

FP-Growth Implementation for Market Basket Analysis in Building Materials Store in Surabaya

Valencia Melita Christy^{1*}, Indra Maryati^{2*}

* Information Systems, Universitas Ciputra Surabaya
vchristy@student.ciputra.ac.id¹, indra.maryati@ciputra.ac.id²

Article Info

Article history:

Received 2026-01-16

Revised 2026-02-18

Accepted 2026-02-27

Keyword:

Association Rules,
FP-Growth,
Market Basket Analysis,
Marketing Strategy,
Retail Store

ABSTRACT

Building materials store generally manage large volumes of transaction data with unique characteristics such as high product variety and uneven purchase distribution making these data often not fully used for business decisions effectively. This study analyses customer purchasing patterns and proposes product bundling recommendations specifically for building material sector using Market Basket Analysis with FP-Growth algorithm. The research uses 46,533 transactions records collected from a building materials store in Surabaya between July and September 2025. FP-Growth was selected for its efficiency in handling sparse data without candidate generation compared to traditional algorithm like Apriori. Due to the uneven distribution of product purchases, a minimum support of 0.5% and a minimum confidence of 0.3 were selected experimentally to obtain relevant and interpretable results. The strength of the relationships between products was evaluated using support, confidence, and lift values, identifying strong association in complementary items like nails and cement mixtures. To validate the results, interviews with the store owner and questionnaires for customers were conducted. CSAT was used to assess the accuracy of the purchasing patterns, while conversion rate measured customer interest in the proposed bundles. Results indicate a high validation score with CSAT reaching 75% and a potential conversion rate of up to 81% for specific bundles. These findings demonstrate that the proposed data-driven approach is highly suitable for practical implementation in optimizing retail marketing strategies.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Dalam beberapa tahun terakhir, pertumbuhan sektor konstruksi dan properti di Indonesia mengalami peningkatan yang signifikan. Perkembangan ini mendorong sektor ritel bahan bangunan untuk semakin bersaing untuk memenuhi kebutuhan pelanggan sekaligus mampu melawan persaingan harga yang sengit. Dalam persaingan, kemajuan teknologi digital juga mulai diterapkan untuk pengelolaan bisnis sehingga diperlukan pemahaman yang baik terhadap perilaku konsumen untuk merancang strategi pemasaran[1].

Sektor ritel bahan bangunan menghadapi tantangan besar dalam pengelolaan data yang beragam dalam jumlah besar. Kompleksitas data transaksi ini sering menyebabkan kesulitan dalam strategi penjualan yang kurang optimal. Padahal, apabila data tersebut diolah dengan baik akan

ditemukan pola asosiasi antar material yang sering dibeli secara bersamaan. Dari pola tersebut akan memberikan pengetahuan yang sangat krusial dalam merancang strategi pemasaran[2]. Namun dalam praktiknya, banyak pemilik toko masih mengandalkan intuisi dalam menentukan strategi penjualannya.

Salah satu cabang dalam *Data Mining* yang berfokus kepada pola asosiasi antar item dalam data transaksi penjualan adalah *Association Rule Mining*. Teknik ini dapat menemukan item yang memiliki keterkaitan berdasarkan frekuensi kemunculan dalam satu transaksi penjualan seperti dalam kasus nyata bahan bangunan, *Waterproof* sering dibeli bersamaan dengan Bata Ringan[3]. Penerapan *Association Rule Mining* secara spesifik pada sektor ritel dikenal sebagai *Market Basket Analysis* untuk membantu perusahaan untuk menemukan hubungan dalam memahami perilaku konsumen

dan merancang strategi pemasaran. *Market Basket Analysis* dapat diimplementasikan dengan beberapa algoritma seperti Apriori, FP-Growth, dan ECLAT.

Algoritma Apriori, yang menjadi dasar untuk *Market Basket Analysis*, pada awalnya dikembangkan oleh Agrawal dan Srikant pada tahun 1994[4], namun Apriori memiliki kekurangan dalam efisiensi sehingga pada tahun 2000 algoritma FP-Growth [5] dikembangkan oleh Han Jiawei untuk mengatasi hal tersebut. Selain itu, algoritma ECLAT juga diperkenalkan dengan pendekatan data vertikal untuk meningkatkan efisiensi pada beberapa kasus tertentu.

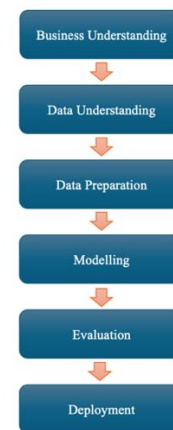
Penelitian sebelumnya[6] yang membandingkan ketiga algoritma ini mengevaluasi berdasarkan parameter jumlah transaksi penjualan, *minimum support*, dan penggunaan memori, dan waktu eksekusi. Hasilnya menunjukkan FP-Growth memiliki efisiensi komputasi terbaik terutama pada dataset besar karena dapat struktur pohon (*Frequent Pattern Tree*) yang meminimalisir proses pemindaian ulang data. Kemudian dalam studi literatur yang dilakukan pada tahun 2022[7] juga memperkuat penemuan sebelumnya dimana FP-Growth dan ECLAT memiliki kinerja yang lebih baik dibandingkan Apriori dalam hal waktu eksekusi dan penggunaan memori pada transaksi besar. Dalam penerapannya, algoritma FP-Growth dapat dipakai untuk mendukung strategi bisnis perusahaan. Penelitian terdahulu juga membuktikan bahwa algoritma FP-Growth sebagai algoritma yang banyak dipakai dalam bidang usaha ritel seperti pada usaha toko makanan sehat[2], toko fashion[8] toko kebutuhan sehari-hari[9], atau bahkan toko bahan bangunan[3]. Studi terdahulu mengungkapkan beberapa jenis dukungan yang dapat diberikan adalah dalam penyusunan tata letak desain produk [10], perencanaan strategi pemasaran dan penjualan [11] manajemen stok barang[12].

Berdasarkan beberapa hasil penelitian yang dicari, FP-Growth sudah memiliki keunggulan dalam studi kasus toko ritel bahan bangunan yang memiliki jumlah transaksi yang besar, variasi produk yang banyak, dan kebutuhan dalam menganalisis pengambilan keputusan bisnis sehingga dalam penerapannya penemuan ini dapat membantu pemilik toko untuk memastikan apakah strategi penjualan sudah sesuai dengan perilaku pembelian pelanggan serta menjadi validasi dalam penawaran kepada konsumen.

Studi ini akan menggunakan algoritma FP-Growth dalam menganalisis 46.533 transaksi penjualan dari sebuah toko ritel bahan bangunan di Surabaya dalam jangka waktu bulan Juli hingga September 2025. Analisis ini akan berfokus pada identifikasi kombinasi barang yang sering muncul dalam transaksi penjualan, sehingga dapat menghasilkan informasi yang membantu perusahaan merancang strategi pemasaran.

II. METODE

Studi ini mengadaptasi metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) [12] yang memiliki 6 tahap:



Gambar 1. Metodologi CRISP-DM

A. Business Understanding

Langkah pertama dalam proses metodologi CRISP-DM adalah memahami bisnis sekaligus mengetahui pola pembelian pelanggan. Dengan mengenali pola pembelian pelanggan, perusahaan dapat mengatur strategi pemasaran dengan baik[1].

B. Data Understanding

Data yang digunakan dalam studi ini adalah data transaksi penjualan dari sebuah toko ritel bahan bangunan di Surabaya dengan jangka waktu Juli hingga September 2025. Data mencakup kode transaksi, tanggal transaksi, nama barang, dan kuantitas barang yang dibeli. Data diambil melalui *software* Accurate yang dipakai dalam perusahaan, dan mengekspor-nya ke dalam Excel. Selain itu untuk memvalidasi data tersebut juga dilakukan observasi dan wawancara terhadap pemilik perusahaan.

C. Data Preparation

Data yang diperoleh akan diproses kembali ke beberapa tahapan untuk dapat digunakan pada tahap selanjutnya. Tahap ini dikenal dengan tahap *preprocessing data* yang mencakup pemilahan data, pembersihan data, dan transformasi data[12].

D. Modelling

Tahap berikutnya adalah tahap *Modelling* dengan menggunakan Python untuk menemukan kombinasi antar item yang paling sering muncul dan juga aturan asosiasinya. FP-Growth bekerja dengan cara membuat struktur pohon (*Frequent Pattern Tree*) sehingga proses dapat berjalan lebih cepat dan efisien [8]. Untuk menentukan *minimum support* dan *minimum confidence* terbaik, studi ini akan melalui beberapa percobaan kombinasi nilai *support* dan *confidence* sehingga kombinasi menghasilkan pola asosiasi paling baik. *Minimum support* menunjukkan seberapa sering kombinasi item akan muncul dalam seluruh transaksi dan dapat menjadi patokan dalam memastikan pola sering muncul dalam transaksi[11]. *Minimum confidence* menggambarkan seberapa besar kemungkinan seseorang bertransaksi barang kombinasi, sehingga dapat dipakai sebagai pengukur

hubungan antar produk dalam aturan asosiasi[3]. Dalam membuat algoritma FP-Growth, tahapan yang harus dilalui adalah sebagai berikut:

1) *Penentuan nilai minimum support dan minimum confidence*: Sebelum menentukan nilai *support* dan *confidence* yang sesuai, dapat dilakukan pengujian pada beberapa nilai kombinasi agar pola yang terbentuk tetap baik secara bisnis dan tidak menghasilkan *rule* yang terlalu sedikit maupun terlalu banyak. Untuk menghitung nilai *support* menggunakan frekuensi yang memiliki *itemset* tersebut dan dibagi dengan total jumlah transaksi yang tersedia dengan rumus pada Gambar 2. Sedangkan untuk menghitung nilai *confidence* menggunakan jumlah transaksi yang memiliki *itemset* dibagi oleh jumlah transaksi yang mengandung salah satu item seperti yang ditunjukkan dalam Gambar 3.

$$\text{Support} (X \rightarrow Y) = \frac{\text{Jumlah transaksi mengandung } X \text{ dan } Y}{\text{Total transaksi}}$$

Gambar 2. Rumus Perhitungan Nilai *Support*

$$\text{Confidence} (X \rightarrow Y) = \frac{\text{Jumlah transaksi mengandung } X \text{ dan } Y}{\text{Jumlah transaksi mengandung } X}$$

Gambar 3. Rumus Perhitungan Nilai *Confidence*

2) *Pengujian kombinasi parameter*: Eksperimen dilakukan terhadap beberapa parameter untuk mengukur dampaknya terhadap jumlah *frequent itemset* yang terbentuk, jumlah aturan asosiasi yang dihasilkan, serta waktu komputasi yang dibutuhkan.

3) *Implementasi Algoritma FP-Growth menggunakan Python*: Setelah menggunakan parameter yang sesuai, Library ‘pandas’ di Python dapat digunakan untuk memperbaiki struktur data dan ‘mlxtend’ digunakan untuk proses *frequent itemset mining* dan *rule generation*. Python memiliki kelebihan dalam efisiensi memori penggunaan CPU serta menjadi alat yang superior dalam pengolahan data dan analisis[13], [14]. dan juga hasil *output* dari proses ini menghasilkan daftar kombinasi item yang sering muncul dalam satu transaksi. Hasil ini kemudian digunakan sebagai dasar analisis strategi pemasaran dan rekomendasi *bundling* produk pada tahap berikutnya.

E. Evaluation

Evaluasi dilakukan untuk menjaga kualitas kombinasi item dengan menggunakan beberapa matriks seperti *Lift Ratio* dan evaluasi terhadap tujuan bisnis. *Lift ratio* digunakan untuk melihat seberapa kuat hubungan antar barang dengan cara membagi nilai *confidence* dengan nilai *support* yang dihasilkan. Nilai *lift* yang lebih dari 1 menandakan kombinasi item tersebut sering keluar bersamaan [11]. *Lift ratio* dapat dihitung menggunakan rumus pada Gambar 4.

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

Gambar 4. Rumus Perhitungan Nilai *Lift Ratio*

F. Deployment

Tahap akhir dari proses ini adalah *deployment*. Setelah mendapatkan hasil yang terbaik dari aturan asosiasinya, tahap selanjutnya adalah menginterpretasikan hasil yang dapat digunakan menjadi rekomendasi nyata bagi perusahaan. Untuk mengetahui seberapa efektifnya kombinasi item tersebut, akan dilakukan validasi terhadap pemilik toko melalui wawancara serta kuesioner singkat yang diberikan kepada pelanggan untuk menilai apakah *bundling* produk yang diusulkan relevan dengan kebutuhan mereka. Kuesioner akan menggunakan 5 skala Likert untuk mengukur kepuasan dan seberapa bermanfaat *bundling* produk tersebut.

III. HASIL DAN PEMBAHASAN

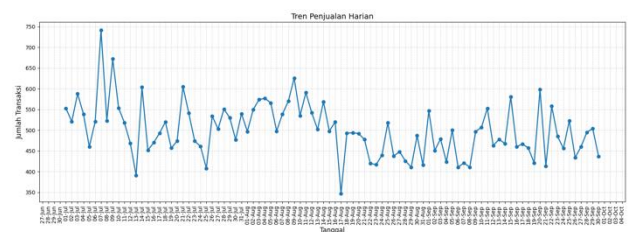
Hasil dapat memberikan kombinasi item dan hasil asosiasi yang paling banyak muncul berdasarkan data transaksi penjualan yang sudah dianalisis.

A. Business Understanding

Studi ini akan berfokus pada strategi penjualan yang akan dilakukan oleh sebuah toko bangunan di Surabaya. Sebelumnya, toko ini menggunakan intuisi dari pemilik toko dalam membuat promo tanpa ada *insight* dari data, namun belum pernah menerapkan produk *bundling* dalam strategi penjualannya. Dengan menggunakan data transaksi penjualan dari beberapa bulan ini, toko akan mendapatkan pengetahuan berbasis data mengenai kombinasi produk yang sering muncul bersama, memahami pola pembelian pelanggan, serta menentukan strategi bisnis berdasarkan data.

B. Data Understanding

Data transaksi penjualan yang diambil merupakan data transaksi dalam kurun waktu 1 Juli hingga 30 September 2025 dengan total 46.533 baris transaksi. Data ini yang menjadi dasar dalam analisis *data mining*. Data yang diambil hanya 3 bulan dan kemungkinan tidak dapat mewakili pola pembelian pembeli dalam satu tahun penuh. Dalam pemahaman mengenai data ini ditemukan bahwa toko bangunan memiliki keunikan yaitu variasi produk yang sangat beragam dengan pembelian berskala besar untuk sekali pembelian. Selain itu, tidak ditemukan adanya lonjakan penjualan dalam musim tertentu yang dibuktikan dalam Gambar 5.



Gambar 5. Grafik Penjualan Harian

Data transaksi penjualan yang dipakai ditunjukkan dalam Tabel 1.

TABEL I
DATA TRANSAKSI PENJUALAN

No. Faktur	Tgl Faktur	Keterangan Barang	Kuantitas	Unit 1 Barang	Jumlah
CASH.0 1.07202 5.0007	01 Jul 2025	DULU X PENTA LITE BASE C	2.00	GLN	444,0 00.00
CASH.0 1.07202 5.0014	01 Jul 2025	KUAS BIASA 3"	1.00	PCS	6,000. 00
CASH.0 1.07202 5.0014	01 Jul 2025	PAKU 4"	2.00	KG	32,00 0.00
...
RJ- 0925- 002647	30 Sep 2025	TEE DRAT DALA M KUNIN GAN 1/2 x 3/4 "	-1.00	PCS	- 21,00 0.00

Data yang dipakai adalah data yang masih mentah dimana 1 baris menandakan 1 barang yang ada di transaksi. Data ini diperoleh dalam bentuk CSV. Kolom yang dipakai untuk menjalankan algoritma FP-Growth adalah: No. Faktur, Tgl Faktur, Keterangan Barang.

TABEL II
PENJELASAN KOLON

Nama Kolom	Jumlah Unique Values	Type Data
No. Faktur	21.802	object
Tgl Faktur	92	object
Keterangan Barang	3.526	object
Kuantitas	322	object
Unit 1 Barang	26	object
Jumlah	1.557	object

Untuk memberikan gambaran yang lebih jelas mengenai data dari studi ini, Tabel II dapat menunjukkan contoh struktur data transaksi penjualan sebelum *preprocessing* dilakukan. Dataset awal memiliki jumlah data 46.533 transaksi, namun ketika dihitung nilai uniknya, jumlah transaksi yang sebenarnya adalah 21.802. Hal ini terjadi dikarenakan satu nomor faktur dapat berisi lebih dari satu barang sehingga satu nomor faktur dapat muncul berkali-kali. Hasil observasi yang dihasilkan lainnya adalah kolom No. Faktur memiliki 3 pola penomoran yaitu dengan kode 'CASH', 'FJ', dan 'RJ'. Kode 'CASH' merupakan transaksi yang dicatat langsung di kasir saat penjualan berlangsung. Sedangkan kode 'FJ' merupakan transaksi yang biasa dicatat manual oleh staf penjualan, terutama untuk pesanan pengiriman. Kode 'RJ' digunakan jika ada penjualan yang

diretur atau dikembalikan ke toko. Selain itu, jumlah unik dari keterangan barang atau nama barang sendiri berjumlah 3.526 dengan rata-rata 2,15 item per transaksi. Kondisi ini menunjukkan kemungkinan penyebaran kurang merata, yang dapat dibuktikan pada Tabel III.

TABEL III
DAFTAR 10 PRODUK DENGAN FREKUENSI TERTINGGI

No	Nama Barang	Jumlah Kemunculan
1	SEMEN GRESIK PPC 40KG	1357
2	SEMEN TIGA RODA PCC 40KG	946
3	LEM RAJAWALI PVAC	916
4	PASIR KEDIRI	689
5	SEALTAPE AIR	572
6	SEMEN MERDEKA	561
7	PASIR LUMAJANG	456
8	PAKU 3"	400
9	PASIR KEDIRI A	389
10	SEMEN HITAM 5 KG AN	385

Tabel III menunjukkan bahwa data memiliki penyebaran frekuensi tidak seimbang, dengan jumlah transaksi yang rendah, sedangkan hanya sedikit produk yang muncul lebih sering. Sebagian contohnya, produk dengan frekuensi tertinggi yaitu 'SEMEN GRESIK PPC 40KG' hanya mencakup sekitar 2% dari seluruh transaksi yang menunjukkan bahwa pola pembelian sangat beragam dan tidak terpusat kepada beberapa produk saja.

Sebagai ringkasan, Tabel IV merupakan tabel statistik deskriptif dari data transaksi ini.

TABEL IV
TABEL STATISTIK DESKRIPTIF

Metrik Statistik	Nilai	Keterangan
Total Record Transaksi	46.533	Total baris data sebelum <i>cleaning</i>
Total Transaksi (Invoice)	21.802	Jumlah struk unik
Total Produk Unik	3.526	Variasi item yang dijual
Rata-rata item per Transaksi	2.15	Rata-rata ukuran keranjang belanja (<i>Basket Size</i>)
Maksimum item per Transaksi	20	Transaksi dengan jumlah unik item terbanyak
Kategori Produk Dominan	Semen, Perekat, Pasir	Berdasarkan frekuensi transaksi
Periode Data	3 bulan	Juli – September 2025

Selain itu, tipe data pada semua kolom masih belum sesuai untuk proses analisis, sehingga diperlukan penyesuaian dan konversi format pada tahap *Data Preparation*.

C. Data Preparation

Dalam mempersiapkan data agar bisa dipakai dalam permodalan data, harus dilakukan pra-proses terlebih dahulu. Proses ini meliputi pemilahan data, pembersihan data, dan transformasi data yang dilakukan melalui Python. Tahapan yang dilakukan sebagai berikut:

1) *Pemilahan data*: Dalam tahap awal untuk mempersiapkan data dilakukan penggantian nama kolom dan penghapusan kolom yang kurang relevan untuk analisis data. Pemilahan data bertujuan untuk menyederhanakan struktur dari dataset dan memastikan proses *modelling* hanya mencakup atribut penting seperti nomor faktur dan keterangan barang.

2) *Pembersihan data*: Untuk memastikan kualitas data, dilakukan penghapusan data transaksi yang tidak relevan untuk analisis. Pada tahap ini ditemukan adanya data transaksi retur sejumlah 546 baris yang ditandai dengan kuantitas bernilai kurang dari 0. Selain itu, ditemukan indikasi duplikasi data, sehingga dilakukan penghapusan 31 baris data duplikat. Proses pembersihan dilakukan dengan menghapus seluruh baris yang memiliki kuantitas negatif, yang secara otomatis menghapus seluruh transaksi retur beserta data ganda. Pembersihan ini dilakukan agar algoritma FP-Growth berjalan dengan baik karena proporsi data yang dibersihkan relatif kecil, hanya sekitar 1,2% dari keseluruhan dataset, sehingga tidak mempengaruhi representativitas data. Total data bersih yang siap digunakan untuk analisis selanjutnya adalah sebanyak 45.956 baris.

3) *Transformasi data*: Data transaksi yang sudah dibersihkan masih memiliki ketidaksesuaian format. Misalnya, pada kolom Jumlah (harga) yang tidak sesuai dengan menggunakan titik dan koma sebagai pemisah sehingga masih dikenali sebagai tipe data *object*. Selain itu, dalam kolom Kuantitas belum berbentuk numerik sehingga perlu dikonversi menjadi tipe data *float*. Kolom Tanggal juga harus diubah menjadi tipe data *datetime* sesuai standar analisis. Setelah dilakukan proses konversi tipe data dan penyesuaian format, dataset menjadi lebih terstruktur dan siap diproses ke tahap transformasi dan *encoding*. Beberapa contoh dataset yang sudah bersih dapat dilihat pada Tabel V.

TABEL V
HASIL DATA CLEANING TRANSAKSI PENJUALAN

No_Faktur	Tgl_Faktur	Nama_Barang	Kuantitas	Jumlah
CASH.01.0 72025.0007	2025-07-01	DULUX PENTALI TE BASE C	2.0	444000. 0
CASH.01.0 72025.0014	2025-07-01	KUAS BIASA 3"	1.0	6000.00
CASH.01.0 72025.0014	2025-07-01	PAKU 4"	2.0	32000.0 0
...

FJ-0925-003822	2025-09-30	SHOWER TIANG AER MBS-5 BL	1.0	278400 0.00
----------------	------------	---------------------------------------	-----	----------------

4) *Pengelompokan item dan encoding*: Data transaksi yang sudah bersih dapat dikelompokkan berdasarkan nomor faktur dan nama barang untuk penyusunan daftar item per transaksi agar dapat digunakan dalam pencarian *frequent itemset*. Setelah pengelompokan, dataset diubah dalam susunan *binary encoding* yang bernilai 1 dimana menandai kemunculan setiap item dalam sebuah transaksi. Encoding diperlukan sebagai input dalam proses *frequent itemset mining* menggunakan algoritma FP-Growth.

D. Modelling

Dalam tahap Modelling, hal pertama yang harus dilakukan adalah menentukan *minimum support* dan *minimum confidence*. Nilai-nilai tersebut didapatkan melalui beberapa percobaan *trial and error* yang dilakukan dengan mempertimbangkan hasil *frequent itemset* yang diberikan beserta hasil *association rule*-nya. Percobaan awal dilakukan dengan cara mengetahui berapa jumlah *itemset* yang ditemukan dan jumlah *association rule*, yang ditunjukkan pada Tabel VI.

TABEL VI
JUMLAH ITEMSET DAN ASSOCIATION RULES

Minimum Support	Minimum Confidence	Jumlah Itemset	Jumlah Rules
0.02 (2%)	0.2	7	0
	0.3	7	0
	0.4	7	0
	0.5	7	0
0.01 (1%)	0.2	27	0
	0.3	27	0
	0.4	27	0
	0.5	27	0
0.005 (0.5%)	0.2	67	8
	0.3	67	5
	0.4	67	0
	0.5	67	0
0.003 (0.3%)	0.2	146	23
	0.3	146	9
	0.4	146	3
	0.5	146	2

Angka *minimum support* dan *minimum confidence* ditentukan berdasarkan pendekatan berbasis data, yang berlandaskan oleh distribusi data yang sudah dianalisis pada tahap *Data Understanding*. Pada Tabel III, terlihat frekuensi yang tidak merata, bahkan produk dengan frekuensi tertinggi hanya mencakup 2% dari total transaksi. Menganalisis dari karakteristik dari data ini, nilai *minimum support* dan *confidence* yang tinggi akan menghilangkan banyak pola pembelian yang sesuai karena hanya sedikit produk yang mampu memenuhi ambang batas tersebut [15]. Angka 0.5%

dipilih karena ini adalah batas minimum yang masuk akal secara bisnis untuk dianggap sebagai 'pola berulang'. Produk yang terjual kurang dari 100 kali dalam 3 bulan dianggap kebetulan. Pengujian dilakukan secara bertahap dengan menyesuaikan nilai *support* dan *confidence* beserta hasil *itemset* dan *association rule*. Pemilihan nilai awal hingga nilai akhir yang dipakai juga didasarkan pada tujuan bisnis, yaitu menemukan produk yang dapat direkomendasikan sebagai *bundling*, sehingga diperlukan keseimbangan antara item yang cukup sering [11], [16].

Berdasarkan Tabel VI yang sudah dihasilkan berdasarkan beberapa percobaan, ditemukan bahwa semakin tingginya nilai *support*, semakin sedikit *itemset* yang dihasilkan. Seperti pada contoh di nilai *minimum support* 2% hanya menghasilkan 7 *itemset* yang tidak dapat menunjukkan pola asosiasi dengan item lain sehingga tidak dapat merepresentasikan hubungan antar produk. Hal yang sama terjadi pada nilai *support* 1% dimana *itemset* muncul sebanyak 27, tetapi tidak ada kemunculan *association rule*.

Setelah menurunkan nilai *support* ke 0.5%, algoritma mulai menghasilkan *itemset* dan *association rule*. Pada nilai *support* 0.5% dan nilai *confidence* 0.2, terdapat kemunculan 8 *association rule*, sedangkan dengan nilai *confidence* 0.3, terdapat 5 *association rule*. Jumlah aturan asosiasi pada nilai *confidence* 0.2 menghasilkan lebih banyak *rule* namun, sebagian *rule* yang terbentuk pada *confidence* 0.2 memiliki tingkat kepercayaan yang lebih rendah sehingga berisiko menghasilkan rekomendasi *bundling* yang kurang kuat tentang hubungan antar produk. Ketika nilai *confidence* dinaikkan menjadi 0.3, jumlah *rule* memang berkurang, tetapi *rule* yang dihasilkan memiliki tingkat kepercayaan yang lebih tinggi dan lebih dapat diandalkan untuk dijadikan dasar rekomendasi. Dengan mempertimbangkan tujuan bisnis, yaitu menemukan pasangan produk yang layak direkomendasikan sebagai *bundling*, diperlukan *rule* yang tidak hanya banyak secara jumlah, tetapi juga cukup kuat dari sisi *confidence*-nya. Sehingga nilai *minimum support* 0.5% dan nilai *minimum confidence* 0.3 dipilih sebagai hasil aturan asosiasi yang terbaik.

TABEL VII
ASSOCIATION RULES

Antecedents	Consequents	Support	Confidence	Lift
PAKU 3"	PAKU 2"	0.006084	0.325	21.301
PAKU 2"	PAKU 3"	0.006084	0.398	21.301
SEMEN ACIAN TR-30 @40KG	LEM RAJAWALI PVAC	0.005710	0.335	7.818
PASIR LUMAJANG	SEMEN GRESIK PPC 40KG	0.008518	0.399	6.284
PASIR KEDIRI A	SEMEN GRESIK PPC 40KG	0.006693	0.367	5.788

Antecedents dan *Consequents* digunakan sebagai penggambaran dalam sebuah hubungan antar produk. *Antecedents* merujuk kepada barang pertama yang menjadi pemicu dalam aturan asosiasi, dan *consequents* sebagai barang kedua yang cenderung ikut dibeli dalam suatu transaksi sehingga dapat diartikan jika barang *antecedents* dibeli, maka barang *consequents* memiliki kecenderungan untuk dibeli bersamaan, sehingga sesuai untuk dijadikan *bundling*. Lima aturan asosiasi yang dihasilkan menunjukkan nilai lift yang tinggi, yang menandakan adanya hubungan pembelian yang kuat. Pola ini dapat dimanfaatkan sebagai dasar rekomendasi *bundling* produk di perusahaan.

5 aturan asosiasi ini membentuk dasar untuk pengambilan keputusan berbasis data yang digunakan dalam pembuatan paket promosi *bundling* yang efektif yang disesuaikan dengan pola layanan dan perilaku pelanggan.

Seluruh eksperimen dan pengolahan data dalam penelitian ini dilakukan menggunakan perangkat komputasi dengan spesifikasi prosesor Apple M4 Pro Chip (14-core CPU, 20-core GPU) dan memori 24GB RAM. Sistem operasi yang digunakan adalah macOS Tahoe 26.2. Implementasi algoritma FP-Growth dibangun menggunakan bahasa pemrograman Python versi 3.14. *Library* pendukung utama yang digunakan meliputi Pandas untuk manipulasi data struktur dataframe dan Mlxtend (Machine Learning Extensions) untuk eksekusi algoritma *frequent pattern mining* dan aturan asosiasi.

E. Deployment

Dalam tahap *Deployment*, hal pertama yang harus dilakukan adalah melakukan interpretasi hasil yang disajikan di Tabel VII. Aturan asosiasi terkuat ditemukan pada kombinasi "Paku 3" dan "Paku 2" dengan nilai lift sebesar 21. Hal ini menunjukkan bahwa pelanggan yang membeli "Paku 3" memiliki kemungkinan 21 kali lebih besar untuk juga membeli "Paku 2" secara bersamaan. Hubungan asosiasi yang sangat kuat dan masuk akal karena kebiasaan pembelian sering membeli paku dengan berbagai ukuran sehingga memenuhi kebutuhan beragam jenis sambungan material. Aturan asosiasi yang kedua sama seperti aturan asosiasi yang pertama, hanya saja perbedaannya terletak pada nilai *confidence* yang hanya berbeda sedikit akibat perbedaan frekuensi kemunculan masing-masing produk. Perbedaan kecil tersebut tidak mengubah kesimpulan bahwa kedua produk ini memiliki hubungan pembelian yang konsisten dan layak dijadikan rekomendasi *bundling*.

Kemudian, aturan asosiasi berikutnya menunjukkan hubungan antara "Semen Acian TR-30 @40KG" dan "Lem Rajawali PVAC" dengan nilai lift sebesar 7,81. Nilai ini mengindikasikan bahwa konsumen yang membeli "Semen Acian TR-30 @40KG" memiliki kemungkinan sekitar 7 kali lebih besar untuk juga membeli "Lem Rajawali PVAC". Pola pembelian ini juga telah dikonfirmasi oleh pemilik perusahaan, karena kedua produk tersebut umumnya digunakan secara bersamaan dalam pekerjaan plamir atau finishing dinding.

Selanjutnya adalah “Pasir Lumajang” dan “Semen Gresik PPC 40KG” serta “Pasir Kediri” dengan “Semen Gresik PPC 40KG”. Ketiga barang ini memiliki hasil nilai lift yang hampir sama yaitu 6 dan 5, dan kedua aturan asosiasi ini juga memiliki kesamaan fungsi yaitu sebagai pondasi dalam pembangunan.

Hubungan antar produk yang ditemukan tidak hanya kuat secara statistik, tetapi juga terbukti relevan secara praktis berdasarkan validasi dari pemilik perusahaan. Rekomendasi *bundling* dapat diterapkan dengan menawarkan produk utama beserta item pelengkapinya dalam bentuk paket harga khusus yang memberikan nilai tambah berupa kenyamanan (*convenience*) dan mempercepat proses transaksi di kasir. Pendekatan ini memanfaatkan metrik asosiasi seperti lift dan *confidence* untuk memastikan bahwa pasangan produk benar-benar memiliki hubungan pembelian yang konsisten. Dengan demikian, *bundling* yang dihasilkan tidak hanya memberikan nilai tambah bagi pelanggan, tetapi juga menguntungkan bisnis karena sesuai dengan pola pembelian aktual di toko bangunan.

Selanjutnya dilakukan validasi terhadap seberapa efektifnya aturan asosiasi tersebut dengan melalui wawancara dan kuesioner. Pendekatan validasi ini dilakukan untuk menjamin validitas dan reliabilitas hasil rekomendasi sebelum diterapkan secara luas. Wawancara dilakukan dengan pemilik toko bangunan yang telah berpengalaman dalam mengelola penjualan dan memahami pola pembelian pelanggan sehari-hari selama beberapa tahun. Kuesioner disebar ke sejumlah responden, yang merupakan pelanggan yang telah melakukan transaksi pada hari yang sama di toko tersebut.

Berdasarkan hasil wawancara, pemilik toko menyatakan bahwa kombinasi produk seperti paku dengan ukuran berbeda memang sering dibeli bersamaan, terutama oleh tukang yang membutuhkan beberapa ukuran paku untuk satu pekerjaan. Hal ini menguatkan hasil *association rule* dengan nilai lift yang tinggi. Kemudian dalam kasus semen acian dan lem rajawali biasa digunakan untuk plamir, dan pasir dan semen gresik digunakan sebagai pondasi bangunan. Konfirmasi dari pemilik toko ini menguatkan validitas hasil *association rule* yang memiliki nilai lift tinggi, serta memastikan bahwa rekomendasi *bundling* logis untuk diterapkan.

Pengumpulan data kuesioner dilakukan secara *online* dan *offline* dengan menggunakan *Google Forms*. Metode ini dipilih untuk memudahkan proses pengumpulan data secara cepat dan efisien, serta memberikan fleksibilitas bagi responden dalam mengisi kuesioner tanpa mengganggu proses transaksi. Pengisian secara *offline* dilakukan dengan pendampingan langsung oleh peneliti di area kasir, sehingga responden yang kurang terbiasa mengakses formulir digital tetap dapat berpartisipasi. Pendampingan ini bertujuan untuk menjaga reliabilitas data, memastikan responden memahami pertanyaan dengan konsisten, serta meminimalisir bias jawaban. Sementara itu, tautan kuesioner juga disediakan

melalui poster yang dicetak dan diletakkan di area kasir toko bangunan.

Penentuan jumlah responden dalam penelitian ini mengacu pada rumus Slovin[17]. Berdasarkan data rata-rata transaksi harian sebanyak 200 pelanggan, dengan periode pengambilan data selama 7 hari, maka jumlah populasi yang diperkirakan sebesar 1.400 pelanggan. Dengan menggunakan tingkat kesalahan (*margin of error*) sebesar 10%, jumlah minimum sampel yang harus dicari adalah sebanyak 94 responden dan hasilnya diperoleh jumlah responden yang terkumpul dalam penelitian ini sebanyak 101 responden, yang dinilai telah cukup mewakili populasi untuk keperluan validasi rekomendasi *bundling* produk. Pengambilan data dilakukan pada bulan Desember 2025.

Pemilihan responden dilakukan menggunakan metode *accidental sampling*, di mana setiap pelanggan yang kebetulan bertransaksi pada saat penelitian berlangsung dipilih sebagai responden. Pendekatan ini digunakan untuk memperoleh gambaran persepsi pelanggan yang lebih objektif terhadap rekomendasi *bundling* berdasarkan hasil *association rule*.

Instrumen kuesioner yang dipakai dalam penelitian ini terdiri dari 2 jenis pertanyaan utama yang mengukur validitas persepsi, yaitu tingkat keakuratan dan relevansi pola pembelian berdasarkan pengalaman responden dan tingkat ketertarikan responden terhadap skema *bundling* yang diusulkan. Tingkat keakuratan dan ketertarikan diukur menggunakan skala Likert lima tingkat, mulai dari sangat tidak setuju, tidak setuju, netral, setuju, dan sangat setuju.

Untuk membedakan antara validasi pola pembelian dan minat bisnis terhadap rekomendasi *bundling*, penelitian ini menggunakan dua metrik evaluasi, yaitu CSAT (*Customer Satisfaction Score*) dan *Conversion Rate*. Hal ini dilakukan untuk membedakan antara validasi pola pembelian secara perilaku dan potensi dalam penerapan di masa depan. CSAT digunakan untuk mengukur tingkat kesesuaian dan keakuratan pola pembelian yang dihasilkan oleh algoritma FP-Growth berdasarkan pengalaman dan kesadaran konsumen di lapangan. Perhitungan CSAT menggunakan rumus yang ditunjukkan pada Gambar 6[11].

$$CSAT = \frac{\text{Jumlah responden setuju dan sangat setuju}}{\text{Total responden}} \times 100\%$$

Gambar 6. Rumus CSAT

TABEL VIII
HASIL KUESIONER CSAT

No	Produk Bundling	Validasi Pola Pembelian
1	Paku 2” & Paku 3”	_73,26% setuju
2	Semen Acian TR-30 & Lem Rajawali PVAC	68,31% setuju
3	Pasir Lumajang atau Pasir Kediri & Semen Gresik	_75,24% setuju

Berdasarkan hasil CSAT pada Tabel VIII, mayoritas responden memberikan penilaian setuju dan sangat setuju

terhadap pola pembelian yang diberikan. Hal ini membuktikan bahwa pola yang dihasilkan oleh algoritma FP-Growth sudah sesuai dengan kebiasaan pelanggan dan dapat dianggap valid.

Sementara itu, *conversion rate* digunakan untuk merepresentasikan minat konsumen dalam melakukan aksi pembelian. Metriks ini mencerminkan sejauh mana rekomendasi *bundling* layak untuk dikejar dan diterapkan oleh perusahaan sebagai strategi bisnis seperti untuk meningkatkan kemungkinan pembelian ulang juga loyalitas[18]. Perhitungan *conversion rate* dilakukan dengan menghitung proporsi responden yang memberikan penilaian positif, yaitu memilih tingkat setuju dan sangat setuju, sebagaimana ditunjukkan pada rumus di Gambar 7.

$$\text{Conversion rate} = \frac{\text{Jumlah responden tertarik (setuju dan sangat setuju)}}{\text{Total responden}} \times 100\%$$

Gambar 7. Rumus *conversion rate*

TABEL IX
HASIL KUESIONER CONVERSION RATE

No	Produk Bundling	Minat Pembelian
1	Paku 2” & Paku 3”	79,20% tertarik
2	Semen Acian TR-30 & Lem Rajawali PVAC	81,18% tertarik
3	Pasir Lumajang atau Pasir Kediri & Semen Gresik	73,26% tertarik

Berdasarkan hasil *conversion rate* pada Tabel IX, sebagian besar responden menunjukkan kecenderungan yang tinggi untuk tertarik terhadap skema bundling yang diusulkan. Hal ini juga membuktikan bahwa rekomendasi *bundling* tidak hanya sesuai dengan pola pembelian pelanggan, tetapi juga memiliki potensi untuk diterapkan secara nyata dalam lingkungan bisnis toko bangunan. Temuan ini memperkuat hasil pemodelan menggunakan algoritma FP-Growth dan menunjukkan keselarasan antara hasil analisis data dan perilaku konsumen secara nyata.

IV. KESIMPULAN

Penelitian ini menggunakan algoritma FP-Growth terbukti dapat dimanfaatkan sebagai landasan pengambilan keputusan untuk strategi *bundling* produk di toko bangunan. Aturan asosiasi yang dihasilkan berhasil mewakili hubungan antar produk yang tidak hanya selaras secara statistik, tetapi juga relevan dengan kebutuhan konsumen di lapangan. Hal ini mengonfirmasi juga bahwa pendekatan *data mining* dapat merepresentasikan perilaku pembelian pelanggan secara akurat.

Validasi hasil pemodelan dilakukan melalui wawancara dengan pemilik toko serta kuesioner kepada pelanggan menggunakan dua metrik evaluasi, yaitu CSAT dan *conversion rate*. Hasil validasi menunjukkan bahwa mayoritas responden menyatakan kesesuaian terhadap pola pembelian yang diusulkan dan memiliki minat terhadap skema *bundling* produk. Temuan ini mengindikasikan bahwa

rekomendasi *bundling* yang dihasilkan tidak hanya kuat secara analitis, tetapi juga relevan dengan kondisi pembelian di lapangan dan berpotensi untuk diterapkan secara nyata dalam operasional toko bangunan. Namun penelitian ini memiliki keterbatasan, yaitu periode dataset yang digunakan hanya mencakup transaksi selama tiga bulan sehingga belum bisa mewakili keseluruhan penjualan di dalam toko ini. Selain itu, penelitian ini juga belum mempertimbangkan aspek profitabilitas sehingga rekomendasi *bundling* yang dihasilkan bertujuan untuk meningkatkan volume penjualan, yang belum tentu memaksimalkan total keuntungan finansial toko secara langsung.

Berdasarkan hasil penelitian ini, disarankan agar pihak toko bangunan dapat memanfaatkan hasil rekomendasi *bundling* produk sebagai bahan pertimbangan dalam strategi penjualan, khususnya untuk menawarkan paket produk yang sesuai dengan kebiasaan pembelian konsumen. Untuk implementasi yang berkelanjutan, disarankan agar algoritma ini diintegrasikan secara langsung ke dalam sistem manajemen basis data kasir atau *Point of Sale (POS)*. Integrasi ini memungkinkan sistem untuk memberikan rekomendasi *bundling* secara *real-time* di layar kasir saat transaksi berlangsung, serta memungkinkan pembaruan pola asosiasi secara otomatis seiring bertambahnya data transaksi harian. Selain itu, penelitian selanjutnya dapat menggunakan data transaksi dengan rentang waktu yang lebih panjang agar pola pembelian dapat dianalisis secara lebih menyeluruh, serta mengombinasikan hasil *association rule* dengan evaluasi lanjutan seperti uji implementasi atau analisis peningkatan penjualan. Pengembangan metode dengan mempertimbangkan faktor lain, seperti harga dan margin keuntungan, juga diharapkan dapat menghasilkan rekomendasi yang lebih optimal.

DAFTAR PUSTAKA

- [1] F. Y. Ramadhani and H. Ali, “Determinasi Keputusan Pembelian: Analisis Strategi Pemasaran, Perilaku Konsumen dan Teknologi,” *Dinasti Information and Technology*, pp. 145–155, 2025.
- [2] K. Singha, P. Parthanadee, A. Kessuvan, and J. Buddhakulsomsiri, “Market Basket Analysis of a Health Food Store in Thailand: A Case Study,” *International Journal of Knowledge and Systems Science*, vol. 15, no. 1, 2024, doi: 10.4018/IJKSS.333617.
- [3] Faris Syaifulloh, Eva Yulia Puspaningrum, and M. Muharram Al Haromayni, “Analisis Pola Pembelian Pelanggan Menggunakan Algoritma Squeezer, Apriori dan FP-Growth Pada Toko Bangunan,” *Modem: Jurnal Informatika dan Sains Teknologi.*, vol. 2, no. 3, pp. 134–147, Jul. 2024, doi: 10.62951/modem.v2i3.153.
- [4] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules,” 1994.
- [5] J. Han, J. Pei, and Y. Yin, “Mining Frequent Patterns without Candidate Generation,” 2000.
- [6] V. Srinadh, “Evaluation of Apriori, FP growth and Eclat association rule mining algorithms,” *Int. J. Health Sci. (Qassim).*, pp. 7475–7485, Apr. 2022, doi: 10.53730/ijhs.v6ns2.6729.
- [7] D. Erlina, P. Suretno, D. Irfansyah, Y. Setiawan, and I. Pratama, “Literatur Sistematis Perbandingan Kinerja Algoritma Apriori, ECLAT, dan FP-Growth dalam Analisis Pola Asosiasi,” *Indonesian Journal of Computer Science and Engineering (IJCSSE)*, 2025, doi: 10.70656/ijcse.v2i01.323.

- [8] U. M. Wulandari, A. T. Suseno, and M. Kholilurrahman, "Market Basket Analysis Using FP-Growth and Apriori on Distro Store Sales Transaction," *MATICS: Jurnal Ilmu Komputer dan Teknologi Informasi (Journal of Computer Science and Information Technology)*, vol. 17, no. 1, pp. 12–18, Mar. 2025, doi: 10.18860/mat.v17i1.28820.
- [9] I. Ismarmiaty and R. Rismayati, "Product Sales Promotion Recommendation Strategy with Purchase Pattern Analysis FP-Growth Algorithm," *Sinkron*, vol. 8, no. 1, pp. 202–211, Jan. 2023, doi: 10.33395/sinkron.v8i1.11898.
- [10] E. Haerani and C. Juliane, "Finding Customer Patterns Using FP-Growth Algorithm for Product Design Layout Decision Support," *SISTEMASI: Jurnal Sistem Informasi*, pp. 402–413, May 2022, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [11] I. Gusti *et al.*, "Implementation of FP-Growth Algorithms for Promo Package Determination in a Scooter Motorcycle Workshop Business," 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [12] D. Kurniawan, M. Sahata Sipayung, R. Ismayanti, M. Rivani Ibrahim, Y. Bintan, and S. Aulia Miranda, "Optimalisasi Strategi Pemenuhan Persediaan Stok Barang Menggunakan Algoritma Frequent Pattern Growth," *METIK JURNAL*, vol. 6, no. 2, pp. 104–114, Dec. 2022, doi: 10.47002/metik.v6i2.387.
- [13] A. Ivana and I. Maryati, "Implementation of Apriori Algorithm in Identifying Purchase Relationships at Bluder Cokro Pakuwon Mall," 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [14] D. Syahirah, P. Priati, and O. P. Martadireja, "Association Rule Mining across Multiple Domains: Systematic Literature Review," *sinkron*, vol. 9, no. 4, pp. 1953–1964, Oct. 2025, doi: 10.33395/sinkron.v9i4.15227.
- [15] M. Ogedengbe, S. Junaidu, and D. Kana, "Adaptive Minimum Support Threshold for Association Rule Mining," *Indonesian Journal of Data and Science*, vol. 5, no. 2, Jul. 2024, doi: 10.56705/ijodas.v5i2.134.
- [16] R. M. T. Nursasongka, I. Fahrurrozi, U. Y. Oktawati, U. Taufiq, U. Farooq, and G. Alfian, "Utilizing association rule mining for enhancing sales performance in web-based dashboard application," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 2, pp. 1105–1113, Nov. 2024, doi: 10.11591/ijeecs.v36.i2.pp1105-1113.
- [17] B. Antoro, "Analisis Penerapan Formula Slovin Dalam Penelitian Ilmiah: Kelebihan, Kelemahan, Dan Kesalahan Dalam Perspektif Statistik," *Jurnal Multidisiplin Sosisal Humaniora*, vol. 1, no. 2, 2024, [Online]. Available: <https://jurnal.ananpublisher.com/index.php/jmsh>
- [18] A. Novianto Margarena *et al.*, "Conversion Rate: Digital Analytics on Online Food Delivery Platform Customers," *Jurnal Akuntansi, Ekonomi dan Manajemen Bisnis*, vol. 11, no. 2, pp. 160–169, 2023.