

## BAB III

### METODE PENELITIAN

#### 3.1 Objek Penelitian

Objek dalam penelitian ini merupakan data penjualan yang diambil dari konveksi *sunny four*. Data yang digunakan yaitu data dari hasil penjualan dari tahun 2020 sampai tahun 2022, berikut adalah tabel perincian pelaksanaan kegiatan penelitian.

Tabel 3. 1 Jadwal Penelitian

Kegiatan	Bulan					
	November	Desember	Januari	Februari	Maret	April
Pengumpula Data						
Analisis Data						
Pengolaha Data						
Pengujian						
Evaluasi						
Tugas akhir						

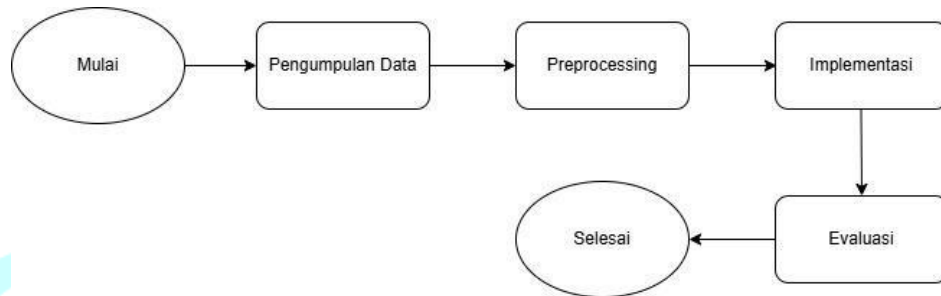
#### 3.2 Peralatan Penelitian

Penelitian ini dilakukan dengan menggunakan peralatan berupa *hardware* dan *software*

1. *Hardware* yang digunakan dalam penelitian ini adalah:
  - a. Laptop Hp 14-dv0068tx
  - b. AMD A6-5200 APU with Radeon (TM) HD Graphic
  - c. Memory size 12 gb
2. *Software* yang digunakan dalam penelitian ini adalah:
  - a. Microsoft word 2013
  - b. Microsoft exel 2013

### 3.3 Prosedur Penelitian

Tahapan penelitian tugas akhir yang akan dilakukan dapat digambarkan sebagai berikut:



Gambar 3. 1 Prosedur Penelitian

#### 3.1.1 Pengumpulan Data

Penelitian dimulai dengan mengumpulkan data hasil penjualan konveksi *sunny four*, data yang dikumpulkan berupa data hasil penjualan dari Januari tahun 2020 sampai Mei 2022 dengan 5 jenis pakaian yang dijual yaitu gamis, baju tidur, daster, pakaian muslim dan kemeja.

Tabel 3. 2 Contoh Data

No	Jenis Produk	Bulan	Jumlah
1	Kemeja	Januari	5
2	Gamis	Januari	4
3	Pakaian Muslim	Januari	10
4	Baju Tidur	Januari	7
5	Daster	Januari	10

#### 3.1.2 PreProcessing

*Pre-processing* data adalah tahap awal dalam pengolahan data yang bertujuan untuk mempersiapkan dataset agar layak dan optimal digunakan dalam analisis atau pembangunan model. Tahap preprocessing merupakan tahapan awal untuk mempersiapkan data agar model yang dibuat bisa lebih baik pada saat training maupun testing. Secara umum, pre-processing data melibatkan serangkaian langkah untuk membersihkan, mengubah, dan

menyiapkan data agar sesuai dengan kebutuhan analisis. Berikut merupakan Langkah - Langkah Pre-Processing Data pada Penelitian ini mencakup .

Tabel 3. 3 Preprocessing Data

Jenis Barang	Januari	Februari	Maret	.....	Desember
Gamis	15	15	60	.....	25
Pakaian Muslim	25	20	15	.....	25
Kemeja	25	15	10	.....	20
baju tidur	25	20	15	.....	30
Daster	35	15	20	.....	35

### 1. Transformasi Data

Pada tahap ini dilakukan transformasi data berupa pemisahan dataset berdasarkan jenis produk, yaitu Kemeja, Pakaian Muslim, Gamis, Baju Tidur, dan Daster. Pemisahan ini bertujuan untuk membangun dan mengevaluasi model prediksi secara terpisah pada setiap jenis produk.

### 2. Pembentukan *Time Index*

Pembentukan time index dilakukan untuk merepresentasikan urutan waktu pada data penjualan. Time index dibuat secara berurutan mulai dari periode awal hingga periode akhir untuk setiap jenis produk. Variabel ini digunakan sebagai indikator waktu dalam analisis deret waktu dan berfungsi sebagai salah satu variabel independen dalam model regresi linear. Pembentukan *time index* secara terpisah untuk setiap jenis produk bertujuan agar urutan waktu tetap konsisten dan tidak tercampur antar kategori produk.

### 3. Rekayasa Fitur (Feature Enggining)

Beberapa fitur baru ditambahkan untuk meningkatkan kualitas data dan memudahkan model dalam mengenali pola, antara lain:

1. MA\_3: Rata-rata bergerak 3 bulan terakhir.
2. Lag\_2: Nilai penjualan 2 bulan sebelumnya.

#### 4. Menangani *Missing Values*

*Missing values* digunakan untuk memeriksa jumlah data yang hilang dalam setiap kolom *dataset*. Dalam analisis data, penting untuk mengetahui apakah terdapat nilai kosong (NaN / *null values*) dalam dataset sebelum melakukan pemrosesan lebih lanjut.

### 3.4 Implementasi

Tahap implementasi dilakukan dengan membangun model prediksi menggunakan algoritma Regresi Linear melalui bahasa pemrograman Python dan library *scikit-learn*. Variabel independen (*X*) yang digunakan meliputi *time\_index*, *lag\_2*, dan *MA\_3*, sedangkan variabel dependen (*y*) adalah jumlah penjualan.

Dataset dibagi menjadi data latih (80%) dan data uji (20%) berdasarkan urutan waktu. Model dibentuk menggunakan fungsi `LinearRegression()` dan dilatih dengan metode `fit(X_train, y_train)` untuk mempelajari hubungan antara variabel input dan jumlah penjualan.

Setelah proses pelatihan selesai, model digunakan untuk melakukan prediksi pada data uji menggunakan `predict(X_test)`. Hasil prediksi selanjutnya digunakan pada tahap evaluasi untuk menghitung nilai error menggunakan RMSE.

### 3.5 Evaluasi

Evaluasi Hasil pengujian model dilakukan untuk menilai sejauh mana model mampu memprediksi variabel yang menjadi tujuan penelitian. Dalam tahap evaluasi ini, prediksi yang dihasilkan oleh model dibandingkan dengan data aktual untuk mengetahui tingkat kesesuaiannya. Beberapa metrik evaluasi yang digunakan antara lain Mean Absolute Error (MAE), Root Mean Square Error (RMSE) dan R-squared ( $R^2$ ), yang masing-masing berfungsi untuk mengukur akurasi dan kesalahan prediksi model. Jika hasil evaluasi menunjukkan bahwa tingkat akurasi masih rendah, maka peneliti dapat melakukan perbaikan, seperti menambah tahapan preprocessing untuk meningkatkan kualitas data atau mencoba metode pemodelan lainnya. Upaya ini dilakukan untuk mengoptimalkan performa model agar mampu menghasilkan prediksi yang lebih akurat dan dapat dijadikan dasar pengambilan keputusan yang lebih tepat berbasis data.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Pengumpulan Data

Tahapan pertama yang dilakukan dalam penelitian ini yaitu pengumpulan data dari hasil penjualan produk di konveksi *Sunny Four*, data yang didapatkan adalah data hasil penjualan produk kemeja, pakaian muslim, Gamis, baju tidur, daster dari bulan Januari dari tahun 2020 sampai bulan Mei tahun 2022, data tersebut akan digunakan untuk melakukan perhitungan prediksi dengan menggunakan algoritma regresi linear.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 145 entries, 0 to 144
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   jenis produk 145 non-null    object
1   Bulan        145 non-null    object
2   Tahun        145 non-null    int64
3   jumlah       145 non-null    int64
dtypes: int64(2), object(2)
memory usage: 4.7+ KB
```

Gambar 4. 1 Isi Dataset

#### 4.2 Preprocessing Data

Tahap *preprocessing* dilakukan dengan mengelompokkan data penjualan produk konveksi Sunny Four berdasarkan jenis produk, yaitu kemeja, pakaian muslim, gamis, baju tidur, dan daster, serta periode bulanan dari Januari 2020 hingga Mei 2022. Data yang sudah dipisahkan ditambahkan fitur *lag-2* sebagai representasi penjualan dua periode sebelumnya serta fitur *moving average* tiga periode (MA(3)) guna menangkap pola tren penjualan. Pada tahap akhir, dilakukan penanganan *missing value* yang muncul akibat pembentukan fitur *lag* dan *moving average* sehingga dataset menjadi lengkap dan siap digunakan dalam proses pemodelan prediksi penjualan.

### 4.2.1 Transformasi Data

Pada tahap ini dilakukan transformasi data berupa pemisahan dataset berdasarkan jenis produk, yaitu Kemeja, Pakaian Muslim, Gamis, Baju Tidur, dan Daster. Pemisahan ini bertujuan untuk membangun dan mengevaluasi model prediksi secara terpisah pada setiap jenis produk.

*Tabel 4. 1 Hasil Transformasi Data*

Sebelum	Sesudah
df_all	df_kemeja
	df_pakaian_muslim
	df_gamis
	df_baju_tidur
	df_daster

Tabel 4.1 Proses transformasi data berupa pemisahan dataset awal (*df\_all*) menjadi beberapa dataframe berdasarkan jenis produk, yaitu *df\_kemeja*, *df\_pakaian\_muslim*, *df\_gamis*, *df\_baju\_tidur*, dan *df\_daster*.

### 4.2.2 Rekayasa Fitur

Pada tahapan ini, dilakukan penambahan fitur untuk memperkaya informasi temporal pada data penjualan. Fitur *month\_num* ditambahkan untuk merepresentasikan urutan bulan dalam satu tahun, sedangkan *time index* digunakan sebagai penanda urutan waktu secara kronologis. Selain itu, dibentuk fitur *lag-2* untuk merepresentasikan nilai penjualan dua periode sebelumnya serta fitur *moving average* tiga periode (MA(3)) guna menangkap pola tren penjualan jangka pendek. Penambahan fitur-fitur ini bertujuan meningkatkan kemampuan model dalam mempelajari pola historis dan tren waktu sehingga menghasilkan prediksi penjualan yang lebih akurat.

```

<class 'pandas.core.frame.DataFrame'>
Index: 29 entries, 0 to 28
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   jenis produk    29 non-null     object
1   Bulan           29 non-null     object
2   Tahun           29 non-null     int64
3   jumlah          29 non-null     int64
4   bulan_num       29 non-null     int64
5   time_index      29 non-null     int64
6   ma_3            27 non-null     float64
7   lag_2           27 non-null     float64
dtypes: float64(2), int64(4), object(2)
memory usage: 2.0+ KB

```

Gambar 4. 2 Dataframe Setelah Penambahan Fitur

Pada Gambar 4.2 menampilkan info *dataset* yang telah siap digunakan untuk tahap implementasi model. Seluruh *dataframe* hasil transformasi memiliki struktur dan kolom yang sama, sesuai dengan yang ditunjukkan pada Gambar 4.2, sehingga data bersifat konsisten dan seragam. Keseragaman struktur ini memastikan bahwa setiap fitur yang digunakan telah tersusun dengan baik dan dapat diproses secara optimal pada tahap pemodelan dan evaluasi prediksi.

#### 4.2.3 Menangani Missing Values

Missing values digunakan untuk memeriksa jumlah data yang hilang dalam setiap kolom *dataset*. Dalam analisis data, penting untuk mengetahui apakah terdapat nilai kosong (NaN / *null values*) dalam dataset sebelum melakukan pemrosesan lebih lanjut.

```

0
jenis produk  0
Bulan         0
Tahun         0
jumlah        0
bulan_num     0
time_index    0
ma_3          0
lag_2         0

dtype: int64

```

Gambar 4. 3 Hasil Missing Values

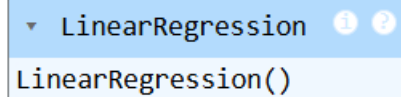
Pada Gambar 4.3 menampilkan ringkasan struktur dataset yang digunakan dalam penelitian. Seluruh variabel bertipe data **int64** dan tidak memiliki nilai

kosong (*missing value*). Variabel yang digunakan terdiri dari variabel utama, yaitu jenis produk, bulan, tahun, dan jumlah, serta variabel turunan seperti `bulan_num`, `time_index`, `ma_3`, dan `lag_2` yang mendukung analisis deret waktu. Dataset ini menunjukkan bahwa data telah melalui tahap pra-pemrosesan dan siap digunakan untuk analisis dan pemodelan.

### 4.3 Implementasi

Setelah seluruh tahap preprocessing selesai, model dilatih menggunakan algoritma regresi linear dari pustaka `scikit-learn` melalui parameter `LinearRegression()`. Proses persiapan data berupa pemabagian dataset menjadi data latih dan data pengujian dengan perbandingan 80 banding 20 atau 80% sebanyak 21 untuk data latih dan 20% sebanyak 6 untuk data uji.

```
model = LinearRegression()
model.fit(X_train, y_train)
```



Gambar 4. 4 Hasil Pembangunan Model Regresi Linear

Pada Gambar 4.5 menunjukkan proses pembuatan dan pelatihan model regresi linear menggunakan bahasa pemrograman Python dengan memanfaatkan library `scikit-learn`. Pada tahap ini, objek model dibentuk dengan memanggil fungsi `LinearRegression()`, kemudian dilakukan proses pelatihan model menggunakan metode `fit()` dengan memasukkan data latih (`X_train`) sebagai variabel independen dan data target (`y_train`) sebagai variabel dependen.

Proses *fitting* ini bertujuan untuk membangun model regresi linear yang mampu mempelajari hubungan antara variabel input dan output berdasarkan data historis. Hasil dari proses ini adalah terbentuknya model regresi linear yang telah memiliki parameter koefisien dan intercept, sehingga model siap digunakan untuk melakukan prediksi pada data uji (testing data) pada tahap selanjutnya.

#### 4.4 Evaluasi

Pada tahap evaluasi, dilakukan pengukuran kinerja model prediksi untuk mengetahui tingkat akurasi hasil peramalan pada masing – masing produk. Evaluasi ini menggunakan beberapa metrik, yaitu *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi ( $R^2$ ), yang bertujuan untuk menilai besar kesalahan prediksi serta kemampuan model dalam menjelaskan variasi data aktual.

Tabel 4. 2 Hasil Evaluasi Model

Produk	MAE	RMSE	R2
Kemeja	1.817729	2.090366	0.856994
Pakaian Muslim	2.065706	2.249839	0.635552
Gamis	1.986012	2.289687	0.555915
Daster	3.327617	3.726705	0.310372
Baju Tidur	3.132719	3.893935	0.503765

Berdasarkan hasil evaluasi kinerja model prediksi time series pada masing-masing produk, diperoleh nilai Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi ( $R^2$ ) yang bervariasi. Produk Kemeja menunjukkan kinerja model terbaik dengan nilai RMSE sebesar 2,09, MAE sebesar 1,82, serta nilai  $R^2$  sebesar 0,857. Nilai  $R^2$  yang mendekati 1 mengindikasikan bahwa model mampu menjelaskan sebagian besar variasi data penjualan Kemeja dengan baik.

Produk Pakaian Muslim dan Gamis juga menunjukkan kinerja model yang cukup baik, dengan nilai RMSE masing-masing sebesar 2,25 dan 2,29, serta nilai  $R^2$  sebesar 0,636 dan 0,556. Hal ini menunjukkan bahwa model masih mampu menangkap pola tren dan fluktuasi penjualan, meskipun tingkat akurasinya berada di bawah produk Kemeja.

Sementara itu, produk Baju Tidur dan Daster memiliki nilai RMSE dan MAE yang relatif lebih besar, yaitu masing-masing di atas 3, serta nilai  $R^2$  yang lebih rendah. Produk Daster menunjukkan nilai  $R^2$  terendah sebesar 0,31, yang

mengindikasikan bahwa kemampuan model dalam menjelaskan variasi data penjualan pada produk ini masih terbatas. Kondisi ini dapat disebabkan oleh pola penjualan yang lebih fluktuatif atau adanya faktor eksternal yang belum terakomodasi dalam model.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model time series yang dibangun paling optimal pada produk Kemeja, cukup baik pada Pakaian Muslim dan Gamis, serta kurang optimal pada produk Baju Tidur dan Daster. Perbedaan kinerja ini menunjukkan bahwa karakteristik data penjualan tiap produk sangat memengaruhi kemampuan model dalam melakukan prediksi.

Setelah model prediksi berhasil dibangun dan di uji tahap selanjutnya adalah melakukan prediksi jumlah penjualan pada periode Juni 2022. Hasil prediksi tersebut kemudian dibandingkan dengan data penjualan aktual periode sebelumnya untuk mengevaluasi kemampuan model dalam mengikuti pola dan tren penjualan yang di gunakan dalam proses pelatihan.

Tabel 4. 3 Hasil Prediksi

Produk	Prediksi Juni 2022
Kemeja	21
Pakaian Muslim	30
Gamis	30
Baju Tidur	28
Daster	31

Pada tabel 4.3 menampilkan hasil prediksi jumlah penjualan pada bulan Juni 2022 untuk masing-masing produk berdasarkan model prediksi yang telah dibangun. Berdasarkan hasil tersebut, produk Daster memiliki nilai prediksi penjualan tertinggi sebesar 31,97, diikuti oleh Pakaian Muslim dan Gamis dengan nilai prediksi masing-masing sebesar 30,78 dan 30,23. Sementara itu, produk Baju Tidur diprediksi memiliki jumlah penjualan sebesar 28,26, dan produk Kemeja menunjukkan nilai prediksi terendah yaitu 21,14. Hasil prediksi ini memberikan gambaran mengenai estimasi penjualan tiap produk pada periode yang akan datang serta dapat digunakan sebagai dasar dalam pengambilan keputusan terkait perencanaan stok dan strategi penjualan.