

Analisis Pola Asosiasi Produk Menggunakan Algoritma Apriori dan FP-Growth pada Dataset Online Retail

TUGAS BESAR DATA MINING

Oleh

Muhammad Farhan Akbar 714220004

Balqis Rosa Sekamayang 714220006

Sindy Maulina 714220018



Universitas Logistik & Bisnis Internasional

DIPLOMA IV TEKNIK INFORMATIKA

SEKOLAH VOKASI

UNIVERSITAS LOGISTIK & BISNIS INTERNASIONAL

BANDUNG

HALAMAN PERNYATAAN ORISINALITAS

Laporan tugas besar ini adalah hasil karya kami sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah kami nyatakan dengan benar. Bilamana di kemudian hari ditemukan bahwa karya tulis ini menyalahi peraturan yang ada berkaitan dengan etika dan kaidah penulisan karya ilmiah yang berlaku, maka kami bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Yang menyatakan,

Nama : Muhammad Farhan Akbar

NIM : 714220004

Tanda Tangan :

Tanggal : 10 Juli 2025

Mengetahui,

Ketua : Muhammad Farhan Akbar Muhlis

Dosen Pengampu Mata Kuliah Data Mining : Nisa Hanum Harani., S.Kom., M.T., SFPC

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan laporan akhir Tugas Besar Data Mining yang berjudul **“Analisis Pola Asosiasi Produk Menggunakan Algoritma Apriori dan FP-Growth pada Dataset Online Retail”** dengan lancar dan tepat waktu.

Laporan ini disusun sebagai salah satu syarat untuk memenuhi tugas akhir pada mata kuliah Data Mining di Program Studi Informatika, Universitas Logistik dan Bisnis Internasional (ULBI). Penulisan laporan ini bertujuan untuk menerapkan dan mengimplementasikan teknik data mining, khususnya pada metode *association rule mining* dalam konteks dunia nyata menggunakan dataset e-commerce.

Dalam proses penyusunan laporan ini, penulis menyadari bahwa tanpa bantuan, dukungan, dan arahan dari berbagai pihak, laporan ini tidak akan terselesaikan dengan baik. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar- besarnya kepada:

1. **Ibu Nisa Hanum Harani, S.Kom., M.T.,CDSP, SFPC Dosen Pengampu Mata Kuliah Data Mining**, yang telah memberikan ilmu, bimbingan, dan arahan selama proses perkuliahan dan pelaksanaan tugas besar.
2. **Orang tua dan keluarga**, atas segala doa, motivasi, dan dukungan moral maupun material yang tidak pernah henti-hentinya diberikan.
3. **Rekan-rekan mahasiswa**, yang turut membantu, berdiskusi, dan memberikan semangat selama pelaksanaan tugas besar ini.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang. Semoga laporan ini dapat memberikan manfaat bagi pembaca, khususnya dalam bidang data mining dan analisis pola pembelian konsumen.

Bandung, 10 Juli 2025

Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI LAPORAN AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Logistik & Bisnis Internasional, saya yang bertanda tangan di bawah ini:

Nama: Muhammad Farhan Akbar Muhlis

NPM: 714220004

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Logistik Bisnis Internasional, Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalti Free Right*) atas karya ilmiah saya yang berjudul:

Analisis Pola Asosiasi Produk Menggunakan Algoritma Apriori dan FP-Growth pada Dataset Online Retail

Beserta perangkat yang ada (jika diperlukan). Dengan Hak ini Universitas Logistik Bisnis Internasional Hayati berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Bandung

Pada tanggal : 10 Juli 2025

Yang menyatakan

Muhammad Farhan Akbar Muhlis

714220004

ABSTRAK

Penelitian ini bertujuan untuk menemukan pola asosiasi produk dari data transaksi e-commerce menggunakan pendekatan data mining. Dataset yang digunakan adalah Online Retail dari UCI Machine Learning Repository, yang memuat data transaksi penjualan dari Desember 2010 hingga Desember 2011. Metode yang digunakan dalam penelitian ini adalah algoritma Apriori dan FP-Growth untuk menemukan asosiasi antar produk. Proses preprocessing dilakukan melalui pembersihan data, pengelompokan transaksi, dan transformasi ke format one-hot encoding. Hasil penelitian menunjukkan bahwa algoritma FP-Growth lebih efisien dibandingkan Apriori dari sisi waktu komputasi dan jumlah aturan yang dihasilkan. Pola asosiasi yang ditemukan diharapkan dapat dimanfaatkan oleh pelaku bisnis untuk strategi pemasaran dan pengelolaan stok yang lebih optimal.

Kata Kunci: data mining, market basket analysis, apriori, fp-growth, asosiasi produk

ABSTRACT

This study aims to discover product association patterns from e-commerce transaction data using a data mining approach. The dataset used is Online Retail from the UCI Machine Learning Repository, containing transaction data from December 2010 to December 2011. The methods applied are the Apriori and FP-Growth algorithms to extract associations between products. Preprocessing includes data cleaning, transaction grouping, and transformation into a one-hot encoded format. The results show that FP-Growth is more efficient than Apriori in terms of computation time and the number of rules generated. The discovered association rules are expected to help businesses improve their marketing strategies and stock management.

Keywords: *data mining, market basket analysis, apriori, fp-growth, product association*

DAFTAR ISI

| | |
|--|-----|
| HALAMAN PERNYATAAN ORISINALITAS..... | ii |
| KATA PENGANTAR | iii |
| HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI LAPORAN AKHIR UNTUK KEPENTINGAN AKADEMIS | iv |
| ABSTRAK | v |
| <i>ABSTRACT</i> | vi |
| DAFTAR ISI | vii |
| DAFTAR TABEL..... | ix |
| DAFTAR GAMBAR..... | x |
| DAFTAR RUMUS | xi |
| DAFTAR NOTASI | xii |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah..... | 1 |
| 1.3 Tujuan Penelitian | 2 |
| 1.4 Manfaat Penelitian | 2 |
| 1.5 Ruang Lingkup | 2 |
| BAB II TINJAUAN PUSTAKA | 3 |
| 2.1 Landasan Teori dan Konsep | 3 |
| 2.1.1 Data Mining..... | 3 |
| 2.1.2 Market Basket Analysis (MBA)..... | 3 |
| 2.2 Algoritma Association Rule Mining | 4 |
| 2.2.1 Algoritma Apriori | 4 |
| 2.2.2 Algoritma FP-Growth | 4 |
| 2.3 Diagram Alur Konsep Metodologi | 5 |
| 2.4 State Of The Art dalam Association Rule Mining | 6 |
| BAB III METODOLOGI..... | 8 |
| 3.1 Kerangka Penelitian | 8 |
| 3.2 Deskripsi Dataset | 8 |
| 3.3 Preprocessing Data | 8 |
| 3.4 Pemodelan dan Algoritma yang Digunakan | 9 |
| 3.5 Evaluasi dan Validasi Hasil | 9 |
| 3.5.1 Evaluasi Aturan Asosiasi | 9 |

| | | |
|--|---|-----------|
| 3.5.2 | Visualisasi dan Analisis Perbandingan | 10 |
| 3.5.3 | Catatan Evaluasi..... | 11 |
| BAB IV HASIL DAN PEMBAHASAN | | 12 |
| 4.1 | Visualisasi eksploratif (EDA) | 12 |
| 4.1.1 | Struktur Data | 12 |
| 4.1.2 | Cek Missing Value dan Duplikasi Data | 12 |
| 4.1.3 | Statistik Deskriptif..... | 13 |
| 4.1.4 | Visualisasi Data | 14 |
| 4.2 | Hasil preprocessing dan pemodelan | 15 |
| 4.2.1 | Hasil Preprocessing | 16 |
| 4.2.2 | Persiapan Pemodelan..... | 17 |
| 4.2.3 | Penerapan Algoritma Apriori dan FP-Growth | 18 |
| 4.2.4 | Hasil Pemodelan dan Visualisasi | 18 |
| 4.3 | Tabel Hasil eksperimen/model | 21 |
| 4.4 | Analisis keunggulan dan keterbatasan | 22 |
| BAB V KESIMPULAN DAN SARAN | | 24 |
| 5.1 | Kesimpulan..... | 24 |
| 5.2 | Saran | 24 |
| DAFTAR PUSTAKA..... | | 26 |
| LAMPIRAN | | 27 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 1 Struktur Data | 12 |
| Tabel 2 Analisis Keunggulan dan Keterbatasan..... | 22 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2. 1 Diagram Alur..... | 5 |
| Gambar 3. 1 Visualisasi dan Analisis Perbandingan..... | 10 |
| Gambar 4. 1 Missing Value dan Duplikasi data..... | 13 |
| Gambar 4. 2 Statistik Deskriptif | 13 |
| Gambar 4. 3 Distribusi Quantity & UnitPrice..... | 14 |
| Gambar 4. 4 Boxplot Quantity & Boxplot UnitPrice..... | 15 |
| Gambar 4. 5 Heatmap Korelasi..... | 15 |
| Gambar 4. 6 Hasil Preprocessing 10 data | 16 |
| Gambar 4. 7 Hasil Preprocessing Encoding..... | 17 |
| Gambar 4. 8 Persiapan Pemodelan | 17 |
| Gambar 4. 9 Penerapan Algoritma Apriori | 18 |
| Gambar 4. 10 Penerapan Algoritma FP-Growth..... | 18 |
| Gambar 4. 11 Aturan Asosiasi (Apriori) | 19 |
| Gambar 4. 12 Aturan Asosiasi FP-Growth..... | 20 |
| Gambar 4. 13 Perbandingan Support Itemsets..... | 20 |
| Gambar 4. 14 Tabel Hasil eksperimen/model | 21 |

DAFTAR RUMUS

DAFTAR NOTASI

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring berkembangnya era digital, volume data yang dihasilkan oleh aktivitas perdagangan elektronik (e-commerce) semakin meningkat secara eksponensial. Data tersebut menyimpan berbagai informasi penting mengenai perilaku konsumen, preferensi pembelian, serta pola-pola transaksi yang jika dianalisis secara tepat dapat digunakan untuk mendukung keputusan bisnis strategis. Salah satu pendekatan yang digunakan untuk menggali pola tersembunyi dalam data transaksi adalah data mining, khususnya teknik association rule mining [1].

Market Basket Analysis (MBA) merupakan salah satu aplikasi dari association rule mining yang bertujuan untuk menemukan asosiasi atau korelasi antara item yang sering dibeli secara bersamaan dalam satu transaksi. Teknik ini dapat membantu pelaku bisnis dalam menyusun strategi pemasaran seperti rekomendasi produk, penyusunan layout toko, dan promosi bundling [2].

Pada penelitian ini, digunakan dataset "Online Retail" yang berasal dari UCI Machine Learning Repository. Dataset tersebut memuat transaksi penjualan dari sebuah perusahaan *e-commerce* yang berbasis di Inggris, dan mencakup informasi tentang produk, pelanggan, dan waktu transaksi dari tahun 2010 hingga 2011. Penelitian ini memanfaatkan dua algoritma asosiasi, yaitu Apriori dan FP-Growth, untuk menggali pola pembelian produk yang sering muncul bersamaan. Dengan menerapkan proses preprocessing yang tepat, hasil analisis diharapkan dapat memberikan wawasan yang berguna bagi optimalisasi strategi penjualan dan pemasaran [3].

1.2 Rumusan Masalah

1. Bagaimana proses preprocessing dilakukan untuk menyiapkan dataset Online Retail dalam analisis data mining?
2. Bagaimana penerapan algoritma Apriori dan FP-Growth dalam menemukan pola asosiasi antar produk?
3. Seberapa kuat dan relevan pola asosiasi yang ditemukan terhadap pengambilan keputusan bisnis?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Melakukan preprocessing data pada dataset Online Retail agar siap digunakan dalam proses mining.
2. Menerapkan algoritma Apriori dan FP-Growth untuk mengekstraksi aturan asosiasi dari data transaksi.
3. Mengevaluasi pola asosiasi yang ditemukan dan menginterpretasikan hasilnya dalam konteks bisnis ritel.

1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Manfaat Teoritis: Menambah wawasan dan pengetahuan dalam bidang data mining, khususnya dalam penerapan teknik asosiasi pada data transaksi e-commerce.
2. Manfaat Praktis: Memberikan insight yang dapat dimanfaatkan pelaku usaha untuk meningkatkan efektivitas strategi pemasaran dan pengelolaan stok barang melalui analisis pola pembelian konsumen.

1.5 Ruang Lingkup

Ruang lingkup penelitian ini dibatasi pada analisis data transaksi yang bersumber dari dataset Online Retail yang tersedia secara publik di UCI Machine Learning Repository. Fokus utama penelitian adalah pada itemset yang muncul dalam satu invoice atau transaksi, tanpa mempertimbangkan waktu pembelian, segmentasi pelanggan, atau analisis prediktif. Penelitian ini menggunakan dua algoritma asosiasi, yaitu Apriori dan FP-Growth, serta mengevaluasi hasilnya berdasarkan metrik support, confidence, dan lift.

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori dan Konsep

2.1.1 Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi terkait dari berbagai database besar [4]. Tujuannya adalah untuk membantu pengambilan keputusan berdasarkan data dengan cara yang lebih efisien dan akurat. Proses data mining umumnya melibatkan tahapan seperti pembersihan data, integrasi data, seleksi data, transformasi data, penambangan pola, evaluasi pola, dan penyajian informasi.

2.1.2 Market Basket Analysis (MBA)

Salah satu metode yang dapat digunakan untuk menganalisis pola perilaku belanja konsumen adalah Market Basket Analysis. Analisis ini merupakan salah satu metode dalam penambangan data (data mining) yang bertujuan untuk menemukan produk-produk yang sering dibeli bersamaan dari data transaksi [5].

MBA termasuk dalam teknik Association Rule Mining, di mana aturan yang dihasilkan dievaluasi menggunakan metrik seperti support, confidence, dan lift. Dengan memahami keterkaitan antar produk, perusahaan dapat meningkatkan strategi pemasaran, penempatan barang, bundling produk, serta meningkatkan pengalaman pelanggan. Contohnya, dalam sebuah toko ritel, produk-produk yang sering dibeli bersamaan dapat ditempatkan dalam rak yang berdekatan.

MBA juga dapat digunakan tidak hanya di sektor retail, tetapi juga dalam bidang lain seperti analisis log data, medis, dan sistem rekomendasi konten, selama terdapat pola keterkaitan antar item dalam satu grup peristiwa atau transaksi.

2.2 Algoritma Association Rule Mining

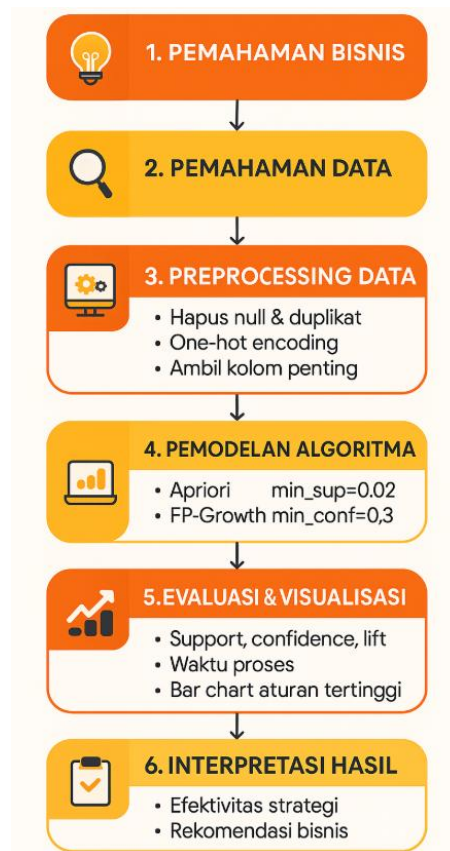
2.2.1 Algoritma Apriori

Algoritma Apriori merupakan sebuah metode yang digunakan dalam mencari pola hubungan antar satu atau lebih item dalam suatu dataset [6]. Algoritma klasik dalam association rule mining yang bekerja dengan prinsip bahwa semua subset dari itemset yang sering muncul juga harus sering muncul. Algoritma ini menggunakan pendekatan iteratif (level-wise search) dengan memperluas itemset satu per satu (dari 1-itemset ke 2-itemset dan seterusnya). Untuk menentukan itemset yang sering muncul (frequent itemsets), Apriori membandingkan support dari itemset terhadap ambang batas minimum support. Apriori cocok untuk dataset berukuran kecil hingga menengah, namun bisa menjadi lambat pada dataset besar karena eksplorasi kombinasi item yang banyak.

2.2.2 Algoritma FP-Growth

Algoritma fp-growth merupakan salah satu algoritma yang merupakan bagian dari Teknik association rule yang bertujuan untuk mencari dan menentukan suatu kumpulan data yang sering muncul pada sebuah tambang data [7]. Algoritma alternatif dari Apriori yang lebih efisien dalam menemukan frequent itemsets. FP-Growth membangun struktur data yang disebut FP-Tree (Frequent Pattern Tree) yang menyimpan informasi frekuensi item dalam bentuk pohon kompak. Dengan struktur ini, algoritma dapat menemukan pola frequent itemsets tanpa eksplisit membangkitkan kandidat itemset seperti pada Apriori, sehingga mengurangi waktu komputasi secara signifikan. FP-Growth lebih cocok digunakan untuk dataset besar dan kompleks karena performanya yang lebih baik.

2.3 Diagram Alur Konsep Metodologi



Gambar 2. 1 Diagram Alur

Dalam penelitian ini, proses analisis data dilakukan melalui enam tahapan utama, dimulai dari pemahaman bisnis hingga interpretasi hasil. Tahap pertama adalah pemahaman bisnis, yaitu mengenali permasalahan dan tujuan utama dari analisis, yakni untuk menemukan pola hubungan antar produk dalam data transaksi guna mendukung pengambilan keputusan bisnis yang lebih efektif. Setelah memahami konteks bisnis, langkah berikutnya adalah pemahaman data, di mana dilakukan eksplorasi terhadap struktur dataset yang tersedia, seperti jenis atribut, nilai-nilai yang terkandung, dan format transaksi penjualan yang relevan.

Tahapan ketiga adalah preprocessing data, yang mencakup penghapusan data kosong (null) dan duplikat untuk menjaga kualitas data, penerapan teknik one-hot encoding agar data transaksi dapat diproses oleh algoritma, serta pemilihan atribut penting seperti kode produk, nama produk, dan nomor transaksi. Selanjutnya, dilakukan pemodelan algoritma dengan menerapkan dua metode Association Rule Mining, yaitu

algoritma Apriori dengan minimum support 0.02, dan algoritma FP-Growth dengan minimum confidence 0.3. Tahap ini bertujuan untuk menemukan aturan asosiasi berdasarkan frekuensi kombinasi item dalam transaksi.

Tahap kelima adalah evaluasi dan visualisasi, yang melibatkan pengukuran metrik seperti support, confidence, dan lift untuk menilai kekuatan hubungan antar produk. Selain itu, juga dilakukan analisis waktu proses untuk membandingkan efisiensi dari masing-masing algoritma, serta visualisasi aturan asosiasi dalam bentuk bar chart untuk mempermudah interpretasi. Tahap terakhir adalah interpretasi hasil, di mana dilakukan penafsiran terhadap aturan-aturan yang dihasilkan untuk menilai efektivitas strategi bisnis yang mungkin diterapkan. Hasil ini digunakan sebagai dasar untuk memberikan rekomendasi bisnis, seperti strategi pemasaran produk, penempatan produk di toko, atau penyusunan promosi bundling yang lebih menarik bagi konsumen.

2.4 State Of The Art dalam Association Rule Mining

Beberapa penelitian telah dilakukan terkait penerapan algoritma Apriori dan FP-Growth dalam pengaturan tata letak barang atau analisis pola belanja konsumen di supermarket. Harahap et al. (2022) membandingkan algoritma Apriori dan FP-Growth menggunakan RapidMiner dalam menentukan tata letak barang. Hasilnya, FP-Growth mampu menghasilkan 78 rules dengan akurasi 78,06%, lebih unggul dibandingkan Apriori yang hanya menghasilkan 10 rules dengan akurasi 12,8%. Penelitian serupa dilakukan kembali oleh penulis yang sama, menegaskan keunggulan FP-Growth dalam membantu pengaturan tata letak barang [8].

Sementara itu, Made Leo Radhitya et al. (2024) melakukan analisis tata letak produk berdasarkan pola belanja konsumen menggunakan algoritma Apriori dan High Utility Itemset Mining. Penelitian ini berhasil memperoleh nilai support sebesar 85%, confidence 86%, dan lift ratio 1,01, serta mencatatkan keuntungan tertinggi sebesar Rp567.000. Rekomendasi penataan produk disusun berdasarkan hasil aturan asosiasi yang terbentuk, sehingga dapat membantu peningkatan strategi penjualan [9].

Penelitian oleh Edwin Omol et al. (2024) di Kenya juga memanfaatkan algoritma Apriori yang diimplementasikan dengan bahasa pemrograman Python untuk analisis Market Basket. Penelitian ini menghasilkan 42 aturan asosiasi dan mampu memberikan insight penting terkait pola pembelian konsumen yang dapat digunakan sebagai dasar pengambilan keputusan dalam strategi pemasaran dan penempatan produk [10].

Selanjutnya, Bekri Murdianto dan Arief Jananto (2023) menggunakan metode CRISP-DM dengan bantuan tools RStudio untuk membandingkan algoritma Apriori dan FP-Growth dalam memberikan rekomendasi penataan display barang. FP-Growth kembali menunjukkan performa lebih unggul dengan menghasilkan 53 aturan asosiasi dan akurasi sebesar 1245%, dibandingkan Apriori yang hanya menghasilkan 12 aturan dengan akurasi 7,8%. Hal ini memperkuat temuan sebelumnya bahwa FP-Growth lebih layak digunakan untuk kasus-kasus rekomendasi penataan barang [11].

Terakhir, penelitian yang dilakukan oleh Mahtum et al. (2024) pada Mohare Supermarket juga memperkuat keunggulan algoritma FP-Growth dibanding Apriori. Dalam penelitian tersebut, FP-Growth menghasilkan 78 aturan asosiasi dengan nilai support sebesar 2,516 dan akurasi 780%, sedangkan Apriori hanya mampu menghasilkan 10 aturan dengan akurasi 12,8%. Temuan ini menegaskan bahwa FP-Growth lebih efektif dalam membantu analisis dan pengaturan tata letak barang pada lingkungan retail [12].

Dari berbagai penelitian tersebut, dapat disimpulkan bahwa algoritma FP-Growth secara konsisten memberikan hasil yang lebih baik dibandingkan algoritma Apriori, baik dari segi jumlah aturan yang dihasilkan, akurasi, maupun efektivitas dalam mendukung strategi bisnis, khususnya dalam penataan barang dan analisis pola belanja konsumen.

BAB III METODOLOGI

3.1 Kerangka Penelitian

Penelitian ini didasarkan pada pendekatan sistematis dan terstruktur, dengan mengikuti siklus CRISP-DM (Cross-Industry Standard Process for Data Mining) yang telah diakui secara luas dalam dunia akademik dan industri. Pendekatan ini terdiri atas beberapa tahapan utama, yaitu: pemahaman bisnis, pemahaman data, persiapan data (preprocessing), pemodelan, evaluasi, dan deployment. CRISP-DM dipilih karena fleksibilitas dan kestabilannya dalam berbagai skenario aplikasi data mining.

3.2 Deskripsi Dataset

Dataset yang digunakan adalah Online Retail yang bersumber dari UCI Machine Learning Repository [4], terdiri atas 541.909 baris dan 8 atribut. Dataset ini merepresentasikan data transaksi penjualan dari sebuah perusahaan retail yang berbasis di Inggris, dengan periode transaksi antara Desember 2010 hingga Desember 2011. Atribut yang tersedia mencakup informasi faktur, kode produk, nama produk, kuantitas, tanggal transaksi, harga per unit, ID pelanggan, dan negara.

Adapun atribut yang digunakan secara khusus dalam penelitian ini adalah:

- InvoiceNo: ID transaksi
- Description: Nama produk
- CustomerID: ID pelanggan

Atribut lain seperti UnitPrice dan Quantity tidak dimanfaatkan secara langsung dalam proses asosiasi karena fokus utama adalah pada pola kebersamaan pembelian, bukan prediksi nilai atau jumlah.

3.3 Preprocessing Data

Proses preprocessing dilakukan untuk meningkatkan kualitas data dan memastikan validitas hasil analisis. Tahapan preprocessing meliputi:

1. Pembersihan Data:
 - Menghapus nilai kosong pada kolom CustomerID dan Description.
 - Menghapus baris duplikat agar tidak mempengaruhi frekuensi itemset.
2. Transformasi Data:
 - Data transaksi dikelompokkan berdasarkan InvoiceNo.

- Setiap transaksi dikonversi ke dalam format one-hot encoded menggunakan TransactionEncoder dari pustaka Python mlxtend. Hasilnya berupa matriks boolean yang menunjukkan apakah suatu produk terdapat dalam sebuah transaksi.

3. Reduksi Data:

- Hanya tiga atribut penting yang diambil: InvoiceNo, Description, dan CustomerID.
- Reduksi dilakukan untuk efisiensi pemrosesan dan relevansi terhadap tujuan asosiasi.

Dataset final terdiri dari sekitar 25.000 transaksi unik, dalam format CSV, siap digunakan dalam model Apriori dan FP-Growth.

3.4 Pemodelan dan Algoritma yang Digunakan

Dua algoritma utama digunakan dalam penelitian ini, yaitu:

1. Apriori:

- Merupakan algoritma berbasis enumerasi kandidat.
- Menemukan frequent itemsets dengan memanfaatkan prinsip downward closure.
- Meskipun mudah dipahami, performanya cenderung lambat pada dataset besar.

2. FP-Growth:

- Menggunakan struktur data Frequent Pattern Tree (FP-Tree).
- Tidak menghasilkan kandidat itemset, sehingga lebih efisien.
- Cocok untuk dataset besar seperti yang digunakan dalam penelitian ini.

3. Kedua algoritma diterapkan dengan parameter:

- Minimum Support: 0.02
- Minimum Confidence: 0.3

3.5 Evaluasi dan Validasi Hasil

3.5.1 Evaluasi Aturan Asosiasi

Evaluasi dilakukan dengan menilai kualitas aturan asosiasi yang dihasilkan dari algoritma Apriori dan FP-Growth menggunakan tiga metrik utama:

- **Support:** Menyatakan proporsi transaksi yang mengandung kombinasi item tertentu. Semakin tinggi support, semakin sering kombinasi tersebut muncul dalam data.
- **Confidence:** Mengukur probabilitas item consequents muncul dalam transaksi yang juga mengandung item antecedents. Confidence yang tinggi menunjukkan aturan yang lebih dapat dipercaya.
- **Lift:** Menunjukkan kekuatan hubungan antar item dibandingkan dengan jika keduanya muncul secara acak. Nilai lift > 1 menunjukkan adanya asosiasi positif.

Untuk meningkatkan relevansi aturan, dilakukan filtering rule, yaitu hanya aturan-aturan dengan nilai confidence dan/atau lift tertentu yang ditampilkan. Hal ini bertujuan agar hasil analisis lebih fokus pada pola-pola yang signifikan secara statistik.

3.5.2 Visualisasi dan Analisis Perbandingan

Visualisasi dilakukan dengan menampilkan tabel aturan asosiasi hasil dari Apriori dan FP-Growth, yang mencakup:

- Kombinasi antecedents \rightarrow consequents
- Nilai support, confidence, dan lift
- Perbandingan antara algoritma Apriori dan FP-Growth untuk aturan yang sama

Dari hasil tersebut terlihat bahwa kedua algoritma mampu menghasilkan aturan dengan kualitas metrik yang identik untuk pola yang sama. Namun, FP-Growth cenderung lebih efisien secara waktu komputasi karena tidak perlu menghasilkan kandidat itemset secara eksplisit seperti pada Apriori.

| Algoritma | antecedents | consequents | support | confidence | lift |
|-----------|---|--------------------------------------|----------|------------|-----------|
| Apriori | (60 TEATIME FAIRY CAKE CASES) | (PACK OF 72 RETROSPOT CAKE CASES) | 0.020457 | 0.503012 | 7.889220 |
| FP-Growth | (60 TEATIME FAIRY CAKE CASES) | (PACK OF 72 RETROSPOT CAKE CASES) | 0.020457 | 0.503012 | 7.889220 |
| Apriori | (ROSES REGENCY TEACUP AND SAUCER , GREEN REGENCY TEACUP AND SAUCER) | (PINK REGENCY TEACUP AND SAUCER) | 0.024316 | 0.712747 | 19.690386 |
| FP-Growth | (ROSES REGENCY TEACUP AND SAUCER , GREEN REGENCY TEACUP AND SAUCER) | (PINK REGENCY TEACUP AND SAUCER) | 0.024316 | 0.712747 | 19.690386 |
| Apriori | (WOODEN FRAME ANTIQUE WHITE) | (WOODEN PICTURE FRAME WHITE FINISH) | 0.029032 | 0.573850 | 10.480142 |
| FP-Growth | (WOODEN FRAME ANTIQUE WHITE) | (WOODEN PICTURE FRAME WHITE FINISH) | 0.029032 | 0.573850 | 10.480142 |
| Apriori | (WOODEN PICTURE FRAME WHITE FINISH) | (WOODEN FRAME ANTIQUE WHITE) | 0.029032 | 0.530201 | 10.480142 |
| FP-Growth | (WOODEN PICTURE FRAME WHITE FINISH) | (WOODEN FRAME ANTIQUE WHITE) | 0.029032 | 0.530201 | 10.480142 |
| Apriori | (RED HANGING HEART T-LIGHT HOLDER) | (WHITE HANGING HEART T-LIGHT HOLDER) | 0.028174 | 0.664740 | 5.391559 |
| FP-Growth | (RED HANGING HEART T-LIGHT HOLDER) | (WHITE HANGING HEART T-LIGHT HOLDER) | 0.028174 | 0.664740 | 5.391559 |
| Apriori | (PAPER CHAIN KIT VINTAGE CHRISTMAS) | (PAPER CHAIN KIT 50'S CHRISTMAS) | 0.027807 | 0.640339 | 10.560411 |
| FP-Growth | (PAPER CHAIN KIT VINTAGE CHRISTMAS) | (PAPER CHAIN KIT 50'S CHRISTMAS) | 0.027807 | 0.640339 | 10.560411 |
| Apriori | (PACK OF 60 PINK PAISLEY CAKE CASES) | (PACK OF 72 RETROSPOT CAKE CASES) | 0.020702 | 0.500741 | 7.853597 |
| FP-Growth | (PACK OF 60 PINK PAISLEY CAKE CASES) | (PACK OF 72 RETROSPOT CAKE CASES) | 0.020702 | 0.500741 | 7.853597 |
| Apriori | (LUNCH BAG WOODLAND) | (LUNCH BAG RED RETROSPOT) | 0.026888 | 0.523242 | 6.428120 |
| FP-Growth | (LUNCH BAG WOODLAND) | (LUNCH BAG RED RETROSPOT) | 0.026888 | 0.523242 | 6.428120 |
| Apriori | (LUNCH BAG PINK POLKADOT) | (LUNCH BAG RED RETROSPOT) | 0.032462 | 0.557388 | 6.846629 |
| FP-Growth | (LUNCH BAG PINK POLKADOT) | (LUNCH BAG RED RETROSPOT) | 0.032462 | 0.557388 | 6.846629 |
| Apriori | (LUNCH BAG DOLLY GIRL DESIGN) | (LUNCH BAG SPACEBOY DESIGN) | 0.021804 | 0.588430 | 9.597695 |
| FP-Growth | (LUNCH BAG DOLLY GIRL DESIGN) | (LUNCH BAG SPACEBOY DESIGN) | 0.021804 | 0.588430 | 9.597695 |
| Apriori | (JUMBO STORAGE BAG SUKI) | (JUMBO BAG RED RETROSPOT) | 0.027133 | 0.560051 | 5.565396 |
| FP-Growth | (JUMBO STORAGE BAG SUKI) | (JUMBO BAG RED RETROSPOT) | 0.027133 | 0.560051 | 5.565396 |
| Apriori | (JUMBO BAG STRAWBERRY) | (JUMBO BAG RED RETROSPOT) | 0.025847 | 0.632684 | 6.287174 |
| ... | | | | | |
| Apriori | (JUMBO BAG PINK POLKADOT) | (JUMBO BAG RED RETROSPOT) | 0.033932 | 0.626697 | 6.227681 |
| FP-Growth | (JUMBO BAG PINK POLKADOT) | (JUMBO BAG RED RETROSPOT) | 0.033932 | 0.626697 | 6.227681 |
| Apriori | (GREEN REGENCY TEACUP AND SAUCER , PINK REGENCY TEACUP AND SAUCER) | (ROSES REGENCY TEACUP AND SAUCER) | 0.024316 | 0.842887 | 16.461512 |
| FP-Growth | (GREEN REGENCY TEACUP AND SAUCER , PINK REGENCY TEACUP AND SAUCER) | (ROSES REGENCY TEACUP AND SAUCER) | 0.024316 | 0.842887 | 16.461512 |

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...

Gambar 3. 1 Visualisasi dan Analisis Perbandingan

Menampilkan perbandingan hasil aturan asosiasi yang diperoleh dari algoritma Apriori dan FP-Growth setelah dilakukan filtering rule berdasarkan nilai confidence dan lift. Terlihat bahwa beberapa aturan yang ditemukan oleh Apriori juga ditemukan oleh FP-Growth dengan nilai metrik yang identik. Contohnya adalah kombinasi (ROSES REGENCY TEACUP AND SAUCER, GREEN TEACUP...) \rightarrow (PINK REGENCY TEACUP AND SAUCER) yang memiliki nilai lift sangat tinggi, menunjukkan asosiasi yang kuat. Hal ini menunjukkan bahwa meskipun kedua algoritma menggunakan pendekatan yang berbeda, kualitas aturan yang dihasilkan tetap konsisten.

3.5.3 Catatan Evaluasi

Karena Association Rule Mining termasuk dalam kategori unsupervised learning, maka evaluasi tidak menggunakan metrik seperti accuracy, precision, recall, F1-score, maupun confusion matrix. Fokus utama evaluasi adalah pada metrik rule interestingness (support, confidence, lift) dan relevansi aturan terhadap konteks bisnis atau pola transaksi yang ingin ditemukan.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Visualisasi eksploratif (EDA)

Tahap eksplorasi data atau Exploratory Data Analysis (EDA) bertujuan untuk memahami struktur, pola, dan karakteristik umum dari dataset Online Retail. Proses ini penting untuk memastikan kualitas data sebelum dilakukan pemodelan menggunakan algoritma asosiasi.

4.1.1 Struktur Data

Dataset yang digunakan berasal dari transaksi penjualan retail online di kawasan Eropa pada tahun 2010–2011. Dataset ini memiliki 8 atribut utama dengan total 541.909 baris data.

Tabel 1 Struktur Data

| Nama Kolom | Tipe Data | Deskripsi |
|-------------------|------------------|-----------------------------|
| InvoiceNo | String | Nomor faktur transaksi |
| StockCode | String | Kode unik produk |
| Description | String | Nama atau deskripsi produk |
| Quantity | Int | Jumlah produk yang dibeli |
| InnvoiceDate | Datetime | Tanggal dan waktu transaksi |
| UnitPrice | Float | Harga per unit produk |
| Customer Id | Float | ID pelanggan |
| Country | String | Negara asal transaksi |

4.1.2 Cek Missing Value dan Duplikasi Data

Berdasarkan hasil eksplorasi awal, ditemukan adanya nilai kosong (missing values) pada dua kolom penting dalam dataset. Kolom CustomerID memiliki sebanyak 135.080 baris kosong, sedangkan kolom Description memiliki 1.454 baris kosong. Nilai kosong pada kolom CustomerID menunjukkan bahwa sebagian transaksi tidak dapat diidentifikasi pemiliknya, sedangkan kosongnya kolom Description dapat menyebabkan hilangnya informasi produk yang dijual.


```

InvoiceNo      0
StockCode      0
Description    1454
Quantity      0
InvoiceDate    0
UnitPrice      0
CustomerID     135080
Country        0
dtype: int64
InvoiceNo StockCode Description Quantity \
622 536414 22139 NaN 56
1443 536544 21773 DECORATIVE ROSE BATHROOM BOTTLE 1
1444 536544 21774 DECORATIVE CATS BATHROOM BOTTLE 2
1445 536544 21786 POLKADOT RAIN HAT 4
1446 536544 21787 RAIN PONCHO RETROSPOT 2

InvoiceDate UnitPrice CustomerID Country
622 2010-12-01 11:52:00 0.00 NaN United Kingdom
1443 2010-12-01 14:32:00 2.51 NaN United Kingdom
1444 2010-12-01 14:32:00 2.51 NaN United Kingdom
1445 2010-12-01 14:32:00 0.85 NaN United Kingdom
1446 2010-12-01 14:32:00 1.66 NaN United Kingdom
Baris duplikat: 5268
InvoiceNo StockCode Description Quantity \
517 536409 21866 UNION JACK FLAG LUGGAGE TAG 1
...
527 2010-12-01 11:45:00 2.10 17908.0 United Kingdom
537 2010-12-01 11:45:00 2.95 17908.0 United Kingdom
539 2010-12-01 11:45:00 4.95 17908.0 United Kingdom
555 2010-12-01 11:45:00 2.05 17908.0 United Kingdom

```

Gambar 4. 1 Missing Value dan Duplikasi data

Selain itu, terdeteksi adanya 5.268 baris duplikat dalam dataset. Duplikasi ini berpotensi mengganggu keakuratan analisis karena satu transaksi bisa tercatat lebih dari sekali. Untuk memastikan kualitas data dan akurasi model yang akan digunakan pada tahap selanjutnya, diperlukan proses pembersihan data. Proses ini meliputi penghapusan data yang memiliki nilai kosong pada kolom penting serta penghapusan duplikat agar tidak memengaruhi hasil analisis asosiasi nantinya.

4.1.3 Statistik Deskriptif

```

Total data transaksi: 541909
Pelanggan unik: 4372
Produk unik: 4223
      Quantity      UnitPrice
count  541909.000000  541909.000000
mean      9.552250      4.611114
std     218.081158     96.759853
min    -80995.000000  -11062.060000
25%       1.000000      1.250000
50%       3.000000      2.080000
75%      10.000000      4.130000
max     80995.000000  38970.000000
Quantity      3.00
UnitPrice      2.08
dtype: float64

```

Gambar 4. 2 Statistik Deskriptif

Dataset Online Retail terdiri dari 541.909 data transaksi, dengan 4.372 pelanggan unik dan 4.223 produk unik. Rata-rata jumlah produk yang dibeli

(Quantity) adalah 9,55, dengan median 3. Nilai Quantity berkisar dari -80.995 hingga 80.995, menunjukkan kemungkinan adanya pengembalian barang.

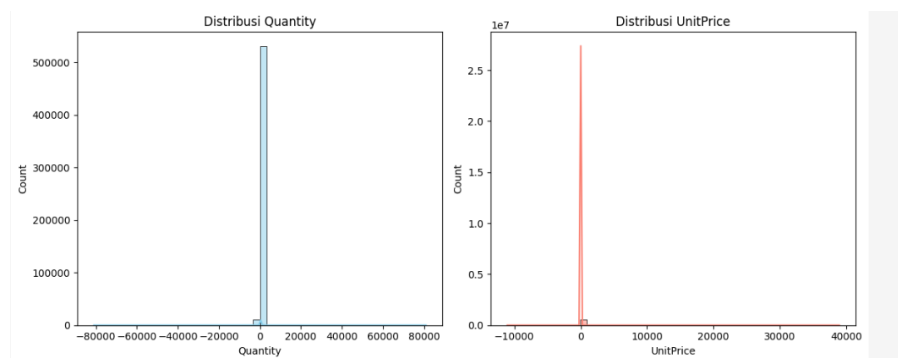
Harga per unit produk (UnitPrice) memiliki rata-rata 4,61 dan median 2,08, dengan nilai ekstrem dari -11.062 hingga 38.970. Nilai-nilai negatif pada Quantity dan UnitPrice menunjukkan adanya data tidak valid yang perlu dibersihkan sebelum proses analisis atau pemodelan lebih lanjut.

4.1.4 Visualisasi Data

Visualisasi data digunakan untuk memahami pola distribusi dan hubungan antar fitur numerik dalam dataset. Tahapan ini penting untuk mengidentifikasi potensi outlier, distribusi tidak normal, serta hubungan antar variabel yang dapat memengaruhi hasil pemodelan.

a. Distribusi Quantity dan UnitPrice

Histogram ini menunjukkan distribusi kedua fitur yang sangat miring ke kanan, dengan konsentrasi nilai pada rentang rendah. Terlihat pula beberapa nilai ekstre yang mengindikasikan adanya outlier signifikan.

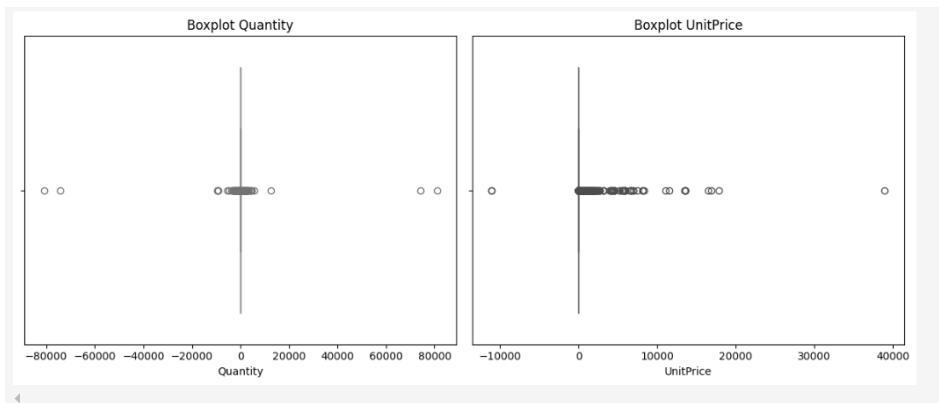


Gambar 4. 3 Distribusi Quantity & UnitPrice

Histogram menunjukkan bahwa sebagian besar transaksi memiliki jumlah barang dan harga satuan yang kecil. Distribusi keduanya sangat tidak normal (skewed ke kanan), dengan beberapa nilai ekstrem yang jauh lebih besar dari mayoritas data. Hal ini memperkuat dugaan adanya outlier, terutama pada Quantity yang mencapai ± 80.000 dan UnitPrice hingga ± 39.000

b. Boxplot Quantity dan UnitPrice

Boxplot pada Gambar 4.2 memperjelas keberadaan outlier pada kedua fitur. Sebagian besar nilai terkonsentrasi di tengah, sementara banyak nilai yang berada jauh di luar rentang interkuartil.

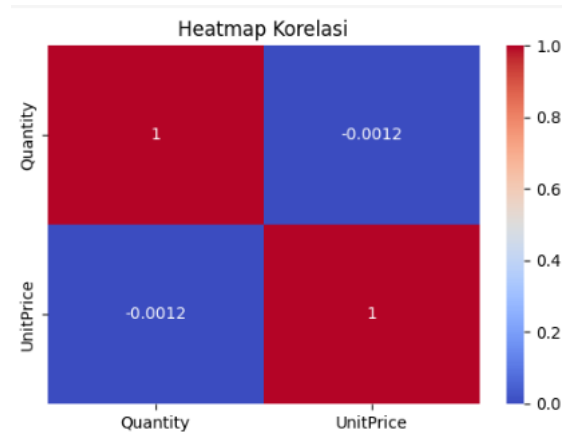


Gambar 4. 4 Boxplot Quantity & Boxplot UnitPrice

Boxplot memperkuat hasil histogram. Banyak titik data berada di luar batas normal (di atas Q3 dan di bawah Q1), mengindikasikan keberadaan outlier signifikan. Kondisi ini berpotensi mengganggu hasil pemodelan jika tidak ditangan

c. Korelasi antar fitur numerik

Gambar 4.3 menunjukkan heatmap korelasi antara Quantity dan UnitPrice yang sangat rendah, hampir mendekati nol. Hal ini mengindikasikan bahwa tidak terdapat hubungan linier yang berarti antara kedua fitur ini.



Gambar 4. 5 Heatmap Korelasi

Hasil visualisasi heatmap menunjukkan korelasi sangat lemah antara Quantity dan UnitPrice, dengan nilai mendekati nol (~ -0.0012). Artinya, harga suatu produk tidak berpengaruh besar terhadap jumlah pembelian. Ini penting karena fitur-fitur tersebut dapat dianalisis secara independen.

4.2 Hasil preprocessing dan pemodelan

Bagian ini menjelaskan proses pembersihan data (preprocessing) dan langkah-langkah pemodelan yang digunakan untuk menemukan pola asosiasi dalam transaksi pembelian.

4.2.1 Hasil Preprocessing

Proses preprocessing dilakukan untuk membersihkan dan menyiapkan data sebelum dimasukkan ke dalam proses pemodelan frequent pattern mining. Dataset awal yang digunakan merupakan data transaksi penjualan retail online yang berisi berbagai atribut seperti InvoiceNo, Description, Quantity, UnitPrice, CustomerID, dan Country. Pada tahap awal, dilakukan penghapusan nilai kosong pada kolom Description dan CustomerID, karena nilai kosong tersebut tidak relevan untuk dianalisis. Tercatat sebanyak 1.454 baris pada kolom Description dan 135.080 baris pada kolom CustomerID dihapus dari dataset. Selain itu, sebanyak 5.268 baris duplikat juga dihapus untuk menghindari redundansi data.

| | InvoiceNo | StockCode | Description | Quantity | \ |
|---|-----------|-----------|-------------------------------------|----------|---|
| 0 | 536365 | 85123A | WHITE HANGING HEART T-LIGHT HOLDER | 6 | |
| 1 | 536365 | 71053 | WHITE METAL LANTERN | 6 | |
| 2 | 536365 | 84406B | CREAM CUPID HEARTS COAT HANGER | 8 | |
| 3 | 536365 | 84029G | KNITTED UNION FLAG HOT WATER BOTTLE | 6 | |
| 4 | 536365 | 84029E | RED WOOLLY HOTTIE WHITE HEART. | 6 | |
| 5 | 536365 | 22752 | SET 7 BABUSHKA NESTING BOXES | 2 | |
| 6 | 536365 | 21730 | GLASS STAR FROSTED T-LIGHT HOLDER | 6 | |
| 7 | 536366 | 22633 | HAND WARMER UNION JACK | 6 | |
| 8 | 536366 | 22632 | HAND WARMER RED POLKA DOT | 6 | |
| 9 | 536367 | 84879 | ASSORTED COLOUR BIRD ORNAMENT | 32 | |

| | InvoiceDate | UnitPrice | CustomerID | Country |
|---|---------------------|-----------|------------|----------------|
| 0 | 2010-12-01 08:26:00 | 2.55 | 17850.0 | United Kingdom |
| 1 | 2010-12-01 08:26:00 | 3.39 | 17850.0 | United Kingdom |
| 2 | 2010-12-01 08:26:00 | 2.75 | 17850.0 | United Kingdom |
| 3 | 2010-12-01 08:26:00 | 3.39 | 17850.0 | United Kingdom |
| 4 | 2010-12-01 08:26:00 | 3.39 | 17850.0 | United Kingdom |
| 5 | 2010-12-01 08:26:00 | 7.65 | 17850.0 | United Kingdom |
| 6 | 2010-12-01 08:26:00 | 4.25 | 17850.0 | United Kingdom |
| 7 | 2010-12-01 08:28:00 | 1.85 | 17850.0 | United Kingdom |
| 8 | 2010-12-01 08:28:00 | 1.85 | 17850.0 | United Kingdom |
| 9 | 2010-12-01 08:34:00 | 1.69 | 13047.0 | United Kingdom |

Gambar 4. 6 Hasil Preprocessing 10 data

Setelah itu, data dengan nilai negatif pada Quantity dan UnitPrice juga dibersihkan karena dianggap tidak merepresentasikan transaksi penjualan yang valid. Kemudian, untuk memfokuskan model pada produk yang paling relevan, dilakukan filtering terhadap 100 produk dengan frekuensi penjualan tertinggi. Langkah selanjutnya adalah mengelompokkan produk berdasarkan InvoiceNo, sehingga setiap transaksi direpresentasikan sebagai kumpulan produk yang dibeli secara bersamaan. Transformasi ini menghasilkan struktur list of list, yang kemudian diubah menjadi bentuk one-hot encoding menggunakan TransactionEncoder dari pustaka mlxtend.

Contoh encoding:

| | 4 PURPLE FLOCK DINNER CANDLES | 50'S CHRISTMAS GIFT BAG LARGE |
|---|-------------------------------|-------------------------------|
| 0 | False | False |
| 1 | False | False |
| 2 | False | False |
| 3 | False | False |
| 4 | False | False |

| | DOLLY GIRL BEAKER | I LOVE LONDON MINI BACKPACK \ |
|---|-------------------|-------------------------------|
| 0 | False | False |
| 1 | False | False |
| 2 | False | False |
| 3 | False | False |
| 4 | False | False |

| | I LOVE LONDON MINI RUCKSACK | NINE DRAWER OFFICE TIDY \ |
|---|-----------------------------|---------------------------|
| 0 | False | False |
| 1 | False | False |
| 2 | False | False |
| 3 | False | False |
| 4 | False | False |

| | OVAL WALL MIRROR DIAMANTE | RED SPOT GIFT BAG LARGE \ |
|-----|---------------------------|---------------------------|
| 0 | False | False |
| 1 | False | False |
| ... | | |
| 3 | False | False |
| 4 | False | False |

Gambar 4. 7 Hasil Preprocessing Encoding

Dimana setiap kolom merepresentasikan satu produk dan nilainya berupa boolean (True jika produk dibeli dalam transaksi tersebut, False jika tidak).

4.2.2 Persiapan Pemodelan

Pada tahap ini, data hasil preprocessing digunakan sebagai input untuk proses pemodelan dengan algoritma Apriori dan FP-Growth. Sebelum itu, dilakukan pemilihan 100 produk terpopuler berdasarkan frekuensi kemunculannya agar model lebih fokus pada item yang memiliki relevansi tinggi. Selanjutnya, data dikelompokkan berdasarkan nomor invoice untuk membentuk list transaksi yang akan diolah menjadi format one-hot encoding menggunakan TransactionEncoder.

```

1 # Hitung frekuensi item (produk)
2 top_items = df['Description'].value_counts().head(100).index.tolist()
3
4 # Filter data hanya yang termasuk dalam 100 produk terbanyak
5 df_filtered = df[df['Description'].isin(top_items)]
6 # Group per transaksi
7 transactions = df_filtered.groupby('InvoiceNo')['Description'].apply(list).tolist()
8
9 # One-hot encoding
10 te = TransactionEncoder()
11 te_ary = te.fit(transactions).transform(transactions)
12 df_encoded = pd.DataFrame(te_ary, columns=te.columns_)
13
14

```

Gambar 4. 8 Persiapan Pemodelan

4.2.3 Penerapan Algoritma Apriori dan FP-Growth

Setelah data transaksi dikonversi ke dalam format biner melalui proses one-hot encoding, langkah selanjutnya adalah mencari frequent itemsets dengan menggunakan dua algoritma, yaitu Apriori dan FP-Growth. Keduanya diterapkan dengan nilai minimum support sebesar 0.02, yang berarti hanya itemset yang muncul minimal pada 2% dari total transaksi yang akan dipertimbangkan. Selanjutnya, dari itemset yang terbentuk, dihasilkan aturan asosiasi (association rules) dengan menerapkan nilai minimum confidence sebesar 0.3.

```
1 # ===== APRIORI =====
2 frequent_apriori = apriori(df_encoded, min_support=0.02, use_colnames=True)
3 rules_apriori = association_rules(frequent_apriori, metric="confidence", min_threshold=0.3)
4
5 # === FILTERING RULE (contoh: lift > 1.2 dan confidence > 0.5) ===
6 filtered_rules_apriori = rules_apriori[
7     (rules_apriori['lift'] > 1.2) & (rules_apriori['confidence'] > 0.5)
8 ]
9
10 print("Filtered Rules APRIORI:")
11 print(filtered_rules_apriori[['antecedents', 'consequents', 'support', 'confidence', 'lift']].head())
12
```

Gambar 4. 9 Penerapan Algoritma Apriori

```
1
2 frequent_fp = fpgrowth(df_encoded, min_support=0.02, use_colnames=True)
3 rules_fp = association_rules(frequent_fp, metric="confidence", min_threshold=0.3)
4
5 # === FILTERING RULE ===
6 filtered_rules_fp = rules_fp[
7     (rules_fp['lift'] > 1.2) & (rules_fp['confidence'] > 0.5)
8 ]
9
10 print("Filtered Rules FP-GROWTH:")
11 print(filtered_rules_fp[['antecedents', 'consequents', 'support', 'confidence', 'lift']].head())
```

Gambar 4. 10 Penerapan Algoritma FP-Growth

Untuk memastikan kualitas dan relevansi aturan yang dihasilkan, dilakukan penyaringan tambahan dengan kriteria confidence lebih dari 0.5 dan lift lebih dari 1.2, sehingga hanya aturan yang memiliki kekuatan asosiasi yang tinggi yang disertakan dalam hasil akhir.

4.2.4 Hasil Pemodelan dan Visualisasi

Setelah dilakukan proses pencarian frequent itemsets dan pembuatan aturan asosiasi menggunakan algoritma Apriori dan FP-Growth, diperoleh hasil yang

menunjukkan pola keterkaitan antar produk yang sering dibeli bersamaan oleh pelanggan.

a. Aturan Asosiasi (Apriori)

Dari proses pemodelan menggunakan algoritma Apriori, diperoleh lima aturan asosiasi terbaik setelah dilakukan penyaringan berdasarkan nilai support minimal 2 persen, confidence lebih dari 50 persen, dan lift di atas 1,2. Aturan-aturan ini menunjukkan keterkaitan pembelian produk secara bersamaan yang cukup signifikan.

Contohnya, produk ALARM CLOCK BAKELIKE GREEN sangat sering dibeli bersamaan dengan ALARM CLOCK BAKELIKE RED, dengan nilai confidence sebesar 66,3 persen dan lift sebesar 11,93. Hal ini menunjukkan bahwa apabila seorang pelanggan membeli jam alarm hijau, maka kemungkinan besar ia juga membeli jam alarm merah. Selain itu, hubungan antara produk 60 TEATIME FAIRY CAKE CASES dan PACK OF 72 RETROSPOT CAKE CASES juga menunjukkan pola pembelian bersama yang kuat, dengan lift sebesar 7,88.

| Filtered Rules APRIORI: | | | |
|-------------------------|-------------------------------|-----------------------------------|-----------|
| | antecedents | consequents | support \ |
| 0 | (60 TEATIME FAIRY CAKE CASES) | (PACK OF 72 RETROSPOT CAKE CASES) | 0.020457 |
| 3 | (ALARM CLOCK BAKELIKE PINK) | (ALARM CLOCK BAKELIKE GREEN) | 0.021069 |
| 4 | (ALARM CLOCK BAKELIKE RED) | (ALARM CLOCK BAKELIKE GREEN) | 0.032829 |
| 5 | (ALARM CLOCK BAKELIKE GREEN) | (ALARM CLOCK BAKELIKE RED) | 0.032829 |
| 7 | (ALARM CLOCK BAKELIKE PINK) | (ALARM CLOCK BAKELIKE RED) | 0.024683 |
| | confidence | lift | |
| 0 | 0.503012 | 7.889220 | |
| 3 | 0.546900 | 11.037372 | |
| 4 | 0.590959 | 11.926565 | |
| 5 | 0.662546 | 11.926565 | |
| 7 | 0.640700 | 11.533298 | |

Gambar 4. 11 Aturan Asosiasi (Apriori)

Secara umum, aturan-aturan yang dihasilkan oleh Apriori cenderung melibatkan produk yang memiliki kemiripan kategori, seperti dekorasi rumah atau perlengkapan dapur, sehingga mendukung validitas hasil pemodelan

b. Aturan Asosiasi (FP-Growth)

Pemodelan menggunakan algoritma FP-Growth dilakukan dengan parameter yang sama seperti Apriori, yaitu minimum support sebesar 0,02 dan minimum confidence sebesar 0,3. Setelah proses ekstraksi frequent itemsets dan pembentukan aturan asosiasi, dilakukan penyaringan berdasarkan nilai confidence lebih dari 0,5 dan lift di atas 1,2.

Dari hasil yang diperoleh, ditemukan lima aturan asosiasi terbaik. Seperti halnya Apriori, FP-Growth juga menunjukkan bahwa produk-produk seperti ALARM CLOCK BAKELIKE RED dan ALARM CLOCK BAKELIKE GREEN memiliki asosiasi yang kuat. Salah satu aturan menunjukkan bahwa jika pelanggan membeli ALARM CLOCK BAKELIKE GREEN, maka

kemungkinan besar mereka juga akan membeli ALARM CLOCK BAKELIKE RED, dengan confidence sebesar 66,3 persen dan lift 11,93.

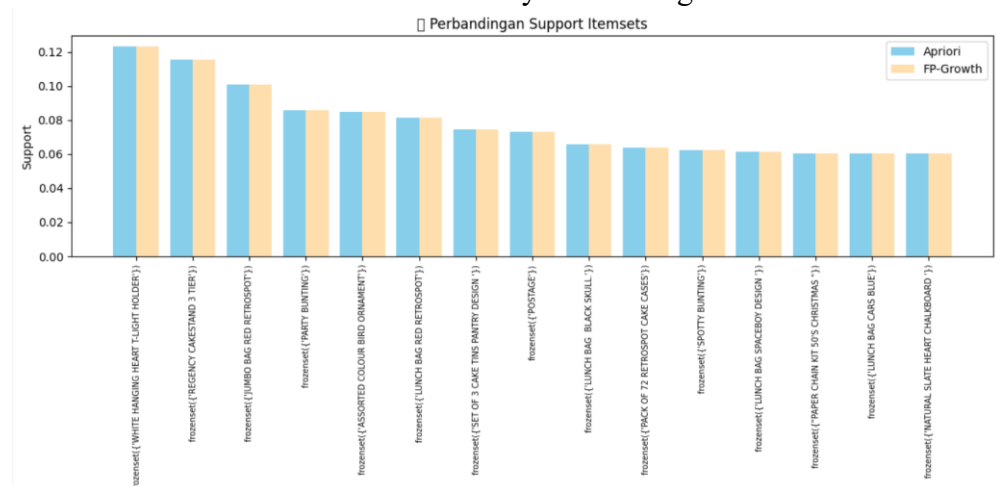
| Filtered Rules FP-GROWTH: | | | |
|---------------------------|-------------------------------------|-------------------------------|-----------|
| | antecedents | consequents | |
| 0 | (ALARM CLOCK BAKELIKE RED) | (ALARM CLOCK BAKELIKE GREEN) | |
| 1 | (ALARM CLOCK BAKELIKE GREEN) | (ALARM CLOCK BAKELIKE RED) | |
| 3 | (ALARM CLOCK BAKELIKE PINK) | (ALARM CLOCK BAKELIKE RED) | |
| 5 | (ALARM CLOCK BAKELIKE PINK) | (ALARM CLOCK BAKELIKE GREEN) | |
| 6 | (WOODEN PICTURE FRAME WHITE FINISH) | (WOODEN FRAME ANTIQUE WHITE) | |
| | support | confidence | lift |
| 0 | 0.032829 | 0.590959 | 11.926565 |
| 1 | 0.032829 | 0.662546 | 11.926565 |
| 3 | 0.024683 | 0.640700 | 11.533298 |
| 5 | 0.021069 | 0.546900 | 11.037372 |
| 6 | 0.029032 | 0.530201 | 10.480142 |

Gambar 4. 12 Aturan Asosiasi FP-Growth

Selain itu, FP-Growth juga menemukan asosiasi antara produk WOODEN PICTURE FRAME WHITE FINISH dan WOODEN FRAME ANTIQUE WHITE dengan nilai lift sebesar 10,48. Hal ini menunjukkan bahwa algoritma FP-Growth mampu menangkap pola pembelian produk dekoratif rumah yang juga berkorelasi erat.

c. Perbandingan Support Itemsets

Untuk mengevaluasi konsistensi antara kedua algoritma, dilakukan analisis perbandingan terhadap itemsets yang dihasilkan. Ditemukan bahwa terdapat 149 itemsets yang identik antara Apriori dan FP-Growth, dan tidak ditemukan perbedaan signifikan ($>1\%$) dalam nilai support-nya. Ini menandakan bahwa meskipun proses komputasi keduanya berbeda, hasil akhir dalam hal itemsets dan frekuensinya cenderung stabil.



Gambar 4. 13 Perbandingan Support Itemsets

Visualisasi dalam bentuk grafik batang memperlihatkan nilai support dari 15 itemsets teratas. Tampak bahwa perbedaan nilai support antara kedua

algoritma sangat kecil, memperkuat hasil bahwa keduanya dapat digunakan secara bergantian tanpa mengorbankan akurasi pola.

4.3 Tabel Hasil eksperimen/model

Pada tahap ini dilakukan perbandingan hasil antara algoritma Apriori dan FP-Growth berdasarkan aturan asosiasi yang dihasilkan setelah melalui proses filtering menggunakan parameter minimum support dan confidence. Hasil menunjukkan bahwa kedua algoritma menghasilkan pola-pola asosiasi yang relatif serupa, baik dari sisi kombinasi produk yang terlibat maupun nilai evaluasi metrik seperti support, confidence, dan lift. Sebagai contoh, aturan seperti “60 TEATIME FAIRY CAKE CASES → PACK OF 72 RETROSPOT CAKE CASES” dan “LUNCH BAG WOODLAND → LUNCH BAG RED RETROSPOT” muncul pada kedua algoritma dengan nilai evaluasi yang identik. Hal ini mengindikasikan bahwa baik Apriori maupun FP-Growth mampu mendeteksi asosiasi yang sama kuatnya ketika parameter yang digunakan serupa.

TABEL PERBANDINGAN ATURAN APRIORI vs FP-GROWTH (Setelah Filtering):

| Algoritma | antecedents | consequents | support | confidence | lift |
|---|-------------------------------|--------------------------------------|----------|------------|-----------|
| Apriori | (60 TEATIME FAIRY CAKE CASES) | (PACK OF 72 RETROSPOT CAKE CASES) | 0.020457 | 0.503012 | 7.889220 |
| FP-Growth | (60 TEATIME FAIRY CAKE CASES) | (PACK OF 72 RETROSPOT CAKE CASES) | 0.020457 | 0.503012 | 7.889220 |
| Apriori (ROSES REGENCY TEACUP AND SAUCER , GREEN REGENCY TEACUP AND SAUCER) | | (PINK REGENCY TEACUP AND SAUCER) | 0.024316 | 0.712747 | 19.690386 |
| FP-Growth (ROSES REGENCY TEACUP AND SAUCER , GREEN REGENCY TEACUP AND SAUCER) | | (PINK REGENCY TEACUP AND SAUCER) | 0.024316 | 0.712747 | 19.690386 |
| Apriori (WOODEN FRAME ANTIQUE WHITE) | | (WOODEN PICTURE FRAME WHITE FINISH) | 0.029032 | 0.573850 | 10.480142 |
| FP-Growth (WOODEN FRAME ANTIQUE WHITE) | | (WOODEN PICTURE FRAME WHITE FINISH) | 0.029032 | 0.573850 | 10.480142 |
| Apriori (WOODEN PICTURE FRAME WHITE FINISH) | | (WOODEN FRAME ANTIQUE WHITE) | 0.029032 | 0.530201 | 10.480142 |
| FP-Growth (WOODEN PICTURE FRAME WHITE FINISH) | | (WOODEN FRAME ANTIQUE WHITE) | 0.029032 | 0.530201 | 10.480142 |
| Apriori (RED HANGING HEART T-LIGHT HOLDER) | | (WHITE HANGING HEART T-LIGHT HOLDER) | 0.028174 | 0.664740 | 5.391559 |
| FP-Growth (RED HANGING HEART T-LIGHT HOLDER) | | (WHITE HANGING HEART T-LIGHT HOLDER) | 0.028174 | 0.664740 | 5.391559 |
| Apriori (PAPER CHAIN KIT VINTAGE CHRISTMAS) | | (PAPER CHAIN KIT 50'S CHRISTMAS) | 0.027807 | 0.640339 | 10.560411 |
| FP-Growth (PAPER CHAIN KIT VINTAGE CHRISTMAS) | | (PAPER CHAIN KIT 50'S CHRISTMAS) | 0.027807 | 0.640339 | 10.560411 |
| Apriori (PACK OF 60 PINK PAISLEY CAKE CASES) | | (PACK OF 72 RETROSPOT CAKE CASES) | 0.020702 | 0.500741 | 7.853597 |
| FP-Growth (PACK OF 60 PINK PAISLEY CAKE CASES) | | (PACK OF 72 RETROSPOT CAKE CASES) | 0.020702 | 0.500741 | 7.853597 |
| Apriori (LUNCH BAG WOODLAND) | | (LUNCH BAG RED RETROSPOT) | 0.026888 | 0.523242 | 6.428120 |
| FP-Growth (LUNCH BAG WOODLAND) | | (LUNCH BAG RED RETROSPOT) | 0.026888 | 0.523242 | 6.428120 |
| Apriori (LUNCH BAG PINK POLKADOT) | | (LUNCH BAG RED RETROSPOT) | 0.032462 | 0.557308 | 6.846629 |
| FP-Growth (LUNCH BAG PINK POLKADOT) | | (LUNCH BAG RED RETROSPOT) | 0.032462 | 0.557308 | 6.846629 |
| Apriori (LUNCH BAG DOLLY GIRL DESIGN) | | (LUNCH BAG SPACEBOY DESIGN) | 0.021804 | 0.588430 | 9.597695 |
| FP-Growth (LUNCH BAG DOLLY GIRL DESIGN) | | (LUNCH BAG SPACEBOY DESIGN) | 0.021804 | 0.588430 | 9.597695 |
| Apriori (JUMBO STORAGE BAG SUKI) | | (JUMBO BAG RED RETROSPOT) | 0.027133 | 0.560051 | 5.565396 |
| FP-Growth (JUMBO STORAGE BAG SUKI) | | (JUMBO BAG RED RETROSPOT) | 0.027133 | 0.560051 | 5.565396 |
| Apriori (JUMBO BAG STRAWBERRY) | | (JUMBO BAG RED RETROSPOT) | 0.025847 | 0.632684 | 6.287174 |
| ... | | | | | |
| Apriori (JUMBO BAG PINK POLKADOT) | | (JUMBO BAG RED RETROSPOT) | 0.033932 | 0.626697 | 6.227681 |
| FP-Growth (JUMBO BAG PINK POLKADOT) | | (JUMBO BAG RED RETROSPOT) | 0.033932 | 0.626697 | 6.227681 |
| Apriori (GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER) | | (ROSES REGENCY TEACUP AND SAUCER) | 0.024316 | 0.842887 | 16.461512 |
| FP-Growth (GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER) | | (ROSES REGENCY TEACUP AND SAUCER) | 0.024316 | 0.842887 | 16.461512 |

Gambar 4. 14 Tabel Hasil eksperimen/model

Meskipun demikian, terdapat perbedaan dari sisi pendekatan teknis. Apriori bekerja dengan eksplorasi kombinasi item secara eksplisit sehingga dapat menjadi lambat saat menangani data dalam jumlah besar. Sebaliknya, FP-Growth membangun struktur data berupa pohon (FP-tree) yang memungkinkan pencarian frequent itemsets dilakukan secara efisien tanpa harus membentuk kandidat kombinasi terlebih dahulu. Dengan demikian, dalam hal efisiensi waktu dan performa, FP-Growth cenderung lebih unggul, terutama pada dataset berskala besar.

4.4 Analisis keunggulan dan keterbatasan

Setelah dilakukan perbandingan hasil antara algoritma Apriori dan FP-Growth, perlu dilakukan analisis lebih lanjut terkait kelebihan dan kekurangan dari masing-masing algoritma. Hal ini bertujuan untuk memahami karakteristik teknis serta efektivitas penggunaannya dalam konteks data yang dianalisis. Kedua algoritma memiliki kemampuan dalam menemukan frequent itemsets dan menghasilkan aturan asosiasi yang relevan. Namun, perbedaan dalam pendekatan teknis menyebabkan adanya variasi dalam efisiensi waktu komputasi dan kompleksitas proses. Berikut merupakan ringkasan keunggulan dan keterbatasan masing-masing algoritma dalam bentuk tabel.

Tabel 2 Analisis Keunggulan dan Keterbatasan

| Aspek | Apriori | Fp-Growth |
|---------------------|--|---|
| keunggulan | Proses kerja lebih mudah dipahami dan diimplementasikan | Waktu komputasi lebih cepat karena tidak membentuk kandidat eksplisit |
| | Fleksibel dalam mengontrol level kombinasi itemset | Struktur data FP-tree efisien dan hemat memori |
| | Hasilnya konsisten pada dataset kecil hingga menengah | Mampu menangani dataset besar dengan pola kompleks secara efisien |
| keterbatasan | Kurang efisien untuk dataset besar | Implementasi lebih kompleks dan sulit dipahami oleh pengguna |
| | Membentuk banyak kandidat itemsets, meningkatkan beban komputasi | Kurang fleksibel dalam membatasi kombinasi item tertentu secara eksplisit |

Algoritma Apriori lebih cocok digunakan dalam kasus analisis data skala kecil hingga menengah, karena prosesnya yang eksplisit dan mudah ditelusuri. Namun, karena algoritma ini membentuk kandidat itemset pada setiap iterasi, maka prosesnya bisa menjadi lambat dan memakan banyak memori ketika diterapkan pada dataset besar atau dengan jumlah item yang tinggi. Di sisi lain, FP-Growth mengatasi kelemahan tersebut dengan cara membangun struktur data berupa pohon (FP-tree), yang memungkinkan pencarian frequent itemsets dilakukan tanpa eksplorasi kombinasi secara eksplisit. Hal ini menjadikan FP-Growth lebih cepat dan efisien, terutama dalam skala data besar. Meskipun demikian, struktur pohon dan alur kerja FP-Growth relatif lebih kompleks dan memerlukan pemahaman mendalam untuk implementasinya.

Oleh karena itu, pemilihan algoritma sangat bergantung pada karakteristik data dan kebutuhan analisis. Jika efisiensi waktu menjadi prioritas dan data berskala besar, FP-

Growth lebih direkomendasikan. Namun jika analisis membutuhkan fleksibilitas dan keterlacakan proses, Apriori bisa menjadi pilihan yang lebih tepat.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dengan menggunakan algoritma Apriori dan FP-Growth terhadap dataset Online Retail, dapat diambil beberapa kesimpulan sebagai berikut:

1. Proses preprocessing yang tepat—meliputi pembersihan nilai kosong, penghapusan data duplikat, dan transformasi data ke dalam format one-hot encoding—sangat penting dalam mempersiapkan data untuk proses association rule mining.
2. Kedua algoritma, baik Apriori maupun FP-Growth, mampu menghasilkan aturan asosiasi yang valid dan relevan dalam konteks pola belanja konsumen, dengan nilai support, confidence, dan lift yang tinggi.
3. FP-Growth terbukti lebih efisien dibandingkan Apriori dalam hal waktu komputasi dan konsumsi memori karena tidak membentuk kandidat itemset secara eksplisit.
4. Hasil aturan asosiasi menunjukkan bahwa beberapa produk cenderung dibeli bersamaan, terutama yang berasal dari kategori produk sejenis. Hal ini dapat dimanfaatkan untuk strategi bundling, pengelompokan display barang, maupun peningkatan layanan rekomendasi.
5. Evaluasi terhadap kedua algoritma menunjukkan hasil yang hampir identik dari sisi itemset dan metrik evaluasi, sehingga keduanya dapat digunakan bergantian tergantung kebutuhan dan kompleksitas data.

5.2 Saran

Beberapa saran yang dapat diberikan untuk pengembangan lebih lanjut dari penelitian ini antara lain:

1. Penelitian selanjutnya dapat memperluas cakupan analisis dengan mempertimbangkan dimensi waktu (time series) atau segmentasi pelanggan (misalnya berdasarkan lokasi atau frekuensi pembelian).
2. Implementasi visualisasi interaktif berbasis web untuk hasil aturan asosiasi dapat membantu pelaku usaha memahami pola belanja dengan lebih intuitif dan real-time.
3. Penambahan algoritma lain seperti Eclat atau metode berbasis machine learning dapat menjadi alternatif pembanding untuk mendapatkan model yang lebih adaptif dan akurat.

4. Disarankan untuk menggunakan data transaksi yang lebih terkini agar hasil analisis lebih relevan terhadap tren dan preferensi konsumen saat ini.
5. Untuk penggunaan komersial, integrasi sistem rekomendasi berbasis aturan asosiasi ke dalam platform e-commerce secara langsung akan memberikan dampak nyata terhadap peningkatan pengalaman pelanggan dan konversi penjualan.

DAFTAR PUSTAKA

- [1] A. N. Rahmi and Y. A. Mikola, "Implementasi Algoritma Apriori Untuk Menentukan Pola Pembelian Pada Customer (Studi Kasus: Toko Bakoel Sembako)," *Information System Journal*, vol. 4, no. 1, pp. 14–19, 2021.
- [2] A. Fergina, P. A. Negara, A. Sujjada, and I. Sanjaya, "Implementasi Algoritma Apriori dan FP-Growth untuk Menganalisis Pola Pembelian Produk Skincare dan Kosmetik," *Jurnal Ilmiah KOMPUTASI*, vol. 23, no. 3, pp. 433–442, 2024.
- [3] F. Prasetyo and H. Hasugian, "Analisis pola pembelian produk makanan menggunakan algoritma FP-Growth untuk strategi penjualan," *IDEALIS: InDonEsiA journal Information System*, vol. 7, no. 1, pp. 11–20, 2024.
- [4] A. Yani, Z. Azmi, and D. Suherdi, "Implementasi Data Mining Menganalisa Data Penjualan Menggunakan Algoritma K-Means Clustering," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 2, no. 2, pp. 315–323, 2023.
- [5] M. Sholihin, P. D. Sari, A. Ikhsan, and A. Rahman, "Membangun Dashboard Analisis Perilaku Konsumen dengan pendekatan Market Basket Analysis," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 2, pp. 1379–1386, 2025.
- [6] I. A. Darmawan, M. F. Randy, I. Yunianto, M. M. Mutoffar, and M. T. P. Salis, "Penerapan Data Mining Menggunakan Algoritma Apriori Untuk Menentukan Pola Golongan Penyandang Masalah Kesejahteraan Sosial," *Sebatik*, vol. 26, no. 1, pp. 223–230, 2022.
- [7] F. Achmad, O. Nurdiawan, and Y. A. Wijaya, "Analisa Pola Transaksi Pembelian Konsumen Pada Toko Ritel Kesehatan Menggunakan Algoritma Fp-Growth," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 168–175, 2023.
- [8] A. Harahap, A. L. R. Perangin-Angin, K. Kumar, and S. P. Tamba, "Analisis Penerapan Data Mining Dalam Penentuan Tata Letak Barang Menggunakan Algoritma Apriori dan FP-Growth," *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 5, no. 2, pp. 291–300, 2022.
- [9] M. L. Radhitya, N. K. M. Widiyanti, M. D. P. Asana, B. K. Wijaya, and I. G. I. Sudipa, "Product Layout Analysis Based on Consumer Purchasing Patterns Using Apriori Algorithm," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 6, no. 3, pp. 1701–1711, 2024.
- [10] E. J. Omol, D. A. Onyango, L. W. Mburu, and P. A. Abuonji, "Apriori algorithm and market basket analysis to uncover consumer buying patterns: Case of a Kenyan supermarket," *Buana Information Technology and Computer Sciences (BIT and CS)*, vol. 5, no. 2, pp. 51–63, 2024.
- [11] B. Murdianto and A. Jananto, "INDONESIA Pola Asosiasi Untuk Rekomendasi Penataan Display Barang Menggunakan Algoritma Apriori dan FP-Growth (Study Kasus Gamefantasia Ada Swalayan Pati)," *Elkom: Jurnal Elektronika dan Komputer*, vol. 16, no. 1, pp. 109–120, 2023.
- [12] R. Mahtum and Z. Fatah, "OPTIMALISASI PENENTUAN TATA LETAK BARANG MENGGUNAKAN ALGORITMA FP-GROWTH PADA DATA TRANSAKSI PENJUALAN," *Jurnal TIMES*, vol. 13, no. 2, pp. 183–191, 2024.

LAMPIRAN

1. Lampiran A – *Dataset* dan Informasi Terkait

A. Lampiran A1 – Deskripsi *Dataset*

Sumber Data : Kaggle

Jumlah Data : 541.910 Data

Jumlah Atribut : 8

Deskripsi Atribut :

- Atribut 1: Invoice No
- Atribut 2: Stock Code
- Atribut 3: Description
- Atribut 4: Quantity
- Atribut 5: Invoice Date
- Atribut 6: Unit Price
- Atribut 7: Customer ID
- Atribut 8: Country

B. Lampiran A2 – Contoh *Dataset* Mentah (Raw)

| | A | B | C | D | E | F | G | H |
|----|---------|-----------|--------------------|----------|------------------|-----------|------------|----------------|
| | voiceNo | StockCode | Description | Quantity | InvoiceDate | UnitPrice | CustomerID | Country |
| 1 | 536365 | 85123A | WHITE HANGING HE | 6 | 01/12/2010 08:26 | 2,55 | 17850 | United Kingdom |
| 2 | 536365 | 71053 | WHITE METAL LANTI | 6 | 01/12/2010 08:26 | 3,39 | 17850 | United Kingdom |
| 3 | 536365 | 84406B | CREAM CUPID HEAR | 8 | 01/12/2010 08:26 | 2,75 | 17850 | United Kingdom |
| 4 | 536365 | 84029G | KNITTED UNION FLA | 6 | 01/12/2010 08:26 | 3,39 | 17850 | United Kingdom |
| 5 | 536365 | 84029E | RED WOOLLY HOTTI | 6 | 01/12/2010 08:26 | 3,39 | 17850 | United Kingdom |
| 6 | 536365 | 22752 | SET 7 BABUSHKA NE | 2 | 01/12/2010 08:26 | 7,65 | 17850 | United Kingdom |
| 7 | 536365 | 21730 | GLASS STAR FROSTE | 6 | 01/12/2010 08:26 | 4,25 | 17850 | United Kingdom |
| 8 | 536366 | 22633 | HAND WARMER UNI | 6 | 01/12/2010 08:28 | 1,85 | 17850 | United Kingdom |
| 9 | 536366 | 22632 | HAND WARMER REC | 6 | 01/12/2010 08:28 | 1,85 | 17850 | United Kingdom |
| 10 | 536367 | 84879 | ASSORTED COLOUR | 32 | 01/12/2010 08:34 | 1,69 | 13047 | United Kingdom |
| 11 | 536367 | 22745 | POPPY'S PLAYHOUS | 6 | 01/12/2010 08:34 | 2,1 | 13047 | United Kingdom |
| 12 | 536367 | 22748 | POPPY'S PLAYHOUS | 6 | 01/12/2010 08:34 | 2,1 | 13047 | United Kingdom |
| 13 | 536367 | 22749 | FELTCRAFT PRINCES | 8 | 01/12/2010 08:34 | 3,75 | 13047 | United Kingdom |
| 14 | 536367 | 22310 | IVORY KNITTED MUI | 6 | 01/12/2010 08:34 | 1,65 | 13047 | United Kingdom |
| 15 | 536367 | 84969 | BOX OF 6 ASSORTEC | 6 | 01/12/2010 08:34 | 4,25 | 13047 | United Kingdom |
| 16 | 536367 | 22623 | BOX OF VINTAGE JIC | 3 | 01/12/2010 08:34 | 4,95 | 13047 | United Kingdom |
| 17 | 536367 | 22622 | BOX OF VINTAGE AL | 2 | 01/12/2010 08:34 | 9,95 | 13047 | United Kingdom |
| 18 | 536367 | 21754 | HOME BUILDING BL | 3 | 01/12/2010 08:34 | 5,95 | 13047 | United Kingdom |
| 19 | 536367 | 21755 | LOVE BUILDING BLO | 3 | 01/12/2010 08:34 | 5,95 | 13047 | United Kingdom |
| 20 | 536367 | 21777 | RECIPE BOX WITH N | 4 | 01/12/2010 08:34 | 7,95 | 13047 | United Kingdom |
| 21 | 536367 | 48187 | DOORMAT NEW ENI | 4 | 01/12/2010 08:34 | 7,95 | 13047 | United Kingdom |
| 22 | 536368 | 22960 | JAM MAKING SET W | 6 | 01/12/2010 08:34 | 4,25 | 13047 | United Kingdom |
| 23 | 536368 | 22913 | RED COAT RACK PAF | 3 | 01/12/2010 08:34 | 4,95 | 13047 | United Kingdom |
| 24 | 536368 | 22912 | YELLOW COAT RACK | 3 | 01/12/2010 08:34 | 4,95 | 13047 | United Kingdom |
| 25 | 536368 | 22914 | BLUE COAT RACK PA | 3 | 01/12/2010 08:34 | 4,95 | 13047 | United Kingdom |
| 26 | 536369 | 21756 | BATH BUILDING BLC | 3 | 01/12/2010 08:35 | 5,95 | 13047 | United Kingdom |
| 27 | 536370 | 22728 | ALARM CLOCK BAKE | 24 | 01/12/2010 08:45 | 3,75 | 12583 | France |
| 28 | 536370 | 22727 | ALARM CLOCK BAKE | 24 | 01/12/2010 08:45 | 3,75 | 12583 | France |
| 29 | 536370 | 22726 | ALARM CLOCK BAKE | 12 | 01/12/2010 08:45 | 3,75 | 12583 | France |
| 30 | 536370 | 21724 | PANDA AND BUNNIE | 12 | 01/12/2010 08:45 | 0,85 | 12583 | France |
| 31 | 536370 | 21883 | STARS GIFT TAPE | 24 | 01/12/2010 08:45 | 0,65 | 12583 | France |
| 32 | 536370 | 10002 | INFLATABLE POLITIC | 48 | 01/12/2010 08:45 | 0,85 | 12583 | France |
| 33 | 536370 | 21791 | VINTAGE HEADS ANI | 24 | 01/12/2010 08:45 | 1,25 | 12583 | France |
| 34 | 536370 | | | | | | | |

2. Lampiran B – Proses *Preprocessing*

A. Lampiran B1 – Data Cleaning

Langkah – langkah pembersihan:

Penanganan nilai kosong :

- Description: 1.454 baris
- CustomerID: 135.080 baris

Duplikasi data : 5.268 baris dengan nilai kosong pada CustomerID dihapus.

Outlier : -

B. Lampiran B2 – Tranformasi Data

Jenis transformasi:

Normalisasi/Standarisasi: -

Encoding: Deskripsi produk diubah menjadi transaksi per invoice (list of items).

~~Binning~~ Discretization: Transaksi diubah ke format one-hot encoding untuk input ke algoritma Apriori dan FP-Growth.

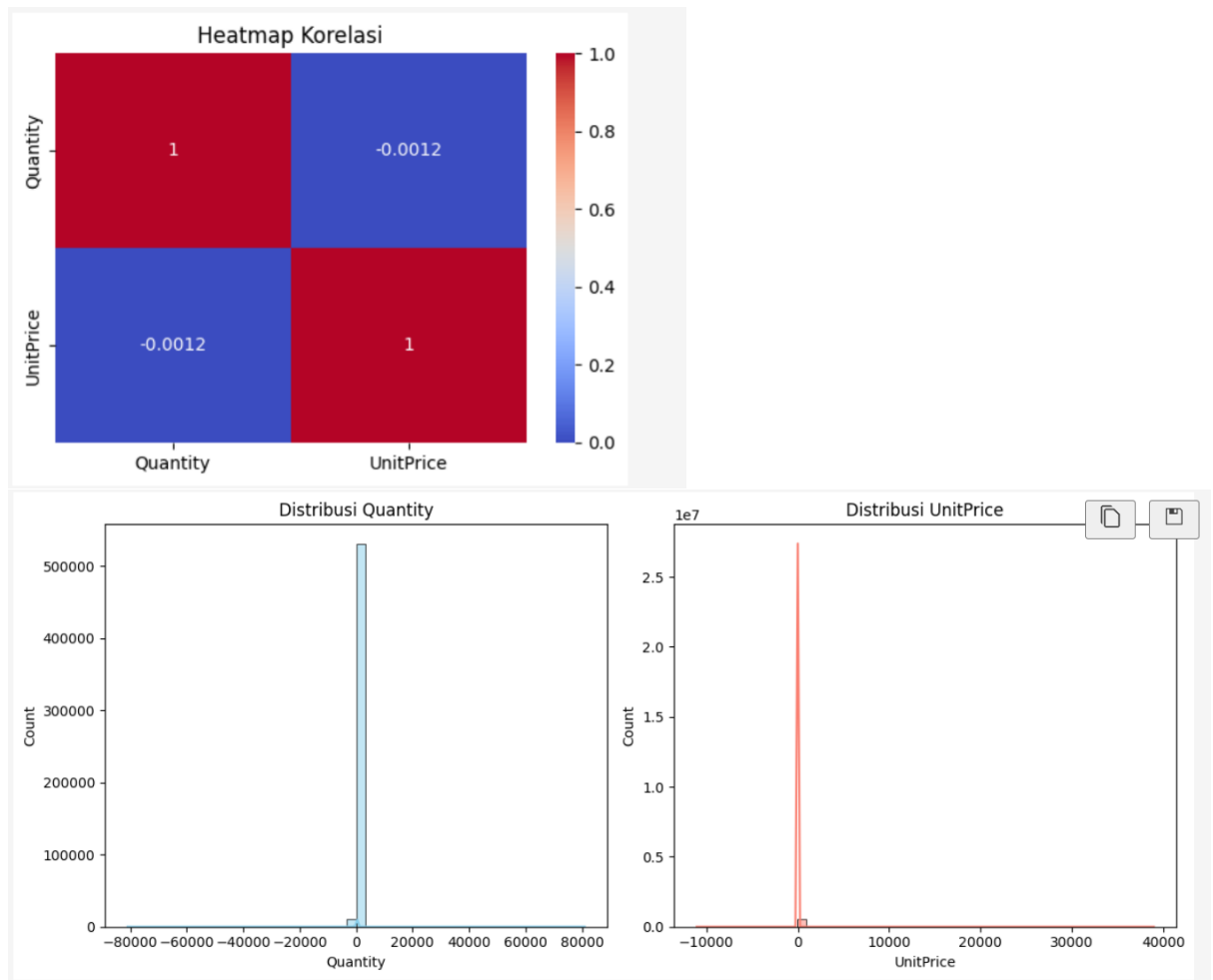
3. Lampiran C – Eksplorasi Data & Visualisasi (EDA)

A. Lampiran C1 – Statistik Deskriptif

```
Total data transaksi: 541909
Pelanggan unik: 4372
Produk unik: 4223
      Quantity      UnitPrice
count  541909.000000  541909.000000
mean      9.552250      4.611114
std     218.081158     96.759853
min    -80995.000000  -11062.060000
25%       1.000000      1.250000
50%       3.000000      2.080000
75%      10.000000      4.130000
max     80995.000000  38970.000000
Quantity    3.00
UnitPrice    2.08
dtype: float64
```

B. Lampiran C2 – Grafik dan Visualisasi

(Tambahkan visualisasi EDA: histogram, boxplot, scatterplot, heatmap)



4. Lampiran D – Pemodelan dan Evaluasi

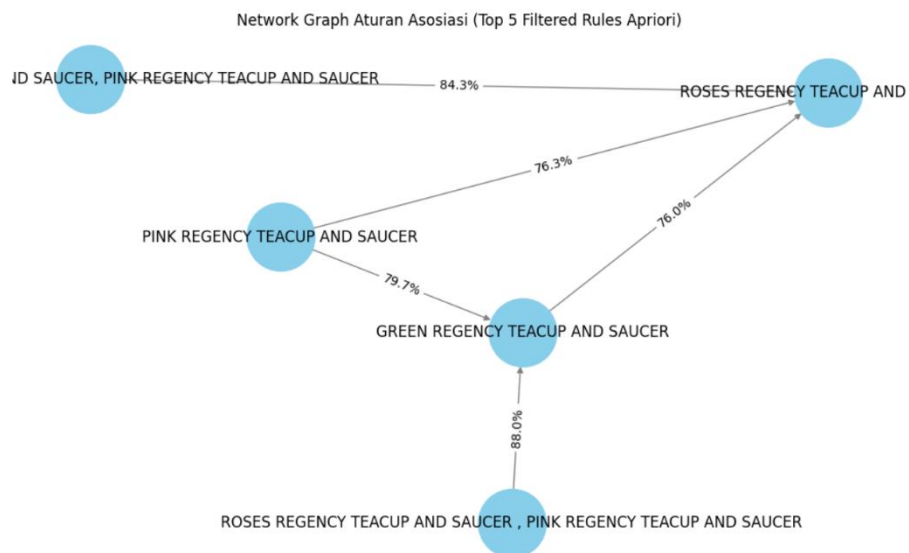
A. Lampiran D1 – Rincian Model

Model yang digunakan: Apriori & FP-Growth

Parameter model:

- $\text{Min_support} = 0.02$
- $\text{Min_confidence} = 0.5$
- $\text{Lift} > 1.2$

B. Lampiran D2 – Hasil Evaluasi Model



5. Lampiran E – Kode Program

A. Lampiran E1 – Script Python/R/Notebook

Scrip data pemodelan

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import networkx as nx
4 from mlxtend.frequent_patterns import apriori, fpgrowth, association_rules
5 from mlxtend.preprocessing import TransactionEncoder
6
7 # Baca data hasil preprocessing
8 df = pd.read_csv('../data/processed/dataset_final.csv')
9
10
11 # Hitung frekuensi item (produk)
12 top_items = df['Description'].value_counts().head(100).index.tolist()
13
14 # Filter data hanya yang termasuk dalam 100 produk terbanyak
15 df_filtered = df[df['Description'].isin(top_items)]
16 # Group per transaksi
17 transactions = df_filtered.groupby('InvoiceNo')['Description'].apply(list).tolist()
18
19 # One-hot encoding
20 te = TransactionEncoder()
21 te_ary = te.fit(transactions).transform(transactions)
22 df_encoded = pd.DataFrame(te_ary, columns=te.columns_)
23
24
25 # Filter data hanya yang termasuk dalam 100 produk terbanyak
26 df_filtered = df[df['Description'].isin(top_items)]
27
28 # Group per transaksi
29 transactions = df_filtered.groupby('InvoiceNo')['Description'].apply(list).tolist()
30
31 # One-hot encoding
32 te = TransactionEncoder()
33 te_ary = te.fit(transactions).transform(transactions)
34 df_encoded = pd.DataFrame(te_ary, columns=te.columns_)
35
36 # ===== APRIORI =====
37 frequent_apriori = apriori(df_encoded, min_support=0.02, use_colnames=True)
38 rules_apriori = association_rules(frequent_apriori, metric="confidence", min_threshold=0.3)
39
40 # ==== FILTERING RULE (contoh: lift > 1.2 dan confidence > 0.5) ====
41 filtered_rules_apriori = rules_apriori[
42     (rules_apriori['lift'] > 1.2) & (rules_apriori['confidence'] > 0.5)
43 ]
44
45 print("Filtered Rules APRIORI:")
46 print(filtered_rules_apriori[['antecedents', 'consequents', 'support', 'confidence', 'lift']].head())
47
48 frequent_fp = fpgrowth(df_encoded, min_support=0.02, use_colnames=True)
49 rules_fp = association_rules(frequent_fp, metric="confidence", min_threshold=0.3)
50
51
52
53
54

```

Link script pemodelan atau github <https://github.com/sindymaulina04/tugas-besar-datamining-kelompok5>

B. Lampiran E2 – Struktur Folder Proyek

- /data
- /src
- Output
- models