

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 DATA MINING**

*Data mining* merupakan salah satu tahap proses dari KDD dengan menelusuri data yang ada untuk membangun sebuah hubungan yang berarti pola, dan kecenderungan dengan memeriksa sekumpulan data yang tersimpan dengan menggunakan teknik pengendalian pola seperti teknik statistik dan matematika [15]. *Data mining* terdiri dari algoritma inti yang memungkinkan seseorang untuk mendapatkan wawasan dan pengetahuan dari data yang sangat besar [16].

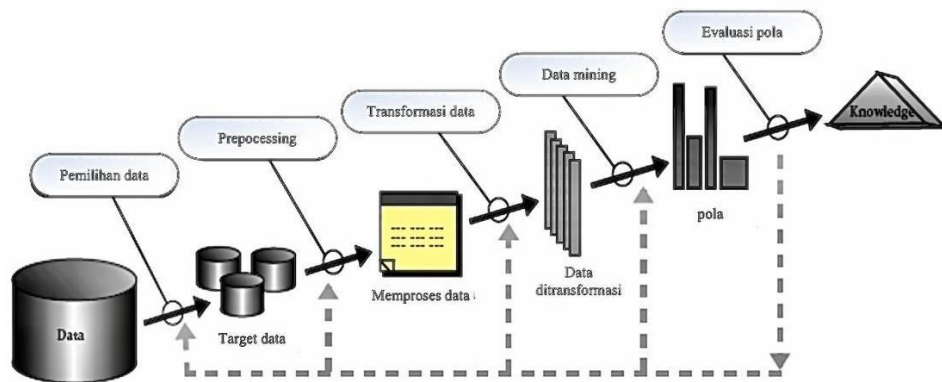
Dalam definisinya, *data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar [17].

*Data mining* merupakan suatu proses ekstraksi atau penggalian data dan informasi yang besar, yang belum diketahui sebelumnya, namun dapat di pahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting [18]. *Data mining* merupakan proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar [19].

*Data mining* merupakan proses ekstraksi data untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar guna mendapatkan informasi yang berguna.

### 2.1.1 Tahapan Dalam Data Mining

*Knowledge Discovery in Database (KDD)* digambarkan sebagai proses penggunaan metode *data mining* untuk menemukan informasi bermanfaat dan pola didalam data, sedangkan *data mining* menggunakan algoritma untuk mengidentifikasi pola-pola didalam data yang didapatkan dari proses KDD sehingga KDD adalah salah satu proses menyeluruh yang meliputi *data mining*. Beberapa tahapan dalam proses KDD dapat dilihat pada gambar 2.1.



**Gambar 2.1 Tahapan Proses KDD [20].**

#### 1. Pemilihan Data

Merupakan proses pemilihan atau penyeleksian data dari sekumpulan data operasional yang dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Dalam tahapan ini, kita memilih data-data seperti apa saja yang kita butuhkan untuk diproses lebih lanjut.

#### 2. *Preprocessing*

Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak

sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang kita miliki. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil *data mining* nantinya. Pembersihan data juga akan mempengaruhi performa dari sistem *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

### 3. Transformasi Data

Beberapa teknik *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut *binning*. Disini juga dilakukan pemilihan data yang diperlukan oleh teknik *data mining* yang dipakai. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil data mining nantinya karena ada beberapa karakteristik dari teknik-teknik *data mining* tertentu yang tergantung pada tahapan ini.

### 4. *Data Mining*

Teknik, metode atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

## 5. Evaluasi Pola

Dalam proses ini, pola-pola yang telah diidentifikasi oleh sistem kemudian diterjemahkan ke dalam bentuk knowledge yang mudah dimengerti oleh pihak yang berkepentingan.

### 2.1.2 Pengelompokan *Data Mining*

*Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas/pekerjaan yang dapat dilakukan, yaitu [21]:

#### 1. Deskripsi

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

#### 2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan baris data (*record*) lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

#### 3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa

metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

#### 4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

#### 5. Pengklasteran (clustering)

Pengklasteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

#### 6. Asosiasi

Tugas asosiasi dalam *data mining* adalah untuk menemukan atribut yang muncul dalam satu waktu. Salah satu implementasi dari asosiasi adalah *market basket analysis* atau analisis keranjang belanja.

## 2.2 ASOSIASI

*Association rules* (aturan asosiasi) merupakan salah satu dari pengetahuan yang paling penting dalam data mining yang dapat diartikan sebagai hubungan dan ketergantungan antara sejumlah item (itemsets) dengan menentukan nilai *support* dan *confidence* pada basis data. Aturan asosiasi ingin memberikan informasi tersebut dalam bentuk hubungan “if-then” atau “jika-maka” [22].

*Association rule* adalah bentuk jika “kejadian sebelumnya” kemudian “konsekuensinya” (*If antecedent, then consequent*), yang diikuti dengan

perhitungan aturan *support* dan *confidence*. Bentuk umum dari *association rule* adalah *Antecedent*  $\rightarrow$  *Consequent*, dimana tidak ada batasan dalam jumlah item item pada bagian *antecedent* ataupun *consequent* dalam sebuah rule. *Association rule* memiliki dua tahap pengerjaan, yaitu [23]:

- a. Mencari kombinasi yang paling sering terjadi dari suatu itemset.
- b. Mendefinisikan *Condition* dan *Result* (untuk conditional association rule).

Dalam hal analisis asosiasi digunakan untuk menemukan semua aturan asosiasi yang memenuhi syarat minimum untuk *support* dan syarat minimum untuk *confidence*. Tahapan metode dasar asosiasi terbagi menjadi dua tahap [24]:

1. Analisa Pola Frekuensi Tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai Support dalam database. Nilai Support sebuah item diperoleh dengan rumus 1 dan 2:

$$Support(A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ transaksi} \quad (1)$$

sedangkan nilai Support dari 2 item diperoleh dari rumus 2 berikut:

$$Support(A,B) = P(A \cap B)$$

$$Support(A,B) = \frac{\Sigma Transaksi\ mengandung\ A\ dan\ B}{\Sigma Transaksi} \quad (2)$$

2. Pembentukan Aturan Asosiasi

Setelah semua pola frekuensi tinggi ditemukan, 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 dari rumus 3:

$$Confidence = P\left(\frac{B}{A}\right) = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi A}} \dots\dots\dots (3)$$

### 2.2.1 Algoritma Asosiasi

Berikut beberapa algoritma yang di gunakan untuk asosiasi pada *data mining* yaitu:

#### 1. Apriori

Melibatkan langkah-langkah seperti menghasilkan itemsets kandidat dan memindai dataset. Apriori lebih mudah untuk diimplementasikan dan dimengerti.

#### 2. FP-Growth

Melibatkan konstruksi FP-Tree dan menghasilkan conditional FP-Trees, yang dapat lebih efisien terutama pada dataset dengan ukuran besar. FP-Growth memerlukan pemahaman yang lebih mendalam tentang struktur pohon.

## 2.3 ALGORITMA APRIORI

Algoritma apriori termasuk jenis aturan asosiasi pada *data mining* atau *market basket analysis*. Algoritma apriori salah satu algoritma yang berfungsi untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah item yang sering muncul pada sebuah *data mining* dan memiliki frekuensi *support* diatas ambang tertentu yang disebut dengan istilah minimum *support* dan minimum *confidence*. Pola frekuensi tinggi ini digunakan untuk menyusun aturan assosiasi dan juga beberapa teknik *data mining* [25].

Terdapat dua proses utama dalam membentuk kandidat itemset pada algoritma Apriori yaitu [26]:

1. *Join* (penggabungan)

Dalam proses ini, setiap item dikombinasikan dengan item lain sampai tidak dapat terbentuk kombinasi lagi.

2. *Pruning* (pemangkasan)

Pada proses ini, hasil kombinasi item akan dipangkas berdasarkan minimum *support* yang telah ditentukan.

## **2.4 POLA PEMBELIAN**

Pola adalah sistem cara kerja ataupun bentuk struktur yang pasti. Pembelian konsumen merupakan suatu proses tindakan yang dilakukan oleh konsumen untuk membeli produk berupa barang atau jasa [27]. Maka pola pembelian konsumen bisa diartikan sebagai bentuk struktur dari kegiatan belanja konsumen yang pasti. Dari pola pembelian yang dapat diprediksi inilah pembuat keputusan dapat membuat strategi pemasaran yang lebih efektif [28].

## **2.5 DATA PREPROCESSING**

*Preprocessing* data merupakan langkah penting dalam proses penemuan pengetahuan, karena keputusan-keputusan yang berkualitas harus didasarkan pada data yang berkualitas. *Preprocessing* data seringkali digunakan untuk mengurangi kesalahan data dan sistematis bias dalam data mentah sebelum analisis apapun terjadi [29].



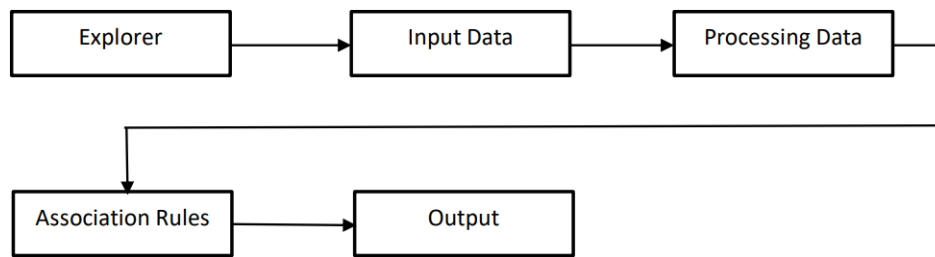
Data *preprocessing* sendiri disebut sebagai langkah utama dan yang paling kritis dalam proses *Knowledge Data Discovery* (KDD). Teknik ini dibutuhkan agar dapat menghilangkan *noise* dari suatu *raw data* (data mentah) untuk mengambil suatu informasi dari data tersebut [30]. Oleh karena itu *preprocessing* data sangat penting karena dapat meningkatkan kualitas data serta meningkatkan akurasi dan efisiensi proses *mining* [31].

## 2.6 WEKA

WEKA (*Waikato Environment for Knowledge Analysis*), adalah kumpulan algoritma pembelajaran mesin untuk tugas-tugas data mining. Algoritma baik dapat diterapkan langsung ke dataset atau disebut dari kode Java itu sendiri. Weka berisi alat untuk data pra-pengolahan, klasifikasi, regresi, clustering, aturan asosiasi, dan visualisasi. Hal ini juga cocok untuk mengembangkan skema pembelajaran mesin baru [32].

WEKA adalah perangkat lunak yang paling terkenal untuk melakukan tugas-tugas ML (*machine learning*) dan DM (*data mining*). Algoritmanya dapat diterapkan secara langsung ke sebuah dataset dari antarmukanya sendiri atau digunakan dalam kode Java Anda sendiri [33].

Proses pengolahan data yang umum pada perangkat lunak WEKA bisa dilihat pada gambar 2.2.



**Gambar 2.2 Proses Pengolahan Data Pada WEKA**

1. Pada tampilan utama Weka, pilih tab "Explorer".
2. Pilih tab "Preprocess" di bagian atas, lalu klik tombol "Open file" dan pilih dataset.
3. Pilih tab "Associate" dan pilih associator "Apriori".
4. Hasil pengolahan data muncul di layar (window). Output WEKA bisa berupa tulisan atau grafik.

## 2.7 PENELITIAN SEJENIS

Penelitian sejenis merupakan tinjauan informasi dan pengetahuan mengenai penelitian yang sama dengan penelitian ini, tinjauan tersebut dapat dilihat pada tabel 2.1.

**Tabel 2.1 Penelitian Sejenis**

No	Penulis dan Tahun	Masalah	Metode	Hasil
1.	HapsariDita Anggraeni, Ragil Saputra, Beta Noranita (2013) [34].	Apotek sebagai salah satu organisasi yang menghasilkan data penjualan setiap hari, belum dapat memaksimalkan pemanfaatan data	Algoritma apriori.	Algoritma yang digunakan sebagai proses utama dari analisis keranjang pasar adalah algoritma apriori dengan menggunakan parameter minimum <i>support</i> , minimum <i>confidence</i> , dan periode bulan

		tersebut. Data penjualan hanya disimpan tanpa dilakukan analisis lebih lanjut.		transaksi penjualan untuk menemukan aturan asosiasi. Aplikasi <i>data mining</i> menghasilkan aturan asosiasi antar item pada bulan Februari 2012 yaitu konsumen melakukan transaksi pembelian obat jenis obat darah dan analgesik secara bersamaan dengan <i>support</i> sebesar 2,08% dan <i>confidence</i> sebesar 45,45%.
2.	Robi Yanto , Riri Khoiriah (2015) [35].	Data penjualan obat selama disimpan dalam aplikasi Microsoft Excel sehingga data penjualan yang semakin hari semakin banyak tersebut hanya berfungsi sebagai arsip dan laporan penjualan obat bagi perusahaan dan tidak dapat dimanfaatkan perusahaan untuk pengembangan dalam penentuan pola pembelian obat.	Algoritma apriori.	Dari proses perhitungan dengan algoritma apriori ditentukan nilai minimum <i>support</i> 50 % dan nilai minimum <i>confidence</i> 70 % maka dihasilkan kecenderungan obat yang dibeli oleh konsumen yaitu: Jika membeli amoxicilin maka membeli obat asamefenamat dengan <i>confidence</i> 75 %. Jika Membeli Cefadroxil Maka Membeli Sanmol dengan <i>confidence</i> 77 %.
3.	Lilis Sry Rahayu Situmorang , Meri Sri Wahyuni, M.Syaifuddin (2022) [36].	Untuk saat ini proses penjualan di Apotek Pelita 3 masih bersifat manual yaitu dengan cara tulis tangan, sehingga pihak apotek masih sulit dalam mendapatkan informasi data pola penjualan obat.	Algoritma FP-growth	Berdasarkan hasil nilai perhitungan <i>support</i> 2 itemset menggunakan algoritma FP-growth, <i>frequent</i> itemset yang memenuhi minimum <i>confidence</i> yaitu 40% . Dibentuk rule aturan asosiasi, jika membeli obat Becom C maka secara bersamaan akan membeli obat Paracetamol dengan nilai <i>support</i> 56,7% dan <i>confidence</i> 70,83%. dan Jika membeli obat Becom C maka secara bersamaan akan membeli obat Ambroxol dengan nilai <i>support</i> 40% dan <i>confidence</i> 50%.

4.	Muhammad Sholihul hamdi, I Gede Putu Wirarama Wadashwara Wirawan, Fitri Bimantoro (2022) [37].	Dari proses transaksi penjualan yang terjadi setiap harinya seringkali pemilik apotek memanfaatkan data hasil transaksi hanya untuk laporan data tahunan dan tidak menyadari akan fungsi lain dari data tersebut dapat dimanfaatkan untuk mendapatkan informasi guna meningkatkan penjualan.	Algoritma apriori.	Dari hasil pemrosesan data transaksi penjualan obat pada Apotek Gilda Farma menggunakan algoritma apriori yang telah dilakukan menghasilkan 6 rules yang berupa kombinasi obat yang paling sering dibeli pada Apotek Gilda Farma. Dalam penelitian ini rule paling kuat yang dihasilkan adalah kombinasi obat Allopurinol dengan Piroxicam yang memiliki nilai <i>confidence</i> 0.191011 adapun untuk nilai rule terendah berada pada kombinasi obat Paracetamol dengan Amoxycillin dengan nilai <i>confidence</i> yang dihasilkan 0.061809.
5.	Rizal Rachman, Nanang Hunaifi (2020) [38].	Seiring pertumbuhan industri, informasi tentang produknya menjadi kebutuhan bagi perusahaan. Salah satu kebutuhan penting adalah informasi tentang penjualan obat-obatan dan informasi tentang persiapan atau produksi obat-obatan. Informasi mengenai berapa banyak obat yang akan diproduksi merupakan hal yang sangat penting karena hal ini berkaitan dengan berapa banyak penjualan yang terjadi dalam kurun waktu tertentu atau target pasar yang akan dicapai.	Algoritma apriori dan struktur FP-Tree.	Berdasarkan data transaksi penjualan obat di pabrik Farma kimia Jakarta, dilakukan analisis menggunakan algoritma Apriori dengan dukungan parameter <i>support</i> 10% dan <i>confidence</i> minimum 50%. Hasil penelitian menghasilkan 7 aturan Asosiasi dengan kombinasi item terbesar hingga 2 item. Dan FP-Tree adalah struktur penyimpanan data terkompresi. FP-Tree dibangun dengan memetakan setiap catatan transaksi ke setiap jalur spesifik di FP-tree.
6.	Feronikawati Lasmaria Sitorus, Agus	Apotek Metro Medika Center adalah Apotek yang terletak di Pejaten Barat yang masih	Algoritma apriori.	Penelitian ini dilakukan dengan menerapkan algoritma apriori dengan melihat obat/item yang memenuhi

	Salim (2021) [13].	menggunakan cara manual, dengan setiap transaksi harus ditulis ke dalam buku besar, ini menyebabkan kesulitan karena begitu banyak data transaksi yang ada dan akan sulit untuk mencari data yang sebelumnya, dan akan ada kemungkinan data tersebut hilang karena masih manual. Dengan perkembangan teknologi maka dibutuhkan pengolahan data yang besar dengan menggunakan <i>data mining</i> dengan metode asosiasi.		minimum <i>support</i> dan minimum <i>confidence</i> . <i>Data mining</i> menghasilkan aturan asosiasi antar item pada bulan Januari 2019 sampai dengan Desember 2019, diketahui pola penjualan obat Jika membeli paracetamol, maka akan membeli amoxcillin dan ponstan dengan nilai <i>support</i> 33% dan nilai <i>confidence</i> 66,67%.
7.	Siti Awaliyah Rachmah Sutomo, Frisma Handayanna (2020) [14].	Pengolahan data pada Apotek Sekar Adi masih sangat sederhana, banyaknya data pada transaksi penjualan di setiap pembelian obat dapat menimbulkan penumpukan data dan berbagai macam masalah salah satu contohnya seperti persediaan stok obat, dengan teknik <i>data mining</i> bisa di analisis perilaku konsumen dalam melakukan transaksi pola pembelian obat.	Algoritma apriori.	Proses penentuan pola pembelian obat dapat dilakukan dengan menerapkan metode algoritma apriori, penentuan pola pembelian obat dapat dilakukan dengan melihat hasil dari kecenderungan konsumen membeli obat berdasarkan kombinasi 3 <i>itemset</i> . Dengan menghitung Analisis Pola Frekuensi Tinggi dan Pembentukan Aturan Asosiasi, minimal <i>support</i> 30% terdapat kombinasi 3 <i>itemset</i> yaitu molagit per tab(M1), vit c tablet (V2), dan paracetamol 500 mg tablet (P2) dengan hasil <i>support</i> yang didapat 33,33%, dan minimal <i>confidence</i> 65% terdapat 6 final <i>association rules</i> .
8.	Olfiana Wati, Ratna Juita, Marlinda (2024) [39].	Kendala yang terjadi di apotek di kota Nabire saat ini kurangnya persediaan obat yang	Algoritma apriori.	Hasil akhir ditemukan 4 aturan asosiasi dengan nilai minimum <i>support</i> 30% dan nilai minimum <i>confidence</i> 50% .

		<p>memadai yaitu dimana sering sekali penjualan obat yang diinginkan konsumen atau masyarakat tidak ada atau habis sehingga mengakibatkan konsumen tersebut pindah dari satu apotek ke apotek lainnya. Hal tersebut membuat lambatnya pelayanan terhadap konsumen dan mengurangi tingkat penjualan di apotek.</p>		<p>Jika membeli Asam Mefenamat dan Grantusif maka akan membeli Bodrex dengan <i>support</i> 34,21%, <i>confident</i> 58,81%. Jika membeli Asam Mefenamat dan Dexa maka membeli Bodrex dengan nilai <i>support</i> 36,09%, nilai <i>confident</i> 62,04%. Jika membeli Asam Mefenamat dan Dexa maka membeli Paracetamol dengan nilai <i>support</i> 31,42%, nilai <i>confident</i> 68,85%. Jika membeli Asam Mefenamat dan Dexa maka membeli Amoxilin dengan nilai <i>support</i> 31,01%, nilai <i>confident</i> 68,04%.</p>
9.	<p>Afrizal Ardiansyah, Ahmad Turmudi Zy, Agung Nugroho (2023) [40].</p>	<p>Penjualan obat di klinik merupakan aspek penting dalam pelayanan kesehatan berkualitas. Namun, terdapat beberapa masalah yang perlu diatasi untuk mengoptimalkan efisiensi, kepuasan pasien, dan kesuksesan bisnis klinik. Masalah-masalah yang teridentifikasi mencakup pengelolaan persediaan obat yang tidak efisien, kurangnya pemahaman tentang preferensi konsumen terhadap obat-obatan, kekurangan strategi pemasaran yang tepat berdasarkan pola penjualan obat, dan kurangnya pengambilan keputusan</p>	<p>Algoritma apriori.</p>	<p>Hasil penelitian ini menyimpulkan bahwa dua obat yang sering dibeli oleh pasien di klinik adalah Mecobalamindan Omeprazole 20mg. Pasien yang membeli Mecobalamin cenderung juga membeli Omeprazole 20mg dengan <i>confidence</i> 66%. Begitu juga sebaliknya, pasien yang membeli Omeprazole 20mg kemungkinan besar juga membeli Mecobalamin dengan <i>confidence</i> 57.1%.</p>

		berbasis data yang mendalam.		
10.	Indra Supriadi, Dudih Gustian, Falentino Sembiring (2021) [41].	Pola kecenderungan pembelian obat setiap orang berbeda tergantung kepada obat yang cocok ataupun penyakit yang di derita. Banyak menumpuknya data pada setiap transaksi penjualan obat dapat menimbulkan masalah contohnya seperti persediaan stok obat yang tidak bisa dikontrol dan data transaksi penjualan. Supaya tidak mengalami keterlambatan dalam pengadaan obat dan juga mengantisipasi stok obat yang berlebihan maka pihak apotek perlu menentukan strategi untuk mempermudah pengontrolan persediaan obat.	Algoritma apriori.	Penerapan algoritma apriori pada 350 data transaksi penjualan Obat di Apotek Rsu Jampangkulon menggunakan tools RapidMiner dengan ditentukan minimum <i>support</i> 5% dan <i>confidence</i> 60% menghasilkan 10 rules asosiasi, dengan begitu pihak Apotek Rsu Jampangkulon dapat mengetahui hubungan pola frekuensi penjualan obat yang sering dibeli secara bersamaan oleh konsumen serta obat yang paling banyak terjual .