

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pesatnya perkembangan industri ritel daring (*e-commerce*) saat ini berdampak langsung pada akumulasi data transaksi dalam volume yang sangat besar. Aktivitas operasional harian yang menghasilkan data dalam skala masif tersebut memicu tantangan baru bagi pihak manajemen, di mana tumpukan data transaksi sering kali hanya berakhir sebagai arsip digital tanpa pemanfaatan lebih lanjut untuk mendukung keputusan strategis pemasaran. Salah satu representasi data transaksi berskala besar ini adalah *Online Retail Dataset* dari *UCI Machine Learning Repository* yang memuat riwayat transaksi ritel wilayah *United Kingdom* (UK) dengan fokus penjualan produk hadiah dan perlengkapan rumah tangga.

Untuk mengubah tumpukan data transaksi mentah menjadi informasi strategis, diperlukan pendekatan berbasis *data mining* melalui metode *Market Basket Analysis* (MBA). MBA diaplikasikan secara spesifik untuk mengungkap karakteristik perilaku belanja konsumen melalui identifikasi kecenderungan pembelian produk secara bersamaan dalam satu keranjang belanja. Secara ilmiah, MBA merupakan implementasi nyata dari teknik *Association Rule Mining* (penambangan aturan asosiasi) guna memformulasikan kombinasi produk (*itemset*) yang memiliki tingkat kemunculan tinggi berdasarkan riwayat transaksi historis. (Romadhon & Kodar, n.d.).

Dalam ranah penambangan aturan asosiasi, Algoritma Apriori merupakan metodologi standar yang paling sering diadopsi karena memiliki struktur logika yang eksplisit. Algoritma ini bekerja dengan melakukan pencarian kombinasi *itemset* secara berulang melalui pembangkitan kandidat kombinasi produk (*k-itemset*) berdasarkan ambang batas nilai *minimum support* dan *minimum confidence*. Namun, algoritma ini memiliki keterbatasan komputasi yang signifikan ketika dihadapkan pada dataset riil berukuran besar dengan jumlah transaksi mencapai ratusan ribu baris. Keterbatasan utama Apriori terletak pada mekanisme pemindaian basis data (*database scanning*) yang dilakukan berulang kali pada setiap iterasi, sehingga memicu pembengkakan penggunaan memori (*Random Access Memory*) dan waktu eksekusi prosesor.

Sebagai landasan pijak akurasi metodologi, penelitian ini mengkaji beberapa penelitian terdahulu yang relevan. Penelitian pertama oleh Darmaastawan menerapkan Algoritma Apriori untuk mengoptimalkan penataan letak barang pada Apotek Mooladhara Denpasar,

di mana ditemukan pola asosiasi kuat antara produk Asam Mefenamat dan Amoxicillin dengan nilai *confidence* sebesar 65,2%. Selanjutnya, Alma'arif dkk. mengimplementasikan Algoritma Apriori pada platform toko daring untuk membangkitkan rekomendasi produk otomatis berbasis nilai *confidence* historis pelanggan (Darmaastawan, 2023). Penelitian ketiga oleh Asharil dkk. menggunakan *Market Basket Analysis* dengan Algoritma Apriori untuk menganalisis pendapatan usaha ritel dan menunjukkan bahwa pola asosiasi yang terbentuk berkorelasi langsung terhadap puncak lonjakan omzet penjualan bulanan perusahaan (Ashari et al., 2022). Lebih lanjut, Kristania dkk. memanfaatkan Algoritma Apriori pada PT. Duta Kencana Swaguna untuk memformulasikan strategi penjualan produk makanan beku (*frozen food*) dengan perolehan nilai *lift ratio* mencapai 1,09 (Listanto & Meisella Kristania, 2022). Terakhir, Amalia dkk. menguji performa Algoritma Apriori pada CV Rey Gasendra menggunakan komparasi manual dan *tools* RapidMiner, yang menghasilkan rekomendasi persediaan stok barang terintegrasi untuk produk elektronik.

Berdasarkan analisis kritis terhadap sejumlah literatur terdahulu tersebut, ditemukan adanya *Research Gap* (celah penelitian) yang nyata sebagai berikut:

1. **Keterbatasan Evaluasi Performa Komputasi:** Mayoritas penelitian terdahulu didominasi oleh dataset berskala kecil hingga menengah (di bawah 10.000 transaksi) dan cenderung berasumsi bahwa Algoritma Apriori dapat diterapkan pada seluruh karakteristik data tanpa mengukur batas atas kompleksitas ruang memori komputer serta waktu eksekusi.
2. **Ketiadaan Analisis Sensitivitas Parameter:** Penelitian sebelumnya umumnya langsung menetapkan satu nilai parameter *support* dan *confidence* tanpa melakukan pengujian sensitivitas multi-parameter untuk menemukan titik keseimbangan informasi yang optimal.
3. **Keterbatasan Visualisasi dan Interpretasi Bisnis:** Luaran penelitian terdahulu sebagian besar berhenti pada penyajian angka statistik aturan asosiasi dalam bentuk tabel mentah, tanpa dilengkapi visualisasi jaringan yang interaktif maupun penerjemahan nyata ke dalam skema tata letak produk ritel.

Mengingat *Online Retail Dataset* dalam penelitian ini memiliki volume yang sangat besar yaitu mencapai 467.054 baris transaksi bersih, pengujian awal menggunakan Algoritma Apriori sebagai baseline berpotensi besar mengalami kegagalan sistem akibat keterbatasan memori (*out-of-memory error*). Sebagai solusi komputasi ilmiah atas

kelemahan struktural tersebut, penelitian ini mengintegrasikan Algoritma *FP-Growth* (*Frequent Pattern Growth*) sebagai alternatif pembanding sekaligus solusi optimalisasi komputasi. *FP-Growth* mengatasi kelemahan struktural Apriori dengan cara memadatkan database transaksi ke dalam struktur pohon dinamis yang disebut *FP-Tree*, sehingga pemindaian basis data hanya perlu dilakukan sebanyak dua kali tanpa perlu membangkitkan jutaan kandidat kombinasi produk buatan yang membebani memori komputer. Melalui batasan dan kondisi tersebut, **Kebaruan (Novelty)** yang ditawarkan dalam penelitian ini dirumuskan ke dalam poin-poin sebagai berikut:

1. **Evaluasi Skalabilitas Komputasi Ritel Skala Besar:** Menguji secara empiris batas ketahanan dan tingkat skalabilitas komputasi Algoritma Apriori ketika dihadapkan pada dataset transaksi ritel berskala besar.
2. **Analisis Komparatif Performa Apriori dan *FP-Growth*:** Menyajikan perbandingan kuantitatif yang objektif antara Algoritma Apriori dan Algoritma *FP-Growth* dari aspek efisiensi waktu eksekusi (*execution time*) dan alokasi kapasitas memori RAM.
3. **Penerapan Analisis Sensitivitas dan Visualisasi Aturan Asosiasi:** Melakukan simulasi pengujian multi-parameter untuk menemukan ambang batas koordinat aturan terbaik, serta memvisualisasikannya melalui komponen *Network Graph* dan *scatter Plot* guna mempermudah identifikasi sebaran aturan.
4. **Formulasi Rekomendasi Bisnis Terjemahan Kuantitatif:** Mentransformasikan nilai metrik *Support*, *Confidence*, dan *Lift Ratio* menjadi rekomendasi tata letak produk digital (*product placement*), rancangan paket hemat (*bundling*), serta skema pemasaran silang (*cross-selling*) secara konkret.

Berdasarkan urgensi permasalahan tersebut, penelitian ini dilakukan dengan judul: **"Analisis Pola Pembelian Konsumen Menggunakan Market Basket Analysis dengan Perbandingan Algoritma Apriori dan *FP-Growth* pada Online Retail Dataset"**.

1.2. Rumusan Masalah

Dengan uraian latar belakang masalah maka dapat diidentifikasi permasalahan yang ada sebagai berikut.

1. Bagaimana karakteristik pola pembelian produk konsumen dari *Online Retail Dataset* yang diidentifikasi melalui parameter *frequent itemset*?
2. Bagaimana perbandingan performa komputasi antara Algoritma Apriori dan Algoritma *FP-Growth* dalam menangani dataset transaksi berskala besar berdasarkan indikator waktu eksekusi (*execution time*), penggunaan memori (RAM),

dan jumlah aturan (*rules*) yang dihasilkan?

3. Aturan asosiasi (*association rules*) apa saja yang dihasilkan dan bagaimana interpretasi bisnis dari aturan tersebut berdasarkan metrik *Support*, *Confidence*, dan *Lift Ratio* untuk strategi penataan produk?

1.3. Tujuan dan Manfaat

Tujuan dari penelitian ini yaitu:

1. Menganalisis karakteristik pola pembelian produk konsumen dari *Online Retail Dataset* berdasarkan frekuensi kemunculan *frequent item*.
2. Membandingkan efisiensi performa komputasi antara Algoritma Apriori dan Algoritma *FP-Growth* berdasarkan indikator waktu eksekusi (*execution time*), penggunaan memori (*RAM*) dan jumlah aturan (*rules*) yang dihasilkan dalam menangani dataset transaksi berskala besar.
3. Merumuskan rekomendasi strategi penataan produk (*store layout*) digital melalui interpretasi bisnis aturan asosiasi (*association rules*) yang memenuhi ambang kasus metrik *Support*, *Confidence*, dan *Lift Ratio*.

Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat praktis berupa penyediaan informasi analitis mengenai peta kebiasaan belanja pelanggan. Informasi ini dapat digunakan oleh pengelola industri ritel daring sebagai instrumen dasar pengambilan keputusan berbasis data (*data-driven decision*) dalam merancang program promosi, paket penjualan, dan tata letak beranda web yang lebih optimal.

1.4. Batasan Masalah

Guna lebih memfokuskan permasalahan yang diteliti maka penulis menetapkan batasan masalah sebagai berikut:

1. Data yang diteliti bersumber dari *Online Retail Dataset* (UCI Machine Learning Repository), di mana pemrosesan data dibatasi secara spesifik pada transaksi wilayah *United Kingdom* (UK) dengan alasan akademik bahwa wilayah tersebut memiliki kepadatan volume transaksi terbesar (>90%), variabilitas produk tertinggi, serta tingkat konsistensi pencatatan data paling stabil dibandingkan wilayah lainnya.
2. Fokus Penelitian ini dibatasi pada analisis data transaksi penjualan produk ritel daring untuk kategori produk hadiah (*giftware*) dan perlengkapan kebutuhan keluarga ringan.
3. Metode utama yang digunakan sebagai *baseline* adalah *Association Rule Mining*

dengan menempatkan Algoritma Apriori sebagai model dasar serta Algoritma *FP-Growth* sebagai solusi optimalisasi komputasi untuk mengatasi kendala keterbatasan memori.

4. Parameter evaluasi kualitas aturan asosiasi dibatasi pada pengujian nilai *Support*, *Confidence*, dan kekuatan keterikatan aturan melalui *Lift Ratio*.
5. Proses pra-pemrosesan data (*data preprocessing*) dan eksperimen komparasi algoritma dilakukan menggunakan bahasa pemrograman Python dalam lingkungan perangkat lunak Visual Studio Code dengan *Jupyter Notebook* dan perangkat keras berspesifikasi RAM 8.
6. Output akhir yang dihasilkan berupa pola aturan asosiasi terpilih (Network Graph dan Scatter Plot) yang dilengkapi dengan rekomendasi strategi penataan produk (*product placement*), paket penjualan (*bundling*), dan pemasaran silang (*cross-selling*).

1.5. Sistematika Penulisan

1. BAB 1 PENDAHULUAN Bab ini menjelaskan tentang latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penulisan dan sistematikan penulisan.
2. BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI Pada bab ini menjelaskan tinjauan Pustaka dari penelitian sebelumnya.
3. BAB III METODE PENELITIAN Bab ini berisikan tentang uraian alur penelitian, pada bab ini juga dijelaskan dengan detail langkah-langkah yang harus dilalui untuk mencapai tujuan penelitian dan menghasilkan simpulan akhir penelitian.
4. BAB IV HASIL DAN PEMBAHASAN Bab ini menjelaskan penjabaran hasil denagn pengujian menggunakan Software dan pembahasan yang bersifat terpadu dan tidak dipecah menjadi sub bab tersendiri
5. BAB V PENUTUP Bab ini akan memuat kesimpulan dari penelitian yang dilakukan dan saran untuk penelitian selanjutnya.