

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Penelitian Terkait

Penelitian mengenai penerapan algoritma Apriori dalam analisis pola pembelian konsumen telah dilakukan secara ekstensif di berbagai sektor ritel, menunjukkan efektivitas dan adaptabilitas algoritma ini dalam konteks yang beragam. *Market Basket Analysis* telah menjadi teknik analisis data yang mapan dan banyak digunakan dalam bidang pemasaran dengan tujuan utama untuk memeriksa data transaksi konsumen guna mengidentifikasi perilaku dan preferensi pembelian [16].

Selanjutnya, penelitian komprehensif mengenai implementasi algoritma Apriori untuk menganalisis pola pembelian konsumen di toko serba ada, yang berhasil menunjukkan efektivitas yang signifikan dalam membantu pengelola toko merancang strategi penataan produk dan promosi yang lebih efektif. Hasil penelitian mengungkapkan bahwa algoritma Apriori mampu mengidentifikasi asosiasi produk dengan tingkat *confidence* yang mencapai 80%, memberikan dasar yang kuat bagi pengelola toko untuk mengembangkan strategi *bundling* produk dan optimalisasi tata letak barang berdasarkan pola pembelian yang teridentifikasi [17].

Dalam konteks sektor ritel lebih luas mengenai aplikasi algoritma Apriori untuk merekomendasikan *bundling* produk di sektor ritel, yang berhasil mengidentifikasi 877 *itemsets* yang terdiri dari *frequent 1-itemsets* hingga *frequent 4-itemsets*. Penelitian ini memberikan kontribusi signifikan dalam mengembangkan metodologi untuk prediksi produk *bundle* dalam program promosi, dengan menggunakan parameter *minimum support* 0,05% dan 0,06% serta *minimum confidence* 75%. Hasil analisis menunjukkan efektivitas algoritma dalam menghasilkan rekomendasi strategis yang aplikatif untuk pengembangan strategi *bundling* produk dan penataan produk yang lebih efisien dalam konteks bisnis ritel modern [18].

Transformasi *digital* dalam sektor *e-commerce* juga menjadi fokus penelitian yang mengimplementasikan algoritma Apriori untuk menganalisis pola pembelian konsumen di *marketplace* Shopee, khususnya pada Jaktimstore. Penelitian ini menggunakan data transaksi dari Desember 2023 hingga Mei 2024 dengan menerapkan metodologi CRISP-DM, dan berhasil menghasilkan 53 aturan asosiasi dengan parameter *minimum support* 20% dan *minimum confidence* 80%. Temuan penelitian menunjukkan bahwa algoritma Apriori efektif untuk mengidentifikasi hubungan produk yang dapat digunakan untuk memberikan rekomendasi kepada pelanggan guna meningkatkan penjualan, yang sangat relevan dalam konteks transformasi *digital retail* [19].

Eksplorasi dalam industri fashion tradisional dilakukan dalam menerapkan *market basket analysis* menggunakan algoritma Apriori pada penjualan produk batik di toko ritel, memberikan perspektif unik tentang aplikasi algoritma dalam industri dengan karakteristik budaya yang khas. Penelitian ini menemukan kombinasi *items* yang memenuhi nilai *support* sebesar 30% dengan *confidence* 100%, menunjukkan bahwa meskipun produk batik memiliki karakteristik khusus sebagai produk budaya, algoritma Apriori tetap efektif dalam mengidentifikasi pola pembelian yang dapat dimanfaatkan untuk pengembangan strategi pengelolaan stok dan promosi berbasis data asosiasi [20].

Penerapan algoritma ECLAT dalam *market basket analysis* untuk data *groceries* dengan 38.766 *record* data transaksi. Hasil penelitian menunjukkan bahwa kombinasi 1 *itemset* dengan minimal *support* 20 dan 250 mampu menghasilkan data yang relevan, sementara kombinasi 2 *itemset* hanya dapat menghasilkan analisis pada nilai minimal *support* 10 karena keterbatasan data pembelian. Waktu eksekusi algoritma ECLAT menunjukkan kondisi 1 *itemset* dengan waktu 0,0402 detik dan eksekusi terlambat pada kondisi 2 *itemset* sebesar 8,773 detik [21].

Penelitian tentang analisis pola pembelian konsumen di Toserba Yusuf Semarang menggunakan algoritma Apriori berhasil mengidentifikasi pola pembelian dengan tingkat *confidence* mencapai 75%, dan memberikan

rekomendasi konkret untuk implementasi diskon *bundling* serta pengelolaan inventaris yang lebih efisien. Salah satu temuan menarik dari penelitian ini adalah identifikasi hubungan antara kentang goreng dan telur dengan tingkat *confidence* sebesar 0,19, *support* 0,039, dan *lift ratio* 1,308, yang membuktikan kemampuan algoritma Apriori dalam menganalisis pola pembelian konsumen [22].

Inovasi dalam implementasi *digital* yang menganalisis pola pembelian konsumen di toko ritel DMART menggunakan algoritma Apriori berbasis website. Penelitian ini menunjukkan efektivitas implementasi algoritma Apriori dalam lingkungan *digital*, memberikan fleksibilitas lebih besar dalam pemrosesan data transaksi *real-time* dan penyajian hasil analisis secara interaktif. Kontribusi penelitian ini sangat signifikan dalam konteks digitalisasi retail [23].

Dalam konteks *industry fashion e-commerce* yang mengimplementasikan penggunaan metode *market basket analysis* dengan algoritma Apriori untuk menganalisis pola belanja konsumen pada *online shop* Amerta Fashion. Penelitian ini mengungkapkan bahwa konsumen tidak hanya membeli satu item dalam satu transaksi, dan berhasil mengidentifikasi bahwa Boyfriend Jeans Highwaist dan Wrap Top Summer merupakan dua jenis item yang paling populer dibeli konsumen. Temuan ini menunjukkan efektivitas algoritma Apriori dalam mengidentifikasi produk-produk yang menjadi driving force dalam keseluruhan transaksi pembelian [24].

Penelitian dalam bidang restoran dilakukan dengan mengintegrasikan *Market Basket Analysis* dan *Causal Loop Diagram* untuk mengatasi tantangan pengelolaan restoran yang kompleks. Penelitian ini menggunakan algoritma Apriori dengan parameter *minimum support* 0,01 dan *confidence* 0,50 untuk mengidentifikasi hubungan antar produk, yang kemudian menjadi dasar pengembangan *Causal Loop Diagram* untuk memetakan interaksi kausal antar variabel utama [25].

Berdasarkan kajian penelitian terdahulu, algoritma Apriori telah terbukti efektif dalam berbagai konteks ritel, mulai dari toko serba ada, marketplace online, produk budaya tradisional, hingga fashion e-commerce. Namun, penerapan spesifik pada sektor toko komputer dengan karakteristik produk teknis yang kompleks dan interdependensi tinggi antar komponen masih memerlukan eksplorasi lebih mendalam. Kesenjangan ini menciptakan peluang penelitian untuk mengadaptasi dan mengoptimalkan algoritma Apriori pada konteks toko komputer, sehingga dapat memberikan kontribusi berupa pemahaman mendalam tentang pola pembelian konsumen di sektor teknologi dan pengembangan strategi bisnis yang lebih efektif.

## **2.2 Landasan Teori**

### **2.2.1 Data Mining**

*Data mining* adalah proses teknik untuk mengekstraksi pola dan informasi tersembunyi dari kumpulan data besar melalui analisis statistik, pembelajaran mesin, dan teknik *database modern* [26]. Teknik ini mencakup berbagai metode seperti klasifikasi, klusterisasi, dan *association rule mining* yang dapat membantu organisasi dalam pengambilan keputusan berbasis data.

### **2.2.2 Market Basket Analysis**

*Market Basket Analysis* (MBA) adalah teknik *association rule mining* yang digunakan *retailer* untuk mengidentifikasi pola pembelian produk yang sering dibeli bersama dalam satu transaksi [27]. Temuan MBA dapat digunakan untuk strategi *cross-selling*, penataan *layout* toko, dan *bundling* produk guna meningkatkan penjualan dan kepuasan pelanggan.

### **2.2.3 Algoritma Apriori**

Algoritma Apriori adalah metode dasar dalam *frequent itemset mining* yang dikembangkan oleh Agrawal dan Srikant pada 1994 [28]. Algoritma ini beroperasi menurut prinsip Apriori *Property*: jika suatu itemset tidak memenuhi ambang *support*, maka seluruh superset-nya juga tidak akan memenuhi ambang tersebut.

Prinsip tersebut dirumuskan sebagai berikut: “Jika sebuah itemset tidak *frequent*, maka tidak ada superset-nya yang *frequent*”. Properti ini memungkinkan penerapan *pruning* pada ruang pencarian, sehingga menurunkan kompleksitas komputasi secara signifikan.

Algoritma Apriori beroperasi *level-wise*, dimulai dari *frequent 1-itemsets*, lalu secara iteratif membentuk kandidat  $(k+1)$ -itemsets dari *frequent k-itemsets* hingga tidak diperoleh *frequent itemsets* baru. Pada iterasi ke- $k$ , algoritma membentuk *candidate k-itemsets* ( $C_k$ ), yaitu semua kombinasi  $k$  item yang memenuhi ambang *support*. Setelah frekuensi dihitung, hanya itemset dengan *support*  $\geq$  ambang yang dipertahankan sebagai *frequent k-itemsets* ( $L_k$ ). Itemset di luar  $L_k$  di-*prune* dan dieliminasi pada level berikutnya.

Di setiap iterasi, *support* setiap kandidat itemset dihitung menurut persamaan berikut:

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

Dimana  $X$  adalah itemset yang dievaluasi. Itemset dengan *support* di bawah ambang akan dieliminasi, sehingga hanya *frequent itemsets* yang digunakan untuk membentuk kandidat pada level berikutnya.

Keunggulan utama Apriori terletak pada kesederhanaan dan kemampuannya menghasilkan seluruh *frequent itemsets* secara tuntas. Algoritma ini juga mudah diimplementasikan dan memberikan hasil yang dapat diinterpretasikan dengan baik untuk keperluan bisnis. Namun, algoritma ini kurang efisien pada dataset besar dengan ambang *support* rendah karena memerlukan banyak *database scans* dan dapat menghasilkan banyak kandidat. Meskipun demikian, pada dataset moderat dan aplikasi bisnis yang membutuhkan interpretabilitas, Apriori tetap efektif dan andal.

### 2.2.4 Association Rules

Setelah diperoleh *frequent itemsets*, langkah berikutnya adalah membentuk dan mengevaluasi *association rules*. *Association rules* adalah keluaran utama *market basket analysis* yang mengekspresikan implikasi “jika X maka Y” ( $X \rightarrow Y$ ), di mana X disebut *antecedent* dan Y disebut *consequent kuantitatif*. Kualitas dan kekuatan *association rules* dievaluasi menggunakan dua metrik utama yaitu *confidence* dan *lift* [29].

*Confidence* mengukur frekuensi kemunculan *consequent* Y dalam transaksi yang mengandung *antecedent* X, dinyatakan sebagai:

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}$$

Nilai *confidence* berkisar dari 0 hingga 1; nilai mendekati 1 menunjukkan bahwa kemunculan X hampir selalu diikuti oleh Y. *Confidence* dapat diinterpretasikan sebagai probabilitas bersyarat  $P(Y|X)$ , yaitu kemungkinan konsumen membeli Y setelah membeli X. Dalam aplikasi bisnis, *confidence* tinggi mengindikasikan kekuatan prediktif yang baik untuk strategi *cross-selling* dan sistem rekomendasi.

*Lift* adalah metrik yang membandingkan *confidence* aturan dengan *support consequent* secara independen:

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Confidence}(X \rightarrow Y)}{\text{Support}(Y)}$$

Interpretasi *lift* menunjukkan derajat hubungan antara dua item: apabila *lift* sama dengan 1, maka X dan Y bersifat independen secara statistik; nilai *lift* lebih dari 1 mengindikasikan adanya asosiasi positif, yaitu kemunculan X meningkatkan probabilitas kemunculan Y dibandingkan dengan kemungkinan acak; sebaliknya, apabila *lift* kurang dari 1, maka terdapat asosiasi negatif, di mana kemunculan X justru menurunkan probabilitas kemunculan Y.

Pembentukan *association rules* dilakukan setelah semua *frequent itemsets* diperoleh dengan algoritma Apriori. Setiap *frequent itemset* berukuran  $\geq 2$  dapat menghasilkan beberapa aturan dengan mempartisi

itemset tersebut menjadi *antecedent* dan *consequent*. Hanya aturan yang memenuhi ambang *confidence* minimum dipertahankan sebagai *strong association rules*. Penggunaan *confidence* dan *lift* bersama-sama memastikan aturan tidak hanya prediktif, tetapi juga signifikan dan tidak terjadi secara kebetulan.

#### **2.2.5 Preprocessing Data untuk Market Basket Analysis**

*Preprocessing* data merupakan tahap krusial yang meliputi pembersihan (*data cleaning*), transformasi (*data transformation*), integrasi (*data integration*), dan reduksi data (*data reduction*) untuk memastikan kualitas data transaksi yang akan dianalisis oleh algoritma Apriori. Pada MBA, data transaksi diubah menjadi format *boolean matrix* agar algoritma dapat menghitung *frequent itemsets* dengan efisien [30].

#### **2.2.6 Python dan Pustaka mlxtend**

*Python* menjadi bahasa pilihan dalam riset *data mining* berkat pustaka *mlxtend* yang menyediakan modul apriori dan *association\_rules* untuk penambangan *frequent itemsets* dan pembuatan aturan asosiasi dengan parameter yang dapat disesuaikan [31]. Versi terbaru *mlxtend* menambahkan metrik baru seperti *Jaccard* dan *Certainty* untuk memperkaya evaluasi aturan asosiasi [32].

#### **2.2.7 Karakteristik Industri Ritel Teknologi**

Industri ritel teknologi, khususnya toko komputer, memiliki kompleksitas unik berupa kompatibilitas teknis antar komponen (*motherboard*, RAM, GPU) yang mempengaruhi pola pembelian konsumen. Konsumen cenderung membeli produk yang kompatibel secara teknis, sehingga model MBA pada sektor ini harus mempertimbangkan ketergantungan komponen teknis selain faktor preferensi konsumen umum [33].