

## Penerapan Market Basket Analysis pada Data Transaksi Online Retail Menggunakan Algoritma Apriori dan FP-Growth dengan Kerangka CRISP-DM

Abiyu Alfia Budi Putro<sup>1</sup>, Muhammad Iqbal Mustofa<sup>2</sup>, Nauval Mirano<sup>3</sup>

<sup>1,2,3</sup>Program Studi Sistem Informasi, Universitas Pamulang

[abiyyualfia32@gmail.com](mailto:abiyyualfia32@gmail.com), [mustofaiqbal2222@gmail.com](mailto:mustofaiqbal2222@gmail.com), [novalmirano2@gmail.com](mailto:novalmirano2@gmail.com)

---

### Article Info

#### Article history:

Received June 1, 2026

Revised June 10, 2026

Accepted June 20, 2026

---

#### Keywords:

Data Mining, Customer Intelligence, Market Basket Analysis, Apriori, FP-Growth, CRISP-DM

---

### ABSTRACT

*This study uses Market Basket Analysis (MBA) to identify customer purchasing patterns in online retail data. Both Apriori and FP-Growth algorithms generated 218 identical association rules, but FP-Growth showed faster computational performance. The results can be applied to product bundling, cross-selling, and more effective inventory management.*

*This is an open access article under the CC BY-SA license.*



---

### Article Info

#### Article history:

Received June 1, 2026

Revised June 10, 2026

Accepted June 20, 2026

---

#### Kata kunci:

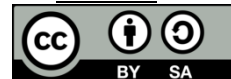
Data Mining, Customer Intelligence, Market Basket Analysis, Apriori, FP-Growth, CRISP-DM

---

### ABSTRAK

Penelitian ini menggunakan Market Basket Analysis (MBA) untuk menemukan pola pembelian pelanggan pada data retail online. Algoritma Apriori dan FP-Growth menghasilkan 218 aturan asosiasi yang sama, namun FP-Growth lebih cepat dalam proses komputasi. Hasilnya dapat digunakan untuk strategi bundling produk, cross-selling, dan pengelolaan persediaan yang lebih efektif.

*This is an open access article under the CC BY-SA license.*



---

### Corresponding Author:

Abiyu Alfia Budi Putro

Universitas Pamulang

Email: [abiyyualfia32@gmail.com](mailto:abiyyualfia32@gmail.com)

---

## Pendahuluan

Persaingan di industri *retail online* saat ini semakin ketat. Hal ini menuntut perusahaan untuk tidak hanya fokus mencari pelanggan baru, tetapi juga harus mengoptimalkan *Customer Intelligence* untuk meningkatkan *Average Order Value* (AOV). Setiap harinya, ada ribuan transaksi yang terjadi di platform *e-commerce*, namun data tersebut seringkali dibiarkan begitu saja tanpa dieksplorasi secara maksimal. Oleh karena itu, diperlukan pendekatan *Data Mining* untuk mengubah data mentah tersebut menjadi pengetahuan yang berguna dalam mengambil keputusan strategis. Masalah yang ingin dikaji dalam penelitian ini adalah untuk mengetahui pola kombinasi produk apa saja yang memiliki tingkat asosiasi tertinggi pada transaksi pelanggan, dan bagaimana pola tersebut bisa diterapkan sebagai rekomendasi strategi bagi bisnis *e-commerce*. Berdasarkan masalah tersebut, penelitian ini bertujuan untuk menemukan *association rules* yang menunjukkan probabilitas pelanggan membeli suatu

barang setelah membeli barang lainnya. Hasil dari penelitian ini diharapkan bisa membantu strategi penempatan barang (*slotting*), *cross-selling*, dan promosi *bundling*.

Terdapat beberapa landasan teori yang digunakan sebagai dasar penelitian, meliputi konsep *Data Mining*, *Market Basket Analysis* (MBA), algoritma Apriori, algoritma FP-Growth, *Association Rule*, CRISP-DM, dan *E-Commerce*. *Market Basket Analysis* adalah salah satu teknik data mining yang paling banyak digunakan di bidang pemasaran. Teknik ini bertujuan mencari item yang sering dibeli secara bersamaan oleh pelanggan (Han et al., 2000). Dalam memproses analisis ini, metode yang digunakan adalah algoritma Apriori dan FP-Growth. Algoritma Apriori bekerja dengan cara melakukan pemindaian (*scan*) database secara berulang kali untuk menghasilkan kandidat *itemset*. Di sisi lain, algoritma *Frequent Pattern Growth* (FP-Growth) memproses data dengan memampatkan database ke dalam struktur data *FP-Tree* tanpa perlu melakukan iterasi kandidat secara berulang. Hal ini membuat FP-Growth menjadi jauh lebih cepat dan hemat memori, meskipun aturan (*rules*) yang dihasilkan oleh kedua algoritma ini tetap sama.

Dalam menentukan seberapa kuat dan andal aturan asosiasi ( $X \rightarrow Y$ ) yang ditemukan, terdapat tiga parameter metrik evaluasi utama yang digunakan, yaitu *Support*, *Confidence*, dan *Lift*. Pertama, metrik *Support* digunakan untuk menunjukkan seberapa populer atau seberapa sering suatu *itemset* muncul di seluruh data transaksi. Kedua, *Confidence* digunakan untuk mengukur probabilitas bersyarat, yaitu seberapa sering item Y dibeli ketika pelanggan sudah membeli item X. Ketiga, metrik *Lift* berfungsi untuk mengukur seberapa kuat kecenderungan item X dan Y dibeli bersamaan, di mana nilai *Lift* yang lebih besar dari 1 ( $Lift > 1$ ) menandakan adanya asosiasi yang positif antar produk. Dengan menggunakan gabungan algoritma dan ketiga metrik evaluasi ini, proses analisis diharapkan bisa menemukan pola pembelian pelanggan dengan akurat, sehingga dapat dimanfaatkan sebagai strategi nyata untuk mengembangkan bisnis pada platform *e-commerce*.

## Metode

Penelitian ini menggunakan metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) dengan tahapan sebagai berikut:

1. *Business Understanding*: Memahami kebutuhan manajemen untuk merancang paket *bundling* dan meningkatkan retensi lewat penawaran yang relevan.
2. *Data Understanding*: *Dataset Online Retail II* dari Kaggle, merekam seluruh faktur transaksi retail online berbasis UK dengan lebih dari 540.000 baris data awal.
3. *Data Preparation*: Menghilangkan transaksi retur (faktur berawalan C), membersihkan missing values, memfilter kuantitas/harga  $\leq 0$ , memfokuskan data pada pelanggan UK, dan merestrukturisasi data menjadi format basket biner matriks.
4. *Modeling*: Algoritma Apriori dan FP-Growth diterapkan dengan minimum support 2% dan minimum Lift 1. Kinerja kedua algoritma dibandingkan berdasarkan waktu eksekusi komputasi.

## Hasil dan Pembahasan

Hasil penerapan metode *Market Basket Analysis* (MBA) pada data transaksi *Online Retail II* menggunakan algoritma Apriori dan FP-Growth. Seluruh proses analisis dalam penelitian ini dilakukan berdasarkan tahapan-tahapan dalam CRISP-DM, yang meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, serta evaluasi hasil. Secara lebih detail, penjelasan dalam bagian ini akan mencakup proses pengolahan data, aturan

asosiasi yang diperoleh, dan perbandingan kinerja dari kedua algoritma tersebut. Pada bagian akhir, hasil analisis akan diinterpretasikan menjadi bentuk rekomendasi strategi bisnis yang dapat diterapkan pada platform *e-commerce*.

**a. Eksperimen dan Evaluasi**

Setelah preprocessing, data menyusut menjadi 485.123 baris transaksi valid yang mencakup 18.019 Faktur (*Invoice*) dan 3.996 variasi produk. Kedua algoritma mengidentifikasi 401 frequent itemsets yang ekuivalen dan mengekstraksi 218 aturan asosiasi yang persis sama. Namun secara komputasi:

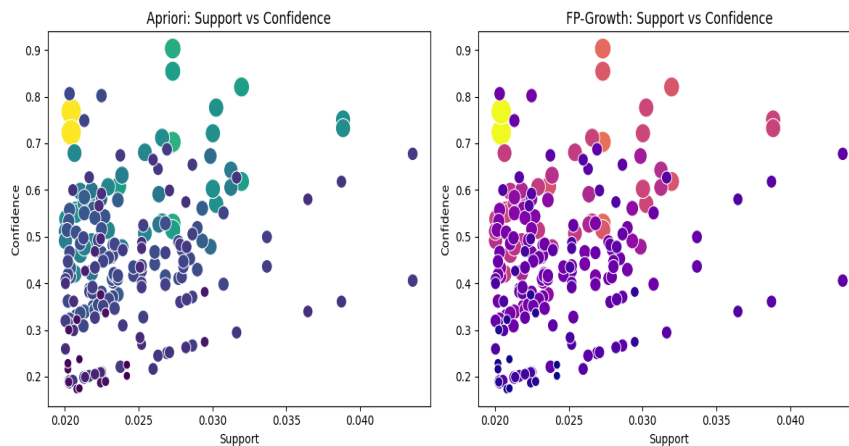
1. Apriori membutuhkan waktu eksekusi lebih lama karena iterasi kandidat berulang.
2. FP-Growth memproses data secara instan, sekitar 3-5 kali lebih cepat dari Apriori.

Tabel 1. Top Aturan Asosiasi dengan Lift Tertinggi

Antecedents	Consequents	Sup.	Conf.	Lift
Wooden Heart Xmas Scandinavian	Wooden Star Xmas Scandinavian	2.04%	72.29%	27.19
Roses Regency Teacup & Saucer	Pink Regency Teacup & Saucer	3.02%	77.63%	14.63
Green Regency Teacup & Saucer	Pink Regency Teacup & Saucer	3.19%	61.80%	15.86
Jumbo Bag Pears	Jumbo Bag Apples	2.06%	67.88%	13.82

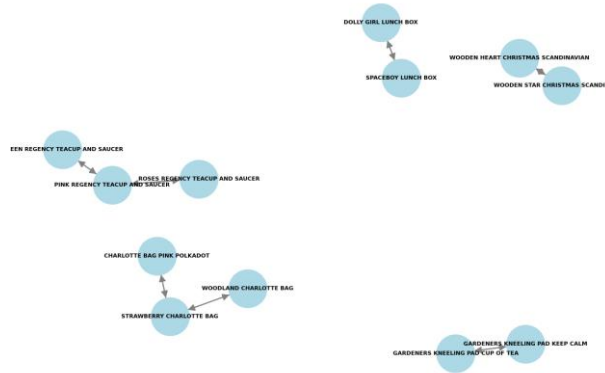
**b. Visualisasi Aturan Asosiasi**

Gambar 1 menunjukkan scatter plot perbandingan nilai Support vs Confidence antara model Apriori dan FP-Growth. Titik-titik aturan yang dihasilkan kedua algoritma identik sempurna.



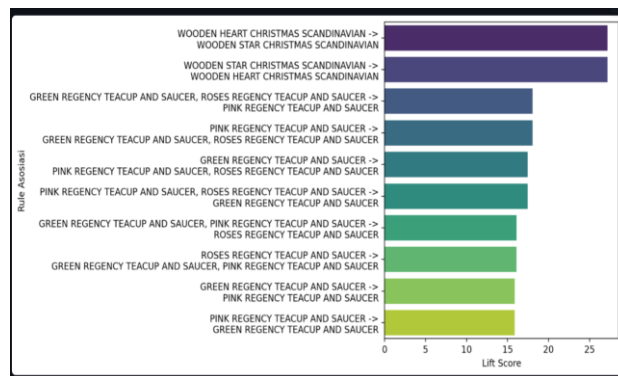
Gambar 1: Scatter Plot Support vs Confidence: Apriori (Kiri) vs FP-Growth (Kanan)

Gambar 2 menampilkan *network graph* yang merepresentasikan 20 aturan asosiasi teratas. Tanda panah menunjukkan arah asosiasi antar produk.



Gambar 2: Network Graph Top 20 Aturan Asosiasi

Gambar 3 menampilkan 10 aturan asosiasi teratas berdasarkan nilai Lift Score, dipimpin oleh pasangan Wooden Heart dan Wooden Star Christmas Scandinavian dengan Lift 27.19.



Gambar 3: Top 10 Aturan Asosiasi berdasarkan Lift Score

### c. Interpretasi

Tingginya nilai Lift pada produk Wooden Heart dan Wooden Star (27.19) membuktikan bahwa pembelian kedua barang tersebut sangat tidak independen. Pelanggan yang mencari satu ornamen natal memiliki kepastian tinggi (Confidence 72.29%) untuk membeli pasangannya. Pola serupa terlihat pada kategori *Regency Teacup*, di mana warna-warna komplementer (Pink, Green, Roses) mendominasi relasi cross-selling.

### d. Rekomendasi Bisnis

1. Promosi *Bundling*: Membuat SKU khusus paket Christmas Scandinavian Set dan *Regency Teacup Set* dengan diskon kecil bila dibeli bersamaan.
2. *Web Placement*: Integrasi rule ke UI *e-commerce* untuk *pop-up* rekomendasi otomatis saat pelanggan menambahkan produk ke keranjang.
3. Efisiensi Logistik: Menempatkan barang dengan Lift tinggi pada rak bersebelahan di gudang untuk mempercepat proses *picking*.

### Kesimpulan

Penelitian ini membuktikan bahwa FP-Growth sangat efektif untuk data transaksi e-commerce bervolume besar. Sebanyak 218 aturan asosiasi berhasil diekstraksi dari 485.123 baris data, merefleksikan pola thematic purchases seperti koleksi natal dan set cangkir teh. Pola yang diekstraksi sebaiknya diintegrasikan ke dalam arsitektur Sistem Informasi

perusahaan melalui Recommendation Engine yang berjalan di latar belakang situs belanja. Sementara itu, untuk penelitian selanjutnya disarankan agar dapat mengombinasikan FP-Growth dengan segmentasi klasterisasi (K-Means/RFM Analysis) agar perusahaan mengetahui tidak hanya produk yang di-bundle, tetapi juga segmen pelanggan yang paling tepat sebagai target.

### **Daftar Pustaka**

- Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. *ACM SIGMOD Record*, 29(2), 1-12.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. SPSS Inc, 9(13), 1-73.
- Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software*, 3(24), 638.
- Chen, Y. L., Tang, K., Shen, R. J., & Hu, Y. H. (2005). Market basket analysis in a multiple store environment. *Decision Support Systems*, 40(2), 339-354.