BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian dalam penelitian ini adalah data penjualan harian pada toko ritel kosmetik. Yang mencerminkan aktivitas penjualan dalam periode tertentu. Data tersebut mencakup informasi yang relevan dengan proses prediksi, seperti jumlah penjualan, waktu transaksi, serta atribut pendukung lainnya yang berkaitan dengan karakteristik penjualan produk kosmetik.

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari catatan transaksi internal toko ritel kosmetik. Pemilihan objek penelitian ini didasarkan pada karakteristik penjualan ritel kosmetik yang bersifat fluktuatif dan dipengaruhi oleh berbagai faktor, seperti promosi, kategori produk, dan waktu penjualan, sehingga sesuai untuk dianalisis menggunakan pendekatan data mining dan algoritma *machine learning*.

3.1.1 Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan selama periode Desember 2025 hingga April 2026. Pengolahan dan analisis data dilakukan di lingkungan Universitas Buana Perjuangan Karawang menggunakan perangkat lunak pendukung analisis data. Rincian tahapan pelaksanaan penelitian dapat dilihat secara lengkap pada Tabel 3.1 yang menyajikan alokasi waktu tiap aktivitas.

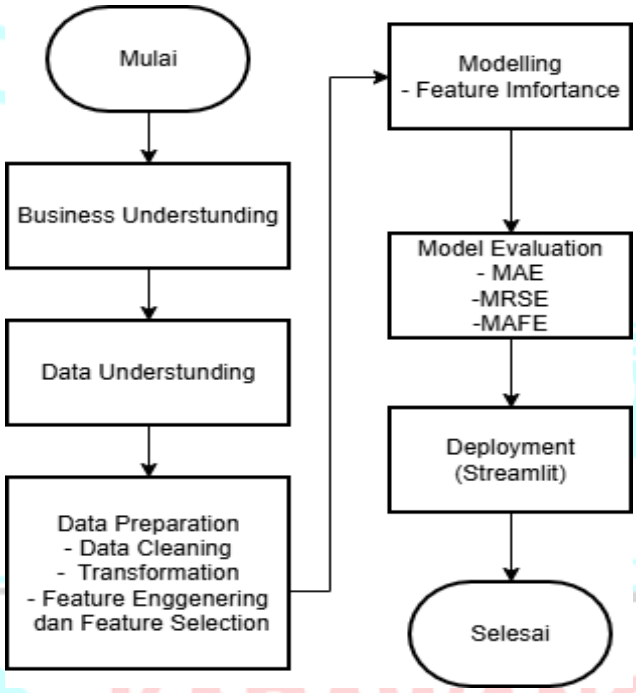
Tabel 3. 1 *Timeline* Penelitian

No	Kegiatan	2025								2026								
		11				12				1				2				3
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1
1	*Business Understunding*	■	■	■	■	■	■	■	■									
2	*Data Understunding*					■	■	■	■	■								
3	*Data Preparation*									■	■	■	■	■	■	■	■	
4	*Modelling*											■	■	■	■	■	■	
5	*Evaluation*															■	■	
6	*Deployment*															■	■	■

3.2 Prosedur Penelitian

Prosedur penelitian ini mengacu pada kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*) merupakan standar umum dalam penelitian data mining. CRISP-DM terdiri dari enam tahapan utama, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*.

Alur penelitian secara keseluruhan diilustrasikan dalam bentuk diagram yang ditampilkan pada Gambar 3.1.



Gambar 3. 1 Flowchart Alur Penelitian

3.2.1 Business Understunding

Tahap *Business Understanding* bertujuan untuk merumuskan permasalahan penelitian dan menetapkan tujuan pengembangan model prediksi.

A. Permasalahan yang terdapat pada toko ritel kosmetik

Setelah melakukan analisis terdapat point bisnis pada toko ritel kosmetik yang dapat diambil, berikut adalah rincianya yang di tulis secara sfesifik :

1) Ketidakpastian jumlah penjualan di masa mendatang

Toko ritel kosmetik sering mengalami fluktuasi penjualan harian/bulanan yang menyebabkan kesulitan dalam menentukan persediaan barang dan menentukan keputusan bisnis.

2) Risiko kelebihan dan kekurangan stok produk kosmetik

Produk kosmetik memiliki masa simpan dan tren yang cepat berubah, sehingga kesalahan prediksi penjualan dapat mengakibatkan:

a) *Overstock* (stok menumpuk dan berisiko kedaluwarsa)

b) *Understock* (kehabisan produk yang banyak diminati pelanggan).

3) Belum diketahuinya faktor yang paling berpengaruh terhadap penjualan

Data penjualan menyimpan berbagai variabel seperti harga, diskon, kategori produk, brand, dan waktu penjualan, namun belum dianalisis untuk mengetahui:

a) Faktor apa yang paling dominan memengaruhi penjualan,

b) Variabel mana yang paling penting dalam proses prediksi.

4) Pengambilan keputusan bisnis yang masih bersifat intuitif

Penentuan strategi penjualan, promosi, dan pengadaan stok masih banyak dilakukan berdasarkan pengalaman, bukan berdasarkan analisis data historis.

B. Tujuan Pengembangan model prediksi

Setelah dilakukan analisis *point* bisnis pada toko ritel, maka berikut tujuan pengembangan model prediksi :

1) Membangun model prediksi penjualan menggunakan algoritma Random Forest.
2) Mengukur performa model prediksi menggunakan metrik evaluasi yang sesuai.
3) Menghasilkan informasi feature importance untuk mengetahui variabel yang paling berpengaruh terhadap penjualan produk kosmetik.

3.2.2 Data Understanding

Tahap *Data Understanding* dilakukan untuk memahami karakteristik data penjualan yang digunakan dalam penelitian. Pada tahap ini dilakukan pengumpulan data, pemeriksaan struktur data, serta eksplorasi awal untuk mengidentifikasi pola, distribusi data, dan potensi permasalahan seperti data hilang atau nilai ekstrem. Hasil dari tahap ini menjadi dasar dalam menentukan langkah pengolahan data pada tahap selanjutnya.

A. Pengumpulan Data

Pada tahap pengumpulan data, dataset yang diambil berasal dari data transaksi di toko ritel kosmetik yang ada di salah satu darah karawang, mulai dari Januari sampai Desember 2025. Dataset ini berisi 21.053 baris data dan memiliki 16 atribut yang berbeda.

Tabel 3. 2 Tabel Sample Dataset

Outlet	code	Date	time	transaction_id	customer_id	product_name	category	brand	quantity	unit_price	discount_percent	discount_amount	net_price	payment_method	cashier
Galuh Mas	9az5	1-Jan-25	10:11	0123RX88	876192xxxx	YOU NoutriWear Cushion Shade C05 Tan 12g	Skincare	YOU	1	169000	0	0	169000	Cash	Andreyani Safitri
Galuh Mas	9az5	1-Jan-25	10:28	3787XA01	887530xxxxx	Pixy Lip Cream 01 Chic Rose	lips	Pixy	3	52000	15	7800	44200	QRIS	Andreyani Safitri
Galuh Mas	9az5	1-Jan-25	10:29	0791RL14	804080xxxxx	Somethinc AHA BHA PHA Peeling Solution	Skincare	Somethinc	1	81287	30	24386	56901	QRIS	Andreyani Safitri
Galuh Mas	9az5	1-Jan-25	10:32	3181AS61	888961xxxx	Barenbliss Peach Lip Tint No 02 Peach 5g	lips	Barenbliss	1	67600	0	0	67600	QRIS	Andreyani Safitri

B. Deskripsi Data

Berikut ini penjelasan table pada sample dataset yang terdiri dari 16 variabel :

Tabel 3. 3 Tabel Penjelasan Sample Dataset

No	Nama Atribut	Fungsi Atribut	Peran dalam Analisis
1	outlet	Identifikasi lokasi toko	Variabel kategorik, memerlukan encoding untuk analisis
2	code	Kode unik produk	Identifikasi produk, dapat di-encoding atau digunakan untuk grouping
3	date	Tanggal transaksi	Analisis temporal, perlu ekstraksi fitur (hari, bulan, tahun)
4	time	Waktu transaksi	Analisis pola waktu, perlu ekstraksi fitur (jam, periode hari)
5	transaction_id	ID unik transaksi	Identifikasi record, bukan variabel prediktif
6	customer_id	ID pelanggan	Analisis perilaku pelanggan, dapat digunakan untuk segmentasi
7	product_name	Nama produk	Deskripsi produk, dapat di-encoding atau digabung dengan category
8	category	Kategori produk	Variabel kategorik penting, memerlukan encoding (one-hot/label)
9	brand	Merek produk	Variabel kategorik, memerlukan encoding untuk analisis
10	quantity	Jumlah produk dibeli	Variabel numerik, potensial sebagai input model
11	unit_price	Harga per unit	Variabel numerik, potensial sebagai input model
12	discount_percent	Persentase diskon	Variabel numerik, potensial sebagai input model
13	discount_amount	Nominal diskon	Variabel numerik, potensial sebagai input model
14	net_price	Harga bersih setelah diskon	Variabel numerik, potensial sebagai target atau input
15	payment_method	Metode pembayaran	Variabel kategorik, memerlukan encoding untuk analisis
16	cashier	Nama kasir	Identifikasi kasir, dapat digunakan untuk analisis performa

C. Eksplorasi Data

Berdasarkan dataset transaksi dengan 21.052 baris data dan 16 kolom, eksplorasi data dilaksanakan melalui statistik deskriptif untuk mengidentifikasi:

1) Volume data keseluruhan dataset mencakup 21.052 record transaksi dengan informasi outlet, produk, pelanggan, dan pembayaran yang menyediakan basis memadai untuk analisis dan pemodelan.
2) Rentang dan nilai tengah variabel numerik analisis dilakukan pada variabel *quantity*, *unit_price*, *discount_percent*, *discount_amount*, dan *net_price* untuk mengetahui nilai minimum, maksimum, dan rata-rata yang menggambarkan spektrum harga, pola diskon, dan volume transaksi.
3) Distribusi transaksi berdasarkan kategori dan metode pembayaran Pemetaan sebaran transaksi berdasarkan *category* dan *payment_method* untuk mengidentifikasi preferensi produk dan kecenderungan metode pembayaran konsumen.
4) Pola transaksi dalam dimensi temporal analisis atribut *date* dan *time* untuk mengungkap pola transaksi berdasarkan periode harian, jam operasional, dan tren bulanan yang menunjukkan periode puncak aktivitas penjualan.

Tahap ini menghasilkan pemahaman komprehensif terhadap karakteristik data dan mengidentifikasi potensi korelasi antar variabel untuk optimalisasi pemodelan.

3.2.3 Data Preparation

Tahap *Data Preparation* merupakan proses pengolahan data agar siap digunakan dalam pemodelan. Kegiatan yang dilakukan pada tahap ini meliputi pembersihan data, penanganan data yang tidak lengkap, transformasi variabel, serta rekayasa fitur. Selain itu, data dibagi menjadi data latih dan data uji untuk memastikan bahwa model dapat dievaluasi secara objektif.

a) Data Cleaning

Data *cleaning* dilakukan untuk meningkatkan kualitas data dengan cara mengatasi permasalahan yang terdapat dalam dataset. Proses ini meliputi penanganan data hilang (*missing values*), penghapusan data duplikat, serta identifikasi dan penanganan nilai ekstrem (*outlier*) yang berpotensi mengganggu proses pemodelan.

Dengan melakukan pembersihan data, diharapkan data yang digunakan dalam penelitian lebih akurat dan mencerminkan kondisi penjualan yang sebenarnya. Contoh *Pseudocode Data Cleanning* :

```
BEGIN
INPUT: Dataset_transaksi
FOR setiap kolom dalam Dataset_transaksi:
    Cek missing value
    IF missing value ditemukan:
        IF kolom numerik:
            Isi dengan nilai median
        ELSE:
            Isi dengan nilai modus
END FOR
Cek data duplikat
IF duplikat ditemukan:
    Hapus data duplikat
Deteksi outlier pada quantity dan net_price
IF outlier ditemukan:
    Tangani menggunakan metode IQR

OUTPUT Dataset_bersih
END
```

Tabel 3. 4 Contoh Data *Before Cleanning*

transaction_id	category	quantity	net_price
4925NW1	Parfum	1	118750
4925NW1	Parfum	1	118750
5001AB2	Skincare	Nan	75000
6002CD5	Parfum	1	9500000

Tabel 3. 5 Contoh Data *After cleanning*

transaction_id	category	quantity	net_price
4925NW1	Parfum	1	118750
5001AB2	Skincare	2	75000

Berdasarkan table 3. 4 Contoh data sebelum *cleaning* dan table 3.5 Contoh data sesudah *cleaning* menunjukan contoh hasil output sebagai berikut:

1. Data duplikat dengan *transaction_id* yang sama dihapus
2. *Missing value* pada *quantity* diisi menggunakan nilai rata-rata
3. Nilai *outlier* pada *net_price* yang sangat tinggi dihapus menggunakan metode IQR
4. Dataset akhir lebih stabil dan representatif terhadap kondisi penjualan

b) Transformasi data

Transformasi data dilakukan agar format dan skala data sesuai dengan kebutuhan algoritma *machine learning*.

Contoh *Pseudocode* Transformasi data:

```
BEGIN
INPUT: Dataset_transaksi
// Ekstraksi tanggal
FOR setiap transaksi:
    year  = extract(date, "year")
    month = extract(date, "month")
END FOR
Hapus kolom date
// Encoding kategorikal
        Encode category menjadi category_encoded
        Encode payment_method menjadi payment_encoded
Hapus kolom category dan payment_method
OUTPUT: Dataset_transaksi_tertransformasi
END
```

Tabel 3. 6 Contoh *Before transformation*

transaction_id	date	category	payment_method	quantity	net_price
8001GH1	01-12-2025	Parfum	E-Wallet	2	120000

Tabel 3. 7 Contoh *After Transformation*

transaction_id	year	month	category_encoded	payment_encoded	quantity	net_price
8001GH1	2025	1	0	1	2	120000

Tabel 3.6 dan Tabel 3.7 merupakan contoh dari proses *transformation*. Table 3.7 menunjukan hasil *output* sesudah transformasi, transformasi yang dilakukan sebagai berikut:

1) Kolom *date* ditransformasi menjadi *year* dan *month* untuk menangkap pola waktu transaksi.
2) Kolom *category* dan *payment_method* diubah ke bentuk numerik melalui proses *encoding*.

c) Feature Engeneering dan Feature Selection

a) Feature engineering

Feature engineering merupakan proses pembuatan fitur baru dari fitur yang sudah ada guna memperkaya informasi dalam data sehingga kemampuan prediksi model terhadap penjualan dapat meningkat. *Feature engineering* dilakukan dengan menambahkan fitur baru, yaitu:

a. total_sales = quantity × net_price
b. *is_discount* (0 = tidak ada diskon, 1 = ada diskon)

Tabel 3. 8 Contoh *Output Feature Engineering*

transaction_id	quantity	net_price	unit_price	total_sales	is_discount
8001GH1	2	120000	60000	120000	1
8002GH2	1	100000	100000	100000	0

Table 3.8 merupakan contoh output dari feature engineering. Pada table tersebut menunjukan hasil *feature engineering* yang menghasilkan dua fitur baru, yaitu total_sales sebagai representasi nilai transaksi dan is_discount sebagai indikator adanya diskon, yang selanjutnya digunakan dalam proses pemodelan.

Contoh *Psedocode Feature Engineering*:

```
BEGIN
INPUT: Dataset_tertransformasi
FOR setiap transaksi:
    total_sales = quantity × net_price
    IF discount > 0:
        is_discount = 1
    ELSE:
        is_discount = 0
END FOR
OUTPUT : Dataset_feature_engineering
END
```

b) Feature Selection

Feature selection dilakukan setelah tahap *feature engineering* untuk memilih atribut yang paling berpengaruh terhadap variabel target total_sales.

Contoh *pseudocode Feature Selection :*

```
BEGIN
INPUT: Dataset_feature_engineering
Hitung korelasi fitur terhadap target
Pilih fitur dengan korelasi tertinggi
Buang fitur tidak relevan
Selected_Features =
    [quantity, net_price, Category, total_sales, is_discount, month]
OUTPUT: Dataset_final siap pemodelan
END
```

Tabel 3. 9 Contoh *Output Feature Selection*

quantity	unit_price	category	is_discount	month	total_sales
2	60000	skincare	1	1	110000
1	100000	makeup	0	1	100000

Table 3.9 menunjukan contoh hasil output *Feature selection* . pada table tersebut menunjukan variabel *quantity*, *unit_price*, *category*, *is_discount*, dan

month terpilih sebagai variabel input utama untuk memprediksi total_sales karena memiliki pengaruh signifikan terhadap nilai penjualan.

d). Agregrasi data

Agregasi data adalah proses mengelompokkan dan merangkum data transaksi detail menjadi data yang lebih sederhana dan informatif. Pada penelitian ini, data transaksi penjualan yang awalnya berupa setiap item pembelian diubah menjadi total penjualan harian. Tujuan agregasi data adalah agar data lebih mudah dianalisis dan sesuai untuk proses prediksi time series, karena model membutuhkan data berdasarkan urutan waktu.

Tabel 3. 10 Sebelum Agregrasi Data

date	transaction	Name_Produk	Qty	Unit_Price	Total
01-01-2025	TRX001	Lipstik	2	50.000	100.000
01-01-2025	TRX002	Bedak	1	75.000	75.000
01-01-2025	TRX003	Serum	1	120.000	120.000
02-01-2025	TRX004	Sabun Wajah	2	35.000	70.000
02-01-2025	TRX005	Toner	1	60.000	60.000

Dibawah ini merupakan Tabel sesudah dilakukan proses agregasi.

Tabel 3. 11 Sesudah Agregasi (Per Hari)

Tanggal	Total Sales
01-01-2025	295.000
02-01-2025	130.000

Melalui proses ini, beberapa transaksi dalam satu tanggal dijumlahkan menjadi satu nilai total sales per hari. Dengan demikian, data yang semula berjumlah ribuan baris transaksi menjadi lebih ringkas dalam bentuk data harian.

e). Train – Test split

Dataset dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*) dengan perbandingan 80:20.

```
BEGIN
INPUT: Dataset_final
Bagi data:
    80% data training
    20% data testing
OUTPUT :Training_Data, Testing_Data
END
```

Hasil dari proses *Train-test Split* , Data latih digunakan untuk melatih model agar dapat mengenali pola dalam data, sedangkan data uji digunakan untuk mengevaluasi kinerja model pada data yang belum pernah dilihat. Pada tahap pemodelan, data penjualan harian dibagi menjadi data training dan testing menggunakan metode time-based split atau pembagian berdasarkan urutan waktu. Teknik ini dilakukan dengan mengurutkan data berdasarkan tanggal, kemudian membagi data menjadi 80% data training dan 20% data testing.

3.2.4 Modelling

Tahap Modeling bertujuan membangun model prediksi penjualan ritel kosmetik menggunakan algoritma Random Forest Regressor. Random Forest dipilih karena merupakan algoritma *ensemble learning* berbasis pohon keputusan yang efektif dalam menangani data berjumlah besar serta variabel input yang kompleks dan bervariasi. Pada tahap ini, proses pembangunan model juga dilakukan dengan hyperparameter tuning menggunakan GridSearchCV untuk mencari kombinasi parameter terbaik sehingga performa model menjadi lebih optimal. Melalui proses tuning tersebut, model dapat menghasilkan prediksi penjualan yang lebih akurat dan stabil.

Contoh *Pseudocode model Random Forest*:

```
INPUT:
X_train ← data fitur latih
y_train ← data target latih

PROSES:
1. Menentukan algoritma yang digunakan:
model ← Random Forest Regressor

2. Menentukan parameter yang akan diuji (Hyperparameter Tuning):
n_estimators ← [100, 200, 300]
max_depth ← [5, 10, 15, None]
min_samples_split ← [2, 5, 10]
min_samples_leaf ← [1, 2, 4]

3. Melakukan pencarian parameter terbaik menggunakan GridSearchCV:
a. Bentuk beberapa kombinasi parameter
b. Latih model pada data latih untuk setiap kombinasi
c. Evaluasi performa masing-masing kombinasi
d. Pilih kombinasi dengan nilai terbaik

4. Membentuk model Random Forest terbaik:
best_model ← hasil parameter terbaik

5. Melatih model terbaik menggunakan data latih:
best_model.fit(X_train, y_train)

OUTPUT:
model_random_forest_terbaik ← model Random Forest terlatih
best_parameter ← parameter terbaik hasil tuning

END
```

Selanjutnya dilakukan penghitungan nilai MAE, RMSE, dan MAFE untuk mengukur tingkat kesalahan prediksi. Hasil evaluasi metrik MAE dan MAFE digunakan untuk mengetahui rata-rata kesalahan absolut, sedangkan RMSE digunakan untuk mengidentifikasi keberadaan kesalahan prediksi yang bernilai besar. Selain itu, R^2 Score digunakan sebagai validasi model untuk mengukur seberapa baik model mampu menjelaskan variasi data target. Semakin mendekati nilai 1, maka semakin baik kemampuan model dalam melakukan prediksi.

Contoh *Pseudocode Hasil Model* :

```
BEGIN
INPUT: y_actual, y_predicted
        MAE  = hitung_MAE(y_actual, y_predicted)
        RMSE = hitung_RMSE(y_actual, y_predicted)
        MAPE = hitung_MAPE(y_actual, y_predicted)
OUTPUT: MAE, RMSE, MAPE
END
```

Table 3.12 merupakan contoh output evaluasi. Pada table tersebut menunjukkan hasil evaluasi menggunakan metrik MAE, RMSE, dan MAPE, diperoleh nilai kesalahan prediksi yang relatif rendah. Nilai MAE sebesar 0,24 menunjukkan rata-rata kesalahan absolut yang kecil, sedangkan RMSE sebesar 0,31 mengindikasikan tidak adanya kesalahan prediksi yang ekstrem. Selain itu, nilai MAPE sebesar 0,18 menunjukkan bahwa rata-rata kesalahan prediksi termasuk dalam kategori baik.

Tabel 3. 12 Hasil Model Random Forest

Metrik	**Nilai**
MAE	0,24
RMSE	0,31
MAPE	0,18

A. *Feature Importance* (Hasil Tambahan *Modelling*)

Feature importance dilakukan pada tahap *modeling* karena nilai kepentingan fitur dihasilkan langsung dari model yang telah dilatih dan digunakan untuk menjelaskan faktor utama yang memengaruhi prediksi penjualan.

Contoh *Pseudocode Feature Imfortance* :

```
BEGIN
INPUT: model_random_forest
Ambil nilai feature importance dari model
Urutkan fitur berdasarkan nilai kepentingannya
OUTPUT : daftar fitur dan nilai importance
END
```

Table 3. 13 merupakan contoh *output feature imfortance*. Pada table tersebut menunjukkan hasil *feature imfortance quantity* (0,41) merupakan faktor paling dominan dalam memprediksi total penjualan, diikuti oleh *net_price* (0,32) dan *is_discount* (0,14). Sementara itu, variabel *month* (0,08), *category_encoded* (0,04), dan *day* (0,01) memiliki kontribusi yang relatif kecil. Hal ini menunjukkan bahwa kuantitas dan harga produk merupakan penentu utama dalam prediksi nilai penjualan kosmetik.

Tabel 3. 13 Contoh *Output Feature Imfortance*

Fitur	Nilai Importance
quantity	0.41
net_price	0.32
is_discount	0.14
month	0.08
category_encoded	0.04
day	0.01

3.2.5 Evaluation

Evaluasi model dilakukan untuk mengukur tingkat akurasi dan performa model prediksi yang telah dibangun. Proses ini bertujuan mengetahui seberapa dekat hasil prediksi terhadap nilai aktual menggunakan data uji.

Tabel 3. 14 Contoh Output Hasil Evaluasi

Total Sales Aktual (Rp)	Total Sales Prediksi (Rp)
118.750	120.300
115.000	113.800
185.000	182.600
28.800	30.100
75.050	74.200

Tabel 3. 14 merupakan contoh *output* hasil *modelling*. Pada tabel tersebut menunjukkan hasil bahwa model yang dibangun mampu mempelajari pola yang terdapat pada data latih dan menghasilkan prediksi yang mendekati nilai aktual pada data uji. Hal ini mengindikasikan bahwa algoritma yang digunakan efektif dalam menangkap hubungan antarfitur dan target, serta memiliki kemampuan generalisasi yang baik.

3.2.6 Deployment

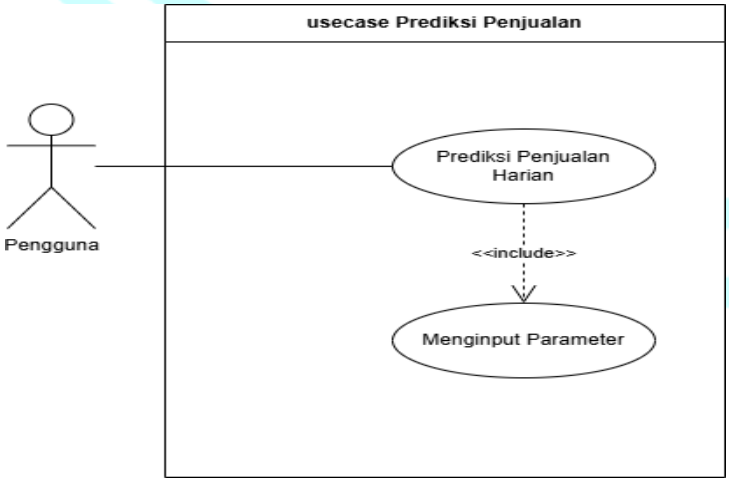
Tahap *Deployment* merupakan tahap penyajian hasil model prediksi. Hasil penelitian disajikan dalam bentuk laporan dan visualisasi, serta dapat diimplementasikan dalam aplikasi sederhana berbasis *web* sebagai alat bantu analisis. Tahap ini bertujuan agar hasil penelitian tidak hanya bersifat akademis, tetapi juga memiliki nilai praktis bagi pengambilan keputusan.

Tahap implementasi sistem merupakan bagian dari proses *Deployment* dalam metodologi CRISP-DM, yaitu menerapkan model prediksi penjualan yang telah dibangun ke dalam bentuk aplikasi yang dapat digunakan secara langsung oleh pengguna. Sistem ini dikembangkan berbasis web menggunakan Streamlit sebagai antarmuka pengguna dan Python sebagai bahasa pemrograman utama. Model prediksi yang digunakan adalah *Random Forest Regressor* yang sebelumnya telah melalui tahap pelatihan dan evaluasi.

Melalui implementasi ini, pengguna dapat melakukan prediksi total penjualan harian berdasarkan parameter yang dimasukkan, serta melihat informasi pendukung seperti produk terlaris dan *feature importance.*

A. *UseCase* Diagram

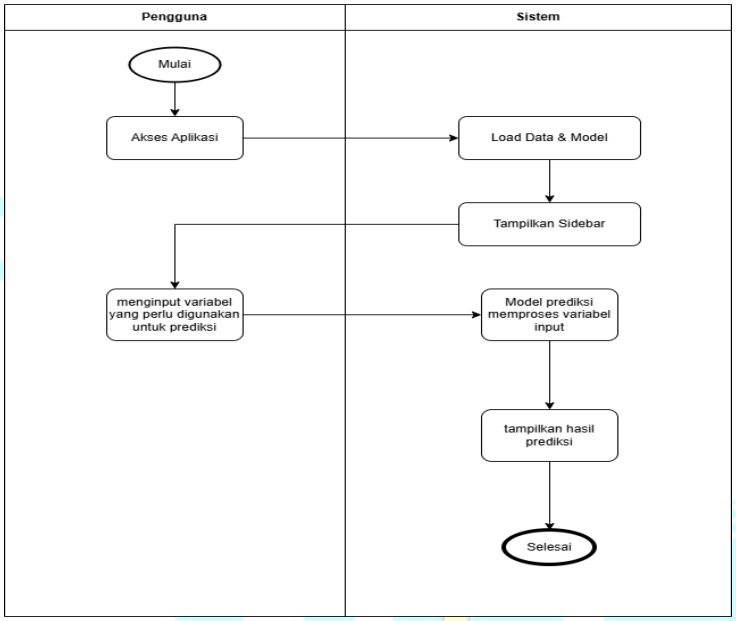
UseCase Diagram menggambarkan interaksi utama antara pengguna (Admin/Manajer) dengan sistem. Pengguna memiliki akses Melakukan Prediksi Penjualan (membutuhkan input parameter), *Usecase* dapat dilihat pada Gambar 3.2.



Gambar 3. 2 *UseCase* Prediksi Penjualan

B. Activity Diagram Sistem

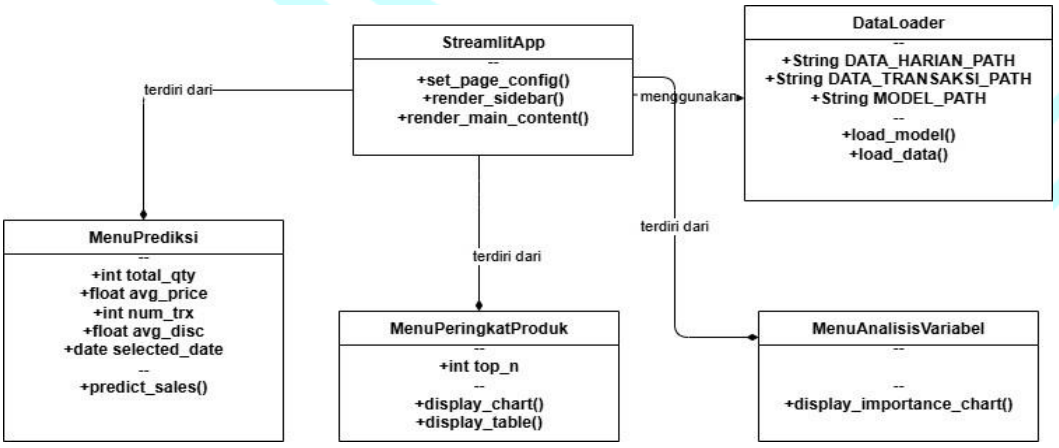
Activity diagram ini menunjukkan alur aplikasi prediksi penjualan antara pengguna dan sistem. Proses dimulai saat pengguna membuka aplikasi, seperti menghitung prediksi.



Gambar 3. 3 Activity Diagram Prediksi Penjualan

C. Class Diagram

Diagram ini menunjukkan bahwa aplikasi dibuat dengan konsep modular dan terstruktur, di mana setiap kelas memiliki tanggung jawab masing-masing sehingga sistem lebih mudah dikembangkan, diuji, dan dipelihara.

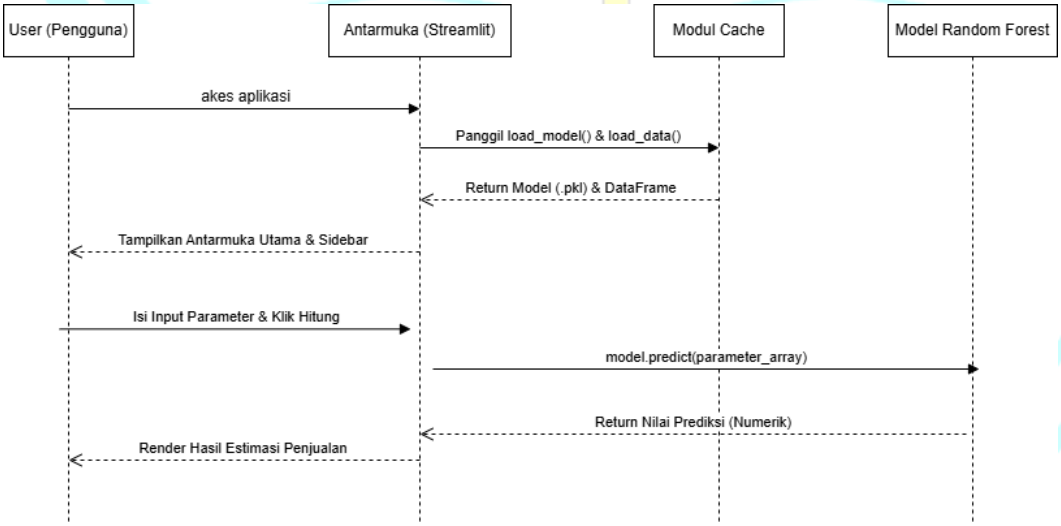


Gambar 3. 4 Class Diagram Prediksi Penjualan

Berdasarkan Gambar 3.4, Class diagram pada sistem ini terdiri dari beberapa kelas utama. StreamlitApp berfungsi menjalankan aplikasi dan mengatur tampilan menu. DataLoader digunakan untuk memuat data dan model machine learning. MenuPrediksi berfungsi melakukan prediksi penjualan berdasarkan input pengguna. MenuPeringkatProduk menampilkan daftar produk terlaris dalam bentuk tabel dan grafik. MenuAnalisisVariabel digunakan untuk menampilkan pengaruh variabel terhadap hasil prediksi melalui *feature importance* model. prediksi. Kelas ini menunjukkan *feature importance* dari model machine learning.

D. Sequence Diagram

Alur proses fitur prediksi penjualan pada aplikasi Streamlit dimulai saat pengguna membuka aplikasi, kemudian sistem memanggil fungsi load_model() dan load_data() melalui modul *cache* untuk mengambil model Random Forest dan dataset. Setelah antarmuka utama ditampilkan, pengguna memilih menu prediksi penjualan dan mengisi parameter input. Selanjutnya sistem mengirim data input ke model Random Forest melalui fungsi predict(). Model mengembalikan nilai prediksi penjualan, lalu hasil estimasi ditampilkan kembali kepada pengguna pada antarmuka aplikasi.



Gambar 3. 5 Squence Diagram Prediksi Penjualan