

**“Perbandingan Algoritma Apriori dan Bayesian Network dalam  
Analisis Penjualan Retail: Pendekatan Frekuensi vs Probabilistik”**



**YOLANDA ARNASPEN  
210401069**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MUHAMMADIYAH RIAU  
PEKANBARU  
2024**

## DAFTAR ISI

<b>HALAMAN PERNYATAAN ORISINALITAS .....</b>	<b>Error! Bookmark not defined.</b>
<b>HALAMAN PENGESAHAN SKRIPSI.....</b>	<b>Error! Bookmark not defined.</b>
<b>KATA PENGANTAR.....</b>	<b>Error! Bookmark not defined.</b>
<b>ABSTRAK.....</b>	<b>Error! Bookmark not defined.</b>
<b>ABSTRACT .....</b>	<b>Error! Bookmark not defined.</b>
<b>DAFTAR ISI.....</b>	<b>2</b>
<b>DAFTAR GAMBAR.....</b>	<b>1</b>
<b>DAFTAR TABEL .....</b>	<b>2</b>
<b>BAB 1 PENDAHULUAN.....</b>	<b>2</b>
1.1 Latar Belakang .....	3
1.2 Identifikasi Masalah .....	5
1.3 Rumusan Masalah .....	6
1.4 Batasan Masalah .....	6
1.5. Tujuan Penelitian .....	6
1.6 Manfaat Penelitian .....	7
1.7 Sistematika Penulisan.....	7
<b>BAB 2 TINJAUAN PUSTAKA .....</b>	<b>9</b>
2.1 Landasan Teori.....	9
2.1.1 Data Mining dengan Analisis Penjualan Retail .....	9
2.1.2 Algoritma Apriori.....	10
2.1.3 Algoritma Bayesian Network.....	13
2.1.4 Perbandingan Pendekatan Frekuensi dan Probabilistik.....	17
2.2 Penelitian Terdahulu .....	21
2.2.1 Penemuan Promosi Mingguan .....	21
2.2.2 Pengaturan Tata Letak Produk .....	22
2.2.3 Pengelolaan Stok .....	23
<b>BAB 3 METODOLOGI PENELITIAN .....</b>	<b>29</b>
3.1 Identifikasi Masalah.....	30

3.2	Pengumpulan Data.....	30
3.3	Preprocessing Data .....	32
3.4	Penerapan Algoritma Apriori .....	34
3.5	Penerapan Algoritma Bayensian Network .....	35
3.6	Analisis dan Perbandingan Hasil.....	38
<b>BAB 4 HASIL DAN PEMBAHASAN.....</b>		<b>Error! Bookmark not defined.</b>
4.1	Subbab .....	<b>Error! Bookmark not defined.</b>
4.1.1	Import Library .....	<b>Error! Bookmark not defined.</b>
<b>BAB 5 PENUTUP .....</b>		<b>Error! Bookmark not defined.</b>
5.1	Kesimpulan .....	<b>Error! Bookmark not defined.</b>
5.2	Saran.....	<b>Error! Bookmark not defined.</b>
<b>DAFTAR PUSTAKA.....</b>		<b>39</b>

## **DAFTAR GAMBAR**

## **DAFTAR TABEL**

Tabel 2.1 Penelitian Terdahulu	4
--------------------------------	---

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Dalam industri retail yang terus berkembang, pengambilan keputusan yang didasarkan pada data menjadi sangat penting untuk meningkatkan efisiensi kondisi bisnis yang berubah-ubah. Salah satu cara untuk mencapai ini adalah dengan menggunakan teknik analisis data untuk mengidentifikasi pola perilaku konsumen. Dengan menggunakan pola dari algoritma Apriori dan Bayesian Network dalam analisis penjualan retail. Algoritma Apriori adalah salah satu teknik data mining yang paling populer dan banyak digunakan dalam Market Basket Analysis (MBA). Algoritma ini secara efektif dapat menemukan pola pembelian yang sering terjadi bersamaan dengan memanfaatkan data transaksi historis. Dengan Apriori, perusahaan dapat mengetahui kombinasi produk apa saja yang sering dibeli bersama-sama oleh konsumen. Misalnya, jika konsumen membeli item A, kemungkinan besar mereka juga akan membeli item B. Informasi ini dapat digunakan untuk mengatur tata letak produk, menyusun strategi promosi bundling, atau menentukan pengelolaan stok (Al et al., 2023). Pendekatan Apriori sangat cocok digunakan dalam konteks retail dengan transaksi yang besar, di mana tujuan utamanya adalah untuk menemukan frequent itemsets dan membentuk association rules. Misalnya, dalam sebuah toko komputer, penelitian menggunakan Apriori dapat mengungkap pola bahwa pelanggan yang membeli roti juga kemungkinan besar membeli slay, dengan confidence dan lift yang tinggi (Series, 2021). Algoritma ini membantu toko dalam membuat keputusan yang lebih baik tentang tata letak produk, promosi bundling, dan pengelolaan stok. Namun, algoritma Apriori memiliki beberapa keterbatasan. Misalnya, algoritma ini lebih fokus pada pola frekuensi tanpa memperhitungkan hal lain yang mungkin mempengaruhi keputusan pembelian, seperti diskon, promosi, atau harga produk. Hal ini bisa menyebabkan keterbatasan dalam situasi di mana kondisi pasar berubah cepat, misalnya saat ada promosi khusus atau perubahan harga secara tiba-tiba.

Sementara itu, Bayesian Network menawarkan pendekatan yang lebih canggih dan probabilistik untuk analisis penjualan retail. Berbeda dengan Apriori, Bayesian Network mempertimbangkan berbagai faktor dinamis yang memengaruhi penjualan, seperti harga, jenis promosi, lokasi produk dalam toko, dan perilaku pesaing. Bayesian Network dapat memodelkan hubungan kausal antara variabel-variabel ini, memungkinkan toko untuk memperkirakan penjualan produk dalam berbagai skenario, terutama saat ada promosi. Sebagai contoh, dalam sebuah studi kasus di supermarket, Bayesian Network digunakan untuk memprediksi penjualan produk selama periode promosi. Promosi sering kali menciptakan lonjakan permintaan yang tidak dapat diprediksi oleh model statistik konvensional. Dengan Bayesian Network, toko dapat memodelkan pengaruh promosi terhadap penjualan, serta bagaimana berbagai faktor seperti lokasi produk dan diskon mempengaruhi hasil akhir. Bayesian Network juga memungkinkan integrasi antara data kuantitatif dan kualitatif, memberikan model yang lebih komprehensif dan adaptif. Bayesian Network juga unggul dalam menangani ketidakpastian dan dapat memperbarui prediksi penjualan saat informasi baru tersedia. Ini sangat penting dalam dunia retail yang dinamis, di mana faktor-faktor seperti promosi mendadak atau perubahan perilaku pesaing dapat mempengaruhi hasil penjualan (Hamza et al., 2021).

Ketika kita membandingkan kedua pendekatan ini, kita dapat melihat bahwa keduanya memiliki kekuatan dan kelemahan masing-masing. Apriori lebih baik digunakan untuk menemukan pola pembelian frekuensi tinggi yang berulang secara historis. Ini sangat efektif dalam Market Basket Analysis, di mana fokus utama adalah menemukan produk yang sering dibeli bersamaan. Namun, Algoritma ini kurang fleksibel dalam menyesuaikan diri dengan perubahan yang dinamis, seperti promosi dan harga yang berubah-ubah. Sementara itu, Bayesian Network lebih baik dalam mengatasi ketidakpastian dan perubahan di pasar. Dengan kemampuannya untuk memodelkan hubungan kausal antara berbagai faktor yang mempengaruhi penjualan, Bayesian Network memberikan pendekatan yang lebih komprehensif dan probabilistik. Ini membuatnya ideal untuk digunakan dalam skenario di mana faktor eksternal, seperti

promosi atau perubahan harga, memainkan peran penting dalam memengaruhi keputusan pembelian konsumen (Hamza et al., 2021).

Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada fokus analisis yang lebih mendalam. Penelitian sebelumnya umumnya hanya menganalisis pola pembelian menggunakan algoritma asosiasi seperti Apriori, FP-Growth dan lain-lain. Dalam penelitian ini, dilakukan perbandingan antara algoritma Apriori dan Bayesian Network untuk menilai seberapa baik kinerja masing-masing metode dalam mengidentifikasi pola pembelian dan memprediksi penjualan. Selain itu, penelitian ini juga menganalisis faktor-faktor yang mempengaruhi promosi mingguan, pengaturan tata letak dan pengelolaan stok. Harapan dari dilakukannya penelitian ini adalah perusahaan dapat lebih memahami pola belanja pelanggan yang terjadi di toko, sehingga dapat meningkatkan keuntungan, mengembangkan strategi penjualan yang lebih efektif, dan mengelola persediaan dengan lebih baik.

## **1.2 Identifikasi Masalah**

Berdasarkan latar belakang tersebut, beberapa masalah yang dapat diidentifikasi adalah sebagai berikut:

1. Menentukan pola dari transaksi konsumen menggunakan algoritma Apriori dan Bayesian Network untuk mengidentifikasi kesamaan antar pembelian pada setiap transaksi yang dilakukan.
2. Membandingkan efektivitas algoritma Apriori dan Bayesian Network dalam analisis penjualan retail.



### 1.3 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana menentukan pola dari transaksi konsumen menggunakan algoritma Apriori dan Bayesian Network untuk mengidentifikasi kesamaan antar pembelian pada setiap transaksi yang dilakukan.
2. Bagaimana efektivitas algoritma Apriori dan Bayesian Network dalam analisis penjualan retail.

### 1.4 Batasan Masalah

Berdasarkan rumusan masalah diatas, penulis membatasi permasalahan tersebut sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah " Groceries dataset" dari Kaggle, yang terdiri dari 38765 baris data. Dataset ini dapat diakses melalui tautan [dataset](#).
2. Pola dari data transaksi konsumen yang diteliti akan digunakan untuk menghasilkan strategi promosi mingguan, pengaturan tata letak produk, serta memberikan saran dan penentuan jumlah barang yang optimal.

### 1.5. Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Menganalisis performa algoritma Apriori dan Bayesian Network dalam memprediksi penjualan retail.
2. Menentukan pola penjualan retail, termasuk promosi mingguan, pengaturan tata letak produk yang berubah setiap bulan untuk produk tertentu, dan pengelolaan stok.

## 1.6 Manfaat Penelitian

Hasil penelitian yang diharapkan penulis yakni dapat memberikan manfaat adalah sebagai berikut:

1. Menyediakan kajian komprehensif tentang perbandingan dua algoritma penting dalam data mining untuk analisis penjualan retail.
2. Membantu manajemen retail dalam membuat keputusan yang lebih baik terkait produk yang dipromosikan, penataan produk, dan pengelolaan stok berdasarkan hasil analisis pola pembelian.

## 1.7 Sistematika Penulisan

Sistematika penulisan Skripsi ini terdiri dari pokok-pokok permasalahan yang dibahas dan berikut uraian singkat masing-masing bagian.

### **BAB 1 PENDAHULUAN**

Berisi tentang pendeskripsian umum dari penelitian yang meliputi Latar Belakang, Identifikasi Masalah, Perumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian dan Sistematika Penulisan

### **BAB 2 TINJAUAN PUSTAKA**

Menguraikan teori – teori yaitu algoritma Apriori, Bayesian Network, serta studi sebelumnya terkait analisis penjualan retail.

### **BAB 3 METODOLOGI PENELITIAN**

Bagian ini menjelaskan tentang proses penelitian yang dilakukan di Universitas Muhammadiyah Riau, berdasarkan jenis penelitian, waktu penelitian, sumber data, tahapan penelitian dan analisa kebutuhan.

**BAB 4 HASIL DAN PEMBAHASAN**

Menyajikan hasil dari implementasi dan perbandingan algoritma Apriori dan Bayesian Network, serta pembahasan mengenai hasil tersebut.

**BAB 5 PENUTUP**

Berisi tentang kesimpulan dan saran yang diajukan agar dapat menjadi bahan pertimbangan.

## **BAB 2**

### **TINJAUAN PUSTAKA**

#### **2.1 Landasan Teori**

Pada era digital yang ditandai dengan perkembangan pesat teknologi informasi, analisis data menjadi salah satu kunci utama untuk meningkatkan daya saing dalam industri retail. Berbagai metode data mining digunakan untuk memahami pola pembelian konsumen dan memprediksi penjualan, di antaranya adalah algoritma Apriori dan Bayesian Network. Kedua pendekatan ini memiliki karakteristik dan pendekatan analisis yang berbeda Apriori berfokus pada frekuensi kemunculan item dalam transaksi, sedangkan Bayesian Network menggunakan model probabilistik untuk memprediksi penjualan dengan mempertimbangkan ketidakpastian.

##### **2.1.1 Data Mining dengan Analisis Penjualan Retail**

Data mining adalah proses ekstraksi informasi penting dan bermanfaat dari kumpulan data yang besar dan kompleks. Proses ini melibatkan penggunaan teknik analitis dan algoritma untuk mengidentifikasi pola, tren, dan hubungan yang tersembunyi dalam data yang mungkin tidak terlihat melalui metode analisis tradisional. Dalam beberapa dekade terakhir, data mining telah menjadi elemen penting dalam pengambilan keputusan berbasis data di berbagai bidang, termasuk pemasaran, kesehatan, keuangan, dan tentunya retail.

Dalam konteks penjualan retail, data mining digunakan untuk memahami perilaku konsumen, mengoptimalkan inventaris, meningkatkan strategi pemasaran, dan pada akhirnya meningkatkan penjualan dan profitabilitas. Dengan kemajuan teknologi dan digitalisasi, retailer memiliki akses ke sejumlah besar data transaksi pelanggan yang dapat digunakan untuk menganalisis berbagai aspek bisnis. Data ini mencakup data penjualan, data pelanggan, waktu dan lokasi pembelian, serta informasi demografis dan preferensi konsumen.

### **2.1.1.1 Penerapan Data Mining dalam Retail**

Penerapan data mining dalam industri retail melibatkan beberapa teknik:

1. **Market Basket Analysis (Analisis Keranjang Belanja):** Teknik yang paling umum digunakan dalam data mining untuk retail. Teknik ini bertujuan untuk menemukan keterkaitan antar produk yang sering dibeli secara bersamaan dalam satu transaksi. Dengan menggunakan metode ini, retailer dapat menentukan produk-produk yang dapat ditempatkan berdekatan di rak, membuat penawaran paket produk, atau menjalankan promosi bundle untuk meningkatkan penjualan. Misalnya, jika analisis menunjukkan bahwa pelanggan yang membeli roti juga pasti akan membeli selai, maka kedua produk tersebut dapat ditempatkan berdekatan untuk mendorong penjualan yang lebih tinggi. Algoritma Apriori dan FP-Growth adalah beberapa algoritma yang sering digunakan dalam Market Basket Analysis (Kantardzic, 2020) (Science, 2021).
2. **Optimasi Penempatan Produk:** Penempatan produk di toko sangat mempengaruhi perilaku belanja pelanggan. Dengan menggunakan data mining, retailer dapat menentukan lokasi yang optimal untuk produk tertentu, misalnya dengan menempatkan barang yang sering dibeli bersama di lokasi yang berdekatan. Algoritma seperti Apriori dapat digunakan untuk menganalisis transaksi historis dan menemukan produk mana yang sering dibeli secara bersamaan, sehingga retailer dapat merancang tata letak toko yang lebih efektif untuk meningkatkan penjualan (Lee, 2020).

### **2.1.2 Algoritma Apriori**

Algoritma Apriori adalah salah satu algoritma yang paling terkenal dalam data mining untuk menemukan aturan asosiasi antara item dalam sebuah dataset transaksi. Algoritma ini diperkenalkan oleh Rakesh Agrawal dan Ramakrishnan Srikant pada tahun 1994. Algoritma Apriori didasarkan pada prinsip dasar bahwa semua subset dari itemset yang sering muncul harus juga sering muncul. Dengan kata lain, jika sebuah

itemset dinyatakan sebagai itemset yang sering muncul (frequent itemset), maka semua subset dari itemset tersebut juga harus muncul dengan frekuensi yang sama atau lebih tinggi.

#### **2.1.2.1 Konsep Dasar Algoritma Apriori**

Algoritma Apriori bekerja berdasarkan dua langkah utama:

1. **Generasi Frequent Itemsets:** Langkah pertama dari algoritma ini adalah menemukan semua frequent itemsets, yaitu kumpulan item yang sering muncul bersamaan dalam dataset transaksi. Itemsets ini dihasilkan berdasarkan ambang batas support yang telah ditentukan sebelumnya, yaitu frekuensi minimum kemunculan itemset tersebut dalam seluruh transaksi. Kategori dari sebuah itemset adalah rasio jumlah transaksi yang mengandung itemset tersebut terhadap total jumlah transaksi dalam dataset.
2. **Pembentukan Aturan Asosiasi:** Setelah frequent itemsets diidentifikasi, langkah selanjutnya adalah membentuk aturan asosiasi. Aturan asosiasi dinyatakan dalam bentuk " $A \rightarrow B$ ," yang berarti jika item A ada dalam transaksi, maka item B kemungkinan besar juga ada. Pembentukan aturan asosiasi ini melibatkan penghitungan nilai support dan confidence untuk menentukan kekuatan hubungan antara item-item tersebut.

#### **2.1.2.2 Penghitungan Support dan Confidence**

Dalam algoritma Apriori, dua metrik utama digunakan untuk menilai aturan asosiasi, yaitu support dan confidence:

1. **Support:** Mengukur seberapa sering itemset muncul dalam dataset transaksi. Support dihitung dengan membagi jumlah transaksi yang mengandung itemset tersebut dengan total jumlah transaksi. Secara matematis, support dapat dinyatakan sebagai:

$$Support(A \rightarrow B) = \frac{Jumlah\ transaksi\ yang\ mengandung\ A \cup B}{Total\ Transaksi}$$

2. Confidence: Mengukur seberapa sering item B muncul dalam transaksi yang juga mengandung item A. Ini memberikan indikasi seberapa kuat hubungan antara item-item tersebut. Confidence dihitung sebagai:

$$Confidence(A \rightarrow B) = \frac{Jumlah\ transaksi\ yang\ mengandung\ A \cup B}{Jumlah\ transaksi\ yang\ mengandung\ A}$$

3. Lift: Mengukur sejauh mana item B lebih mungkin muncul dalam transaksi yang mengandung item A dibandingkan jika item A dan item B bersifat independen. Nilai lift menunjukkan apakah hubungan antar item bersifat acak atau signifikan.

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

Untuk sebuah aturan agar dianggap kuat, support dan confidence harus memenuhi atau melampaui ambang batas minimum yang telah ditetapkan oleh retailer sedangkan lift harus lebih dari 1 agar menunjukkan bahwa hubungan antara item-item tersebut tidak terjadi secara acak tetapi memiliki keterkaitan kuat. Lift yang lebih besar dari 1 juga mengindikasikan bahwa kehadiran suatu item meningkatkan kemungkinan munculnya item lain dalam transaksi.

### 2.1.2.3 Proses Kerja Algoritma Apriori

Proses kerja dari algoritma Apriori dapat diuraikan sebagai berikut:

1. Inisialisasi: Pada iterasi pertama, algoritma memeriksa setiap item secara individual untuk menentukan item mana yang memenuhi ambang batas support minimum. Item-item ini disebut sebagai item tunggal yang sering muncul (frequent 1-itemsets).
2. Generasi Kandidat Itemsets: Pada iterasi selanjutnya, algoritma menghasilkan kandidat itemsets dari frequent itemsets sebelumnya dengan menggabungkan dua itemsets yang memiliki elemen pertama yang sama. Misalnya, jika {A} dan {B} adalah frequent 1-itemsets, maka mereka dapat digabungkan menjadi {A, B} sebagai kandidat frequent 2-itemsets.

3. Pemeriksaan Kandidat: Kandidat itemsets kemudian diperiksa untuk melihat apakah mereka memenuhi ambang batas support minimum. Jika ya, maka itemsets tersebut ditambahkan ke daftar frequent itemsets.
4. Pengecekan Subset: Algoritma akan memverifikasi bahwa semua subset dari setiap kandidat itemset adalah frequent. Jika tidak, kandidat tersebut dieliminasi.
5. Iterasi Lanjut: Proses di atas diulang sampai tidak ada lagi kandidat frequent itemsets yang dapat dihasilkan.
6. Algoritma berhenti ketika tidak ada lagi frequent itemsets baru yang dapat dihasilkan.

### **2.1.3 Algoritma Bayesian Network**

Jaringan Bayesian adalah model grafis probabilistik yang digunakan untuk merepresentasikan hubungan kausal antara sekumpulan variabel. Dalam jaringan ini, variabel direpresentasikan sebagai node, dan hubungan antara variabel tersebut digambarkan dengan edge (panah) yang menunjukkan pengaruh langsung dari satu variabel terhadap yang lain. Model ini didasarkan pada teori probabilitas, terutama Teorema Bayes, yang memungkinkan kita menghitung probabilitas bersyarat suatu kejadian berdasarkan informasi yang diketahui. Jaringan Bayesian memungkinkan kita menggabungkan data kuantitatif dan informasi kualitatif untuk membuat prediksi, memperkirakan ketidakpastian, serta menganalisis hubungan sebab-akibat.

#### **2.1.3.1 Konsep Dasar Jaringan Bayesian**

Jaringan Bayesian menggabungkan dua komponen utama:

1. Struktur Jaringan: Struktur ini terdiri dari node dan edge. Node mewakili variabel (baik variabel acak maupun deterministik), sedangkan edge menggambarkan pengaruh antara variabel. Setiap edge diarahkan dari satu node ke node lain, menunjukkan bahwa node asal mempengaruhi node tujuan. Struktur jaringan dapat mencerminkan hubungan sebab-akibat atau korelasi antara variabel.



2. Distribusi Probabilitas Bersyarat: Setiap node dalam jaringan memiliki distribusi probabilitas bersyarat yang menggambarkan pengaruh dari node induknya (jika ada). Distribusi ini mencakup probabilitas bahwa suatu variabel berada dalam keadaan tertentu, mengingat keadaan dari variabel lain yang terhubung dengannya. Distribusi ini dinyatakan dengan tabel probabilitas bersyarat (Conditional Probability Table/CPT) yang menyajikan probabilitas dari semua kombinasi variabel.

### 2.1.3.2 Teorema Bayes

Teorema Bayes merupakan dasar dari jaringan Bayesian, yang memungkinkan kita menghitung probabilitas bersyarat berdasarkan probabilitas terbalik. Teorema ini dinyatakan dengan rumus:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Di sini, (  $P(A|B)$  ) adalah probabilitas dari A diberikan B, sedangkan (  $P(B|A)$  ) adalah probabilitas dari B diberikan A. (  $P(A)$  ) dan (  $P(B)$  ) adalah probabilitas awal (prior) dari A dan B masing-masing. Teorema Bayes memungkinkan kita untuk memperbarui kepercayaan terhadap suatu hipotesis (A) berdasarkan bukti baru (B).

### 2.1.3.3 Penerapan Jaringan Bayesian dalam Analisis Penjualan Retail

Dalam konteks analisis penjualan retail, jaringan Bayesian digunakan untuk memodelkan hubungan antara berbagai faktor yang mempengaruhi penjualan, seperti harga, promosi, dan lokasi produk di toko. Model ini memungkinkan retailer untuk memprediksi kemungkinan penjualan berdasarkan kombinasi berbagai faktor, serta memperhitungkan dampak dari perubahan tertentu pada penjualan. Misalnya, jaringan Bayesian dapat digunakan untuk memprediksi dampak diskon terhadap penjualan suatu produk, atau untuk mengevaluasi bagaimana perubahan harga memengaruhi volume penjualan.

Salah satu keuntungan utama jaringan Bayesian adalah kemampuannya untuk menangani data yang tidak lengkap atau memiliki ketidakpastian tinggi. Dalam banyak kasus, retailer mungkin tidak memiliki data lengkap untuk semua faktor yang memengaruhi penjualan, tetapi jaringan Bayesian masih dapat digunakan untuk membuat prediksi berdasarkan informasi yang tersedia dan memperhitungkan ketidakpastian tersebut (Hamza et al., 2021) (Kanhare & Sahni, 2021).

#### **2.1.3.4 Proses Kerja Algoritma Bayesian Network**

1. Membangun Struktur Bayesian Network

Langkah pertama adalah membangun struktur jaringan yang menunjukkan hubungan antar item yang dijelaskan oleh Item Description. Struktur ini berisi node yang mewakili item dan edges yang menunjukkan hubungan probabilistik antara item. Contohnya, jika data menunjukkan bahwa pelanggan yang membeli "kopi" juga cenderung membeli "donat", maka node "kopi" dan "donat" akan dihubungkan oleh edge.

2. Menghitung Parameter Probabilitas

Model di-fit menggunakan data transaksi yang berisi Member\_number (identitas pembeli) dan Date (tanggal transaksi) serta Item Description (deskripsi item yang dibeli). Probabilitas dihitung untuk menentukan seberapa besar peluang item tertentu dibeli jika item lain telah dibeli. Misalnya, menghitung probabilitas bahwa "teh" dibeli ketika "biskuit" sudah dibeli.

3. Melakukan Inferensi

Inferensi dilakukan untuk memprediksi pembelian item yang tidak diketahui berdasarkan data transaksi yang tersedia. Algoritma ini bisa memprediksi apakah pelanggan akan membeli item tertentu jika ada riwayat pembelian item lain. Misalnya, memprediksi apakah pelanggan akan membeli "kue" jika mereka membeli "kopi".

#### 4. Evaluasi Model dengan Precision, Recall, dan F1 Score

- a) Precision: Mengukur seberapa akurat model dalam memprediksi item yang benar-benar dibeli dibandingkan dengan semua item yang diprediksi akan dibeli. Rumusnya:

$$Precision = \frac{True\ Positives}{True\ positives + False\ Positives}$$

Contoh: Jika model memprediksi bahwa pelanggan akan membeli "kopi", tetapi hanya setengah dari prediksi itu yang benar, precision akan rendah.

- b) Recall: Mengukur seberapa baik model dapat menemukan semua item yang benar-benar dibeli. Rumusnya:

$$Recall = \frac{True\ Positives}{True\ positives + False\ Positives}$$

Contoh: Jika model berhasil memprediksi sebagian besar item yang benar-benar dibeli tetapi ada beberapa item yang terlewat, recall akan menurun.

- c) F1 Score: Merupakan rata-rata harmonis dari precision dan recall, digunakan untuk memberikan keseimbangan antara keduanya. Rumusnya:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

F1 Score tinggi menunjukkan bahwa model memiliki keseimbangan yang baik antara precision dan recall.

#### 5. Menggunakan Hasil Evaluasi

Setelah precision, recall, dan F1 score dihitung, hasil ini digunakan untuk menilai performa model. Model yang optimal membantu retailer dalam merancang promosi yang lebih tepat sasaran, mengatur tata letak produk, dan mengelola stok dengan lebih efisien berdasarkan pola pembelian yang diprediksi.

Melalui proses ini, retailer dapat memahami hubungan antar item dan meningkatkan strategi pemasaran serta penjualan berdasarkan data transaksi yang dimiliki (Gede et al., 2023).

### **2.1.3.5 Perbandingan Algoritma Apriori dengan Algoritma Bayesian**

Jika dibandingkan dengan algoritma Apriori, jaringan Bayesian menawarkan pendekatan yang lebih fleksibel dalam menangani ketidakpastian dan mempertimbangkan berbagai faktor. Sementara algoritma Apriori berfokus pada penemuan pola yang sering muncul (frequent itemsets) tanpa memperhitungkan probabilitas bersyarat antar item, jaringan Bayesian mampu memodelkan hubungan sebab-akibat dan menghitung dampak dari berbagai faktor yang saling berkaitan.

Secara keseluruhan, jaringan Bayesian adalah algoritma yang kuat untuk analisis probabilistik dalam berbagai konteks, terutama di industri retail. Dengan menggabungkan data historis dan informasi tambahan, model ini dapat memberikan prediksi yang lebih akurat dan membantu dalam pengambilan keputusan yang lebih baik.

### **2.1.4 Perbandingan Pendekatan Frekuensi dan Probabilistik**

Pendekatan frekuensi dan probabilistik adalah dua metode utama dalam data mining yang digunakan untuk menganalisis data transaksi dalam retail, khususnya dalam menemukan pola pembelian dan memprediksi penjualan. Meskipun kedua pendekatan ini bertujuan untuk mengidentifikasi pola dan hubungan antar item, keduanya memiliki perbedaan mendasar dalam cara bekerja, kelebihan, dan kelemahannya.

#### **2.1.4.1 Pendekatan Frekuensi**

Pendekatan frekuensi, yang banyak digunakan dalam algoritma seperti Apriori, bertujuan untuk menemukan itemset atau kelompok item yang sering muncul bersama dalam transaksi. Metode ini didasarkan pada konsep frekuensi atau jumlah kemunculan suatu itemset dalam dataset. Dengan menetapkan ambang batas minimum untuk support (frekuensi minimum kemunculan itemset), pendekatan ini mengidentifikasi

itemset yang dianggap penting atau relevan berdasarkan seberapa sering item tersebut muncul bersama.

#### Kelebihan Pendekatan Frekuensi:

1. Mudah Dipahami dan Diimplementasikan: Pendekatan frekuensi cukup sederhana dalam konsep dan dapat dengan mudah diimplementasikan untuk berbagai jenis dataset transaksi. Algoritma seperti Apriori atau FP-Growth menggunakan prinsip dasar dari penghitungan frekuensi itemset untuk menemukan asosiasi yang relevan.
2. Cepat untuk Dataset Kecil hingga Sedang: Dalam situasi dengan jumlah transaksi yang terbatas dan item yang tidak terlalu banyak, pendekatan frekuensi dapat sangat efisien dalam menemukan pola yang signifikan.
3. Dapat Menghasilkan Aturan yang Kuat dan Mudah Diinterpretasikan: Aturan asosiasi yang dihasilkan dari pendekatan ini umumnya langsung dan mudah dipahami oleh pengguna, seperti "Jika membeli A, maka kemungkinan besar juga akan membeli B".

#### Kekurangan Pendekatan Frekuensi:

1. Masalah Skalabilitas pada Dataset Besar: Ketika ukuran dataset menjadi sangat besar, jumlah kombinasi item yang harus diperiksa meningkat secara eksponensial, yang dapat menyebabkan waktu komputasi yang lama dan kebutuhan memori yang tinggi.
2. Mengabaikan Informasi Probabilistik dan Ketidakpastian: Pendekatan ini tidak memperhitungkan ketidakpastian atau variabel kontekstual yang dapat memengaruhi pola pembelian. Sebagai contoh, kondisi musiman, perubahan harga, atau promosi tidak dapat diperhitungkan dengan baik dalam model berbasis frekuensi.
3. Cenderung Menghasilkan Banyak Aturan yang Kurang Relevan: Karena hanya didasarkan pada frekuensi kemunculan itemset, algoritma ini sering menghasilkan aturan asosiasi yang banyak, tetapi tidak semuanya relevan atau memberikan wawasan yang berguna bagi pengguna.

#### 2.1.4.2 Pendekatan Probabilistik

Pendekatan probabilistik, yang diterapkan dalam algoritma seperti jaringan Bayesian, berfokus pada penghitungan probabilitas bersyarat dan ketidakpastian antar variabel. Dalam pendekatan ini, bukan hanya frekuensi kemunculan item yang dipertimbangkan, tetapi juga probabilitas bahwa suatu item akan muncul berdasarkan item lain yang ada dalam transaksi. Dengan mempertimbangkan probabilitas bersyarat, pendekatan probabilistik dapat menghasilkan prediksi yang lebih tepat dan dapat disesuaikan dengan kondisi tertentu.

Kelebihan Pendekatan Probabilistik:

1. Mempertimbangkan Ketidakpastian dan Variabilitas: Pendekatan probabilistik memungkinkan kita untuk memasukkan ketidakpastian ke dalam model, yang sangat berguna dalam situasi di mana data tidak lengkap atau ada banyak faktor yang mempengaruhi perilaku konsumen. Misalnya, model ini dapat mengakomodasi informasi yang tidak pasti tentang faktor eksternal seperti cuaca, promosi, atau perubahan harga.
2. Kemampuan untuk Menangani Data yang Tidak Lengkap: Karena pendekatan probabilistik tidak memerlukan data yang lengkap untuk semua variabel, metode ini lebih fleksibel dalam membuat prediksi meskipun hanya sebagian informasi yang tersedia.
3. Modeling Sebab-Akibat: Pendekatan probabilistik memungkinkan pengguna untuk memodelkan hubungan sebab-akibat antara variabel, bukan hanya korelasi sederhana. Hal ini memudahkan dalam mengevaluasi dampak dari perubahan pada satu variabel terhadap variabel lain.

Kekurangan Pendekatan Probabilistik:

1. Kompleksitas Komputasi yang Lebih Tinggi: Pendekatan ini umumnya lebih kompleks daripada pendekatan frekuensi, baik dari sisi pembangunan model maupun pemrosesan datanya. Perhitungan probabilitas bersyarat, terutama dalam

model yang besar dengan banyak variabel, membutuhkan lebih banyak sumber daya komputasi.

2. Kesulitan dalam Menentukan Probabilitas Awal (Prior): Untuk beberapa kasus, menentukan nilai probabilitas awal untuk setiap variabel bisa menjadi tantangan, terutama jika tidak ada cukup data historis yang mendukung estimasi tersebut.
3. Tergantung pada Struktur Model: Keakuratan pendekatan probabilistik sangat tergantung pada struktur model yang digunakan. Jika struktur yang dibangun tidak mencerminkan hubungan yang benar antara variabel, hasil prediksinya dapat menyesatkan.

#### **2.1.4.3 Perbandingan Antara Pendekatan Frekuensi dan Probabilistik**

Perbedaan utama antara pendekatan frekuensi dan probabilistik terletak pada bagaimana pola dan hubungan antar item ditentukan:

1. Basis Penentuan Pola: Pendekatan frekuensi menentukan pola berdasarkan kemunculan bersama item dalam dataset, sedangkan pendekatan probabilistik menggunakan informasi probabilitas untuk memperkirakan kemungkinan terjadinya suatu item atau kejadian berdasarkan informasi lain.
2. Penggunaan Data Tambahan: Dalam pendekatan frekuensi, hanya data transaksi yang digunakan untuk menemukan pola. Sebaliknya, pendekatan probabilistik dapat menggabungkan berbagai jenis data, seperti data kuantitatif dan kualitatif, serta memperhitungkan variabel kontekstual atau data eksternal yang memengaruhi pola pembelian.
3. Kemampuan Menangani Ketidakpastian: Pendekatan probabilistik lebih unggul dalam menangani ketidakpastian dan data yang tidak lengkap dibandingkan dengan pendekatan frekuensi. Hal ini disebabkan oleh kemampuannya dalam menggabungkan informasi probabilistik dengan data yang ada, sementara pendekatan frekuensi hanya mengandalkan frekuensi kemunculan item.
4. Output dan Interpretasi: Hasil dari pendekatan frekuensi sering berupa aturan asosiasi yang dapat langsung digunakan, seperti "Jika membeli X, maka

kemungkinan besar juga membeli Y." Sementara itu, pendekatan probabilistik memberikan output berupa distribusi probabilitas yang lebih komprehensif dan memerlukan interpretasi yang lebih mendalam untuk menentukan implikasi bisnis.

#### **2.1.4.5 Kapan Menggunakan Pendekatan Frekuensi atau Probabilistik?**

Pemilihan antara pendekatan frekuensi dan probabilistik tergantung pada kebutuhan spesifik dan karakteristik dataset:

1. Pendekatan Frekuensi lebih cocok digunakan ketika dataset besar tersedia, dan kita ingin menemukan pola item yang sering muncul tanpa mempertimbangkan banyak faktor eksternal.
2. Pendekatan Probabilistik sebaiknya digunakan ketika ada banyak faktor kontekstual yang perlu diperhitungkan, atau ketika data yang tersedia tidak lengkap dan mengandung ketidakpastian.

## **2.2 Penelitian Terdahulu**

Penelitian terdahulu memberikan landasan penting dalam memahami penerapan berbagai metode analisis data untuk meningkatkan kinerja bisnis retail. Berbagai studi telah dilakukan untuk mengeksplorasi penggunaan algoritma seperti Apriori dan Jaringan Bayesian dalam analisis penjualan, dengan fokus pada identifikasi pola pembelian, prediksi penjualan, dan pengoptimalan strategi pemasaran dengan penerapan dalam menemukan promosi mingguan, pengaturan tata letak produk, serta pengelolaan stok. Bagian ini akan membahas beberapa penelitian yang relevan terkait penerapan kedua algoritma tersebut.

### **2.2.1 Penemuan Promosi Mingguan**



Algoritma Apriori, dengan pendekatan berbasis frekuensi, sangat efektif untuk menemukan pola pembelian yang sering terjadi dalam periode tertentu. Ini memungkinkan retailer untuk menentukan produk mana yang sering dibeli bersama selama promosi mingguan, sehingga promosi dapat difokuskan pada produk-produk tersebut untuk meningkatkan penjualan (Noviana et al., 2023) (Market basket analysis ...)(tang2020).

Sementara itu, Jaringan Bayesian memberikan pendekatan yang lebih dinamis dengan mempertimbangkan berbagai faktor kontekstual, seperti musim, promosi sebelumnya, dan tren pasar. Hal ini memungkinkan retailer untuk memprediksi dampak promosi terhadap penjualan dengan lebih akurat, terutama ketika ada perubahan kondisi eksternal yang signifikan. Jaringan Bayesian dapat menggabungkan data historis dan informasi baru untuk memperbarui prediksi secara berkala (Forecasting sales)(kanhere2021).

### **2.2.2 Pengaturan Tata Letak Produk**

Penelitian menunjukkan bahwa Algoritma Apriori dapat digunakan untuk menganalisis asosiasi antar produk dan menemukan produk yang sering dibeli bersama, sehingga produk tersebut dapat ditempatkan berdekatan di rak. Ini membantu meningkatkan pembelian impulsif dan optimasi tata letak toko, yang berdampak positif pada penjualan produk terkait(tang2020)(yingzhuo2021).

Di sisi lain, Jaringan Bayesian memungkinkan analisis yang lebih mendalam terkait dampak perubahan tata letak terhadap penjualan berdasarkan data historis dan faktor-faktor lain seperti penempatan promosi atau lokasi spesifik di toko. Dengan memodelkan hubungan antara tata letak produk dan volume penjualan, Jaringan Bayesian dapat memberikan rekomendasi penempatan produk yang lebih efektif seiring perubahan kondisi setiap bulan(Forecasting sales)(tang2020).

### 2.2.3 Pengelolaan Stok

Algoritma Apriori dapat membantu dalam menentukan item yang sering dibeli bersamaan sehingga stok produk tersebut dapat dikelola dengan lebih baik, seperti memastikan ketersediaan barang yang terkait untuk mencegah stockout atau overstock (ogurtsov2019).

Di sisi lain, Jaringan Bayesian memberikan pendekatan yang lebih prediktif dalam pengelolaan stok. Dengan mempertimbangkan ketidakpastian dan variabilitas dalam data permintaan, Jaringan Bayesian dapat memberikan perkiraan yang lebih akurat tentang kebutuhan stok di masa depan, memungkinkan perencanaan stok yang lebih efisien dan mengurangi biaya penyimpanan berlebih atau kekurangan stok (Forecasting sales)(kanhere2021).

Tabel 2. 1 Penelitian Terdahulu

No	Nama Peneliti	Judul Penelitian	Tujuan Penelitian	Hasil dan Kesimpulan
1	Kanhere et al. (2021)	Clustering Based Approach to Enhance Association Rule Mining	Mengurangi ukuran dataset menggunakan teknik clustering untuk mengoptimalkan algoritma Apriori.	Pendekatan clustering berhasil mengurangi data hingga 7% dari total item, mempercepat proses komputasi.
2	Hamza et al. (2021)	Forecasting Sales with Bayesian Networks	Memodelkan dampak promosi terhadap penjualan dengan menggunakan jaringan Bayesian.	Jaringan Bayesian lebih akurat dibandingkan dengan metode tradisional dalam memprediksi

No	Nama Peneliti	Judul Penelitian	Tujuan Penelitian	Hasil dan Kesimpulan
				penjualan selama promosi.
3	Ünvan (2020)	Market Basket Analysis with Association Rules	Menggunakan algoritma Apriori dan FP-Growth untuk analisis keranjang belanja di supermarket.	Algoritma FP-Growth memberikan hasil yang lebih baik karena efisiensi dalam menemukan aturan asosiasi.
4	Ogurtsov & Dorrer (2019)	Application of Association Rules Learning for Studying Store History	Menggunakan aturan asosiasi untuk menganalisis riwayat transaksi di jaringan retail besar.	Pengelompokan barang berhasil meningkatkan efisiensi perencanaan inventaris di toko.
5	Tang et al. (2020)	Apply Apriori Algorithm in Supermarket Layout Research	Menerapkan algoritma Apriori untuk mengoptimalkan tata letak produk di supermarket.	Penataan produk berdasarkan hasil analisis Apriori meningkatkan peluang pembelian impulsif.
6	Bassamzadeh & Ghanem (2017)	Bayesian Approach for Forecasting Electricity Demand	Memodelkan permintaan listrik dengan menggabungkan data historis dan informasi kualitatif.	Jaringan Bayesian menghasilkan prediksi yang lebih akurat dibandingkan

No	Nama Peneliti	Judul Penelitian	Tujuan Penelitian	Hasil dan Kesimpulan
				dengan model regresi sederhana.
7	Arora et al. (2019)	Using Bayesian Networks for Risk Prediction in Healthcare	Memprediksi risiko penyakit berdasarkan data medis dan informasi risiko terkait.	Jaringan Bayesian dapat memperkirakan risiko dengan lebih baik meskipun data yang tersedia tidak lengkap.
8	Yu et al. (2011)	Comparison of Bayesian Networks and Regression Models in Retail Forecasting	Membandingkan jaringan Bayesian dengan model regresi dalam memprediksi penjualan retail.	Jaringan Bayesian lebih unggul dalam menangani variabilitas dan faktor eksternal yang mempengaruhi penjualan.
9	Yingzhuo & Xuewen (2021)	Research on Community Consumer Behavior Based on Association Rules	Menganalisis perilaku konsumen di supermarket komunitas menggunakan algoritma Apriori.	Algoritma yang dioptimalkan meningkatkan efisiensi dalam mendeteksi pola pembelian.
10	Abolghasemi & Alvandi (2021)		Memodelkan elastisitas penjualan menggunakan jaringan Bayesian	Bayesian Networks berhasil memprediksi pengaruh

No	Nama Peneliti	Judul Penelitian	Tujuan Penelitian	Hasil dan Kesimpulan
			berdasarkan perubahan harga.	perubahan harga dengan lebih baik dibandingkan model lain.
11	Yu et al. (2011)	Comparison of Bayesian Networks and Regression Models in Retail Forecasting	Membandingkan jaringan Bayesian dengan model regresi dalam prediksi penjualan retail.	Jaringan Bayesian lebih efektif menangani variabilitas data dan faktor eksternal.
12	Sun & Huang (2020)	Improving Product Placement Using Apriori Analysis	Menggunakan Apriori untuk menata produk di supermarket berdasarkan aturan asosiasi.	Hasil analisis membantu merancang tata letak yang meningkatkan peluang pembelian impulsif.
13	Reski Noviana et al. (2023)	Market Basket Analysis Menggunakan Algoritma Apriori dan FP Growth untuk Menentukan Pola Pembelian Konsumen	Mengidentifikasi pola pembelian konsumen dengan menggunakan algoritma Apriori dan FP-Growth untuk meningkatkan strategi penjualan.	Algoritma Apriori menghasilkan kombinasi yogurt, sausage, whole milk dengan support 0.00147, confidence 25.58%, lift 1.61986. Algoritma FP-

No	Nama Peneliti	Judul Penelitian	Tujuan Penelitian	Hasil dan Kesimpulan
				Growth lebih cepat (0.15892 detik dibanding Apriori 3.1765 detik).
14	Wang et al. (2019)	Enhanced Market Basket Analysis Using Modified Apriori	Mengembangkan algoritma Apriori untuk meningkatkan efisiensi analisis keranjang belanja.	Algoritma yang dimodifikasi mempercepat proses pencarian frequent itemsets hingga 30%.

Berdasarkan hasil perbandingan Algoritma Apriori dan Jaringan Bayesian dari berbagai penelitian terdahulu, dapat disimpulkan bahwa Algoritma Apriori lebih unggul dalam mengidentifikasi pola pembelian yang sering terjadi dan sangat cocok digunakan untuk promosi mingguan dan pengaturan tata letak produk di toko. Algoritma ini dapat dengan cepat menemukan kombinasi produk yang sering dibeli bersama, sehingga retailer dapat merancang strategi promosi dan penempatan produk yang lebih efektif. Namun, Apriori cenderung terbatas dalam menangani ketidakpastian dan variabilitas data, serta kurang fleksibel ketika harus memperhitungkan berbagai faktor kontekstual yang memengaruhi penjualan. Sebaliknya, Jaringan Bayesian menawarkan keunggulan dalam memodelkan hubungan antar variabel dan menangani ketidakpastian. Algoritma ini mampu memprediksi dampak perubahan faktor eksternal, seperti promosi atau harga, dan memberikan perkiraan yang lebih akurat untuk pengelolaan stok serta penyesuaian strategi pemasaran. Jaringan Bayesian dapat mempertimbangkan berbagai faktor yang memengaruhi penjualan secara bersamaan dan memperbarui prediksi berdasarkan informasi baru, membuatnya lebih dinamis dibandingkan Apriori.

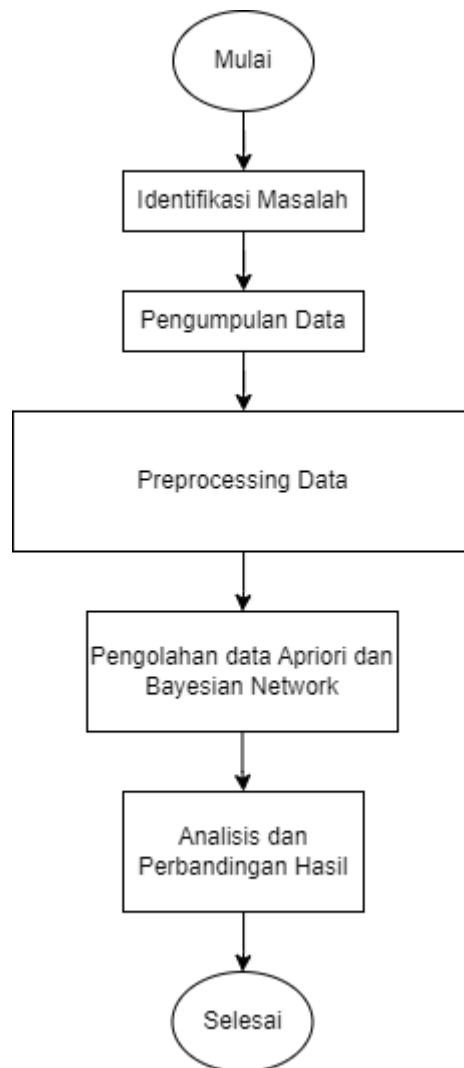
Hasil dari penelitian terdahulu menunjukkan bahwa dengan mengombinasikan kedua pendekatan ini, retailer dapat mengoptimalkan strategi penjualan dan

manajemen inventaris secara lebih menyeluruh. Algoritma Apriori dapat digunakan untuk menemukan pola pembelian frekuensi tinggi, sementara Jaringan Bayesian dapat dimanfaatkan untuk melakukan prediksi yang lebih dinamis dan menangani perubahan berdasarkan faktor-faktor kontekstual. Dengan pendekatan ini, perusahaan dapat meningkatkan efisiensi operasional, mengurangi risiko kekurangan atau kelebihan stok, dan merancang strategi promosi yang lebih efektif.

### **BAB 3**

## **METODOLOGI PENELITIAN**

Bab ini memaparkan tahapan-tahapan proses yang dilakukan selama penelitian tugas akhir untuk memberikan informasi yang tepat dan relevan terhadap permasalahan yang dikaji. Metodologi penelitian ini mencakup kerangka kerja sistematis dari awal penelitian hingga tercapainya tujuan yang diharapkan. Tahapan ini akan menjelaskan secara rinci setiap langkah yang diambil dalam penelitian. Seperti yang terlihat pada gambar berikut.



Gambar 3. 1 Kerangka Kerja



### 3.1 Identifikasi Masalah

Identifikasi masalah dalam penelitian ini berfokus pada pola dari data transaksi algoritma Apriori dan Bayesian Network serta membandingkan algoritma Apriori dan Bayesian Network dalam menganalisis pola penjualan retail. Apriori bekerja berdasarkan frekuensi kemunculan item, cocok untuk menemukan kombinasi produk yang sering dibeli bersama, sedangkan Bayesian Network memodelkan hubungan probabilistik, memperhitungkan ketidakpastian dan perubahan kondisi seperti promosi atau harga. Data transaksi retail dari Kaggle digunakan, mencakup atribut seperti Member\_Number, Date, dan Item Description, untuk menemukan pola pembelian dan memprediksi penjualan. Penelitian ini bertujuan untuk menentukan algoritma mana yang lebih efektif dalam mengoptimalkan strategi penjualan dan pengelolaan stok.

### 3.2 Pengumpulan Data

Tahapan ini mencakup pengumpulan data transaksi retail yang relevan dari berbagai sumber, dalam hal ini data diambil dari Kaggle yang berjudul " Groceries dataset" dengan link [dataset](#). Data yang dikumpulkan meliputi:

1. Member\_number (untuk identifikasi pelanggan),
2. Date (untuk melacak waktu terjadinya transaksi), dan
3. Item Description (untuk mencatat produk yang dibeli dalam setiap transaksi).

Tabel 3. 1 Dataset

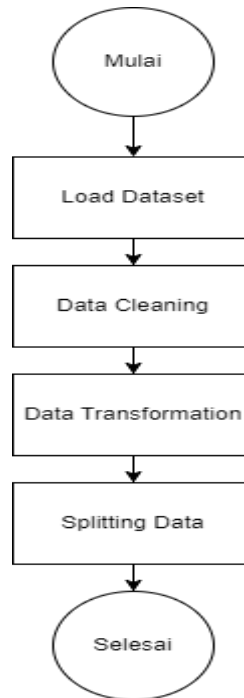
Member_number	Date	itemDescription
1808	21-07-2015	tropical fruit
2552	5/1/2015	whole milk
2300	19-09-2015	pip fruit
1187	12/12/2015	other vegetables
3037	1/2/2015	whole milk

4941	14-02-2015	rolls/buns
4501	8/5/2015	other vegetables
3803	23-12-2015	pot plants
2762	20-03-2015	whole milk
4119	12/2/2015	tropical fruit
1340	24-02-2015	citrus fruit
2193	14-04-2015	beef
1997	21-07-2015	frankfurter
4546	3/9/2015	chicken
4736	21-07-2015	butter

Data ini akan digunakan sebagai bahan dasar untuk analisis menggunakan kedua algoritma yang akan diuji dalam penelitian.

### 3.3 Preprocessing Data

Preprocessing data yang dilakukan dalam penelitian Perbandingan Algoritma Apriori dan Bayesian Network dalam Analisis Penjualan Retail.



Gambar 3. 2 Preprocessing Data

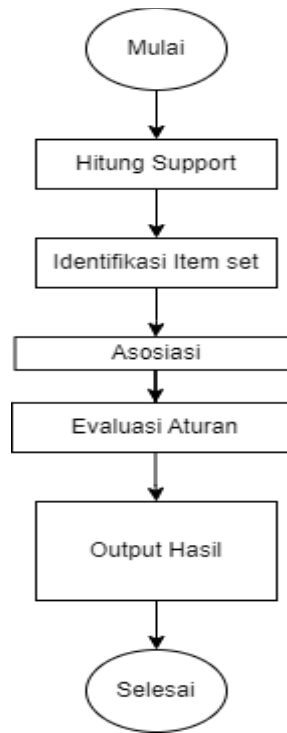
Berikut adalah penjelasan dari setiap tahap:

1. Start (Mulai): Penelitian dimulai dari memuat data transaksi penjualan retail yang akan dianalisis.
2. Load Dataset (Memuat Data Transaksi Penjualan Retail): Pada tahap ini, data transaksi retail yang telah dikumpulkan dari Kaggle, dimuat ke dalam sistem untuk dianalisis lebih lanjut. Data yang dimuat mencakup atribut seperti Member\_number, Date, Item Description.
3. Data Cleaning (Pembersihan Data):
  - a. Menghilangkan Duplikasi: Langkah ini bertujuan untuk menghapus data transaksi yang tercatat lebih dari satu kali (duplikat), sehingga hasil analisis tetap akurat dan objektif.

- b. Menangani Data Hilang: Transaksi yang tidak lengkap atau atribut yang hilang diperbaiki atau dihilangkan agar dataset menjadi bersih dan siap untuk analisis lebih lanjut.
- 4. Data Transformation (Transformasi Data):
  - a. Apriori: Data diubah menjadi item-set, yang berarti setiap transaksi diwakili oleh kumpulan produk yang dibeli bersama. Format ini memudahkan algoritma Apriori dalam menemukan pola pembelian yang sering terjadi.
  - b. Bayesian Network: Data diolah menjadi bentuk yang dapat memodelkan hubungan probabilistik antar variabel, seperti waktu transaksi, promosi, atau item yang dibeli.
- 5. Splitting Data (Pemecahan Data): Setelah transformasi selesai, data dipecah menjadi dua bagian:
  - a. Apriori: Data yang telah diproses dan diubah menjadi item-set digunakan untuk menemukan pola pembelian frekuensi tinggi.
  - b. Bayesian Network: Data yang relevan untuk analisis probabilistik digunakan untuk membangun model jaringan Bayesian, memperhitungkan berbagai faktor seperti ketidakpastian dan hubungan antar produk.
- 6. End (Selesai): Tahapan preprocessing selesai setelah data siap untuk digunakan oleh kedua algoritma, dan analisis selanjutnya bisa dilakukan untuk menemukan pola penjualan atau prediksi berdasarkan metode yang digunakan.

Flowchart ini membantu merinci tahapan preprocessing agar data transaksi siap untuk dianalisis oleh kedua algoritma tersebut, baik untuk menemukan pola pembelian (Apriori) maupun prediksi probabilistik (Bayesian Network).

### 3.4 Penerapan Algoritma Apriori



Gambar 3. 3 Flowchart Apriori

Penerapan algoritma Apriori dalam flowchart dimulai dengan menggunakan data yang telah diproses untuk menemukan pola pembelian yang sering terjadi. Algoritma ini menghitung support untuk mengidentifikasi item atau kombinasi item yang muncul dalam dataset, sehingga menghasilkan item-set frekuensi tinggi. Selanjutnya, algoritma membuat aturan asosiasi berdasarkan kombinasi produk yang ditemukan, dengan parameter seperti confidence dan lift. Hasil dari penerapan algoritma ini memberikan informasi tentang pola pembelian, seperti "Jika produk A dibeli, maka produk B juga akan dibeli dengan probabilitas tertentu." Informasi ini dapat digunakan untuk merancang strategi promosi mingguan, mengatur tata letak produk di toko, dan mengoptimalkan pengelolaan stok, sehingga retailer dapat meningkatkan penjualan dan memanfaatkan pola pembelian yang teridentifikasi.

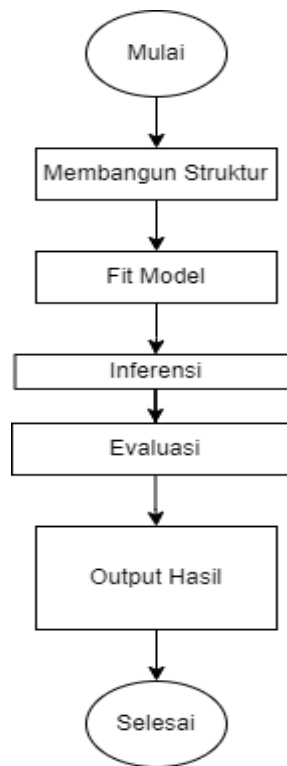
1. Start (Mulai): Setelah Data yang telah di preprocessing, maka diterapkannya lah algoritma Apriori

2. Hitung Support: Di tahap ini, algoritma menghitung support untuk setiap item dalam dataset. Support mengukur seberapa sering item atau kombinasi item muncul dalam transaksi, dan ini penting untuk menentukan item-set yang signifikan.
3. Identifikasi Itemsets: Setelah support dihitung, algoritma mengidentifikasi frequent itemsets, yaitu item-set yang memenuhi ambang batas support yang telah ditentukan. Langkah ini penting untuk menemukan pola pembelian yang sering terjadi.
4. Association Rules: Pada tahap ini, algoritma menghasilkan aturan asosiasi berdasarkan frequent itemsets yang ditemukan. Aturan ini menunjukkan hubungan antara produk, contohnya "Jika produk A dibeli, maka produk B juga cenderung dibeli".
5. Evaluate Rules (Evaluasi Aturan): Hasil aturan asosiasi dievaluasi menggunakan metrik seperti:
  - a. Support: Seberapa sering aturan tersebut benar di seluruh transaksi.
  - b. Confidence: Kemungkinan produk B dibeli ketika produk A dibeli.
  - c. Lift: Mengukur seberapa banyak probabilitas item B dibeli ketika item A dibeli dibandingkan dengan probabilitas item B dibeli secara independen.
6. Output Results (Hasil): Setelah evaluasi, hasil analisis ditampilkan. Ini mencakup pola-pola pembelian yang diidentifikasi dan aturan asosiasi yang dapat digunakan untuk strategi pemasaran.
7. End (Selesai): Proses penerapan algoritma Apriori berakhir di sini. Hasil dan wawasan yang diperoleh dari analisis dapat digunakan untuk membuat keputusan strategis dalam meningkatkan penjualan dan pengelolaan produk di retail.

### **3.5 Penerapan Algoritma Bayensian Network**

Penerapan algoritma Bayesian Network dimulai dengan membangun struktur jaringan yang menggambarkan hubungan probabilistik antara produk dalam dataset.

Setelah struktur ditentukan, model di-fit menggunakan data untuk menghitung parameter probabilitas antar variabel. Selanjutnya, inferensi dilakukan untuk memprediksi nilai variabel yang tidak diketahui, seperti kemungkinan produk B dibeli jika produk A sudah dibeli. Hasil dari analisis ini dievaluasi untuk menilai akurasi model, dan informasi yang diperoleh digunakan untuk merancang promosi, mengatur tata letak produk, serta mengelola stok dengan lebih efektif. Dengan demikian, Bayesian Network membantu retailer memahami hubungan antar produk dan memanfaatkan pola pembelian untuk meningkatkan penjualan.



Gambar 3. 4 Flowchart Bayesian

1. Start (Mulai): Setelah data telah diproses, langkah selanjutnya adalah membangun dan menerapkan model Bayesian Network untuk menganalisis hubungan antar variabel dalam dataset.
2. Membangun Struktur: Di tahap ini, struktur jaringan Bayesian dibangun dengan menentukan node dan edges. Node mewakili variabel atau item, sedangkan edges menunjukkan hubungan atau ketergantungan antara variabel. Misalnya, jika produk A sering dibeli bersama produk B, maka akan ada edge dari node A ke

node B. Struktur ini bisa ditentukan berdasarkan pengetahuan domain atau analisis data awal.

3. Fit Model: Setelah struktur jaringan ditetapkan, langkah berikutnya adalah “fit” model dengan menggunakan data yang ada. Pada tahap ini, algoritma menghitung parameter probabilitas untuk setiap node dalam jaringan, yang menjelaskan hubungan antar variabel berdasarkan data yang digunakan. Pustaka seperti pgmpy di Python memungkinkan untuk melakukan ini dengan estimator seperti BayesianEstimator.
4. Inferensi: Dengan model yang telah terlatih, tahap ini melakukan inferensi untuk memprediksi nilai variabel yang tidak diketahui berdasarkan informasi yang ada. Inferensi dapat dilakukan menggunakan metode seperti Variable Elimination untuk menentukan probabilitas dari suatu item (misalnya, produk) diberikan bahwa item lain telah dibeli. Misalnya, probabilitas bahwa produk B dibeli jika produk A sudah dibeli.
5. Evaluasi Hasil: Hasil dari inferensi dievaluasi untuk memahami seberapa baik model dalam memprediksi hubungan antar variabel. Ini termasuk analisis kesalahan dan membandingkan prediksi dengan data nyata untuk menilai akurasi model. Metrik yang digunakan untuk evaluasi bisa mencakup akurasi, precision, dan recall, tergantung pada tujuan analisis.
6. Output Results (Hasil): Setelah evaluasi, hasil analisis ditampilkan. Ini mencakup hubungan probabilistik yang diidentifikasi antara produk dan pola pembelian yang dapat digunakan untuk strategi pemasaran. Misalnya, model dapat menunjukkan bahwa jika produk A dibeli, maka ada kemungkinan tertentu bahwa produk B juga akan dibeli, membantu retailer dalam merancang promosi yang lebih efektif.
7. End (Selesai): Proses penerapan algoritma Bayesian Network berakhir di sini. Hasil dan wawasan yang diperoleh dari analisis dapat digunakan untuk membuat keputusan strategis dalam pengelolaan stok, penempatan produk, dan promosi, sehingga retailer dapat lebih memahami perilaku konsumen dan memanfaatkan pola pembelian yang teridentifikasi untuk meningkatkan penjualan.



### **3.6 Analisis dan Perbandingan Hasil**

Setelah menerapkan kedua algoritma, hasil dari Apriori dan Bayesian Network dianalisis dan dibandingkan berdasarkan akurasi pola yang ditemukan serta waktu komputasi. Analisis ini bertujuan untuk menentukan algoritma yang paling tepat diterapkan dalam konteks retail, terutama dalam aspek bundling promosi, pengelolaan stok, dan penempatan produk di toko. Perbandingan dilakukan untuk melihat sejauh mana setiap algoritma mampu mengidentifikasi hubungan antar produk secara efektif, dengan mempertimbangkan efisiensi waktu dan sumber daya komputasi. Hasil analisis ini akan digunakan untuk merekomendasikan algoritma yang paling sesuai dengan kebutuhan operasional retail, sehingga dapat mendukung pengambilan keputusan yang lebih optimal dalam strategi pemasaran dan manajemen inventaris.

## DAFTAR PUSTAKA

- Al, F., Hassan, M., Human, A., Haque, A., & Charles, N. (2023). ScienceDirect ScienceDirect An Apriori Association Rule Analysis to Suicidal Behaviour An Apriori Association Rule Analysis to detect Health Algorithm-Based An Apriori Algorithm-Based Association Rule Analysis to detect Suicidal Behaviour Human Suicidal Behaviour. *Procedia Computer Science*, 219, 1279–1288. <https://doi.org/10.1016/j.procs.2023.01.412>
- Gede, P., Cipta, S., Informatika, J. T., Bayes, N., & Penjualan, P. (2023). *Penerapan Algoritma Naive Bayes Untuk Prediksi Penjualan Produk Terlaris Pada CV Akusara Jaya Abadi*. 10(4), 518–534.
- Hamza, M., Abolghasemi, M., & Alvandi, A. O. (2021). *Forecasting sales with Bayesian networks : a case study of a supermarket product in the presence of promotions*. December, 974–980.
- Kanhere, S., & Sahni, A. (2021). Clustering Based Approach to Enhance Association Rule Mining. *PROCEEDING OF THE 28TH CONFERENCE OF FRUCT ASSOCIATION*, 142–150.
- Kantardzic, M. (2020). *APPENDIX A INFORMATION ON. 1*, 559–587.
- Lee, P. (2020). *Apply Apriori Algorithm in Supermarket Layout*. 521–524. <https://doi.org/10.1109>
- Noviana, R., Hermawan, A., & Avianto, D. (2023). *Market Basket Analysis Menggunakan Algoritma Apriori dan FP Growth untuk Menentukan Pola Pembelian Konsumen*. 7, 1474–1482. <https://doi.org/10.30865/mib.v7i3.6304>
- Science, C. (2021). *Research on Community Consumer Behavior Based on Association Rules Analysis*. Icsp, 6–9.
- Series, C. (2021). Market basket analysis using apriori algorithm to find consumer

patterns in buying goods through transaction data ( case study of Mizan computer retail stores ) Market basket analysis using apriori algorithm to find consumer patterns in buying goods throu. *Market Basket Analysis Using Apriori Algorithm to Find Consumer Patterns in Buying Goods through Transaction Data (Case Study of Mizan Computer Retail Stores) To*, 1–13.  
<https://doi.org/10.1088/1742-6596/1722/1/012020>