

**LAPORAN
PENEMUAN POLA KETERKAITAN PRODUK DALAM DATA
PENJUALAN RETAIL ONLINE MENGGUNAKAN ALGORITMA
APRIORI**



RISET INFORMATIKA D081

Dosen pengampu :

Dr. Basuki Rahmat, S.Si., M.T.

Dibuat Oleh:

Bagas Maulana Akbar (21081010307)

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UPN "VETERAN" JAWA TIMUR

2024

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital telah mengubah berbagai aspek kehidupan, termasuk cara orang berbelanja. Saat ini, aktivitas perdagangan elektronik atau e-commerce semakin menjadi pilihan utama konsumen karena kemudahan dan kecepatan yang ditawarkannya. Setiap transaksi yang dilakukan melalui platform e-commerce menghasilkan data dalam jumlah besar. Data ini mencakup informasi seperti jenis produk yang dibeli, jumlah transaksi, pola pembelian, hingga preferensi konsumen. Jika data ini dianalisis dengan benar, maka dapat memberikan wawasan yang berharga bagi perusahaan untuk memahami kebutuhan konsumen dan meningkatkan strategi pemasaran.

Namun, mengingat besarnya volume data transaksi yang terus bertambah, analisis manual tidak lagi menjadi metode yang efisien dan akurat. Oleh karena itu, diperlukan pendekatan yang berbasis teknologi untuk menggali informasi berharga dari data ini. Salah satu metode yang efektif adalah algoritma Apriori, sebuah algoritma dalam data mining yang dirancang untuk menemukan pola keterkaitan antarproduk yang sering dibeli bersama dalam data transaksi. Algoritma ini bekerja dengan mengidentifikasi item atau kombinasi item yang sering muncul dalam transaksi, sehingga menghasilkan wawasan mendalam mengenai pola belanja konsumen. Hasil analisis tersebut dapat dimanfaatkan untuk mendukung strategi pemasaran berbasis data, seperti rekomendasi produk dan promosi silang (cross-selling).

Penelitian ini bertujuan untuk menerapkan algoritma Apriori pada data transaksi retail online guna menemukan pola keterkaitan produk. Dataset yang digunakan dalam penelitian ini diambil dari UCI Machine Learning Repository, yang mencakup ratusan ribu transaksi. Analisis difokuskan pada data pelanggan di Jerman untuk memberikan hasil yang lebih spesifik dan relevan. Dengan parameter minimum

support sebesar 5% dan minimum confidence sebesar 50%, penelitian ini diharapkan dapat mengidentifikasi pola pembelian signifikan yang mampu mendukung pengembangan strategi pemasaran yang lebih efektif.

Berdasarkan hal tersebut, penelitian ini mengangkat beberapa rumusan masalah, yaitu: bagaimana algoritma Apriori dapat diterapkan untuk menemukan pola keterkaitan produk dalam data transaksi retail online, apa saja pola pembelian signifikan yang ditemukan, dan bagaimana hasil analisis ini dapat dimanfaatkan untuk mendukung strategi pemasaran berbasis data. Penelitian ini juga memiliki tujuan utama, yakni menggunakan algoritma Apriori untuk mengekstraksi pola keterkaitan produk, mengidentifikasi pola pembelian produk yang sering muncul bersama, serta memberikan rekomendasi strategi pemasaran berbasis data.

Penelitian ini diharapkan dapat memberikan manfaat baik secara akademis maupun praktis. Secara akademis, penelitian ini dapat menambah wawasan terkait penerapan algoritma Apriori dalam analisis data transaksi retail online. Adapun secara praktis, hasil penelitian ini dapat dimanfaatkan oleh perusahaan retail online untuk mengembangkan strategi pemasaran yang lebih terarah, seperti penentuan rekomendasi produk, promosi silang, dan pengelolaan inventaris yang lebih efisien. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi nyata dalam mendukung keberlanjutan bisnis retail online di era digital.

1.2 Rumusan Masalah

Untuk memberikan arahan yang jelas dalam penelitian ini, berikut adalah beberapa rumusan masalah yang diangkat:

1. Bagaimana algoritma Apriori dapat diterapkan untuk menemukan pola keterkaitan produk dalam data transaksi retail online?
2. Apa saja pola pembelian signifikan yang ditemukan berdasarkan analisis algoritma Apriori?

3. Bagaimana hasil analisis dapat dimanfaatkan untuk mendukung strategi pemasaran berbasis data?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk menjawab rumusan masalah yang telah disebutkan, tujuan utama penelitian meliputi:

1. Menggunakan algoritma Apriori untuk mengekstraksi pola keterkaitan produk dari data transaksi retail online.
2. Mengidentifikasi pola pembelian produk yang sering muncul bersama.
3. Memberikan rekomendasi strategi pemasaran berbasis data berdasarkan hasil analisis.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat dalam dua aspek utama, yaitu:

1. Manfaat Akademis
 - Menambah literatur terkait penerapan algoritma Apriori dalam data mining.
 - Memberikan contoh kasus nyata bagaimana algoritma data mining digunakan untuk menyelesaikan masalah bisnis.
2. Manfaat Praktis
 - Menyediakan rekomendasi pola keterkaitan produk yang dapat digunakan perusahaan retail online untuk meningkatkan strategi pemasaran.
 - Mendukung pengembangan promosi yang lebih efektif, seperti rekomendasi produk berdasarkan pola pembelian konsumen.

1.5 Batasan Masalah

Agar penelitian ini lebih berfokus dan tidak meluas dari inti pembahasan, maka penelitian ini membataskan beberapa macam batasan masalah, antara lain:

1. Cakupan Data

Penelitian ini hanya menggunakan data transaksi retail online yang berasal dari dataset UCI Machine Learning Repository. Analisis difokuskan pada

pelanggan dari Jerman untuk menyederhanakan ruang lingkup dan menghasilkan wawasan yang lebih relevan untuk segmen pasar tertentu.

2. Parameter Algoritma:

- Minimum support ditetapkan sebesar 5% untuk menyaring itemset yang jarang muncul.
- Minimum confidence ditetapkan sebesar 50% untuk membentuk aturan asosiasi yang memiliki tingkat kepercayaan tinggi.

3. Teknik Data Mining

Penelitian ini hanya menggunakan algoritma Apriori sebagai metode untuk menemukan pola keterkaitan antarproduk, tanpa membandingkannya dengan algoritma lain seperti FP-Growth atau Eclat.

4. Tujuan Analisis

Fokus analisis adalah mengidentifikasi frequent itemsets dan membentuk aturan asosiasi untuk mendukung strategi pemasaran, seperti promosi silang (cross-selling) dan rekomendasi produk, tanpa mengintegrasikan hasil analisis ke dalam sistem rekomendasi otomatis.

5. Alat Analisis

Penelitian ini memanfaatkan Python dan pustaka terkait (mlxtend, pandas, numpy, matplotlib) untuk proses data mining dan visualisasi, tanpa menggunakan perangkat lunak atau platform data mining lainnya.

BAB II

TINJAUAN PUSTAKA

2.1 Data Mining

Data mining merupakan bagian dari proses analisis data yang bertujuan untuk menggali informasi berharga dari dataset yang sangat besar. Proses ini melibatkan teknik-teknik canggih dari berbagai disiplin ilmu, seperti statistika, kecerdasan buatan, dan machine learning, untuk mengidentifikasi pola atau hubungan yang tersembunyi. Dalam konteks bisnis, data mining digunakan untuk mempermudah pengambilan keputusan strategis, seperti perencanaan promosi, rekomendasi produk, hingga pengelolaan inventaris.

Salah satu aplikasi utama data mining dalam retail adalah penerapan aturan asosiasi untuk memahami pola pembelian konsumen. Melalui analisis data transaksi, perusahaan dapat mengidentifikasi produk yang sering dibeli bersama, yang kemudian dapat dijadikan dasar untuk penawaran paket bundling atau promosi silang (cross-selling). Data mining juga mendukung pengembangan sistem rekomendasi berbasis data, yang secara signifikan dapat meningkatkan pengalaman belanja pelanggan dan loyalitas mereka terhadap merek tertentu.

Tahapan data mining yang terintegrasi dalam proses Knowledge Discovery in Databases (KDD) meliputi seleksi data, preprocessing, transformasi, analisis, dan evaluasi hasil. Dengan pendekatan ini, data yang awalnya tidak terstruktur dapat diolah menjadi informasi yang dapat dipahami dan digunakan untuk pengambilan keputusan.

2.2 Algoritma Apriori

Algoritma Apriori adalah salah satu metode dalam data mining yang digunakan untuk menemukan aturan asosiasi (association rules) antaritem dalam data transaksi. Algoritma ini dirancang untuk mengidentifikasi frequent itemsets, yaitu kombinasi produk yang sering muncul bersama dalam transaksi. Proses algoritma Apriori dimulai dengan mengidentifikasi itemset tunggal yang memenuhi ambang batas frekuensi tertentu, yang disebut minimum support. Selanjutnya, algoritma mengkombinasikan

itemset tersebut untuk menemukan pola dengan panjang yang lebih besar hingga tidak ada lagi pola yang memenuhi kriteria minimum support. Setelah itemsets ditemukan, langkah berikutnya adalah membentuk aturan asosiasi yang memenuhi ambang batas minimum confidence, yaitu tingkat kepercayaan pada keterkaitan antaritem.

Apriori juga menggunakan metrik evaluasi seperti lift ratio untuk mengukur kekuatan hubungan antarproduk. Lift ratio >1 menunjukkan bahwa hubungan antarproduk lebih kuat daripada yang diharapkan secara acak. Keunggulan algoritma Apriori terletak pada kemampuannya untuk menemukan pola yang relevan dari data transaksi yang besar, menjadikannya alat yang efektif untuk mendukung pengambilan keputusan di bidang pemasaran, pengelolaan inventaris, dan rekomendasi produk.

2.3 Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) adalah proses sistematis untuk menemukan pola yang valid, baru, dan berguna dari data. Proses ini mencakup beberapa tahapan, yaitu:

1. Seleksi Data: Memilih data yang relevan dari dataset utama untuk analisis lebih lanjut.
2. Preprocessing dan Pembersihan Data: Menghilangkan data yang tidak konsisten, noise, atau duplikasi untuk memastikan kualitas data.
3. Transformasi Data: Mengubah data ke dalam format yang sesuai untuk analisis, seperti matriks biner untuk algoritma Apriori.
4. Data Mining: Menerapkan algoritma atau teknik tertentu untuk mengekstraksi pola dari data.
5. Evaluasi dan Interpretasi: Memeriksa pola yang ditemukan untuk memastikan relevansi dan manfaatnya.

Dalam penelitian ini, KDD menjadi kerangka kerja utama untuk menganalisis data transaksi retail online menggunakan algoritma Apriori.

2.4 Penelitian Terdahulu

Penelitian terdahulu menunjukkan bahwa algoritma Apriori telah banyak digunakan untuk mengidentifikasi pola pembelian konsumen dalam berbagai konteks, seperti retail offline, e-commerce, dan bahkan di sektor kesehatan. Sandy et al. (2023) mengaplikasikan algoritma Apriori untuk menemukan pola pembelian dalam sebuah toko swalayan dan berhasil mengidentifikasi kombinasi produk yang relevan untuk promosi silang. Saputra dan Iskandar (2023) mengaplikasikan metode ini dalam retail online untuk menentukan rekomendasi bundling produk.

Penelitian lain oleh Wijaya et al. (2023) menunjukkan bahwa algoritma Apriori sangat efektif dalam mengungkap pola pembelian yang signifikan pada data transaksi retail besar. Penelitian ini juga menyoroti pentingnya parameter seperti support dan confidence dalam memastikan relevansi pola yang ditemukan. Penelitian terdahulu memberikan dasar yang kuat untuk implementasi algoritma Apriori dalam konteks data transaksi retail online, khususnya dalam mengidentifikasi pola keterkaitan produk yang dapat mendukung strategi pemasaran berbasis data.

2.5 Kontribusi Penelitian

Penelitian ini berbeda dari penelitian terdahulu dengan fokus pada analisis data transaksi retail online menggunakan dataset dari UCI Machine Learning Repository. Dengan parameter minimum support sebesar 5% dan minimum confidence sebesar 50%, penelitian ini bertujuan untuk menemukan pola keterkaitan produk yang relevan bagi pelanggan di Jerman. Hasil penelitian ini tidak hanya memberikan wawasan mendalam tentang preferensi konsumen tetapi juga memberikan kontribusi nyata dalam strategi pemasaran berbasis data, seperti rekomendasi produk dan promosi silang.

BAB III

DESAIN DAN IMPLEMENTASI SISTEM

3.1 Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksploratif. Pendekatan kuantitatif dipilih karena penelitian ini mengandalkan data transaksi retail online dalam jumlah besar untuk menganalisis pola keterkaitan antarproduk. Metode eksploratif digunakan karena tujuan penelitian adalah menemukan pola-pola baru yang belum teridentifikasi sebelumnya, tanpa diawali dengan hipotesis tertentu.

Algoritma Apriori diterapkan untuk mengekstraksi pola keterkaitan produk yang dapat mendukung pengambilan keputusan strategis dalam bisnis retail online. Proses penelitian mencakup berbagai tahapan, seperti seleksi data, preprocessing, transformasi, analisis data mining, hingga evaluasi hasil. Dengan desain ini, penelitian diharapkan dapat memberikan hasil yang mendalam, relevan, dan dapat diterapkan dalam skenario bisnis nyata.

Selain itu, penelitian ini mempertimbangkan parameter-parameter penting seperti support, confidence, dan lift ratio untuk mengevaluasi pola yang ditemukan. Pola yang memenuhi kriteria evaluasi akan dijadikan dasar untuk memberikan rekomendasi strategi pemasaran yang lebih efektif.

3.2 Sumber Data

Dataset yang digunakan dalam penelitian ini bersumber dari UCI Machine Learning Repository dengan url dataset <https://archive.ics.uci.edu/dataset/352/online+retail>, salah satu basis data yang sering digunakan dalam penelitian data mining. Dataset ini berisi 541.910 data transaksi retail online yang mencakup periode waktu tertentu. Setiap transaksi memiliki atribut-atribut berikut:

1. InvoiceNo: ID unik yang merepresentasikan setiap transaksi.
2. StockCode: Kode produk untuk mengidentifikasi jenis barang yang dibeli.
3. Description: Deskripsi produk.

4. Quantity: Jumlah barang yang dibeli dalam setiap transaksi.
5. InvoiceDate: Tanggal dan waktu transaksi dilakukan.
6. UnitPrice: Harga satuan produk.
7. CustomerID: ID pelanggan (jika tersedia).
8. Country: Negara asal pelanggan.

Dataset ini dipilih karena mencakup data yang relevan dengan kebutuhan penelitian, yaitu menganalisis pola pembelian dalam transaksi retail online. Untuk menyederhanakan ruang lingkup analisis, penelitian ini difokuskan pada pelanggan dari Jerman, menghasilkan subset data sebanyak 9.495 transaksi.

3.3 Tahapan Penelitian

Penelitian ini mengikuti kerangka kerja **Knowledge Discovery in Databases (KDD)** yang terdiri dari lima tahap utama: seleksi data, preprocessing, transformasi, data mining, dan evaluasi hasil. Berikut penjelasan setiap tahapan secara terperinci:

3.3.1 Seleksi Data

Seleksi data merupakan tahap awal yang penting untuk memastikan data yang digunakan relevan dengan tujuan penelitian. Dari total 541.910 transaksi, data difokuskan pada transaksi pelanggan dari Jerman. Langkah ini dilakukan untuk:

- Mengurangi kompleksitas analisis dengan membatasi ruang lingkup.
- Memberikan hasil analisis yang lebih spesifik dan relevan untuk segmen pasar tertentu.

Hasil seleksi menghasilkan subset sebanyak 9.495 transaksi yang digunakan untuk tahapan berikutnya.

3.3.2 Preprocessing Data

Preprocessing bertujuan untuk membersihkan dan menyelaraskan data agar siap digunakan dalam analisis. Beberapa langkah yang dilakukan pada tahap ini adalah:

1. Menghapus data yang tidak relevan

Data transaksi dengan kode produk seperti “POST” dan “M” yang tidak merepresentasikan barang fisik dihapus.

2. Menangani missing values

Transaksi tanpa ID pelanggan (CustomerID) dihapus untuk menjaga konsistensi dataset.

3. Normalisasi teks

Atribut Description diubah menjadi huruf kecil untuk memastikan keseragaman data.

4. Validasi data

Memastikan setiap atribut memiliki format dan nilai yang sesuai dengan deskripsi datanya.

3.3.3 Transformasi Data

Setelah preprocessing, data diubah menjadi format matriks biner. Format ini merepresentasikan setiap transaksi sebagai baris dan setiap produk sebagai kolom, dengan nilai 1 menunjukkan keberadaan produk dalam transaksi, dan 0 menunjukkan ketiadaan. Proses ini memudahkan algoritma Apriori untuk mengidentifikasi frequent itemsets dan membentuk aturan asosiasi.

3.3.4 Data Mining

Tahap ini merupakan inti dari penelitian, di mana algoritma Apriori diterapkan untuk menemukan pola keterkaitan produk. Langkah-langkah yang dilakukan meliputi:

1. Identifikasi Frequent Itemsets

- Frequent itemsets adalah kombinasi produk yang sering muncul bersama dalam transaksi.
- Parameter minimum support sebesar 5% digunakan untuk menyaring itemsets yang jarang muncul.

2. Pembuatan Association Rules

- Aturan asosiasi dibentuk dari frequent itemsets yang telah memenuhi kriteria minimum confidence sebesar 50%.

- Association rules menunjukkan hubungan antara produk, misalnya: “Jika pelanggan membeli produk A, maka kemungkinan besar mereka juga membeli produk B.”

3.3.5 Evaluasi Hasil

Pola keterkaitan yang ditemukan dievaluasi menggunakan lift ratio. Lift ratio >1 menunjukkan bahwa hubungan antarproduk lebih signifikan daripada yang diharapkan secara acak. Hasil evaluasi ini divisualisasikan dalam bentuk tabel dan grafik untuk mempermudah interpretasi dan penyampaian informasi.

3.4 Alat dan Bahan

Penelitian ini memanfaatkan alat dan bahan sebagai berikut:

1. Perangkat Lunak
 - Python dengan pustaka pendukung seperti mlxtend, pandas, numpy, dan matplotlib untuk analisis data dan visualisasi.
 - Jupyter Notebook sebagai lingkungan kerja interaktif.
2. Dataset
 - Data transaksi retail online yang diambil dari UCI Machine Learning Repository.
3. Perangkat Keras
 - Komputer dengan spesifikasi minimal prosesor Intel Core i5 dan RAM 8 GB untuk mendukung proses analisis data yang besar.

3.5 Parameter Evaluasi

Algoritma Apriori dalam penelitian ini dievaluasi menggunakan tiga metrik utama:

1. Support
 - Mengukur frekuensi kemunculan kombinasi produk dalam transaksi.

- Rumus:

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

2. Confidence

- Mengukur seberapa besar kemungkinan pelanggan membeli produk tertentu jika mereka telah membeli produk lainnya.

- Rumus:

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

3. Lift Ratio

- Menilai kekuatan hubungan antaritem dibandingkan kemunculannya secara independen.

- Rumus:

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

3.6 Analisis Hasil

Hasil analisis berupa frequent itemsets dan association rules akan disajikan dalam bentuk tabel dan grafik. Pola-pola tersebut kemudian dievaluasi untuk memberikan rekomendasi strategi pemasaran, seperti:

- Promosi silang (cross-selling).
- Penawaran bundling produk.
- Rekomendasi produk berbasis data.