

BAB II

LANDASAN TEORI

2.1. Data Science

Data science adalah disiplin ilmu yang mencakup pengumpulan, analisis, interpretasi, presentasi, dan pengelolaan data [1]. Dengan menggunakan pendekatan multidisiplin, data science menggabungkan konsep statistika, matematika, dan pemrograman komputer untuk mengidentifikasi pola, tren, dan wawasan yang terkandung dalam data. Para profesional data science menggunakan berbagai metode seperti machine learning, analisis prediktif, dan pengolahan bahasa alami untuk mengurai kompleksitas data. Tujuan utama dari data science adalah untuk menghasilkan informasi yang dapat digunakan untuk mendukung pengambilan keputusan dan inovasi di berbagai bidang, mulai dari bisnis dan industri hingga ilmu pengetahuan dan kesehatan. Selain itu, data science juga memiliki peran penting dalam memahami perilaku pengguna, meningkatkan efisiensi operasional, dan mengidentifikasi peluang baru. Dengan terus berkembangnya teknologi dan volume data yang terus meningkat, data science menjadi semakin integral dalam membentuk dunia modern.

Data science dan data mining saling terkait erat dalam upaya untuk menggali wawasan berharga dari dataset yang besar dan kompleks. Data mining adalah bagian dari data science yang fokus pada proses ekstraksi pola dan informasi yang tersembunyi dalam data. Dalam konteks data science, data mining memainkan peran penting sebagai salah satu teknik analisis data yang digunakan

untuk mengidentifikasi hubungan, asosiasi, atau tren yang dapat memberikan wawasan penting.

Data mining menggunakan berbagai metode, seperti clustering, association rule mining, dan classification, untuk mengurai data menjadi struktur yang lebih mudah dimengerti. Ini memungkinkan para profesional data science untuk mengidentifikasi pola kompleks dan menghasilkan pemahaman mendalam tentang perilaku data. Penerapan data mining dalam data science tidak hanya membantu dalam pengambilan keputusan yang lebih baik, tetapi juga memungkinkan peramalan yang lebih akurat, pengelompokan data yang lebih efisien, dan pengembangan model prediktif yang lebih kuat.

Dengan integrasi data mining dalam kerangka data science, organisasi dapat lebih efektif mengoptimalkan nilai dari dataset mereka. Ini membuka peluang untuk inovasi, penemuan pengetahuan baru, dan pemahaman yang lebih baik tentang dinamika bisnis atau fenomena di berbagai bidang. Dengan kata lain, data mining menjadi instrumen penting dalam kotak alat data science yang memberikan kemampuan untuk mengurai, mengenali, dan memahami makna di balik setiap baris dan kolom data.

2.1.1. Data Mining

Data mining adalah proses ekstraksi informasi yang berharga dan berpotensi bermanfaat dari dalam kumpulan data yang besar dan kompleks [2] [3]. Tujuan utama dari data mining adalah untuk mengidentifikasi pola, hubungan, atau wawasan yang tidak terlihat secara langsung melalui analisis konvensional [4] [5]. Metode data mining mencakup teknik-teknik seperti clustering, association rule

mining, classification, dan regression analysis [6]. Proses ini melibatkan langkah-langkah seperti pengumpulan data, preprocessing data, pemodelan, evaluasi model, dan interpretasi hasil. Data mining digunakan secara luas dalam berbagai industri, termasuk pemasaran, keuangan, ilmu kesehatan, dan manufaktur. Contoh penerapan data mining termasuk identifikasi perilaku konsumen, deteksi kecurangan keuangan, pengelompokan pelanggan, dan prediksi tren pasar. Dengan kemampuannya untuk mengurai kompleksitas data dan menghasilkan wawasan yang berharga, data mining memainkan peran krusial dalam mendukung pengambilan keputusan yang informasional dan strategis.

Data mining dengan model asosiasi merupakan pendekatan yang sangat berguna dalam menganalisis dataset untuk mengidentifikasi hubungan atau pola asosiatif antaritem [7]. Model asosiasi dalam data mining fokus pada penemuan aturan yang menyatakan kecenderungan atau hubungan antara satu set item dengan item lainnya dalam dataset transaksional. Salah satu algoritma yang populer dalam model asosiasi adalah algoritma Apriori. Proses data mining dengan model asosiasi melibatkan beberapa langkah, termasuk pembentukan kandidat itemset, perhitungan frekuensi munculnya itemset, pemilihan itemset dengan frekuensi di atas ambang batas, dan pembentukan aturan asosiasi. Aplikasi dari model asosiasi dalam data mining sangat beragam, mulai dari analisis pola pembelian konsumen, rekomendasi produk, hingga deteksi kecenderungan dalam data transaksional bisnis. Model asosiasi memberikan wawasan berharga yang dapat membantu organisasi membuat keputusan strategis, meningkatkan pengalaman pengguna, dan mengoptimalkan proses bisnis. Dengan menggunakan

model asosiasi, perusahaan dapat mengidentifikasi pola hubungan yang tidak terlihat secara langsung, memungkinkan mereka untuk mengambil langkah-langkah yang lebih terarah dan efisien berdasarkan tren dan asosiasi yang terungkap dari analisis data.

2.1.2. Database dan Data Processing

Database dan data processing merupakan elemen kunci dalam pengelolaan informasi di era digital saat ini. Sebuah database adalah koleksi terstruktur dari data yang disimpan secara elektronik, memungkinkan untuk pengorganisasian, penyimpanan, dan pengambilan informasi dengan efisien. Sistem manajemen basis data (DBMS) digunakan untuk mengelola, memelihara, dan menyediakan akses ke database. Proses pemrosesan data melibatkan transformasi data dari berbagai sumber menjadi bentuk yang lebih berguna dan informatif. Ini melibatkan langkah-langkah seperti pengumpulan data, validasi, normalisasi, transformasi, dan pelaporan. Kombinasi database dan data processing memberikan fondasi bagi aplikasi dan sistem informasi yang dapat mendukung berbagai kebutuhan organisasi, mulai dari manajemen inventaris dan transaksi keuangan hingga analisis bisnis dan pengembangan aplikasi canggih berbasis data. Dengan adanya teknologi ini, organisasi dapat dengan cepat mengakses, menyimpan, dan memanipulasi data untuk mendukung pengambilan keputusan yang tepat waktu dan efektif.

Database dan data preprocessing memainkan peran kunci dalam konteks data mining, membentuk dasar yang kuat untuk ekstraksi wawasan dan penemuan pola dalam dataset yang besar dan kompleks. Sebuah database menyimpan data

dalam struktur yang terorganisir, menyediakan landasan untuk mengakses dan menyimpan informasi dengan efisien. Sebelum memulai proses data mining, langkah pertama yang penting adalah memahami struktur dan karakteristik dataset dalam database.

Data preprocessing, di sisi lain, adalah langkah yang kritis dalam persiapkan data sebelum diterapkan pada algoritma data mining. Proses ini mencakup langkah-langkah seperti pembersihan data untuk mengatasi nilai yang hilang atau tidak valid, penghilangan duplikasi, normalisasi untuk menyesuaikan rentang nilai, dan transformasi data untuk memastikan konsistensi dan relevansi. Selain itu, data preprocessing juga mencakup pemilihan atribut yang relevan dan teknik reduksi dimensi untuk mengurangi kompleksitas dataset. Menyiapkan data dengan benar memastikan bahwa algoritma data mining dapat bekerja secara efektif dan menghasilkan hasil yang akurat. Oleh karena itu, kombinasi antara sebuah database yang baik dan proses data preprocessing yang cermat menciptakan fondasi yang kuat untuk penggalian informasi yang berharga dan pemahaman yang lebih dalam dari data.

Dalam rangkaian data mining, penggunaan database yang efisien dan teknik data preprocessing yang tepat memastikan bahwa analisis data dapat dilakukan dengan akurat dan efektif, membuka peluang untuk menemukan pola yang bermanfaat, tren, dan wawasan penting bagi pengambilan keputusan yang informasional dan strategis.

2.1.3. Visualisation

Visualisasi data adalah teknik untuk menyajikan informasi dan data dalam bentuk grafis yang mudah dipahami. Tujuannya adalah untuk membuat data kompleks menjadi lebih mudah diinterpretasikan, memungkinkan pengguna untuk mengidentifikasi pola, tren, dan wawasan yang mungkin sulit dilihat dalam bentuk tabel atau angka. Dengan menggunakan berbagai jenis grafik seperti diagram batang, diagram lingkaran, peta panas, dan grafik garis, visualisasi membantu memvisualisasikan hubungan antara variabel dan memfasilitasi pemahaman yang lebih baik. Selain itu, alat visualisasi modern juga menggunakan teknologi interaktif untuk memungkinkan pengguna berinteraksi dengan data, menggali lebih dalam ke dalam informasi yang tersedia. Visualisasi data memiliki peran penting dalam berbagai bidang, termasuk bisnis, ilmu pengetahuan, pemasaran, dan pemerintahan. Dengan cara ini, visualisasi data tidak hanya mempermudah komunikasi informasi kompleks, tetapi juga mendukung pengambilan keputusan yang lebih baik dan inovasi melalui pemahaman yang lebih mendalam tentang pola data.

Visualisasi memainkan peran kunci dalam data mining dengan memfasilitasi interpretasi dan pemahaman yang lebih baik terhadap pola dan tren dalam dataset yang kompleks. Dalam konteks data mining, visualisasi memungkinkan para analis dan pengambil keputusan untuk mewakili informasi yang kompleks secara intuitif dan efektif. Grafik, diagram, dan visualisasi interaktif membantu menggambarkan hubungan antarvariabel, distribusi data, serta deteksi outlier dengan cara yang mudah dimengerti. Visualisasi juga membantu dalam eksplorasi

data, memungkinkan pengguna untuk mengidentifikasi pola secara visual sebelum menerapkan algoritma data mining. Pemilihan metode visualisasi yang tepat dapat membimbing pemahaman terhadap aspek-aspek tertentu dari dataset, seperti sebaran data, korelasi, atau cluster yang mungkin ada.

Selain itu, visualisasi juga penting dalam menyajikan hasil dari analisis data mining secara jelas dan persuasif kepada pemangku kepentingan yang mungkin tidak memiliki latar belakang teknis. Dengan menggunakan visualisasi yang efektif, tim manajerial atau pihak non-teknis dapat dengan cepat memahami temuan penting dan membuat keputusan yang terinformasi. Dengan menyediakan pandangan yang lebih intuitif terhadap informasi yang tersembunyi dalam data, visualisasi memainkan peran krusial dalam mendukung tujuan utama data mining, yaitu mengungkap wawasan yang dapat digunakan untuk mendukung pengambilan keputusan yang lebih baik dan inovasi dalam berbagai bidang.

2.1.4. Statistik

Statistik adalah cabang ilmu matematika yang berkaitan dengan pengumpulan, analisis, interpretasi, presentasi, dan pengorganisasian data. Tujuannya adalah untuk menyajikan pola atau tren yang ada dalam data, membantu pengambilan keputusan, dan memberikan dasar untuk membuat kesimpulan yang dapat diterima secara ilmiah. Metode statistik mencakup konsep probabilitas, inferensi statistik, dan analisis regresi, yang digunakan untuk membuat generalisasi atau prediksi dari sampel data terbatas ke populasi yang lebih besar. Statistik diterapkan dalam berbagai bidang seperti ekonomi, sains, kesehatan, dan industri untuk menguji hipotesis, mengukur variabilitas, dan

mengidentifikasi hubungan antarvariabel. Dengan bantuan teknologi modern, analisis statistik semakin dapat diakses dan diterapkan dalam berbagai konteks, memberikan kontribusi signifikan dalam pengambilan keputusan yang informasional dan ilmiah.

Statistik memiliki peran sentral dalam data mining, membentuk landasan matematis untuk analisis dan interpretasi pola dalam dataset yang kompleks. Dalam konteks data mining, konsep dan teknik statistik digunakan untuk merangkum, menyajikan, dan menafsirkan informasi yang terdapat dalam data. Metode statistik seperti mean, median, dan deviasi standar digunakan untuk memberikan gambaran ringkas tentang distribusi dan karakteristik sentral dari atribut dalam dataset. Selain itu, uji hipotesis dan analisis varians dapat digunakan untuk menentukan apakah perbedaan antara kelompok dalam dataset signifikan secara statistik. Ini membantu dalam pengambilan keputusan yang didasarkan pada bukti statistik, mengidentifikasi pola yang lebih dari sekadar kebetulan.

Pada tahap eksplorasi data, statistik deskriptif membantu pengguna untuk memahami struktur data dan mengidentifikasi nilai-nilai ekstrem atau anomali. Selanjutnya, pada tahap pemodelan, teknik statistik seperti regresi dapat membantu memahami hubungan antarvariabel dan memprediksi nilai yang tidak diketahui. Dengan menggabungkan statistik dalam data mining, para analis dapat menyelidiki struktur dataset, mengukur kepastian dalam temuan, dan membuat generalisasi yang didukung oleh analisis statistik. Oleh karena itu, statistik bukan hanya merupakan alat bantu, melainkan fondasi yang memberikan kerangka kerja

kritis untuk interpretasi yang tepat dan pemahaman mendalam terhadap data dalam proses data mining.

2.1.5. Pattern Recognition

Pattern recognition, atau pengenalan pola, adalah bidang studi dalam ilmu komputer dan kecerdasan buatan yang berfokus pada identifikasi dan interpretasi otomatis terhadap pola dalam data. Tujuan dari pengenalan pola adalah untuk mengembangkan algoritma dan model yang memungkinkan mesin mengenali dan mengategorikan pola, baik dalam bentuk gambar, sinyal, teks, atau bentuk data lainnya. Bidang ini mengambil inspirasi dari kognisi dan persepsi manusia, dengan tujuan mereplikasi kemampuan otak manusia untuk mengenali dan menginterpretasi pola kompleks dengan mudah. Pengenalan pola melibatkan berbagai teknik, termasuk pembelajaran mesin, jaringan saraf tiruan, dan metode statistik, untuk melatih sistem agar dapat mengidentifikasi pola dan membuat prediksi atau klasifikasi berdasarkan pola tersebut. Aplikasi dari pengenalan pola melibatkan berbagai bidang, mulai dari pengenalan gambar dan suara hingga diagnosis medis dan deteksi kecurangan. Seiring berkembangnya teknologi, pengenalan pola semakin memainkan peran penting dalam otomatisasi tugas yang memerlukan pemahaman dan interpretasi terhadap pola-pola rumit dalam berbagai dataset.

Pattern recognition, atau pengenalan pola, memainkan peran integral dalam data mining dengan fokus pada identifikasi dan interpretasi pola yang tersembunyi atau signifikan dalam dataset. Dalam konteks data mining, pattern recognition mencakup teknik-teknik yang memungkinkan sistem untuk secara otomatis

mengenali pola-pola kompleks dalam data, baik itu berupa gambar, teks, suara, atau bentuk data lainnya. Algoritma machine learning, seperti klasifikasi dan clustering, digunakan untuk melatih sistem agar dapat mengenali dan memahami hubungan antarvariabel atau objek dalam dataset.

Penggunaan pattern recognition dalam data mining memungkinkan para analis untuk mengungkap struktur, tren, atau anomali yang mungkin sulit atau tidak dapat diidentifikasi secara manual. Contohnya, dalam analisis citra medis, pattern recognition dapat membantu mendeteksi pola karakteristik penyakit atau kelainan. Selain itu, dalam bisnis dan pemasaran, dapat digunakan untuk mengidentifikasi pola perilaku konsumen dan preferensi produk. Melalui penerapan pattern recognition, data mining menjadi lebih dari sekadar pengolahan data, melainkan suatu pendekatan yang memungkinkan sistem untuk belajar dan mengenali pola-pola yang muncul dalam dataset yang besar dan kompleks. Dengan demikian, pattern recognition tidak hanya memfasilitasi analisis yang lebih mendalam, tetapi juga mendukung pengambilan keputusan yang lebih informasional dan responsif terhadap dinamika data yang terus berkembang.

2.2. Model Asosiasi

Model asosiasi adalah konsep dalam data mining yang digunakan untuk mengidentifikasi hubungan dan asosiasi antara berbagai item atau variabel dalam suatu dataset [8]. Metode ini berfokus pada penemuan aturan asosiasi yang menyatakan keterkaitan antara satu set item atau kejadian dengan kejadian lainnya [9]. Algoritma asosiasi umumnya digunakan untuk menganalisis pola pembelian, perilaku konsumen, dan aspek lain dari data transaksional [10]. Contoh model

asosiasi yang terkenal adalah algoritma Apriori, yang bekerja dengan prinsip apriori, yaitu jika suatu itemset sering muncul, maka itemset yang lebih kecil yang terkandung di dalamnya juga sering muncul. Model asosiasi memainkan peran penting dalam analisis bisnis dan pemasaran, membantu organisasi untuk memahami hubungan antarproduk atau kejadian, sehingga dapat diambil keputusan strategis, seperti pengaturan harga, penempatan produk, atau kampanye pemasaran yang lebih efektif. Dengan menerapkan model asosiasi, perusahaan dapat mengoptimalkan strategi mereka berdasarkan wawasan yang diperoleh dari pola asosiasi yang teridentifikasi dalam data mereka.

2.3. Algoritma Apriori

Algoritma Apriori adalah salah satu algoritma asosiasi yang digunakan dalam data mining untuk menemukan pola-pola asosiasi dalam dataset transaksional [11]. Dikembangkan oleh Rakesh Agrawal dan Ramakrishnan Srikant pada tahun 1994, algoritma ini memanfaatkan konsep "apriori property", yaitu jika suatu itemset sering muncul, maka itemset yang lebih kecil yang terkandung di dalamnya juga sering muncul. Langkah-langkah utama algoritma Apriori melibatkan pembentukan kandidat itemset, pencarian frekuensi itemset dengan mendiskritikan dataset, dan penggunaan aturan asosiasi untuk mengekstrak pola yang signifikan. Algoritma ini berguna untuk menganalisis data transaksional dan mengidentifikasi hubungan antara berbagai item yang sering muncul bersama-sama. Penerapan algoritma Apriori umumnya ditemukan dalam konteks toko ritel, di mana informasi pola pembelian dapat digunakan untuk menyusun strategi pemasaran yang lebih efektif atau mengoptimalkan susunan produk di

dalam toko. Meskipun algoritma Apriori efektif dalam mengekstrak pola asosiasi, kemunculan algoritma lain seperti FP-growth telah memberikan alternatif yang lebih efisien untuk dataset yang besar.

Algoritma Apriori adalah algoritma klasik dalam data mining yang digunakan untuk mengekstrak aturan asosiasi dari dataset transaksional [12]. Dikembangkan oleh Rakesh Agrawal dan Ramakrishnan Srikant pada tahun 1994, algoritma ini memanfaatkan prinsip "apriori property," yang menyatakan bahwa jika suatu itemset sering muncul, maka subset dari itemset tersebut juga harus sering muncul. Langkah-langkah utama algoritma Apriori melibatkan pembentukan kandidat itemset, pencarian frekuensi itemset dengan mendiskritikan dataset, dan pemilihan kandidat itemset berdasarkan ambang batas support.

Pertama, algoritma membentuk kandidat 1-itemset dari semua item yang ada dalam dataset transaksional. Selanjutnya, dengan menggunakan prinsip apriori, algoritma menghasilkan kandidat-kandidat itemset berikutnya dengan menggabungkan itemset yang sering muncul [13]. Setiap kandidat itemset kemudian dihitung frekuensinya dalam dataset, dan itemset yang memenuhi ambang batas support dianggap sebagai itemset frequent. Algoritma Apriori terkenal karena kemampuannya dalam menangani sejumlah besar data dan mengekstrak aturan asosiasi dengan efisien. Meskipun demikian, performanya dapat dipengaruhi oleh ukuran dan kompleksitas dataset. Seiring waktu, beberapa variasi dan perbaikan pada algoritma Apriori telah dikembangkan untuk meningkatkan efisiensi dan kinerjanya dalam skenario tertentu. Dengan kontribusinya yang signifikan, algoritma Apriori tetap menjadi bagian penting

dalam pemahaman dan analisis pola asosiasi dalam dataset transaksional dalam dunia data mining. Berikut adalah langkah-langkah umum algoritma Apriori:

1. Pembentukan Itemset Satu Item (1-itemset)

Pembentukan 1-itemset adalah langkah awal yang kritis dalam algoritma Apriori pada proses data mining. Pada tahap ini, algoritma menciptakan himpunan itemset yang terdiri dari satu elemen atau item tunggal yang ditemukan dalam dataset transaksional. Proses ini melibatkan pemindahan setiap elemen unik dari transaksi ke dalam 1-itemset. Dengan kata lain, setiap item yang ada dalam transaksi menjadi satu kandidat 1-itemset. Langkah ini membantu membentuk dasar untuk pembentukan kandidat itemset yang lebih besar pada langkah-langkah berikutnya dalam algoritma Apriori.

Contoh sederhana dapat menjelaskan langkah ini dengan lebih baik. Misalnya, dalam sebuah dataset transaksional yang mencatat pembelian di sebuah toko online, jika terdapat produk A, B, C, dan D dalam berbagai transaksi, pembentukan 1-itemset akan menghasilkan empat kandidat 1-itemset: {A}, {B}, {C}, dan {D}. Proses ini membantu algoritma Apriori untuk memulai identifikasi item yang memiliki tingkat kemunculan yang cukup tinggi, yang kemudian akan digunakan untuk membentuk kandidat itemset yang lebih besar pada langkah-langkah berikutnya. Dengan pembentukan 1-itemset sebagai titik awal, algoritma Apriori dapat mengoptimalkan pencarian pola asosiasi yang signifikan dalam dataset transaksional secara efisien.

2. Pembentukan Kandidat Itemset Dua Item (2-itemset)

Langkah berikutnya dalam algoritma Apriori setelah pembentukan 1-itemset adalah pembentukan kandidat 2-itemset. Pada tahap ini, algoritma menggabungkan itemset-itemset tunggal yang ditemukan pada langkah sebelumnya untuk membentuk himpunan kandidat 2-itemset. Prinsip Apriori yang mendasar adalah bahwa jika suatu itemset sering muncul, maka subset dari itemset tersebut juga harus sering muncul. Oleh karena itu, algoritma menggabungkan dua 1-itemset yang berbeda untuk membentuk 2-itemset.

Misalkan pada tahap sebelumnya, kita memiliki kandidat 1-itemset $\{A\}$, $\{B\}$, $\{C\}$, dan $\{D\}$. Maka, pada tahap pembentukan kandidat 2-itemset, algoritma akan menggabungkan setiap pasangan kandidat 1-itemset untuk membentuk himpunan 2-itemset. Contohnya, gabungan dari $\{A\}$ dan $\{B\}$ menghasilkan $\{A, B\}$, dan seterusnya. Langkah ini membantu dalam mengekplorasi hubungan atau asosiasi antara dua item dalam dataset transaksional.

Proses pembentukan kandidat 2-itemset ini adalah langkah krusial yang memungkinkan algoritma Apriori untuk membangun asosiasi yang lebih kompleks dan mendalam. Itemset-itemset yang dihasilkan pada tahap ini akan diuji kemunculannya dalam dataset, dan hanya itemset-itemset yang memenuhi ambang batas support yang ditentukan yang akan dipertahankan untuk langkah-langkah berikutnya. Dengan cara ini, algoritma Apriori dapat mengidentifikasi pola asosiasi yang kuat dan relevan dengan efisien pada dataset transaksional yang besar.

3. Pencarian Frekuensi 2-itemset

Pencarian frekuensi 2-itemset merupakan langkah penting dalam algoritma Apriori yang bertujuan untuk mengidentifikasi itemset yang memiliki kemunculan atau dukungan (support) yang memenuhi ambang batas yang telah ditetapkan. Setelah membentuk kandidat 2-itemset dari langkah sebelumnya, algoritma Apriori menghitung seberapa sering setiap 2-itemset muncul dalam dataset transaksional.

Proses ini melibatkan pemindaian seluruh dataset untuk mencari setiap kemunculan dari kandidat 2-itemset yang telah dibentuk sebelumnya. Jumlah kemunculan ini kemudian digunakan untuk menghitung frekuensi atau dukungan dari masing-masing 2-itemset. Dukungan didefinisikan sebagai jumlah transaksi yang mengandung 2-itemset tertentu dibagi dengan total jumlah transaksi dalam dataset.

Dengan mengidentifikasi frekuensi 2-itemset, algoritma Apriori dapat menyaring itemset yang tidak memenuhi ambang batas support yang ditentukan sebelumnya. Itemset-itemset yang memenuhi syarat ini akan dipertahankan untuk langkah-langkah berikutnya dalam algoritma, sementara yang tidak memenuhi akan dihapus. Pencarian frekuensi 2-itemset membantu mempersempit fokus analisis pada itemset-itemset yang relevan dan signifikan dalam konteks data mining, mengurangi kompleksitas pencarian pola asosiasi secara keseluruhan.

4. Pemilihan Kandidat Itemset K (k-itemset)

Setelah menemukan frekuensi atau dukungan dari kandidat 2-itemset, langkah berikutnya dalam algoritma Apriori adalah pemilihan kandidat itemset.

Pemilihan ini melibatkan penggabungan k-itemset yang memiliki $\binom{k-1}{k}$ item yang sama untuk membentuk kandidat baru. Misalkan pada tahap sebelumnya, algoritma telah menemukan kandidat 2-itemset $\{A, B\}$, $\{B, C\}$, dan $\{A, C\}$ dengan dukungan yang memenuhi ambang batas. Pada langkah ini, algoritma akan menggabungkan itemset-itemset tersebut untuk membentuk kandidat 3-itemset baru, seperti $\{A, B, C\}$. Proses ini dilakukan secara berulang untuk membentuk kandidat-kandidat k-itemset yang potensial.

Pemilihan kandidat itemset k memungkinkan algoritma Apriori untuk secara iteratif memperluas ruang pencarian untuk pola asosiasi yang lebih kompleks dan mendalam. Meskipun kompleksitasnya meningkat dengan setiap penambahan item, prinsip apriori tetap berlaku, yaitu jika suatu itemset sering muncul, maka subset dari itemset tersebut juga harus sering muncul. Selanjutnya, pada langkah-langkah berikutnya, algoritma akan menghitung dukungan dari kandidat-kandidat k-itemset yang baru dibentuk dan mempertahankan hanya yang memenuhi ambang batas yang ditetapkan. Proses ini berulang hingga tidak ada lagi kandidat k-itemset yang dapat dibentuk atau ditemukan yang memenuhi ambang batas support. Dengan demikian, pemilihan kandidat itemset k merupakan langkah kunci dalam membangun tingkat kompleksitas algoritma Apriori dan mengidentifikasi pola asosiasi yang semakin lengkap dan terperinci pada setiap iterasinya.

5. Pencarian Frekuensi K-itemset

Pencarian frekuensi K-itemset adalah tahap yang menentukan dalam algoritma Apriori di mana dukungan (support) dari kandidat itemset dengan

ukuran $\setminus(K)$ dihitung untuk memastikan keberlanjutan identifikasi pola asosiasi yang kompleks. Setelah kandidat-kandidat itemset K dibentuk pada langkah sebelumnya, algoritma Apriori melakukan pemindaian dataset untuk menghitung frekuensi kemunculan setiap kandidat K-itemset. Dukungan K-itemset diukur dengan menghitung jumlah transaksi yang mengandung K-itemset tersebut dan membaginya dengan total jumlah transaksi dalam dataset.

Langkah ini memegang peran krusial dalam menyaring pola-pola yang signifikan dan relevan dalam data. Hanya K-itemset yang memenuhi ambang batas support yang ditentukan yang dipertahankan untuk langkah-langkah berikutnya dalam algoritma. Dengan cara ini, algoritma Apriori secara efisien mempersempit fokus analisis pada pola asosiasi yang memiliki dukungan yang cukup tinggi, menghindari pencarian pola yang kurang signifikan.

Pencarian frekuensi K-itemset memungkinkan algoritma untuk melanjutkan proses iteratifnya dengan itemset yang semakin kompleks, membangun asosiasi yang semakin mendalam. Ini memungkinkan penemuan pola asosiasi yang lebih kompleks dan kontekstual yang mungkin tidak terlihat ketika hanya mempertimbangkan itemset yang lebih kecil. Dengan demikian, langkah ini menjadi tonggak dalam meningkatkan kedalaman dan kekayaan analisis pola asosiasi dalam data mining menggunakan algoritma Apriori.

6. Iterasi dan Pengulangan

Iterasi dan pengulangan adalah aspek kunci dalam algoritma Apriori yang memungkinkan pendalaman dan peningkatan kualitas hasil identifikasi pola asosiasi. Setelah pencarian frekuensi K-itemset selesai, langkah ini melibatkan

pembentukan kandidat itemset untuk ukuran K+1, menghitung dukungan masing-masing kandidat, dan membuang itemset yang tidak memenuhi ambang batas support yang telah ditetapkan. Proses ini diulangi secara iteratif hingga tidak ada lagi kandidat itemset yang dapat dibentuk atau dipertahankan yang memenuhi kriteria dukungan.

Selama setiap iterasi, algoritma Apriori secara cerdas memanfaatkan prinsip apriori, yaitu jika suatu itemset sering muncul, maka subset dari itemset tersebut juga harus sering muncul. Ini memungkinkan algoritma untuk fokus pada pembentukan dan evaluasi itemset yang memiliki potensi dukungan tinggi, meminimalkan upaya pada itemset yang memiliki kemunculan rendah. Proses iteratif dan pengulangan ini membentuk dasar kekuatan algoritma Apriori dalam menangani dataset yang besar dan kompleks. Melalui setiap iterasi, algoritma secara progresif memperluas ruang pencarian, menyaring itemset yang kurang relevan, dan meningkatkan kompleksitas analisis pola asosiasi. Dengan pendekatan yang sistematis, algoritma ini secara efisien mengidentifikasi dan mempertahankan itemset-itemset yang signifikan, memungkinkan penemuan wawasan yang lebih dalam dan berarti dari data. Keberhasilan algoritma Apriori bergantung pada kesinambungan iterasi ini untuk mendapatkan pemahaman yang mendalam terhadap pola asosiasi dalam dataset transaksional.

7. *Pembentukan Aturan Asosiasi*

Pembentukan aturan asosiasi merupakan tahap penting dalam algoritma Apriori yang mengarah pada hasil akhir yang sangat bermanfaat: aturan yang menggambarkan hubungan dan pola asosiasi antaritem dalam dataset. Setelah

identifikasi itemset-itemset yang memenuhi ambang batas support, langkah berikutnya adalah menciptakan aturan asosiasi dengan mempertimbangkan hubungan antaritem di dalam itemset tersebut.

Pertama, algoritma Apriori menghasilkan aturan untuk setiap itemset yang telah teridentifikasi. Aturan asosiasi dinyatakan dalam bentuk " $A \Rightarrow B$," yang berarti jika suatu transaksi mengandung itemset A, maka transaksi tersebut cenderung juga mengandung itemset B. Selanjutnya, algoritma menghitung metrik seperti confidence dan lift untuk menilai kekuatan dan relevansi dari masing-masing aturan.

Confidence mengukur seberapa sering itemset B muncul dalam transaksi yang juga mengandung itemset A, sementara lift menunjukkan seberapa besar peningkatan kemungkinan munculnya itemset B ketika itemset A telah muncul. Aturan dengan confidence dan lift yang tinggi dianggap lebih kuat dan relevan dalam konteks data mining.

Proses ini dilakukan untuk setiap itemset yang memenuhi ambang batas support, dan aturan yang dihasilkan dapat memberikan wawasan yang mendalam tentang keterkaitan antaritem dalam dataset. Hasil akhirnya adalah sejumlah aturan asosiasi yang dapat digunakan untuk membuat prediksi, memberikan rekomendasi, atau memahami pola perilaku dalam suatu domain tertentu. Dengan langkah ini, algoritma Apriori memberikan kontribusi penting dalam analisis dan interpretasi dataset transaksional, memungkinkan pemahaman yang lebih baik tentang hubungan antaritem dan memberikan landasan yang kuat untuk pengambilan keputusan yang terinformasi.

2.3.1. Uji Performa

2.4. Alat Bantu Program/Tools Pendukung

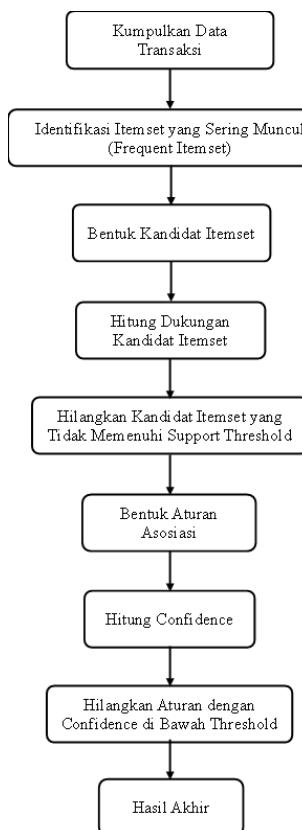
2.5.1. Orange

Orange adalah sebuah platform open-source yang digunakan untuk analisis data dan machine learning. Dikembangkan oleh laboratorium AI di Universitas Ljubljana, Slovenia, Orange memberikan alat visual yang intuitif untuk eksplorasi, pemrosesan, dan pemodelan data. Salah satu keunggulan utama dari Orange adalah antarmuka pengguna grafisnya yang memungkinkan pengguna, termasuk mereka yang tidak memiliki latar belakang teknis yang mendalam, untuk dengan mudah merancang dan melaksanakan analisis data kompleks. Aplikasi Orange digunakan dalam berbagai konteks, termasuk riset akademis, industri, dan pengembangan proyek-proyek machine learning. Orange menyediakan berbagai komponen atau "widget" yang dapat digabungkan dalam suatu alur kerja visual. Widget ini mencakup alat-alat untuk pre-processing data, eksplorasi data, pembuatan model machine learning, evaluasi model, dan visualisasi hasil.

Dengan Orange, pengguna dapat melakukan tugas-tugas seperti clustering, klasifikasi, regresi, dan analisis komponen utama dengan mudah. Selain itu, platform ini mendukung integrasi dengan berbagai teknik machine learning dan statistik yang kuat, memungkinkan pengguna untuk merancang dan mengimplementasikan solusi analisis data yang komprehensif. Kesederhanaan penggunaan dan fleksibilitas Orange membuatnya menjadi pilihan yang populer

bagi para profesional dan peneliti yang ingin menjelajahi dan memahami data dengan cepat dan efektif.

2.5.2. Lembar Kerja



1. Kumpulkan Data Transaksi

Proses pengumpulan data transaksi merupakan langkah kritis dalam analisis data yang bertujuan untuk mengidentifikasi pola-pola asosiasi di antara item atau produk dalam suatu konteks bisnis. Data transaksi yang terkumpul mencakup informasi mengenai setiap pembelian yang dilakukan oleh pelanggan atau entitas lainnya. Ini bisa melibatkan sejumlah besar transaksi, termasuk pembelian produk, layanan, atau interaksi bisnis lainnya. Pada tahap ini, informasi yang dikumpulkan mungkin mencakup detail seperti jenis produk atau layanan yang dibeli, jumlah

pembelian, waktu transaksi, dan identitas pelanggan. Data transaksi yang lengkap dan akurat menjadi kunci untuk memahami pola-pola konsumen, preferensi, dan kecenderungan pembelian. Pengumpulan data yang cermat dan komprehensif menjadi dasar yang solid untuk analisis lebih lanjut.

Teknologi modern, seperti sistem manajemen basis data atau perangkat lunak analisis data, dapat memudahkan proses pengumpulan dan penyimpanan data transaksi dengan efisiensi. Pengelolaan data transaksi yang baik membuka pintu untuk pengembangan model analisis yang dapat memberikan wawasan berharga terkait perilaku pelanggan, tren pasar, dan peluang bisnis. Seiring dengan pertumbuhan besar-besaran data, perusahaan kini semakin mengenali pentingnya memanfaatkan data transaksi dengan bijak. Dengan melibatkan metode-metode analisis yang tepat, pengumpulan data transaksi dapat menjadi landasan untuk mengidentifikasi pola-pola yang dapat memberikan keunggulan kompetitif dan mendukung pengambilan keputusan yang lebih informasional dan cerdas.

2. Identifikasi Itemset yang Sering Muncul (Frequent Itemset)

Identifikasi Itemset yang Sering Muncul (Frequent Itemset) adalah tahap kritis dalam analisis data menggunakan model asosiasi. Setelah data transaksi berhasil dikumpulkan, langkah selanjutnya adalah mengidentifikasi kelompok item atau produk yang sering muncul bersama-sama dalam transaksi. Frequent Itemset mencerminkan pola-pola yang konsisten dan signifikan dalam perilaku pembelian atau interaksi lainnya yang terdapat dalam dataset.

Proses identifikasi Frequent Itemset melibatkan perhitungan frekuensi kemunculan setiap kombinasi item dalam dataset. Itemset yang memiliki

frekuensi di atas suatu ambang batas yang ditentukan disebut sebagai Frequent Itemset. Ambang batas ini, biasanya disebut sebagai support threshold, adalah parameter yang dapat disesuaikan sesuai dengan kebutuhan analisis dan tujuan bisnis. Mengidentifikasi Frequent Itemset membantu menyaring item atau produk yang mungkin memiliki keterkaitan atau asosiasi yang signifikan dalam suatu konteks. Analisis ini memberikan pemahaman mendalam tentang hubungan antar-item yang dapat memberikan wawasan berharga terkait tren konsumen, strategi pemasaran yang efektif, atau peningkatan efisiensi operasional.

Teknik-teknik seperti algoritma Apriori atau FP-Growth sering digunakan untuk melakukan identifikasi Frequent Itemset dengan efisien. Hasil dari langkah ini menjadi dasar untuk tahap selanjutnya dalam model asosiasi, seperti pembentukan aturan asosiasi dan perhitungan tingkat kepercayaan (confidence) yang memberikan pemahaman lebih lanjut tentang pola-pola hubungan antar-item. Identifikasi Frequent Itemset merupakan langkah kunci dalam membuka potensi wawasan berharga yang dapat mendukung pengambilan keputusan dan strategi bisnis.

3. Bentuk Kandidat Itemset

Bentuk Kandidat Itemset merupakan tahap berikutnya dalam proses analisis data menggunakan model asosiasi setelah berhasil mengidentifikasi Frequent Itemset. Pada tahap ini, tujuan utamanya adalah membentuk kombinasi potensial dari item atau produk yang mungkin membentuk pola-pola baru atau asosiasi yang belum teridentifikasi. Proses pembentukan kandidat itemset dilakukan dengan eksplorasi kombinasi semua item yang ada dalam dataset transaksi.

Algoritma-asosiasi seperti Apriori atau FP-Growth sering digunakan untuk membentuk kandidat itemset dengan efisien. Apriori, sebagai contoh, bekerja dengan menyusun kandidat itemset dari itemset yang lebih kecil yang telah terbukti sering muncul bersama dalam transaksi. Dengan pendekatan ini, pembentukan kandidat itemset menjadi lebih efisien, mengurangi kompleksitas perhitungan dan mempercepat proses analisis.

Setelah kandidat itemset terbentuk, tahap selanjutnya melibatkan perhitungan dukungan (support) untuk masing-masing kandidat. Support adalah ukuran seberapa sering kandidat itemset muncul dalam dataset transaksi. Langkah ini memungkinkan analis untuk memfilter kandidat itemset yang memenuhi batas support threshold yang telah ditentukan sebelumnya. Bentuk Kandidat Itemset menjadi langkah kritis dalam menyusun dasar bagi analisis selanjutnya. Dengan memahami kombinasi potensial dari item-item yang ada, organisasi dapat mengidentifikasi pola-pola asosiasi yang belum terungkap sebelumnya dan mendapatkan wawasan baru yang dapat meningkatkan pemahaman terhadap perilaku konsumen atau dinamika bisnis. Proses pembentukan kandidat itemset membuka jalan untuk pemodelan yang lebih lanjut, seperti pembentukan aturan asosiasi dan pengembangan strategi berbasis data yang lebih canggih.

4. Hitung Dukungan Kandidat Itemset

Setelah membentuk kandidat itemset, tahap selanjutnya dalam analisis data menggunakan model asosiasi adalah menghitung dukungan (support) untuk masing-masing kandidat itemset. Dukungan merupakan metrik yang mengukur seberapa sering suatu kandidat itemset muncul dalam dataset transaksi.

Perhitungan dukungan ini sangat penting karena membantu mengidentifikasi itemset yang memiliki dampak dan relevansi yang signifikan dalam konteks analisis. Proses menghitung dukungan kandidat itemset melibatkan pencatatan kemunculan setiap kandidat dalam setiap transaksi. Jumlah transaksi yang mengandung suatu kandidat itemset dibagi dengan total jumlah transaksi untuk mendapatkan nilai dukungan. Nilai dukungan ini kemudian digunakan untuk menentukan apakah suatu kandidat itemset dapat dianggap sebagai Frequent Itemset atau tidak.

Dalam prakteknya, batas dukungan (support threshold) sering ditetapkan sebelumnya sebagai parameter yang dapat disesuaikan sesuai dengan kebutuhan analisis dan tujuan bisnis. Kandidat itemset yang memiliki dukungan di atas batas ini dianggap sebagai Frequent Itemset dan dianggap relevan untuk langkah-langkah analisis selanjutnya. Perhitungan dukungan kandidat itemset membantu menyaring itemset yang memiliki dampak yang lebih besar dalam dataset transaksi. Informasi ini dapat memberikan wawasan yang lebih akurat tentang hubungan antar-item yang mungkin membentuk dasar bagi keputusan bisnis yang lebih baik dan strategi pemasaran yang lebih efektif. Proses ini adalah tahap kunci dalam menyaring informasi yang tidak relevan dan memfokuskan analisis pada pola-pola yang memiliki nilai bisnis yang lebih tinggi.

5. Hilangkan Kandidat Itemset yang Tidak Memenuhi Support Threshold

Langkah berikutnya dalam proses analisis data menggunakan model asosiasi adalah menghilangkan kandidat itemset yang tidak memenuhi batas dukungan (support threshold) yang telah ditetapkan sebelumnya. Setelah menghitung

dukungan untuk masing-masing kandidat itemset, langkah ini menjadi krusial dalam menyaring itemset yang kurang relevan atau kurang signifikan dalam dataset transaksi. Penghapusan kandidat itemset yang tidak memenuhi support threshold bertujuan untuk memfokuskan analisis pada pola-pola asosiasi yang lebih kuat dan lebih bermakna. Batas dukungan sendiri adalah ambang batas yang ditentukan sebelumnya, seringkali berdasarkan kebijakan bisnis atau kebutuhan analisis tertentu. Itemset yang tidak mencapai dukungan minimal ini dianggap kurang relevan dalam konteks analisis asosiasi.

Proses penghilangan kandidat itemset yang tidak memenuhi support threshold dapat meningkatkan efisiensi analisis dan membantu memfokuskan perhatian pada informasi yang lebih berharga. Hal ini dapat menghindarkan analis dari pengolahan data yang kurang relevan dan mempercepat langkah-langkah analisis selanjutnya. Selain meningkatkan efisiensi, penghilangan kandidat itemset yang tidak memenuhi support threshold juga memungkinkan fokus analisis pada pola-pola yang dapat memberikan wawasan yang lebih mendalam terkait dengan perilaku konsumen, preferensi produk, atau dinamika bisnis lainnya. Oleh karena itu, langkah ini memberikan kontribusi signifikan dalam menyusun dasar bagi pembentukan aturan asosiasi dan langkah-langkah analisis selanjutnya dalam model asosiasi.

6. Bentuk Aturan Asosiasi

Setelah langkah-langkah sebelumnya seperti pengumpulan data transaksi, identifikasi frequent itemset, dan penghitungan dukungan kandidat itemset, tahap berikutnya dalam model asosiasi adalah membentuk aturan asosiasi. Aturan

asosiasi menggambarkan keterkaitan antar-item atau produk dalam dataset transaksi dan memberikan pandangan yang lebih rinci tentang pola-pola hubungan yang telah diidentifikasi. Pembentukan aturan asosiasi melibatkan kombinasi itemset yang telah diidentifikasi sebelumnya, terutama frequent itemset. Aturan asosiasi biasanya bersifat "jika... maka..." dan menyajikan hubungan antara satu set item (yang disebut sebagai antecedent) dengan set item lainnya (yang disebut sebagai consequent). Proses pembentukan aturan ini memberikan pandangan yang lebih spesifik tentang keterkaitan antar-item, yang dapat memberikan wawasan berharga untuk pengambilan keputusan bisnis.

Satu aspek penting dari pembentukan aturan asosiasi adalah menetapkan tingkat kepercayaan (confidence) untuk setiap aturan. Tingkat kepercayaan mengukur sejauh mana suatu aturan dianggap dapat diandalkan atau valid berdasarkan data transaksi yang ada. Tingkat kepercayaan ini dapat disesuaikan sesuai dengan ambang batas tertentu yang ditetapkan, dan aturan-asosiasi dengan tingkat kepercayaan di bawah batas tersebut dapat dihapus. Penting untuk mencatat bahwa pembentukan aturan asosiasi dapat menghasilkan banyak aturan, dan analis perlu memilih aturan yang paling relevan dan bermakna. Beberapa kriteria pemilihan dapat mencakup tingkat kepercayaan, jumlah item dalam aturan, dan relevansi dengan tujuan analisis bisnis. Proses ini memungkinkan pemodelan yang lebih tepat terhadap pola-pola yang terkandung dalam data transaksi dan memberikan wawasan yang lebih mendalam terkait dengan perilaku konsumen atau dinamika bisnis.

7. Hitung Confidence

Perhitungan confidence adalah tahap penting dalam model asosiasi yang membantu mengukur sejauh mana suatu aturan asosiasi dapat diandalkan berdasarkan data transaksi yang ada. Confidence dinyatakan sebagai persentase atau rasio antara dukungan (support) dari aturan asosiasi tersebut dan dukungan dari antecedent (item atau produk yang muncul di sisi "jika") pada aturan tersebut.

Formula umumnya adalah:

$$\text{Confidence (Aturan)} = \frac{\text{Support (Aturan)}}{\text{Support (Antecedent)}}$$

Aturan-asosiasi yang memiliki confidence tinggi menunjukkan bahwa kemungkinan item atau produk pada sisi "maka" (consequent) akan dibeli jika item pada sisi "jika" (antecedent) juga dibeli sangat tinggi. Semakin tinggi nilai confidence, semakin dapat diandalkan aturan tersebut sebagai indikator hubungan antar-item.

Analisis confidence membantu memahami kekuatan hubungan antar-item dan dapat digunakan untuk mendukung pengambilan keputusan bisnis. Pada umumnya, tingkat confidence yang tinggi di atas batas threshold yang ditetapkan menunjukkan bahwa aturan-asosiasi tersebut memiliki dampak yang kuat dan relevan. Meskipun demikian, tingkat confidence yang sangat tinggi juga dapat menunjukkan bahwa aturan tersebut mungkin terlalu spesifik atau bahkan terlalu sering terjadi dalam dataset. Dalam praktiknya, penetapan batas confidence harus seimbang agar aturan-asosiasi yang dihasilkan relevan, dan hasil interpretasi tetap praktis dan dapat diaplikasikan dalam konteks bisnis. Perhitungan confidence menjadi langkah kunci dalam menyaring dan mengevaluasi aturan-asosiasi

sehingga hasil akhirnya memberikan wawasan yang berguna untuk strategi pemasaran, penataan produk, atau pengambilan keputusan bisnis yang lebih baik.

8. Hilangkan Aturan dengan Confidence di Bawah Threshold

Langkah selanjutnya dalam proses analisis data menggunakan model asosiasi adalah menghilangkan aturan dengan confidence di bawah threshold yang telah ditetapkan. Confidence merupakan ukuran seberapa dapat diandalkannya suatu aturan asosiasi berdasarkan data transaksi yang ada. Dengan menetapkan batas confidence, analis dapat menyaring aturan-asosiasi yang dihasilkan dan memfokuskan perhatian pada aturan yang lebih kuat dan lebih bermakna. Proses penghilangan aturan dengan confidence di bawah threshold adalah esensial karena membantu menyederhanakan dan memfokuskan hasil analisis. Aturan-asosiasi dengan confidence rendah mungkin kurang dapat diandalkan atau memiliki dampak yang lebih terbatas dalam konteks aplikasi bisnis. Dengan menghilangkan aturan-asosiasi ini, analis dapat menyajikan hasil yang lebih terfokus dan relevan untuk pengambilan keputusan.

Menentukan batas confidence yang tepat dapat bervariasi tergantung pada kebutuhan dan tujuan analisis bisnis. Sebagai contoh, batas confidence yang lebih tinggi dapat digunakan ketika kesalahan atau kesalahan interpretasi yang tinggi tidak dapat diterima. Sebaliknya, batas confidence yang lebih rendah mungkin lebih cocok untuk analisis eksploratif atau ketika jumlah data terbatas. Penghilangan aturan dengan confidence di bawah threshold juga berkontribusi pada keberlanjutan dan keberlanjutan interpretasi aturan-asosiasi. Dengan menyajikan aturan-asosiasi yang memiliki tingkat kepercayaan yang lebih tinggi,

analisis dapat memberikan informasi yang lebih dapat diandalkan dan bermakna bagi pemangku kepentingan bisnis. Proses ini mengoptimalkan hasil analisis asosiasi untuk memberikan wawasan yang lebih mendalam terkait dengan perilaku konsumen, strategi pemasaran, atau peningkatan efisiensi operasional.

9. Hasil Akhir

Hasil akhir dari proses analisis data menggunakan model asosiasi merupakan rangkuman keseluruhan dari temuan dan wawasan yang dihasilkan dari langkah-langkah sebelumnya. Inilah tahap akhir yang menyajikan informasi yang dapat dijadikan dasar untuk pengambilan keputusan bisnis atau strategi yang lebih baik. Hasil akhir ini mencakup aturan-asosiasi yang relevan dan memiliki tingkat kepercayaan yang memadai, yang telah disaring berdasarkan dukungan dan confidence yang telah ditentukan sebelumnya. Dalam hasil akhir, aturan-asosiasi yang tersisa sering kali disajikan secara terstruktur dan diprioritaskan berdasarkan kriteria tertentu seperti tingkat kepercayaan tertinggi atau kombinasi item yang paling relevan. Informasi ini dapat memberikan wawasan mendalam tentang pola hubungan antar-item yang ditemukan dalam dataset transaksi, seperti kecenderungan pembelian produk bersamaan atau asosiasi produk yang mungkin kurang terduga.

Hasil akhir juga dapat mencakup visualisasi grafis atau laporan yang memudahkan interpretasi dan komunikasi hasil analisis kepada pemangku kepentingan bisnis. Visualisasi ini dapat memberikan gambaran yang lebih jelas tentang struktur hubungan antar-item, memfasilitasi pemahaman, dan mendukung pengambilan keputusan yang informasional. Selain aturan-asosiasi, hasil akhir

juga bisa mencakup rekomendasi atau implikasi bisnis yang dapat diambil dari analisis tersebut. Pemahaman terhadap pola-pola asosiasi dapat membantu perusahaan merumuskan strategi pemasaran yang lebih efektif, meningkatkan pengalaman pelanggan, atau mengoptimalkan penataan produk di dalam toko.

2.5. Metodologi Penelitian

2.6.1. Penelitian Terdahulu

Referensi Penelitian	1
Judul	PENERAPAN METODE DATA MINING TERHADAP DATA TRANSAKSI PENJUALAN MENGGUNAKAN ALGORITMA APRIORI (STUDI KASUS: TOKO FASENTRO
Nama	Adhitia Erfina1), Melawati2), Nunik Destria Arianti3)
Tahun	2020
Hasil	Penerapan algoritma Apriori dalam penelitian tentang penentuan kombinasi item dan tata letak barang berdasarkan kecenderungan pembelian konsumen telah membuktikan diri sebagai solusi efektif bagi Toko Fasentro Fancy dalam

	mengembangkan strategi pemasaran guna meningkatkan penjualan. Melalui analisis hasil aturan asosiasi, teridentifikasi bahwa konsumen yang membeli produk Pranaya (P24TVC) dan Toyobu Fodu cenderung juga membeli produk Roberto Cavali, dengan nilai support dan confidence tertinggi. Implementasi algoritma Apriori memberikan wawasan yang berharga bagi toko ini untuk mengoptimalkan tata letak produk dan menyusun paket penawaran yang lebih menarik, yang dapat meningkatkan daya tarik dan efisiensi strategi pemasaran mereka [14].
Referensi Penelitian	2
Judul	Penerapan Data Mining pada Tata Letak Buku Di Perpustakaan Sintong Bingei Pematangsiantar dengan Metode Apriori
Nama	Yulia Andini1,* , Jaya Tata Hardinata2, Yuegilion Pranayama Purba3

Tahun	2022
Hasil	<p>Metode Apriori telah berhasil digunakan untuk menentukan pola penataan tata letak buku di perpustakaan, menyediakan sistem yang mempermudah identifikasi kombinasi buku yang sering ditempatkan bersama untuk memenuhi kebutuhan pengunjung. Hasil implementasi menggunakan RapidMiner menunjukkan bahwa kombinasi tata letak buku paling tinggi terjadi antara Ilmu Murni dan Ilmu Sosial, dengan dukungan (support) sebesar 50% dan tingkat kepercayaan (confidence) sebesar 86%. Selain itu, ditemukan juga kombinasi antara Karya Umum dan Ilmu Murni dengan support 41% dan confidence 83%, serta Karya Umum dan Ilmu Sosial dengan support 41% dan confidence 83%. Informasi ini dapat digunakan oleh perpustakaan</p>

	untuk merancang tata letak yang lebih efisien dan meningkatkan kemudahan akses pengunjung terhadap berbagai kategori buku yang sering dicari bersamaan [15].
Referensi Penelitian	3
Judul	Application of the FP-Growth Method to Determine Drug Sales Patterns
Nama	Badrul Anwar1)*, Ambiyar2), Fadhilah3
Tahun	2023
Hasil	Berdasarkan penelitian tersebut, penggunaan metode FP-Growth dalam data mining berhasil mengidentifikasi pola penjualan obat di apotek, yang memungkinkan pengaturan tata letak obat yang lebih optimal. Dengan minimum support dan confidence sebesar $\geq 15\%$, ditemukan pola penjualan obat berpasangan yang sering terjadi, membantu apotek untuk mempercepat proses pelayanan kepada konsumen. Hasil penelitian ini

	<p>menunjukkan bahwa metode FP-Growth efektif dalam meningkatkan efisiensi penjualan obat dan dapat digunakan sebagai panduan untuk pengaturan strategi penjualan yang lebih baik di apotek [16].</p>
--	---