

BAB II

LANDASAN TEORI

2.1 Data Mining

Data mining adalah proses mengekstraksi dan mengidentifikasi informasi terkait suatu data besar menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning (Yani et al., 2023). Data mining merupakan sebuah proses untuk mengetahui informasi atau pola dari gabungan data yang tersembunyi dengan tujuan untuk menemukan, menggali pengetahuan dari data atau informasi yang dimiliki Perusahaan (Napitupulu et al., 2024). Data mining ialah proses analisis data untuk menemukan pola tertentu dalam jumlah data besar dan bertujuan untuk menghasilkan informasi yang dapat digunakan serta dikembangkan lebih lanjut sehingga dapat membantu dalam pengambilan Keputusan (Nasyuha et al., 2021). Data mining digunakan untuk menganalisis data dalam jumlah besar agar dapat diketahui hubungan antar data serta dapat disajikan dalam bentuk yang mudah dipahami, sehingga diperoleh sebuah pola dari hasil pemrosesan data tersebut (Pangestu & Ridwan, 2022). Dari beberapa pengertian diatas dapat disimpulkan bahwa data mining merupakan proses analisis data dalam jumlah besar untuk menemukan pola atau informasi tersembunyi, hal tersebut dapat membantu pengambilan keputusan, memahami hubungan antar data, dan memberikan informasi yang mudah dipahami. Proses analisis dapat dilakukan dengan teknik seperti statistik, matematika, kecerdasan buatan, dan *machine learning*.

Data mining, sering juga disebut *knowledge discovery in database* (KDD), yaitu proses mengumpulkan dan menggunakan data historis untuk menemukan hubungan, pola, atau keteraturan dalam kumpulan data dengan jumlah besar (Sibarani, 2020). Data Mining bertujuan untuk mengungkapkan informasi yang tidak terlihat secara langsung dan mungkin tidak terduga sebelumnya, sehingga dapat digunakan untuk berbagai tujuan, termasuk prediksi, analisis, dan pengetahuan baru. Data mining dibagi menjadi dua sifat yaitu deskriptif dan prediktif. Deskriptif dilakukan untuk mendeskripsikan data sehingga dapat menemukan pola yang paling dipahami, sedangkan prediktif untuk melakukan

prediksi dengan model pengetahuan yang ada. Data mining memiliki beberapa karakteristik umum, sebagai berikut (Sibarani, 2020):

1. Data mining berfokus pada penemuan pola atau informasi yang tersembunyi dalam data yang sebelumnya tidak diketahui.
2. Data mining sering menggunakan volume data yang besar untuk analisis.
3. Penggunaan data besar biasanya meningkatkan tingkat kepercayaan terhadap hasil yang diperoleh.
4. Data mining sangat berguna untuk pengambilan keputusan, khususnya dalam perencanaan strategi.

Berdasarkan tugasnya data mining dibagi menjadi beberapa kelompok, antara lain (Suwandi et al., 2021):

1. Deskripsi : proses untuk menggambarkan pola atau karakteristik yang ada dalam data tanpa mengubah atau membuat prediksi tentang data tersebut. Contohnya, menyajikan informasi mengenai jumlah penjualan produk per bulan atau distribusi usia pelanggan.
2. Estimasi : berhubungan dengan perkiraan suatu nilai numerik berdasarkan data yang sudah ada, seringkali dilakukan untuk memperkirakan parameter atau sifat tertentu dari populasi berdasarkan sampel. Contohnya, estimasi harga rumah berdasarkan berbagai faktor seperti lokasi, luas, dan fasilitas yang ada.
3. Prediksi : analisis yang berfokus pada memprediksi nilai atau hasil yang belum terjadi di masa depan berdasarkan pola atau tren yang ada dalam data historis. Contohnya, menggunakan data penjualan historis untuk memprediksi penjualan di bulan berikutnya.
4. Klasifikasi : teknik untuk mengelompokkan data ke dalam kategori atau kelas yang telah ditentukan sebelumnya. Proses ini menggunakan fitur atau atribut dari data untuk memprediksi kategori mana yang paling relevan. Contohnya, memprediksi apakah sebuah email adalah spam atau bukan spam berdasarkan fitur-fitur seperti kata-kata yang terdapat dalam email tersebut.
5. Clustering : proses mengelompokkan data ke dalam grup-grup yang lebih kecil berdasarkan kesamaan atau kedekatannya satu sama lain. Berbeda dengan klasifikasi, pada clustering, tidak ada kategori atau kelas yang sudah ditentukan

sebelumnya. Contohnya, mengelompokkan pelanggan dalam grup berdasarkan perilaku pembelian mereka untuk melakukan segmentasi pasar yang lebih baik.

6. Asosiasi : analisis yang mencari hubungan atau pola yang sering muncul antara item-item dalam satu transaksi atau kejadian dalam periode waktu tertentu. Contohnya, jika seseorang membeli roti, maka kemungkinan besar mereka juga akan membeli mentega. Pola ini dapat ditemukan menggunakan algoritma seperti Apriori.

2.2 Aturan Asosiasi

Aturan Asosiasi (*Association Rules*) merupakan aturan yang dibuat untuk mengetahui hubungan yang terjadi antar item pada suatu dataset (Panjaitan et al., 2020). Aturan Asosiasi sering dikenal sebagai salah satu teknik data mining yang mendasari berbagai teknik data mining lainnya (ichlas et al., 2021). Teknik ini umumnya dimanfaatkan dalam dunia bisnis yang biasa disebut dengan analisis keranjang belanja (Novianti & Elisa, 2020). Tujuan dari analisis asosiasi yaitu untuk mencari pola antar item yang muncul bersamaan di antara banyaknya transaksi yang ada, dimana setiap transaksinya terdiri dari beberapa item (Dewi Sri Mulyani et al., 2023). Selain itu, aturan asosiasi juga dapat membantu pengambilan keputusan bisnis, baik dalam hal mengidentifikasi pola pembelian *customer*, persediaan stok barang, penentuan susunan barang di rak dan sebagai bahan rekomendasi strategi promosi produk (Devita Sari & Khoiriah, 2022). Dalam penentuan Aturan Asosiasi terdapat dua ukuran yang digunakan yaitu (Panjaitan et al., 2020):

1. Minimal Support

Minimal Support adalah batas frekuensi kejadian (count support) dari seluruh item set (support) dalam keseluruhan transaksi. Nilai support item (X) dapat ditemukan dengan menggunakan persamaan 2. 1:

$$Support(X) = \frac{\text{jumlah transaksi yang mengandung } X}{\text{total transaksi}} \dots (2.1)$$

Sedangkan nilai support item (X,Y) dapat ditemukan dengan menggunakan persamaan 2. 2:

$$Support(X,Y) = \frac{\text{jumlah transaksi yang mengandung } X \text{ dan } Y}{\text{total transaksi}} \dots (2.2)$$

2. Minimal Confidence

Frekuensi munculnya item Y dalam transaksi yang menyertakan item X. Untuk menentukan nilai *confidence* dapat menggunakan persamaan 2. 3:

$$Confidence (X|Y) = \frac{jumlah\ transaksi\ yang\ mengandung\ X\ dan\ Y}{total\ transaksi\ yang\ mengandung\ X}(2.3)$$

2.3 Algoritma Apriori

Algoritma Apriori merupakan algoritma dasar yang dirancang untuk mengidentifikasi *frequent itemset* dalam proses pembuatan aturan asosiasi (Qoni'ah & Priandika, 2020). Kemampuannya dalam menemukan pola dengan frekuensi tinggi membuat algoritma ini banyak diterapkan dalam analisis data transaksi, seperti analisis keranjang belanja (*market basket analysis*), untuk mengungkap pola pembelian *customer*. Selain itu, algoritma apriori juga menggunakan dua ukuran utama, yaitu *support* untuk mengukur seberapa sering itemset muncul, dan *confidence* untuk menentukan kekuatan hubungan antar item dalam aturan asosiasi. (Prasetyo et al., 2020). Proses utama dalam algoritma apriori untuk memperoleh *frequent itemset* meliputi (Santoso et al., 2016):

1. Join(penggabungan)

Proses ini dilakukan dengan mengombinasikan item satu dengan item lainnya hingga tidak memungkinkan terbentuknya kombinasi baru.

2. Prune(pemangkasan)

Pemangkasan dilakukan pada hasil kombinasi item, di mana item-item tersebut dipangkas berdasarkan nilai minimum support yang telah ditetapkan sebelumnya.

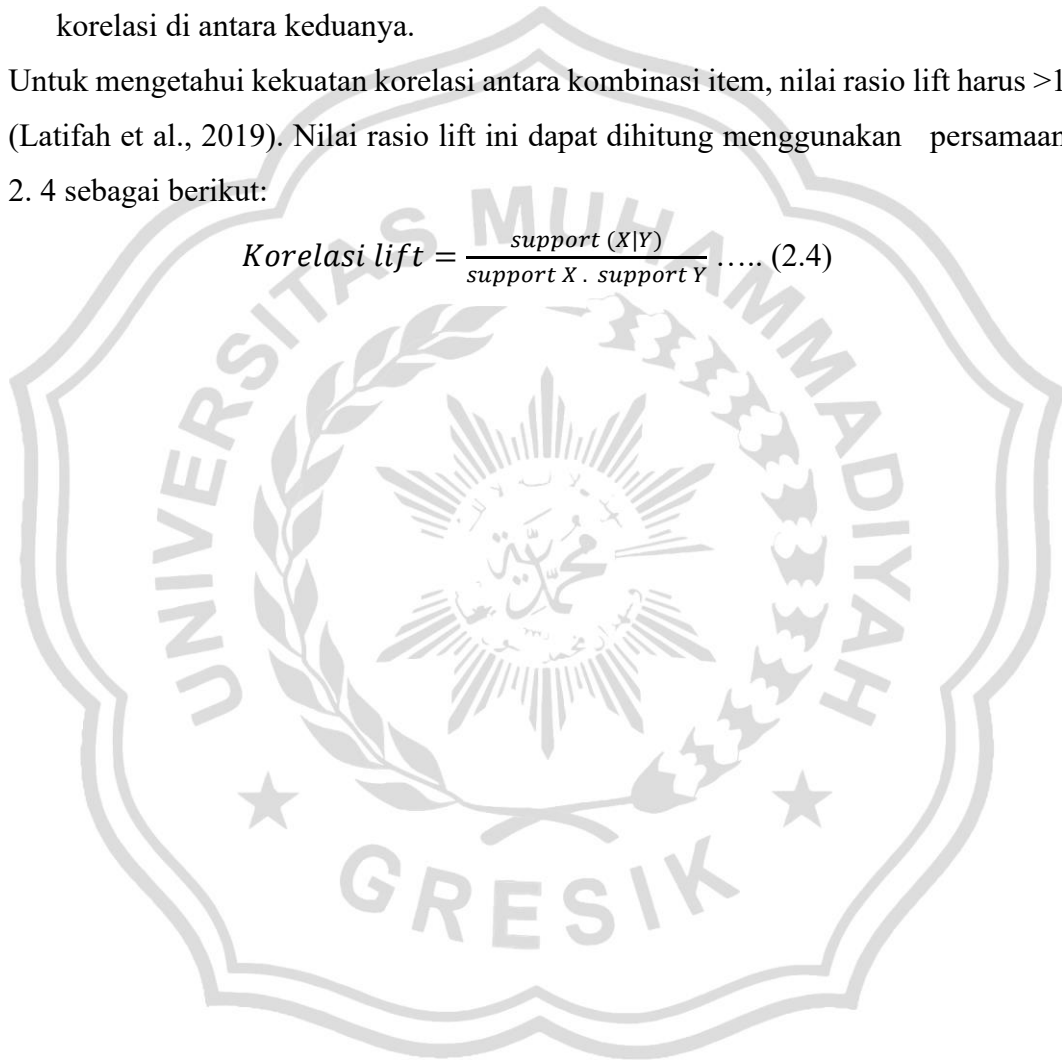
2.4 Korelasi Lift

Korelasi *Lift* atau rasio *lift* adalah suatu ukuran yang digunakan untuk menilai kekuatan hubungan atau kaitan antara item-item yang terbentuk. Nilai rasio Lift digunakan untuk menilai sejauh mana kekuatan suatu *rule*. Terdapat tiga kemungkinan yang dihasilkan pada rasio *lift* (Prabowo & Ramdani, 2020):

1. Jika nilai lift < 1 , maka kemunculan item X memiliki korelasi negatif dengan kemunculan item Y. Artinya, kemunculan satu item cenderung mengarah pada ketidakhadiran item lainnya.
2. Jika nilai lift > 1 , maka item X dan item Y memiliki korelasi positif. Ini berarti bahwa kemunculan satu item cenderung diikuti oleh kemunculan item lainnya.
3. Jika nilai lift $= 1$, maka item X dan item Y bersifat independen, dan tidak ada korelasi di antara keduanya.

Untuk mengetahui kekuatan korelasi antara kombinasi item, nilai rasio lift harus > 1 (Latifah et al., 2019). Nilai rasio lift ini dapat dihitung menggunakan persamaan 2.4 sebagai berikut:

$$\text{Korelasi lift} = \frac{\text{support}(X|Y)}{\text{support } X \cdot \text{support } Y} \dots (2.4)$$



2.5 Penelitian Terkait

Sebagai penguatan topik penelitian, penulis melakukan analisis dari beberapa riset penelitian sebelumnya, sebagai berikut :

No	Landasan Literatur	Metode Penelitian	Permasalahan	Hasil Penelitian
1	Judul Penelitian : Penerapan Algoritma Apriori untuk Mencari Pola Transaksi Penjualan Berbasis Web pada Cafe Sakuyan Side Penulis : T Kurniana, A Lestari, dan E D Oktaviyani Tahun : 2023	Algoritma Apriori	Data transaksi penjualan di Cafe Sakuyan Side hanya digunakan untuk laporan keuangan dan tidak dimanfaatkan untuk meningkatkan strategi penjualan.	Terdapat beberapa aturan asosiasi yang valid sehingga dapat digunakan untuk menyusun strategi penjualan yang lebih efektif, seperti rekomendasi paket menu dan cross-selling produk tambahan.
2	Judul Penelitian : Implementasi Data Mining Untuk Mengetahui Pola Pembelian Obat Menggunakan Algoritma Apriori Penulis : NF Ulfha dan R Amin Tahun : 2020	Algoritma Apriori	Persaingan bisnis di sektor apotek menuntut strategi pemasaran yang lebih efektif.	Hasil dari penelitian ini menunjukkan dengan menggunakan algoritma Apriori, data transaksi penjualan obat dapat diolah untuk menemukan pola pembelian yang bermanfaat. Sehingga dapat digunakan untuk menyusun strategi pemasaran produk.

No	Landasan Literatur	Metode Penelitian	Permasalahan	Hasil Penelitian
3	Judul Penelitian : Algoritma Apriori Dalam Identifikasi Pola Pembelian Konsumen Pada Produk Minuman Penulis : VC Nisa dan FN Khasanah Tahun : 2023	Algoritma Apriori	kurangnya pemahaman mengenai pola pembelian konsumen yang dapat digunakan untuk merancang strategi pemasaran yang lebih efektif.	Penelitian ini berhasil menemukan pola pembelian produk yang sering dibeli secara bersamaan dengan memanfaatkan data transaksi penjualan menggunakan metode algoritma Apriori.
4	Judul Penelitian : Analysis of Consumer Purchase Patterns Using Apriori Algorithm at Ryo Jaya Buah Store Penulis : R. Ricardo, Y. Reswan, dan D. Abdullah Tahun : 2024	Algoritma Apriori	Ketidak tahuan mengenai pola pembelian konsumen membuat Toko Ryo Jaya Buah belum memanfaatkan data transaksi secara maksimal	Penelitian ini menunjukkan bahwa dengan menggunakan Algoritma Apriori, pola pembelian konsumen di Toko Ryo Jaya Buah dapat diketahui serta pola ini sangat berguna untuk merancang strategi pemasaran.
5	Judul Penelitian : Implementasi Data Mining Dengan Metode Algoritma Apriori Dalam	Algoritma Apriori	Toko Ayam Presto Cabe Hijo Cabang Medan mengalami masalah penjualan menu	Hasil penelitian menunjukkan bahwa dengan menggunakan Algoritma Apriori, restoran dapat menyusun

No	Landasan Literatur	Metode Penelitian	Permasalahan	Hasil Penelitian
	Menetapkan Paket Menu Yang Akan Dipromosikan Penulis : LM Napitupulu dan M Zunaidei Tahun : 2024		yang tidak merata, di mana beberapa menu tidak laku habis dan menyebabkan penumpukan bahan baku yang akhirnya rusak.	paket menu promosi yang lebih efektif, mengurangi pemborosan bahan baku, dan meningkatkan penjualan.
6	Judul Penelitian : Association Rule Mining using Apriori Algorithm for Extracting Product Sales Patterns in Groceries Penulis : M Kavitha dan S Subbaiah Tahun : 2020	Algoritma Apriori	<i>Sales</i> ingin menemukan hubungan antara produk yang sering dibeli secara bersamaan oleh konsumen.	Hasil penelitian ini menunjukkan bahwa Algoritma Apriori dapat digunakan untuk mengekstraksi pola pembelian produk di toko bahan makanan, hal ini membantu <i>Sales</i> dalam memahami hubungan antar produk yang sering dibeli bersama, untuk merancang strategi penjualan yang lebih baik.
7	Judul Penelitian : Implementasi Metode Apriori pada Data Mining untuk Pola Pembelian	Algoritma Apriori	Pengelolaan data yang tidak tepat mengakibatkan Toko Matahari Kota Lubuklinggau	Hasil penelitian menunjukkan algoritma Apriori dapat menemukan pola pembelian yang dapat digunakan

No	Landasan Literatur	Metode Penelitian	Permasalahan	Hasil Penelitian
	Barang pada Toko Matahari Kota Lubuklinggau Penulis : CA Suwandi, R Yanto, D Apriadi. Tahun : 2021		sulit untuk mengetahui pola pembelian barang yang sering terjadi	untuk merancang strategi pemasaran yang lebih efektif.
8	Judul Penelitian : Penentuan Aturan Asosiasi Pola Pembelian Pada Minimarket Dengan Algoritma Apriori Penulis : A. Novianti dan E. Elisa Tahun : 2020	Algoritma Apriori	Peningkatan data yang terus bertambah dan penempatan barang yang belum disesuaikan dengan pola pembelian konsumen menjadi hambatan dalam mengoptimalkan strategi penjualan.	Dengan algoritma Apriori, penelitian ini berhasil menemukan bahwa produk seperti Telur dan Mie sering dibeli bersamaan. Pola ini dapat membantu toko dalam mengatur penempatan barang dan merancang strategi pemasaran yang sesuai.
9	Judul Penelitian : Implementasi Algoritma Apriori Pada Penentuan Kombinasi Menu Umkm XYZ Sukoharjo	Algoritma Apriori	Stok menu dan bahan yang tidak optimal akibat beberapa menu menumpuk	Hasil yang didapat dari analisis data penjualan UMKM XYZ menggunakan algoritma Apriori adalah sistem yang dapat menentukan

No	Landasan Literatur	Metode Penelitian	Permasalahan	Hasil Penelitian
	Penulis : Donni Prabowo dan Fitri Ramdani Tahun : 2020		dari persediaan sebelumnya yang tidak terjual	kombinasi menu untuk meningkatkan efisiensi penentuan menu yang harus dipromosikan.
10	Judul Penelitian : Implementasi Algoritma Apriori Untuk Menentukan Pola Pembelian Pada Customer (Studi Kasus: Toko Bakoel Sembako) Penulis : AN Rahmi dan YA Mikola Tahun : 2021	Algoritma Apriori	Tidak adanya pemahaman yang jelas mengenai pola pembelian konsumen dapat mempengaruhi strategi pemasaran dan pengelolaan stok barang di Toko Bakoel Sembako.	Hasil dari penelitian ini adalah penerapan algoritma Apriori menghasilkan beberapa aturan asosiasi yang menunjukkan pola-pola pembelian konsumen di Toko Bakoel Sembako, yang dapat digunakan untuk meningkatkan strategi pemasaran dan pengelolaan stok.