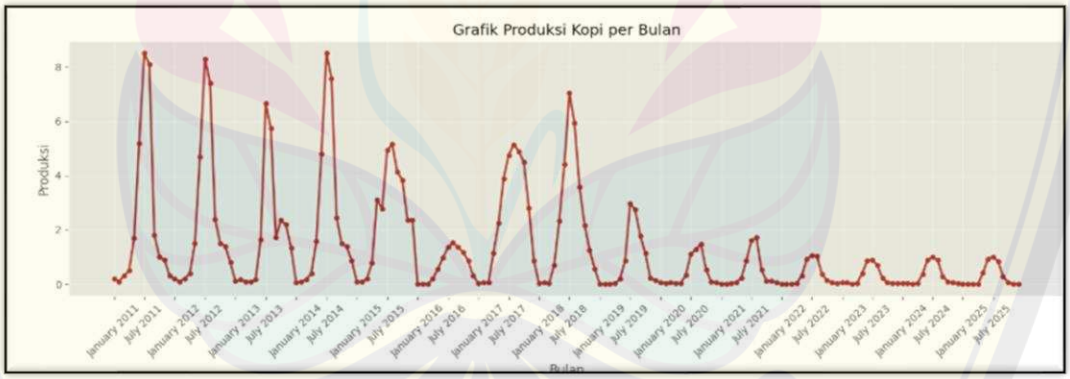
BAB 4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Penelitian ini menggunakan data *time-series* produksi kopi, suhu, curah hujan, dan kelembaban di Indonesia periode Januari 2011–Desember 2025. Data tersebut disusun dalam format bulanan untuk analisis menggunakan metode *Long Short-Term Memory* (LSTM).

4.1.1 Data Hasil Produksi Kopi Indonesia

Produksi kopi Indonesia menunjukkan pola fluktuatif dengan karakteristik musiman. Puncak produksi terjadi pada tahun 2011, 2012, 2014, dan 2018, kemudian cenderung menurun setelah 2018. Pola musiman terlihat dari peningkatan produksi pada sekitar Mei–Juli yang diikuti penurunan hingga akhir dan awal tahun berikutnya. Pada periode 2020–2025, produksi relatif lebih stabil meskipun pola musiman tetap terlihat. Karakteristik tersebut menunjukkan bahwa data produksi kopi sesuai digunakan dalam pemodelan deret waktu (*time-series*) menggunakan metode *Long Short-Term Memory* (LSTM).

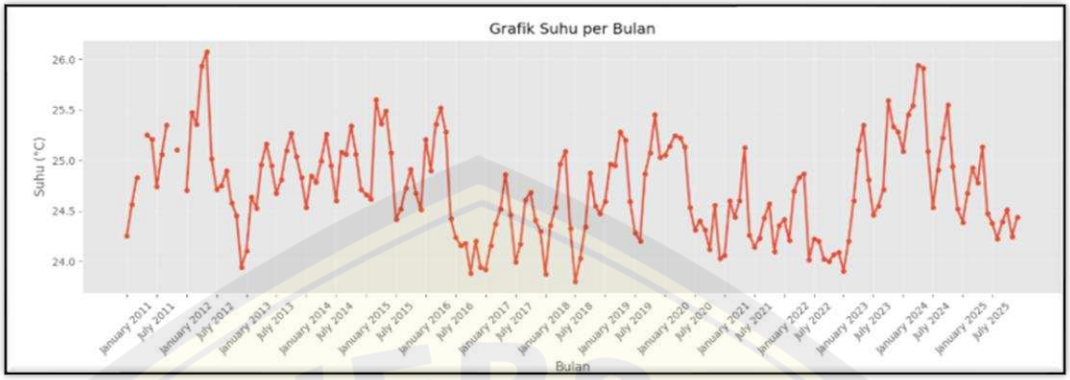


Gambar 4. 1Grafik Hasil Produksi Kopi

4.1.2 Data Suhu

Berdasarkan Gambar 4.2, suhu rata-rata menunjukkan pola yang berfluktuasi selama periode pengamatan. Nilai suhu berada pada kisaran sekitar 24–26°C. Suhu tertinggi terlihat pada akhir tahun 2011 hingga awal tahun 2012 serta kembali meningkat pada tahun 2023–2024. Sementara itu, suhu yang relatif lebih rendah terjadi pada periode 2016–2018 dan kembali terlihat pada tahun 2021–2022. Secara

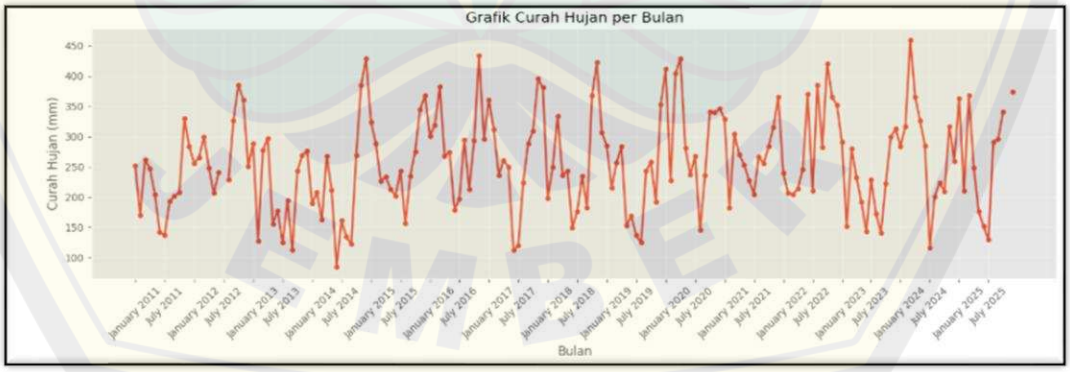
umum, suhu tidak menunjukkan tren kenaikan maupun penurunan yang konsisten, melainkan cenderung stabil dengan fluktuasi dari tahun ke tahun.



Gambar 4. 2Grafik Suhu

4.1.3 Data Curah Hujan

Berdasarkan Gambar 4.3, curah hujan menunjukkan pola yang berfluktuasi cukup tinggi selama periode pengamatan. Nilai curah hujan mengalami kenaikan dan penurunan yang tidak beraturan, dengan beberapa puncak tertinggi terlihat pada tahun 2014, 2016, 2019, 2020, 2022, dan 2024, sedangkan nilai yang relatif rendah muncul pada beberapa periode seperti tahun 2013, 2017, 2019, 2023, dan awal 2025. Berbeda dengan data produksi yang memiliki pola musiman yang jelas, grafik curah hujan tidak menunjukkan pola musiman yang konsisten maupun tren kenaikan atau penurunan yang berkelanjutan, sehingga menggambarkan variabilitas curah hujan yang cukup tinggi dari tahun ke tahun.

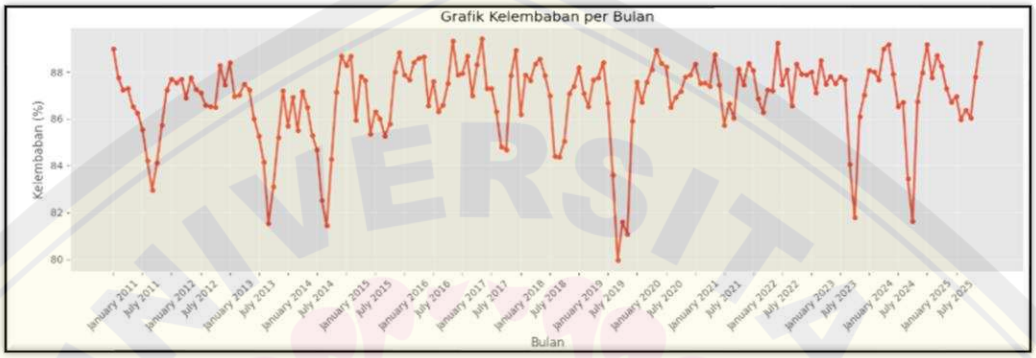


Gambar 4. 3 Grafik Curah Hujan

4.1.4 Data Kelembaban

Berdasarkan Gambar 4.4, kelembaban udara menunjukkan pola yang relatif fluktuatif sederhana selama periode pengamatan dengan kisaran nilai sekitar 80–

89%. Meskipun demikian, terdapat beberapa penurunan yang cukup tajam, terutama pada tahun 2013, 2014, 2019, 2023, dan 2024. Setelah mengalami penurunan, nilai kelembaban kembali meningkat dan berada pada kisaran yang relatif tinggi. Secara umum, kelembaban tidak menunjukkan tren kenaikan maupun penurunan yang konsisten, melainkan berfluktuasi dengan tingkat perubahan yang relatif kecil dibandingkan variabel curah hujan.



Gambar 4. 4Grafik Kelembaban

4.2 Pengolahan Data

Tahap pengolahan data dilakukan sebelum proses pemodelan menggunakan metode *Long Short-Term Memory* (LSTM). Tahapan ini bertujuan untuk memastikan data berada dalam kondisi yang baik sehingga dapat diproses secara optimal oleh model *deep learning*.

4.2.1 Data *Preprocessing* dan *Cleaning*

Pada tahap ini dilakukan *preprocessing* dan *cleaning* data untuk memastikan dataset berada dalam kondisi yang siap digunakan pada proses analisis dan pemodelan LSTM.

a. Informasi Dataset

Tahap pertama pada *preprocessing* adalah menampilkan informasi dataset untuk mengetahui struktur data yang akan digunakan pada penelitian. Informasi yang ditampilkan meliputi jumlah data, jumlah atribut, tipe data setiap kolom, serta statistik deskriptif dataset menggunakan fungsi *head()*, *info()*, *describe()*, dan shape.

	bulan	suhu (rata-rata) (°C)	curah hujan (mm/day)	kelembaban udara (rata-rata) (%)	Hasil Produksi (ribu ton)
0	January 2011	24.256	252.128	88.984	0,20
1	February 2011	24.564	170.708	87.744	0,10
2	March 2011	24.834	261.594	87.232	0,30
3	April 2011	65.352	247.262	87.290	0,50
4	May 2011	25.254	204.768	86.508	1,70

Gambar 4. 5Tampilan Dataset

b. Mengubah Nama Kolom

Pada tahap ini dilakukan perubahan nama kolom agar lebih sederhana dan konsisten sehingga memudahkan proses pemrograman. Perubahan nama kolom tidak mengubah isi data, melainkan hanya menyesuaikan penamaan atribut agar lebih mudah digunakan pada proses analisis dan pemodelan.

	bulan	suhu (rata-rata) (°C)	curah hujan (mm/day)	kelembaban udara (rata-rata) (%)	Hasil Produksi (ribu ton)
0	January 2011	24.256	252.128	88.984	0,20
1	February 2011	24.564	170.708	87.744	0,10
2	March 2011	24.834	261.594	87.232	0,30
3	April 2011	65.352	247.262	87.290	0,50
4	May 2011	25.254	204.768	86.508	1,70

Gambar 4. 6Sebelum Nama Kolom Diubah

	tanggal	suhu	curah_hujan	kelembapan	produksi
0	January 2011	24.256	252.128	88.984	0,20
1	February 2011	24.564	170.708	87.744	0,10
2	March 2011	24.834	261.594	87.232	0,30
3	April 2011	65.352	247.262	87.290	0,50
4	May 2011	25.254	204.768	86.508	1,70

Gambar 4. 7Setelah Nama Kolom Diubah

c. Mengecek *Missing value*

Pengecekan *missing value* dilakukan untuk memastikan tidak terdapat data yang kosong pada setiap atribut. Berdasarkan hasil pengecekan pada notebook bahwa tidak ditemukan *missing value* pada seluruh variabel, sehingga data dapat langsung digunakan pada tahap selanjutnya tanpa proses imputasi.

```
tanggal       0
suhu          0
curah_hujan   0
kelembapan    0
produksi      0
dtype: int64
```

Gambar 4. 8Hasil Pengecekan *Missing value*

d. Mengonversi Format Tanggal menjadi *Datetime*

Pada tahap ini kolom tanggal dikonversi ke format *datetime* menggunakan library *Pandas*. Proses ini dilakukan agar data dapat dikenali sebagai data deret waktu (*time series*), sehingga memudahkan proses pengurutan data dan analisis pada tahapan selanjutnya.

```
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   bulan                           180 non-null    object
 1   suhu (rata-rata) (°C)           180 non-null    float64
 2   curah hujan (mm/day)            180 non-null    float64
 3   kelembaban udara (rata-rata) (%) 180 non-null   float64
 4   Hasil Produksi (ribu ton)       180 non-null    object
dtypes: float64(3), object(2)
```

Gambar 4. 9Sebelum Mengubah Format data

```
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   tanggal      180 non-null    datetime64[ns]
 1   suhu         180 non-null    float64
 2   curah_hujan  180 non-null    float64
 3   kelembapan   180 non-null    float64
 4   produksi     180 non-null    object
dtypes: datetime64[ns](1), float64(3), object(1)
```

Gambar 4. 10Setelah Mengubah Format Data

e. Sorting Data

Setelah format tanggal berhasil dikonversi, data diurutkan berdasarkan waktu dari periode paling lama hingga paling baru. Pengurutan ini bertujuan agar urutan

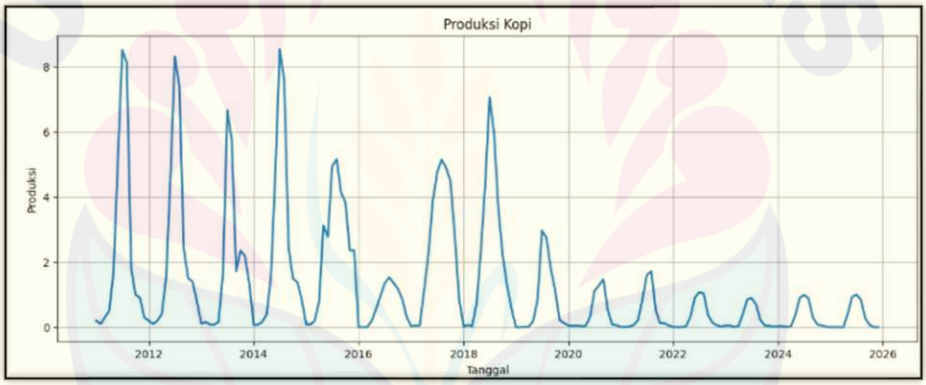
data sesuai dengan kronologi waktu sehingga dapat digunakan pada proses pemodelan LSTM yang memerlukan data berurutan.

f. Konversi Data Numerik

Mengubah atribut numerik ke tipe data yang sesuai agar dapat diproses oleh model. Konversi dilakukan pada variabel produksi, suhu, curah hujan, dan kelembaban sehingga seluruh data numerik memiliki format yang konsisten dan siap digunakan pada proses analisis berikutnya.

4.2.2 *Exploratory Data Analysis* (EDA)

Pada tahap *Exploratory Data Analysis* (EDA), dilakukan visualisasi terhadap data produksi kopi untuk mengetahui pola dan tren produksi selama periode pengamatan. Visualisasi ini bertujuan memberikan gambaran awal mengenai karakteristik data *time series* sebelum dilakukan proses *feature engineering* dan pemodelan menggunakan *Long Short-Term Memory* (LSTM).



Gambar 4. 11EDA

4.2.3 *Feature engineering*

Pada tahap ini dilakukan *feature engineering* untuk menghasilkan fitur tambahan yang dapat membantu model LSTM dalam mempelajari pola temporal dan musiman pada data. Fitur yang dibentuk meliputi produksi_lag1, produksi_lag2, produksi_lag3, serta representasi siklik bulan menggunakan bulan_sin dan bulan_cos. Setelah proses ini selesai, dataset siap digunakan pada tahap pemodelan.

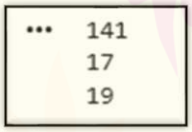
4.2.4 Variabel Penelitian

Berdasarkan hasil *feature engineering*, variabel yang digunakan dalam penelitian terdiri atas produksi, suhu, curah hujan, kelembaban, produksi_lag1,

produksi_lag2, produksi_lag3, bulan_sin, dan bulan_cos. Variabel-variabel tersebut digunakan sebagai fitur masukan model LSTM untuk memprediksi produksi kopi pada periode berikutnya.

4.2.5 Pembagian Data (Data Splitting)

Pada tahap ini dataset dibagi menggunakan metode *time-series split* dengan proporsi 80% data *training*, 10% data *validation*, dan 10% data *testing*. Pembagian dilakukan berdasarkan urutan waktu sehingga data periode awal digunakan sebagai data pelatihan, periode berikutnya sebagai data validasi, dan periode terakhir sebagai data pengujian. Hasil pembagian menghasilkan 141 data *training*, 17 data *validation*, dan 19 data *testing*. Pembagian ini dilakukan sebelum proses normalisasi untuk mencegah terjadinya data leakage, sehingga informasi dari data validasi dan data pengujian tidak memengaruhi proses pelatihan model.



Gambar 4. 12Hasil Split Data

4.2.6 Normalisasi Data

Setelah pembagian data, dilakukan normalisasi menggunakan *Min-Max Scaling* untuk menyamakan rentang nilai fitur dan variabel target. Penelitian ini menggunakan *feature_scaler* untuk seluruh fitur masukan dan *target_scaler* untuk variabel target (produksi). Proses *fit_transform()* hanya diterapkan pada data *training*, sedangkan data *validation* dan *testing* menggunakan *transform()* dengan parameter dari data *training* guna mencegah data leakage.

4.2.7 *Sliding window*

Pada tahap ini data diubah menjadi pasangan input dan target berbentuk *sequence* menggunakan metode *Sliding window*. Window yang digunakan sepanjang 12 bulan, sehingga model memanfaatkan informasi produksi dan variabel iklim selama 12 bulan sebelumnya untuk memprediksi produksi kopi tiga bulan ke depan (3-step ahead forecasting). Hasil dari proses ini kemudian digunakan sebagai masukan pada tahap pembentukan dataset LSTM.

4.2.8 Pemodelan *Long Short-Term Memory* (LSTM)

Tahap ini bertujuan untuk membangun model *Long Short-Term Memory* (LSTM) yang digunakan untuk mempelajari pola hubungan antara data historis produksi kopi dan variabel iklim sehingga dapat menghasilkan prediksi produksi pada periode berikutnya.

a. Membentuk Dataset LSTM

Data hasil proses sliding window selanjutnya dibentuk menjadi dataset yang siap digunakan oleh model Long Short-Term Memory (LSTM). Dataset terdiri atas data masukan (input) yaitu X_train, X_val, dan X_test, serta data target (output) yaitu y_train, y_val, dan y_test. Berdasarkan proses pembentukan dataset diperoleh sebanyak 127 sampel data training, 15 sampel data validation, dan 17 sampel data testing. Setiap data masukan memiliki dimensi (12, 9) yang menunjukkan bahwa setiap sampel terdiri atas 12 time step dengan 9 fitur pada setiap time step. Sementara itu, data target memiliki dimensi (1) yang merepresentasikan satu nilai produksi kopi yang diprediksi, yaitu produksi kopi tiga bulan setelah periode terakhir pada jendela masukan (3-step ahead forecasting). Dataset yang telah dibentuk kemudian digunakan pada proses pelatihan, validasi, dan pengujian model LSTM.

```
====================================
TRAIN
(127, 12, 9)
(127, 1)
====================================
VALIDATION
(15, 12, 9)
(15, 1)
====================================
TEST
(17, 12, 9)
(17, 1)
```

Gambar 4. 13Hasil Pembentukan Dataset

b. *Callback*

Pada proses pelatihan digunakan *callback* *EarlyStopping* dan *ReduceLROnPlateau* yang memantau nilai *validation loss* (val_loss). *EarlyStopping* menghentikan pelatihan apabila tidak terjadi peningkatan selama 20 *epoch* serta mengembalikan bobot terbaik model, sedangkan *ReduceLROnPlateau* menurunkan *learning rate* sebesar 50% apabila nilai *validation loss* tidak membaik

selama 8 *epoch*. Penggunaan kedua *callback* ini bertujuan untuk mengoptimalkan proses pelatihan dan mengurangi risiko overfitting..

c. Membangun Model LSTM

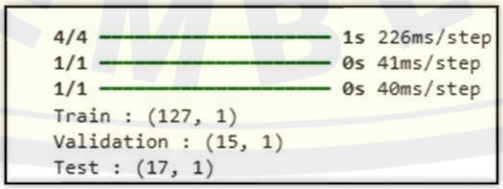
Model *Long Short-Term Memory* (LSTM) dibangun menggunakan dua layer LSTM. Layer pertama terdiri atas 64 unit dengan return_*sequence*s=True, *dropout*=0,2, dan recurrent_*dropout*=0,2. Layer kedua terdiri atas 32 unit. Setelah masing-masing layer LSTM ditambahkan *Dropout* sebesar 0,2. Selanjutnya digunakan *Dense layer* sebanyak 16 neuron dengan fungsi aktivasi *ReLU* dan *Dense output layer* sebanyak 1 neuron. Model dikompilasi menggunakan *optimizer Adam* dengan *learning rate* 0,005 dan fungsi loss *Mean Squared Error* (MSE).

d. *Training* Model

Pada tahap ini, model LSTM dilatih menggunakan data *training* dan dievaluasi menggunakan data *validation* selama proses pelatihan. Proses *training* menggunakan maksimum 100 *epoch*, *batch size* 8, serta menerapkan *callback EarlyStopping* dan *ReduceLROnPlateau*. Pengaturan tersebut bertujuan agar model memperoleh hasil yang optimal sekaligus mencegah terjadinya overfitting.

4.2.9 Prediksi

Setelah proses *training* selesai, model LSTM digunakan untuk melakukan prediksi pada data *training*, *validation*, dan *testing*. Hasil prediksi masih berada dalam bentuk data yang telah dinormalisasi sehingga belum dapat diinterpretasikan secara langsung. Proses prediksi menghasilkan 127 data prediksi untuk data *training*, 15 data prediksi untuk data *validation*, dan 17 data prediksi untuk data *testing*, dengan masing-masing menghasilkan satu nilai prediksi pada setiap sampel. Hasil prediksi tersebut selanjutnya digunakan pada proses *inverse transform* dan evaluasi model.



Gambar 4. 14Hasil Prediksi

4.2.10 Invers Transform

Hasil prediksi dan data aktual kemudian dikembalikan ke skala nilai produksi sebenarnya menggunakan *inverse transform* pada target scaler. Tahapan ini bertujuan agar hasil prediksi dapat dibandingkan secara langsung dengan data aktual dan digunakan pada proses evaluasi model.

4.2.11 Perbandingan dengan Baseline Persistence

Sebagai pembanding, penelitian ini menggunakan Baseline Persistence (Naive Model) yang memprediksi nilai produksi berdasarkan nilai produksi terakhir. Hasil pengujian menunjukkan bahwa Baseline Persistence memperoleh nilai R^2 sebesar -1,8285, sedangkan model LSTM memperoleh nilai R^2 yang lebih tinggi. Nilai R^2 negatif menunjukkan bahwa Baseline Persistence kurang mampu memodelkan pola produksi kopi, sehingga performanya lebih rendah dibandingkan model LSTM.

```
Baseline R²
-1.8285263862557426
```

Gambar 4. 15Metode Perbandingan Sederhana

4.2.12 Evaluasi Model

Setelah proses prediksi selesai, dilakukan evaluasi untuk mengetahui performa model LSTM dengan hasil sebagai berikut :

Tabel 4. 1Hasil Evaluasi Model

MAE	0,1138 ribu ton	Rata-rata kesalahan prediksi model sebesar 0,1138 ribu ton terhadap data aktual.
RMSE	0,1456 ribu ton	Nilai RMSE lebih besar daripada MAE, yang menunjukkan adanya beberapa prediksi dengan kesalahan yang lebih besar dibandingkan rata-rata.
MAPE	17694181402021706.00%	Nilai MAPE sangat besar karena terdapat nilai aktual yang mendekati nol sehingga menghasilkan persentase kesalahan yang sangat tinggi.

R2	0,8435	Model mampu menjelaskan sekitar 84,35% variasi data produksi kopi.

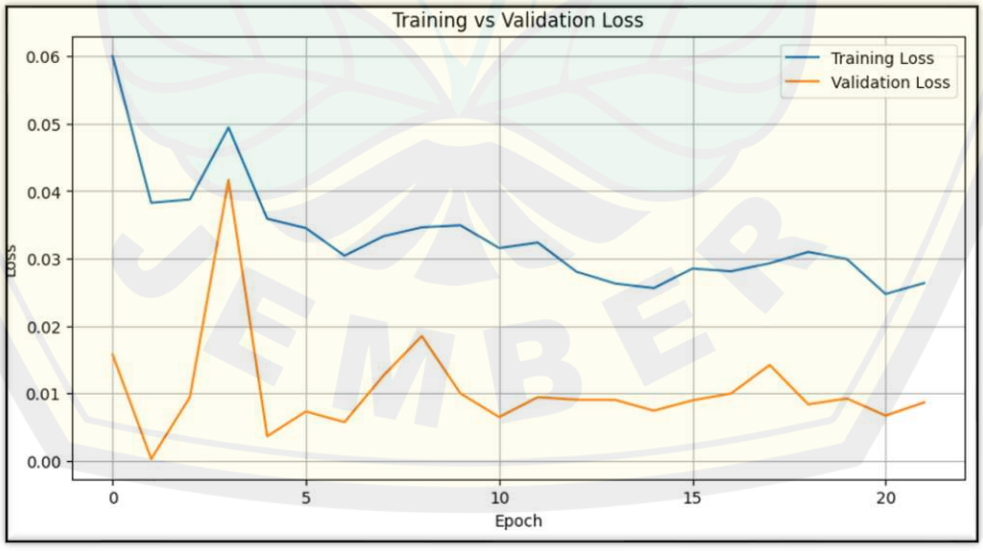
```
MAE  : 0.1138
RMSE : 0.1456
MAPE : 17694181402021706.00%
R²   : 0.8435
```

Gambar 4. 16Hasil Evaluasi Model

4.2.13 Visualisasi Hasil

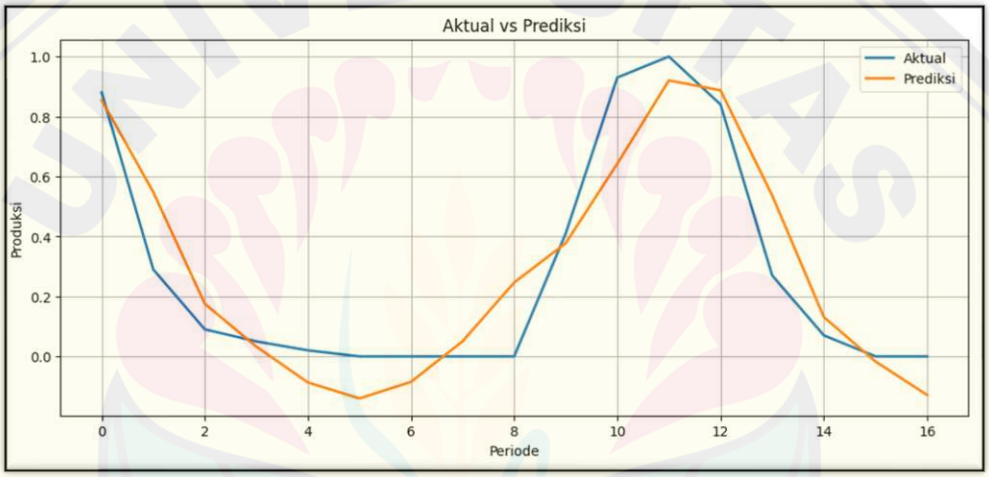
Visualisasi dilakukan dengan membandingkan data aktual dan hasil prediksi pada data *testing* untuk melihat kemampuan model LSTM dalam mengikuti pola data. Hasilnya sebagai berikut:

Berdasarkan Gambar 4.16, nilai training loss menunjukkan kecenderungan menurun dari awal hingga akhir proses pelatihan meskipun masih mengalami fluktuasi pada beberapa epoch. Sementara itu, validation loss juga mengalami fluktuasi, namun secara umum berada pada nilai yang relatif rendah dibandingkan training loss. Hal ini menunjukkan bahwa model mampu mempelajari pola data selama proses pelatihan dan mempertahankan performa yang cukup baik pada data validasi. Proses pelatihan berhenti pada sekitar epoch ke-22 karena mekanisme EarlyStopping, sehingga model menggunakan bobot terbaik yang diperoleh selama pelatihan untuk menghindari pelatihan yang berlebihan (overtraining).



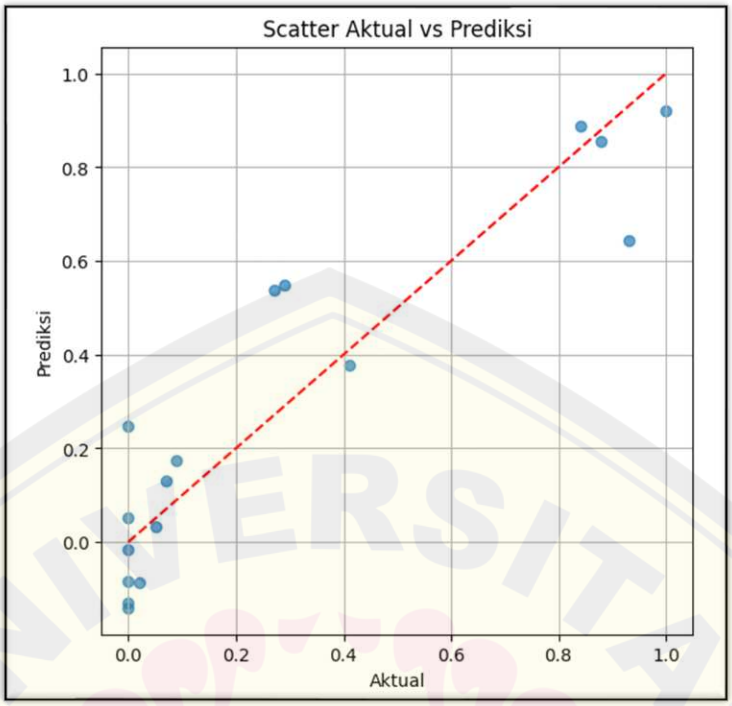
Gambar 4. 17Visualisasi *Training* vs *Validation loss*

Berdasarkan Gambar 4.17, hasil prediksi model LSTM secara umum mampu mengikuti pola perubahan data aktual pada data pengujian. Kurva prediksi berhasil menggambarkan tren penurunan dan peningkatan produksi kopi pada sebagian besar periode pengamatan. Meskipun demikian, masih terdapat beberapa perbedaan antara nilai prediksi dan nilai aktual, terutama pada periode ketika terjadi perubahan produksi yang cukup signifikan. Perbedaan tersebut menunjukkan bahwa model belum mampu menangkap seluruh fluktuasi data secara sempurna. Namun secara keseluruhan, hasil prediksi masih memiliki pola yang mendekati data aktual, sehingga menunjukkan bahwa model memiliki kemampuan yang baik dalam memprediksi produksi kopi berdasarkan data historis dan variabel iklim.



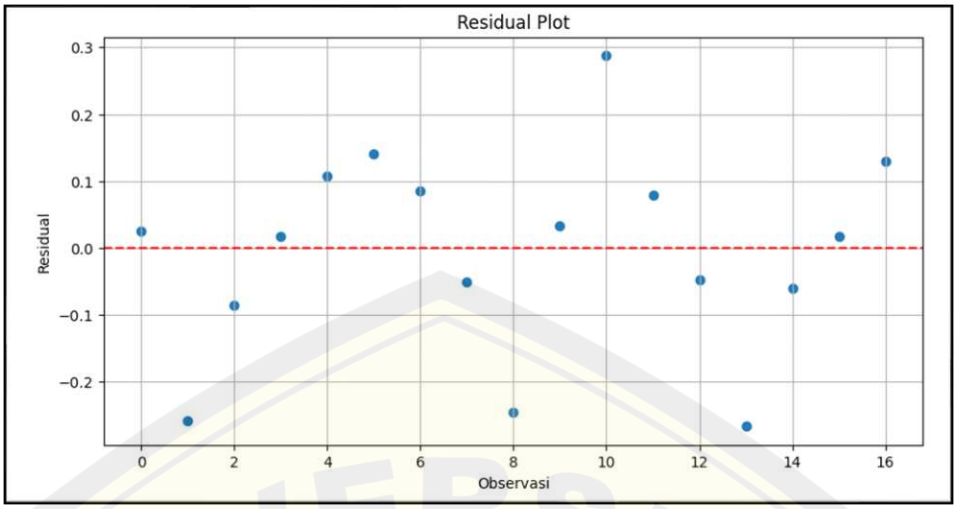
Gambar 4. 18Visualisasi Aktual vs Prediksi

Berdasarkan Gambar 4.18, scatter plot menunjukkan hubungan antara nilai aktual dan nilai prediksi pada data pengujian. Sebagian besar titik berada di sekitar garis diagonal, yang menunjukkan bahwa hasil prediksi memiliki kesesuaian yang cukup baik dengan nilai aktual. Namun, masih terdapat beberapa titik yang menyimpang dari garis diagonal, terutama pada nilai produksi yang rendah maupun pada beberapa nilai produksi yang tinggi. Penyimpangan tersebut menunjukkan bahwa model masih menghasilkan kesalahan prediksi pada beberapa observasi. Meskipun demikian, secara keseluruhan pola sebaran titik menunjukkan hubungan yang cukup kuat antara nilai aktual dan nilai prediksi. Hal ini didukung oleh nilai koefisien determinasi (R^2) sebesar 0,8435, yang menunjukkan bahwa model mampu menjelaskan sekitar 84,35% variasi data produksi kopi pada data pengujian.



Gambar 4. 19Scater Aktual vs Prediksi

Berdasarkan Gambar 4.19, nilai residual tersebar di sekitar garis nol tanpa membentuk pola tertentu. Penyebaran residual terjadi baik di atas maupun di bawah garis nol, yang menunjukkan bahwa kesalahan prediksi tidak bersifat sistematis. Sebagian besar residual memiliki nilai yang relatif kecil, meskipun masih terdapat beberapa observasi dengan residual yang lebih besar. Kondisi tersebut menunjukkan bahwa model secara umum mampu menghasilkan prediksi yang cukup baik, walaupun masih terdapat beberapa data yang belum dapat diprediksi secara o ptimal. Dengan demikian, pola penyebaran residual mengindikasikan bahwa model telah mampu menangkap sebagian besar pola pada data produksi kopi tanpa menunjukkan kecenderungan kesalahan yang terarah.



Gambar 4. 20Residual Aktual vs Prediksi

4.2.14 Interpretasi Shap

Setelah model LSTM dilatih dan dievaluasi, dilakukan interpretasi menggunakan SHAP (*SHapley Additive exPlanations*) untuk mengetahui kontribusi setiap fitur terhadap hasil prediksi serta mengurangi sifat *black-box* pada model LSTM.

a. Persiapan Data

Sebelum menghitung nilai SHAP, disiapkan *background data* dari sebagian data *training* sebagai acuan (*baseline*) *GradientExplainer* serta data *testing* untuk menghitung kontribusi setiap fitur terhadap hasil prediksi model.

```
Background : (50, 12, 9)
Explain : (17, 12, 9)
```

Gambar 4. 21Hasil Pembentukan Data Shap

b. Proses Perhitungan SHAP

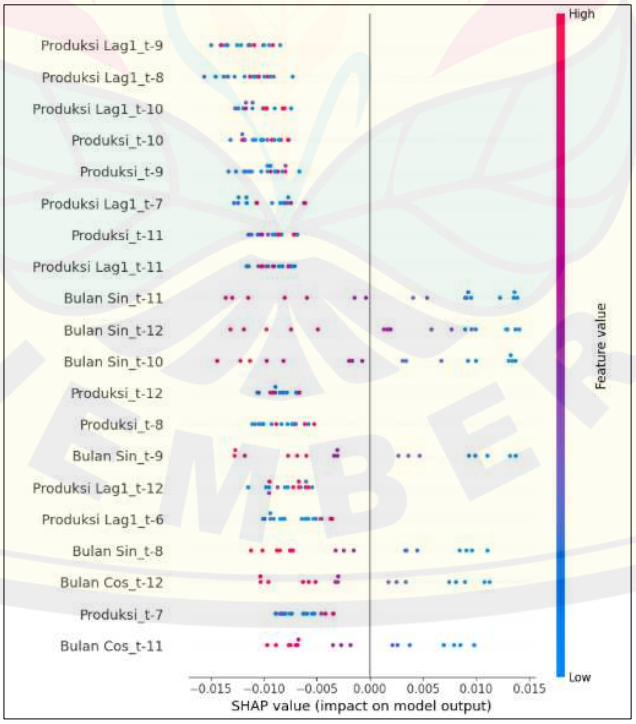
Nilai SHAP berhasil dihitung menggunakan *GradientExplainer* pada model LSTM yang telah dilatih. Hasil perhitungan menghasilkan nilai kontribusi setiap fitur terhadap prediksi produksi kopi. Nilai SHAP yang diperoleh masih berbentuk array multidimensi sehingga dilakukan penyesuaian bentuk data sebelum proses visualisasi.

c. SHAP Skenario 1 (Menggunakan Seluruh Fitur)

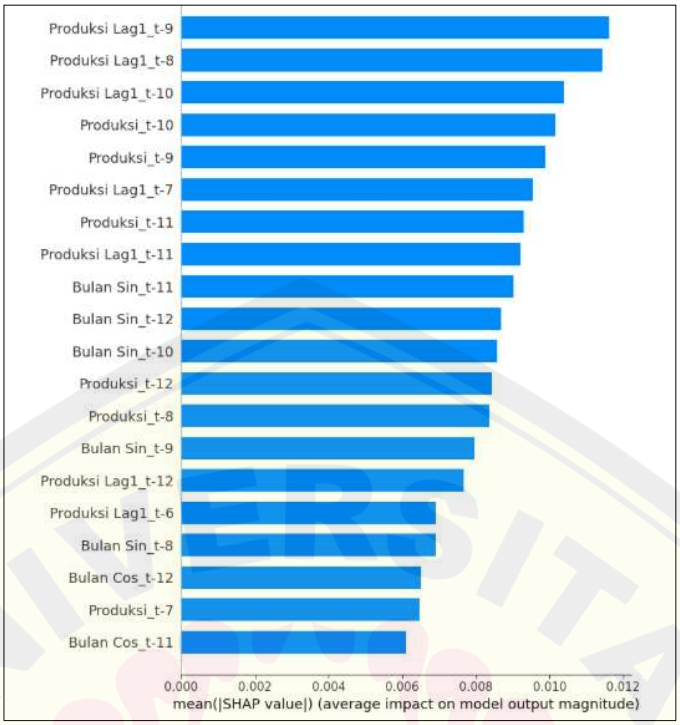
Pada skenario pertama, visualisasi SHAP dilakukan dengan menggunakan seluruh fitur yang menjadi masukan model. Hasil visualisasi menunjukkan tingkat

kontribusi masing-masing fitur terhadap prediksi produksi kopi berdasarkan nilai SHAP yang dihasilkan.

Berdasarkan Gambar 4.20 dan Gambar 4.21, hasil interpretasi menggunakan SHAP menunjukkan bahwa fitur produksi historis (lag produksi) merupakan faktor yang paling dominan dalam memengaruhi hasil prediksi model LSTM. Hal tersebut terlihat dari nilai mean absolute SHAP tertinggi yang dimiliki oleh fitur Produksi Lag1_t-9, diikuti oleh Produksi Lag1_t-8, Produksi Lag1_t-10, Produksi_t-10, dan Produksi_t-9. Hasil ini menunjukkan bahwa informasi produksi pada periode sebelumnya memiliki kontribusi terbesar dalam proses pembentukan prediksi. Selain variabel produksi historis, beberapa fitur musiman yang direpresentasikan oleh Bulan Sin dan Bulan Cos juga memberikan kontribusi terhadap hasil prediksi, meskipun pengaruhnya lebih kecil dibandingkan variabel produksi historis. Sementara itu, variabel iklim seperti suhu memiliki nilai mean absolute SHAP yang relatif rendah sehingga kontribusinya terhadap prediksi model lebih kecil dibandingkan variabel produksi historis. Secara keseluruhan, hasil SHAP menunjukkan bahwa model LSTM lebih mengandalkan informasi produksi historis dibandingkan variabel iklim dalam memprediksi produksi kopi.



Gambar 4. 22Visualisasi *SHAP Summary Plot* (Seluruh Fitur)



Gambar 4. 23Visualisasi *SHAP Bar Plot*

	Feature	Mean |SHAP|
0	Produksi Lag1_t-9	1.162289e-02
1	Produksi Lag1_t-8	1.143029e-02
2	Produksi Lag1_t-10	1.040349e-02
3	Produksi_t-10	1.017028e-02
4	Produksi_t-9	9.895106e-03
...	...	...
103	Suhu_t-4	3.117297e-06
104	Suhu_t-5	2.745372e-06
105	Suhu_t-1	1.633946e-06
106	Suhu_t-2	1.395182e-06
107	Suhu_t-3	6.802049e-07

108 rows × 2 columns

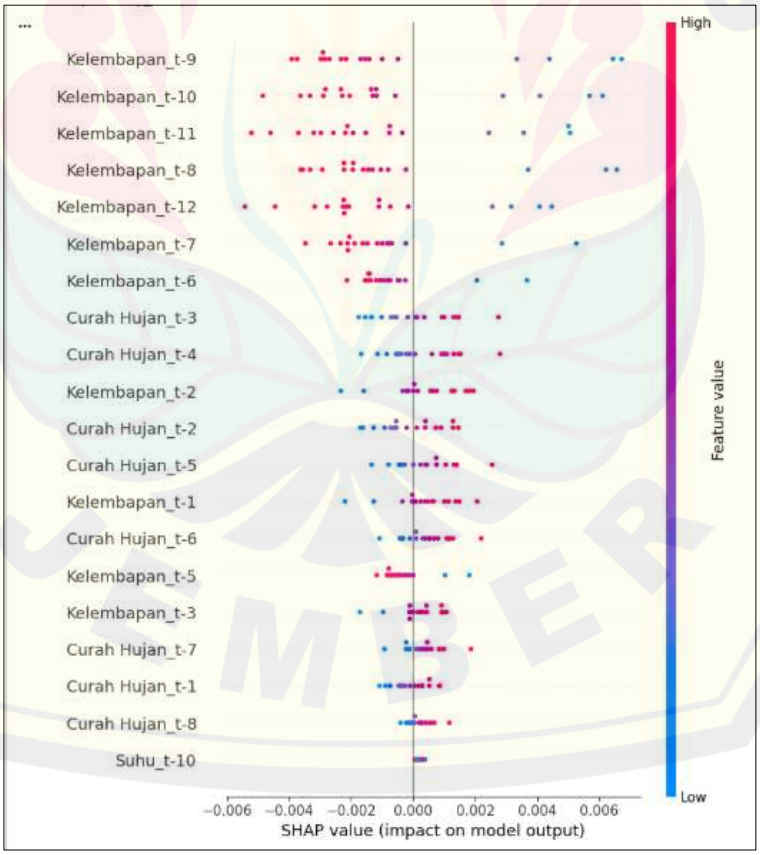
Gambar 4. 24Hasil Perhitungan Mean Absolute SHAP

d. SHAP Skenario 2 (Variabel Iklim)

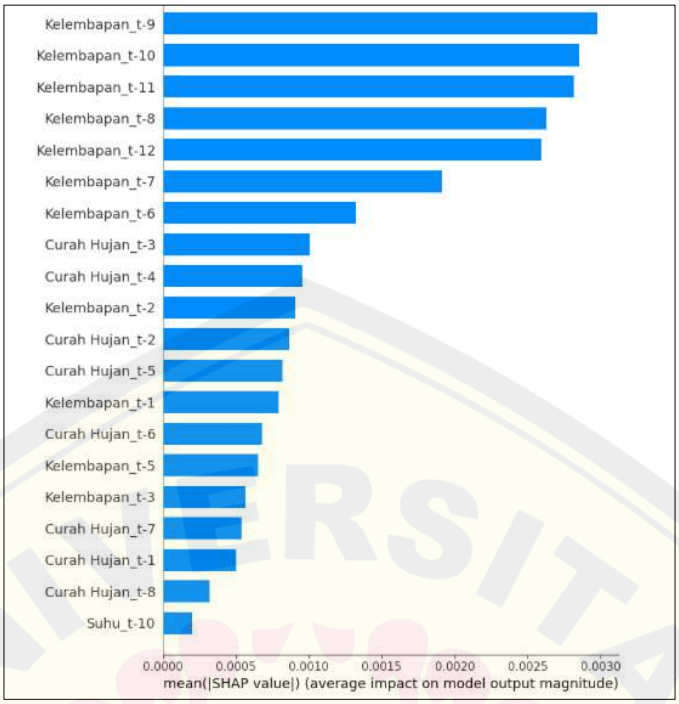
Pada skenario kedua, visualisasi difokuskan pada variabel iklim sehingga pengaruh suhu, curah hujan, dan kelembaban terhadap hasil prediksi dapat diamati

secara lebih jelas. Analisis ini dilakukan untuk mengetahui kontribusi masing-masing variabel iklim dalam proses prediksi produksi kopi.

Berdasarkan Gambar 4.22 dan Gambar 4.23, hasil interpretasi SHAP menunjukkan bahwa variabel kelembapan merupakan faktor iklim yang memberikan kontribusi terbesar terhadap prediksi model. Hal ini ditunjukkan oleh nilai mean absolute SHAP tertinggi yang dimiliki oleh fitur Kelembapan_t-9, Kelembapan_t-10, Kelembapan_t-11, Kelembapan_t-8, dan Kelembapan_t-12. Selain itu, variabel curah hujan juga memberikan kontribusi terhadap hasil prediksi, namun dengan pengaruh yang lebih rendah dibandingkan kelembapan. Sementara itu, variabel suhu memiliki nilai mean absolute SHAP paling kecil sehingga kontribusinya terhadap prediksi model relatif rendah. Secara keseluruhan, hasil SHAP menunjukkan bahwa kelembapan merupakan variabel iklim yang paling berpengaruh terhadap prediksi produksi kopi, diikuti oleh curah hujan, sedangkan suhu memberikan kontribusi paling kecil.



Gambar 4. 25Visualisasi SHAP Summary Plot (Variabel Iklim)



Gambar 4. 26Visualisasi *SHAP Bar Plot* (Variabel Iklim)

	Feature	Mean SHAP
11	Kelembapan_t-9	0.002976
8	Kelembapan_t-10	0.002855
5	Kelembapan_t-11	0.002819
14	Kelembapan_t-8	0.002629
2	Kelembapan_t-12	0.002595
17	Kelembapan_t-7	0.001912
20	Kelembapan_t-6	0.001322
27	Curah Hujan_t-3	0.001005
24	Curah Hujan_t-4	0.000952
32	Kelembapan_t-2	0.000908

Gambar 4. 27Hasil Perhitungan Mean Absolute SHAP