

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Data Mining

Data mining adalah proses ekstraksi informasi yang bermanfaat dari data besar dengan menggunakan teknik-teknik matematika, statistika, dan kecerdasan buatan. Informasi yang dihasilkan dari proses data mining dapat digunakan untuk membantu pengambilan Keputusan dan prediksi di berbagai bidang, seperti bisnis, ilmu pengetahuan, teknologi, Kesehatan, dan lain – lain. Definisi data mining telah berkembang seiring dengan perkembangan teknologi dan penelitian. Data mining merupakan sebuah proses penggalian data atau penyaringan data dengan memanfaatkan Kumpulan data yang memiliki ukuran cukup besar melalui serangkaian proses untuk mendapatkan informasi(Asyuti & Setyawan, 2023).

Data mining merupakan suatu metode yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. Data mining juga merupakan metode pengolahan data berskala besar sehingga data mining memiliki peran yang sangat penting dalam bidang keuangan, ilmu, teknologi, industri dan cuaca.

Data mining sangat dibutuhkan terutama dalam mengelola data yang sangat besar untuk memudahkan aktivitas suatu transaksi. Alasan mengapa data mining sangat menarik perhatian industri informasi bahkan sampai ke toko – toko dalam beberapa tahun sebelumnya karena data mining melakukan kegiatan ekstraksi atau lebih dikenal menambang pengetahuan dari data yang berjumlah besar, kemudian

informasi ini yang nantinya sangat berguna untuk pengembangan. (Erwansyah et al., 2021)

2.1.2 Aturan Asosiasi (*Association Rule*)

Aturan asosiasi adalah cara penambangan data yang digunakan untuk mendapatkan banyaknya pola kombinasi items, atau kumpulan items yang banyak muncul pada sekumpulan data tertentu. Aturan asosiasi bertujuan untuk menemukan pola yang sering muncul pada beberapa items sehingga dapat direkomendasikan pada suatu Perusahaan dalam mendukung bisnis tertentu. Asosiasi juga merupakan proses penambangan data yang digunakan dalam mengidentifikasi semua aturan asosiasi yang telah memenuhi kriteria dukungan dan nilai kepercayaan dalam database. Dimulai dengan mencari kombinasi items yang sering muncul dalam Kumpulan dataset yang harus memenuhi nilai dukungan.(Octavia et al., 2023).

Association rule memiliki dua tahap proses yaitu:

1. Temukan semua *frequent itemsets*: anggota *frequent itemsets* setidaknya memiliki jumlah dalam *database* sama dengan *minimum support* yang telah ditentukan di awal.
2. Temukan semua aturan asosiasi yang kuat dari *frequent itemsets* aturan asosiasi yang dihasilkan harus memenuhi *minimum support* dan *minimum confidence*.

Association Rule adalah teknik data mining untuk menemukan aturan assosiatif dengan kombinasi item. Asosiasi adalah aturan yang mendukung

pengolahan data mengidentifikasi i korelasi, dan mempertimbangkan urutan pengambilan materi sebagai suatu basis pembangkit aturan (Syefira Salsabila, 2022).

Association rule adalah implikasi dari $X \Rightarrow Y$, dimana $X \subset I$, $Y \subset I$, dan $X \cap Y = \emptyset$. Aturan $X \Rightarrow Y$ berada di dalam himpunan transaksi D dengan kepercayaan (*Confidence*) c, jika c% dan transaksi dalam D yang ada X terdapat juga Y. Aturan $X \Rightarrow Y$ memiliki dukungan (*support*) s di dalam set transaksi D, jika s% transaksi dalam D terdapat $X \cup Y$.

Support itemset (I) adalah jumlah transaksi T yang didalamnya terdapat (I). *Confidence* mengukur seberapa besar ketergantungan suatu *item* dengan *item* yang lainnya. *Frequent itemset* adalah *itemset* yang memiliki nilai support lebih atau sama besar dengan nilai *minimum support* yang ditentukan (Agrawal & Srikant, n.d.). Di misalkan suatu aturan $R : X \Rightarrow Y$, maka:

$$Support (X \Rightarrow Y) = Support (\{X\} \cup \{Y\})$$

$$Support (X \Rightarrow Y) = \frac{Support (\{X\} \cup \{Y\})}{Support (\{X\})}$$

Sebuah aturan (*rule*) biasanya terdiri dari dua bagian, yaitu kondisi (*condition*) dan hasil (*result*).

2.1.3 Algoritma Apriori

Algoritma apriori merupakan algoritma yang sangat terkenal untuk menentukan pola frekuensi tinggi. Algoritma Apriori ini termasuk jenis aturan asosiasi pada data mining. Pola frekuensi tinggi adalah pola-pola items didalam

suatu basis data yang memiliki frekuensi atau support. Pola frekuensi tinggi inilah yang dapat digunakan untuk Menyusun aturan assosiatif dan juga beberapa Teknik data mining lainnya(Orisky Sitra Arifah Destiyati & Eko Aribowo, 2015).

Algoritma apriori digunakan untuk menemukan *association rules* yang memenuhi batas nilai *Support* dan *Confidence*. Cara kerja apriori adalah dengan menganalisa kumpulan item yang diambil atau dipilih secara bersamaan pada beberapa transaksi. Penghitung kumpulan pola data yang muncul di dalam database melalui beberapa iterasi atau perulangan. Iterasi i menghitung semua kumpulan data i (kumpulan data yang mengandung elemen i) yang sering muncul. Setiap iterasi terdiri dari 2 langkah yaitu *candidate generation* (penentuan kandidat) dan *candidate counting and selection* (pemilihan dan penghitungan kandidat) (Fadhilah et al., 2020) . Penting atau tidaknya suatu asosiasi dapat diketahui dengan dua tolak ukur yaitu *Support* dan *Confidence*. *Support* (nilai penunjang) adalah persentase kombinasi item tersebut dalam database, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar *item* dalam aturan asosiasi(Kusrini & Emha Taufiq Luthfi, 2009).

Terminologi dasar dalam algoritma apriori ada beberapa istilah penting yang perlu dipahami:

1. *Frequent Itemset*

Frequent Itemset adalah sekumpulan item yang sering muncul bersamaan dalam suatu dataset transaksi. Untuk menentukan apakah sebuah *itemset* sering muncul, digunakan ukuran yang disebut *support*.

3. *Support*

Support mengukur seberapa sering suatu *item* atau *itemset* muncul dalam dataset. Dalam konteks transaksi, support dari suatu itemset X adalah persentase dari total transaksi yang mengandung *itemset* X.

Rumus *Support* untuk *itemset* X:

$$\text{Support}(X) = \frac{\text{Jumlah transaksi yang mengandung X}}{\text{Total Jumlah Transaksi}}$$

4. *Confidence*

Confidence mengukur seberapa sering item B muncul dalam transaksi yang juga mengandung item A. Dengan kata lain, *confidece* mengukur kekuatan aturan asosiasi $A \rightarrow B$ yang menunjukkan beberapa kali item B muncul ketika A ada dalam transaksi. Rumus *Confidence* untuk aturan $A \rightarrow B$:

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

5. *Lift*

Lift adalah ukuran seberapa besar pengaruh *item antecedent* (A) terhadap item *consequent* (B). Nilai *lift* di atas 1 menunjukkan bahwa item A dan B memiliki hubungan yang lebih kuat dari yang diharapkan.

Rumus *lift* untuk aturan $A \rightarrow B$:

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

Penerapan algoritma Apriori melibatkan beberapa langkah untuk menemukan *frequent itemsets* dan membentuk aturan asosiasi (*association rules*), diantaranya:

1. Menentukan Parameter minimum *Support* dan *Confidence*

Menentukan parameter sebelum menjalankan algoritma ada dua parameter penting, yaitu:

- *Minimum Support*: Nilai minimum persentase transaksi yang harus mengandung suatu *itemset* agar *itemset* tersebut dianggap *frequent*.
- *Minimum Confidence*: Nilai minimum persentase keyakinan yang harus dipenuhi oleh aturan asosiasi. Sebagai contoh bisa menetapkan minimum *Support* sebesar 50% dan minimum *Confidence* sebesar 70%.

2. Menemukan *Frequent Itemsets*

Melibatkan pencarian *itemsets* yang sering muncul dalam *dataset*. Apriori menggunakan metode iteratif yang dimulai dengan *itemset* tunggal dan kemudian membangun *itemset* yang lebih besar, menggunakan prinsip bahwa subset dari *frequent itemset* juga harus *frequent*.

- **Iterasi 1: Menghitung *Support* untuk item tunggal (1-itemset).** Algoritma memulai dengan menghitung *Support* untuk setiap item tunggal dalam *dataset*. Jika item tersebut memenuhi minimum *Support*, item tersebut akan disimpan sebagai *Frequent 1 – itemset*.
- **Iterasi 2: Menghasilkan 2-Itemsets.** Setelah *frequent 1-itemsets* ditemukan, algoritma menghasilkan kombinasi *itemsets* dari 2-*itemsets*

dan menghitung *support* untuk setiap kombinasi. *Itemsets* yang memenuhi minimum *support* akan dilanjutkan ke langkah berikutnya.

- **Iterasi N: Menghasilkan N-Itemsets.** Proses ini berlanjut dengan membentuk itemsets yang lebih besar (3-itemsets, 4-itemsets, dst.) sampai tidak ada lagi kombinasi yang memenuhi minimum *support*.

3. Membentuk Aturan Asosiasi (*Association Rule*)

Setelah *frequent itemsets* ditemukan, langkah selanjutnya adalah membentuk **aturan asosiasi**. Untuk setiap *frequent itemset*, algoritma membentuk aturan dalam bentuk $A \rightarrow B$, di mana A dan B adalah subset dari itemset. *Confidence* untuk setiap aturan dihitung, dan hanya aturan yang memenuhi minimum *confidence* yang dianggap valid.

2.1.4 Preprocessing Data

Preprocessing data adalah tahap awal dalam analisis data yang penting untuk meningkatkan kualitas dan keakuratan hasil model. Proses ini mencakup beberapa teknik yang digunakan untuk membersihkan, menyaring, dan menyiapkan data agar siap digunakan dalam algoritma machine learning atau analisis statistik. Proses ini meliputi beberapa tahapan penting, seperti penghapusan noise (noise removal), normalisasi data, tokenisasi, penghapusan stopwords, dan stemming. Semua tahapan ini bertujuan untuk mengurangi kompleksitas dan meningkatkan efisiensi dalam analisis data.

Preprocessing teks sangat berpengaruh pada hasil analisis sentimen dalam bahasa Indonesia. Proses ini mencakup transformasi huruf (case folding), penghapusan tanda baca, serta filter untuk menghapus kata-kata yang tidak

memiliki makna penting dalam analisis sentimen seperti stopwords. Penelitian tersebut menunjukkan bahwa dengan menerapkan teknik preprocessing seperti penghapusan noise dan penggunaan stemming serta penghapusan stopwords, akurasi model dapat meningkat secara signifikan (WILIAM FAJAR DICKA SUDIYONO et al., 2019).

Dalam penelitian lainnya, pengolahan data juga dilakukan dengan mengeliminasi variabel yang tidak relevan atau memiliki nilai yang tidak akurat, seperti pada studi yang dilakukan dengan data absensi mahasiswa yang menganalisis kedisiplinan mahasiswa berdasarkan kehadiran kelas. Di sini, proses preprocessing berfungsi untuk membersihkan data dari ketidakakuratan dan mengeliminasi data yang tidak relevan agar model yang dibangun lebih akurat. Secara umum, preprocessing data membantu mengoptimalkan input data agar sesuai dengan kebutuhan model dan meningkatkan keakuratan hasil analisis

2.1.5 Pengembangan Aplikasi

Landasan teori mengenai pengembangan aplikasi dapat merujuk pada berbagai metode dan prinsip dasar yang digunakan dalam menciptakan perangkat lunak. Salah satu pendekatan umum adalah pengembangan berbasis prototipe (Prototyping Model), yang menekankan pembuatan model awal dari aplikasi sebelum pengembangan lebih lanjut. Prototipe ini memungkinkan pengujian dan evaluasi di awal untuk memenuhi kebutuhan pengguna secara lebih tepat.

Selain itu, pentingnya pemahaman terhadap siklus hidup pengembangan perangkat lunak juga sangat ditekankan. Salah satu model yang banyak digunakan

adalah model Waterfall, yang memiliki tahapan yang jelas dari analisis kebutuhan hingga pemeliharaan aplikasi. Namun, model agile semakin populer dalam pengembangan aplikasi modern karena fleksibilitasnya dalam menanggapi perubahan kebutuhan dan kecepatan dalam iterasi pengembangan (Purnomo, 2017).

Praktiknya, pengembangan aplikasi sering kali melibatkan analisis kebutuhan sistem dengan melibatkan pengguna untuk memastikan bahwa aplikasi yang dikembangkan memenuhi kebutuhan fungsional dan non-fungsional. Pengujian juga sangat krusial untuk memastikan kualitas perangkat lunak, dengan menggunakan berbagai teknik seperti black-box testing dan white-box testing (Kurniadi et al., 2023).

2.2 Penelitian Terkait (State Of The Art)

Tabel 2. 1 State Of The Art

No	Penulis dan Tahun	Judul	Permasalahan	Hasil dan Simpulan
1	Dewi Listriani, Anif Hanifa Setyaningrum, dan Fenty Eka M. A (2016)	Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Analisa Pola Belanja Konsumen (Studi Kasus Toko Buku Gramedia Bintaro)	Mengetahui kebiasaan pelanggan atau perilaku pelanggan mengenai apa saja buku yang sering dibeli berdasarkan data transaksi penjualan yang telah tersimpan dalam <i>database</i>	Pola asosiasi yang terbentuk dengan nilai minimum <i>support</i> 5% dan nilai minimum <i>confidence</i> 15% menghasilkan 7 aturan asosiasi. <i>strong rules</i> yang didapatkan adalah <i>schoolbooks indonesia curriculum, children's book</i> dengan nilai <i>support</i> 11,23% dan nilai <i>confidence</i> 30,66%
2	Mohammad Badrul (2016)	Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan	Data transaksi yang menumpuk dan belum diolah dengan baik	Penerapan algoritma Apriori yang dilakukan melalui perangkat lunak yang dirancang terbukti menunjukkan hasil yang sama dibandingkan dengan

				melakukan perhitungan secara manual ataupun dengan perangkat lunak sejenis yang dibuktikan melalui perhitungan nilai <i>support</i> dan <i>confidence</i> yang menunjukkan hasil yang sama
3	Asep Budiman Kusdinar, Daris Riyadi, dan Asriyanik (2020)	Implementasi Algoritma Apriori Pada Penyusunan Menu Makanan Rumah Makan Prasmanan	Penataan setiap menu makanan yang ada di rumah makan masih belum efektif, sehingga diperlukan pola dengan mencari keterhubungan dari setiap menu dengan menu lainnya	Menu makanan disusun berdasarkan hasil nilai <i>confident</i> yang paling besar, ditemukan nilai <i>confident</i> terbesar dari menu ayam goreng dan tumis terong dengan nilai <i>confident</i> sebesar 47,54 sehingga tumis terong ditaruh berdekatan dengan ayam goreng pada susunan menu makanan prasmanan
4	Alfie Nur Rahmi dan Yosaphat Ananda Mikola (2021)	Implementasi Algoritma Apriori Untuk Menentukan Pola Pembelian Pada <i>Customer</i>	Toko Bakoel Sembako adalah salah satu toko ritel yang selama ini belum melakukan pemanfaatan data untuk penentuan strategi bisnis toko	Diperoleh hasil dari 30 data transaksi yang diolah, yang memenuhi minimal <i>support</i> 3 dan minimal <i>confidence</i> 70% diperoleh 1 aturan yang terbentuk yaitu

		(Studi Kasus : Toko Bakoel Sembako)		jika pelanggan membeli produk Sedap mie goreng maka akan membeli telur 250gr
5	Anthony Anggrawan, Mayadi, dan Christofer Satria (2021)	Menentukan Akurasi Tata Letak Barang dengan Menggunakan Algoritma Apriori dan Algoritma <i>FP-Growth</i>	Melakukan analisis yang bertujuan untuk mencari kemiripan barang berdasarkan item pembelian dijadikan sebagai acuan dalam tata letak barang dan mengidentifikasi kesamaan barang yang dibeli ketika menambah stok barang	<i>FP-Growth</i> menghasilkan <i>rule</i> yang lebih baik dibandingkan dengan algoritma apriori dengan total <i>rule</i> sebanyak 6, sedangkan algoritma apriori menghasilkan 4 <i>rule</i> , dan untuk pengujian dengan evaluasi hasil <i>rule</i> dari masing masing algoritma, algoritma <i>FP-Growth</i> memiliki hasil yang terbaik dengan <i>lift ratio</i> 1.27908
6	Sanjai Dekariski Saragih (2021)	Analisa Pola Penjualan Alat Pancing Menggunakan Algoritma Apriori	Banyaknya data transaksi yang tidak dimanfaatkan secara maksimal	Menghasilkan jumlah minimum <i>support</i> 30% dan minimum <i>confidence</i> 70% menghasilkan 10 aturan asosiasi yang terbentuk. Data tersebut bisa dijadikan sebagai informasi kepada pemilik toko

				<p>untuk mengembangkan strategi penjualan dan membentuk pola penjualan baru yang lebih baik agar penjualan di toko tersebut dapat berjalan secara maksimal</p>
7	<p>Dewi Anisa Istiqomah, Yuli Astuti, dan Siti Nurjanah (2022)</p>	<p>Implementasi Algoritma <i>Fp-Growth</i> Dan Apriori Untuk Persediaan Produk</p>	<p>Persediaan produk belum dikelola dengan baik serta data transaksi penjualan terus menumpuk dalam jumlah besar setiap harinya dan hanya disimpan menjadi data yang belum diketahui manfaatnya</p>	<p>Aturan asosiasi yang dihasilkan oleh algoritma Apriori juga dihasilkan oleh algoritma <i>FP-Growth</i>. Algoritma Apriori menghasilkan 2 aturan asosiasi dan algoritma <i>FP-Growth</i> menghasilkan 10 aturan asosiasi dengan minimum support 0,05 dan minimum confidence 0,7. Berdasarkan pola hubungan yang dihasilkan, implementasi algoritma <i>FP-Growth</i> dan Apriori dapat membantu Toko Emyra Bedding untuk memantau stok barang yang sering dibeli oleh</p>

				pelanggan sehingga tidak akan terjadi kelangkaan pasokan
8	Muhammad Uliil Albab dan Deny Hidayatulla (2022)	Penerapan Algoritma Apriori pada Sistem Informasi Inventori Toko	Masih menggunakan sistem manual yang kurang efektif	Menghasilkan aplikasi sistem informasi inventori yang dapat mempermudah pengolahan data barang masuk dan keluar yang mempermudah pemilik toko. Penggunaan algoritma apriori untuk mendapatkan keterkaitan antar barang pada hasil penjualan untuk ditemukannya pola asosiasi pada kombinasi antar barang, dari data mining tersebut pemilik dapat mengetahui barang yang perlu diperbanyak atau dikurangi stoknya guna meningkatkan omset penjualan
9	Zaenal Abidin, Arsyah Kharisma Amartya,	Penerapan Algoritma Apriori Pada Penjualan Suku Cadang	Data transaksi hanya digunakan sebagai arsip dan belum diolah dengan baik	Algoritma apriori menghasilkan 13 aturan asosiasi diantaranya 2 aturan

	dan Arliyanti Nurdin (2022)	Kendaraan Roda Dua (Studi Kasus: Toko Prima Motor Sidomulyo)		asosiasi untuk merk suzuki, 6 aturan asosiasi untuk merk honda dan 5 aturan asosiasi untuk merk yamaha yang memenuhi syarat minimum dua parameter, yaitu parameter <i>support</i> dan <i>confidence</i> dan diuji dengan menggunakan pengujian <i>lift ratio</i> untuk menentukan aturan asosiasi yang dihasilkan valid atau tidak valid. Item yang paling laku untuk merk suzuki adalah item dengan kode B01 yaitu Cool Starter Satria FU, untuk merk Honda adalah item dengan kode C14 yaitu Seal Shock Depan Tiger, untuk merk Yamaha adalah item dengan kode D01 yaitu Cool Starter Jupiter Z, Vega ZR, Mio J Hal ini dapat meminimalisir kekosongan
--	--------------------------------	--	--	--

				persediaan dari masing-masing item yang paling laku dari setiap produk dari 3 merk suku cadang tersebut.
10	Taufiq Martiwansyah, Debby Alita, Auliya Rahman Isnain, dan Nuroji (2023)	Penerapan Algoritma Apriori untuk Menentukan Tata Letak Barang (Studi Kasus: Swalayan S&M Mart)	Penataan letak barang masih belum dikelola dengan baik serta transaksi penjualan belum dimanfaatkan dengan baik untuk menunjang optimalisasi proses penjualan	Dengan menggunakan data transaksi penjualan pada tahun 2020-2021 sebanyak 2.024 transaksi dan minimal <i>support</i> 30% dan <i>confidence</i> 80% yang sudah ditetapkan didapatkan hasil 15 <i>rules</i> dengan 2 kombinasi item menghasilkan produk yang sering dibeli bersamaan.

2.3 Kebaruan Penelitian

Tabel 2. 2 Kebaruan Penelitian

NO	Penulis dan Tahun	Judul Penelitian	Ruang Lingkup Penelitian									
			Metode		Pengujian				Tujuan			
			Apriori	FP - Growth	Support	Confidence	Cross-Validation	Rule Quality	Execution Time Testing	Analisis	Software	
1	Dewi Listriani, Anif Hanifa Setyaningrum, dan Fenty Eka M. A (2016)	Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Analisa Pola Belanja Konsumen (Studi Kasus Toko Buku Gramedia Bintaro)	✓		✓	✓						✓

2	Mohammad Badrul (2016)	Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan	√		√	√	√				√
3	Asep Budiman Kusdinar, Daris Riyadi, dan Asriyanik (2020)	Implementasi Algoritma Apriori Pada Penyusunan Menu Makanan Rumah Makan Prasmanan	√			√		√			√
4	Alfie Nur Rahmi dan Yosaphat Ananda Mikola (2021)	Implementasi Algoritma Apriori Untuk Menentukan Pola Pembelian Pada <i>Customer</i> (Studi Kasus : Toko Bakoel Sembako)	√		√	√		√		√	
5	Anthony Anggrawan, Mayadi, dan Christofer Satria (2021)	Menentukan Akurasi Tata Letak Barang dengan Menggunakan Algoritma Apriori dan Algoritma <i>FP-Growth</i>	√	√				√		√	√
6	Sanjai Dekariski Saragih (2021)	Analisa Pola Penjualan Alat Pancing Menggunakan Algoritma Apriori	√		√	√		√		√	
7	Dewi Anisa Istiqomah, Yuli Astuti, dan Siti Nurjanah (2022)	Implementasi Algoritma <i>Fp-Growth</i> Dan Apriori Untuk Persediaan Produk	√	√	√	√					√
8	Muhammad Ulil Albab dan Deny Hidayatulla (2022)	Penerapan Algoritma Apriori pada Sistem Informasi Inventori Toko	√		√	√					√

9	Zaenal Abidin, Arsyah Kharisma Amartya, dan Arliyanti Nurdin (2022)	Penerapan Algoritma Apriori Pada Penjualan Suku Cadang Kendaraan Roda Dua (Studi Kasus: Toko Prima Motor Sidomulyo)	√		√	√		√		√	√
10	Taufiq Martiwansyah, Debby Alita, Auliya Rahman Isnain, dan Nuroji (2023)	Penerapan Algoritma Apriori untuk Menentukan Tata Letak Barang (Studi Kasus: Swalayan S&M Mart)	√		√	√					√
11	Defri Crisna Pramudi (2024)	IMPLEMENTASI ALGORITMA APRIORI PENENTUAN ETALASE DI TOKO ALAT PANCING	√		√	√		√		√	√