorang konsumen atau pelanggan. Penerapan teknik *data mining* ini diharapkan dapat membantu optimalisasi stok dan strategi promosi serta memenuhi kebutuhan

pelanggan. Dengan hal tersebut penjual dapat menentukan strategi penjualan untuk menghasilkan keuntungan yang lebih agar meningkat. Menurut (Saputro, Jumasa, & Murhadi, 2020) menjelaskan bahwa langkah-langkah dalam algoritma Apriori meliputi menentukan *threshold minimum support*, menghitung *support* untuk semua *itemset* yang mungkin, dan menghasilkan *itemset* yang valid. Dalam penelitian ini, algoritma Apriori akan diterapkan untuk menganalisis data transaksi musiman dan menemukan pola pembelian yang signifikan. Penelitian oleh (Junfeng, Guo, Fang, Wang, & Fan, 2022) juga menunjukkan bagaimana pola musiman dapat dianalisis menggunakan algoritma Apriori, memberikan bukti tambahan tentang efektivitas metode ini dalam konteks analisis transaksi musiman.

2.4. Time Window

Time window adalah konsep yang digunakan untuk membagi data transaksi berdasarkan waktu, musim, atau periode tertentu. Dalam analisis musiman, *time window* membantu dalam memahami bagaimana pola transaksi berubah seiring waktu. (Zhenyu, Zhu, Gao, & Xu., 2021) Time menyatakan bahwa dengan menggunakan *time window*, peneliti dapat menganalisis data dalam rentang waktu tertentu untuk mengidentifikasi tren dan pola yang mungkin tidak terlihat dalam analisis data secara keseluruhan. Pada penelitian (Akhriza & Mumpuni, 2021) menunjukkan bahwa dengan membagi data ke dalam *time window*, analisis dapat dilakukan untuk setiap periode, sehingga memberikan wawasan yang lebih mendalam tentang perubahan perilaku konsumen.

Dalam penelitian ini, pendekatan *time window* diterapkan dengan menggunakan lebar jendela dua bulan dan pergeseran satu bulan. Artinya, analisis dilakukan dengan durasi dua bulan dan pergeseran satu bulan. Contohnya, analisis pertama dilakukan pada bulan Februari, lalu Februari–Maret, kemudian Maret, dan seterusnya. Strategi ini bertujuan untuk mengamati pola pembelian yang bersifat musiman dan mendeteksi perubahan pola antar bulan secara lebih detail.

2.5. Transaksi Musiman

Transaksi musiman adalah pola pembelian yang terjadi secara berulang dalam periode tertentu, seperti bulanan atau tahunan. Pola ini sering dipengaruhi oleh faktor eksternal seperti musim, hari libur, atau acara khusus yang dapat mempengaruhi perilaku konsumen. Pada penelitian (Junfeng, Guo, Fang, Wang, & Fan, 2022) menjelaskan bahwa memahami transaksi musiman sangat penting bagi perusahaan untuk merencanakan persediaan dan strategi pemasaran yang efektif. Penelitian ini akan mengeksplorasi bagaimana pola transaksi musiman dapat diidentifikasi dan dianalisis menggunakan teknik *data mining*.

2.6. KNIME

KNIME merupakan *platform open-source* yang digunakan untuk analisis dan pemrosesan data secara visual. Platform ini memungkinkan pengguna membangun alur kerja (*workflow*) dengan pendekatan berbasis *node*, di mana setiap *node* mewakili fungsi spesifik dalam proses analitik, seperti impor data, pembersihan, analisis, hingga visualisasi. (Benjamín, Valderrama-Chauca, Cari-Mogrovejo, Apaza-Huanca, & Sanchez-Ilabaca, 2022) menjelaskan bahwa KNIME mendukung berbagai teknik analisis data, mulai dari statistik dasar hingga algoritma pembelajaran mesin, termasuk *association rule mining* menggunakan

algoritma Apriori. Platform ini kompatibel dengan berbagai sumber data, seperti file CSV, basis data, dan layanan web.

Menurut (Frank, 2023) Proses kerja KNIME umumnya terdiri atas lima tahap, yaitu:

1. *import* data menggunakan *reader nodes*
2. pra-pemrosesan data seperti filter dan penanganan nilai hilang
3. penerapan algoritma analitik
4. visualisasi hasil analisis
5. ekspor *output* ke berbagai format

KNIME telah digunakan secara luas dalam penelitian dan praktik bisnis. Misalnya, penelitian (Frank, 2023) menunjukkan keberhasilan KNIME dalam analisis prediktif dan efisiensi operasional di sektor industri dan kesehatan. Keunggulan utama KNIME terletak pada antarmuka grafis yang intuitif, integrasi dengan berbagai bahasa pemrograman seperti R dan *Python*, serta dukungan komunitas pengguna yang aktif.