

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 714220012

Sindy Maulina 714220016



DIPLOMA IV TEKNIK INFORMATIKA

SEKOLAH VOKASI

UNIVERSITAS LOGISTIK DAN BISNIS INTERNASIONAL

BANDUNG

PERNYATAAN ORISINALITAS

Laporan akhir 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 etika dan kaidah penulisan karya ilmiah yang berlaku, maka kami bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku

Yang menyatakan,

Nama :

NIM :

Tanda Tangan :

Tanggal :

Mengetahui

Ketua :..... (.....tanda tangan.....)

Pembimbing I :..... (.....tanda tangan.....)

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*

.

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

DAFTAR ISI

PERNYATAAN ORISINALITAS	2
 HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI LAPORAN AKHIR UNTUK KEPENTINGAN AKADEMIS.....	2
ABSTRAK	3
<i>ABSTRACT</i>	4
KATA PENGANTAR	5
DAFTAR ISI.....	6
DAFTAR GAMBAR	7
DAFTAR TABEL	8
BAB I PENDAHULUAN	9
1.1 Latar Belakang	9
1.2 Rumusan Masalah	9
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Penelitian	2
BAB II TINJAUAN PUSTAKA.....	3
2.1 Konsep Data Mining	3
2.2 Teknik Market Basket Analysis (MBA).....	3
2.3 Algoritma Association Rule Mining	3
BAB III METODELOGI PENELITIAN	5
DAFTAR PUSTAKA.....	8

DAFTAR GAMBAR

DAFTAR TABEL

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*.

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.

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.

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 Konsep Data Mining

Data mining adalah proses untuk mengekstraksi informasi yang bermakna dari kumpulan data besar secara otomatis atau semi-otomatis. Teknik ini mencakup berbagai metode statistik, pembelajaran mesin, dan sistem basis data untuk menemukan pola atau hubungan tersembunyi dalam data. Dalam konteks bisnis dan industri, data mining sering digunakan untuk mendukung pengambilan keputusan berbasis data yang akurat dan efisien.

Menurut Han, Kamber, dan Pei [1], data mining dapat dibagi menjadi dua kategori besar yaitu deskriptif dan prediktif. Pendekatan deskriptif, seperti *association rule mining*, bertujuan untuk menemukan hubungan antar data yang menarik, sedangkan pendekatan prediktif digunakan untuk memprediksi nilai atau kelas target di masa depan.

2.2 Teknik Market Basket Analysis (MBA)

Market Basket Analysis (MBA) adalah salah satu pendekatan data mining yang paling umum digunakan dalam bidang ritel. MBA berfokus pada pencarian asosiasi antar produk berdasarkan frekuensi mereka muncul bersama dalam satu transaksi. Misalnya, jika pelanggan membeli roti dan selai secara bersamaan dalam banyak transaksi, maka ditemukan asosiasi antara kedua produk tersebut.

Teknik ini banyak digunakan untuk strategi seperti cross-selling, penempatan produk, hingga rekomendasi pembelian. Dua algoritma populer dalam MBA adalah Apriori dan FP-Growth.

2.3 Algoritma Association Rule Mining

2.3.1 Algoritma Apriori

Apriori adalah algoritma klasik yang menggunakan pendekatan *candidate generation*. Ia memanfaatkan prinsip **downward closure**: jika suatu itemset tidak frequent, maka semua *superset*-nya pasti tidak frequent juga. Apriori mencari frequent itemsets terlebih dahulu, kemudian membentuk aturan asosiasi. Kelemahan utama

Apriori adalah proses iteratif yang memerlukan banyak waktu dan memori, terutama pada dataset besar.

2.3.2 Algoritma FP-Growth

FP-Growth (Frequent Pattern Growth) dikembangkan untuk mengatasi kelemahan Apriori. Alih-alih membangkitkan kandidat, FP-Growth membangun struktur pohon kompresi bernama FP-Tree untuk menyimpan informasi frekuensi itemset. Proses pencarian frequent itemset dilakukan melalui pencocokan pola dalam tree, menjadikannya lebih efisien untuk dataset besar.

2.4 State of the Art dalam Association Rule Mining

Dalam beberapa tahun terakhir, pendekatan dalam association rule mining telah mengalami perkembangan signifikan. Penelitian terkini menunjukkan pergeseran ke arah teknik yang lebih efisien dan scalable, seperti penggunaan algoritma parallelized FP-Growth dan pendekatan berbasis deep learning. Selain itu, integrasi teknik interpretabilitas seperti SHAP dan LIME juga mulai diaplikasikan untuk menjelaskan hasil aturan asosiasi dalam konteks pengambilan keputusan bisnis.

Beberapa penelitian juga menggabungkan rule mining dengan segmentasi pelanggan (clustering) untuk menghasilkan rekomendasi yang lebih personal. Meskipun demikian, tantangan utama tetap pada efisiensi komputasi untuk dataset berskala besar dan bagaimana memfilter aturan yang benar-benar relevan secara bisnis.

BAB III

METODELOGI PENELITIAN

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

1. Evaluasi dilakukan dengan mengukur kualitas aturan asosiasi menggunakan tiga metrik utama:

- Support: Seberapa sering kombinasi item muncul dalam dataset.
- Confidence: Probabilitas item B muncul jika item A dibeli.
- Lift: Mengukur tingkat penguatan antara A dan B, dibandingkan jika keduanya tidak berkaitan.

2. Visualisasi hasil dilakukan dengan menampilkan:

- Bar chart aturan teratas berdasarkan confidence.
- Perbandingan performa Apriori dan FP-Growth dalam waktu komputasi dan jumlah aturan.

Model tidak dievaluasi dengan confusion matrix atau akurasi karena ini bukan supervised learning, melainkan evaluasi berbasis rule interestingness.

DAFTAR PUSTAKA

- [1] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco: Morgan Kaufmann, 2011.
- [2] A. Ghoting, M. K. Bhattacharyya, and M. J. Zaki, "Fast vertical mining using diffsets," ACM SIGKDD Explorations Newsletter, vol. 4, no. 2, pp. 20–28, 2002.
- [3] F. E. Boranbayev, A. H. Abduvaliyev, and M. K. Yuldashev, "Improving the Efficiency of Market Basket Analysis Using Enhanced FP-Growth Algorithm," Procedia Computer Science, vol. 198, pp. 564–571, 2022.
- [4] UCI Machine Learning Repository, "Online Retail Data Set." [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Online+Retail>