

BAB 3. METODOLOGI PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan metode kuantitatif, yaitu pendekatan yang berfokus pada pengolahan data numerik dan analisis statistik untuk memperoleh hasil yang objektif dan terukur (van Klompenburg et al., 2020). Metode ini digunakan untuk mengidentifikasi hubungan antarvariabel melalui analisis berbasis data (Kulyal & Saxena, 2022). Pendekatan kuantitatif memanfaatkan data yang terstruktur dan terukur sehingga memungkinkan pengujian hipotesis serta penarikan kesimpulan secara umum (van Klompenburg et al., 2020). Selain itu, metode kuantitatif menggunakan teknik analisis yang sistematis untuk memastikan validitas dan reliabilitas hasil penelitian (Swamy et al., 2025). Oleh karena itu, penelitian ini menggunakan metode kuantitatif untuk memprediksi produksi kopi yang dipengaruhi faktor iklim secara objektif dan terukur berdasarkan data yang tersedia (Jiabul Hoque et al., 2024).

3.2 Objek Penelitian

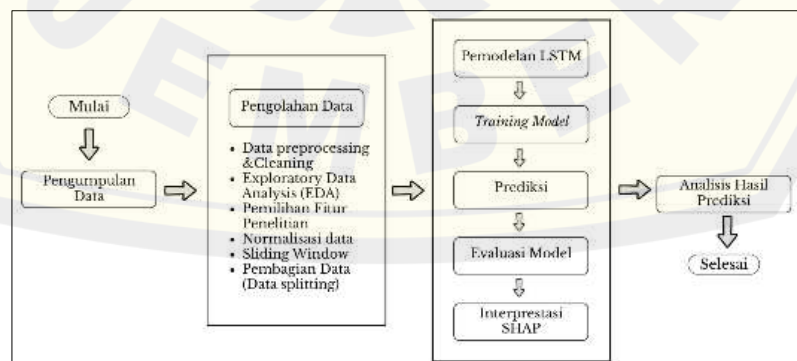
Objek penelitian yang digunakan adalah komoditas kopi di Indonesia.

3.3 Tempat dan Waktu Penelitian

Penelitian dilakukan di Fakultas Ilmu Komputer Universitas Jember menggunakan data yang diperoleh dari (BPS, n.d.) dan (NASA POWER, n.d.) . Dataset yang digunakan merupakan data bulanan periode Januari 2011 hingga Desember 2025.

3.4 Tahapan Penelitian

Tahapan penelitian yang dilakukan dapat diamati pada gambar 3.1 dibawah ini:



Gambar 3. 1 Tahapan Penelitian

3.4.1 Teknik Pengumpulan Data

Penelitian ini diawali dengan pengumpulan data dan informasi untuk memprediksi hasil produksi dengan dua teknik yaitu studi literatur dan observasi.

a. Studi Literatur

Studi literatur dilakukan dengan mengumpulkan informasi dari jurnal, artikel, website, dan referensi lain yang relevan sebagai landasan penelitian.

b. Observasi

Observasi dilakukan dengan mengumpulkan data historis produksi kopi perkebunan besar di Indonesia serta data faktor iklim periode 2011–2025 sebagai atribut dalam model prediksi (Tabel 3.1). Data iklim diperoleh dari NASA POWER menggunakan lima titik koordinat yang mewakili wilayah penghasil kopi di Indonesia, yaitu Aceh, Sumatra Utara, Bengkulu, Lampung, dan Sumatra Selatan., kemudian dirata-ratakan setiap bulan untuk mewakili representasi kondisi iklim nasional agar sesuai dengan data target berupa produksi kopi nasional (Hoffmann et al., 2016).

Tabel 3. 1 Atribut Data

No	Atribut	Parameter
1.	Bulan/Tahun	10 tahun terakhir
2.	Hasil Produksi	Ribu ton
3.	Suhu	Rata-rata suhu bulanan (°C per bulan)
4.	Curah Hujan	Rata-rata curah hujan bulanan (mm per bulan)
5.	Kelembaban	Rata-rata kelembaban (% per bulan)

3.4.2 Pengolahan Data

Pada tahap ini dilakukan pengolahan data sebagai persiapan sebelum proses pelatihan dan pengujian model.

a. *Data preprocessing & Cleaning*

Tahapan ini bertujuan untuk menyiapkan data supaya dapat mencegah data dari kesalahan, ketidakkonsistenan, dan hal hal yang dapat mempengaruhi hasil prediksi

1. Mengubah nama kolom agar lebih sederhana.
2. Mengecek *missing value*.

3. Mengonversi format tanggal menjadi *datetime*.
4. Mengurutkan data berdasarkan waktu.
5. Mengubah format data numerik.

b. *Exploratory Data Analysis (EDA)*

Dilakukan visualisasi data produksi kopi untuk memahami pola dan tren data.

c. *Feature engineering*

Feature engineering menghasilkan fitur turunan (*derived features*) yang tidak tersedia secara langsung pada dataset asli. Fitur-fitur tersebut dibentuk agar model LSTM mampu mempelajari hubungan temporal dan pola musiman produksi kopi. Pada penelitian ini fitur turunan terdiri atas :

1. Membentuk fitur produksi_lag1, produksi_lag2, dan produksi_lag3.
2. Membentuk fitur siklik bulan_sin dan bulan_cos untuk merepresentasikan pola musiman.

d. *Variabel Penelitian*

Variabel yang digunakan terdiri atas produksi historis, suhu, curah hujan, dan kelembaban. Selain itu, proses *feature engineering* menghasilkan fitur turunan berupa produksi_lag1, produksi_lag2, produksi_lag3, bulan_sin, dan bulan_cos.

e. *Pembagian Data (Data splitting)*

Untuk mencegah data leakage, dataset dibagi menggunakan metode *time-series split* dengan proporsi 80% *training*, 10% *validation*, dan 10% *testing* berdasarkan urutan waktu, sehingga data periode awal digunakan untuk pelatihan, periode berikutnya untuk validasi, dan periode terakhir untuk pengujian.

f. *Normalisasi data*

Melakukan normalisasi menggunakan *Min-Max Scaling* ke rentang 0–1. Parameter *MinMaxScaler* dihitung hanya dari data *training* yang akan digunakan pada data *validation* serta *testing* untuk mencegah data leakage.

g. *Sliding window*

Membentuk data *time-series (sliding window)* dengan menyusun data menjadi pasangan input-output berbasis urutan (*sequence*). Pada penelitian ini digunakan window sepanjang 12 bulan, sehingga model memanfaatkan data produksi dan variabel iklim selama 12 bulan sebelumnya untuk memprediksi produksi kopi tiga

bulan ke depan (3-step ahead forecasting). Sliding window sepanjang 12 bulan dipilih karena produksi kopi memiliki pola musiman tahunan, sehingga model memerlukan informasi satu siklus musim penuh untuk mempelajari hubungan antara kondisi iklim dan produksi kopi.

3.4.3 Pemodelan LSTM

Berdasarkan Tabel 3.x, model Long Short-Term Memory (LSTM) pada penelitian ini terdiri atas dua layer LSTM. Layer pertama menggunakan 64 unit dengan parameter `return_sequences=True` agar seluruh urutan keluaran diteruskan ke layer LSTM berikutnya. Layer kedua terdiri atas 32 unit yang berfungsi mengekstraksi pola temporal tingkat lanjut dari data time series. Setelah masing-masing layer LSTM diterapkan dropout sebesar 0,2 untuk mengurangi risiko overfitting. Selanjutnya, representasi fitur diteruskan ke Dense layer dengan 16 neuron dan fungsi aktivasi ReLU sebelum menghasilkan satu neuron pada output layer yang digunakan untuk memprediksi produksi kopi tiga bulan ke depan (3-step ahead forecasting).

Tabel 3. 2Arsitektur LSTM

LAYER	KONFIGURASI
Input	Sequence (12, 9)
LSTM 1	64 unit, <code>return_sequences=True</code>
Dropout	0,2
LSTM 2	32 unit
Dropout	0,2
Dense	16 neuron, ReLU

3.4.4 Training model

Proses pelatihan menggunakan *batch size* sebesar 8 dan maksimum 100 *epoch*. Selain itu diterapkan *callback EarlyStopping* untuk menghentikan pelatihan ketika *validation loss* tidak mengalami perbaikan serta *ReduceLROnPlateau* untuk menurunkan *learning rate* secara otomatis apabila proses pelatihan mulai stagnan.

3.4.5 Prediksi

Setelah pelatihan selesai, model LSTM digunakan untuk memprediksi data *training*, *validation*, dan *testing* menggunakan model terbaik. Hasil prediksi yang masih dalam bentuk data ternormalisasi kemudian dikembalikan ke skala asli melalui proses *inverse transform* agar dapat diinterpretasikan.

3.4.6 Invers Transform

Tahap *inverse transform* dilakukan menggunakan *target_scaler* untuk mengembalikan hasil prediksi dan data aktual yang telah dinormalisasi ke skala aslinya sehingga dapat digunakan pada evaluasi model.

3.4.7 Evaluasi Model

Evaluasi model dilakukan menggunakan metrik MAE, MSE, RMSE, dan R^2 dengan membandingkan hasil prediksi terhadap data aktual.

3.4.8 Visualisasi Hasil

Tahap visualisasi dilakukan untuk mengevaluasi performa model menggunakan grafik *training loss* dan *validation loss*, grafik perbandingan data aktual dan prediksi, *scatter plot*, serta *residual plot*.

3.4.9 Interpretasi SHAP

Pada tahap ini SHAP menghitung kontribusi setiap fitur berdasarkan konsep *Shapley Value* dengan mengukur perubahan hasil prediksi pada berbagai kombinasi fitur. Semakin besar nilai absolut SHAP, semakin besar kontribusi fitur terhadap hasil prediksi. Pada penelitian ini, nilai SHAP dihitung menggunakan *GradientExplainer* untuk mengestimasi kontribusi setiap fitur pada model LSTM.

Interpretasi dilakukan melalui dua skenario, yaitu dengan dan tanpa variabel produksi historis. Skenario pertama mengevaluasi kontribusi seluruh fitur, sedangkan skenario kedua berfokus pada pengaruh variabel iklim (suhu, curah hujan, dan kelembaban). Selanjutnya, dihitung *mean absolute SHAP value* untuk menentukan tingkat kepentingan setiap fitur. Hasil interpretasi disajikan menggunakan *SHAP Summary Plot*, *SHAP Bar Plot*, dan tabel *mean absolute SHAP value*.

Tabel 3. 3Arsitektur SHAP

Explainer	GradientExplainer
Background Data	Sebagian data training sebagai baseline
Data yang Dianalisis	Data testing
Input Shape	Sequence (12, 9)
Output SHAP	Nilai kontribusi setiap fitur terhadap prediksi
Post-processing	Mengubah output SHAP multidimensi menjadi bentuk 2D untuk visualisasi
Visualisasi	SHAP Summary Plot, SHAP Bar Plot
Analisis Kepentingan Fitur	Mean Absolute SHAP Value

3.4.10 Analisis Hasil Prediksi

Pada tahap akhir penelitian ini dilakukan proses penarikan kesimpulan dan analisis terhadap hasil prediksi produksi komoditas kopi di Indonesia akibat variabilitas iklim menggunakan LSTM dan SHAP.