

# Evaluasi Kinerja Algoritma Apriori dan FP-Growth untuk Association Rule Mining pada Data Transaksi Ritel

Performance Evaluation of Apriori and FP-Growth Algorithms for Association Rule Mining on  
Retail Transaction Data

Fanny Soewignyo<sup>1</sup>, Tonny Irianto Soewignyo<sup>2</sup>, Wilsen Grivin Mokodaser<sup>\*3</sup>, Argha Orion  
Silitonga<sup>4</sup>

<sup>1,2</sup>Fakultas Ekonomi dan Bisnis, Universitas Klabat, Airmadidi, Sulawesi Utara

<sup>3,4</sup>Fakultas Ilmu Komputer, Universitas Klabat, Airmadidi, Sulawesi Utara

E-mail : f.soewignyo@unklab.ac.id<sup>1</sup>, tonnysoewignyo@unklab.ac.id<sup>2</sup>, wilsenm@unklab.ac.id<sup>\*3</sup>,  
arghal@unklab.ac.id<sup>4</sup>

\*Corresponding author

Received 20 October 2025; Revised 30 October 2025; Accepted 7 November 2025

**Abstrak** - Ledakan data transaksi ritel yang tererekam melalui sistem *Point of Sale* (POS) dan platform daring menuntut metode analisis yang efektif untuk menggali pola pembelian konsumen. Association Rule Mining merupakan pendekatan populer untuk menemukan keterkaitan antarproduk, dengan algoritma Apriori dan FP-Growth sebagai dua metode yang paling banyak digunakan. Penelitian ini bertujuan memberikan gambaran empiris mengenai efektivitas kedua algoritma tersebut pada data transaksi ritel yang nyata. Metode yang digunakan meliputi tahapan *data understanding* untuk mengenali struktur data, *data cleaning* untuk menghapus nilai kosong dan menyeragamkan format, serta data transformation menggunakan *TransactionEncoder* untuk mengubah data mentah menjadi format biner (*one-hot encoded*). Selanjutnya algoritma Apriori dan FP-Growth dijalankan dengan parameter yang sama untuk menghasilkan frequent itemsets dan aturan asosiasi. Evaluasi kinerja dilakukan dengan mengukur waktu pemrosesan, jumlah aturan yang dihasilkan, serta nilai *support*, *confidence*, dan *lift* tertinggi. Hasil penelitian menunjukkan bahwa kedua algoritma menghasilkan jumlah aturan yang sama (63 aturan) dengan *support* tertinggi 0,06, *confidence* tertinggi 0,51, dan *lift* tertinggi 3,29, tetapi waktu pemrosesan berbeda signifikan (Apriori 0,39 detik, FP-Growth 6,95 detik). Kesimpulannya, *association rule mining* efektif mengungkap pola pembelian konsumen, dan algoritma Apriori lebih efisien untuk dataset kecil hingga menengah, sedangkan FP-Growth lebih sesuai untuk dataset yang jauh lebih besar.

**Keywords** - Association Rules, Apriori, FP-Growth, Frequent Itemset, Transaksi Ritel.

**Abstract** - The surge in retail transaction data captured through Point of Sale (POS) systems and online platforms demands effective analytical methods to uncover consumer purchasing patterns. Association Rule Mining is a popular approach for identifying relationships between products, with Apriori and FP-Growth being two of the most widely used algorithms. This study aims to provide an empirical overview of the effectiveness of both algorithms on real-world retail transaction data. The methodology includes several stages: data understanding to recognize the data structure, data cleaning to remove missing values and standardize formats, and data transformation using TransactionEncoder to convert raw data into a binary (*one-hot encoded*) format. Subsequently, the Apriori and FP-Growth algorithms are executed with identical parameters to generate frequent itemsets and association rules. Performance evaluation is conducted by measuring processing time, the number of rules generated, and the highest values of support, confidence, and lift. The results show that both algorithms produced the same number of rules (63 rules), with the highest support at 0.06, confidence at 0.51, and lift at 3.29. However, processing times differed significantly: Apriori took 0.39 seconds, while FP-Growth required 6.95 seconds. In conclusion, association rule mining effectively reveals consumer purchasing

patterns. The Apriori algorithm is more efficient for small to medium-sized datasets, whereas FP-Growth is better suited for significantly larger datasets.

**Keywords** - Association Rules, Apriori, FP-Growth, Frequent Itemset, retail transaction.

## 1. PENDAHULUAN

Perkembangan teknologi informasi dan transformasi digital telah mengubah wajah industri ritel secara fundamental.[1] Saat ini hampir setiap transaksi penjualan terekam secara elektronik melalui sistem *Point of Sale* (POS), *platform e-commerce*, dan aplikasi loyalti pelanggan, sehingga organisasi ritel mengakumulasi volume data transaksi yang sangat besar dan terus bertambah.[2] Data tersebut tidak hanya berisi catatan jumlah dan nilai penjualan, tetapi juga informasi tentang kombinasi produk yang dibeli dalam satu transaksi, waktu pembelian, dan identitas pelanggan (jika tersedia). Ketersediaan data transaksi yang terstruktur ini membuka peluang besar untuk analisis kuantitatif yang dapat memberikan wawasan operasional dan strategis bagi manajemen ritel.

Pertumbuhan pesat industri ritel di Indonesia, seperti yang diuraikan dalam jurnal analisis perkembangan bisnis ritel di Indonesia, menunjukkan bahwa semakin banyak transaksi ritel terekam secara digital melalui sistem Point of Sale (POS) dan *platform* daring. Kondisi ini menciptakan basis data transaksi yang sangat besar dan kaya informasi tentang perilaku belanja konsumen. Di satu sisi fenomena ini membuka peluang untuk meningkatkan layanan, promosi, dan tata kelola persediaan, tetapi di sisi lain menuntut metode analisis yang lebih canggih agar pola pembelian yang tersembunyi dapat terungkap secara efektif.[3]

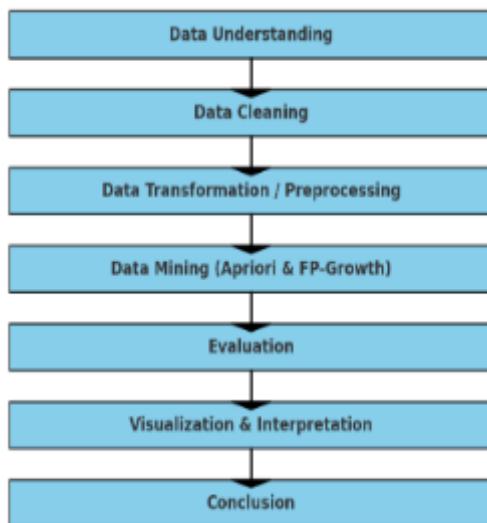
Bagi industri ritel yang kompetitif saat ini, POS tidak hanya menjadi alat pencatat transaksi, tetapi juga menjadi sumber utama data yang mendukung *business intelligence*. Inilah mengapa POS merupakan titik awal yang penting dalam penelitian terkait association rule mining dan market basket analysis, karena dari sinilah data pembelian konsumen diperoleh untuk dianalisis lebih lanjut. Dalam praktiknya, POS modern sering terhubung langsung dengan sistem manajemen persediaan (*inventory management*), sistem CRM (*customer relationship management*), dan *platform* analitik. Dengan demikian, setiap transaksi yang masuk tidak hanya memperbarui stok secara otomatis, tetapi juga memperkaya basis data pelanggan. Data ini kemudian dapat dianalisis untuk berbagai tujuan seperti mengidentifikasi pola pembelian konsumen, mengoptimalkan stok barang, merancang promosi yang lebih efektif, hingga menghasilkan laporan kinerja toko secara instan.[4] Menggali pola pembelian konsumen dari kumpulan data transaksi menjadi aspek krusial dalam pengelolaan bisnis ritel modern karena langsung berdampak pada keputusan bisnis yang bersifat taktis dan strategis.[5] Dengan memahami produk-produk yang sering dibeli bersamaan, manajer pemasaran dapat merancang kampanye promosi yang lebih tepat sasaran dan paket bundling yang meningkatkan penjualan silang (*cross-selling*). Tim merchandising dapat mengoptimalkan penempatan barang di rak dan layout toko untuk memaksimalkan impulse purchase, sementara tim rekomendasi di *platform* daring dapat meningkatkan relevansi saran produk sehingga menaikkan konversi dan retensi pelanggan. Selain itu, wawasan pola pembelian membantu pengendalian persediaan, perencanaan logistik, dan penentuan kebijakan harga sehingga keputusan menjadi lebih berbasis data (*data-driven*) dan berpotensi meningkatkan efisiensi serta profitabilitas.[6]. Dalam konteks teknis, *Market Basket Analysis* (MBA) dan *Association Rule Mining* (ARM) merupakan pendekatan utama untuk menemukan keterkaitan antar produk pada level transaksi.[7] MBA mengekstrak *frequent itemsets* kombinasi produk yang sering muncul bersama dan menerjemahkannya menjadi aturan asosiatif berbentuk "Jika A maka B" yang dapat dievaluasi menggunakan metrik seperti support (seberapa sering kombinasi muncul), confidence (probabilitas konsekuensi muncul bila antecedent hadir), dan lift (seberapa kuat hubungan dibandingkan kemunculan acak).[8] Algoritma klasik seperti Apriori dan FP-Growth adalah metode populer untuk mengekstraksi *frequent*

itemsets;[9] Apriori bekerja dengan pendekatan kandidat-generasi sedangkan FP-Growth membangun struktur FP-Tree untuk efisiensi pencarian.[10][11] Hasil dari teknik-teknik ini memberikan dasar empiris bagi rekomendasi produk, perancangan promosi bersama, dan penataan produk membuktikan peranan sentral MBA/ARM dalam memanfaatkan data transaksi ritel untuk pengambilan keputusan bisnis yang lebih baik[12]. Penelitian terkait telah dilakukan menggunakan algoritma apriori untuk rekomendasi paket produk Hasil penelitian menunjukkan bahwa nilai parameter minimum dukungan dan minimum kepercayaan meningkat secara signifikan. Jumlah rekomendasi yang dibuat berkurang sebesar 80%, tetapi kualitas asosiasi yang terbentuk meningkat. Nilai rata rata dukungan meningkat sebesar 35% dan nilai rata rata dukungan meningkat sebesar 32%. Selain itu, jumlah rekomendasi yang dihasilkan dipengaruhi oleh penambahan parameter kategori pada algoritma Apriori. Dimana empat set dengan kategori Lampu, Kabel, Fitting, dan Stop Kontak digabungkan menghasilkan satu rekomendasi yang cocok dengan dukungan rata-rata 0.056 dan keyakinan rata-rata 0.529[13]. Penelitian selanjutnya tentang menemukan pola asosiasi pada data retail fashion dengan dataset dengan 3.400 transaksi pelanggan dari platform Kaggle diolah menggunakan RapidMiner dengan parameter minimum support 0,1 dan confidence 0,6. Preprocessing data, normalisasi, transformasi one-hot encoding, dan pengujian dengan operator W-Apriori adalah semua teknik yang digunakan. Hasilnya menunjukkan pola signifikan, seperti bahwa pembelian tas dan loafers sangat terkait dengan raincoat (keyakinan 74%). Algoritma Apriori dapat digunakan untuk melakukan promosi, membuat rekomendasi produk, dan mengatur layout toko. Ini juga terbukti efektif dalam mengidentifikasi pola kebiasaan pembelian pelanggan. Perilaku konsumen saat berbelanja ditunjukkan oleh aturan asosiasi yang dikumpulkan, yang menunjukkan hubungan erat antara beberapa produk. Untuk membuat keputusan strategis yang berorientasi data, data ini dapat digunakan untuk membuat fitur rekomendasi yang lebih baik, menata produk yang lebih efisien, dan membuat promosi yang relevan dan tepat sasaran[14]. Penggunaan algoritma FP-Growth untuk mengoptimalkan penempatan produk untuk menarik perhatian pelanggan dan memaksimalkan ruang yang tersedia adalah tantangan bagi Prima Dewata. Dimulai dengan pengumpulan data, pra-pemrosesan, penggunaan Algoritma FP-Growth, dan analisis hasilnya. Menurut hasil pemrosesan data, teh Sariwangi adalah produk yang paling banyak dibeli oleh pengunjung Prima Dewata dari Januari hingga November 2024. Ada empat aturan utama yang dapat digunakan sebagai acuan dalam menyusun barang: 1. Teh Sariwangi dan Sampo Clear Men (dukungan: 0,682 dan keyakinan: 0,867); 2. Teh Sariwangi dan Pasta Gigi Indomie Pepsodent (dukungan: 0,171 dan keyakinan: 0,876); dan 3. Teh Sariwangi dan Indom (dukungan: 0,172) dan keyakinan: 0,876)[15]. Penggunaan algoritma Fp-Growth selanjutnya didapatkan bahwa dengan jumlah dataset yang digunakan mencakup 508 penjualan, masing-masing dengan nama produk, jumlah, harga satuan, dan pendapatan total. Hasil penelitian menunjukkan bahwa produk tertentu, seperti Milo Milk dan Milo Ori, memiliki pola pembelian yang kuat dengan keyakinan sebesar 70%. Pola ini memberikan wawasan untuk strategi pemasaran, seperti menerapkan promosi bundling atau diskon untuk produk terkait. Selain itu, penelitian ini menemukan bahwa algoritma FP-Growth lebih baik dalam menganalisis data transaksi daripada metode konvensional. Kesimpulannya, algoritma FP-Growth dapat membantu bisnis menemukan pola pembelian pelanggan, meningkatkan efektivitas pemasaran, dan mengoptimalkan manajemen stok. Rekomendasi untuk penelitian lanjutan adalah untuk memperluas karakteristik dataset dan mendapatkan hasil yang lebih lengkap dengan menggunakan metode tambahan seperti algoritma Apriori[16].

Oleh karena itu bedasarkan pembahasan latar belakang dan penelitian terkait diatas, maka penelitian tentang *Performance Evaluation* dari Apriori and FP-Growth algoritma untuk penambangan aturan asosiasi pada data transaksi ritel menjadi penting dilakukan untuk memberikan gambaran empiris mengenai efektivitas kedua algoritma pada data transaksi ritel yang nyata. Hasil evaluasi ini tidak hanya memberikan kontribusi akademis berupa pembandingan sistematis antara dua algoritma *Association Rule Mining*, tetapi juga memberikan panduan praktis bagi pengelola bisnis ritel dalam memilih algoritma yang paling tepat untuk

menganalisis pola pembelian pelanggan di tengah pertumbuhan data transaksi yang masif. Dengan demikian, penelitian ini berperan strategis dalam mendukung pengambilan keputusan berbasis data (*data-driven decision making*) di sektor ritel modern.

## 2. METODE PENELITIAN



Gambar 1. Tahapan penelitian

### 1. Data Understanding

Tahapan awal penelitian ini dimulai dengan memahami data transaksi ritel yang akan dianalisis. Pada tahap ini peneliti memuat dataset transaksi dari sumber yang tersedia dan mengamati struktur serta isi data, seperti jumlah baris, kolom, nama atribut, dan jenis data yang dikandungnya. Pemahaman ini penting untuk mengetahui konteks setiap kolom, misalnya bahwa satu baris merepresentasikan pembelian satu item pada satu transaksi, sehingga peneliti dapat merancang langkah pengolahan data yang tepat pada tahapan selanjutnya.

### 2. Data Cleaning

Setelah memahami karakteristik data, langkah berikutnya adalah melakukan pembersihan data untuk memastikan data siap dianalisis. Proses pembersihan yang dilakukan meliputi penghapusan baris yang memiliki nilai kosong (NaN).<sup>[17]</sup> penyeragaman format teks untuk nama produk, dan penghapusan spasi berlebih agar tidak terjadi penggandaan item akibat perbedaan penulisan. Pada tahap ini data yang dihasilkan menjadi lebih konsisten dan akurat, sehingga risiko kesalahan analisis akibat kualitas data yang buruk dapat diminimalkan.

### 3. Data Transformation / Preprocessing

Data transaksi yang telah bersih kemudian diubah ke dalam format yang sesuai dengan kebutuhan algoritma. Peneliti mengelompokkan item berdasarkan nomor transaksi (InvoiceNo) sehingga setiap transaksi direpresentasikan sebagai satu list berisi produk-produk yang dibeli bersama. Selanjutnya list transaksi tersebut diubah menjadi bentuk matriks biner (one-hot encoded) menggunakan TransactionEncoder, di mana setiap kolom merepresentasikan satu produk dan setiap baris menunjukkan transaksi tertentu dengan nilai 1 jika produk muncul dan 0 jika tidak.<sup>[18]</sup> Transformasi ini merupakan prasyarat agar algoritma Apriori dan FP-Growth dapat dijalankan.

#### 4. Data Mining (Apriori and FP-Growth)

Tahap inti penelitian adalah penerapan algoritma data mining untuk menemukan pola pembelian konsumen. Peneliti menjalankan algoritma Apriori dengan parameter minimum support tertentu untuk mendapatkan frequent itemsets beserta aturan asosiatifnya.[19] Algoritma FP-Growth dijalankan dengan parameter yang sama untuk memastikan hasilnya dapat dibandingkan secara adil. Dari kedua algoritma ini diperoleh sekumpulan frequent itemsets dan aturan asosiasi yang masing-masing dilengkapi metrik support, confidence, dan lift yang menggambarkan kekuatan hubungan antar produk[20]. Dalam penggunaan algoritma FP-Growth maka akan digunakan rumus sebagai berikut:

Untuk penggunaan support dimana Mengukur seberapa sering kombinasi item tertentu muncul dalam database, yang dinyatakan dalam bentuk persentase yang besar suatu itemset dalam suatu proses. Seperti pada persamaan berikut:

$$\text{Support } (A) = \frac{\text{Jumlah Transaksi Mengandung } A}{\text{Total Transaksi}}$$
$$\text{Support } (A \cap B) = \frac{\text{Jumlah Transaksi Mengandung } A \text{ dan } B}{\text{Total Transaksi}}$$

Sementara confidance dimana nilai kepastian untuk memperlihatkan relasi, antara dua item – item yang terdapat dalam kondisi asosiasi. Secara umum dapat dituliskan dalam persamaan berikut

$$\text{Confidence } (A \rightarrow B) = P(A|B) = \frac{\text{Jumlah Transaksi Mengandung } A \text{ dan } B}{\text{Jumlah Transaksi Mengandung } A}$$

Ukuran di atas nantinya akan Support dan confidence berguna dalam menentukan association rules yang menarik dengan membandingkannya dengan batasan yang telah ditentukan. Batasan tersebut biasanya mencakup min\_support dan min\_confidence.

Sementara ini dalam penerapan algoritma apriori akan menggunakan rumus perhitungan support dan confidence seperti FP-Growth dan tambahan lift rasio diterapkan pada aturan asosiasi yang sudah diperoleh. Tujuan perhitungan ini adalah untuk mengevaluasi validitas aturan asosiasi, yang berarti bahwa aturan asosiasi berlaku jika nilai lift rasio lebih besar dari satu seperti persamaan dibawah ini:

$$\text{Lift Rasio } (A, B) = \frac{\text{Confidence } (A, B)}{\text{Support } (B)}$$

#### 5. Evaluation

Setelah algoritma dijalankan, peneliti melakukan evaluasi terhadap hasil yang diperoleh. Evaluasi dilakukan dengan mengukur waktu komputasi masing-masing algoritma, menghitung jumlah aturan yang dihasilkan, dan mengekstrak nilai support, confidence, dan lift tertinggi.[13] Hasil evaluasi tersebut disajikan dalam bentuk tabel untuk memudahkan pembandingan kinerja kedua algoritma, sehingga dapat diketahui algoritma mana yang lebih cepat dan bagaimana kualitas aturan yang dihasilkan pada dataset transaksi ritel yang sama.

#### 6. Visualization and Interpretation

Untuk memperjelas hasil yang diperoleh, peneliti membuat visualisasi dalam bentuk grafik. Visualisasi yang digunakan antara lain grafik batang untuk menampilkan sepuluh itemset dengan nilai support tertinggi, serta scatter plot yang menggambarkan hubungan antara confidence dan lift pada aturan yang dihasilkan dengan ukuran titik sesuai besarnya support.[21] Visualisasi ini membantu pembaca untuk lebih mudah memahami pola pembelian yang ditemukan dan membandingkan karakteristik hasil dari kedua algoritma secara visual.

#### 7. Conclusion

Tahapan terakhir adalah menyusun kesimpulan dari seluruh rangkaian analisis. Pada tahap ini peneliti menuliskan temuan utama terkait algoritma mana yang lebih efisien dan

seberapa baik kualitas aturan yang dihasilkan. Kesimpulan juga mencakup implikasi praktis bagi pengelola bisnis ritel, misalnya rekomendasi pemilihan algoritma untuk kondisi data tertentu, serta masukan untuk penelitian lanjutan agar metode analisis yang digunakan semakin optimal di masa depan.

### 3. HASIL DAN PEMBAHASAN

Setelah melewati setiap tahapan penelitian yang telah ditentukan maka diperoleh hasil sebagai berikut:

#### 1. Data Understanding

	shrimp	almonds	avocado	vegetables	green	whole	west	yams	cottage	energy	tomato
				mix	grapes	flour			cheese	drink	juice
0	burgers	meatballs		eggs		NaN	NaN	NaN	NaN	NaN	NaN
1	chutney		NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	turkey	avocado		NaN		NaN	NaN	NaN	NaN	NaN	NaN
3	mineral		milk	energy	whole	green	NaN	NaN	NaN	NaN	NaN
4	water			bar	wheat	tea	NaN	NaN	NaN	NaN	NaN
	low fat						NaN	NaN	NaN	NaN	NaN
	yogurt							NaN	NaN	NaN	NaN

Gambar 2. Struktur data

Pada penelitian ini menggunakan daftar produk belanja pertransaksi yang berjumlah 7500 baris dan 20 kolom. Pada tahap Data Understanding, gambar 2 dilakukan memuat dan memeriksa dataset transaksi ritel yang digunakan sebagai sumber analisis. Data yang diperoleh berasal dari pencatatan sistem penjualan (Point of Sale) dan memuat sejumlah atribut yang menggambarkan transaksi, misalnya nomor faktur atau transaksi (InvoiceNo), deskripsi barang (ItemDescription), kuantitas, dan tanggal transaksi. Dengan melakukan eksplorasi awal menggunakan fungsi-fungsi deskriptif (misalnya *head()*, *info()*, dan *shape()* pada Python), diketahui struktur dasar dataset, jumlah baris dan kolom, serta jenis data pada masing-masing atribut. Pemeriksaan ini juga mengungkap adanya potensi data kosong, inkonsistensi penulisan nama produk, dan duplikasi yang harus ditangani pada tahap berikutnya.

Hasil dari tahapan ini adalah pemahaman yang lebih jelas mengenai karakteristik data transaksi ritel yang digunakan, termasuk frekuensi transaksi, variasi produk yang dijual, dan pola awal item yang sering muncul. Pengetahuan awal ini menjadi dasar untuk menentukan parameter analisis seperti nilai minimum support dan minimum confidence yang akan digunakan pada algoritma Apriori dan FP-Growth. Dengan demikian, tahap Data Understanding memastikan bahwa peneliti memiliki gambaran yang tepat mengenai data sebelum melakukan pembersihan, transformasi, dan penerapan algoritma association rule mining.

#### 2. Data Cleaning

	shrimp	almonds	avocado	vegetables	green
				mix	grapes
0	burgers	meatballs		eggs	
1	chutney				
2	turkey	avocado			
3	mineral		milk	energy	whole
	water			bar	wheat
					rice
					green
					tea
4	low fat				
	yogurt				

Gambar 3. Hasil Data Cleaning

Pada tahap Data Cleaning, dataset transaksi ritel yang telah dipahami sebelumnya dibersihkan untuk memastikan kualitasnya sebelum dilakukan analisis lebih lanjut. Proses pembersihan yang dilakukan meliputi penghapusan baris yang memiliki nilai kosong (*missing values*), penghapusan spasi berlebih serta penyeragaman format teks pada nama produk agar tidak terjadi duplikasi item akibat perbedaan penulisan, serta penghapusan data duplikat jika ditemukan transaksi ganda. Tahapan ini menghasilkan jumlah data yang akan diproses sebanyak 7500 baris dan 19 kolom.

Gambar 3 diatas menampilkan contoh data transaksi dalam bentuk mentah sebelum dilakukan proses encoding. Setiap baris merepresentasikan satu transaksi, di mana isinya adalah daftar item yang dibeli bersama dalam transaksi tersebut. Misalnya, transaksi pertama berisi *burgers*, *meatballs*, dan *eggs*, transaksi kedua hanya berisi *chutney*, sedangkan transaksi ketiga berisi *turkey* dan *avocado*. Pada bagian atas tabel terlihat daftar item unik seperti *shrimp*, *almonds*, *avocado*, *vegetables mix*, dan *green grapes*, yang menunjukkan kumpulan produk yang ada di seluruh dataset. Namun, isi tabel masih berupa nama barang dalam bentuk teks, bukan nilai boolean. Data mentah seperti ini nantinya akan diproses menggunakan *TransactionEncoder()* agar diubah menjadi bentuk one-hot encoded (bernilai *True* atau *False*), sehingga bisa digunakan oleh algoritma *association rule mining* seperti Apriori atau FP-Growth untuk menemukan pola keterkaitan antar item dalam transaksi.

### 3. Data Transformation / Preprocessing

	almonds	antioxydant juice	asparagus	avocado	babies food	bacon	barbecue sauce	black tea	blueberries	body spray	...
0	False	False	False	False	False	False	False	False	False	False	...
1	False	False	False	False	False	False	False	False	False	False	...
2	False	False	False	True	False	False	False	False	False	False	...
3	False	False	False	False	False	False	False	False	False	False	...
4	False	False	False	False	False	False	False	False	False	False	...

5 rows x 119 columns

Gambar 4. Tabel hasil Transaction Encoder

Pada tahap Data Transformation / Preprocessing, dataset transaksi ritel yang sudah bersih diubah ke dalam format yang sesuai untuk dianalisis menggunakan algoritma Apriori dan FP-Growth. Proses ini dimulai dengan mengelompokkan item berdasarkan nomor transaksi (*InvoiceNo*) sehingga setiap transaksi direpresentasikan sebagai sebuah daftar (*list*) produk yang dibeli bersamaan. Selanjutnya, daftar transaksi ini diubah menjadi bentuk matriks biner (*one-hot encoded*) menggunakan *TransactionEncoder*, di mana setiap kolom mewakili satu produk dan setiap baris mewakili satu transaksi, dengan nilai 1 jika produk muncul dalam transaksi tersebut dan 0 jika tidak.

Gambar 4 merupakan hasil dari *TransactionEncoder()* pada association rule mining merepresentasikan data transaksi dalam bentuk one-hot encoded. Setiap baris pada tabel menunjukkan satu transaksi, sedangkan setiap kolom mewakili satu item unik yang muncul di seluruh transaksi. Nilai yang terisi pada tabel berupa True atau False, di mana True berarti item tersebut terdapat dalam transaksi tertentu, dan False berarti item tersebut tidak ada dalam transaksi tersebut. Dengan cara ini, data yang semula berbentuk daftar item dalam list dapat diubah menjadi format numerik biner yang sesuai untuk diproses lebih lanjut oleh algoritma seperti Apriori atau FP-Growth. Misalnya, dari data awal berupa daftar `['avocado', 'yogurt cake'], ['bacon', 'avocado'], ['almonds', 'blueberries', 'zucchini']`, hasil encoding akan menghasilkan tabel dengan kolom *almonds*, *avocado*, *bacon*, *blueberries*, dan *zucchini*, serta menandai transaksi mana saja yang mengandung item tersebut. Jika jumlah kolom pada tabel mencapai 119, artinya terdapat 119 item unik dalam keseluruhan transaksi. Bentuk representasi ini kemudian menjadi dasar untuk mencari *frequent itemsets* maupun membangun *association*

rules yang menggambarkan hubungan antar item dalam data transaksi.

#### 4. Data Mining (Apriori and FP-Growth)

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	(burgers)	(eggs)	0.087200	0.179733	0.028800	0.330275	1.837585
1	(turkey)	(eggs)	0.062533	0.179733	0.019467	0.311301	1.732014
2	(turkey)	(mineral water)	0.062533	0.238267	0.019200	0.307036	1.288624
3	(avocado)	(mineral water)	0.033200	0.238267	0.011467	0.345382	1.449559
4	(milk)	(mineral water)	0.129600	0.238267	0.048000	0.370370	1.554436

Gambar 5. Hasil penerapan algoritma FP-Growth

Pada gambar 5 menunjukkan setiap baris menunjukkan satu aturan asosiasi yang ditemukan oleh FP-Growth. Kolom *antecedents* berisi item yang menjadi premis aturan, sedangkan *consequents* adalah item yang diprediksi muncul bersamaan. Nilai *support* menunjukkan seberapa besar proporsi transaksi yang memuat kombinasi item tersebut, sedangkan *confidence* menunjukkan probabilitas konsumen membeli *consequents* jika sudah membeli *antecedents*. Selain itu, nilai *lift* digunakan untuk menilai kekuatan hubungan antar item: nilai lebih besar dari 1 berarti ada hubungan positif, atau kedua item cenderung dibeli bersama lebih sering daripada secara acak. Contohnya, aturan (*milk*) dengan (*mineral water*) memiliki *confidence* sekitar 37% dengan *lift* 1,55, yang artinya pelanggan yang membeli susu lebih berpeluang juga membeli air mineral. Aturan lain seperti (*avocado*) → (*mineral water*) juga memiliki *lift* 1,44, sehingga menunjukkan kecenderungan konsumen membeli keduanya bersama.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	(avocado)	(mineral water)	0.033200	0.238267	0.011467	0.345382	1.449559
1	(burgers)	(eggs)	0.087200	0.179733	0.028800	0.330275	1.837585
2	(cake)	(mineral water)	0.081067	0.238267	0.027467	0.338816	1.422002
3	(cereals)	(mineral water)	0.025733	0.238267	0.010267	0.398964	1.674442
4	(chicken)	(mineral water)	0.060000	0.238267	0.028000	0.380000	1.594852

Gambar 6. Hasil penerapan algoritma Apriori

Hasil pada gambar 6 tersebut merupakan keluaran dari algoritma Apriori yang menampilkan aturan asosiasi antar item dalam data transaksi. Setiap baris menunjukkan sebuah aturan, di mana kolom *antecedents* berisi item awal dan kolom *consequents* berisi item yang cenderung dibeli bersamaan. Nilai *support* menunjukkan persentase transaksi yang mengandung kombinasi item tersebut, sedangkan *confidence* menggambarkan probabilitas kemunculan item konsekuensi jika item antecedent sudah ada. Selain itu, nilai *lift* mengukur kekuatan asosiasi dengan membandingkan kemungkinan kemunculan bersamaan dengan kemungkinan acak; nilai lebih dari satu menandakan adanya hubungan positif antar item. Sebagai contoh, aturan (*burgers*) → (*eggs*) memiliki *confidence* 33% dan *lift* 1,83, yang berarti pelanggan yang membeli burger memiliki peluang lebih besar untuk juga membeli telur dibandingkan dengan peluang acak. Demikian juga, aturan (*cake*) → (*mineral water*) memiliki *confidence* 33,9% dengan *lift* 1,42, sehingga ada kecenderungan konsumen yang membeli kue juga membeli air mineral. Informasi ini sangat bermanfaat dalam analisis keranjang belanja untuk membantu strategi pemasaran, seperti rekomendasi produk, penempatan barang di rak toko, atau penawaran bundling.

## 5. Evaluation

Tabel 1. Hasil Evaluasi *Performance*

Algoritma	Runtime (Second)	Jumlah Aturan	Support Tertinggi	Confidance Tertinggi	Lift Tertinggi
Apriori	0.39	63	0.06	0.51	3.29
FP-Growth	6.95	63	0.06	0.51	3.29

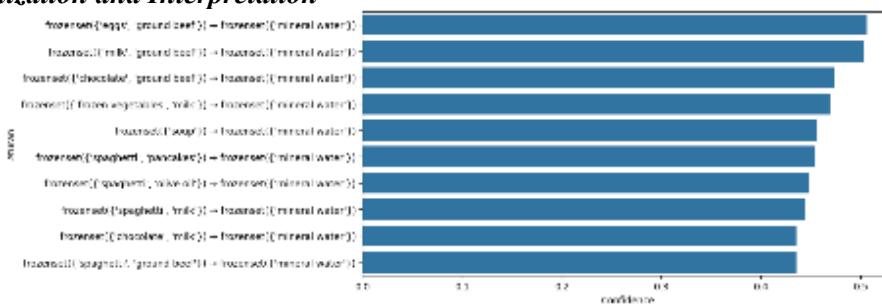
Tabel 1 di atas memperlihatkan hasil evaluasi kinerja dua algoritma *association rule mining*, yaitu Apriori dan FP-Growth, pada data transaksi ritel. Kolom *Runtime (Second)* menunjukkan lama waktu pemrosesan masing-masing algoritma untuk menghasilkan *frequent itemsets* dan aturan asosiasi. Pada percobaan ini, algoritma Apriori hanya membutuhkan sekitar 0,39 detik, sedangkan FP-Growth memerlukan 6,95 detik. Perbedaan ini menunjukkan bahwa pada dataset ini Apriori lebih cepat dibandingkan FP-Growth, walaupun secara teoritis FP-Growth dirancang untuk lebih efisien pada dataset yang jauh lebih besar.

Kolom *Jumlah Aturan* menunjukkan total aturan asosiasi yang dihasilkan. Kedua algoritma menghasilkan jumlah aturan yang sama, yaitu 63 aturan, karena parameter *minimum support* dan *minimum confidence* yang digunakan identik. Kolom *Support Tertinggi* menunjukkan proporsi transaksi terbesar yang memuat kombinasi item tertentu. Nilainya 0,06 atau sekitar 6% transaksi, yang berarti itemset paling sering muncul terdapat pada sekitar enam persen dari seluruh transaksi.

Kolom *Confidence* tertinggi menunjukkan probabilitas tertinggi kemunculan konsekuensi jika antecedent muncul. Nilai 0,51 menandakan bahwa aturan terkuat memiliki peluang benar sekitar 51% ketika antecedent dipenuhi. Kolom *Lift Tertinggi* menggambarkan kekuatan hubungan antaritem dibandingkan kemunculan acak. Nilai 3,29 berarti bahwa kombinasi item pada aturan terkuat dibeli bersama 3,29 kali lebih sering daripada yang diharapkan secara acak, menunjukkan adanya asosiasi positif yang signifikan.

Dengan demikian, tabel ini menunjukkan bahwa meskipun kualitas aturan yang dihasilkan Apriori dan FP-Growth identik (jumlah aturan, support tertinggi, *confidence* tertinggi, dan lift tertinggi sama), waktu pemrosesannya berbeda. Hal ini penting untuk dipertimbangkan dalam memilih algoritma yang paling sesuai dengan karakteristik data transaksi yang dianalisis.

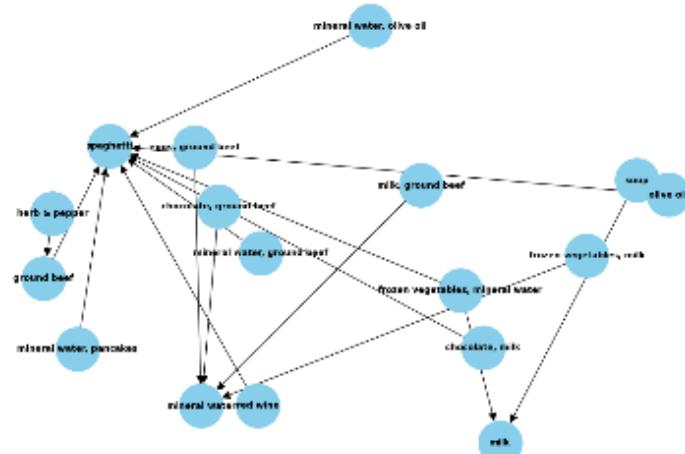
## 6. Visualization and Interpretation



Gambar 7 Visualisasi algoritma Apriori

Dari gambar 7 terlihat bahwa hampir semua aturan memiliki *consequents* yang sama, yaitu mineral water, artinya air mineral sering muncul bersamaan dengan berbagai produk lain dalam transaksi. Contohnya, aturan '{eggs', 'ground beef}' ke {'mineral water'} dan {'milk', 'ground beef'} ke {'mineral water'} memiliki *confidence* paling tinggi, sekitar 50%, yang menunjukkan bahwa separuh dari konsumen yang membeli kombinasi tersebut juga membeli air mineral. Aturan lain seperti {'chocolate', 'ground beef'} ke {'mineral water'} atau {'spaghetti', 'milk'} ke {'mineral water'} juga menunjukkan pola serupa dengan *confidence* yang sedikit lebih rendah. Grafik ini memperlihatkan bahwa mineral water adalah produk yang sangat sering dibeli

bersama produk lain, sehingga bisa dianggap sebagai item “pelengkap” dalam banyak transaksi. Informasi semacam ini berguna untuk strategi penjualan, misalnya menempatkan air mineral dekat dengan produk lain yang sering dibeli bersamanya atau membuat paket promosi bundling.



Gambar 8. Network Graph Aturan Asosiasi (FP-Growth)

Pada gambar 8 menunjukkan Setiap lingkaran biru (*node*) merepresentasikan kombinasi item (misalnya *milk*, *ground beef*, *spaghetti*, atau *frozen vegetables*, *mineral water*). Sementara itu, panah (*edge*) menunjukkan arah aturan asosiasi, artinya jika suatu item atau kombinasi item dibeli, maka kemungkinan besar item yang ditunjuk oleh panah juga akan dibeli bersamaan. Misalnya, aturan dari *eggs*, *ground beef* ke *spaghetti* menunjukkan bahwa konsumen yang membeli telur dan daging cincang cenderung juga membeli spaghetti. Pola serupa juga terlihat pada *frozen vegetables*, *milk* ke *soup* atau *chocolate*, *milk* ke *mineral water*.

Grafik ini memperlihatkan bahwa ada beberapa produk yang berperan sebagai hub utama (*node* dengan banyak panah masuk atau keluar), seperti *spaghetti*, *milk*, dan *mineral water*. Hal ini berarti produk-produk tersebut sering muncul dalam berbagai aturan assosiasi, sehingga bisa dianggap sebagai item yang populer atau sering dibeli bersama item lain. Sebagai contoh, *spaghetti* tidak hanya dipengaruhi oleh kombinasi telur dan daging cincang, tetapi juga muncul dalam hubungan dengan produk lain seperti minyak zaitun (*mineral water*, *olive oil*, *spaghetti*). Grafik ini membantu memahami pola belanja konsumen. Produk seperti *spaghetti*, *milk*, dan *mineral water* dapat dijadikan fokus strategi pemasaran, misalnya ditempatkan berdekatan dengan item yang sering berasosiasi dengannya atau dijadikan paket promosi untuk meningkatkan penjualan.

## 7. Analisa Hasil

Analisis dimulai dengan memahami struktur dan karakteristik dataset transaksi ritel yang terdiri dari 7.500 transaksi dengan berbagai atribut penjualan. Pemeriksaan awal menunjukkan bahwa data memiliki variasi produk yang cukup besar dan pola pembelian yang berpotensi membentuk keterkaitan item yang signifikan. Dari proses ini juga ditemukan adanya beberapa masalah kualitas data seperti nilai kosong, duplikasi, dan ketidakkonsistenan penulisan nama produk yang dapat memengaruhi hasil analisis jika tidak ditangani. Tahap pembersihan kemudian dilakukan untuk menghilangkan nilai hilang, menormalkan format teks, serta menghapus transaksi duplikat sehingga diperoleh data yang lebih bersih dan siap diolah. Hasil pembersihan memastikan bahwa tidak ada informasi yang bias atau redundan yang dapat mengganggu proses penemuan rule asosiasi.

Setelah data bersih diperoleh, data diubah ke dalam format yang dapat diproses algoritma association rule mining dengan mengelompokkan produk berdasarkan transaksi dan melakukan one-hot encoding. Hasil transformasi menghasilkan matriks biner dengan setiap kolom mewakili

item unik dan setiap baris mewakili transaksi, sehingga memungkinkan perhitungan frekuensi kemunculan kombinasi item secara efisien. Proses ini memastikan bahwa struktur data sesuai dengan kebutuhan algoritma Apriori dan FP-Growth untuk mengekstraksi pola pembelian konsumen. Tahap data mining menunjukkan bahwa kedua algoritma menghasilkan jumlah aturan asosiasi yang sama, yaitu 63 aturan, dengan nilai support, confidence, dan lift yang identik. Hal ini mengindikasikan bahwa parameter yang digunakan sudah tepat dan dataset memiliki pola yang cukup jelas. Namun demikian, Apriori memberikan waktu pemrosesan yang jauh lebih cepat dibanding FP-Growth pada dataset berukuran sedang ini, meskipun secara teoritis FP-Growth dirancang untuk lebih efisien pada dataset besar. Aturan yang ditemukan memperlihatkan bahwa *mineral water* merupakan produk yang paling sering muncul sebagai konsekuensi dalam berbagai kombinasi, menunjukkan bahwa produk tersebut berperan sebagai item pelengkap umum dalam transaksi konsumen. Selain itu, produk seperti susu, telur, ground beef, dan spaghetti tampak sebagai item yang sering memiliki hubungan asosiasi, mencerminkan pola belanja konsumen yang konsisten.

Visualisasi hasil memperkuat temuan tersebut, di mana grafik menunjukkan dominasi hubungan pembelian menuju *mineral water* dan beberapa item yang berfungsi sebagai hub dalam network pembelian. Interpretasi ini menunjukkan bahwa terdapat pola komplementer antarproduk yang dapat dimanfaatkan untuk strategi pemasaran seperti cross-selling, penempatan produk di rak toko, maupun penyusunan paket bundling. Secara keseluruhan, setiap tahapan menghasilkan pemahaman yang lebih mendalam tentang perilaku belanja pelanggan dan mengonfirmasi kemampuan algoritma association rule mining dalam memberikan insight bisnis berbasis data.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian, penerapan algoritma Apriori dan FP-Growth pada data transaksi ritel berhasil mengungkap pola pembelian konsumen melalui *frequent itemsets* dan aturan asosiasi. Transformasi data menjadi bentuk biner (*one-hot encoded*) terbukti memudahkan kedua algoritma untuk menghitung metrik *support*, *confidence*, dan *lift* secara tepat sehingga pola keterkaitan antarproduk dapat diidentifikasi. Aturan yang dihasilkan menunjukkan adanya kecenderungan konsumen membeli produk tertentu secara bersamaan, misalnya kombinasi susu dan air mineral atau *burger* dan telur. Produk-produk seperti *mineral water*, *milk*, dan *spaghetti* muncul berulang kali sebagai item yang berasosiasi dengan berbagai produk lain, menandakan perannya sebagai item pelengkap atau pusat (*hub*) dalam transaksi.

Evaluasi kinerja menunjukkan bahwa kedua algoritma menghasilkan jumlah aturan yang sama (63 aturan) dengan nilai *support* tertinggi sekitar 0,06, *confidence* tertinggi sekitar 0,51, dan *lift* tertinggi 3,29. Hal ini mengindikasikan bahwa kualitas aturan yang diperoleh dari Apriori dan FP-Growth pada dataset ini relatif identik. Namun, terdapat perbedaan signifikan pada waktu pemrosesan: Apriori hanya memerlukan sekitar 0,39 detik sedangkan FP-Growth memerlukan 6,95 detik untuk menghasilkan aturan dengan parameter yang sama. Temuan ini menunjukkan bahwa pada dataset dengan ukuran dan karakteristik seperti penelitian ini, algoritma Apriori lebih efisien secara komputasi dibandingkan FP-Growth, meskipun secara teoritis FP-Growth dirancang lebih cepat untuk dataset yang jauh lebih besar.

#### DAFTAR PUSTAKA

- [1] A. J. Soekandar and P. Pratiwi, "Difusi inovasi untuk keberlanjutan bisnis ritel kecil: Strategi pemasaran digital," *J. Ilm. Manajemen, Ekon. Dan Bisnis*, vol. 2, no. 1, pp. 81–99, 2023.
- [2] B. Maulana and M. H. Fikri, "Implementasi dan Pelatihan Aplikasi Point Of Sales (POS) Berbasis Website di Toko Hycal \_ Store," *ABDIMAS J. Pengabdi. Masy.*, vol. 3, no. 5, pp. 153–156, 2025.
- [3] D. Meliana, J. Riswati, and D. Astuti, "Analisis Perkembangan Bisnis Ritel Di Indonesia,"

- [4] *J. Bus. Econ. Manag. E-ISSN 3063-8968*, vol. 1, no. 3, pp. 235–243, 2025.
- [5] R. Fakhira, “Implementasi dan pelatihan aplikasi PO S untuk optimalisasi pengelolaan usaha pada toko MahyaLINE casual style,” *ABDIMAS J. Pengabdi. Masy.*, vol. 3, no. 4, pp. 133–139, 2025.
- [6] D. Apriana and C. Yuliansyah, “Mengoptimalkan Penjualan Online Melalui Teknik Data Mining (Studi Kasus E-Commerce),” *AL-MIKRAJJ. Stud. Islam Dan Hum. (E-ISSN 2745-4584)*, vol. 4, no. 02, pp. 514–527, 2024.
- [7] A. Herdiansyah, “Analisis Pola Pembelian Makanan Menggunakan Algoritma Apriori Di Restoran Simpang Raya,” *J. Penelit. Dan Pengkaj. Ilm. Eksakta*, vol. 4, no. 2, pp. 201–210, 2025.
- [8] D. Ursulum and T. Palaoag, “MARKET 4.0: EXPLORING MARKET BASKET ANALYSIS (MBA) ALGORITHMS FOR CO-MARKET INTELLIGENT APPLICATION,” *Proc. Eng.*, vol. 6, no. 4, pp. 1523–1530, 2024.
- [9] M. M. Alawadh and A. M. Barnawi, “A survey on methods and applications of intelligent market basket analysis based on association rule.,” *J. Big Data*, vol. 4, no. 1, 2022.
- [10] D. Dwiputra, A. M. 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,” *J. World Sci.*, vol. 2, no. 8, pp. 1229–1248, 2023.
- [11] S. H. Lina, M. O. Kadang, D. Patulak, R. D. Liklikwatil, D. Ramadhani, and N. S. A. P. Hamid, “PENERAPAN ALGORITMA FP-GROWTH UNTUK PEMAHAMAN MENDALAM POLA PEMINJAMAN KOLEKSI BUKU DI PERPUSTAKAAN,” *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 4, no. 2, pp. 114–123, 2025.
- [12] C. Rustam, S. Defit, and G. W. Nurcahyo, “Penerapan Data Mining Menggunakan Algoritma FP-Growth Dalam Analisis Data Penjualan,” *J. KomtekInfo*, pp. 205–212, 2024.
- [13] F. S. Zikri and M. Ikhsan, “The Comparison Between The Apriori Algorithm And The FP-Growth Algorithm In Determining Frequent Pattern,” *INOVTEK Polbeng-Seri Inform.*, vol. 10, no. 2, pp. 615–625, 2025.
- [14] N. A. Hibnastiar, A. F. Setiawan, and E. H. Susanto, “Penerapan Algoritma Apriori dalam Menentukan Rekomendasi Paket Produk: Implementation of Apriori Algorithm for Product Bundle Recommendations,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 1, pp. 321–331, 2025.
- [15] S. D. Rahmawati, A. B. Oktavia, F. S. A. Putri, and D. L. Fithri, “Penerapan Algoritma Apriori Untuk Menemukan Pola Asosiasi Pada Data Penjualan Retail Fashion,” *J. Sist. Inf. dan Sist. Komput.*, vol. 10, no. 2, pp. 289–295, 2025.
- [16] N. B. Uly, M. A. A. Lobo, N. D. A. Hary, and Y. Y. Hinggiranja, “IMPLEMENTASI ALGORITMA FP-GROWTH PADA PENEMPATAN PRODUK DI PRIMA DEWATA,” *J. Sist. Inf. dan Inform.*, vol. 8, no. 2, pp. 440–447, 2025.
- [17] S. Erpian, R. Astuti, W. Prihartono, and R. Hamonangan, “IMPLEMENTASI ALGORITMA FP-GROWTH UNTUK MENINGKATKAN PEMASARAN PRODUK MINUMAN BOBA BERBASIS PYTHON (STUDI KASUS: KEDAI NGENYOD’S DESA BOGOR INDRAMAYU),” *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, 2025.
- [18] Y. N. R. Putro, A. Afriansyah, and R. Bagaskara, “Penggunaan Algoritma Gaussian Naïve Bayes & Decision Tree Untuk Klasifikasi Tingkat Kemenangan Pada Game Mobile Legends,” *JUKI J. Komput. dan Inform.*, vol. 6, no. 1, pp. 10–26, 2024.
- [19] A. Fergina, P. A. Negara, A. Sujjada, and I. Sanjaya, “Implementasi Algoritma Apriori dan FP-Growth untuk Menganalisis Pola Pembelian Produk Skincare dan Kosmetik,” *J. Ilm. KOMPUTASI*, vol. 23, no. 3, pp. 433–442, 2024.
- [20] R. D. Parinduri, S. Defit, and G. W. Nurcahyo, “Implementasi Algoritma Apriori dalam Data Mining untuk Optimalisasi Stok Obat di Apotik,” *J. KomtekInfo*, pp. 89–97, 2024.
- [21] P. H. Marpaung, N. Sahara, A. Rambe, and F. Suryani, “Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Sibolga,”

- J. Penelit. Multidisiplin Bangsa*, vol. 1, no. 3, pp. 117–122, 2024.
- [21] A. F. Achmad, A. Rahim, and N. A. Verdikha, “Implementasi Data Mining Algoritma Apriori Pada Data Transaksi Pik Store,” *J. Inform. Polinema*, vol. 11, no. 2, pp. 112–203, 2025.