

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Landasan Teori**

##### **2.1.1 Data Mining**

Data mining adalah bidang ilmiah yang mempelajari metode menemukan suatu hubungan atau kombinasi dalam data besar (Wibowo & Jananto, 2020). Data mining adalah penggunaan algoritme dan statistik, mesin *learning* dan sistem manajemen basis data untuk mengekstrak informasi dari kumpulan data besar (Handayani & Rosyid, 2021). Data mining yaitu proses menemukan pola atau informasi dari data terpilih dengan menggunakan metode tertentu (Azura & Safii, 2021). Data mining atau penambangan data adalah langkah menganalisis data besar untuk memperoleh hubungan antar data guna merangkumnya menjadi data yang lebih mudah dipahami dan berguna (Yani & Nazir, 2022). Dari definisi diatas dapat dipahami bahwa data mining adalah ilmu komputer komprehensif yang menganalisis data besar, tujuannya adalah dengan menggunakan metode tertentu untuk menemukan hubungan antar data dan memperoleh informasi terkini.

Data mining mempunyai beberapa ciri, antara lain data mining berkaitan dengan penemuan pola data yang tersembunyi atau yang sebelumnya tidak diketahui, data yang diolah dalam data mining biasanya berukuran besar, semakin besar data yang digunakan maka semakin akurat pula hasil yang didapat, berguna untuk keputusan penting, terutama dalam hal strategi (Saputra & Sibarani, 2020).

Data mining secara sistematis memiliki tiga langkah utama yaitu (Wahono, 2023):

1. Eksplorasi/pemrosesan awal data

Pada proses ini dilakukan pembersihan terhadap data mentah, melakukan normalisasi terhadap data, transformasi data, menangani data yang kurang benar, memilih fitur dan lainnya.

2. Membangun model dan melakukan validasi terhadapnya

Membangun dan memvalidasi model berarti menganalisis model yang berbeda

dan memilih model yang memiliki daya prediksi yang baik. Metode yang digunakan dalam tahap ini antara lain klasifikasi, regresi, analisis *cluster*, deteksi anomali, aturan asosiasi, analisis pola urutan, dll. Pada beberapa referensi, deteksi anomali juga dimasukkan dalam tahap pencarian. Namun deteksi anomali juga dapat digunakan sebagai algoritmedasar, khususnya untuk mencari informasi tertentu.

### 3. Penerapan

Tahap penerapan merupakan menerapkan suatu model pada data baru untuk memperoleh hasil perkiraan/prediksi terhadap kasus yang diteliti.

Berdasarkan tugas yang dilakukan, data mining dikelompokkan menjadi beberapa kelompok, yaitu (Wahono, 2023) :

#### 1. Deskripsi

Terkadang peneliti dan analis mencoba suatu metode untuk menemukan pola dalam data. Contohnya, pada saat pemilihan presiden seorang petugas pemilu mungkin tidak menyadari suatu fakta bahwa calon presiden yang tidak profesional hanya menerima sedikit dukungan.

#### 2. Estimasi

Estimasi hampir sama dengan klasifikasi, namun untuk variabel target estimasi lebih bersifat numerik dibandingkan kategorikal. Model estimasi dibangun dengan menggunakan catatan yang lengkap, yang nilainya diperoleh dari variabel target dan digunakan sebagai prediktor. Nilai-nilai variabel prediktor kemudian digunakan sebagai kilas balik estimasi variabel target. Misalnya, ketika tekanan darah sistolik pasien di rumah sakit dinilai, peninjauannya didasarkan pada usia, jenis kelamin, kadar *natrium* darah, dan berat badan pasien. Selama proses pembelajaran diperoleh model estimasi berdasarkan hubungan tekanan darah *sistolik* dengan nilai variabel prediktor. Contoh lainnya adalah memperkirakan indeks kumulatif mahasiswa pascasarjana berdasarkan indeks kumulatif mahasiswa sarjana.

#### 3. Prediksi

Peramalan hampir sama dengan klasifikasi dan estimasi, namun hasil prediksinya adalah pada masa yang akan datang. Prediksi dalam bidang bisnis dan penelitian, seperti prediksi kenaikan harga beras dalam tiga bulan ke depan, prediksi persentase peningkatan kecelakaan lalu lintas pada saat batas kecepatan

minimum dinaikkan pada tahun berikutnya. Metode yang digunakan dalam klasifikasi dan estimasi juga dapat diterapkan pada prediksi jika diberikan situasi atau keadaan yang tepat.

#### 4. Klasifikasi

Proses klasifikasi melibatkan pencarian variabel kategori. Misalnya, ada tiga kategori pendapatan: pendapatan tinggi, pendapatan menengah, dan pendapatan rendah. Dalam bisnis dan penelitian, contoh klasifikasi antara lain mendiagnosis penyakit pada pasien dan menilai kualitas kredit nasabah yang mengajukan pinjaman hipotek.

#### 5. Pengklusteran

*Clustering* adalah pengelompokan objek – objek yang direkam, diamati atau diperhatikan untuk membentuk kelas – kelas objek yang serupa. Kumpulan data disebut *cluster* jika mempunyai karakteristik yang sama dan tidak berhubungan satu sama lain. Tidak ada variabel target dalam *clustering*, itulah perbedaan antara *clustering* dan klasifikasi. Contoh kluster bisnis adalah perusahaan yang tidak memiliki sumber daya pemasaran yang signifikan sehingga menargetkan pemasaran melalui kelompok konsumen.

#### 6. Asosiasi

Tugas asosiasi dalam penambangan data melibatkan pencarian atribut yang terjadi bersamaan. Dalam dunia bisnis, korelasi sering disebut dengan analisis keranjang belanja. Misalnya, asosiasi bisnis dan peneliti mempelajari berapa banyak pelanggan perusahaan yang diharapkan memberikan tanggapan positif terhadap usulan insentif untuk peningkatan layanan dan menemukan bahwa ada pembeli di supermarket yang juga membeli barang.

### 2.1.2 Aturan Asosiasi

Aturan asosiasi (*Association rule*) adalah salah satu teknik penambangan data (data mining) untuk mengetahui aturan kombinasi (Pare et al., 2023). Aturan asosiasi sering dikenal dengan nama analisis keranjang belanja

(*Market Basket Analysis*) karena mampu menganalisis isi keranjang para konsumen yang melakukan transaksi. Contohnya, kemampuan konsumen untuk membelipasta gigi dan sikat gigi, atau sabun mandi dan sampo secara bersamaan, sehingga pemilik supermarket dapat merencanakan tata letak produknya dan merencanakan model promosi yang kemungkinan besar akan menarik konsumen untuk membelinya, seperti menawarkan diskon kepada konsumen jika mereka membeli beberapa produk secara bersamaan. Aturan ini terkenal karena tidak hanya mampu menganalisis data pembelian yang berlaku pada transaksi langsung, tetapi pembelian atau pesanan *online* juga dapat dianalisis menggunakan aturan asosiasi ini (Arhami & Nasir, 2020).

Pada umumnya, tujuan dari aturan asosiasi adalah untuk memperoleh kumpulan itemset yang muncul secara bersama pada proses transaksi. Satu setitem terdiri dari *i* buah item yang disebut dengan *i*-itemset. Persentase kombinasi yang terdiri dari satu itemset disebut dengan *support* itemset dan rasio kuatnya hubungan antar item dalam aturan asosiasi disebut *confidence* (nilai kepastian). Jika dimisalkan, kumpulan items:  $I = \{I_1, I_2, \dots, I_m\}$ , transaksi:  $DB = \{t_1, t_2, \dots, t_n\}$ ,  $t_j \subseteq I$ , Itemset:  $\{I_{i1}, I_{i2}, \dots, I_{ik}\} \subseteq I$ . Maka suatu kaidah asosiasi dapat dinyatakan dalam pernyataan IF...THEN... (Arhami & Nasir, 2020).

Dalam melakukan analisis asosiasi terdapat dua metode yaitu (Fahrudin, 2019):

#### 1. Analisis pola frekuensi tinggi

Analisis pola frekuensi tinggi merupakan langkah yang bertujuan untuk menemukan kombinasi item yang memenuhi nilai minimum *support* yang telah ditentukan untuk suatu kumpulan data. Nilai *support* item dapat ditentukan menggunakan persamaan 2.1.

$$Support(A) = \frac{Jumlah\ transaksi\ yang\ mengandung\ A}{Total\ transaksi} \dots \dots \dots (2.1)$$

Atau menggunakan persamaan 2.2 untuk menghitung *support* dua buah item.

$$Support(A \cap B) = \frac{Jumlah\ transaksi\ yang\ mengandung\ A\ dan\ B}{Total\ transaksi} \dots \dots (2.2)$$

#### 2. Pembentukan aturan asosiasi

Pembentukan aturan asosiasi terjadi setelah semua pola frekuensi tinggi ditemukan. Selanjutnya, cari aturan asosiasi yang memenuhi nilai minimum *confidence*. Jika presentase kemunculan item A dan B secara bersamaan disebut *support*, maka frekuensi munculnya item B dalam transaksi yang menyertakan A disebut *confidence*. Persamaan 2.3 merupakan rumus untuk menentukan nilai *confidence*.

$$Confidence = P(B|A) = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Jumlah transaksi mengandung A}} \dots (2.3)$$

### 2.1.3 Algoritme Apriori, Modified Apriori dan FP-Growth

Algoritme apriori merupakan salah satu algoritme dalam aturan asosiasi untuk menemukan *frequent* itemset (Saefudin & Fernando, 2020). *Frequent* itemset atau pola frekuensi adalah pola item – item dalam suatu *database* yang frekuensi atau *support*nya melebihi batas rata – rata dapat disebut dengan minimum *support* (Tiara Antesia, 2020). Algoritme apriori memiliki dua proses utama yaitu (Susilo & Sunardi, 2019):

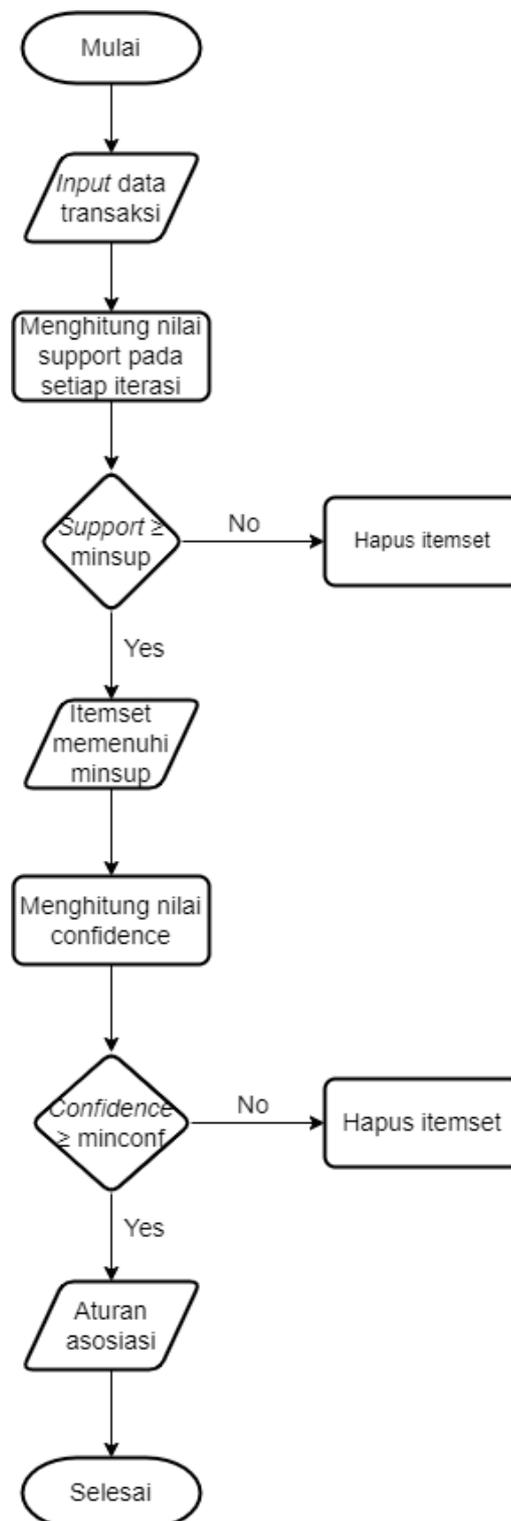
#### 1. *Join* (Penggabungan)

Pada proses penggabungan/*join* terjadi kombinasi antar item hingga tidak ada lagi kombinasi yang terbentuk. Untuk mencari Lk pada suatu set dari kandidat k-itemset, dilakukan dengan menggabungkan Lk-1 dengan dirinya sendiri. Hasil dari penggabungan tersebut dinotasikan sebagai Ck. Didalam aturan *join* tidak boleh dihasilkan kandidat yang sama.

#### 2. *Prune* (Pemangkasan)

*Prune*/pemangkasan merupakan tahapan atau proses dimana hasil kombinasi item yang sudah dikombinasikan dilakukan pemangkasan dengan minsup yang telah ditetapkan diawal.

Langkah-langkah dalam perhitungan algoritme apriori dapat dilihat pada gambar 2.1

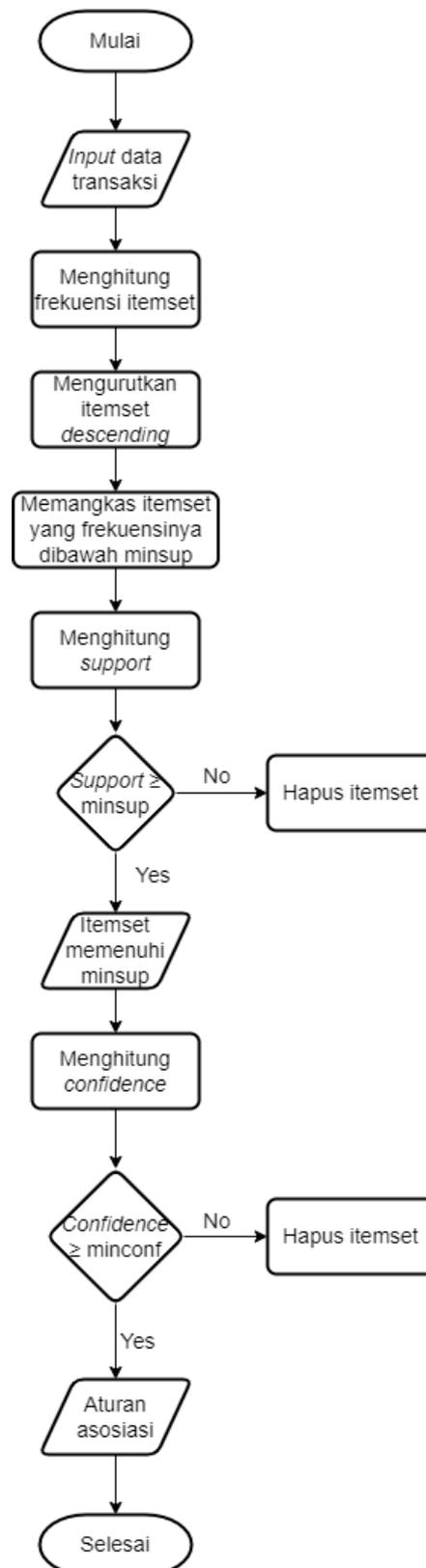


**Gambar 2.1** Diagram alur (*flowchart*) algoritme apriori  
Pada gambar 2.1 merupakan langkah perhitungan algoritme apriori :

1. Menentukan nilai minimum *support*

2. Menghitung nilai *support* pada 1 itemset
3. Memangkas itemset yang nilai *support* nya dibawah minsup
4. Hasil pemangkasan digabungkan membentuk 2 itemset
5. Menghitung nilai *support* pada 2 itemset
6. Memangkas itemset yang nilai *support* nya dibawah minsup
7. Hasil pemangkasan digabungkan membentuk 3 itemset
8. Menghitung nilai *support* pada 3 itemset
9. Memangkas itemset yang nilai *support* nya dibawah minsup
10. Menghitung nilai *confidence* untuk mendapatkan aturan asosiasi
11. Nilai *confidence* diatas nilai minconf maka menjadi aturan asosiasi

Algoritme *modified* apriori adalah pembaruan atau modifikasi dari algoritme apriori yang digunakan dalam pengumpulan data untuk mencari polaasosiasi berdasarkan hubungan antar item untuk mendapatkan kombinasi elemen. Perbedaan algoritme *modified* apriori dengan algoritme apriori terletak pada cara perhitungan nilai *support*. Pada algoritme apriori, nilai *support* dihitung secara langsung pada kumpulan data dan setiap iterasi melibatkan banyak perhitungan *support* yang membutuhkan banyak waktu komputasi pada kumpulan data yang besar. Sedangkan pada algoritme *modified* apriori nilai *support* dihitung menggunakan informasi *support* yang telah dihitung sebelumnya, sehingga mengurangi jumlah perhitungan *support* yang diperlukan pada setiap iterasi (Fey, 2023). Langkah perhitungan algoritme *modified* apriori dapat dilihat pada gambar 2.2.



**Gambar 2.2** Diagram alur (*flowchart*) algoritme *modified apriori*

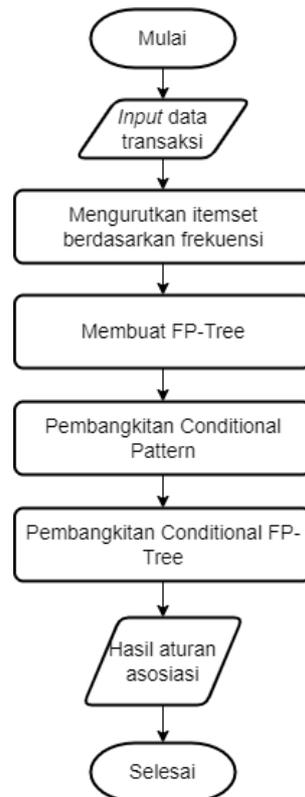
Pada gambar 2.2 merupakan langkah perhitungan algoritme *modified apriori*:

1. Menghitung frekuensi itemset
2. Mengurutkan itemset berdasarkan nilai frekuensi yang tertinggi
3. Memangkas itemset yang nilai frekuensinya dibawah minimum *support* pada setiap iterasi
4. Menghitung nilai *support*
5. Menghitung *confidence*
6. Diperoleh aturan asosiasi

Algoritme *fp-Growth* adalah algoritme yang berfungsi untuk menentukan kumpulan data yang sering muncul (*frequent* itemset) (Munanda & Monalisa, 2021). Algoritme *fp-Growth* merupakan algoritme apriori yang dikembangkan, sehingga kekurangan dari algoritme apriori dapat diperbaiki algoritme *fp-growth*. Perbedaan antara algoritme apriori dengan algoritme *fp-growth* terletak pada pencarian *frequent* itemset. Didalam algoritme apriori diperlukan *generate candidate* untuk mendapatkan *frequent* itemset, sedangkan pada algoritme *fp-growth* untuk mendapatkan *frequent* itemset diterapkan konsep pembangunan *tree* atau biasa disebut dengan *fp-tree* (Suhada & Ratag, 2020). Algoritme *fp-growth* telah diterapkan pada berbagai bidang untuk mencari informasi yang tersembunyi dalam *database*. Seperti dalam penelitian Erlin Elisa dan Nurul Azwanti yang menjelaskan bahwa algoritme *fp-growth* dapat membantu pemilik UD. Maju Bersama untuk mengambil keputusan dalam mengolah persediaan LPG (Harlan Kurnia AR & Nurmaliana Pohan, 2022). Algoritme *fp-growth* dibagi menjadi tiga langkah utama, yaitu (Munanda & Monalisa, 2021):

1. Tahap pembangkitan *conditional pattern base*
2. Tahap pembangkitan *conditional fp-tree*
3. Tahap pencarian *frequent* itemset

Untuk lebih jelasnya mengenai langkah perhitungan algoritme *fp-growth* dapat dilihat pada gambar 2.3.



**Gambar 2.3** Diagram alur (*flowchart*) algoritme *fp-growth*

Pada gambar 2.3 merupakan langkah perhitungan algoritme *fp-growth* :

1. Menghitung frekuensi itemset
2. Mengurutkan itemset berdasarkan prioritas
3. Membuat *fp-tree*
4. Pembangkitan *conditional pattern*
5. Pembangkitan *conditional fp-tree*

#### 2.1.4 Korelasi *lift*

Korelasi *lift* atau *lift ratio* merupakan suatu rasio untuk mengukur kekuatan hubungan/relasi antar item yang terbentuk. Nilai korelasi *lift* digunakan sebagai patokan atau tolak ukur untuk mengetahui bahwa aturan asosiasi yang terbentuk akurat atau tidak akurat. Pada *lift ratio* terdapat tiga kemungkinan yang dihasilkan, yaitu (Prabowo & Ramdani, 2020):

1. Jika *lift ratio*  $< 1$ , menyatakan bahwa item A dan B muncul lebih jarang dari yang diharapkan. Kemunculan item A berkorelasi negatif terhadap kemunculan item B.

2. Jika *lift ratio*  $> 1$ , menyatakan bahwa item A dan B muncul lebih sering dari yang diharapkan. Kemunculan item A berkorelasi positif terhadap kemunculan item B.
3. Jika *lift ratio*  $= 1$ , menyatakan bahwa item A dan B munculnya hampir bersamaan sehingga tidak ada korelasi antara kedua item.

Untuk mengetahui korelasi *lift* digunakan persamaan 2.4 sebagai berikut:

$$\text{Korelasi lift} = \frac{\text{Confidence (A, B)}}{\text{Benchmark Confidence (A, B)}} \dots \dots \dots (2.4)$$

Nilai *benchmark confidence* dapat dihitung dengan persamaan 2.5 sebagai berikut:

$$\text{Benchmark Confidence} = \frac{\text{Jumlah transaksi yang mengandung B}}{\text{Total transaksi}} \dots \dots \dots (2.5)$$

## 2.2 Tinjauan Pustaka

Terdapat beberapa penelitian sebelumnya yang membahas mengenai aturan asosiasi dengan menerapkan algoritme *modified* apriori antara lain:

Penelitian yang berjudul “*Penerapan Metode Association Rule Mining Untuk Asosiasi Ulasan Terhadap Aspek Tempat Wisata Jawa Timur Park 3*” oleh Aisyatul Maulidah, Fitri A. Bachtiar mendapatkan kesimpulan bahwa algoritme apriori dapat diterapkan untuk mengelola data ulasan pengunjung *Jatim Park 3* yang berjumlah besar agar dapat menemukan hubungan antar kata yang sering muncul pada ulasan. Aturan asosiasi yang diperoleh menunjukkan hubungan antara kata kunci dengan setiap aspek layanannya, yang memiliki nilai *lift ratio* diatas 1 (Maulidah & Bachtiar, 2021).

Penelitian dengan judul “*Implementasi Data Mining Menggunakan Asosiasi Dengan Algoritme Apriori Untuk Mendapatkan Pola Rekomendasi Belanja Produk Pada Toko Avis Mobile*” oleh Kamil Erwansyah, Beni Andika, Rudi Gunawan mendapatkan kesimpulan bahwa algoritme apriori dapat menentukan pola rekomendasi belanja pada *Toko Avis Mobile* dan dihasilkan 15 aturan asosiasi (Erwansyah & Andika, 2021).

Penelitian dengan judul “*Rekomendasi Artikel Menggunakan Algoritma Apriori Pada Website*” oleh Ivan Yusuf Rahadika, Amak Yunus Eko Prasetyo, Moh. Ahsan mendapatkan kesimpulan bahwa penelitian ini menerapkan

algoritme apriori untuk memberikan rekomendasi artikel pada website Art7Wing. Terdapat 4 aturan asosiasi yang memenuhi nilai *confidence* (Yusuf & Yunus, 2019).

Penelitian selanjutnya dilakukan oleh Hita Maulidiya, Arief Jananto yang berjudul “*Asosiasi Data Mining Menggunakan Algoritme Apriori Dan FP-Growth Sebagai Dasar Pertimbangan Penentuan Paket Sembako*” mendapatkan kesimpulan bahwa dengan menerapkan metode apriori diperoleh 8 *rule* dengan waktu 0,03s dan tingkat akurasi 35% serta kombinasi yang dihasilkan yaitu 2 item. Sedangkan, dengan menggunakan metode *fp-growth* diperoleh 14 *rule* dengan waktu 0,09s dan tingkat akurasi 284% serta kombinasi yang dihasilkan yaitu 3 item. Setelah membandingkan kedua algoritme tersebut, algoritme yang paling baik dalam penentuan paket sembako adalah algoritme *fp-growth*. Meskipun memiliki waktu pengujian lebih lama tetapi nilai akurasinya tiga kali lipat dibandingkan dengan algoritme apriori (Maulidiya & Jananto, 2020).

Penelitian selanjutnya dilakukan oleh Aditya Asrorul Hidayat, Nirwana Hendrastuty, Styawati yang berjudul “*Penerapan Algoritme Apriori Pada Apotek Shaqeena Untuk Memprediksi Penjualan Berbasis Android*” mendapatkan kesimpulan bahwa sistem yang menerapkan algoritme apriori berbasis *android* berhasil dibangun dengan hasil pengujian sistem sebesar 91,64%. Oleh karena itu sistem dapat memprediksi barang yang sering dibeli dan dapat menyesuaikan stok barang sesuai dengan data transaksi penjualan barang masuk dan barang keluar (Teknologi & Sistem, 2023).

Penelitian selanjutnya dilakukan oleh Pratama Haryandi, Yuni Widiastiwi, Nurul Chamidah yang berjudul “*Penerapan Algoritme Apriori Untuk Mencari Pola Penjualan Produk Herbal (Studi Kasus : Toko Hanawan Gemilang)*” mendapatkan kesimpulan bahwa dengan menerapkan algoritme apriori diperoleh lima aturan asosiasi yang memenuhi *minsup* dan *minconf* serta memiliki nilai korelasi rasio lebih dari 1 (Haryandi et al., 2021).

Penelitian berikutnya dilakukan oleh Iftitah Dwi Ulumiyah, Herman Yuliansyah dengan judul “*Analisis Pola Asosiasi Judul Artikel Publikasi*

*Berdasarkan Data Google Scholar Menggunakan Algoritma Apriori*” mendapatkan kesimpulan bahwa algoritme apriori dapat menemukan pola pasangan kata yang sering dipasangkan secara bersamaan pada judul publikasi dosen informatika dan sistem informasi universitas ahmad Dahlan. Berdasarkan 10 kali percobaan dengan menggunakan nilai *support* dan *confidence* yang berbeda, diperoleh hasil *rules* terbanyak dan memungkinkan untuk dianalisis sebanyak 23 *rules*. Nilai tertinggi untuk *support*, *confidence* dan *lift ratio* dalam 23 *rules* tersebut adalah 13%, 88% dan 12.62 berturut-turut (Dwi Ulumiyah & Yuliansyah, 2022).

Penelitian selanjutnya dilakukan oleh Jayakumar Kaliappan, S. Mohan Sai, K. Shaily Preetham dengan judul “*Weblog And Retail Industries Analysis Using A Robust Modified Apriori Algorithm*” menyatakan bahwa algoritme *modified* apriori lebih baik dibandingkan algoritme apriori karena algoritme tersebut lebih efisien dalam menemukan aturan asosiasi dalam dataset yang besar (Kaliappan & Sai, 2019).

Penelitian selanjutnya dilakukan oleh Dianti Anggraeni, Fetty Tri Anggraeny, Intan Yuniar Purbasari dengan judul “*Penerapan Algoritma Modifikasi (Tree-Based Approach) Untuk Menemukan Pola Navigasi Pengguna Website*” menyatakan bahwa penerapan algoritme modifikasi apriori dapat menemukan pola navigasi pengguna website dan sistem yang dikembangkan mampu membentuk aturan asosiasi berdasarkan variasi nilai parameter minimum *support* dan minimum *confidence* yang diuji coba (Anggraeni & Anggraeny, 2019).

Penelitian selanjutnya dilakukan oleh T Kurniana, A. Lestari, E D Oktaviyani dengan judul “*Penerapan Algoritma Apriori Untuk Mencari Pola Transaksi Penjualan Berbasis Web Pada Café Sakuyan Side*” menyatakan bahwa algoritme apriori mampu menemukan pola transaksi penjualan Café Sakuyan Side dan aturan asosiasi yang dihasilkan dapat digunakan sebagai rekomendasi pembuatan paket menu dan *cross-selling* (Kurniana & Lestari, 2023).

Pada penelitian selanjutnya yang dilakukan oleh Poepy Tri Handayani

dan Harunur Rosyid dengan judul “*Implementasi Data Mining Menggunakan Algoritme Apriori Untuk Menentukan Aturan Asosiasi Pada Penjualan Suku Cadang Sepeda Motor Di PT. Ahass Agung Motor Malang*” menyatakan bahwa algoritme apriori dapat membantu perusahaan untuk mengatur strategi penjualan dan penataan barang. Aturan asosiasi yang diperoleh sebanyak 12 aturan dengan nilai korelasi  $lift \geq 1$ . Nilai minimum *support* dan minimum *confidence* yang digunakan pada penelitian ini adalah 30% dan 60% hal ini berdasarkan hasil pengujian yang menyatakan bahwa nilai minimum *support* dan minimum *confidence* tersebut merupakan yang terbaik karena menghasilkan *rule* terbanyak dan memiliki nilai korelasi  $lift \geq 1$  yang berarti bahwa aturan yang terbentuk valid (Handayani & Rosyid, 2021).

Berdasarkan analisis penelitian terdahulu, maka penelitian implementasi algoritme *modified* apriori untuk penentuan paket barang berdasarkan pola transaksi penjualan dengan studi kasus Toko Raya Digital dapat dilakukan.