

# Pengembangan Data Warehouse Berdasarkan Data Histori Penjualan dan Persediaan Barang di Nuhsantara Merchandise

1<sup>st</sup> Akmal Dwi Utomo

*Sistem Informasi  
Telkom University*

Surabaya, Indonesia

[akmaldwi@student.telkomuniversity.ac.id](mailto:akmaldwi@student.telkomuniversity.ac.id)  
id

2<sup>nd</sup> Sri Hidayati

*Sistem Informasi  
Telkom University*

Surabaya, Indonesia

[srihidayati@telkomuniversity.ac.id](mailto:srihidayati@telkomuniversity.ac.id)

3<sup>rd</sup> Agus Sulisty

*Sistem Informasi  
Telkom University*

Surabaya, Indonesia

[sulisty@telkomuniversity.ac.id](mailto:sulisty@telkomuniversity.ac.id)

**Abstrak -** Dalam dunia bisnis modern yang semakin digital, pengelolaan data penjualan dan persediaan secara terintegrasi dan analitis menjadi kebutuhan penting bagi perusahaan ritel seperti Nuhsantara Merchandise. Perusahaan ini menghadapi tantangan dalam mengelola data penjualan dan persediaan barang dari berbagai platform secara manual, serta belum memiliki sistem yang mampu mengidentifikasi pola pembelian pelanggan sebagai dasar strategi rekomendasi dan promosi produk. Oleh karena itu penelitian ini bertujuan untuk mengembangkan sistem data warehouse menggunakan metode Nine Step Kimball, mengembangkan dashboard visual interaktif dengan Power BI, serta menerapkan algoritma *FP-Growth* untuk mengidentifikasi pola pembelian pelanggan berdasarkan data transaksi aktual. Metode penelitian ini mencakup tahapan pengembangan data warehouse berbasis skema bintang, proses ETL menggunakan Pentaho Data Integration, dan visualisasi data OLAP dalam Power BI. Selain itu, dilakukan analisis asosiasi menggunakan algoritma *FP-Growth* dalam platform Streamlit, dengan parameter minimum support sebesar 0.02 dan confidence sebesar 0.7. Hasil analisis dari 48.030 transaksi menghasilkan 287 *frequent itemsets*, 184 *association rules*, dan 10 *simplified rules* yang digunakan sebagai dasar rekomendasi produk secara otomatis. Hasil divisualisasikan dalam bentuk tabel dan kartu interaktif memungkinkan interpretasi hasil yang mudah dipahami oleh pengguna non-teknis. Sistem yang dibangun terbukti mampu meningkatkan efisiensi dalam pemantauan penjualan, pengelolaan stok, serta memberikan rekomendasi strategis untuk promosi dan bundling produk. Dengan demikian, sistem ini dapat menjadi solusi untuk mendukung pengambilan keputusan manajerial berbasis data aktual dan memberikan nilai tambah bagi perusahaan.

**Kata kunci—** Data Warehouse, Nine Step Kimball, Power BI, *FP-Growth*, Streamlit, Sistem Rekomendasi.

## I. PENDAHULUAN

Dalam dunia bisnis modern yang semakin digital, pemanfaatan data telah menjadi kunci utama dalam menyusun strategi yang efektif, khususnya sektor ritel yang dinamis dan kompetitif. Pemanfaatan data yang tepat memungkinkan perusahaan menyusun strategi penjualan, pengelolaan stok, hingga promosi dengan lebih akurat berbasis fakta. Salah satu teknologi yang berperan besar adalah data *warehouse*, yang mampu menyediakan informasi akurat untuk mendukung analisis serta pengambilan keputusan, khususnya terkait penjualan dan pengelolaan

persediaan barang. Sistem data *warehouse* mampu membantu meningkatkan pengolahan data transaksi dengan menerapkan teknik pengoptimalan *database*, yang sangat diperlukan di sektor ritel [1]. Dengan dukungan data yang terintegrasi, perusahaan seperti Nuhsantara Merchandise dapat memperoleh wawasan yang lebih baik mengenai tren penjualan dan persediaan barang.

Nuhsantara Merchandise, sebuah perusahaan ritel di Surabaya yang berfokus pada penjualan suvenir, menghadapi tantangan dalam mengelola dan memantau data penjualan serta persediaan barang secara efisien. Berdasarkan wawancara dengan Manager Gudang perusahaan, diketahui bahwa pencatatan transaksi masih dilakukan secara manual melalui Microsoft Excel, tanpa sistem terintegrasi antar platform penjualan. Kondisi ini menyulitkan manajemen dalam memperoleh gambaran menyeluruh mengenai performa penjualan dan kondisi stok, sehingga kerap terjadi ketidakseimbangan stok seperti overstock maupun stockout yang berdampak langsung pada efisiensi operasional dan kepuasan pelanggan. Data yang digunakan dalam penelitian ini mencakup histori transaksi penjualan dan persediaan barang selama Agustus 2024 hingga Januari 2025, yang terdiri dari 170.237 transaksi penjualan, 5.684 data barang masuk, dan 495 data barang keluar.

Selain tantangan dalam pengelolaan stok, perusahaan juga belum pernah melakukan analisis terhadap pola pembelian pelanggan secara menyeluruh, khususnya untuk mengetahui produk yang sering dibeli bersamaan. Selama ini, evaluasi penjualan hanya dilakukan berdasarkan kuantitas penjualan produk individu, tanpa memperhatikan keterkaitan antar produk. Padahal, analisis asosiasi produk dapat memberikan wawasan penting untuk pengembangan strategi promosi seperti bundling, rekomendasi produk, dan pengaturan display yang lebih efektif. Untuk mengatasi hal tersebut, algoritma *FP-Growth* dipilih karena mampu mengidentifikasi kombinasi item yang sering muncul dalam transaksi tanpa proses pembangkitan kandidat seperti *Apriori* [3]. Beberapa studi terdahulu menunjukkan bahwa penerapan *FP-Growth* dapat meningkatkan efektivitas strategi penjualan dan efisiensi pengelolaan stok [4], [5].

Penelitian ini menggunakan metode Nine Step Kimball yang memiliki pendekatan terstruktur untuk mengembangkan data warehouse sesuai kebutuhan bisnis ritel. Selain itu,

algoritma *FP-Growth* diterapkan untuk menemukan pola asosiasi pembelian dari data histori penjualan selama enam bulan terakhir. Kombinasi dari kedua pendekatan ini diharapkan menghasilkan sistem analitik yang komprehensif, aplikatif, dan dapat mendukung manajemen dalam merumuskan strategi penjualan dan pengelolaan stok. Melalui sistem visualisasi dan integrasi data yang akurat, hasil penelitian ini diharapkan dapat meminimalkan risiko overstock dan stockout, sekaligus meningkatkan efisiensi operasional serta daya saing Nuhsantara Merchandise di tengah persaingan industri ritel yang kompetitif.

## II. KAJIAN TEORI

### A. Data Warehouse

Definisi dari data *warehouse* menurut Ralph Kimball adalah sebuah sistem yang mengumpulkan informasi dari berbagai sumber data, membersihkan data tersebut, menyesuaikan formatnya, dan menyimpannya ke dalam tempat penyimpanan yang bersifat dimensional, yang kemudian mendukung pelaksanaan *query* dan analisis untuk keperluan pembuatan keputusan [8]. Data *warehouse* mengumpulkan informasi dari berbagai sumber, menyamakan formatnya, lalu mengolahnya untuk analisis data melalui dimensi data [9]. Terdapat empat karakteristik utama dari data *warehouse*, yaitu [10]:

1. **Berorientasi Subjek**, yaitu pengembangan data *warehouse* dengan fokus pada subjek utama, seperti pelanggan, barang, pengiriman, penjualan, dan lain-lain. Informasi yang tidak relevan dalam analisis akan dihapus.
2. **Terintegrasi**, artinya data *warehouse* berasal dari berbagai sumber data seperti database transaksi yang saling terkait, berkas excel atau csv, serta data transaksi lain yang berasal dari platform online.
3. **Time Variant**, data yang tersimpan dalam data *warehouse* adalah data yang berasal dari periode waktu yang panjang untuk memastikan keakuratan data yang dihasilkan.
4. **Non Volatile**, data *warehouse* tidak sama dengan database transaksi, karena data *warehouse* hanya berfungsi untuk membaca data, bukan untuk memperbarui atau menghapusnya.

### B. Metode Nine Step Kimball

*Nine Step Kimball* merupakan pendekatan untuk mengembangkan data *warehouse* yang diciptakan oleh Kimball dan Ross, yang bertujuan untuk membantu pengembang dalam membangun data *warehouse* secara bertahap dari bawah ke atas dengan efektif [8]. Manfaat dari penerapan metode *Nine Step Kimball* mencakup desain yang mudah dimengerti oleh pengguna bisnis, kebutuhan sumber daya manusia yang lebih sedikit untuk membangun dan menjaga operasional data *warehouse* secara efisien, serta biaya awal pengembangan yang relatif rendah dengan pengeluaran berikutnya yang tidak terlalu berbeda [11]. Metode *Nine Step Kimball* meliputi 9 langkah sebagai berikut [10]:

- a) *Choosing the process* (Pemilihan proses)
- b) *Choosing the grain* (Pemilihan grain)
- c) *Identifying and conforming the dimensions* (Identifikasi dan penyesuaian dimensi)
- d) *Choosing the facts* (Pemilihan fakta)

- e) *Storing pre-calculation in the fact table* (Menyimpan pra-kalkulasi dalam tabel fakta)
- f) *Rounding out the dimensions table* (Menyempurnakan tabel dimensi)
- g) *Choosing the duration of the database* (Menentukan durasi basis data)
- h) *Tracking slowly changing dimensions* (Melacak perubahan dimensi secara lambat)
- i) *Deciding the query priorities and the query model* (Menentukan prioritas kueri dan model kueri)

### C. Extract Transform Load (ETL)

Proses ETL adalah langkah pertama yang diambil untuk membuat data *warehouse*, di mana pada tahap ini data dipindahkan dari *database* operasional menuju data *warehouse*. Proses ini akan dilakukan secara berkala untuk memperbarui data *warehouse* [12]. Proses pengembangan *Extract, Transform, Load* (ETL) dilaksanakan sebelum pengembangan data *warehouse* selesai. Tujuan dari proses ETL (*Extract, Transform, Load*) adalah untuk mengumpulkan data dari berbagai sumber, mengubah format dan strukturnya, serta memuatnya ke dalam data *warehouse* atau sistem penyimpanan data lainnya [13].

### D. Online Analytical Processing (OLAP)

OLAP (*Online Analytical Processing*) merupakan cara untuk memberikan jawaban dari permintaan analisis yang bisa mengumpulkan, menyimpan, dan mengolah data multidimensi untuk analisis dengan cepat [14]. Dalam analisis data, OLAP berperan penting dalam data *warehouse* dan *business intelligence* untuk mempermudah analisis data karena data disajikan dalam beberapa dimensi yang berbeda. OLAP memungkinkan data *warehouse* untuk digunakan secara efisien untuk tujuan analisis *online*, memberikan respon cepat terhadap pertanyaan analitis kompleks yang berulang [15].

### E. Data Mining

Data *mining* adalah sebuah proses yang dapat dilakukan untuk menemukan hubungan dari data yang belum diketahui oleh pengguna dengan menyajikannya dengan cara yang lebih mudah untuk dipahami agar dapat menjadi dasar dalam sebuah pengambilan keputusan [16]. Proses ini sering disebut sebagai *knowledge discovery* atau *pengenalan pola* karena bertujuan untuk menemukan pola-pola tersembunyi dalam data yang besar dan kompleks [17]. Data *mining* juga merupakan istilah yang digunakan untuk menemukan pengetahuan tersembunyi di dalam *database* dengan memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin [18]. Tujuan utama data *mining* adalah untuk mengekstrak data yang berharga dan mengubahnya menjadi bentuk yang dapat diterima oleh pengguna [19]. Data *mining* digunakan untuk mengekstraksi dan mengidentifikasi informasi berguna dan pengetahuan terikat dari *database* besar guna mendukung pengambilan keputusan strategis [20].

### F. Association Rule Mining

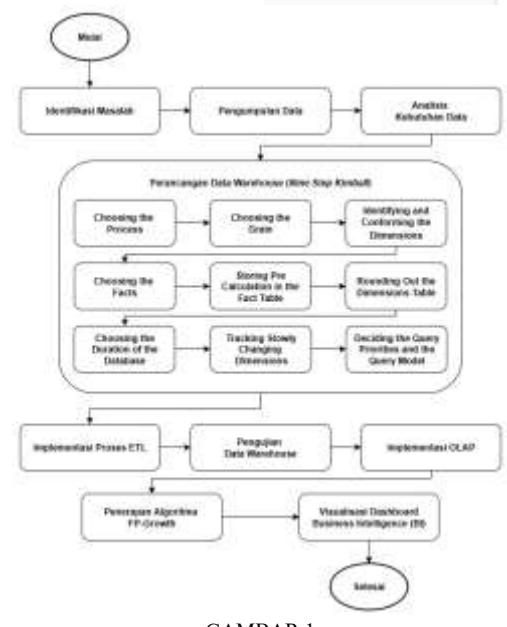
*Association Rule Mining* (ARM) merupakan salah satu teknik dalam bidang *data mining* yang berfungsi untuk mengidentifikasi keterkaitan atau pola hubungan tersembunyi antara item dalam sebuah kumpulan data

transaksi. Teknik ini banyak digunakan dalam analisis keranjang belanja (*market basket analysis*) untuk menemukan pola pembelian barang yang sering terjadi secara bersamaan. Informasi dari hasil analisis ini dapat dimanfaatkan oleh perusahaan untuk menentukan strategi penempatan produk, pengelompokan barang, maupun sistem rekomendasi yang berbasis data historis [5].

### G. Algoritma *FP-Growth*

Algoritma *FP-Growth* adalah salah satu metode dalam *data mining* yang digunakan untuk mengidentifikasi pola itemset dengan frekuensi kemunculan yang tinggi (*frequent itemset*) dalam sekumpulan data transaksi. Berbeda dengan algoritma *Apriori* yang mengharuskan pembangkitan atau iterasi kandidat itemset secara eksplisit, *FP-Growth* membangun struktur pohon khusus yang disebut *Fp-tree* untuk mengekstraksi itemset secara langsung tanpa proses pembangkitan kandidat. Pendekatan ini menjadikan *FP-Growth* lebih efisien dalam pengolahan data berskala besar karena mampu mengurangi kompleksitas perhitungan kombinasi item yang tidak relevan [5]. *FP-Growth* ini merupakan pengembangan dari *Apriori* yang lebih efisien dan skalabel, khususnya untuk mengelola dataset skala besar. *Frequent itemset* dalam *FP-Growth* dilakukan dengan membangun struktur data yang dikenal sebagai *Fp-tree*. Efisiensi *FP-Growth* didukung oleh proses komputasi yang terstruktur melalui tiga tahapan utama, yakni pembentukan *conditional pattern base*, pembangunan *conditional Fp-tree*, dan pencarian *frequent itemset*. Ketiga tahapan tersebut memungkinkan algoritma ini melakukan filtrasi terhadap itemset yang memenuhi nilai ambang minimum *support* dan *confidence* dengan lebih sistematis. Dalam implementasinya, algoritma *FP-Growth* juga didukung oleh pustaka pemrograman seperti *mlxtend*, yang banyak digunakan dalam proses transformasi data untuk keperluan analisis asosiasi [21].

## III. METODE



GAMBAR 1  
(ALUR PENELITIAN)

Sistematika penyelesaian masalah yang digunakan dalam penelitian ini merujuk pada metode Nine Step Kimball yang

terorganisir dan sesuai dengan praktik pengembangan data warehouse. Untuk memperkaya hasil analisis dan mendukung pengambilan keputusan strategis, penelitian ini juga menerapkan metode data mining menggunakan Algoritma *FP-Growth*. Metode ini digunakan untuk menemukan pola asosiasi produk berdasarkan data transaksi historis yang tersimpan dalam data warehouse. Hasil dari seluruh proses tersebut divisualisasikan melalui dashboard Business Intelligence (BI) yang mencakup dua fokus utama, yaitu analisis penjualan dan persediaan produk serta visualisasi hasil analisis *association rule* menggunakan algoritma *FP-Growth*.

Berikut uraian alur penelitian yang diterapkan:

### A. Identifikasi Masalah

Langkah awal dari penelitian ini adalah mengidentifikasi masalah yang dihadapi oleh Nuhsantara Merchandise terkait dengan pengelolaan data penjualan dan persediaan barang. Ditemukan bahwa pengolahan data dilakukan secara manual atau tersebar di berbagai platform *e-commerce* dan penjualan *offline*. Hal ini menyebabkan kesulitan dalam mengintegrasikan data dari sumber yang berbeda, yang berakibat pada rendahnya efisiensi dalam pengambilan keputusan berdasarkan data. Perusahaan juga mengalami kesulitan dalam memantau kinerja persediaan barang secara efektif, yang mengakibatkan seringnya terjadi kelebihan atau kekurangan dalam hal persediaan barang. Selain itu, perusahaan juga kesulitan dalam memahami pola pembelian produk yang dapat digunakan untuk strategi *bundling* atau promosi, perusahaan belum memiliki sistem analitik untuk menyajikan insight mendalam. Hal ini menyulitkan dalam membuat strategi *bundling* dan rekomendasi produk. Masalah ini menjadi landasan untuk mengembangkan solusi berupa data *warehouse* dan *dashboard* analitik yang dapat menyediakan laporan dan analisis yang terintegrasi untuk mendukung pengambilan keputusan yang lebih baik.

### B. Pengumpulan Data

Setelah masalah teridentifikasi, langkah selanjutnya adalah mengumpulkan data mengenai penjualan dan persediaan barang dari berbagai platform yang digunakan oleh perusahaan, seperti Shopee, Tokopedia, dan saluran penjualan lainnya. Proses pengumpulan data melibatkan pengambilan data dari laporan transaksi yang disediakan oleh platform tersebut, lalu menggabungkannya ke dalam format yang seragam. Teknik pengumpulan data yang digunakan dalam penelitian ini adalah observasi dan wawancara. Observasi dilakukan dengan mengamati dan menganalisis aspek yang terkait dengan gudang dan proses bisnis di Nuhsantara Merchandise, sedangkan wawancara dilakukan dengan berkomunikasi langsung dengan Bapak Mario Sebastian Avelino selaku manager gudang pada tanggal 26 November 2024 di Nuhsantara Merchandise. Data yang dipakai pada penelitian ini yaitu data penjualan dan persediaan barang di Nuhsantara Merchandise.

### C. Analisis Kebutuhan Data

Berikut adalah dokumentasi yang mendeskripsikan kebutuhan data, dokumentasi ini berisi rincian mengenai spesifikasi data untuk mendukung format database yang diinginkan. Tujuan dari dokumentasi ini yaitu untuk

menyajikan informasi data yang diperlukan secara terorganisir dan mendetail, sehingga memudahkan peneliti dalam mengumpulkan data serupa. Selain itu, keberadaan dokumentasi ini akan memfasilitasi perbandingan antara data yang sebenarnya dengan kebutuhan yang telah ditetapkan. Berikut ini merupakan sampel struktur data penjualan, data barang masuk, dan barang keluar:

No	No Transaksi	Alasan	Item	No Trans	Tanggal	Kode Produk	Nama Produk	Kode Dimensi	Nama Dimensi	No
1	144520240829 9707	BUKU	POKE	81231011	8/11/2024	AB123456	POKE B	PER001	PER001	1
2	144520240829 9707	LAUDAN	POC	81231011	8/11/2024	AB123456	POKE B	PER001	PER001	1
3	145120240829 9707	TOPENG	TAKE	81231011	8/11/2024	TAKE001	TOPENG	PER001	PER001	1

GAMBAR 2  
(SAMPEL DATA PENJUALAN)

No	Kd	Tanggal	Item	No Transaksi	Tanggal	Kode Produk	Nama Produk	Qty	Net	No
1	AB123456	8/11/2024	POKE	144520240829 9707	8/11/2024	AB123456	POKE B	1	100	1
2	AB123456	8/11/2024	LAUDAN	144520240829 9707	8/11/2024	AB123456	POKE B	1	100	2
3	TAKE001	8/11/2024	TOPENG	145120240829 9707	8/11/2024	TAKE001	TOPENG	1	300	3
4	TAKE001	8/11/2024	TOPENG	145120240829 9707	8/11/2024	TAKE001	TOPENG	1	300	4

GAMBAR 3  
(SAMPEL DATA BARANG MASUK)

No	Kd	Tanggal	Item	No Transaksi	Tanggal	Kode Produk	Nama Produk	Qty	Net	No
1	AB123456	8/11/2024	POKE	144520240829 9707	8/11/2024	AB123456	POKE B	100	10000	1
2	TAKE001	8/11/2024	TOPENG	145120240829 9707	8/11/2024	TAKE001	TOPENG	300	30000	2
3	TAKE001	8/11/2024	TOPENG	145120240829 9707	8/11/2024	TAKE001	TOPENG	300	30000	3
4	TAKE001	8/11/2024	TOPENG	145120240829 9707	8/11/2024	TAKE001	TOPENG	300	30000	4

GAMBAR 4  
(SAMPEL DATA BARANG KELUAR)

#### D. Perancangan Data Warehouse (Nine Step Kimball)

Dalam penelitian ini, perancangan data *warehouse* dilakukan dengan mengikuti pendekatan *Nine Step Kimball*. Metode ini dipilih karena dapat memberikan struktur yang jelas dalam mengembangkan data *warehouse*, dari pemilihan proses hingga menentukan prioritas kueri.

#### E. Implementasi Proses ETL

Pada tahap *Extract*, aktivitas ini bertujuan untuk mendapatkan data dari sumber data tertentu, yang dalam konteks ini mencakup data persediaan dan penjualan barang yang terjadi selama enam bulan, yaitu dari Agustus 2024 hingga Januari 2025 dalam format .xlsx. Setelah itu, proses transformasi dilakukan dengan menggunakan Pentaho Data Integration untuk mengonversi data dari sumber ke format yang tepat bagi data *warehouse*. Kemudian, di tahap *Loading*, data yang telah diproses akan dimasukkan ke dalam MySQL. *Database* akan ditempatkan pada server lokal menggunakan MySQL untuk mendukung data *warehouse*. Lokasi ini dipilih untuk efisiensi dan keamanan data.

#### F. Pengujian Data Warehouse

Pengujian data *warehouse* dilakukan untuk memastikan bahwa proses ETL yang telah dibangun berhasil menghasilkan data yang valid, bersih, konsisten, dan sesuai dengan struktur star schema. Selain itu, pengujian ini bertujuan untuk memverifikasi integritas antar tabel dan mengevaluasi performa sistem ETL dalam memuat data historis dari sumber menuju data *warehouse*.

#### G. Implementasi OLAP

Setelah menyelesaikan tahapan pengujian data *warehouse*, peneliti akan melanjutkan ke tahap OLAP untuk melakukan analisis data secara multidimensi. Data dalam data *warehouse* akan diproses menggunakan skema bintang, yang meliputi tabel fakta serta tabel dimensi. Peneliti akan melakukan OLAP dengan menggunakan MySQL untuk mendukung eksplorasi data, contohnya untuk menentukan total penjualan berdasarkan waktu atau menganalisis tren penjualan produk tertentu. Meskipun tidak menggunakan dedicated OLAP engine seperti SSAS, proses analisis data dilakukan dengan pendekatan OLAP melalui view SQL dan eksplorasi multidimensi pada Power BI. Hasil dari analisis OLAP ini akan diintegrasikan ke dalam Power BI, yang memungkinkan visualisasi interaktif dalam bentuk grafik dan tabel. Proses ini akan membantu dalam memahami pola data demi mendukung pengambilan keputusan di Nuhsantara Merchandise.

#### H. Penerapan Algoritma FP-Growth

Penerapan algoritma *FP-Growth* dalam tahap ini bertujuan untuk menemukan pola keterkaitan antar produk yang sering dibeli secara bersamaan oleh konsumen. Pola tersebut diharapkan dapat memberikan wawasan bagi perusahaan dalam menentukan strategi *bundling* produk atau rekomendasi promosi yang lebih efektif. *FP-Growth* dipilih karena kemampuannya yang efisien dalam menemukan *frequent itemsets* tanpa harus menghasilkan kandidat seperti algoritma Apriori, sehingga lebih efisien untuk data dengan jumlah transaksi yang besar. Proses penerapan *FP-Growth* dilakukan dalam tiga tahapan utama, yaitu *data preparation*, *modeling*, dan *evaluation*.

#### I. Visualisasi Dashboard Business Intelligence (BI)

Dashboard BI yang dibangun memiliki dua fungsi utama:

1. Visualisasi data penjualan dan persediaan, yang dikembangkan dengan Power BI dan kemudian disematkan (embed) ke dalam Streamlit menggunakan iframe. Komponen visualisasinya meliputi tren penjualan dan produk masuk/keluar, performa sales dan marketplace, top produk dan kategori, serta metrik indikator utama seperti total transaksi dan produk terjual.
2. Visualisasi hasil data mining *FP-Growth*, yang sepenuhnya dibangun dengan Streamlit, mencakup fitur unggah file transaksi, filtering data, serta visualisasi hasil *frequent itemset* dan *association rules* dalam bentuk tabel.

## IV. HASIL DAN PEMBAHASAN

#### A. Perancangan Data Warehouse

Dalam mengembangkan data *warehouse* untuk Nuhsantara Merchandise, penelitian ini menggunakan pendekatan Nine Step Kimball. Metode ini dipilih karena mendukung pengembangan data *warehouse* yang efisien dan sesuai dengan kebutuhan analisis bisnis yang berorientasi pada data historis penjualan dan persediaan barang.

Berikut adalah langkah-langkah pengembangan yang digunakan dalam penelitian ini:

##### 1. Choosing the process (Pemilihan proses)

Pada langkah awal ini, pemilihan proses didasarkan pada masalah yang muncul terkait pengelolaan persediaan dan penjualan barang. Permasalahan utamanya adalah pengelolaan stok yang belum optimal dan sulitnya menganalisis tren penjualan, sehingga dapat menyebabkan overstock atau stockout serta kurangnya pemahaman tentang performa produk di berbagai platform. Oleh karena itu, proses yang akan dianalisis dalam penelitian ini mencakup proses persediaan barang dan penjualan barang di Nuhsantara Merchandise.

## 2. Choosing the grain (Pemilihan grain)

Grain yang dipilih harus sesuai dengan kebutuhan analisis bisnis dan mencerminkan data yang akan digunakan untuk proses pengambilan keputusan. Untuk grain dari penjualan barang yaitu setiap transaksi penjualan untuk masing-masing kombinasi waktu, akun, produk, dan sales yang terekam dalam satu baris transaksi dari file hasil integrasi berbagai marketplace dan offline. Sedangkan, untuk grain dari persediaan barang yaitu jumlah barang masuk dan keluar harian untuk setiap produk, berdasarkan tanggal transaksi yang telah dandardisasi.

Setelah menentukan grain, langkah berikutnya yaitu membuat BUS Matrix, yang akan membantu memetakan hubungan antara tabel fakta dan tabel dimensi.

TABEL 1  
(BUS MATRIX)

Fakta / Dimensi	Waktu	Akun	Produk	Sales
Penjualan Barang	✓	✓	✓	✓
Persediaan Barang	✓		✓	

Dari tabel BUS Matrix di atas, dapat dilihat bahwa Fact Penjualan membutuhkan informasi dari Dim\_Waktu, Dim\_Akun, Dim\_Prod, dan Dim\_Sales karena analisis dilakukan berdasarkan tanggal transaksi, marketplace, produk yang dijual, dan sales yang bertanggung jawab. Sementara itu, Fact Persediaan hanya memerlukan Dim\_Waktu dan Dim\_Prod, karena data pergerakan barang hanya dicatat berdasarkan tanggal dan produk, tanpa keterlibatan akun atau sales.

## 3. Identifying and conforming the dimensions (Identifikasi dan penyesuaian dimensi)

Dalam penelitian ini, empat dimensi utama telah diidentifikasi dan disesuaikan agar dapat digunakan secara konsisten dalam berbagai tabel fakta.

Dimensi-dimensi tersebut dijelaskan dalam tabel berikut:

TABEL 2  
(DIMENSI DATA WAREHOUSE)

Dimensi	Keterangan
Dim_Waktu	Menyimpan informasi terkait waktu transaksi, seperti tahun, bulan, tanggal, yang akan digunakan dalam analisis tren penjualan dan perubahan stok.
Dim_Akun	Menyimpan informasi mengenai akun/platform tempat transaksi berlangsung, seperti Shopee,

	Tokopedia, Lazada, TikTok Shop, dan Offline. Dimensi ini membantu dalam perbandingan performa antar marketplace.
Dim_Prod	Menyimpan detail produk yang dijual, termasuk kode produk, nama produk, kategori produk, sehingga memungkinkan untuk analisis segmentasi produk.
Dim_Sales	Menyimpan informasi tentang sales yang melakukan transaksi, untuk analisis performa sales di berbagai platform penjualan.

## 4. Choosing the facts (Pemilihan fakta)

Pada langkah ini, tabel fakta yang dipilih harus mencerminkan grain yang telah ditentukan sebelumnya, yaitu transaksi penjualan dan persediaan barang. Data dalam tabel fakta ini akan digunakan untuk analisis lebih lanjut terkait jumlah penjualan, pergerakan stok, serta status ketersediaan barang.

Berikut adalah tabel fakta yang digunakan dalam penelitian ini:

TABEL 3  
(PEMILIHAN FAKTA)

Fakta	Keterangan
Fact_Penjualan	Menyimpan jumlah barang yang terjual berdasarkan waktu, akun marketplace, produk, dan sales.
Fact_Persediaan	Menyimpan informasi jumlah barang masuk dan jumlah barang keluar berdasarkan waktu dan produk.

## 5. Storing pre-calculation in the fact table (Menyimpan hasil perhitungan sementara dalam tabel fakta)

Dalam penelitian ini, hasil perhitungan yang disimpan dalam tabel fakta mencakup dua proses utama, yaitu penjualan barang dan persediaan barang. Berikut hasil perhitungan yang disimpan dalam tabel fakta dijelaskan sebagai berikut:

### 1. Fakta Penjualan

Tabel fakta penjualan menyimpan jumlah unit produk yang berhasil terjual dari setiap transaksi, beserta informasi marketplace, waktu, produk, dan sales. Nilai ini tidak dihitung ulang di Power BI, melainkan sudah diproses melalui ETL dan langsung disimpan dalam bentuk terstruktur di data warehouse.

### 2. Fakta Persediaan

Tabel fakta persediaan menyimpan informasi hasil agregasi harian dan bulanan, seperti jumlah barang masuk dan barang keluar per produk per tanggal. Data ini dihitung di tahap ETL berdasarkan rekap dari data barang masuk dan barang keluar, sehingga pada saat ditampilkan di dashboard, tidak perlu dilakukan perhitungan ulang.

## 6. Rounding out the dimensions table (Menyempurnakan tabel-tabel dimensi)

Tabel dimensi yang telah disempurnakan dalam penelitian ini terdiri dari dimensi waktu, dimensi akun, dimensi produk, dan dimensi sales.

TABEL 4  
(DETAIL TABEL DIMENSI)

Dimensi	Atribut	Tipe Data	Keterangan
Waktu	id_waktu	int	Primary key dimensi waktu
	tanggal	date	Menyimpan tanggal transaksi
	hari	int	Nomor hari dalam bulan (1-31)
	bulan	int	Bulan dalam angka (1-12)
	tahun	int	Tahun transaksi
	kuartal	int	Kuartal dalam setahun (1-4)
	hari_dalam_tahun	int	Urutan hari dalam satu tahun (1-365/366)
	minggu_dalam_tahun	int	Urutan minggu dalam satu tahun (1-52)
Akun	id_akun	int	Primary key dimensi akun
	nama_akun	varchar	Nama akun marketplace
Produk	id_produk	int	Primary key dimensi produk
	kode_produk	varchar	Kode unik setiap produk
	nama_produk	varchar	Nama lengkap produk
Sales	kategori_produk	varchar	Kategori dari produk
	id_sales	int	Primary key dimensi sales
	nama_sales	varchar	Nama sales yang menangani transaksi

## 7. Choosing the duration of the database (Menentukan durasi penyimpanan data)

Pada penelitian ini, durasi penyimpanan data ditentukan berdasarkan kebutuhan analisis bisnis di Nuhsantara Merchandise. Berdasarkan hasil diskusi dengan tim gudang, ditetapkan bahwa periode penyimpanan data dalam data *warehouse* adalah enam bulan terakhir, yaitu dari Agustus 2024 hingga Januari 2025.

## 8. Tracking slowly changing dimensions (Melacak dimensi yang berubah secara perlahan)

Pada penelitian ini, data *warehouse* yang dibangun untuk Nuhsantara Merchandise memiliki beberapa dimensi yang

perlu dianalisis untuk menentukan apakah data dalam dimensi tersebut berubah secara perlahan dan bagaimana strategi terbaik dalam pengelolaannya. Untuk hasilnya keempat dimensi tidak mengalami perubahan karena berisi atribut yang tepat. Karena tidak ada perubahan dalam struktur data, SCD Type 0 digunakan untuk keempat dimensi.

## 9. Deciding the query priorities and the query model (Menentukan prioritas kueri dan model kueri)

Model kueri yang digunakan terbagi menjadi dua jenis utama, yaitu kueri dinamis dan kueri statis. Keduanya dirancang untuk mendukung eksplorasi data secara multidimensi dalam konteks *Online Analytical Processing* (OLAP), yang memungkinkan pengguna melakukan analisis fleksibel berdasarkan waktu, produk, *marketplace*, dan *sales*.

### 1. Dinamis

Kueri dinamis dibangun langsung dari relasi antara tabel fakta dan dimensi di dalam Power BI. Kueri ini bersifat interaktif dan merespons secara langsung terhadap *slicer* yang digunakan oleh pengguna, sehingga hasil visualisasi akan berubah secara otomatis sesuai filter yang aktif. Kueri jenis ini digunakan untuk menghasilkan visualisasi berikut:

- Total produk terjual, total transaksi, dan rata-rata penjualan per transaksi
- Total produk masuk dan keluar
- Tren penjualan dan tren pergerakan stok berdasarkan bulan
- Penjualan per *sales*
- Performa *marketplace*
- Perbandingan barang masuk dan keluar berdasarkan kategori

### 2. Statis (View SQL)

Beberapa visualisasi dalam *dashboard*, khususnya visualisasi berbasis peringkat (*ranking*), menggunakan kueri statis yang dibuat dalam bentuk *View* di MySQL. *View* ini tidak akan terpengaruh oleh *slicer*, sehingga tetap menampilkan data secara konsisten untuk keseluruhan periode data. *View* ini digunakan untuk visualisasi berikut:

- Top 3 Marketplace
- Top 3 Kategori Terlaris
- Top 10 Produk Terlaris
- Top 3 Produk Masuk dan Produk Keluar

## B. Implementasi Proses ETL

Implementasi proses *Extract, Transform, Load* (ETL) dilakukan menggunakan perangkat lunak Pentaho Data Integration (PDI). Proses ini bertujuan untuk mengekstraksi data dari berbagai *file* sumber (Excel), melakukan transformasi data yang mencakup pembersihan, normalisasi, penggabungan data, hingga pemetaan ke dimensi, dan akhirnya memuat hasil ke dalam tabel fakta dan dimensi di *database* etl\_nm. Setiap transformasi dioptimalkan untuk memastikan data yang masuk ke dalam data *warehouse* sudah bersih, terstruktur, dan siap dianalisis melalui OLAP atau visualisasi *dashboard*. Proses ETL dibagi menjadi dua tahap utama, yaitu pengisian tabel dimensi dan pengisian tabel fakta. Seluruh transformasi dirancang untuk menyesuaikan struktur star schema yang telah dibuat sebelumnya.

Berikut adalah hasil implementasi proses ETL:

Table	Action	Row	Type	Options	Size
dim_akun	10 rows selected, 10 total rows	10	Table	Drop	4.1401 MB
dim_produk	10 rows selected, 10 total rows	10	Table	Drop	1.497 MB
dim_sales	10 rows selected, 10 total rows	10	Table	Drop	1.042 MB
fact_penjualan	10 rows selected, 10 total rows	10	Table	Drop	1.042 MB
fact_persediaan	10 rows selected, 10 total rows	10	Table	Drop	1.042 MB
fact_persediaan	10 rows selected, 10 total rows	10	Table	Drop	1.042 MB
fact_persediaan	10 rows selected, 10 total rows	10	Table	Drop	1.042 MB
fact_persediaan	10 rows selected, 10 total rows	10	Table	Drop	1.042 MB
fact_persediaan	10 rows selected, 10 total rows	10	Table	Drop	1.042 MB

GAMBAR 5  
(HASIL ETL)

Secara keseluruhan, proses ETL menghasilkan enam tabel utama dalam *database etl\_nm*, yang terdiri dari empat tabel dimensi (dim\_akun, dim\_produk, dim\_sales, dim\_waktu) dan dua tabel fakta (fact\_penjualan, fact\_persediaan). Seluruh proses transformasi telah dirancang untuk menghasilkan data yang terintegrasi, bersih, dan siap digunakan untuk proses analisis lanjutan di tahap OLAP maupun *dashboard* visualisasi.

### C. Pengujian Data Warehouse

Pengujian dilakukan untuk memastikan bahwa proses ETL yang telah dibangun menghasilkan data yang valid, konsisten, dan sesuai dengan desain star schema. Pengujian difokuskan pada empat aspek utama:

#### 1. Validitas dan kualitas data yang dimuat ke dalam tabel dimensi dan fakta

TABEL 5

(JUMLAH DATA VALID DAN INVALID PADA TABEL DIMENSI)

Tabel Dimensi	Jumlah Data Awal	Data Invalid	Data Akhir
dim_akun	170.207	111	5
dim_produk	183.600	66	1.497
dim_sales	170.207	110	3
dim_waktu	176.383	111	142

Tabel ini menunjukkan jumlah data awal hasil ekstraksi, jumlah data yang teridentifikasi invalid melalui proses *Filter Rows*, serta jumlah data akhir yang valid untuk dimuat ke dimensi. Hasil ini menunjukkan bahwa mekanisme pembersihan data bekerja dengan efektif untuk menyaring baris kosong (NULL) pada *field* penting seperti nama akun, nama produk, nama sales, tanggal yang tidak valid, dan produk duplikat dan tidak lengkap.

#### 2. Penghapusan duplikasi data

TABEL 6

(HASIL DEDUPLIKASI DATA DIMENSI)

Tabel Dimensi	Jumlah Data Sebelum Unique	Data Unique
dim_akun	170.096	5
dim_produk	183.600	1.497
dim_sales	170.097	3
dim_waktu	176.272	142

#### 3. Integritas referensial antar tabel

MySQL returned an empty result set (i.e. zero rows). (Query took 0.0059 seconds.)
SELECT * FROM fact_persediaan WHERE id_produk IS NULL OR id_waktu IS NULL;
Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

GAMBAR 6  
(PENGUJIAN RELASI FOREIGN KEY)

MySQL returned an empty result set (i.e. zero rows). (Query took 0.0059 seconds.)
SELECT * FROM fact_persediaan WHERE id_produk IS NULL OR id_waktu IS NULL;
id_persediaan id_waktu id_produk produk_masuk produk_keluar

GAMBAR 7

(PENGUJIAN RELASI FOREIGN KEY)

Hasil pengujian menunjukkan bahwa tidak ditemukan satupun baris dengan nilai NULL pada *foreign key*. Ini menjadi bukti bahwa semua data transaksi berhasil dipetakan ke entitas dimensi melalui proses DB *Lookup*, dan integritas referensial antar tabel telah terjaga.

#### 4. Performa proses ETL, dilihat dari waktu eksekusi, volume data, dan kecepatan pemrosesan

TABEL 7

(PERFORMA PROSES ETL)

Transformasi	Total Data Input	Waktu Eksekusi	Kecepatan Rata-rata
dim_akun	170.207 baris	15.5 detik	10.800 r/s
dim_produk	183.600 baris	9.4 detik	19.500 r/s
dim_sales	170.207 baris	18.0 detik	9.400 r/s
dim_waktu	176.383 baris	19.8 detik	8.900 r/s
fact_penjualan	170.207 baris	1m 50 detik	1.300 r/s
fact_persediaan	5.683 + 494 baris	4.0 detik	1.100 r/s
<b>Total</b>	876.781 baris	2 menit 57 detik	51000 r/s
<b>Rata-Rata</b>		29,45 detik	8.500 r/s

#### D. Implementasi OLAP (Online Analytical Processing)

Dalam sistem ini, OLAP diterapkan melalui Power BI, yang mengakses langsung tabel-tabel fakta dan dimensi dari data warehouse yang telah dibentuk sebelumnya melalui proses ETL. Visualisasi interaktif yang ditampilkan dalam dashboard Power BI dibangun berdasarkan model relasional antara fact dan dimension table. Interaksi pengguna, seperti penggunaan slicer (penyaringan data berdasarkan waktu, produk, kategori, dan akun marketplace), memungkinkan perubahan tampilan visualisasi secara dinamis berdasarkan filter yang aktif. Hal ini menunjukkan karakteristik utama dari OLAP, yaitu fleksibilitas dalam eksplorasi data untuk mendeteksi pola, tren, dan insight penting dari data historis penjualan dan persediaan barang.

Selain interaksi dinamis melalui slicer, sistem juga mengimplementasikan kueri statis yang dibangun dalam bentuk View SQL. Kueri statis ini digunakan untuk menampilkan visualisasi berbasis peringkat (ranking), seperti Top 3 Marketplace, Top 10 Produk Terlaris, Top 3 Kategori, Top 3 Produk Masuk, dan Top 3 Produk Keluar. Tujuan penggunaan view statis adalah agar hasil visualisasi tetap konsisten meskipun pengguna mengaktifkan filter pada dashboard. View SQL dirancang menggunakan agregasi dan urutan (GROUP BY, ORDER BY, LIMIT) untuk

menghasilkan informasi strategis yang tidak berubah-ubah, dan tetap dapat dijadikan acuan tetap oleh manajemen.

### E. Penerapan Algoritma *FP-Growth*

Pada tahap ini, dilakukan penerapan algoritma *FrequentPattern Growth (FP-Growth)* untuk menganalisis pola asosiasi pembelian konsumen berdasarkan kombinasi produk yang sering muncul bersamaan dalam satu transaksi di Nuhsantara Merchandise.

#### 1. Data Preparation

Data yang digunakan berasal dari data histori tabel transaksi penjualan yang telah dimuat ke dalam data warehouse selama periode Agustus 2024 hingga Januari 2025. Total data transaksi mencakup 161.664 baris dan 48.030 transaksi unik, yang diekspor dari fact\_penjualan. Untuk keperluan data mining, data tersebut diekspor dan diproses dalam bentuk dataset yang berisi dua atribut utama, yaitu:

- a) no\_transaksi (ID transaksi)
- b) nama\_produk atau kategori\_produk (bergantung pada parameter analisis)
- c) tanggal (opsional jika ingin filter waktu)

#### 2. Modeling

Proses modeling dilakukan dengan menerapkan algoritma *FrequentPattern Growth (FP-Growth)* menggunakan pustaka mlxtend.frequent\_patterns untuk menemukan pola asosiasi antar produk berdasarkan kombinasi item yang sering muncul bersama dalam satu transaksi. Model dibangun dengan memanfaatkan hasil transformasi data yang telah melalui proses one-hot encoding, sehingga setiap transaksi direpresentasikan dalam format biner.

Pemodelan ini melibatkan dua parameter utama yang harus ditentukan sebelum proses pelatihan model dilakukan, yaitu:

- a) **Minimum Support:** Merupakan nilai ambang batas minimum kemunculan kombinasi produk dalam keseluruhan data transaksi. Kombinasi produk dengan support lebih tinggi dari nilai ini akan dianggap sebagai *frequent itemset*.
- b) **Minimum Confidence:** Menunjukkan tingkat keyakinan atau probabilitas bahwa suatu produk akan dibeli bersamaan dengan produk lainnya. Hanya aturan dengan confidence melebihi nilai ini yang akan diambil sebagai *association rules*.

Kedua parameter tersebut sangat penting karena berperan sebagai threshold atau batasan untuk menyaring hasil yang relevan. Nilai support dan confidence yang terlalu rendah dapat menghasilkan terlalu banyak aturan yang tidak signifikan, sementara nilai yang terlalu tinggi berpotensi menghilangkan pola penting yang frekuensinya tidak dominan. Eksperimen dilakukan dengan memvariasikan nilai minimum support mulai dari 0,01 hingga 0,05, sedangkan nilai minimum confidence tetap dijaga konsisten pada 0,7. Pemilihan nilai confidence sebesar 0,7 didasarkan pada praktik umum dan penelitian terdahulu yang menunjukkan bahwa rentang ideal confidence berkisar antara 0,6 hingga 0,8 untuk menghasilkan aturan yang bermakna namun tidak terlalu ketat.

Tujuan dari eksperimen ini adalah untuk menentukan konfigurasi parameter yang paling optimal, yaitu

menghasilkan jumlah aturan asosiasi yang memadai namun tetap relevan dan dapat diinterpretasikan dengan mudah oleh pengguna. Seluruh hasil eksperimen dicatat dan dianalisis secara kuantitatif pada tahap evaluasi berikut.

### 3. Evaluation dan Interpretasi Hasil

Tahap evaluasi dilakukan dengan tujuan untuk membandingkan hasil *frequent itemsets* dan *association rules* yang dihasilkan dari berbagai nilai *minimum support*, serta menilai kualitas dan kuantitas aturan yang terbentuk. Fokus evaluasi diarahkan pada tiga aspek utama yaitu jumlah *frequent itemsets*, jumlah *association rules*, dan jumlah *simplified rules* (aturan 1 produk  $\rightarrow$  1 produk), serta dilengkapi dengan visualisasi rekomendasi produk dan interpretasi pola pembelian.

Berikut adalah rangkuman hasil eksperimen:

- a) **Evaluasi 1**, dengan min support 0.01 dan min confidence 0.7 menghasilkan 1.149 *frequent itemsets*, 2.882 *association rules*, dan 11 *simplified rules*. Jumlah aturan yang sangat besar menyebabkan sistem menghasilkan warning karena beban komputasi yang tinggi. Meskipun *simplified rules* tetap dalam jumlah terbatas, banyaknya aturan keseluruhan berpotensi menghasilkan noise dan menyulitkan interpretasi. Hal ini menyulitkan analisis, rawan noise, dan secara teknis menimbulkan warning saat diproses Python.
- b) **Evaluasi 2**, dengan min support 0.02 dan min confidence 0.7 menghasilkan 287 *frequent itemsets*, 184 *association rules*, dan 10 *simplified rules*. Konfigurasi ini memberikan keseimbangan ideal antara kuantitas dan keterbacaan aturan. Jumlah *simplified rules* mendekati maksimum dengan kompleksitas yang masih dapat diterima untuk analisis bisnis. Struktur aturan asosiasi lebih lengkap, memungkinkan kombinasi multi-produk seperti 4-in-1 atau 5-in-1 bundling. Banyak aturan yang menyertakan 2 hingga 4 produk awal (antecedents) terhadap 1 produk target (consequent), sangat relevan untuk strategi cross-selling.
- c) **Evaluasi 3**, dengan min support 0.03 dan min confidence 0.7 menghasilkan 120 *frequent itemsets*, 49 *association rules*, dan 10 *simplified rules*. Meskipun *simplified rules* masih optimal, total aturan asosiasi mulai menurun drastis sehingga berisiko kehilangan pola penting dari data. Struktur kombinasi produk terbatas hanya pada 2 produk terhadap 1 produk. Hal ini menghambat pembuatan rekomendasi bundling yang kompleks seperti 4-in-1. *Association rules* yang dihasilkan terlalu sempit dan kurang fleksibel untuk aplikasi bisnis yang memerlukan penawaran kolektif atau paket produk edukatif terintegrasi.
- d) **Evaluasi 4**, dengan min support 0.04 dan min confidence 0.7 menghasilkan 68 *frequent itemsets*, 14 *association rules*, dan 7 *simplified rules*. Jumlah aturan terlalu sedikit untuk dapat mewakili kompleksitas pola pembelian pelanggan, sehingga dianggap kurang representatif. Potensi analisis sangat terbatas.

e) **Evaluasi 5**, dengan min support 0.05 dan min confidence 0.7 menghasilkan 32 *frequent itemsets*, 11 *association rules*, dan 6 *simplified rules*. Aturan yang terbentuk semakin sedikit, dan variasi produk yang muncul dalam kombinasi pun makin terbatas. Risiko kehilangan informasi strategis meningkat signifikan. Tidak cukup untuk analisis menyeluruh.

Berdasarkan hasil kelima evaluasi di atas, pemilihan nilai minimum support sebesar 0.02 dan minimum confidence sebesar 0.7 diputuskan sebagai konfigurasi paling optimal. Pemilihan konfigurasi tersebut dikarenakan keseimbangan antara kuantitas dan kualitas data, efektivitas informasi pada *simplified rules*, relevansi frekuensi, dan menghindari potensi noise. Konfigurasi tersebut menghasilkan aturan asosiasi yang informatif, kuat, dan terstruktur kompleks, mendukung strategi bundling 3-in-1 hingga 5-in-1 yang relevan secara bisnis, tetap menjaga keterbacaan dan interpretabilitas hasil oleh pengguna non-teknis, Menghindari kelebihan aturan (overload) tanpa mengorbankan nilai analitik. Konfigurasi ini menjadi dasar utama dalam sistem rekomendasi produk edukatif dan penyusunan strategi promosi berdasarkan pola pembelian aktual.

Berikut merupakan hasil *frequent itemset*, *association rules*, dan *simplified rules* dari penentuan parameter minimum support sebesar 0.02 (berarti sebuah kombinasi produk harus muncul minimal 960 kali) serta minimum confidence sebesar 0.7 terhadap 48.030 transaksi, sistem menghasilkan:

### 1. Frequent itemsets

TABEL 8  
(FREQUENT ITEMSETS)

No	Support	Kombinasi Produk	Panjang Kombinasi	Frekuensi
1	0.1676	Flash Card Edukasi Anak Isi 28 Pcs Huruf Alfabet	1	8050
2	0.1467	Flash Card Edukasi Anak Isi 28 Pcs Hewan Air & Darat	1	7045
3	0.1368	Flash Card Edukasi Anak Isi 28 Pcs Angka	1	6569
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
287	0.02	Flash Card Edukasi Anak Isi 28 Pcs Angka, Flash Card Edukasi Anak Isi 28 Pcs Benda Di Sekitar, Flash Card Edukasi Anak Isi 28 Pcs Profesi, Flash Card Edukasi Anak Isi 28 Pcs Transportasi	4	961

### 2. Association Rules

TABEL 9  
(ASSOCIATION RULES)

No	Produk Awal	Produk Disarankan	Support	Confidence	Lift
1	Poster Edukasi Bahan Kertas Art Paper Hewan Karnivora, Poster Edukasi Bahan Kertas Art Paper Transportasi	Poster Edukasi Bahan Kertas Art Paper Hewan Herbivora	0.0206	0.9007	22.4039
2	Poster Edukasi Bahan Kertas Art Paper Hewan Karnivora	Poster Edukasi Bahan Kertas Art Paper Hewan Herbivora	0.0306	0.8867	22.0562
3	Flash Card Edukasi Anak Isi 28 Pcs Angka, Flash Card Edukasi Anak Isi 28 Pcs Huruf Hijaiyah, Flash Card Edukasi Anak Isi 28 Pcs Transportasi	Flash Card Edukasi Anak Isi 28 Pcs Huruf Alfabet	0.0238	0.8764	5.2292
.	.	.	.	.	.
18	Poster Edukasi Bahan Kertas Art Paper Angka, Poster Edukasi Bahan Kertas Art Paper Hewan	Poster Edukasi Bahan Kertas Art Paper Huruf Hijaiyah	0.0317	0.7	6.1122

### 3. Simplified Rules

TABEL 10  
(SIMPLIFIED RULES)

No	Produk Awal	Produk Disarankan	Support	Confidence	Lift
1	Poster Edukasi Bahan Kertas Art Paper Hewan Karnivora	Poster Edukasi Bahan Kertas Art Paper Hewan Herbivora	0.0206	0.9007	22.4039
2	Poster Edukasi Bahan	Poster Edukasi Bahan	0.0306	0.8867	22.0562

	Kertas Art Paper Angka	Kertas Art Paper Huruf Alfabet			
3	Poster Edukasi Bahan Kertas Art Paper Hewan Herbivora	Poster Edukasi Bahan Kertas Art Paper Hewan Karnivora	0.0238	0.8764	5.2292
4	Poster Edukasi Bahan Kertas Art Paper Huruf Alfabet	Poster Edukasi Bahan Kertas Art Paper Angka	0.0925	0.7525	6.3426
5	Poster Edukasi Bahan Kertas Art Paper Huruf Hijaiyah	Poster Edukasi Bahan Kertas Art Paper Huruf Alfabet	0.0849	0.7413	6.0297
6	Poster Edukasi Bahan Kertas Art Paper Sayur	Poster Edukasi Bahan Kertas Art Paper Buah	0.0301	0.7405	9.9996
7	Flash Card Edukasi Anak Isi 28 Pcs Huruf Hijaiyah	Flash Card Edukasi Anak Isi 28 Pcs Huruf Alfabet	0.0862	0.7151	4.2672
8	Poster Edukasi Bahan Kertas Art Paper Huruf Hijaiyah	Poster Edukasi Bahan Kertas Art Paper Angka	0.0814	0.7111	5.9933
9	Poster Edukasi Bahan Kertas Art Paper Warna & Bentuk	Poster Edukasi Bahan Kertas Art Paper Angka	0.0420	0.7070	5.9591
10	Flash Card Edukasi Anak Isi 28 Pcs Angka	Flash Card Edukasi Anak Isi 28 Pcs Huruf Alfabet	0.0961	0.7028	4.1935

## F. Visualisasi Dashboard Business Intelligence (BI)

Visualisasi *dashboard Business Intelligence* (BI) dalam penelitian ini bertujuan untuk memberikan tampilan visual yang interaktif atas data penjualan dan persediaan barang di Nuhsantara Merchandise. Sistem ini dibangun menggunakan dua platform utama, yaitu Power BI untuk *dashboard Sales & Inventory*, serta Streamlit untuk *dashboard* analisis *FP-Growth*. *Dashboard* dirancang agar dapat digunakan oleh pihak manajemen dan operasional perusahaan dalam mengevaluasi performa penjualan, memantau efisiensi distribusi barang, serta mengidentifikasi peluang bisnis dari pola pembelian pelanggan.

Berikut adalah hasil visualisasi *dashboard* business intelligence:

### Selamat Datang di Sistem Analisis Data

#### Nuhsantara Merchandise

Dengan ini meresmiakan sistem analisis data untuk penjualan barang. Data dan inventori yang dapat diakses melalui dashboard, perbaikan sistem distribusi dan sales serta optimasi.

Apakah anda...

PP Gudang

Distributor

Supplier

Analyst

Manager

Analyst



GAMBAR 8  
(LANDING PAGE)

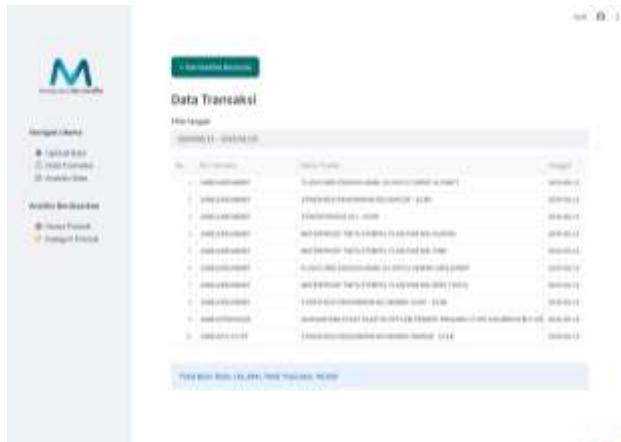
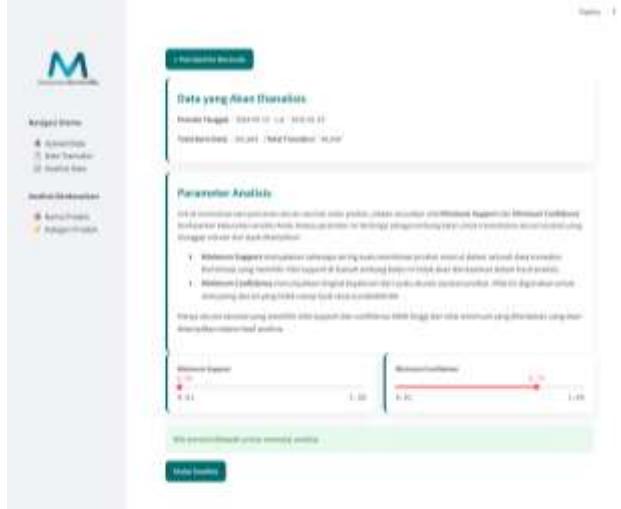


GAMBAR 9  
(INVENTORY)



GAMBAR 10  
(SALES)

GAMBAR 11  
(UPLOAD FILE)

GAMBAR 12  
(DATA TRANSAKSI)GAMBAR 13  
(VALIDASI DAN PARAMETER ANALISIS DATA)GAMBAR 14  
(FREQUENT ITEMSETS)GAMBAR 15  
(ASSOCIATION RULES)GAMBAR 16  
(REKOMENDASI KOMBINASI PRODUK TOP 10)GAMBAR 17  
(SIMPLIFIED RULES)GAMBAR 18  
(REKOMENDASI PRODUK SIMPLIFIED RULES TOP 10)GAMBAR 19  
(REKOMENDASI KOMBINASI PRODUK TANPA PENJELASAN TOP 10)

## V. KESIMPULAN

Penelitian ini dilakukan dengan tujuan untuk mengembangkan sistem *data warehouse* dan menganalisis pola asosiasi pembelian produk menggunakan algoritma *FrequentPattern Growth (FP-Growth)*, dengan studi kasus pada Nuhsantara Merchandise. Berdasarkan implementasi sistem dan pengujian menyeluruh, diperoleh kesimpulan bahwa pengembangan data warehouse berhasil dilakukan menggunakan metode Nine Step Kimball dan dibangun dalam bentuk star schema yang terdiri dari dua tabel fakta dan empat tabel dimensi. Proses ETL berjalan efektif dan efisien, menghasilkan data bersih, valid, tanpa duplikasi, dan memiliki integritas referensial yang baik. Sistem ini telah memenuhi empat karakteristik utama data warehouse.

Selain itu, algoritma *FP-Growth* efektif dalam mengeksplorasi pola asosiasi dari 48.030 transaksi unik.

Setelah dilakukan eksperimen parameter, konfigurasi terbaik ditetapkan pada minimum support 0,02 dan minimum confidence 0,7 yang menghasilkan 287 *frequent itemsets*, 184 *association rules*, dan 10 *simplified rules*. Nilai ini dianggap optimal dari sisi jumlah pola, kemudahan interpretasi, serta relevansi bisnis. Hasil *association rules* memperlihatkan keterkaitan kuat antar produk edukatif anak, terutama flash card dan poster. Temuan ini membuka peluang strategi bundling, seperti paket 3-in-1 hingga 5-in-1. Selain itu, aturan *simplified rules* yang hanya melibatkan satu produk awal dan satu produk disarankan juga mendukung penerapan rekomendasi otomatis yang lebih sederhana dan langsung.

Terakhir, sistem yang dikembangkan dilengkapi dengan antarmuka visual interaktif berbasis Streamlit, yang memungkinkan pengguna non teknis untuk memahami dan mengeksplorasi hasil analisis dengan mudah. Hasil visualisasi dari data warehouse yaitu dalam bentuk dashboard yang dilengkapi dengan filter berdasarkan dimensi, sedangkan untuk hasil dari *FP-Growth* disusun dalam bentuk tabel dan card interaktif yang menyajikan metrik penting seperti support, confidence, dan lift. Dengan demikian, sistem analitik berbasis data warehouse, *FP-Growth*, dan visualisasi Streamlit yang telah dirancang terbukti dapat mendukung pengambilan keputusan berbasis data di Nuhsantara Merchandise.

## REFERENSI

- [1] M. Adreansyah, P. Fathia, and Gustira, "Analisis dan Perancangan Data Warehouse Pada Data Transaksi Supermarket Menggunakan Schema Snowflake," *Sist. Inf. dan Teknol. Inf.*, vol. 2, no. 2, pp. 1–10, 2023.
- [2] B. F. P. Edhya, "Business Intelligence Data Marketing Menggunakan Metode Kimball Dan Etl Dengan Power Bi," *Kurawal - J. Teknol. Inf. dan Ind.*, vol. 5, no. 2, pp. 87–97, 2022, doi: 10.33479/kurawal.v5i2.642.
- [3] D. Rafly, R. Insani, and A. Dzulkarnain, "Penerapan *FP-Growth* untuk Menentukan Rekomendasi Produk pada Kogu Coffee Shop Malang," *INTEGER J. Inf. Technol.*, vol. 9, no. 1, pp. 27–36, 2024.
- [4] M. Fathur Rezki Junaedi, M. Martanto, and U. Hayati, "Analisis Pola Transaksi Pembelian Makanan Dan Minuman Menggunakan Algoritma *FP-Growth*," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 360–367, 2024, doi: 10.36040/jati.v8i1.8429.
- [5] N. Nurahman, A. Maulana, A. Suprianur, and F. U. Faruq, "Algoritma *FP-Growth* Untuk Rekomendasi Menu Minuman Di Jingga Coffee," *J. Glob. Ilm.*, vol. 1, no. 1, pp. 8–17, 2023, doi: 10.55324/jgi.v1i1.2.
- [6] U. Brawijaya, M. Rayhan Pranandika, W. Purnomo, and S. A. Wicaksono, "Pengembangan Data Warehouse Sebagai Pendukung Pelaporan Akademik (Studi Kasus: Departemen Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 4, pp. 2548–964, 2024, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/13625>
- [7] G. Atmaja and R. Rachman, "PERBANDINGAN ALGORITMA APRIORI DAN *FP-GROWTH* PADA ANALISIS PERILAKU KONSUMEN TERHADAP PEMBELIAN DATA ELEKTRONIK," pp. 298–307, 2025.
- [8] R. Kimball and M. Ross, *The Data Warehouse Toolkit*, Second Edi. Robert Ipsen, 2002.
- [9] M. Furqon and M. F. Nugraha, "Perancangan Data Warehouse Sistem Pendaftaran Mahasiswa Menggunakan Online Analytical Prosesing (OLAP) di Universitas Ma'soem," *Intern. (Information Syst. J.*, vol. 4, no. 1, pp. 35–40, 2021, [Online]. Available: <http://journal.stkom.ac.id/index.php/internal>
- [10] I. P. S. Handika, "Penerapan Datawarehouse Dan Business Intelligence Untuk Analisa Persediaan Barang Di Gudang Pt. Abc," *J. Teknol. Inf. Dan Komput.*, pp. 153–162, 2022.
- [11] T. R. Priono, W. Purnomo, and N. Y. Setiawan, "Pengembangan Data Warehouse menggunakan Metode Kimball ( Studi Kasus : Ekspor & Impor Fauna dan Flora Hias Air Laut )," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 8, pp. 3383–3392, 2021, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/9568/4318>
- [12] J. Purwanto and R. Renny, "Perancangan Data Warehouse Rumah Sakit Berbasis Online Analytical Processing (OLAP)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 5, pp. 1077–1088, 2021, doi: 10.25126/jtiik.2021854232.
- [13] A. Ratu, L. Kusneti, and A. Wijaya, "Prototype Data Warehouse Kantor Penilai Publik XYZ Dengan Metode Nine-Step Kimball," *J. Informatics Busines*, vol. 01, no. 03, pp. 89–96, 2023.
- [14] B. Hasibuan, L. Ratnasari, and Soecahyadi, *Perancangan Middleware Dalam Pelayanan Publik*, no. 84. 2020.
- [15] I. P. A. E. Pratama and I. M. S. Raharja, "Business Intelligence Based on Kimball Nine-Steps Methodology for Monitoring the Feasibility of Goods in Market," *Int. J. Adv. Data Inf. Syst.*, vol. 4, no. 2, pp. 135–144, 2023, doi: 10.25008/ijadis.v4i2.1301.
- [16] S. Oktaviani and A. Bahtiar, "Implementasi Algoritma K-Means Dalam Pengelompokan Data Penjualan CV. Widuri Menggunakan Orange," *J. Wahana Inform.*, vol. 2, no. 1, pp. 188–196, 2023.
- [17] R. Riadi and Mesran, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Analisa Penjualan Parfume," *J. Informatics, Electr. Electron. Eng.*, vol. 2, no. 4, pp. 138–145, 2023, doi: 10.47065/jieee.v2i4.1181.
- [18] T. A. Kinaswara, N. R. Hidayati, and F. Nugrahanti, "Rancang Bangun Aplikasi Inventaris Berbasis Website Pada Kelurahan Bantengan | Kinaswara | Prosiding Seminar Nasional Teknologi Informasi dan Komunikasi (SENATIK)," *Pros. Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2, no. 1, pp. 71–75, 2019, [Online]. Available: <http://prosiding.unipma.ac.id/index.php/SENATIK/article/view/1073>
- [19] M. Rafi Nahjan, Nono Heryana, and Apriade Voutama, "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell," *JATI (Jurnal Mhs.*

- [20] *Tek. Inform.*, vol. 7, no. 1, pp. 101–104, 2023, doi: 10.36040/jati.v7i1.6094.
- [21] S. Pujiono, R. Astuti, and F. Muhamad Basysyar, “Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 615–620, 2024, doi: 10.36040/jati.v8i1.8360.
- D. Alfitra, M. Afdal, M. Fronita, and E. Saputra, “Analisa Keranjang Belanja untuk Menentukan Tata Letak Barang Menggunakan Algoritma *FP-Growth* Market Basket Analysis for Determine Goods Layout Using *FP-Growth* Algorithm,” vol. 13, pp. 1651–1661, 2024.