

PENERAPAN DATA MINING APRIORI UNTUK ANALISIS PREFERENSI PRODUK TOKO RITEL

Muhammad Baldy Imalian¹, Anita Diana², Grace Gata³, Rizky Tahara Shita^{4*}

^{1,2,3} Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

⁴ Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹muhammadbaldyimalian@gmail.com, ²anita.diana@budiluhur.ac.id, ³grace.gata@budiluhur.ac.id,
^{4*}rizky.taharashita@budiluhur.ac.id

Abstrak—Di tengah persaingan industri ritel yang ketat, data transaksi yang masif sering kali belum dimanfaatkan secara optimal dan hanya berfungsi sebagai arsip. Hal ini menyebabkan keputusan strategis, seperti promosi, lebih sering didasarkan pada intuisi daripada analisis data. Penelitian ini bertujuan untuk mengatasi masalah tersebut dengan menganalisis preferensi konsumen terhadap produk melalui penerapan metode *data mining*, khususnya algoritma *Apriori*, untuk menemukan pola *association rules* yang tersembunyi. Penelitian ini menyajikan kerangka kerja analitis sistematis dengan mengadopsi metodologi *Cross-Industry Standard Process for Data Mining (CRISP-DM)* dan menggunakan *dataset* publik *Groceries Dataset* yang mencakup transaksi dari tahun 2014 hingga 2015. Proses *preprocessing* data yang komprehensif diterapkan pada 38.765 catatan pembelian mentah, mencakup pembersihan 759 data duplikat, normalisasi data, dan rekayasa fitur untuk mengekstrak informasi temporal, menghasilkan 14.963 transaksi unik yang siap dianalisis. Proses teknis diimplementasikan menggunakan bahasa *Python* dengan *library Pandas* dan *MLxtend*. Dengan menerapkan parameter *minimum support* 0,001 dan *minimum lift* > 1, model berhasil mengidentifikasi 240 *association rules* signifikan yang terbagi menjadi 210 aturan sederhana dan 15 aturan kompleks. Di antara temuan tersebut, aturan kompleks {whole milk, yogurt} → {sausage} menonjol sebagai yang terkuat dengan nilai *lift* sebesar 2,18, mengindikasikan hubungan pembelian bersama yang sangat kuat. Selain itu, ditemukan pula pola *cross-selling* potensial seperti {specialty chocolate} → {citrus fruit} dengan *lift* 1,65. Temuan ini memberikan dasar empiris untuk merancang strategi pemasaran yang dapat ditindaklanjuti, misalnya melalui strategi *product bundling* seperti “Paket Promo American Breakfast”, sementara aturan lainnya mendukung strategi *cross-selling*. Kesimpulannya, penelitian ini membuktikan bahwa algoritma *Apriori* yang diimplementasikan dalam kerangka *CRISP-DM* merupakan pendekatan efektif untuk mentransformasi data transaksi mentah menjadi wawasan strategis yang konkret dan bermanfaat bagi pelaku bisnis ritel.

Kata kunci: Algoritma Apriori, Association Rule, Data Mining, Preferensi Konsumen, Strategi Pemasaran.

IMPLEMENTATION OF APRIORI DATA MINING FOR RETAIL PRODUCT PREFERENCE ANALYSIS

Abstract—Amidst the increasingly competitive retail industry, massive volumes of transaction data often remain underutilized, serving merely as administrative archives, resulting in strategic decisions, particularly regarding promotions, being based on intuition rather than data-driven analysis. This study aims to address this gap by analyzing consumer product preferences through the application of data mining techniques, specifically the Apriori algorithm, to uncover hidden association rules. This research presents a systematic analytical framework adopting the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, utilizing the public “Groceries Dataset” which covers transactions from 2014 to 2015. A comprehensive data preprocessing workflow was applied to 38,765 raw purchase records, including the cleaning of 759 duplicate entries, data normalization, and feature engineering to extract temporal information, resulting in 14,963 unique transactions ready for analysis. The technical implementation was carried out using Python with the Pandas and MLxtend libraries. By applying parameters of a minimum support of 0.001 and a minimum lift greater than 1, the model successfully identified 240 statistically significant association rules, consisting of 210 simple rules and 15 complex rules. Among the findings, the complex rule {whole milk, yogurt} → {sausage} stood out as the strongest, with a lift value of 2.18, indicating a very strong co-purchase relationship. Furthermore, potential cross-selling patterns such as {specialty chocolate} → {citrus fruit} with a lift of 1.65 were also discovered. These findings provide an empirical basis for designing actionable marketing strategies, with the identified breakfast product patterns supporting product bundling initiatives like an “American Breakfast Promo Package,” while other rules support targeted cross-selling campaigns. In conclusion, this study demonstrates that the Apriori algorithm, implemented within the CRISP-DM framework, is an effective approach for transforming raw transaction data into concrete and valuable strategic insights for retail businesses.

Keywords: Apriori Algorithm, Association Rule, Data Mining, Consumer Preference, Marketing Strategy.

1. PENDAHULUAN

Persaingan yang semakin ketat dalam industri ritel menuntut pelaku usaha untuk mengambil keputusan yang cepat, tepat, dan berbasis data. Data transaksi penjualan yang dihasilkan setiap hari sebenarnya menyimpan potensi besar untuk diolah menjadi wawasan strategis. Namun, dalam praktiknya, data tersebut sering kali hanya digunakan sebagai arsip administratif dan belum dimanfaatkan secara optimal untuk mendukung strategi pemasaran maupun pengelolaan stok [1]. Akibatnya, keputusan penting seperti penentuan promosi, tata letak produk, atau penyusunan paket penjualan masih kerap bergantung pada intuisi, yang berisiko tidak sesuai dengan perilaku nyata konsumen [2].

Salah satu pendekatan yang efektif untuk menggali potensi tersebut adalah *data mining*, yaitu proses menemukan pola atau hubungan tersembunyi dalam kumpulan data berukuran besar [3]. Di antara berbagai teknik *data mining*, algoritma *Apriori* banyak digunakan untuk melakukan *association rule mining*, yaitu menemukan keterkaitan antar produk yang sering dibeli secara bersamaan [4]. Hasil analisis *association rules* dapat dimanfaatkan untuk menyusun strategi pemasaran yang lebih tepat sasaran, misalnya melalui *product bundling* dan *cross-selling*, yang terbukti mampu meningkatkan penjualan serta loyalitas pelanggan [5].

Berbagai penelitian terdahulu telah menerapkan algoritma *Apriori* pada beragam konteks bisnis ritel, baik di toko fisik maupun penjualan daring. Misalnya, penelitian [1] membuktikan efektivitas *Apriori* dalam menemukan pola belanja pada toko ritel elektronik, sementara [2] menggunakan pendekatan serupa untuk menganalisis perilaku konsumen di ritel lokal. Penelitian lain juga menunjukkan bahwa penerapan algoritma *Apriori* dapat menghasilkan pola pembelian yang signifikan, mulai dari pengelolaan stok [6] hingga penyusunan strategi penjualan di berbagai sektor [7].

Selain algoritma *Apriori*, terdapat pula metode lain yang banyak digunakan dalam analisis *market basket*, seperti *FP-Growth* dan *clustering-based association*. Algoritma *FP-Growth* dikenal lebih efisien dalam mengolah dataset berukuran sangat besar karena tidak perlu menghasilkan kandidat itemset secara eksplisit, berbeda dengan *Apriori* yang relatif lebih lambat [8], [9]. Sementara itu, pendekatan *clustering* digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik tertentu sebelum dilakukan analisis asosiasi [2], [10].

Namun, meskipun terdapat alternatif tersebut, algoritma *Apriori* tetap relevan dan banyak dipilih karena kesederhanaannya, transparansi proses, serta kemudahan interpretasi aturan yang dihasilkan [1], [4]. Hal inilah yang menjadikannya metode populer di kalangan peneliti maupun praktisi bisnis ritel.

Meskipun demikian, sebagian besar penelitian lebih menekankan pada hasil akhir analisis tanpa menguraikan secara rinci proses *data preprocessing*, padahal tahap ini berperan penting dalam memastikan kualitas dan akurasi pola yang terbentuk [11]. Hal ini menjadi celah penelitian yang ingin dijawab melalui studi ini.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma *Apriori* pada *Groceries Dataset* publik dengan mengadopsi metodologi *Cross-Industry Standard Process for Data Mining (CRISP-DM)*. Fokus penelitian mencakup pembersihan data, normalisasi, dan rekayasa fitur agar data siap dianalisis, serta evaluasi kualitas aturan asosiasi menggunakan metrik *support*, *confidence*, dan *lift*. Kontribusi penelitian ini adalah memberikan kerangka kerja analisis end-to-end yang tidak hanya menghasilkan aturan asosiasi, tetapi juga menerjemahkannya ke dalam rekomendasi strategis yang aplikatif bagi pelaku bisnis ritel dalam merancang strategi pemasaran berbasis data.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan metodologi *Cross-Industry Standard Process for Data Mining (CRISP-DM)*, yang terdiri atas enam tahap utama yang saling berhubungan dan membentuk kerangka kerja analisis yang sistematis [3].

- Business Understanding*: tahap awal untuk mengidentifikasi tujuan bisnis dan mendefinisikan masalah yang akan diselesaikan melalui analisis data [2].
- Data Understanding*: meliputi pengumpulan data, pemeriksaan kualitas, dan eksplorasi awal untuk memahami karakteristik serta menemukan pola awal dalam data [1].
- Data Preparation*: mencakup pembersihan data, transformasi, dan pembuatan atribut baru yang relevan. Pada penelitian ini termasuk penghapusan duplikat, normalisasi format data, dan rekayasa fitur berbasis waktu [6].
- Modeling*: dilakukan dengan algoritma *Apriori* untuk membangun *association rules* yang signifikan dari data transaksi [12].
- Evaluation*: tahap penilaian kualitas aturan menggunakan metrik *support*, *confidence*, dan *lift* untuk memastikan aturan relevan secara statistik maupun bisnis [13].
- Deployment*: hasil analisis diterapkan dalam bentuk rekomendasi strategis, seperti *product bundling* dan *cross-selling*, agar dapat langsung dimanfaatkan oleh pelaku usaha ritel [7].

Keenam tahap CRISP-DM tidak berjalan secara terpisah, tetapi membentuk alur yang saling berhubungan. Hasil dari *business understanding* menjadi dasar dalam memilih *dataset* dan menentukan fokus analisis pada tahap

data understanding. Misalnya, ketika masalah utama yang ditemukan adalah rendahnya pemanfaatan data transaksi ritel, kebutuhan analisis kemudian diarahkan untuk menemukan pola asosiasi produk yang mendukung strategi pemasaran.

Informasi yang diperoleh dari *data understanding* selanjutnya digunakan untuk merancang proses *data preparation*. Sebagai contoh, jika ditemukan adanya duplikasi transaksi atau perbedaan penulisan produk (misalnya “Whole Milk” dan “whole milk”), maka pembersihan dan normalisasi data menjadi prioritas utama.

Data yang telah diproses dengan baik akan menghasilkan *modeling* yang lebih akurat ketika algoritma *Apriori* diterapkan. Aturan asosiasi yang terbentuk kemudian masuk ke tahap *evaluation*, di mana selain diuji dengan metrik *support*, *confidence*, dan *lift*, juga ditinjau relevansinya terhadap kebutuhan bisnis ritel. Hasil evaluasi ini pada akhirnya digunakan dalam tahap *deployment* untuk menyusun rekomendasi nyata, seperti penyusunan paket promosi (*product bundling*) atau strategi penempatan produk (*cross-selling*).



Gambar 1. Tahapan Penelitian Menggunakan Metodologi CRISP-DM

2.2 Sumber Data

Penelitian ini menggunakan *Groceries Dataset* yang bersifat publik dan tersedia pada repositori *machine learning*. Dataset ini berisi catatan transaksi pembelian konsumen di sebuah toko ritel selama periode tahun 2014 hingga 2015 [1]. Format data disajikan per baris transaksi, di mana setiap baris merepresentasikan pembelian satu atau lebih produk oleh pelanggan.

Secara keseluruhan, dataset mentah berjumlah 38.765 catatan transaksi. Setelah dilakukan penghapusan 759 catatan duplikat, diperoleh 38.006 catatan unik. Selanjutnya, melalui proses agregasi berdasarkan *Transaction ID* dan *Member number*, data tersebut dikonsolidasikan menjadi 14.963 transaksi unik yang siap dianalisis. Perbedaan antara *catatan unik* dan *transaksi unik* penting untuk diperhatikan: *catatan unik* merujuk pada baris individual tanpa duplikat, sementara *transaksi unik* merepresentasikan kumpulan produk yang benar-benar dibeli dalam satu transaksi pelanggan.

Alasan pemilihan *Groceries Dataset* adalah: (1) memiliki volume data yang cukup besar sehingga mendukung penerapan *data mining*, (2) struktur datanya sesuai dengan kebutuhan analisis *market basket*, dan (3) bersifat publik, sehingga memudahkan proses replikasi maupun validasi hasil penelitian oleh peneliti lain [6].

Dataset asli berbentuk format “lebar” (*long format*), di mana setiap baris hanya merepresentasikan satu produk dari sebuah transaksi. Contoh potongan data mentah dapat dilihat pada Tabel 1.

Tabel 1. Contoh Data Mentah

<i>id</i>	<i>member number</i>	<i>date</i>	<i>itemDescription</i>
1	1001	2014-01-05	<i>whole milk</i>
2	1001	2014-01-05	<i>yogurt</i>
3	1002	2014-01-06	<i>rolls/buns</i>

Dari tabel tersebut, terlihat bahwa transaksi dengan *member_number* 1001 pada tanggal 5 Januari 2014 sebenarnya mencakup dua produk (“*whole milk*” dan “*yogurt*”). Oleh karena itu, data perlu diagregasi sehingga menjadi satu transaksi unik dengan item {*whole milk*, *yogurt*}.

Tabel 2. Deskripsi *Dataset*

Nama Atribut	Deskripsi	Tipe Data	Jumlah Nilai Unik
<i>id</i>	Nomor urut transaksi	<i>Integer</i>	38.765
<i>Member_number</i>	Nomor anggota pembeli	<i>Integer</i>	5.212
<i>Date</i>	Tanggal transaksi	<i>Date</i>	728
<i>year</i>	Tahun transaksi	<i>Integer</i>	2
<i>month</i>	Bulan transaksi	<i>Integer</i>	12
<i>weekday</i>	Hari transaksi	<i>String</i>	7
<i>itemDescription</i>	Nama produk yang dibeli	<i>String</i>	169

Setelah proses agregasi, dari 38.765 catatan mentah (*raw records*) yang tersisa 38.006 catatan unik, diperoleh total 14.963 transaksi unik yang siap dianalisis. Distingsi ini penting karena *catatan unik* adalah baris individual tanpa duplikat, sedangkan *transaksi unik* mewakili sekumpulan item yang benar-benar dibeli bersama oleh pelanggan.

2.3 Tahap *Data Preprocessing*

Proses *data preprocessing* dilakukan untuk memastikan kualitas data berada pada kondisi optimal sebelum digunakan dalam pemodelan *association rules*. Tahap ini merupakan langkah penting karena kualitas hasil analisis sangat dipengaruhi oleh kebersihan dan konsistensi data [1], [14]. Secara umum, terdapat tiga langkah utama yang dilakukan:

- Penghapusan duplikat**
Dari total 38.765 catatan transaksi mentah, ditemukan 759 catatan duplikat. Setelah dihapus, tersisa 38.006 catatan unik yang lebih representatif [6]. Langkah ini penting agar analisis tidak bias akibat pengulangan data yang sebenarnya tidak mencerminkan perilaku belanja riil pelanggan.
- Normalisasi format data**
Seluruh nilai pada atribut *itemDescription* dikonversi menjadi huruf kecil (*lowercase*), penghapusan spasi ganda dilakukan, serta format tanggal diseragamkan menjadi *YYYY-MM-DD*. Normalisasi ini memastikan konsistensi sehingga data dapat diproses oleh algoritma secara akurat [13]. Contoh transformasi data sebelum dan sesudah normalisasi dapat dilihat pada Tabel 3.

Tabel 3. Contoh Transformasi Data Setelah Normalisasi

<i>id</i>	<i>member number</i>	<i>date</i>	<i>year</i>	<i>month</i>	<i>weekday</i>	<i>itemDescription</i>
1001	1640	2015-01-02	–	–	–	<i>Whole Milk</i>
1001	1640	2015-01-02	2015	01	<i>Friday</i>	<i>whole milk</i>

- Rekayasa fitur (*feature engineering*)**
Dari atribut *date* diturunkan tiga atribut baru, yaitu *year*, *month*, dan *weekday*. Informasi temporal ini memungkinkan analisis yang lebih mendalam, misalnya untuk membandingkan pola pembelian pada hari kerja (*weekday*) dengan akhir pekan (*weekend*), atau tren musiman berdasarkan bulan tertentu. Dengan cara ini, hasil analisis tidak hanya terbatas pada hubungan antar produk, tetapi juga dapat menjelaskan kapan produk-produk tertentu lebih mungkin dibeli secara bersamaan [5].

Tabel 4. Hasil *Preprocessing Data*

Tahap <i>Preprocessing</i>	Jumlah Data Sebelum	Jumlah Data Sesudah
Penghapusan duplikat	38.765	38.006
Normalisasi format	38.006	38.006
Rekayasa fitur	38.006	38.006 (+3 atribut baru: <i>year</i> , <i>month</i> , <i>weekday</i>)

2.4 Pemodelan Algoritma *Apriori*

Pemodelan dilakukan menggunakan algoritma *Apriori* untuk menghasilkan *association rules* dari data transaksi ritel. Parameter utama yang digunakan dalam penelitian ini adalah minimum support = 0,001 dan minimum lift > 1. Selain itu, pada tahap eksperimen juga dipertimbangkan nilai minimum confidence = 0,3 untuk menyaring aturan dengan tingkat kepastian yang lebih tinggi.

Implementasi algoritma dilakukan menggunakan bahasa pemrograman *Python* dengan bantuan *library* *Pandas* untuk pengolahan data serta *MLxtend* untuk membangun model *association rules* [15]. Tiga metrik utama yang digunakan dalam evaluasi aturan adalah *support*, *confidence*, dan *lift*:

- Support*** digunakan untuk mengukur seberapa sering suatu item atau kombinasi item muncul dalam keseluruhan transaksi.
- Confidence*** mengukur probabilitas munculnya item *Y* ketika item *X* ditemukan dalam suatu transaksi.

- c. *Lift* digunakan untuk menilai kekuatan hubungan antara item X dan Y , dengan memperhitungkan rasio nilai *confidence* terhadap probabilitas kemunculan item Y secara keseluruhan.

$$Support(X) = \frac{Jumlah\ Transaksi\ X}{Total\ Transaksi} \quad (1)$$

$$Confidence(X \rightarrow Y) = \frac{Support(X \cup Y)}{Support(X)} \quad (2)$$

$$Lift(X \rightarrow Y) = \frac{Confidence(X \rightarrow Y)}{Support(Y)} \quad (3)$$

2.5 Evaluasi Model

Evaluasi model dilakukan untuk menilai kualitas aturan asosiasi yang terbentuk menggunakan tiga metrik utama, yaitu *support*, *confidence*, dan *lift* [8].

- Support* digunakan untuk melihat frekuensi kemunculan kombinasi produk dalam keseluruhan transaksi.
- Confidence* mengukur tingkat kepastian bahwa suatu konsekuen (Y) muncul apabila anteseden (X) ditemukan.
- Lift* menilai kekuatan hubungan antara X dan Y dengan membandingkan *confidence* terhadap probabilitas kemunculan Y secara acak.

Seleksi aturan terbaik ditentukan berdasarkan nilai *lift* tertinggi serta relevansi aturan terhadap konteks bisnis. Aturan dengan nilai *lift* > 1 menunjukkan adanya keterkaitan positif antar produk, sedangkan aturan dengan nilai *lift* < 1 cenderung tidak relevan [16].

3. HASIL DAN PEMBAHASAN

3.1 Hasil Eksperimen (Data & Aturan Asosiasi)

Proses eksperimen dilakukan dengan menerapkan algoritma *Apriori* pada *Groceries Dataset* yang telah melalui tahap *preprocessing*. Parameter yang digunakan meliputi *minimum support* = 0,001, *minimum confidence* = 0,3, dan *minimum lift* > 1. Hasil pengolahan ini menghasilkan sekitar 240 aturan asosiasi signifikan, yang dapat dikelompokkan menjadi *frequent itemset* utama, aturan sederhana, dan aturan kompleks.

Frequent itemset menggambarkan item atau kombinasi item yang paling sering muncul dalam transaksi. Sepuluh *frequent itemset* teratas ditunjukkan pada Tabel 5.

Tabel 5. *Frequent Itemset* Teratas Berdasarkan Nilai *Support*

<i>Itemset</i>	<i>Support</i>
{whole milk}	0.1579
{other vegetables}	0.1221
{rolls/buns}	0.1100
{soda}	0.0971
{yogurt}	0.0859
{root vegetables}	0.0696
{tropicalfruit}	0.0678
{bottled water}	0.0607
{sausage}	0.0603
{citrus fruit}	0.0531

Hasil Tabel 5 menunjukkan bahwa produk *whole milk*, *other vegetables*, dan *rolls/buns* merupakan produk yang paling sering dibeli, dengan nilai *support* masing-masing 0.1579, 0.1221, dan 0.1100.

Aturan asosiasi sederhana terdiri dari satu *antecedent* menuju satu *consequent*. Aturan ini penting karena menunjukkan keterkaitan langsung antar produk, misalnya kecenderungan konsumen membeli produk tertentu bersamaan dengan produk lain. Contoh aturan sederhana dengan nilai *lift* tertinggi ditunjukkan pada Tabel 6.

Tabel 6. Contoh Aturan Asosiasi Sederhana (*Lift* Tertinggi)

<i>Antecedent</i>	<i>Consequent</i>	<i>Support</i>	<i>Confidence</i>	<i>Lift</i>
{specialty chocolate}	{citrus fruit}	0.0014	0.0879	1.654
{tropical fruit}	{flour}	0.0011	0.0158	1.617
{beverages}	{sausage}	0.0015	0.0927	1.537
{napkins}	{pastry}	0.0017	0.0786	1.519

Selain itu, ditemukan aturan asosiasi kompleks dengan lebih dari satu *antecedent*. Aturan ini sering lebih bermanfaat secara bisnis karena mendukung strategi *product bundling* atau *cross-selling*. Beberapa aturan kompleks terbaik ditunjukkan pada Tabel 7.

Tabel 7. Contoh Aturan Asosiasi Kompleks (Rekomendasi Bundling)

<i>Antecedent</i>	<i>Consequent</i>	<i>Support</i>	<i>Confidence</i>	<i>Lift</i>
{yogurt, whole milk}	{sausage}	0.0015	0.1317	2.183
{sausage, whole milk}	{yogurt}	0.0015	0.1642	1.912
{sausage, yogurt}	{whole milk}	0.0015	0.2558	1.620
{soda, whole milk}	{sausage}	0.0011	0.0920	1.524

Temuan ini menunjukkan bahwa produk *whole milk*, *yogurt*, dan *sausage* berperan sebagai pusat asosiasi karena sering muncul baik sebagai *antecedent* maupun *consequent*. Secara bisnis, hal ini menegaskan bahwa produk-produk tersebut memiliki potensi besar untuk dijadikan acuan dalam strategi pemasaran berbasis rekomendasi, misalnya melalui *cross-selling* maupun *product bundling* [2], [8].

3.2 Analisis Aturan Asosiasi

Aturan asosiasi yang terbentuk dari eksperimen *Apriori* memberikan gambaran perilaku belanja konsumen dalam *Groceries Dataset*. Sebagian besar aturan memiliki nilai *support* yang rendah ($< 0,02$), hal ini wajar karena variasi kombinasi produk dalam ritel sangat banyak. Meski demikian, aturan dengan nilai *lift* tinggi tetap signifikan dan menunjukkan adanya keterkaitan kuat antar produk [1], [8]. Salah satu aturan kompleks yang paling menonjol adalah:

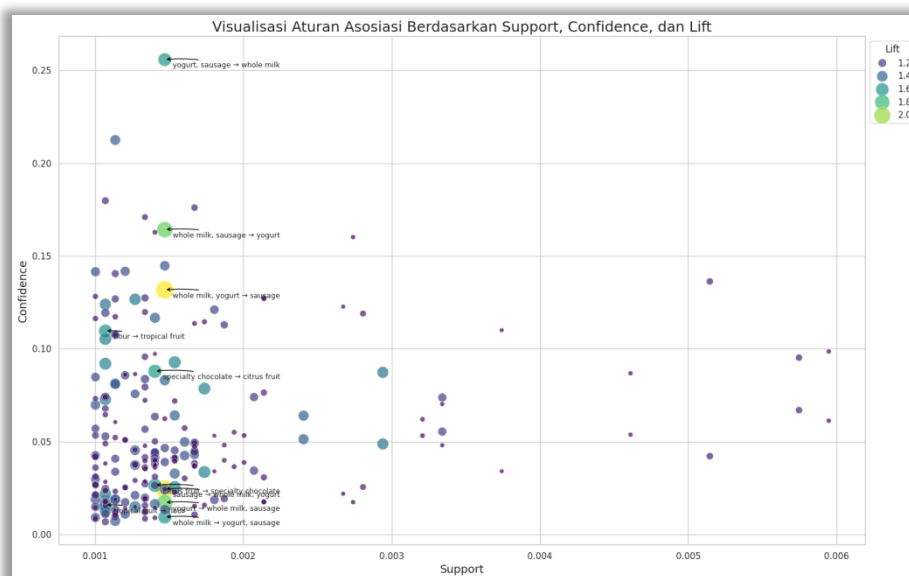
- {*whole milk*, *yogurt*} \rightarrow {*sausage*} dengan nilai *lift* sebesar 2,18. Aturan ini berarti bahwa konsumen yang membeli *whole milk* dan *yogurt* secara bersamaan memiliki kemungkinan 118% lebih besar untuk juga membeli *sausage* dibandingkan konsumen secara umum. Secara perilaku, hal ini dapat dijelaskan karena produk susu, yogurt, dan daging olahan sering dipandang sebagai kebutuhan *household groceries* yang saling melengkapi. Bagi ritel, aturan ini dapat dimanfaatkan untuk menyusun strategi *product bundling* atau *cross-selling*, misalnya dengan menempatkan *sausage* berdekatan dengan produk susu dan yogurt, atau menawarkan paket promosi yang menggabungkan ketiga produk tersebut [2], [12].
- Selain itu, aturan sederhana {*specialty chocolate*} \rightarrow {*citrus fruit*} juga menarik meskipun memiliki *support* yang relatif rendah. Nilai *lift* yang tinggi pada aturan ini menandakan adanya kecenderungan konsumen yang membeli *specialty chocolate* juga memilih *citrus fruit*. Fenomena ini dapat diinterpretasikan sebagai pola konsumsi tertentu, di mana konsumen mengombinasikan produk *snack* atau makanan manis dengan buah segar. Dari sisi pemasaran, informasi ini bisa mendukung strategi promosi *cross-category*, misalnya dengan memberikan diskon pembelian buah ketika konsumen membeli cokelat [13].

Secara keseluruhan, temuan ini menegaskan bahwa meskipun mayoritas aturan memiliki *support* rendah, nilai *lift* yang tinggi menjadi indikator kuat untuk mengidentifikasi aturan yang signifikan. Dengan demikian, analisis aturan asosiasi bukan hanya berfungsi menemukan hubungan statistik, tetapi juga menghasilkan wawasan praktis yang dapat digunakan untuk mendukung strategi pemasaran ritel berbasis data.

3.3 Visualisasi Hasil Pengujian

Visualisasi bertujuan untuk memfasilitasi pemahaman mengenai seberapa kuat dan valid hubungan antar produk yang telah terdeteksi melalui *association rule mining*. Salah satu metode yang efektif adalah *scatter plot* (diagram sebar), yang mampu menggambarkan distribusi aturan berdasarkan tingkat kepentingannya secara visual [17], [18].

Hasil pengujian algoritma *Apriori* divisualisasikan pada *scatter plot* yang memetakan aturan asosiasi berdasarkan tiga metrik evaluasi utama: *support*, *confidence*, dan *lift*. Visualisasi ini memperdalam pemahaman terhadap aturan asosiasi yang telah terbentuk, sekaligus membantu mengidentifikasi pola paling strategis untuk implementasi dalam konteks pemasaran dan pengambilan keputusan di toko ritel.



Gambar 2. Scatter Plot Aturan Asosiasi Berdasarkan *Support*, *Confidence*, dan *Lift*

Pada scatter plot, setiap titik merepresentasikan satu aturan asosiasi, dengan sumbu X (*support*) menunjukkan seberapa sering kombinasi produk dalam aturan tersebut muncul di seluruh transaksi, sumbu Y (*confidence*) menunjukkan tingkat kepastian atau probabilitas aturan tersebut berlaku, dan warna serta ukuran titik mencerminkan *lift*. Semakin terang warnanya (mendekati kuning) dan semakin besar ukuran titik, semakin tinggi nilai *lift*, yang menandakan hubungan asosiasi yang sangat kuat. Dari visualisasi ini, dapat ditarik beberapa interpretasi penting mengenai distribusi dan kekuatan aturan dalam dataset.

- Distribusi Aturan:**
Sebagian besar aturan cenderung terkonsentrasi pada nilai *support* rendah ($< 0,02$), sesuai karakteristik dataset ritel di mana kombinasi produk spesifik jarang memiliki frekuensi tinggi. Hal ini menegaskan bahwa aturan yang paling relevan fokus pada nilai *lift* dan *confidence* tinggi, meskipun *support*-nya rendah.
- Identifikasi Aturan Paling Signifikan:**
Titik-titik menonjol menunjukkan aturan dengan nilai *lift* tertinggi, misalnya aturan kompleks $\{whole\ milk, yogurt\} \rightarrow \{sausage\}$ dengan $lift \approx 2.183$. Aturan ini terlihat jelas sebagai titik terbesar dan paling terang, menandakan pola preferensi konsumen utama yang berhasil diidentifikasi.
- Korelasi vs Popularitas:**
Visualisasi membantu membedakan antara aturan yang hanya populer (*support* tinggi) dengan aturan yang memiliki hubungan asosiasi kuat (*lift* tinggi). Beberapa aturan yang melibatkan *whole milk* (produk populer) mungkin memiliki *support* relatif tinggi, tetapi nilai *lift* mendekati 1, menunjukkan asosiasi yang lebih lemah dibanding aturan yang telah ditandai sebagai signifikan.

Secara keseluruhan, visualisasi hasil pengujian ini memvalidasi model dengan menyoroti aturan-aturan yang paling signifikan secara statistik dan relevan secara bisnis. Hasil ini menjadi dasar rekomendasi strategi pemasaran berbasis *cross-selling* atau *product bundling* [1], [8].

3.4 Diskusi & Implikasi Bisnis

Temuan aturan asosiasi pada *Groceries Dataset* memberikan implikasi strategis yang dapat diterapkan dalam praktik bisnis ritel. Beberapa rekomendasi utama yang dapat diturunkan antara lain:

- Strategi Product Bundling**
Aturan kompleks $\{whole\ milk, yogurt\} \rightarrow \{sausage\}$ menunjukkan peluang untuk merancang paket tematik, misalnya Paket Sarapan “*American Breakfast*”, yang menggabungkan produk susu, yogurt, dan sosis. Bundling ini dapat meningkatkan nilai rata-rata transaksi sekaligus menarik konsumen yang mencari kemudahan belanja.
- Penerapan Cross-Selling**
Aturan sederhana atau kompleks lain, seperti $\{sausage, rolls/buns\} \rightarrow \{whole\ milk\}$ dan $\{rolls/buns, other\ vegetables\} \rightarrow \{soda\}$, dapat dimanfaatkan dalam sistem rekomendasi produk tambahan di *point-of-sale*. Misalnya, saat pelanggan membeli roti dan sayuran, sistem dapat menyarankan minuman *soda* atau *whole milk*.

c. Promosi Bertema dan Musiman

Strategi promosi dapat diarahkan pada periode tertentu dengan memanfaatkan atribut temporal (*month, weekday*). Contohnya, promosi akhir pekan untuk produk *beverages* atau paket musiman untuk *tropical fruit* sesuai pola belanja konsumen.

Rekomendasi ini menegaskan bahwa hasil analisis *association rules* dapat diterjemahkan menjadi strategi bisnis aplikatif. Dengan demikian, penerapan algoritma *Apriori* mendukung pengambilan keputusan berbasis data yang sesuai perilaku nyata konsumen [1], [8].

4. KESIMPULAN

Penelitian ini telah berhasil menerapkan algoritma *Apriori* dengan kerangka kerja *CRISP-DM* pada *Groceries Dataset* periode 2014–2015, yang mencakup transaksi ritel konsumen sehari-hari. Proses awal berupa *data preprocessing* dilakukan secara menyeluruh, mulai dari pembersihan data, penanganan duplikasi, hingga agregasi transaksi, sehingga dari 38.006 catatan mentah berhasil dihasilkan 14.963 transaksi unik yang siap dianalisis. Langkah ini penting untuk memastikan kualitas data, karena kualitas data secara langsung mempengaruhi akurasi dan relevansi *association rules* yang diperoleh.

Pada tahap *association rule mining*, digunakan parameter minimum *support* = 0,001, *confidence* \geq 0,3, dan *lift* > 1. Dengan kriteria ini, diperoleh sekitar 240 *association rules* yang signifikan. Aturan-aturan tersebut mencakup kombinasi produk yang sederhana maupun yang lebih kompleks, menunjukkan keterkaitan antara berbagai item yang dibeli bersama. Salah satu pola terkuat yang muncul adalah $\{whole\ milk, yogurt\} \rightarrow \{sausage\}$, dengan nilai *lift* sebesar 2,18, yang menunjukkan bahwa kombinasi produk ini memiliki kemungkinan dibeli bersama lebih dari dua kali lipat dibandingkan jika dibeli secara independen. Temuan ini memberikan *insight* yang bermanfaat bagi strategi pemasaran, seperti pengembangan paket promosi, dan *cross-selling* di toko untuk meningkatkan pengalaman berbelanja dan penjualan.

Secara keseluruhan, hasil penelitian ini menegaskan bahwa *data mining* berbasis *association rules* merupakan alat yang efektif untuk mendukung pengambilan keputusan di sektor ritel. Penerapan metode ini memungkinkan pihak manajemen memahami pola pembelian konsumen secara lebih mendalam, mengidentifikasi produk-produk yang memiliki hubungan kuat, dan merancang strategi pemasaran yang lebih terarah serta berbasis data. Sebagai pengembangan penelitian lebih lanjut, beberapa arahan disarankan:

- Mengombinasikan analisis asosiasi dengan segmentasi pelanggan menggunakan teknik *clustering* atau profil pelanggan, sehingga strategi pemasaran tidak hanya bersifat umum tetapi juga dapat dipersonalisasi sesuai preferensi kelompok konsumen tertentu.
- Membandingkan performa algoritma *Apriori* dengan metode lain, seperti *FP-Growth*, untuk menilai efisiensi waktu komputasi, skala data, dan kualitas aturan yang dihasilkan, sehingga dapat dipilih metode yang paling optimal untuk dataset besar atau dinamis.
- Menguji model pada data transaksi ritel lokal agar hasil penelitian lebih kontekstual dan aplikatif terhadap pasar Indonesia, memperhatikan perbedaan pola belanja, budaya konsumen, dan tren musiman yang mungkin berbeda dari dataset internasional.

Dengan demikian, penelitian ini tidak hanya memberikan kontribusi pada pemahaman pola pembelian konsumen, tetapi juga menjadi landasan bagi strategi pemasaran berbasis data yang lebih efektif, inovatif, dan terukur di masa mendatang.

DAFTAR PUSTAKA

- [1] K. Brighton And S. Hariyanto, “Penerapan Metode Market Basket Analisis Dengan Algoritma Apriori Pada Toko Ritel Elektronik,” *Bit-Tech*, Vol. 7, No. 1, Pp. 37–46, 2024.
- [2] M. R. Q. Baihaqie, “Analisis Perilaku Konsumen Pada Usaha Ritel Dengan Menggunakan Metode Association Rule-Market Basket Analysis Dan Clustering Sebagai Usulan Strategi Peningkatan Penjualan (Studi Kasus: Intimart Gedongan),” 2023. Accessed: Jul. 20, 2025. [Online]. Available: <https://Dspace.Uii.Ac.Id/Handle/123456789/Dspace.Uii.Ac.Id/123456789/46872>
- [3] 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, Doi: 10.46984/Sebatik.V26i1.1622.
- [4] N. Barkah, E. Sutinah, And N. Agustina, “Metode Asosiasi Data Mining Untuk Analisa Persediaan Fiber Optik Menggunakan Algoritma Apriori,” *Jurnal Kajian Ilmiah*, Vol. 20, No. 3, Pp. 237–248, 2020, Doi: <https://doi.org/10.31599/Jki.V20i3.288>.
- [5] M. Rajagukguk, R. Dewi, E. Irawan, J. T. Hardinata, And I. S. Damanik, “Implementasi Association Rule Mining Untuk Menentukan Pola Kombinasi Makanan Dengan Algoritma Apriori,” *Jurnal Fasikom*, Vol. 10, No. 3, Pp. 248–254, 2020, Doi: [Ttps://Doi.Org/10.37859/Jf.V10i3.2308](https://doi.org/10.37859/Jf.V10i3.2308).
- [6] A. M. Alinafiah, B. C. Octariadi, And S. Sucipto, “Implementasi Impementasi Data Mining Dalam Pengelolaan Stok Obat Menggunakan Metode K-Means Clustering Dan Asosiation Rules Apriori: Analisis Pola Pembelian Dan Hubungan Antar Obat Dalam Pengelolaan Stok Menggunakan K-Means Clustering Dan Association Rules Apriori,” *Jurnal Informatika Polinema*, Vol. 10, No. 4, Pp. 551–558, 2024, Doi: <https://doi.org/10.33795/Jip.V10i4.5523>.
- [7] N. F. Matondang, H. Jaya, And A. Azanuddin, “Penerapan Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Pola Penjualan Barang Elektronik,” *Jurnal Saintikom (Jurnal Sains Manajemen Informatika Dan Komputer)*, Vol. 21, No. 2, Pp. 102–114, 2022, Doi: <https://doi.org/10.53513/Jis.V21i2.5988>.
- [8] D. Dwiputra, A. Mulyo Widodo, H. Akbar, And G. Firmansyah, “Evaluating The Performance Of Association Rules In Apriori And Fp-Growth Algorithms: Market Basket Analysis To Discover Rules Of Item Combinations,” *Journal Of World Science*, Vol. 2, No. 8, Pp. 1229–1248, Aug. 2023, Doi: 10.58344/Jws.V2i8.403.
- [9] M. Fathurrahman, A. R. Pratama, And T. Al-Mudzakir, “Perbandingan Algoritma Apriori Dan Fp Growth Terhadap Market Basket Analysis Pada Data Penjualan Bakery,” *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer Dan Manajemen)*, Vol. 4, No. 2, Pp. 266–274, 2023, Doi: <https://doi.org/10.30645/Kesatria.V4i2.161>.
- [10] M. A. Fathurrohman, “Penentuan Strategi Pengelolaan Coffee Shop Di Yogyakarta Dengan Mengidentifikasi Perilaku Dan Karakteristik Konsumen Menggunakan Metode Association Rules Dan Clustering (Studi Kasus Pada Mahasiswa Yogyakarta),” Universitas Islam Indonesia, 2022. Accessed: Jul. 20, 2025. [Online]. Available: <https://Dspace.Uii.Ac.Id/Handle/123456789/40882>
- [11] N. F. Ulfha And R. Amin, “Implementasi Data Mining Untuk Mengetahui Pola Pembelian Obat Menggunakan Algoritma Apriori,” *Komputasi: Jurnal Ilmiah Ilmu Komputer Dan Matematika*, Vol. 17, No. 2, Pp. 396–402, 2020.
- [12] I. Maryani, O. Revianti, H. M. Nur, And S. Sunanto, “Implementasi Data Mining Pada Penjualan Di Toko Goc Kosmetik Dengan Menggunakan Metode Algoritma Apriori,” *Indonesian Journal On Software Engineering (Ijse)*, Vol. 8, No. 1, Pp. 92–98, 2022, Doi: <https://doi.org/10.31294/Ijse.V8i1.13017>.
- [13] W. Sahara, S. D. Saragih, And A. P. Windarto, “Teknik Asosiasi Data Mining Dalam Menentukan Pola Penjualan Dengan Metode Apriori,” *Tin Terap. Inform. Nusantara*, Vol. 2, No. 12, Pp. 684–689, 2022, Doi: <https://doi.org/10.47065/Tin.V2i12.1577>.
- [14] R. Dzulkarnaen, “Perancangan Aplikasi Data Mining Market Basket Analysis Pada Apotek Permata Dengan Metode Hybrid-Dimension Association Rules,” *Journal Of Information Technology*, Vol. 2, No. 2, Pp. 67–72, 2020.
- [15] S. Raschka, J. Patterson, And C. Nolet, “Machine Learning In Python: Main Developments And Technology Trends In Data Science, Machine Learning, And Artificial Intelligence,” Apr. 01, 2020, *Mdpi Ag*. Doi: 10.3390/Info11040193.
- [16] D. Rizaldi And A. Adnan, “Market Basket Analysis Menggunakan Algoritma Apriori: Kasus Transaksi 212 Mart Soebrantas Pekanbaru,” *Jurnal Statistika Dan Aplikasinya*, Vol. 5, No. 1, Pp. 31–40, 2021.
- [17] H. Xie, “Research And Case Analysis Of Apriori Algorithm Based On Mining Frequent Item-Sets,” *Open J Soc Sci*, Vol. 9, No. 04, P. 458, 2021.
- [18] X. Zhang And J. Zhang, “Analysis And Research On Library User Behavior Based On Apriori Algorithm,” *Measurement: Sensors*, Vol. 27, P. 100802, 2023.