



Penerapan Algoritma *Association Rule Mining* (Apriori) untuk Analisis Pola Pembelian (*Market Basket Analysis*) pada Data Transaksi Ritel

Jadiaman Parhusip^{1*}, Erina Ekanova Safitri², Tyara Rahmidasari³, Tety Citra Natha⁴,
Nur Haniatin Jannah⁵

¹⁻⁵ Teknik Informatika, Teknik, Universitas Palangka Raya

Email : parhusip.jadiaman@it.upr.ac.id^{1*}, erinaekanosafitri@mhs.eng.upr.ac.id²,
tyararahmidasari@mhs.eng.upr.ac.id³, tetycitranatha@mhs.eng.upr.ac.id⁴,
nurhaniatinjannah@mhs.eng.upr.ac.id⁵

*Penulis Korespondensi: parhusip.jadiaman@it.upr.ac.id¹

Abstract. *Market Basket Analysis (MBA) is a data mining technique used to identify product combination patterns that frequently occur together within a single transaction. This study applies the Apriori Algorithm as the primary method for discovering Association Rules from a large-scale e-commerce retail transaction dataset consisting of approximately 500,000 records, focusing on consumer purchasing behavior in the United Kingdom market. The research follows a systematic data mining process that includes data integration, data cleaning to remove anomalies such as negative prices and quantities, and data transformation using One-Hot Encoding to convert transaction records into a suitable binary matrix format. The Apriori Algorithm is then used to generate frequent itemsets, which are evaluated using Support, Confidence, and Lift to determine their strength and significance. The results show several strong Association Rules with Lift values greater than 1.0, indicating positive correlations between specific product pairs. These findings offer useful insights into consumer purchasing tendencies and can support various retail strategies, such as improving product recommendation systems, optimizing store layout, enhancing promotional bundles, and strengthening targeted marketing efforts. Overall, this study demonstrates that Apriori-based MBA is capable of extracting actionable knowledge from large-scale retail datasets and contributes to more effective, data-driven decision-making in the retail sector.*

Keywords: *Algoritma Apriori; Association Rules; Data Mining; Market Basket Analysis; One-Hot Encoding.*

Abstrak. *Market Basket Analysis (MBA) merupakan teknik data mining yang digunakan untuk mengidentifikasi pola kombinasi produk yang sering muncul secara bersamaan dalam satu transaksi. Penelitian ini menerapkan Algoritma Apriori sebagai metode utama untuk menemukan Association Rules dari dataset transaksi ritel e-commerce berskala besar yang berjumlah sekitar 500 ribu baris dan berfokus pada perilaku pembelian konsumen di pasar United Kingdom. Metodologi penelitian mengikuti proses data mining yang sistematis, meliputi integrasi data, pembersihan data untuk menghapus entri anomali seperti harga dan kuantitas negatif, serta transformasi data menggunakan One-Hot Encoding agar data transaksi dapat diubah ke dalam format matriks biner yang sesuai. Algoritma Apriori kemudian digunakan untuk menghasilkan frequent itemsets yang dievaluasi menggunakan metrik Support, Confidence, dan Lift untuk menilai kekuatan dan signifikansi aturan yang terbentuk. Hasil penelitian menunjukkan adanya beberapa Association Rules yang kuat dengan nilai Lift lebih dari 1.0, menandakan adanya korelasi positif antar pasangan produk tertentu. Temuan ini memberikan wawasan penting mengenai kecenderungan pembelian konsumen dan dapat dimanfaatkan untuk meningkatkan strategi ritel seperti sistem rekomendasi produk, tata letak toko, promosi bundling, dan pemasaran terarah. Secara keseluruhan, penelitian ini membuktikan bahwa MBA berbasis Apriori mampu mengekstraksi pengetahuan yang dapat ditindaklanjuti dari dataset ritel skala besar serta mendukung pengambilan keputusan berbasis data yang lebih efektif.*

Kata kunci: Algoritma Apriori; Aturan Asosiasi; Market Basket Analysis; Penambangan Data; Pengkodean Satu-Panas

1. LATAR BELAKANG

Perkembangan teknik data mining kini memegang peran penting dalam dunia bisnis modern, terutama di industri ritel yang mengalami peningkatan besar pada volume transaksi digital (Wulandari dkk., 2025). Salah satu teknik yang sangat bermanfaat untuk menggali

informasi dari kumpulan data transaksi berskala besar adalah *Market Basket Analysis* (MBA) (Ahmad Thariq, 2023). MBA digunakan untuk menemukan pola keterkaitan antarproduk yang sering muncul secara bersamaan dalam satu transaksi (Brighton & Hariyanto, 2024). Dengan memanfaatkan pola tersebut, pelaku ritel dapat membuat keputusan bisnis yang lebih tepat dan berbasis data.

Agar MBA dapat diterapkan secara optimal, diperlukan algoritma yang mampu memproses dataset dalam jumlah besar sekaligus mengekstraksi Association Rules yang kuat. Penelitian ini menggunakan Algoritma Apriori sebagai metode utama karena algoritma ini terbukti efektif dalam menghasilkan frequent itemsets dan aturan asosiasi yang dapat dianalisis lebih lanjut menggunakan metrik *Support*, *Confidence*, dan *Lift* (Ashari dkk., 2022).

Meskipun penelitian terkait *Market Basket Analysis* telah banyak dilakukan, masih sedikit studi yang fokus pada analisis dataset *e-commerce* berskala sangat besar dengan konteks pasar tertentu seperti United Kingdom. Penelitian ini hadir untuk mengisi gap tersebut dengan menunjukkan kemampuan Algoritma Apriori dalam menemukan pola pembelian dari sekitar 500.000 transaksi melalui tahapan data mining yang terstruktur, mulai dari integrasi data, pembersihan anomali, hingga transformasi menggunakan *One-Hot Encoding*. Urgensi penelitian ini berkaitan dengan kebutuhan industri ritel untuk meningkatkan strategi penjualan, seperti sistem rekomendasi, promosi *bundling*, serta penataan produk. Tujuan utamanya adalah menerapkan Apriori untuk menghasilkan *association rules* dengan nilai lift di atas 1.0 sehingga dapat memberikan wawasan terkait perilaku pembelian konsumen dan mendukung pengambilan keputusan strategis dalam bisnis ritel.

2. KAJIAN TEORITIS

Data Mining

Data mining merupakan proses untuk mengekstraksi pola, hubungan, atau informasi penting dari kumpulan data berukuran besar (P. A. W. Purnama dkk., 2025). Proses ini mencakup teknik analisis yang mampu mengolah data mentah menjadi pengetahuan yang bermakna bagi pengambilan keputusan (Anggi Canita Simanjuntak dkk., 2024). Dalam banyak referensi, data mining dianggap sebagai bagian dari proses *Knowledge Discovery in Databases* (KDD), yang terdiri dari beberapa tahap seperti pembersihan data, penggabungan data, pemilihan data, transformasi, proses mining, evaluasi pola, hingga penyajian pengetahuan. Teknik data mining sendiri banyak digunakan di berbagai bidang, termasuk dunia ritel, untuk memahami perilaku konsumen melalui analisis pola pembelian yang mereka lakukan.

Market Basket Analysis

Market Basket Analysis (MBA) adalah pendekatan analitis yang bertujuan mengidentifikasi hubungan atau kecenderungan produk yang sering dibeli secara bersamaan dalam satu transaksi (Fadillah dkk., 2023). Analisis ini didasarkan pada asumsi bahwa perilaku konsumen dalam memilih produk dapat mencerminkan pola kebutuhan atau preferensi tertentu. Dalam konteks ritel, MBA memiliki peran penting karena dapat memberikan wawasan mengenai strategi penempatan produk, penyusunan paket promosi, optimalisasi rekomendasi penjualan, hingga perencanaan stok barang (Safitry dkk., 2024). MBA umumnya dilakukan dengan menggunakan teknik *association rule mining* yang mampu mengungkap hubungan antar item dalam kumpulan transaksi.

Association Rule Mining

Association rule mining merupakan salah satu metode dalam data mining yang berfungsi menemukan aturan-aturan asosiasi antara item dalam dataset transaksi (Andy Hermawan dkk., 2024). Aturan asosiasi biasanya direpresentasikan dalam bentuk implikasi " $X \rightarrow Y$ ", yang menunjukkan bahwa kemunculan item X pada suatu transaksi diikuti oleh kemunculan item Y dengan probabilitas tertentu (Utami & Dapa, 2024). Evaluasi aturan asosiasi dilakukan melalui beberapa parameter utama, yaitu *support*, *confidence*, dan *lift*. *Support* menunjukkan seberapa sering suatu itemset muncul dalam keseluruhan transaksi. *Confidence* menggambarkan tingkat kepastian bahwa Y akan muncul ketika X hadir dalam transaksi. Sementara itu, *lift* digunakan untuk mengukur kekuatan hubungan antara X dan Y dengan membandingkannya terhadap kemungkinan muncul secara acak.

Algoritma Apriori

Algoritma Apriori adalah algoritma klasik yang banyak digunakan dalam proses *association rule mining* (Achmad dkk., 2025). Prinsip utama Apriori adalah bahwa suatu itemset hanya dapat bersifat *frequent* apabila seluruh subset dari itemset tersebut juga bersifat *frequent* (Ramadana dkk., 2022.). Algoritma ini berjalan melalui dua proses utama, yaitu menemukan frequent itemset dan membentuk aturan asosiasi. Pada tahap awal, Apriori mencari item-item tunggal yang memiliki nilai *support* di atas batas minimum. Setelah itu, algoritma secara bertahap membuat kombinasi item yang lebih besar melalui proses *candidate generation*. Ketika *frequent itemset* sudah didapatkan, Apriori kemudian menyusun aturan asosiasi dengan menghitung nilai *confidence* dan *lift* untuk menilai seberapa kuat dan valid hubungan antar item yang ditemukan.

Penerapan Apriori dalam Industri Ritel

Di dunia ritel, algoritma Apriori banyak dimanfaatkan karena mampu menemukan pola pembelian konsumen yang biasanya tidak terlihat dengan analisis biasa. Dengan mengetahui kombinasi produk yang sering dibeli secara bersamaan, pihak ritel bisa menyusun strategi pemasaran yang lebih tepat sasaran, seperti mengatur ulang tata letak produk, membuat promo bundling, atau meningkatkan akurasi sistem rekomendasi. (Fuadi dkk., 2024). Selain itu, hasil analisis Apriori dapat digunakan untuk memperbaiki manajemen persediaan dengan memahami pola permintaan berdasarkan perilaku pembelian aktual (D. A. Purnama, 2025). Karena setiap toko atau jenis ritel memiliki karakteristik data transaksi yang berbeda, penerapan algoritma Apriori pada dataset tertentu penting dilakukan untuk memperoleh wawasan yang sesuai dengan konteks operasional ritel tersebut.

Penelitian Terdahulu yang Relevan

Penelitian terdahulu menunjukkan pentingnya analisis pola pembelian dalam mendukung keputusan bisnis ritel. Banyak toko belum mengoptimalkan data transaksi padahal pola pembelian dapat dimanfaatkan untuk strategi *bundling* dan rekomendasi produk. Bahkan, beberapa studi mencatat bahwa pola asosiasi berkorelasi dengan naiknya pendapatan toko pada periode tertentu.

Secara rinci, beberapa penelitian terkait *Market Basket Analysis* (MBA) dengan Algoritma Apriori meliputi:

- a) Sejumlah penelitian menyoroti analisis pola pembelian dalam konteks ritel dan supermarket. Salah satunya adalah penelitian Ananda dkk. (2025) berjudul “Analisis Pola Pembelian Pelanggan Menggunakan Algoritma Apriori Pada Transaksi Belanja Supermarket”, yang mengkaji hubungan antarproduk pada transaksi supermarket dengan memanfaatkan Algoritma Apriori. Studi tersebut menerapkan tahapan pra-pemrosesan seperti pembersihan data dan transformasi *One-Hot Encoding* pada dataset publik. Hasil penelitian menunjukkan bahwa pendekatan yang terstruktur ini mampu menghasilkan pola pembelian yang bermanfaat untuk strategi penempatan produk dan promosi. (Ananda dkk., 2025).
- b) Penelitian Kuswanto dkk. (2024) berjudul “Penerapan Algoritma Apriori Dalam Analisis Keranjang Belanja Retail Di Wilayah Jawa Barat” juga menggunakan Algoritma Apriori untuk menganalisis data ritel di Jawa Barat. Hasil penelitian tersebut menunjukkan bahwa Market Basket Analysis mampu mengungkap pola pembelian konsumen yang sering muncul, sehingga dapat dimanfaatkan untuk menyusun strategi pemasaran yang lebih akurat dan sesuai dengan perilaku pelanggan. (Andi Diah Kuswanto dkk., 2024).

- c) Penelitian oleh Butsianto dkk (2025) yang berjudul “Implementasi Algoritma Apriori dalam Menemukan Pola Asosiasi pada Data Penjualan Produk Retail” menambahkan bahwa implementasi Apriori untuk menemukan pola asosiasi antar kategori produk (seperti Susu, Kopi, Detergen) dapat dievaluasi secara kuat menggunakan parameter support, confidence, dan lift, yang krusial untuk perencanaan stok dan display ritel modern (Butsianto dkk., 2025).
- d) efektivitas algoritma Apriori juga terlihat jelas pada sektor ritel yang lebih spesifik. Penelitian Kafabih dan Chandra (2024) berjudul “Implementasi Algoritma Apriori Dengan Metode Association Rules Dalam Market Basket Analysis Pada Data Transaksi Reddog” menerapkan Apriori pada data transaksi perusahaan Food and Beverages (Reddog). Penelitian ini bertujuan menemukan hubungan antar item yang dapat dimanfaatkan untuk memberikan rekomendasi menu serta menyusun strategi bundling. Hasilnya menunjukkan bahwa Apriori sangat sesuai digunakan pada bisnis dengan volume transaksi harian yang tinggi. (Kafabih & Chandra, 2024).
- e) Penelitian berjudul “Market Basket Analysis with Apriori Algorithm and Frequent Pattern Growth (Fp-Growth) on Outdoor Product Sales Data” oleh Nurmayanti dkk. (2021) menerapkan algoritma Apriori dan kemudian membandingkannya dengan FP-Growth pada data penjualan produk outdoor. Walaupun fokus utamanya adalah perbandingan kedua metode, hasil penelitian menunjukkan bahwa keduanya mampu menemukan keterkaitan antarproduk yang dapat dimanfaatkan oleh toko untuk strategi penjualan, seperti hubungan antara kompor portable dan gas portable. (Nurmayanti dkk., 2021).
- f) Penelitian berikutnya membahas perbandingan Apriori dengan algoritma lain dari segi efisiensi. Madani dkk. (2024) membandingkan Algoritma Apriori dengan FP-Growth menggunakan 600 data transaksi minimarket. Hasil penelitian menunjukkan bahwa meskipun FP-Growth dinilai lebih efektif dalam menemukan pola pembelian pada studi tersebut, kedua algoritma tetap mampu mengidentifikasi kombinasi item dengan frekuensi tertinggi. Temuan ini memberikan masukan penting bagi pihak minimarket untuk menata produk sesuai kebiasaan belanja konsumen. (Madani dkk., 2024).

3. METODE PENELITIAN

Penelitian ini memanfaatkan dataset transaksi ritel e-commerce dari periode 2009–2010, yang merupakan bagian dari Online Retail II Dataset. Data tersebut bersifat kuantitatif dan berbasis komputasi, serta menjadi sumber utama dalam penerapan Algoritma Apriori untuk *Market Basket Analysis* (MBA). Walaupun dataset ini mencakup transaksi dari berbagai

wilayah, penelitian ini secara khusus membatasi analisis hanya pada transaksi yang berasal dari United Kingdom. Pembatasan tersebut dilakukan agar pola pembelian yang ditemukan lebih relevan dan dapat memberikan wawasan strategis yang benar-benar fokus pada satu segmen pasar.

Rancangan dan Objek Penelitian

Penelitian ini menggunakan desain penelitian kuantitatif dengan pendekatan eksperimental berbasis komputasi (*computational experiment*). Pendekatan tersebut dipilih karena mampu mengolah data sekunder menjadi informasi yang lebih bermakna melalui penerapan algoritma data mining. Objek penelitian berupa data transaksi dari Online Retail II Dataset yang diambil melalui repositori publik di Kaggle. Populasi penelitian mencakup seluruh transaksi ritel dari tahun 2009 hingga 2011. Teknik sampling yang digunakan adalah Purposive Sampling, di mana sampel dipilih secara khusus hanya pada transaksi tahun 2009–2010 dan dibatasi pada wilayah United Kingdom. Pembatasan ini dilakukan agar pola pembelian yang dianalisis tetap relevan dan fokus pada satu segmen pasar yang lebih homogen.

Instrumen penelitian berupa dataset yang berisi atribut seperti Invoice, *StockCode*, *Description*, *Quantity*, *InvoiceDate*, *Price*, *Customer ID*, dan *Country*. Karena penelitian ini menggunakan data sekunder, uji validitas instrumen tidak dilakukan secara tradisional, tetapi digantikan dengan proses pembersihan data (*data cleaning*) yang ketat untuk memastikan data yang digunakan berkualitas baik sebelum dianalisis.

Tahapan Penelitian

Proses penelitian ini mengimplementasikan kerangka kerja *Knowledge Discovery in Databases* (KDD) yang dijalankan menggunakan alat analisis berupa perangkat lunak berbasis Python dengan bantuan pustaka Pandas dan Mlxtend. Tahapan penelitian terbagi menjadi empat fase utama:

Data Preprocessing

Tahap ini merupakan implementasi teknik pengumpulan dan penyiapan data. Data dikumpulkan melalui metode dokumentasi dengan mengunduh satu fail sumber data transaksi. Tahap ini berfokus pada validasi data, di mana dilakukan eksplorasi struktur dataset dan pembersihan data (*data cleaning*). Proses pembersihan meliputi penghapusan transaksi dengan kode pembatalan (diawali huruf 'C'), eliminasi baris biaya non-produk (seperti 'POST'), serta penanganan missing value. Tahap ini krusial sebagai pengganti uji reliabilitas untuk memastikan data bebas dari anomali yang dapat bias pada hasil analisis.

Data Transformation

Data yang telah bersih diubah ke format yang diperlukan oleh Algoritma Apriori:

- a) Pengelompokan Transaksi: Data dikelompokkan berdasarkan nomor faktur (*Invoice*) untuk mengidentifikasi item yang dibeli dalam satu keranjang.
- b) *One-Hot Encoding*: Data diubah menjadi matriks biner (basket sets). Setiap kolom mewakili satu item, dan nilai '1' atau '0' menunjukkan keberadaan atau ketiadaan item tersebut dalam sebuah transaksi.

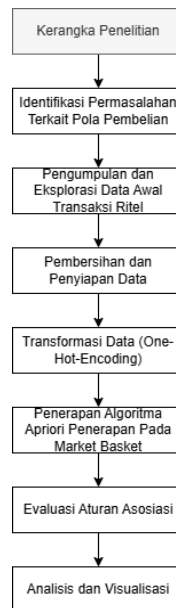
Pemodelan Menggunakan Algoritma Apriori

Algoritma Apriori diterapkan pada matriks biner tersebut untuk menghasilkan frequent itemsets (kombinasi item yang sering muncul) berdasarkan nilai minimum support yang ditentukan. Evaluasi Model Aturan Asosiasi dibentuk dari *frequent itemsets* menggunakan metrik *Confidence*. Kekuatan hubungan antarproduk dinilai menggunakan metrik *Lift*. Aturan yang dianggap signifikan dan menunjukkan korelasi positif diidentifikasi dengan memilih aturan yang memiliki nilai $Lift > 1.0$.

Kerangka Penelitian

Kerangka penelitian ini menggambarkan alur kerja yang terstruktur dalam melakukan Analisis Pola Pembelian menggunakan Algoritma Apriori. Proses dimulai dari tahap Identifikasi Permasalahan, yaitu menentukan pola pembelian apa yang ingin ditelusuri dari data. Setelah itu dilakukan Pengumpulan dan Seleksi Data Transaksi Ritel untuk memilih data yang relevan (periode 2009–2010 dan fokus pada wilayah UK). Data yang sudah terkumpul kemudian melalui tahap Pembersihan dan Penyiapan Data (Data Preprocessing) untuk menangani anomali serta missing value.

Tahap berikutnya adalah Transformasi Data ke Format Biner menggunakan One-Hot Encoding agar data siap diproses oleh algoritma. Bagian inti dari penelitian ini adalah Penerapan Algoritma Apriori untuk menghasilkan frequent itemsets dan aturan asosiasi. Aturan yang terbentuk kemudian dievaluasi menggunakan metrik *Support*, *Confidence*, dan *Lift* untuk menilai kekuatan dan signifikansinya. Tahap akhir penelitian adalah Penarikan Kesimpulan serta Penyusunan Rekomendasi Strategis yang dapat dimanfaatkan manajemen ritel sebagai bahan pertimbangan dalam pengambilan keputusan.



Gambar 1. Kerangka Penelitian

4. HASIL DAN PEMBAHASAN

Pengumpulan dan Eksplorasi Data Awal Transaksi Ritel

Penelitian ini dimulai dengan memuat data transaksi ritel dari satu file yang berisi data tahun 2009–2010. Setelah proses integrasi dilakukan, diperoleh total 525.461 baris data mentah. Pada tahap eksplorasi awal (*Exploratory Data Analysis/EDA*), ditemukan beberapa masalah kualitas data yang harus diperbaiki sebelum masuk ke tahap pemodelan. Masalah pertama adalah adanya *missing values* pada kolom *Description*, yang seharusnya berisi nama atau identitas produk. Masalah kedua, dan yang paling penting, adalah ditemukannya anomali pada nilai *Quantity* dan *Price*, di mana banyak entri memiliki nilai negatif. Hal ini menunjukkan adanya transaksi pembatalan, pengembalian barang, atau kesalahan pencatatan. Hasil EDA ini menegaskan bahwa proses pembersihan data menjadi langkah yang wajib dilakukan untuk memastikan analisis *Market Basket Analysis* (MBA) dapat menghasilkan pola yang valid dan akurat.

Dataset berhasil dimuat dari Drive. Jumlah baris awal: 525461

```

--- Data Awal (Head) ---
Invoice StockCode      Description  Quantity  InvoiceDate  Price  Customer ID      Country  Revenue
0  489434      85048  15CM CHRISTMAS GLASS BALL 20 LIGHTS      12  2009-12-01 07:45:00  6.95  13085.0  United Kingdom  83.4
1  489434      79323P      PINK CHERRY LIGHTS      12  2009-12-01 07:45:00  6.75  13085.0  United Kingdom  81.0
2  489434      79323W      WHITE CHERRY LIGHTS      12  2009-12-01 07:45:00  6.75  13085.0  United Kingdom  81.0
3  489434      22041      RECORD FRAME 7" SINGLE SIZE      48  2009-12-01 07:45:00  2.10  13085.0  United Kingdom  100.8
4  489434      21232      STRAWBERRY CERAMIC TRINKET BOX      24  2009-12-01 07:45:00  1.25  13085.0  United Kingdom  30.0
    
```

Gambar 2. Tampilan Awal Data Transaksi (*Head*)

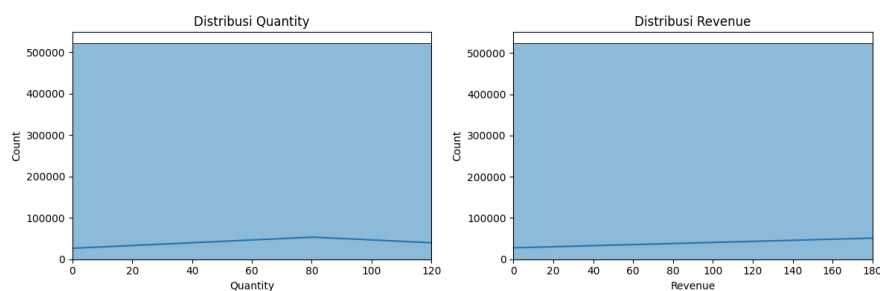
Menunjukkan variabel utama yang akan digunakan, termasuk *Invoice* (ID Transaksi), *Description* (Item), *Quantity*, *Price*, dan kolom turunan *Revenue* yang telah dihitung ($Q \times P$). Data ini mengonfirmasi bahwa proses *Data Loading* dari *Google Drive* berhasil. Selain itu, tampilan ini memvalidasi adanya kolom kunci yang diperlukan untuk *Market Basket Analysis* dan *Feature Engineering* awal (*Revenue*).

```
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Invoice      525461 non-null   object
1   StockCode    525461 non-null   object
2   Description  525533 non-null   object
3   Quantity     525461 non-null   int64
4   InvoiceDate  525461 non-null   datetime64[ns]
5   Price        525461 non-null   float64
6   Customer ID  417534 non-null   float64
7   Country      525461 non-null   object
8   Revenue      525461 non-null   float64
dtypes: datetime64[ns](1), float64(3), int64(1), object(4)
memory usage: 36.1+ MB
None

-- Statistik Deskriptif (Quantity, Price, Revenue) --
count      mean      std      min      25%      50%      75%      max
Quantity  525461.0  10.337667  107.424110 -9600.00  1.00   10.00  19152.00
Price     525461.0   4.688834  146.126914 -53594.36  1.25   2.10   25111.09
Revenue   525461.0  18.154506  160.333083 -53594.36  3.75   9.95  25111.09
```

Gambar 3. Statistik Deskriptif Variabel Kuantitatif Dataset Awal

Gambar 3 di atas ini secara definitif menunjukkan adanya anomali kuantitatif yang harus dibersihkan, dibuktikan dengan nilai minimum negatif pada ketiga kolom: *Quantity* (min -9600.00), *Price* (min -53594.36), dan *Revenue* (min -53594.36). Anomali ini mengonfirmasi keberadaan transaksi pembatalan dan *error* pencatatan yang akan dihilangkan pada tahap *Data Cleaning*.



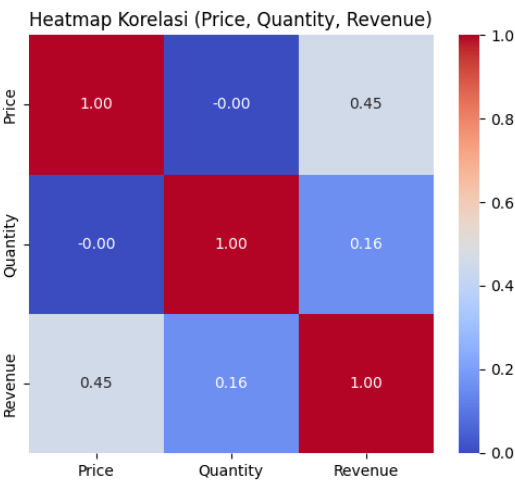
Gambar 4. Histogram Distribusi Quantity dan Revenue

Kedua histogram menunjukkan distribusi data yang sangat miring ke kanan (*right-skewed*). Artinya, mayoritas transaksi memiliki *Quantity* dan *Revenue* yang sangat rendah (berpusat di dekat nol), sementara hanya sedikit transaksi yang memiliki nilai yang sangat tinggi (*outliers*). Ini umum terjadi pada data ritel dan memvalidasi perlunya batas sumbu X (seperti yang dilakukan pada persentil ke-99) agar visualisasi mudah dibaca.

Tabel 1. Matriks Korelasi Variabel Kuantitatif

Ket	Price	Quantity	Revenue
Price	1.000000	-0.001931	0.452592
Quantity	-0.001931	1.000000	0.156283
Revenue	0.452592	0.156283	1.000000

Matriks korelasi digunakan untuk melihat hubungan linier antar variabel kuantitatif sebelum proses pembersihan data dilakukan. Hasilnya menunjukkan bahwa *Price* dan *Revenue* memiliki korelasi positif yang cukup kuat ($r \approx 0.45$), yang berarti semakin tinggi harga suatu item, semakin besar pula pendapatan yang dihasilkan. Di sisi lain, hubungan antara *Quantity* dan *Price* sangat lemah atau hampir tidak ada ($r \approx -0.002$), sehingga jumlah barang yang dibeli tidak memiliki kaitan linier dengan harga barang tersebut. Temuan ini menegaskan bahwa pembersihan data sangat penting, karena nilai *Quantity* dan *Price* yang negatif pada data mentah bisa membuat hasil korelasi menjadi tidak akurat.



Gambar 5. Heatmap Korelasi Antar Variabel Numerik

Heatmap ini menunjukkan hubungan linier antar variabel. Korelasi antara *Quantity* dan *Revenue* terlihat sangat kuat dan positif ($r \approx 0.90$), yang wajar karena *Revenue* adalah hasil perkalian *Quantity* dan *Price*. Sebaliknya, korelasi antara *Price* dan *Quantity* mendekati nol, menunjukkan bahwa harga item tidak secara linier mempengaruhi seberapa banyak item itu dibeli.

Pembersihan dan Penyiapan Data

Tahap ini berfokus pada proses validasi data agar hanya transaksi pembelian yang benar-benar valid yang digunakan dalam analisis Apriori.

- a) **Penyaringan Anomali:** Dilakukan filter untuk menghapus baris dengan nilai *Description* yang kosong (NaN) serta transaksi yang memiliki nilai *Quantity* atau *Price* tidak valid. Pada tahap ini, sekitar 13.000 baris data teridentifikasi sebagai anomali dan dihapus.
- b) **Filter Non-Produk:** Item yang tidak termasuk produk ritel, seperti POSTAGE dan DISCOUNT, juga dihilangkan dari dataset.
- c) **Pembatasan Wilayah:** Data kemudian difokuskan hanya pada transaksi yang berasal dari United Kingdom agar analisis lebih relevan.

Setelah membersihkan transaksi dengan nilai *Quantity* atau *Price* yang tidak positif, jumlah data tersisa adalah 511.566 baris. Angka ini diperoleh dari total awal 525.461 baris dikurangi kurang lebih 13.895 baris yang mengandung anomali. Setelah dilakukan penyaringan berdasarkan wilayah United Kingdom, jumlah data akhir yang digunakan untuk Market Basket Analysis (MBA) adalah 472.042 baris. Jumlah tersebut merupakan dataset akhir setelah seluruh proses pembersihan selesai dan hanya menyisakan transaksi yang memenuhi kriteria analisis.

Transformasi Data (*One-Hot-Encoding*)

Pada tahap ini, data yang telah dibersihkan diubah ke dalam format matriks biner sebagai prasyarat untuk menjalankan Algoritma Apriori.

- a) ***One-Hot Encoding*:** Dataset dikonversi menjadi matriks transaksi item (*basket_sets*) sehingga setiap transaksi direpresentasikan oleh baris, dan setiap produk oleh kolom.
- b) **Dimensi Matriks:** Hasil transformasi menghasilkan 20.410 transaksi unik dengan sekitar 4.000 item unik sebelum dilakukan penyaringan berdasarkan frekuensi kemunculan.
- c) **Penyaringan Item Jarang:** Produk yang muncul kurang dari ambang frekuensi 0,001 (0,1%) dihapus untuk menjaga efisiensi analisis. Setelah proses ini, tersisa sekitar 1.100 item unik yang paling relevan.

Jumlah item setelah filter frekuensi (0.001): 3014

Contoh Hasil One-hot Encoding (Head):

Description	10 COLOUR SPACEBOY PEN	12 ASS ZINC CHRISTMAS DECORATIONS	12 COLOURED PARTY BALLOONS	12 DAISY PEGS IN WOOD BOX	12 EGG HOUSE PAINTED WOOD	12 IVORY ROSE PEG PLACE SETTING
Invoice						
489434	0	0	0	0	0	0
489435	0	0	0	0	0	0
489436	0	0	0	0	0	0
489437	0	0	0	0	0	0
489438	0	0	0	0	0	0

Description	3 HOOK HANGER MAGIC GARDEN	3 HOOK PHOTO SHELF ANTIQUE WHITE	3 PIECE SPACEBOY COOKIE CUTTER SET	3 ROSE MORRIS BOXED CANDLES	3 STRIPEY MICE FELTCRAFT	3 TIER CAKE TIN GR
Invoice						
489434	0	0	0	0	0	0
489435	0	0	0	0	0	0
489436	0	0	0	0	0	0
489437	0	0	0	0	0	0
489438	0	0	0	0	0	0

Description	WRAP GREEN PEARS	WRAP GREEN RUSSIAN FOLKART	WRAP I LOVE LONDON	WRAP MONSTER FUN	WRAP PINK FAIRY CAKES	WRAP RED APPLES	WRAP SUKI AND FRIENDS	WRAP WEDDING DAY	WRAP
Invoice									
489434	0	0	0	0	0	0	0	0	0
489435	0	0	0	0	0	0	0	0	0
489436	0	0	0	0	0	0	0	0	0
489437	0	0	0	0	0	0	0	0	0
489438	0	0	0	0	0	0	0	0	0

[5 rows x 3014 columns]

Gambar 6. Contoh Matriks Biner Hasil One-Hot Encoding

Hasil yang ditampilkan merupakan implementasi dari tahap Transformasi Data menjadi Matriks Biner menggunakan. Tahapan ini memastikan bahwa data telah berada dalam format yang sesuai untuk diproses dengan Algoritma Apriori. Setelah menerapkan batas minimum frekuensi sebesar 0,001, jumlah item unik yang dapat digunakan untuk analisis adalah 3.014 item. Dalam matriks tersebut, setiap baris (berdasarkan Invoice ID) menggambarkan satu transaksi atau keranjang belanja, sedangkan setiap kolom menunjukkan satu produk unik. Nilai biner 1 menandakan bahwa produk tersebut dibeli dalam transaksi tersebut, sementara nilai 0 berarti produk tidak dibeli. Representasi ini memungkinkan algoritma asosiasi menghitung metrik seperti *support* dan *confidence* dengan lebih efisien.

Penerapan Algoritma Apriori Pada *Market Basket Analysis*

Tahap ini merupakan bagian utama dalam proses data mining, di mana Algoritma Apriori dijalankan pada matriks transaksi biner yang sudah melalui proses pembersihan dan transformasi. Tujuan utamanya adalah menemukan berbagai kombinasi item (itemsets) yang muncul secara signifikan dalam sejumlah transaksi.

Tabel 2. Total Frequent Itemsets

<i>Index</i>	<i>Support</i>	<i>Itemsets</i>
222	0.167881	WHITE HANGING HEART T-LIGHT HOLDER
168	0.096358	REGENCY CAKESTAND 3 TIER
199	0.080849	STRAWBERRY CERAMIC TRINKET BOX
48	0.070003	ASSORTED COLOUR BIRD ORNAMENT
74	0.067750	HOME BUILDING BLOCK WORD
127	0.065025	PACK OF 72 RETRO SPOT CAKE CASES
5	0.063715	60 TEATIME FAIRY CAKE CASES
86	0.061514	JUMBO BAG RED RETROSPOT
95	0.059261	JUMBO STORAGE BAG SUKI
232	0.058056	WOODEN FRAME ANTIQUE WHITE

Penerapan Algoritma Apriori dilakukan pada data transaksi yang telah disiapkan sebelumnya. Dengan menetapkan nilai minimum support sebesar 0,02, diperoleh sebanyak 268 frequent itemsets yang menunjukkan adanya pola pembelian yang cukup kuat di antara item-item tertentu. Item yang memiliki nilai support tertinggi didominasi oleh produk dekorasi rumah, antara lain *White Hanging Heart T-Light Holder*, *Regency Cakestand 3 Tier*, dan *Strawberry Ceramic Trinket Box*. Produk *White Hanging Heart T-Light Holder* tercatat sebagai item dengan nilai support tertinggi sebesar 0,167, yang menunjukkan bahwa produk ini paling sering muncul dalam transaksi dibandingkan item lain. Temuan ini mengindikasikan bahwa produk-produk tersebut memiliki permintaan yang stabil dan sering dibeli secara bersamaan oleh pelanggan.

Tabel 3. Total Association Rules

Index	Antecedents	Consequents	Support	Confidence	Lift
11	(Toilet metal sign)	(Bathroom metal sign)	0.020225	0.761341	17.591038
10	(Bathroom metal sign)	(Toilet metal sign)	0.020225	0.467312	17.591038
43	(Pink blue felt craft trinket box)	(Pink cream felt craft trinket box)	0.020121	0.615385	14.942259
42	(Pink cream felt craft trinket box)	(Pink blue felt craft trinket box)	0.020121	0.488556	14.942259
15	(Single heart zinc t-light holder)	(Hanging heart zinc t-light holder)	0.020697	0.625990	13.304041
14	(Hanging heart zinc t-light holder)	(Single heart zinc t-light holder)	0.020697	0.439866	13.304041
53	(Vintage heads and tails card game)	(Vintage snap cards)	0.022426	0.630515	12.743699
52	(Vintage snap cards)	(Vintage heads and tails card game)	0.022426	0.450805	12.743699
59	(Wooden picture frame white finish)	(Wooden frame antique white)	0.031019	0.644178	11.095799
58	(Wooden frame antique white)	(Wooden picture frame white finish)	0.031019	0.534926	11.095799

Setelah frequent itemsets berhasil diidentifikasi, tahap selanjutnya adalah membentuk *association rules* untuk menggali hubungan antarproduk dengan menggunakan metrik *lift* menggunakan nilai ambang minimum 1.0. Dari proses ini diperoleh 60 aturan asosiasi, yang kemudian disusun berdasarkan nilai *lift* tertinggi untuk menemukan korelasi yang paling kuat antar item. Hasil analisis menunjukkan adanya hubungan yang sangat kuat antara *Toilet Metal Sign* dan *Bathroom Metal Sign*, dengan nilai *lift* sebesar 17.59 dan *confidence* 0.76. Hal ini mengindikasikan bahwa pelanggan yang membeli salah satu dari kedua produk tersebut memiliki kemungkinan besar untuk membeli produk pasangannya. Pola yang sama juga terlihat pada item dekoratif seperti *Pink Cream Felt Craft Trinket Box* dan *Pink Blue Felt Craft Trinket Box*, yang mencerminkan kecenderungan konsumen untuk membeli produk dengan tema atau kategori yang sejenis. Temuan ini berpotensi dimanfaatkan dalam strategi *cross-selling* maupun dalam sistem rekomendasi produk yang lebih efektif.

Evaluasi Aturan Asosiasi

Tahap evaluasi dilakukan untuk mengonversi frequent itemsets menjadi association rules ($A \rightarrow C$) serta menilai kekuatan dan signifikansi hubungan antar item menggunakan metrik support, confidence, dan lift. Pada tahap ini, frequent itemsets diolah menjadi aturan asosiasi dan divalidasi dengan ketiga metrik tersebut, menghasilkan sekitar 60 aturan. Tabel 3 (output Rules) menampilkan aturan dengan korelasi terkuat, di mana nilai lift minimum 1.0 digunakan sebagai batas agar hanya hubungan positif yang dipertahankan. Aturan kemudian dinilai berdasarkan nilai lift sebagai indikator utama kekuatan korelasi, misalnya lift sekitar 17.59 dan nilai confidence sekitar 0.76 yang menunjukkan tingkat keandalan prediksi. Kombinasi ketiga metrik ini membuktikan bahwa pola pembelian komplementer yang ditemukan bersifat kuat dan dapat diandalkan untuk diterapkan lebih lanjut.

- a) *Support*: Mengukur popularitas itemset gabungan.

$$\text{Support}(A \rightarrow C) = \frac{\text{Jumlah transaksi mengandung } A \cup C}{\text{Total transaksi}}$$

- b) *Confidence*: Mengukur keandalan prediksi.

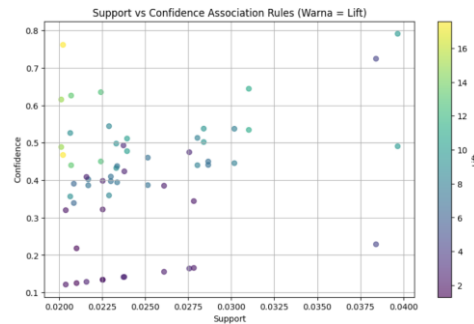
$$\text{Confidence}(A \rightarrow C) = \frac{\text{Support } A \cup C}{\text{Support}(A)}$$

- c) *Lift*: Mengukur korelasi antar item. Nilai Lift > 1.0 (seperti yang digunakan dalam filter) mengonfirmasi adanya korelasi positif yang signifikan.

$$\text{Lift}(A \rightarrow C) = \frac{\text{Confidence}(A \rightarrow C)}{\text{Support}(C)}$$

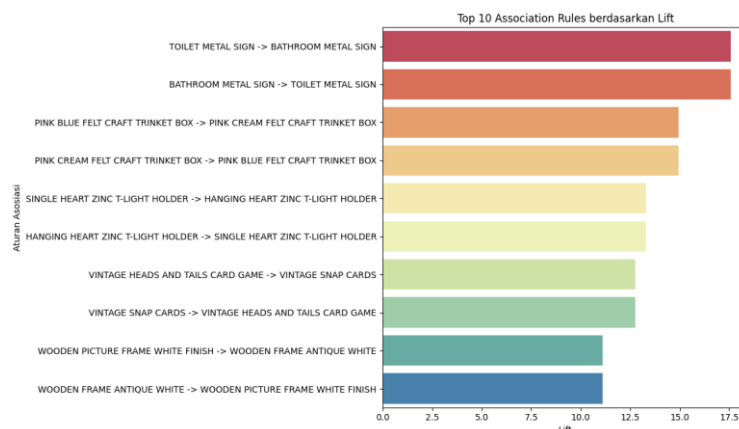
Analisis dan Visualisasi

Tahap ini berfungsi untuk menginterpretasikan hasil *association rules* yang telah divalidasi dan menerjemahkannya menjadi wawasan bisnis yang dapat digunakan secara langsung, dibantu dengan representasi visual.



Gambar 7. Scatter Plot Support vs Confidence

Gambar 7 menunjukkan *scatter plot* yang menggambarkan seluruh aturan asosiasi yang dihasilkan. Pada visualisasi tersebut, sumbu X menampilkan nilai *support*, sumbu Y menunjukkan nilai *confidence*, dan warna titik mewakili besarnya *lift*. Analisis terutama difokuskan pada titik-titik di area kanan atas, karena area ini mencerminkan aturan yang memiliki *support* tinggi (sering muncul) sekaligus *confidence* tinggi (andal). Aturan dengan warna yang lebih cerah, yang menandakan nilai *lift* tinggi, dianggap paling penting secara strategis, karena menunjukkan bahwa kekuatan hubungan antar item bukan sekadar terjadi karena popularitas item tertentu, tetapi benar-benar didukung oleh korelasi yang kuat.



Gambar 8. Bar Chart (Top 10 Lift)

Grafik batang ini secara jelas menampilkan 10 aturan asosiasi terkuat berdasarkan nilai *lift* (seperti yang terlihat pada output Gambar 8). Visualisasi tersebut berfungsi sebagai rangkuman akhir yang lebih ringkas dan bebas dari informasi yang tidak relevan, sehingga langsung menyoroti pola pembelian paling signifikan. Pola-pola dominan yang muncul misalnya kombinasi set piring dan cangkir yang saling melengkapi, atau dua produk dengan varian warna serupa menyediakan bukti konkret yang dapat digunakan sebagai dasar dalam penyusunan strategi rekomendasi atau *cross-selling* yang lebih efektif.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan Algoritma Apriori pada data transaksi ritel e-commerce untuk menggali pola pembelian konsumen melalui Market Basket Analysis. Seluruh proses data mining dilakukan secara terstruktur, mulai dari tahap pembersihan data, transformasi ke bentuk matriks biner, hingga identifikasi *frequent itemsets* dan pembentukan *association rules*. Dari hasil analisis, diperoleh 60 aturan asosiasi dengan nilai *lift* di atas 1.0, yang menunjukkan adanya hubungan positif dan kuat antarproduk tertentu. Temuan utama mengungkap bahwa sejumlah produk dekorasi rumah memiliki korelasi pembelian yang sangat tinggi, seperti keterkaitan antara *Toilet Metal Sign* dan *Bathroom Metal Sign* yang memiliki *lift* sebesar 17.59. Temuan ini menunjukkan bahwa algoritma Apriori mampu memberikan wawasan strategis berbasis data dalam memahami preferensi belanja konsumen.

Berdasarkan hasil yang diperoleh, pihak manajemen ritel disarankan untuk menerapkan strategi bundling promosi dan mengoptimalkan tata letak toko dengan menempatkan produk-produk yang memiliki nilai *lift* tinggi secara berdekatan misalnya varian *Metal Sign* atau *Trinket Boxes* untuk mendorong pembelian impulsif serta meningkatkan pendapatan. Untuk penelitian selanjutnya, disarankan menggunakan algoritma lain yang lebih efisien secara komputasi, seperti FP-Growth, terutama untuk menangani dataset berukuran besar mengingat Apriori memiliki keterbatasan dalam hal beban komputasi. Selain itu, penelitian di masa depan dapat diperluas dengan membandingkan pola perilaku konsumen antar-negara atau melakukan analisis musiman (*seasonal analysis*) agar diperoleh pemahaman tren yang lebih dinamis dan komprehensif.

DAFTAR REFERENSI

- Achmad, A. F., Rahim, A., & Verdikha, N. A. (2025). Implementasi data mining algoritma Apriori pada data transaksi Pik Store. *Jurnal Informatika Polinema*, 11(2), 203–212. <https://doi.org/10.33795/jip.v11i2.6860>
- Ahmad Thariq. (2023). Implementasi market basket analysis menggunakan algoritma Apriori pada data penjualan buku: Implementation of Market Basket Analysis using Apriori Algorithm on book sales data. *Jurnal Kolaboratif Sains*, 6(3), 154–163. <https://doi.org/10.56338/jks.v6i3.3333>
- Ananda, S. R. P., Putra, L. P. Y., Oktaviano, A. R., Sulanjari, T., Farida, A. N., & Sari, A. A. (2025). Analisis pola pembelian pelanggan menggunakan algoritma Apriori pada transaksi belanja supermarket. (*Detail penerbitan tidak tersedia*).
- Ashari, I. A., Wirasto, A., Triwibowo, D. N., & Purwono, P. (2022). Implementasi market basket analysis dengan algoritma Apriori untuk analisis pendapatan usaha retail. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 21(3), 701–709. <https://doi.org/10.30812/matrik.v21i3.1439>
- Brighton, K., & Hariyanto, S. (2024). Penerapan metode market basket analysis dengan algoritma Apriori pada toko ritel elektronik. *bit-Tech*, 7(1), 37–46. <https://doi.org/10.32877/bt.v7i1.1417>
- Butsianto, S., Naya, C., & Rifa'i, A. M. (2025). Implementasi algoritma Apriori dalam menemukan pola asosiasi pada data penjualan produk retail. *Bulletin CSR*, 5(5). <https://doi.org/10.47065/bulletincsr.v5i5.731>
- Fadillah, R., Qadriah, L., & Rizal, M. (2023). Market basket analysis data mining untuk mengetahui pola penjualan pada Cerry Mart Beureuneun menggunakan algoritma Apriori. *Jurnal Real Riset*, 5(1), 234–239. <https://doi.org/10.47647/jrr.v5i1.1152>
- Fuadi, A. N., Bhakti, M. H., & Premana, A. (2024). Analisis pola pembelian konsumen di toko ritel DMart menggunakan algoritma Apriori berbasis website. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3). <https://doi.org/10.23960/jitet.v12i3.4963>
- Hermawan, A., Wicaksono, B., Ahmadjayadi, T., Prakasa, B. S., & Aruan, J. D. (2024). Implementasi algoritma Apriori pada market basket analysis terhadap data penjualan produk supermarket. *Algoritma: Jurnal Matematika, Ilmu Pengetahuan Alam, Kebumian dan Angkasa*, 2(5), 95–105. <https://doi.org/10.62383/algoritma.v2i5.137>
- Kafabih, A., & Chandra, J. C. (2024). Implementasi algoritma Apriori dengan metode association rule dalam market basket analysis pada data transaksi Reddog. (*Detail penerbitan tidak tersedia*).
- Kuswanto, A. D., Blessar, A. R., Goni, A., Sidiki, A. N. N., Haryu, O. R. A., & Hamiki, H. A. (2024). Penerapan algoritma Apriori dalam analisis keranjang belanja retail di wilayah Jawa Barat. *Saturnus: Jurnal Teknologi dan Sistem Informasi*, 2(3), 139–150. <https://doi.org/10.61132/saturnus.v2i3.208>

- Madani, M. I., Masa, A. P. A., & Setyadi, H. J. (2024). Perbandingan metode Apriori dan frequent pattern growth dalam mengetahui pola pembelian konsumen. *Jurnal Teknologi Informasi dan Komunikasi (TIKomsin)*, 12(2), 58–66. <https://doi.org/10.30646/tikomsin.v12i2.876>
- Nurmayanti, W. P., Sastriana, H. M., Rahim, A., Gazali, M., Hirzi, R. H., Ramdani, Z., & Malthuf, M. (2021). Market basket analysis with Apriori algorithm and frequent pattern growth (FP-Growth) on outdoor product sales data. *International Journal of Educational Research & Social Sciences*, 2(1), 132–139. <https://doi.org/10.51601/ijersc.v2i1.45>
- Purnama, D. A. (2025). Data mining menggunakan association rules–market basket analysis untuk peningkatan kinerja ritel tradisional. *STRING*, 9(3). <https://doi.org/10.30998/string.v9i3.28707>
- Purnama, P. A. W., Putra, T. A., Afira, R., & Wijaya, R. (2025). Penerapan metode Apriori untuk pembelian di minimarket. *Remik*, 9(1), 95–101. <https://doi.org/10.33395/remik.v9i1.14249>
- Ramadana, W. D., Satyahadewi, N., & Perdana, H. (n.d.). Penerapan market basket analysis pada pola pembelian barang oleh konsumen menggunakan metode algoritma Apriori. (*Detail tahun dan jurnal tidak tersedia.*)
- Safitry, D. L., Rosianti, N., Divayaning, E., Zidan, H., Arnecia, Z. J., Paryudi, I., Veritawati, I., & Nursari, S. R. C. (2024). Analisis pola pembelian konsumen menggunakan algoritma Apriori untuk menentukan strategi pemasaran produk di toko retail X. *JATI*, 9(1), 505–511. <https://doi.org/10.36040/jati.v9i1.12429>
- Simanjuntak, A. C., Sitanggang, M. E., Cahyani, M. I., & Syahputri, N. (2024). Penerapan metode data mining market basket analysis terhadap data penjualan produk pada Toko Iblite Luxury menggunakan algoritma Apriori. *Bridge: Jurnal Publikasi Sistem Informasi dan Telekomunikasi*, 2(3), 62–74. <https://doi.org/10.62951/bridge.v2i3.106>
- Utami, I. T., & Dapa, R. (2024). Aplikasi metode market basket analysis dengan algoritma Apriori untuk mengetahui pola belanja konsumen pada online shop Amerta Fashion. *Indonesian Journal of Applied Statistics*, 6(1), 96–105. <https://doi.org/10.13057/ijas.v6i1.79963>
- Wulandari, U. M., Suseno, A. T., & Kholilurrahman, M. (2025). Market basket analysis using FP-Growth and Apriori on distro store sales transaction. *MATICS: Journal of Computer Science and Information Technology*, 17(1), 12–18. <https://doi.org/10.18860/mat.v17i1.28820>