

BAB II

LANDASAN TEORI

2.1. Pengertian Data Mining

Data mining adalah proses eksplorasi dan analisis data dalam jumlah besar untuk menemukan pola-pola atau informasi yang tersembunyi, yang kemudian dapat digunakan untuk mendukung pengambilan keputusan yang lebih baik dan efisien. Menurut [1] data mining menggabungkan berbagai teknik dari bidang statistik, pembelajaran mesin, kecerdasan buatan, dan matematika untuk mengekstraksi pengetahuan dari kumpulan data besar. Dalam konteks pengelolaan stok barang, data mining berperan penting dalam mengidentifikasi pola-pola transaksi dan penggunaan barang yang dapat digunakan untuk merencanakan pengadaan dan distribusi barang secara lebih efektif dan efisien.

Data mining adalah proses analisis data dalam jumlah besar untuk menemukan pola atau informasi yang tersembunyi dan dapat digunakan untuk mendukung pengambilan keputusan yang lebih baik dan efisien. Dalam konteks pengelolaan stok barang, data mining digunakan untuk menganalisis data transaksi, pola pembelian, dan kebutuhan persediaan barang. Menurut [2], data mining membantu dalam mengidentifikasi pola-pola tersembunyi yang mungkin tidak dapat ditemukan dengan metode analisis tradisional, sehingga memungkinkan perusahaan untuk mengoptimalkan pengelolaan persediaan barang.

Data mining didefinisikan sebagai proses eksplorasi data dalam jumlah besar untuk menemukan pola tersembunyi yang dapat digunakan untuk mendukung pengambilan keputusan. Menurut [3], *Knowledge Discovery in*

Databases (KDD) adalah proses keseluruhan untuk menemukan pola atau pengetahuan yang bermanfaat, dapat dipahami, dan valid dari data besar. Pendekatan ini melibatkan berbagai tahapan, termasuk pemilihan data, pembersihan, transformasi, penambangan data, dan interpretasi hasil. Data mining dalam konteks PT Telkom Indonesia berfokus pada pengelolaan stok barang dan distribusinya. Proses ini menggali pola asosiasi antara barang-barang yang sering digunakan bersama dalam transaksi pengambilan barang di gudang. Dengan menggunakan data mining, perusahaan dapat memprediksi kebutuhan barang, mengurangi risiko *overstocking* dan *stockout*, serta meningkatkan efisiensi distribusi barang.

Data mining sering kali digunakan untuk menemukan **hubungan yang tidak terduga** antar elemen data, yang sebelumnya mungkin tidak terlihat atau dianggap tidak relevan. Dengan penerapan teknik ini, perusahaan dapat menggali **pengetahuan tersembunyi** dalam database mereka, yang berguna untuk meningkatkan proses bisnis dan operasional perusahaan. Data mining juga dapat menghasilkan informasi yang lebih mendalam dan dapat digunakan untuk memperbaiki keputusan terkait pengelolaan persediaan, seperti memprediksi permintaan barang atau mengenali pola penggunaan barang yang kurang diminati.

Salah satu metode utama dalam data mining adalah **algoritma asosiasi**, yang digunakan untuk menemukan hubungan antar item atau transaksi dalam data. Menurut [4], algoritma **Apriori** adalah salah satu teknik yang paling banyak digunakan untuk analisis asosiasi. Algoritma ini membantu untuk menemukan **itemset yang sering muncul** dalam data transaksi, yang kemudian dapat menghasilkan aturan asosiasi yang menggambarkan hubungan antara item yang

sering dibeli bersama. Misalnya, dalam pengelolaan stok barang di gudang, algoritma ini dapat digunakan untuk menemukan barang-barang yang sering diambil bersama dalam transaksi pengambilan barang di gudang.

Dengan **menemukan pola asosiasi** ini, perusahaan dapat meningkatkan perencanaan persediaan dan mengurangi risiko **kelebihan stok** (*overstocking*) atau **kekurangan stok** (*stockout*). Penerapan algoritma Apriori dalam pengelolaan stok barang dapat memberikan wawasan yang lebih jelas mengenai pola penggunaan barang, yang penting untuk memastikan ketersediaan barang yang optimal di gudang.

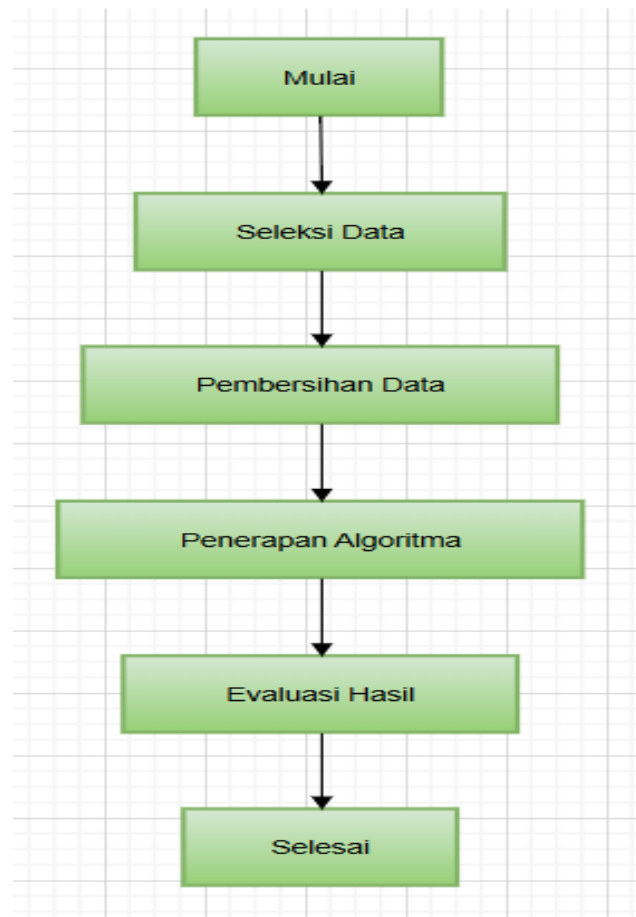
2.2. Proses Data Mining dalam Pengelolaan Stok Barang

Proses data mining untuk analisis stok barang terdiri dari beberapa tahapan yang bertujuan untuk meningkatkan efisiensi dan akurasi dalam pengelolaan persediaan. Berdasarkan [5], *Knowledge Discovery in Databases (KDD)* didefinisikan sebagai proses untuk mengidentifikasi hubungan dan pola dalam dataset besar. Proses ini mencakup langkah-langkah seperti seleksi data, pembersihan data, transformasi data, penambangan data, dan interpretasi hasil. Data mining sendiri merupakan bagian inti dari proses KDD yang berfokus pada penggalian pola atau anomali dari data untuk menghasilkan informasi yang dapat dipahami oleh pengambil Keputusan

2.3. Kerangka Kerja Penelitian

Kerangka kerja penelitian ini mengacu pada tahapan dalam *Knowledge Discovery in Databases (KDD)*, yang mencakup beberapa tahapan: pengumpulan data, *preprocessing*, transformasi data, penerapan algoritma data mining, dan evaluasi hasil. [6] mengungkapkan bahwa proses KDD ini penting untuk

memastikan bahwa data yang dianalisis adalah berkualitas tinggi dan dapat menghasilkan temuan yang relevan aplikatif.



Gambar 2. 1. Kerangka Kerja Penelitian

1. Seleksi Data: Pada tahap seleksi data, data yang relevan akan dipilih untuk digunakan dalam analisis.
2. Pembersihan Data: Data yang telah dikumpulkan kemudian harus dibersihkan untuk memastikan konsistensi dan kualitasnya.
3. Penerapan Algoritma: Setelah data siap, algoritma seperti Apriori akan diterapkan untuk menganalisis data transaksi dan penggunaan barang di gudang. [7] menyatakan bahwa penggunaan teknik ini memungkinkan untuk

menemukan pola yang dapat digunakan untuk perencanaan pengadaan barang.

4. Evaluasi Hasil: Setelah algoritma diterapkan, hasil dari analisis harus dievaluasi dengan menggunakan metrik seperti support, confidence, dan lift untuk menilai kualitas pola yang ditemukan.

2.4. Algoritma Apriori

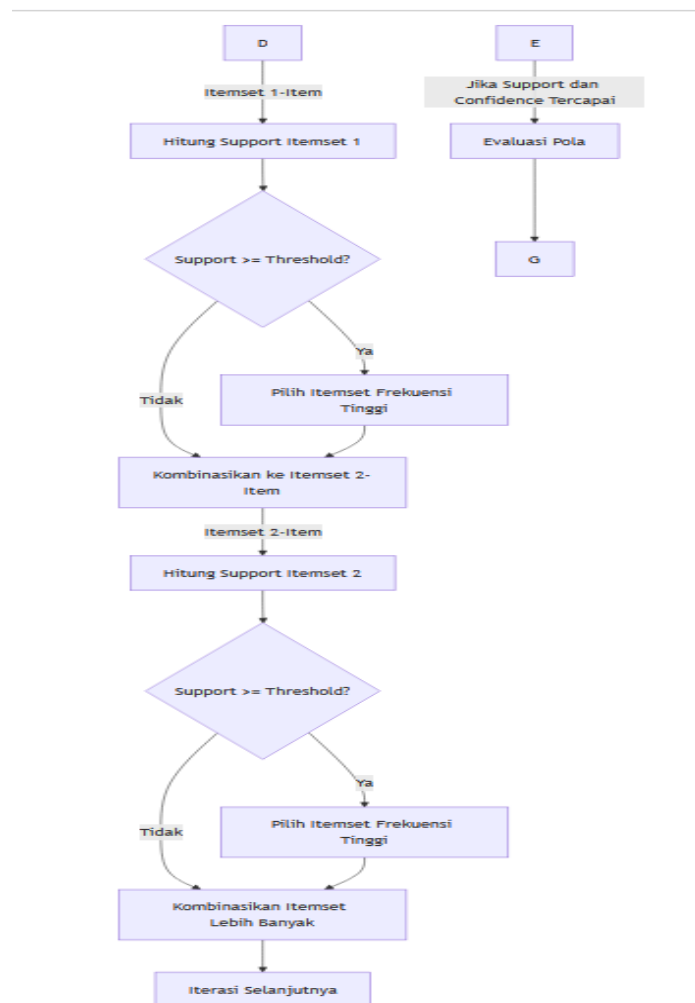
Algoritma Apriori merupakan salah satu algoritma yang paling banyak digunakan dalam data mining untuk mencari pola asosiasi dalam dataset yang besar.

Pada **algoritma Apriori**, dua parameter utama yang digunakan untuk menentukan relevansi suatu itemset adalah **minimum support** dan **minimum confidence**. **Support** mengacu pada proporsi transaksi yang harus mencakup itemset tertentu agar dianggap signifikan dalam dataset, sementara **confidence** mengukur seberapa sering aturan asosiasi yang dihasilkan benar [8] [9]. Kedua parameter ini membantu dalam menyaring itemset dan aturan yang tidak relevan, sehingga hanya pola-pola yang kuat dan bermakna yang dipertimbangkan untuk analisis lebih lanjut.

Setelah parameter support dan confidence ditentukan, algoritma Apriori mulai mencari **frequent itemsets**, yaitu kombinasi item yang sering muncul bersama dalam transaksi. Proses ini dilakukan secara **iteratif**, dimulai dengan itemset tunggal dan kemudian menggabungkan item-item tersebut untuk menemukan kombinasi itemset yang lebih besar. Pada setiap iterasi, algoritma menghitung support untuk itemset yang dihasilkan dan menyaring itemset yang tidak memenuhi kriteria minimum support [10] [11]. Dengan demikian, algoritma

Apriori dapat mengidentifikasi aturan asosiasi yang menunjukkan hubungan antara produk-produk yang sering dibeli bersama, yang sangat berguna untuk strategi pemasaran dan pengelolaan inventaris [12] [13].

2.4.1. Langkah Kerja Algoritma Apriori



Gambar 2. 2. Langkah Kerja Algoritma Apriori

Gambar yang ditampilkan menunjukkan diagram alur (*flowchart*) yang menggambarkan langkah-langkah dalam algoritma Apriori. Algoritma ini digunakan untuk menganalisis pola asosiasi dalam dataset transaksi. Dalam diagram tersebut, terdapat beberapa tahapan yang saling terkait dan berulang, yang mencerminkan proses iteratif dari algoritma ini.

1. Input Dataset, di mana data transaksi dimasukkan untuk dianalisis. Setelah itu, langkah selanjutnya adalah *Menentukan Threshold Support* dan *Confidence*, yang berguna untuk menentukan seberapa kuat hubungan antar item dalam dataset, serta menetapkan parameter-parameter yang akan digunakan dalam proses analisis.
2. Pencarian Itemset Frekuensi Tinggi, algoritma akan mencari itemset yang muncul paling sering dalam dataset dan memenuhi kriteria minimum yang telah ditentukan sebelumnya. Proses ini dilakukan secara bertahap, mulai dari pencarian Itemset 1-Item, yang kemudian dikembangkan menjadi Itemset 2-Item dan seterusnya. Setiap itemset yang memenuhi threshold support akan dipilih untuk langkah selanjutnya.
3. Formasi Aturan (*Rule Formation*), di mana aturan-aturan asosiasi dibuat berdasarkan kombinasi itemset yang memenuhi kriteria support dan confidence. Proses ini menghasilkan aturan yang menunjukkan hubungan antar item yang sering dibeli bersamaan dalam transaksi.
4. Evaluasi Pola dilakukan untuk mengevaluasi kualitas dan relevansi aturan tersebut. Evaluasi dilakukan dengan menggunakan metrik tambahan seperti lift, yang mengukur kekuatan hubungan antar item. Aturan yang telah dievaluasi kemudian disaring dan dipilih berdasarkan kriteria tertentu.

Secara keseluruhan, gambar ini memberikan gambaran yang jelas mengenai tahapan-tahapan dalam algoritma Apriori, dimulai dari input data hingga evaluasi hasil aturan yang ditemukan. Diagram ini menggambarkan alur kerja algoritma secara terstruktur dan mudah dipahami.

2.4.2. Formula Matematika yang Digunakan

Dalam algoritma Apriori, terdapat tiga konsep utama yang digunakan untuk mengukur kualitas dan kekuatan dari aturan asosiasi yang ditemukan, yaitu *support*, *confidence*, dan *lift*.

1. **Support:** *Support* mengukur seberapa sering itemset muncul dalam dataset. *Support* untuk itemset A dihitung dengan rumus berikut:

$$\text{Support}(A) = \frac{\text{Jumlah transaksi yang mengandung itemset A}}{\text{Jumlah total transaksi}}$$

Support yang tinggi menunjukkan bahwa itemset tersebut sering muncul dalam data, yang mengindikasikan bahwa itemset tersebut penting dalam analisis.

2. **Confidence:** *Confidence* mengukur probabilitas bahwa jika item A diambil, maka item B juga akan diambil. Rumus untuk menghitung *confidence* adalah:

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)}$$

***Confidence* yang tinggi menunjukkan bahwa ada kemungkinan besar bahwa ketika item A diambil, item B juga akan diambil.** Ini menggambarkan hubungan yang kuat antara dua item dalam aturan asosiasi.

3. **Lift:** *Lift* mengukur kekuatan hubungan antara dua item yang diambil bersama dibandingkan dengan yang diharapkan jika mereka diambil secara *independen*. *Lift* dihitung dengan rumus berikut:

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Support}(A \rightarrow B)}{\text{Support}(B)}$$

***Lift* yang lebih besar dari 1 menunjukkan bahwa hubungan antara A dan B lebih kuat dibandingkan dengan apa yang diharapkan jika keduanya muncul secara independen.** Metrik ini sangat berguna untuk menilai seberapa signifikan asosiasi antara dua barang.

2.5. Penerapan Algoritma Apriori dalam Pengelolaan Gudang

Penerapan algoritma Apriori dalam pengelolaan gudang bertujuan untuk meningkatkan efisiensi pengadaan dan distribusi barang. Langkah pertama dalam penerapan algoritma Apriori adalah pengumpulan data transaksi, yang berisi informasi mengenai barang-barang yang diambil dalam setiap transaksi. Data ini biasanya berbentuk tabel yang mencatat barang-barang yang dipilih dalam setiap pengambilan.

Setelah data terkumpul, tahap selanjutnya adalah pembersihan data, yang bertujuan untuk menghilangkan data yang tidak lengkap atau tidak relevan. Setelah data siap, algoritma Apriori diterapkan untuk *menemukan frequent itemsets*. Misalnya, jika barang A dan B sering diambil bersama dalam transaksi, algoritma Apriori akan mengidentifikasi {A, B} sebagai frequent itemset. Setelah frequent itemsets ditemukan, aturan asosiasi dapat dibentuk dengan menghitung *confidence* dari setiap aturan yang dihasilkan. *Confidence* mengukur seberapa sering aturan tersebut benar dalam transaksi yang berbeda [14] [15].

Contoh penerapan: Dalam konteks pengelolaan gudang PT Telkom Indonesia, apabila algoritma Apriori menunjukkan bahwa barang A dan B sering diambil bersama, perusahaan dapat menempatkan kedua barang tersebut berdekatan di gudang. Penempatan barang-barang yang sering digunakan bersama di lokasi yang lebih dekat dapat mempercepat proses pengambilan barang dan

meningkatkan efisiensi operasional gudang. Dengan demikian, hasil dari penerapan algoritma Apriori tidak hanya memberikan wawasan mengenai pola pembelian atau penggunaan barang, tetapi juga memberikan dasar untuk pengambilan keputusan strategis dalam pengelolaan stok barang di gudang [16] [17].

2.6. Alat Bantu Pemrograman dan Tools Pendukung

Dalam implementasi algoritma Apriori untuk analisis data transaksi pengambilan barang di gudang PT Telkom Indonesia, beberapa alat bantu pemrograman dan tools pendukung digunakan untuk mempermudah dan mempercepat proses data mining. Alat-alat ini meliputi:

1. RapidMiner: RapidMiner adalah platform perangkat lunak yang menawarkan berbagai alat untuk data mining dan analisis statistik. Dengan menggunakan RapidMiner, peneliti dapat menerapkan algoritma Apriori dalam waktu yang relatif singkat dan mendapatkan hasil yang lebih cepat. Platform ini juga menyediakan fitur visualisasi untuk membantu pengguna dalam memahami dan mengevaluasi hasil analisis.



Gambar 2. 3. Aplikasi RapidMiner

2. Microsoft Excel adalah aplikasi pengolah angka dari Microsoft yang berfungsi untuk mengelola, menganalisis, dan menyajikan data dalam bentuk tabel, grafik, maupun diagram. Aplikasi ini dilengkapi dengan berbagai fitur

seperti rumus, fungsi, pivot table, serta chart yang memudahkan pengguna dalam melakukan perhitungan, pengolahan data, hingga visualisasi informasi secara cepat dan akurat. Excel banyak digunakan di berbagai bidang, mulai dari pendidikan, bisnis, akuntansi, hingga penelitian, karena kemampuannya dalam mengolah data dalam jumlah besar dengan fleksibilitas tinggi.



Gambar 2. 4. Aplikasi Microsoft Excel

2.7. Sistem Informasi dalam Pengelolaan Stok Barang

Penggunaan sistem informasi berbasis komputer dalam pengelolaan stok barang juga sangat penting, karena sistem ini dapat mempercepat proses pencatatan transaksi, mengurangi kesalahan manual, dan meningkatkan akurasi dalam perhitungan stok barang. Sistem informasi yang terintegrasi juga membantu dalam memantau pergerakan barang, mengelola pengadaan, dan menyusun laporan yang dapat digunakan untuk analisis lebih lanjut.

2.8. Metode Penelitian

2.8.1. Penelitian Relevan Terdahulu

Dibawah ini adalah referensi yang penulis gunakan untuk mendukung dan sebagai landasan pembuatan proposal yaitu sebagai berikut:

Tabel 2. 1. Penelitian Relevan Terdahulu

Referensi Penelitian	1
Judul	Implementasi Data Mining Menggunakan Algoritma Apriori dalam Menentukan Persediaan Barang (Studi Kasus: Toko Swapen Jaya Manokwari)
Nama Penulis	Jubita Dongga, Ana, Novalin Koru, Gabriella Lante
Tahun	2023
Hasil Penelitian	Penelitian ini berhasil mengidentifikasi pola pembelian barang yang sering dibeli bersamaan di Toko Swapen Jaya Manokwari menggunakan algoritma Apriori. Ditemukan bahwa kombinasi item dengan support dan confidence tertentu dapat membantu dalam pengambilan keputusan terkait persediaan barang.
Referensi Penelitian	2
Judul	Analisis Data Mining Sistem Inventory Menggunakan Algoritma Apriori

Nama Penulis	I. Riadi, H. Herman, F. Fitriah, S. Suprihatin, A. Muis, M. Yunus
Tahun	2023
Hasil Penelitian	Penelitian ini menganalisis data transaksi penjualan untuk menemukan pola pembelian yang dapat digunakan dalam pengelolaan persediaan. Dengan menggunakan algoritma Apriori, ditemukan beberapa aturan asosiasi yang dapat membantu dalam pengambilan keputusan terkait stok barang.
Referensi Penelitian	3
Judul	Implementasi Data Mining Menggunakan Algoritma Apriori dalam Menentukan Persediaan Barang: Studi Kasus: Toko Sinar Harahap
Nama Penulis	Putri M. S. Tarigan, Jaya T. Hardinata, Hendry Qurniawan, M. Safii, Riki Winanjaya2
Tahun	2020
Hasil Penelitian	Penelitian ini menerapkan algoritma Apriori untuk menganalisis data penjualan di Toko Sinar Harahap. Hasilnya menunjukkan adanya pola pembelian yang dapat digunakan untuk menentukan persediaan barang yang optimal.
Referensi Penelitian	4

Judul	Implementasi Data Mining Algoritma Apriori pada Penjualan Obat di Klinik
Nama Penulis	W. Delrinata, F. B. Siahaan
Tahun	2020
Hasil Penelitian	Penelitian ini berhasil menemukan pola pembelian obat yang sering dibeli bersamaan di sebuah klinik menggunakan algoritma Apriori. Informasi ini bermanfaat dalam mengelola persediaan obat dan meningkatkan efisiensi pelayanan kepada pasien.
Referensi Penelitian	5
Judul	Optimalisasi Strategi Pemenuhan Persediaan Stok Barang Menggunakan Algoritma Apriori
Nama Penulis	Verawati, M. Wishnu
Tahun	2020
Hasil Penelitian	Penelitian ini menerapkan algoritma Apriori untuk menganalisis data penjualan dan menentukan strategi optimal dalam pemenuhan persediaan stok barang. Hasilnya menunjukkan bahwa penggunaan algoritma Apriori dapat meningkatkan efisiensi pengelolaan persediaan.