

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek penelitian ini merupakan jaringan ritel yang memiliki 13 cabang toko dengan kategori operasional *Live*. Data yang digunakan dalam penelitian ini bersumber dari transaksi harian yang tercatat melalui sistem *Point of Sales* (POS) pada masing-masing cabang dengan periode pengambilan data dari Juni hingga September 2025. Seluruh cabang toko diperlakukan sebagai unit penelitian untuk analisis segmentasi kinerja. Nama cabang ditampilkan untuk identifikasi, namun nama merek dan logo tidak dicantumkan guna menjaga kerahasiaan perusahaan.

Tabel 3.1 menyajikan ringkasan rata-rata transaksi harian dari setiap cabang ritel berdasarkan kategori operasional:

Tabel 3. 1 Rata-rata Transaksi Harian per Cabang Toko Ritel (Periode Juni – September 2025)

No	Nama	Category	Average Transaksi Harian (Rp)
1	CIBODAS	Live	27,600,568.22
2	KUNCIRAN	Live	20,809,598.85
3	MUTIARA	Live	26,288,198.78
4	LEGOK	Live	10,752,320
5	CURUG	Live	17,224,190
6	DASANA	Live	25,841,070
7	CITRA	Live	12,283,150
8	RANCAAGONG	Live	19,756,220
9	CIJENGIR	Live	16,510,320
10	UWUNG JAYA	Live	17,313,470
11	VILLA MUTIARA PLUIT	Live	19,280,488
12	PANARUB	Live	11,550,000
13	ADAM MALIK	Live	13,125,000

Data pada Tabel 3.1 Rata-rata Transaksi Harian per Cabang Toko Ritel (Periode Juni – September 2025), menunjukkan bahwa ke-13 cabang yang dianalisis memiliki status Live dan secara kolektif telah menghasilkan data transaksi harian yang aktif dan bervariasi antar lokasi, menjadikannya sampel yang representatif untuk analisis kinerja dan segmentasi.

3.2 Metode Penelitian

Penelitian ini menggunakan pendekatan CRISP-DM untuk menganalisis data transaksi harian pada jaringan toko ritel. Pendekatan ini dipilih karena menyediakan kerangka kerja yang sistematis dan terstruktur, mulai dari pemahaman konteks bisnis hingga penyajian hasil analisis, sehingga mampu mengubah data operasional menjadi informasi strategis yang mendukung proses pengambilan keputusan manajemen. Selain itu, CRISP-DM bersifat fleksibel dan iteratif, memungkinkan penyesuaian proses analisis.

Secara umum, penelitian ini berfokus pada tiga tahap utama, yaitu:

1. Pengumpulan dan pengolahan data transaksi dari sistem POS setiap cabang.
2. Analisis data menggunakan metode *machine learning* (K-Means Clustering dan ARIMA Forecasting) serta validasi statistik untuk menemukan pola dan *insight* tersembunyi.
3. Penyusunan insight dan rekomendasi strategis berbasis hasil analisis melalui dashboard BI interaktif.

Metodologi yang digunakan dalam penelitian ini mengacu pada **CRISP-DM** (**Cross-Industry Standard Process for Data Mining**), yang merupakan standar industri dalam pengolahan dan analisis data. Model ini terdiri dari enam tahapan utama yang dimodifikasi sesuai kebutuhan Business Intelligence, seperti dijelaskan berikut:

3.2.1 Alur Penelitian (CRISP-DM)



Gambar 3. 1 Flowchart Diagram CRISP-DM

1. Pemahaman Bisnis (Business Understanding)
 - A. Menentukan tujuan penelitian, yaitu menghasilkan insight dan rekomendasi strategis dari data transaksi harian toko ritel.
 - B. Mengidentifikasi variabel utama, seperti rata-rata transaksi per hari, kategori toko, serta performa cabang.
2. Pemahaman Data (Data Understanding)
 - A. Mengumpulkan dataset transaksi harian dari seluruh cabang toko ritel.
 - B. Melakukan peninjauan struktur, kelengkapan, dan kualitas data untuk memastikan keakuratan analisis.
3. Persiapan Data (Data Preparation)
 - A. Melakukan pembersihan data (data cleaning), termasuk menghapus duplikasi dan menangani missing values.
 - B. Melakukan transformasi data agar dapat digunakan dalam proses analisis deskriptif dan visualisasi BI.
4. Pemodelan (Modeling)
 - A. Menggunakan teknik *unsupervised learning* K-Means Clustering untuk segmentasi toko dan model ARIMA (*AutoRegressive Integrated Moving Average*) untuk peramalan penjualan. K-Means Clustering digunakan karena merupakan metode *unsupervised learning* yang mampu mengelompokkan toko berdasarkan kemiripan pola penjualan secara objektif tanpa memerlukan label awal. Pendekatan ini menghasilkan segmentasi toko yang homogen berdasarkan karakteristik kinerja penjualan, sehingga memudahkan identifikasi perbedaan performa antar kelompok. Hasil segmentasi tersebut kemudian dimanfaatkan sebagai dasar analisis lanjutan dan mendukung peramalan penjualan menggunakan model ARIMA.
5. Evaluasi (Evaluation)
 - A. Mengevaluasi kualitas dan signifikansi model yang dibangun.

Evaluasi dilakukan secara statistik menggunakan Uji ANOVA dan Uji Chi-Square untuk memvalidasi klaster yang terbentuk dan menguji hipotesis antar variabel.

6. Penyajian dan Rekomendasi (Deployment)
 - A. Menyajikan hasil analisis melalui dashboard Business Intelligence menggunakan Power BI.
 - B. Menyertakan insight otomatis (Natural Language Generation / NLG) agar hasil analisis dapat dibaca dan dipahami dengan mudah oleh pihak manajemen.

3.3 Teknik Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan untuk memperoleh informasi yang akurat dan representatif mengenai aktivitas transaksi harian pada setiap cabang toko ritel. Data yang digunakan bersumber dari sistem internal perusahaan dan observasi langsung terhadap proses operasional toko.

Metode pengumpulan data meliputi:

1. Dokumentasi Internal Perusahaan

Data transaksi harian diperoleh melalui laporan keuangan dan sistem Point of Sales (POS) yang digunakan di setiap cabang. Laporan tersebut berisi detail transaksi harian seperti nomor transaksi, tanggal, nilai transaksi, serta cabang tempat transaksi terjadi.

1.5.1.1 Observasi Lapangan

Observasi dilakukan untuk memahami proses pencatatan dan pelaporan transaksi oleh pihak toko. Setiap cabang memiliki Supervisor (SPV) yang bertanggung jawab melakukan input dan verifikasi laporan harian. Prosesnya dilakukan dengan langkah-langkah sebagai berikut:

- A. SPV mengakses sistem POS pada akhir operasional toko setiap hari.

- B. Data transaksi harian diunduh dalam format laporan Daily Settlement.
 - C. Laporan tersebut kemudian diserahkan ke bagian administrasi pusat melalui sistem internal perusahaan untuk dikompilasi.
2. Database Perusahaan

Setelah seluruh laporan dari cabang diterima, data disimpan dalam database pusat yang terhubung dengan dashboard Business Intelligence. Database ini menjadi sumber utama dalam analisis, karena sudah melewati proses validasi dan penyamaan format data dari masing-masing cabang.

Jenis data yang dikumpulkan meliputi:

- A. Nama cabang toko
- B. Kategori toko (Live, Planned, Temporary)
- C. Rata-rata transaksi harian
- D. Nilai transaksi bersih dan kotor (Net dan Gross Price)
- E. Status pembayaran dan kode cabang

Data yang telah terkumpul kemudian diproses dan diolah pada tahap Data Preparation untuk memastikan kualitas dan konsistensi sebelum dilakukan analisis lebih lanjut.

3.4 Variabel Penelitian

Variabel penelitian digunakan untuk membantu proses analisis dan pengolahan data transaksi harian toko ritel.

Setiap variabel merepresentasikan elemen penting yang berpengaruh terhadap pemahaman performa cabang dan penyusunan insight strategis dalam sistem Business Intelligence.

Tabel 3.2 menjelaskan variabel yang digunakan dalam penelitian ini:

Tabel 3. 2 Variabel Penelitian

Variabel	Jenis	Deskripsi
Form no	Independen	Nomor formulir transaksi sebagai identifikasi unik setiap transaksi.
Transact.no	Independen	Nomor transaksi yang dihasilkan oleh sistem POS untuk setiap pesanan.
Form date	Independen	Tanggal transaksi dilakukan pada cabang toko.
Salesman	Independen	Nama atau kode karyawan yang melakukan penjualan.
Address no	Independen	Kode cabang toko tempat transaksi dilakukan.
Address name	Independen	Nama cabang toko
Deliver to	Independen	Lokasi pengiriman barang
Transaction text	Independen	Jenis transaksi / catatan transaksi

Net Price	Dependen	Nilai bersih transaksi tanpa pajak.
Gross Price	Dependen	Nilai kotor transaksi termasuk pajak
Paid	Independen	Status Pembayaran (K = sudah dibayar)
Variabel	Jenis	Deskripsi
Cost Center	Independen	Kode pusat biaya yang digunakan untuk mengelompokkan transaksi berdasarkan unit bisnis.

Dari variabel-variabel pada tabel 3.2, Net Price dan Gross Price menjadi fokus utama analisis karena menunjukkan nilai performa transaksi harian yang digunakan untuk mengukur tingkat penjualan dan membandingkan kinerja antar cabang. Sementara itu, variabel seperti Address Name, Category, dan Form Date digunakan untuk melakukan agregasi data dan analisis tren berdasarkan waktu dan lokasi cabang.

3.5 Teknik Analisis Data

Teknik analisis data dalam penelitian ini menggunakan pendekatan *hybrid* yang menggabungkan kekuatan analitis Python untuk pemodelan statistik (*backend*) dan kekuatan implementasi *Business Intelligence* Power BI untuk visualisasi interaktif dan *engine NLG* (*frontend*). Alur kerja ini mengacu pada metodologi CRISP-DM, yang difokuskan pada tahap *Modeling* dan *Deployment*

3.5.1 Analisis Statistik

Analisis dilakukan di Python (menggunakan Jupyter Notebook) untuk melakukan pemodelan statistic mendalam guna menemukan insight dari data mentah.

1. Segmentasi Kinerja (K-Means Clustering): Algoritma K-Means Clustering diterapkan untuk mengelompokkan 13 toko aktif berdasarkan tiga variabel kunci: Net Price, Open (Qty), dan Gross Margin %.
2. Validasi Klaster Optimal: Jumlah klaster optimal ($K=4$) ditentukan secara statistik menggunakan Elbow Method dan divalidasi dengan Silhouette Score, yang menghasilkan skor tertinggi 0.480.
3. Uji Signifikansi (ANOVA): Uji F (One-Way ANOVA) dilakukan untuk memvalidasi apakah perbedaan antar klaster (terutama Net Price dan Open (Qty)) signifikan secara statistik.
4. Analisis Perilaku (Uji Chi-Square): Uji Chi-Square diterapkan untuk menguji hipotesis apakah ada perbedaan perilaku pembelian (seperti Product Category atau Paid Method) antar klaster yang berbeda.
5. Output Analisis Backend: Hasil dari analisis Python ini mencakup ID Klaster untuk setiap toko, hasil Uji ANOVA, dan Uji Chi-Square diekspor ke dalam *file* Excel untuk digunakan sebagai *input* utama dalam implementasi BI.

3.5.2 Analisis Proyeksi Penjualan

Proses peramalan (forecasting) juga dilakukan di Python untuk memanfaatkan library statistic yang lebih robust. Model ARIMA (Autoregressive Integrated Moving Average) diterapkan pada data historis (Juni-September) untuk memprediksi Total Net Sales selama dua bulan ke depan (Oktober-November). Hasil prediksi ini diekspor untuk diimpor ke Power BI.

3.5.3 Implementasi BI: Persiapan Data dan Metrik (Power BI)

Tahap ini adalah penerapan hasil dari Python ke dalam Power BI untuk visualisasi dinamis.

1. Integrasi Model Data: Hasil analisis dari Python (Tabel Store Clusters dan Forecast) diimpor ke Power BI dan dihubungkan dengan tabel Transactions dan tabel dimensi Cluster_Names untuk membentuk model data relasional
2. Penggabungan Data (Data Union): Sebuah tabel virtual Data Gabungan dibuat menggunakan fungsi DAX UNION. Tabel ini mengonsolidasikan data aktual dari Transactions dan data prediksi dari Forecast ke dalam satu struktur (Bulan_Grafik, Nilai_Grafik, Tipe_Data) untuk visualisasi tren yang mulus, seperti terlihat pada sintaks di Gambar 3.2

```

1 Data Gabungan =
2 UNION(
3     // Bagian 1: Data Aktual (Transactions)
4     SELECTCOLUMNS(
5         'Transactions',
6         "Bulan_Grafik", 'Transactions'[month],
7         "Nilai_Grafik", 'Store Clusters'[Total Net Sales],
8         "Tipe_Data", "Aktual"
9     ),
10
11     // Bagian 2: Data Prediksi (Forecast)
12     SELECTCOLUMNS(
13         'Forecast',
14         "Bulan_Grafik", 'Forecast'[month],
15         "Nilai_Grafik", 'Forecast'[Predicted_Net_Sales], // Ambil kolom nilai prediksi
16         "Tipe_Data", "Prediksi"
17     )
18 )

```

Gambar 3. 2 Sintaks DAX Tabel Gabungan

3. Pemodelan Metrik Kunci (DAX Measures): Serangkaian *measure* DAX (bukan kolom kalkulasi) dikembangkan untuk menghitung indikator kinerja utama (KPI) yang akan digunakan di seluruh *dashboard*. Metrik ini mencakup Total Net Sales, Total Proyeksi Akhir, dan metrik pendukung klaster seperti Avg Sales per Trans.

a) Metrik Kunci Kinerja Historis dan Clustering

Metrik ini berasal dari tabel transaksi aktual (Transactions) dan digunakan untuk mengukur kinerja toko serta menjadi input utama model *K-Means Clustering* Python.

Tabel 3. 3 Metrik Kinerja

Nama Metrik	Tabel Asal	Keterangan
Total Net Sales	Store Clusters	Total nilai penjualan bersih (pendapatan) yang dihasilkan selama periode aktual, digunakan untuk KPI Nominal.
Open (Qty)	Store Clusters	Jumlah total kuantitas barang yang terjual/diproses, digunakan sebagai sumbu Y pada Scatter Plot untuk mengukur volume.
Net Price	Store Clusters	Nilai harga jual bersih rata-rata per transaksi, digunakan sebagai sumbu X pada Scatter

		Plot untuk mengukur nilai transaksi.
MoM Sales Growth%	Store Clusters	Percentase pertumbuhan nilai penjualan dari bulan ke bulan (<i>Month-over-Month</i>), digunakan untuk menganalisis momentum historis.
Cluster Avg Net Price	Store Clusters	Nilai Net Price rata-rata dari seluruh toko dalam kluster yang sama, digunakan untuk validasi <i>output</i> model K-Means.

b) Metrik Kunci Proyeksi dan Tren (Tabel Data Gabungan)
 Metrik ini adalah hasil perhitungan DAX pada **Tabel Gabungan** (Data Gabungan) dan digunakan untuk membandingkan kinerja Aktual dan Prediksi pada *Line Chart* dan *KPI Cards*.

Tabel 3. 4 Metrik Proyeksi dan Tren

Nama Metrik	Tabel Asal	Keterangan
Total Nilai Prediksi	Data Gabungan	Total nilai penjualan yang diproyeksikan (Oktober–November).

Nama Metrik	Tabel Asal	Keterangan
Total Proyeksi Akhir	Data Gabungan	Nilai gabungan dari Total Penjualan Aktual + Total Nilai Prediksi (Total kinerja yang diproyeksikan Juni–November).
Persen Kontribusi	Cluster_Names	Persentase kontribusi nilai penjualan setiap bulan terhadap total nilai keseluruhan periode Juni–November.
Label Bar Persentase Saja	Cluster_Names	Measure teks yang menampilkan Persen Kontribusi sebagai Custom Label pada Bar Chart.

c) Metrik Pendukung (Digunakan untuk Klustering)

Metrik ini memberikan detail tambahan mengenai kualitas transaksi dan efisiensi di setiap kluster, yang merupakan *measures* agregasi di tabel Store Clusters.

Tabel 3. 5 Metrik Pendukung

Nama Metrik	Tabel Asal	Keterangan
Avg Sales per Trans	Store Clusters	Nilai penjualan rata-rata yang dihasilkan per transaksi.
Promo Transaction %	Store Clusters	Persentase transaksi yang melibatkan penggunaan promosi.
Cluster Avg Margin %	Store Clusters	Rata-rata margin dari seluruh toko dalam kluster yang sama.

3.5.4 Implementasi BI: Engine Rekomendasi Strategis

Pada sub bab 3.5.4 Implementasi BI: Engine Rekomendasi Strategis perlu dilakukan untuk menjawab tujuan utama skripsi (“Penyusunan Insight dan Rekomendasi Strategis”). Engine NLG dinamis ini dibangun di dalam Power BI menggunakan DAX, berdasarkan logika yang divalidasi dari Python.

1. Dekonstruksi Komponen Narasi: Tabel Cluster_Names dirancang untuk menyimpan komponen-komponen teks fundamental, yaitu Peringkat, Insight, dan Rekomendasi untuk setiap kluster.
2. Isolasi Konteks Filter Finansial: Sebuah *measure* DAX bernama [Cluster Avg Net Price (Measure)] dibuat. *Measure* ini menggunakan fungsi CALCULATE(AVERAGE(...), ALL('Store Clusters'[Address Name (Store)])) untuk menonaktifkan konteks

filter eksternal dari toko yang dipilih. Ini adalah langkah krusial untuk memastikan *measure* selalu menampilkan rata-rata klaster yang sebenarnya, bukan nilai toko individual.

3. Sintesis Narasi Dinamis (*Engine DAX*): Sebuah *measure* DAX utama, [NLG_Paragraf_Dinamis], dikembangkan untuk bertindak sebagai *engine* sintesis. *Measure* ini secara *real-time* akan:

- (a) Memanggil variabel teks (Insight Validasi, Rekomendasi Aksi) menggunakan SELECTEDVALUE berdasarkan filter klaster yang aktif.
- (b) Memanggil variabel finansial dinamis ([Cluster Avg Net Price (Measure)]) yang konteks filternya telah terisolasi

Output dinamis dari engine ini akan diimplementasikan dalam dashboard "Analisis Klaster" menggunakan fitur "Tooltip Halaman Laporan", yang akan disajikan pada Bab IV

