

BAB II

TINJAUAN PUSTAKA

2.1 Teori umum

2.1.1 Data mining

Data mining adalah proses yang menggunakan teknik analisis seperti statistik, matematika, kecerdasan buatan, dan machine learning untuk mencari sebuah informasi dan pengetahuan yang penting dari sebuah kumpulan data dalam jumlah yang besar, tujuan dari data mining adalah menemukan pola tersembunyi yang tidak mudah di ketahui secara manual(Candra Susanto et al., n.d.; Program Studi Sistem Informasi et al., 2020)

Secara sederhana data mining adalah pencarian otomatis pola dalam basis data yang besar, menggunakan teknik komputasional campuran dari statistik, pembelajaran mesin dan penggunaan pola. Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola statistik, database, dan visualisasi untuk penanggapan permasalahan pengambilan informasi dari database dengan jumlah yang besar (Fathurrozi et al., 2023)

2.1.2 Pengelompokan Data Mining

Data mining dikelompokkan menjadi 6 menurut (Listanto & Meisella Kristania, 2022)

1. Dekripsi

Dekripsi dari pola dan kecenderungan sering memberikan sebuah penjelasan untuk suatu pola atau kecenderungan dalam data tersebut.

2. Estimasi

Estimasi mirip dengan klasifikasi, perbedaan terletak pada jenis variabel target yang digunakan. Dalam estimasi, variabel target berupa nilai numerik atau angka, bukan termasuk kategori. Permodelan estimasi dibangun berdasarkan data yang lengkap dan menyertakan variabel target yang akan di prediksi. Setelah model terbentuk, proses akan dilanjutkan dengan estimasi atau prediksi nilai variabel target pada data baru yang belum diketahui nilainya.

3. Prediksi

Prediksi memiliki kesamaan dengan klasifikasi dan estimasi, namun fokus utamanya adalah memperkirakan hasil dimasa depan. Artinya, nilai dari variabel atau kejadian yang diprediksi akan terjadi pada priode yang akan mendatang.

4. Classification

Proses yang bertujuan untuk menyimpulkan beberapa definisi dari sebuah grup yang terdapat target variabel kategori.

5. Clustering

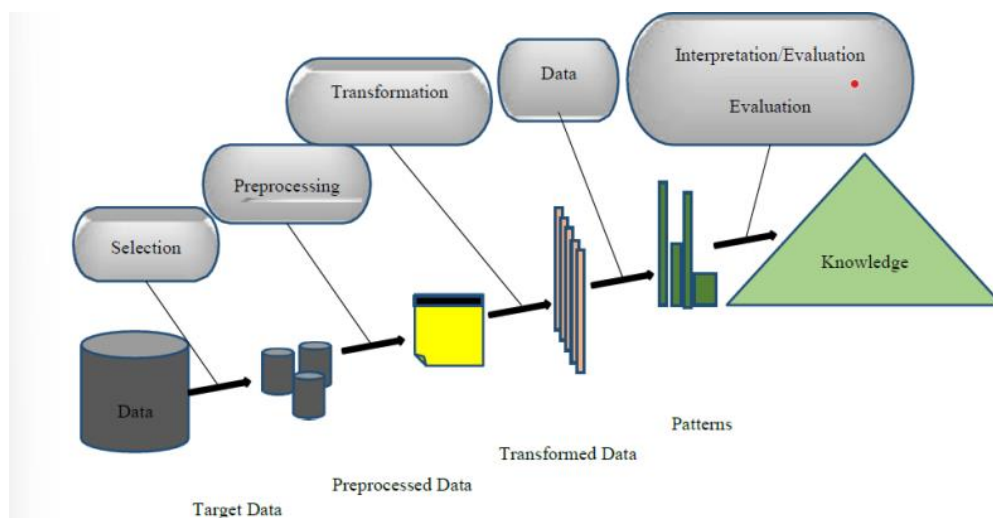
Clustering adalah sebuah proses pengelompokan data atau objek yang berdasarkan kemiripan diantara objek tersebut, Objek yang dikelompokkan kedalam satu kluster memiliki karakteristik yang serupa, sedangkan objek yang di kluster lain berbeda secara singtifikan.

6. Asosiasi

Asosiasi adalah teknik untuk menemukan keterkaitan atau hubungan antara atribut yang muncul secara bersamaan dalam suatu transaksi. Dalam konteks bisnis, teknik ini juga dikenal sebagai market basket analysis atau analisis keranjang belanja.

2.1.3 Knowledge Discovery in Database (KDD)

Data mining merupakan salah satu dari rangkaian Knowledge Discovery in Database (KDD). KDD berkaitan juga dengan proses integrasi teknik untuk menemukan pengetahuan dari data-data yang dikumpulkan. KDD juga mencakup interpretasi dan visualisasi pola yang ditemukan dalam data. (Tachi & Andri, 2021) Berikut 5 tahapan data mining dapat dilihat pada gambar 2.1



Gambar 2.1 Tahapan *Knowledge Discovery in Database (KDD)*

1. Selection

Data yang akan digunakan untuk seleksi data dan atribut yang digunakan untuk proses selanjutnya. Seleksi data dari sekumpulan data operasional dilakukan

sebelum tahapan penggalian informasi. Data hasil dari seleksi akan digunakan untuk proses data mining, dan disimpan dalam berkas yang terpisah dari data operasional.

2. Pre-processing (Cleaning)

Tahapan cleaning merupakan data hasil dari seleksi data yang akan dilakukan pemrosesan pendahuluan dan akan melakukan pembersihan data. Sebelum proses data mining dilakukan cleaning dengan tujuan untuk membuat duplikasi data, atau menghilangkan data yang tidak konsisten.

3. Transformation

Merupakan proses transformasi pada data yang telah dipilih untuk diubah menjadi bentuk data yang sesuai untuk diolah nantinya

4. Data mining

Data mining adalah proses mencari suatu informasi atau pola yang menarik pada data yang telah dipilih menggunakan teknik atau metode tertentu. Pemilihan metode atau algoritma sangatlah penting karena sangat berpengaruh pada tujuan dan proses dari KDD secara keseluruhan.

5. Interpretation (evaluation)

Pada tahap ini dilakukan identifikasi pola-pola yang dihasilkan dari data mining. Pola informasi yang dihasilkan perlu ditampilkan kedalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan

.

2.1.4 Manfaat Data Mining

Penelitian ini akan menggunakan algoritma apriori, dalam memanfaatkan teknik data transaksi penjualan yang didapatkan dari barang yang dibeli oleh konsumen saat berbelanja. Data tersebut digunakan untuk mengambil keputusan dalam pembelian suatu barang yang sangat diminati oleh konsumen dan kurang diminati oleh konsumen. Untuk memperbanyak stok barang yang paling diminati oleh konsumen dan mencukupi kebutuhan konsumen, kemudian mengurangi stok barang yang kurang laku dipasaran dan data tersebut juga dapat digunakan dalam menentukan peletakan barang sesuai permintaan konsumen. Data yang dihasilkan dapat menjadi informasi yang berguna untuk membentuk pola penjualan yang efektif. Teknik analisis kerang pasar merupakan teknik yang beradaptasi pada ilmu data mining. Teknik ini digunakan untuk mencari strategi penjualan barang melalui proses pencarian asosiasi antar item data dari suatu transaksi pembelian barang dari setiap konsumen, kemudian dicari suatu hubungan antar item-item yang dijual dan akan dibeli oleh konsumen. (Daniel Purba et al., n.d.)

2.2 Teori Khusus

2.2.1 Algoritma apriori

Menurut (Qoni'ah & Priandika, 2020) Algoritma apriori merupakan suatu algoritma dasar yang diusulkan oleh Agrawal & Srikant pada tahun 1994 untuk penentuan frequent itemset untuk aturan asosiasi boolean. Algoritma apriori termasuk jenis aturan yang terdapat pada data mining. Aturan yang menyatakan asosiasi atau association rule mining merupakan salah satu teknik data mining

untuk menemukan aturan suatu kombinasi antara item. Salah satu tahap analisis asosiasi pola frekuensi tinggi (frequent pattern mining). Penting atau tidak nya suatu asosiasi dapat diketahui dengan dua tolak ukur, yaitu support dan confidence, Nilai penunjang (support) adalah persentase dari kombinasi itemset tersebut dalam database sedangkan nilai kepastian (confidence) adalah kuatnya hubungan antara item dalam aturan asosiasi.

Analisis pola frekuensi tertinggi dengan algoritma apriori adalah tahap mencari kombinasi item yang memenuhi syarat minimum dari nilai support sebuah item diperoleh dengan menggunakan rumus berikut:

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A} \times 100}{\text{Total transaksi}}$$

Untuk mencari nilai support dari 2 item diperoleh dengan menggunakan rumus:

$$\text{Support (A,B)} = P(A \cap B)$$

$$\text{Support (A,B)} = \frac{\sum \text{Transaksi mengandung A dan B} \times 100}{\sum \text{transaksi}}$$

Pembentukan suatu aturan asosiasi adalah aturan setelah semua pola frekuensi tertinggi ditemukan, kemudian dicari aturan asosiasi yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiasi $A \rightarrow B$ nilai confidence dari aturan $A \rightarrow B$ diperoleh dengan rumus berikut:

$$\text{Confidence} = P(B | A) = \frac{\sum \text{transaksi mengandung A dan B} \times 100}{\sum \text{Transaksi mengandung A}}$$

2.2.2 Tahapan algoritma apriori

Dalam melakukan tahapan algoritma apriori ada 9 langkah :

1. Transformasi data dalam bentuk tabel tabular.
2. Menentukan nilai minimal support dan minimal confidence.
3. Pembentukan kandidat 1-itemset dan kandidat 2-itemset.
4. Pemangkasan atau pembuangan itemset yang memiliki nilai < minimum support (yang diterima adalah frekuensi itemset $\geq$ minimum support).
5. Untuk mencari nilai support diperoleh dengan rumus yang digunakan untuk suatu kombinasi 1-itemset dan kombinasi 2-itemset .

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A} \times 100}{\text{Total transaksi}}$$

Rumus 2.1 untuk mencari kombinasi 1-itemset

$$\text{Support (A,B)} = P \rho = (A \cap B)$$

$$\text{Support (A,B)} = \frac{\sum \text{Transaksi mengandung A dan B} \times 100}{\sum \text{transaksi}}$$

Rumus 2.2 untuk mencari kombinasi 2-itemset

$$\text{Support (A,B,C)} = \frac{\sum \text{Transaksi mengandung A,B,C} \times 100}{\sum \text{transaksi}}$$

Rumus 2.3 Untuk mencari kombinasi 3-itemset

6. Untuk mencari nilai confidence diperoleh dengan rumus digunakan untuk sebuah kombinasi 3-itemset .

$$Confidence = (A, B \Rightarrow C) = \frac{\sum \text{Transaksi mengandung } A, B, C}{\sum \text{Transaksi mengandung } A, B}$$

Rumus 2.4 Untuk mencari *confidence*

7. Selanjutnya akan menentukan nilai minimum confidence, lalu akan melakukan pemangkasan itemset yang kurang nilai minimum confidence (yang diterima adalah nilai confidence $\geq$ minimum confidence).
8. Setelah mendapatkan hasil di antara nilai support dan confidence pilih hasil yang paling besar.
9. Hasil yang paling besar merupakan rule yang dipakai dalam aturan asosiasi final.

2.2.2 Software tanagra

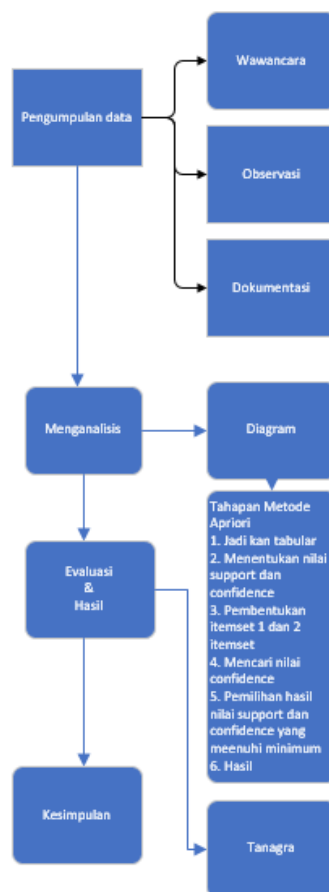
Software data mining bebas untuk tujuan akademik dan penelitian ini mengusulkan beberapa metode data mining dari analisis eksplorasi data, pembelajaran statistik, pembelajaran mesin dan bagian database. Tanagra ialah memberikan penelitian agar mudah dikerjakan pada perangkat lunak data mining. (Rangian & Fitriana, 2022)

Tanagra merupakan perangkat lunak data mining yang memiliki banyak fitur untuk menganalisis data, seperti analisis data eksploratif, statistik, dan mengelolahan database. Tidak seperti perangkat lunak lainnya. Tanagra juga bersifat *open source*, artinya siapapun bisa mengakses. (Listanto & Meisella

Kristania, 2022)

2.3 Kerangka pemikiran

Pelitian ini akan dilakukan sesuai dengan prosedur analisis data seperti yang sudah ada pada proses atau langkah data mining, Namun penelitian ingin menuliskan bagaimana cara kerangka pemikiran dari penelitian ini bisa mempermudah dalam membangun struktur penelitian dengan baik. Kerangka penelitian yang dibuat oleh peneliti.



Gambar 2. 2 Kerangka pemikiran

2.4 Penelitian terdahulu

Melihat masalah dan judul penelitian akan diteliti, maka diperlukan sebuah penerapan tentang penelitian terdahulu menggunakan fenomena yang sama dalam sudut pandang yang berbeda-beda sehingga diharapkan dapat memperluas pengetahuan, secara ringkas, penelitian terdahulu yang mendukung penelitian yang akan dilakukan dapat dilihat dibawa ini:

No	Nama peneliti	Judul peneliti	Hasil
1.	(Nst et al., n.d. 2021)	Impelementasi data mining algoritma apriori untuk Meningkatkan penjualan	Dapat ditemukan menggunakan Algoritma apriori, dengan melihat produk yang memenuhi minimal support 2,4% dan minimal confidence 50%.
2.	(Erfina et al., 2020)	Penerapan metode data mining terhadap data transaksi penjualan menggunakan algoritma apriori (studi kasus : Toko fasentro fancy)	Dari hasil penelitian ini Artinya, untuk menentukan himpunan unsur, ditemukan 4 aturan konjungsi, yaitu pranaya (P24TCV) → Roberto Cavalli dengan keyakinan 83,33%, Levato → Roberto Cavalli dengan

			keyakinan 83,33%, Pranaya (P24TCV) → Toyobo Fudo → Roberto Cavalli dengan keyakinan 85,71%, Roberto Cavalli → Pranaya (P24TCV) → Toyobo Fudo dengan keyakinan 80,00%.
3.	(Aulia Miranda et al., 2022)	Impelementasi assosian rule dalam menganalisis data penjualan sheshop dengan menggunakan algoritma apriori	Berdasarkan dapat dilihat dari 6 aturan terbaik dengan nilai <i>confidence</i> 58% sampai dengan 75% bedasarkan 6 association rule yang diperoleh menghasilkan 2 produk yang sering dibeli dibeli secara bersamaan yaitu Al quran dan tasbih dengan nilai <i>confidence</i> 75% aturan kombinasi telah ditemukan menggunakan association rule dan telah

			diuji dengan menggunakan software rapid miner.
4	(Devita Sari & Khoiriah, 2022)	Penerapan Metode Asosiasi Pada Toko Afifa Dengan Algoritma Apriori	Dengan mengimplementasikan teknik data mining metode asosiasi algoritma apriori didapatkan hasil aturan asosiasi berupa rules menggunakan perpaduan frequent itemset memperoleh nilai support tertinggi adalah 50% dan nilai confidence 88,9% dengan kombinasi produk Sarden ABC dan Aqua dengan nilai Lift Ratio 2,78 menandakan korelasi bersifat positif (kuat).
5.	(Pare et al., 2023)	Penerapan data mining pada transaksi penjualan barang menggunakan metode apriori (studi kasus :	Penerapan data mining pada transaksi penjualan menggunakan algoritma apriori apriori sangat membantu dalam

		Toko BE - MART	mengetahui antara frekuensi item yang ada dengan hasil yang didapatkan yaitu dimana jika konsumen yang membeli softex akan membeli skincare badan/muka menghasilkan nilai support 0.011% dan confidence 1%.
6.	(Fadila Matondang et al., 2022)	Penerapan data mining menggunakan algoritma apriori dalam menentukan pola penjualan elektronik	Dalam tahap pengujian sistem data mining algoritma apriori dapat dilakukan dengan melakukan pengimputan data elektronik, data transaksi, minimum Minimum support dan minimum confidence untuk mendapatkan pola penjualan dan mendapatkan hasil rekomendasi di galaxy

			electronic
7.	(Putri & Sitohang, 2023)	ANALISIS POLA PEMBELIAN KONSUMEN MENGGUNAKAN ALGORITMA APRIORI	Dari hasil penelitian ini barang yang memiliki keterkaitan yaitu Atasan Wanita dan Atasan Laki-Laki dengan nilai support 27,86 % dan nilai confidence 68,00 %. Informasi ini juga dapat digunakan untuk mengetahui barang apa saja yang banyak diminati dan kurang diminati oleh konsumen dalam menentukan pengadaan stok barang apa yang harus diprioritaskan diwaktu mendatang.
8	(Informatika et al., n.d.)	Analisis data mining terhadap food dengan metode apriori pada kopsyahira	Dengan menggunakan metode apriori dengan menggunakan Tanagra dapat membantu dalam mengetahui produk yang

			paling banyak terjual, yang pada akhirnya dapat digunakan untuk Menyusun strategi penjualan dan dapat dilakukan promosi dengan melakukan kombinasi penjualan antara produk yang laku dan yang kurang laku dijual.
9	(Nauw et al., 2023)	Pemanfaatan data mining dalam memprediksi transaksi penjualan menggunakan algoritma apriori	Dapat diketahui produk mana yang sering dibeli secara bersamaan dari catatan transaksi penjualan, yang dapat dijadikan untuk mempertimbangkan taktik penjualan seperti tata letak barang.
10	(Ovilianda & Ginting, 2021)	Penerapan Data Mining Korelasi Penjualan Spare Part Mobil Menggunakan	Dari pengujian data yang telah dilakukan, penerapanaturan asosiasi data mining dengan

		Metode Apriori Algoritma	menggunakan algoritma apriori dapat digunakan untuk mencari korelasi penjualan sparepart mobil menggunakan algoritma apriori (studi kasus: CV. Citra Kencana Mobil). Dengan menentukan rule minimal 3 item set untuk penerapan algoritma k-apriori pada korelasi penjualan sparepart mobil ditemukan maksimum support adalah sebesar 9% yaitu sebanyak 81 rule dengan support dan confidence yang beragam. Dari hasil pengujian penjualan sparepart mobil dengan jumlah data 589 data, ditemukan 81 rule yang terbentuk dan diperoleh Best Rule
--	--	--	--

			tertinggi dan nilai minimum support 1% dan nilai confidence 11% Jika jenis Mobil adalah Avanza/Xenia dan Merk adalah Toyota maka sparepart yang digunakan adalah Filter Udara. Dengan sparepart pendukung di dalam database sebesar 1% dan sparepart kepastian sebesar 11%.
11	(Sudarsono et al., n.d. 2)	Analisis penjualan perlengkapan olahraga pada HS sport menggunakan metode apriori	Disimpulkan jika membeli training club maka akan membeli body protector dengan Nilai <i>confidence</i> 73% jika membeli body protector maka membeli hand protector dengan nilai <i>confidence</i> 79%.
12	(Ardiansyah, 2022)	Impelementasi data mining dengan metode	Dalam pembahasan hasil proses apriori data proses

		apriori untuk analisis pola penjualan barang fashion	menggunakan data tabel transaksi penjualan online shop dan proses menggunakan metode apriori yang dapat menghasilkan nilai minimum <i>support</i> dan minimum <i>confidence</i>
13	(Muharni & Andriyanto, 2024)	Penentuan pola penjualan menggunakan algoritma apriori	Beberapa kesimpulan yang bisa ditarik dari penelitian diatas yaitu skenario penerapan metode asosiasi untuk analisis pola data transaksi penjualan menghasilkan suatu pola aturan terkaitan (asosiasi) baru dari masing-masing item yang ada dataset dengan melalui proses metode asosiasi dalam mendapatkan berbagai kemungkinan aturan keterkaitan item terhadap

			data transaksi penjualan
14	(Setiawan, 2024)	Impelemntasi data mining menggunakan algoritma apriori terhadap pola pembelian konsumen di marketplace shopee jaktimstore	Berdasarkan hasil perhitungan data mining menggunakan algoritma Apriori, data transaksi penjualan dengan batas minimum support sebesar 20% dan batas minimum confidence sebesar 80%.
15	(Osman et al., 2021)	Association rule mining for identification of port state control patterns in Malaysian ports	Penelitian ini menganalisis pola inspeksi kapal di lima pelabuhan Malaysia menggunakan algoritma Apriori. Hasilnya menunjukkan bahwa kapal berbendera tertentu dan kapal berisiko rendah (LRS) cenderung tidak memiliki kekurangan. Temuan ini bisa membantu pelabuhan meningkatkan strategi inspeksi dan keselamatan maritim.

			Namun, masih perlu dikaji lebih lanjut apakah inspeksi di Malaysia terlalu longgar dan apakah kapal yang masuk benar-benar layak laut. Penelitian lanjutan disarankan untuk mempertimbangkan faktor lain seperti usia kapal dan klasifikasi.
--	--	--	---