

BAB 2. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian terdahulu digunakan sebagai landasan dalam penelitian ini. Penelitian (Durai Arumugam et al., 2024) mengombinasikan LSTM dengan *Integrated Gradients* dan SHAP untuk prediksi berbasis *time-series*, mencapai akurasi uji 0,9748 ($\pm 97\%$) serta mampu menginterpretasikan kontribusi variabel, namun masih menggunakan objek hasil produksi pertanian secara umum (*multi-komoditas*). Selanjutnya (Martello et al., 2022) mengestimasi hasil kopi Arabika berbasis citra satelit *time-series* dengan *Random Forest* dan regresi linear ($R^2 = 0,93$), efektif secara spasial tetapi belum memodelkan dependensi temporal jangka panjang. Kemudian (Yan et al., 2025) menunjukkan akurasi sangat tinggi ($R^2 \approx 0,986$) melalui *ensemble tree-based* pada data *multi-negara* dan *multi-komoditas*, namun tidak dirancang untuk dependensi temporal jangka panjang dan tidak diimplementasikan terhadap komoditas khusus. Serta (Rajitha et al., n.d.) membuktikan LSTM efektif menangkap dinamika temporal berbasis data iklim *IoT* dengan kesalahan rendah, tetapi hasil prediksi yang dihasilkan masih bersifat *black-box* sehingga belum memberikan penjelasan yang jelas terkait kontribusi masing-masing variabel. Sementara itu, (Patil et al., 2025) melalui meta-analisis 68 studi menegaskan bahwa variabilitas iklim meningkatkan risiko kehilangan hasil kopi dan agroforestri dapat menurunkannya, namun bersifat konseptual tanpa model prediktif.

Tabel 2. 1 Penelitian Terdahulu

No	Referensi	Dataset	Metode	Hasil
1	et al., 2024	Data pertanian India (1997–2023). tanah, cuaca, iklim, jenis tanaman, historis)	LSTM + <i>Integrated Gradients</i> (IG) + SHAP; iklim, dibandingkan dengan LR, RF, CatBoost	terbaik akurasi uji 0,9748 dan akurasi keseluruhan $\pm 97\%$; model lebih akurat dan <i>interpretable</i>

			RF terbaik dengan
		Data hasil panen kopi	R^2 validasi = 0,93;
		Arabika (2019–2021,	citra fase dormansi
		Brasil) + citra satelit	<i>Random Forest</i> (Juli–Agustus) paling
		PlanetScope resolusi 3 m (RF) & <i>Multiple</i>	optimal untuk
	Martello	<i>et al.</i> (Red, Green, Blue, NIR, Linear Regression	prediksi hasil satu
2	<i>al.</i> , 2022	NDVI, GNDVI) (MLR)	tahun sebelum panen
		Data <i>time-series multi-crop</i> , <i>multi-negara</i>	<i>Random Forest &</i>
		(1990–2013). (Suhu,	<i>Bagging</i> terbaik
		kelembapan, <i>Hybrid ML</i> : RF, dengan akurasi 0,986	
		kelembapan tanah, <i>Bagging</i> , (R^2 = 0,986); <i>error</i>	
	Yan	<i>et al.</i> , curah hujan, hasil <i>XGBoost</i> , GB, DT, rendah (MAE ±345–	
3	2025	historis) KNN, LR	349)
		<i>Systematic</i> Agroforestri	
		68 studi global (2000– <i>Review</i> (PRISMA menurunkan risiko	
		2024) tentang kopi & 2020) + <i>Meta-</i> kehilangan hasil 31%,	
		perubahan iklim. (Suhu, <i>analysis (random-</i> hama/penyakit 24%,	
	Patil	<i>et al.</i> , curah hujan, naungan, <i>effects</i> , <i>REML</i>) + dan penurunan	
4	2025	hama, penyakit) <i>Meta-regression</i>	kualitas 19%.
		(A) Data <i>IoT</i> farm-	
		level (2014–2018, sensor	
		harian: suhu,	
		kelembaban, soil	<i>IoT</i> dataset: MAPE
		moisture, hujan) (B)	±7% (lebih baik dari
		Data global FAO–World LSTM (2-layer) <i>baseline</i> ±9–12%);	
		Bank (1990–2013, <i>multi-</i> dibandingkan Global dataset: R^2	
		<i>negara & multi-crop</i>). <i>Linear Regression</i> 0,90–0,96 (lebih	
	Rajitha	<i>et al.</i> , (Curah hujan, suhu, & <i>baseline time-</i> tinggi dari regresi	
5	2025	pestisida) <i>series</i>	linear)

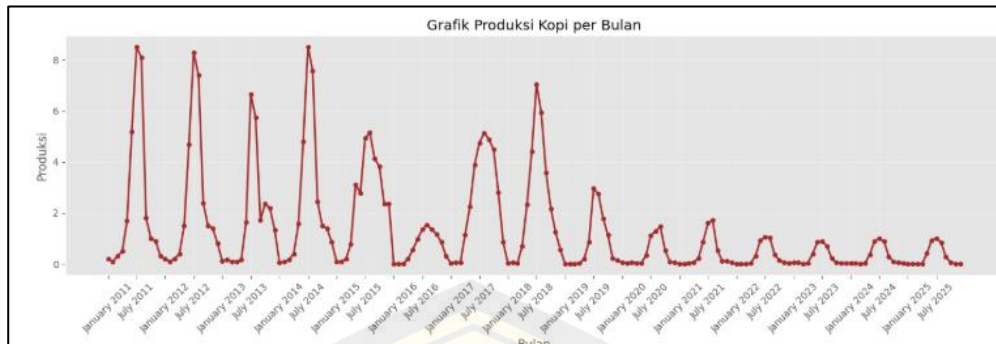
Berdasarkan uraian tersebut, penelitian terdahulu menunjukkan bahwa LSTM efektif untuk prediksi data *time-series*. Namun, sebagian besar penelitian masih dilakukan *multi* wilayah, sehingga belum berfokus pada satu wilayah, khususnya di

Indonesia. Selain itu, model yang digunakan masih bersifat *black box*, sementara sebagian besar penelitian juga menggunakan data tahunan atau estimasi hasil secara umum, sehingga belum menganalisis pengaruh suhu, curah hujan, dan kelembaban bulanan terhadap perubahan produksi kopi secara temporal. Oleh karena itu, penelitian ini mengintegrasikan LSTM dan SHAP untuk menghasilkan prediksi yang akurat sekaligus menjelaskan kontribusi setiap variabel iklim terhadap produksi kopi pada wilayah penelitian.

2.2 Fluktuasi Hasil Produksi Kopi

Produksi komoditas pertanian terutama komoditas kopi memiliki hasil produksi yang terbilang fluktuatif dan tidak stabil (Bacsi et al., 2022). Kondisi ini menunjukkan adanya pola kenaikan dan penurunan hasil produksi yang terjadi secara berkala dalam sistem pertanian (Swamy et al., 2025). Fluktuasi hasil produksi tersebut merupakan bagian dari dinamika pertanian yang dipengaruhi oleh perubahan kondisi produksi antar periode (Jiabul Hoque et al., 2024). Akibatnya, hasil produksi tidak selalu konsisten, melainkan mengikuti tren tertentu yang dapat berubah dari waktu ke waktu (Swamy et al., 2025).

Perubahan produksi yang tidak stabil ini berdampak langsung pada sektor pertanian, terutama dalam hal ketersediaan hasil dan kestabilan sistem produksi (Bacsi et al., 2022). Kondisi tersebut juga menimbulkan ketidakpastian dalam perencanaan produksi dan pengambilan keputusan di bidang pertanian (Jiabul Hoque et al., 2024). Dalam praktiknya, fluktuasi hasil produksi kopi dapat memengaruhi kondisi ekonomi pelaku pertanian, dimana perubahan hasil produksi dapat berdampak pada pendapatan dan keberlanjutan usaha pertanian (Bacsi et al., 2022). Oleh karena itu, ketidakstabilan produksi menjadi tantangan bagi petani dalam mengelola hasil produksi dan menentukan strategi produksi yang tepat (Jiabul Hoque et al., 2024). Berikut merupakan grafik naik turun produksi kopi di Indonesia dari tahun 2016 – 2025 dapat dilihat pada gambar 2.1.



Gambar 2. 1 Grafik Hasil Produksi Kopi Tahun 2016 - 2025

2.3 Hubungan Iklim dengan Produksi Kopi

Produksi kopi dipengaruhi oleh kondisi iklim. Oleh karena itu, analisis faktor iklim diperlukan untuk mengetahui pengaruhnya terhadap tingkat produksi kopi.

2.3.1 Pengaruh Suhu Terhadap Kopi

Suhu merupakan faktor iklim yang memengaruhi pertumbuhan dan produksi tanaman (Jiabal Hoque et al., 2024). Tanaman kopi memerlukan suhu yang sesuai agar menghasilkan produksi optimal (Ed-Daoudi et al., 2025). Kenaikan suhu dapat menurunkan produktivitas tanaman karena mengganggu pertumbuhan (Swamy et al., 2025). Suhu yang tinggi juga meningkatkan kebutuhan air dan menyebabkan stres tanaman sehingga berpotensi menurunkan produksi kopi (Jiabal Hoque et al., 2024). Maka suhu menjadi salah satu variabel untuk memprediksi hasil pertanian karena berpengaruh signifikan terhadap produktivitas (Kulyal & Saxena, 2022).

2.3.2 Pengaruh Curah Hujan Terhadap Kopi

Kopi merupakan salah satu tanaman yang sangat berpengaruh terhadap curah hujan (Sharma et al., 2023). Tanaman kopi membutuhkan distribusi curah hujan yang stabil agar dapat tumbuh dan berproduksi secara optimal (Ed-Daoudi et al., 2025). Perubahan pola curah hujan dapat menyebabkan variasi hasil produksi kopi dari waktu ke waktu (Swamy et al., 2025). Curah hujan yang tidak menentu dapat mengganggu proses pertumbuhan tanaman sehingga berdampak pada penurunan hasil produksi. Selain itu, kekurangan curah hujan dapat menyebabkan kekeringan yang berdampak pada menurunnya produktivitas tanaman kopi. Sebaliknya, curah hujan yang berlebihan juga dapat menyebabkan kerusakan tanaman dan menurunkan hasil produksi (Jiabal Hoque et al., 2024).

2.3.3 Pengaruh Kelembaban Terhadap Kopi

Dalam sistem pertanian, kondisi kelembaban berperan dalam menentukan lingkungan tumbuh tanaman yang optimal (Ed-Daoudi et al., 2025). Kelembaban berperan dalam menjaga keseimbangan air pada tanaman yang berpengaruh terhadap proses fisiologis seperti transpirasi dan penyerapan air (P. Zhang et al., 2022). Tingkat kelembaban yang rendah dapat meningkatkan laju penguapan air dari tanaman sehingga berpotensi menyebabkan stres air dan mengganggu pertumbuhan tanaman kopi (Jiabul Hoque et al., 2024). Sebaliknya, kelembaban yang terlalu tinggi dapat menciptakan kondisi lingkungan yang kurang optimal bagi tanaman dan memengaruhi hasil produksi (Ed-Daoudi et al., 2025).

2.4 Prediksi

Prediksi adalah proses memperkirakan suatu peristiwa di masa depan berdasarkan data dan informasi yang tersedia (KBBI, n.d.). Dalam ilmu data, prediksi digunakan untuk mengidentifikasi pola pada data guna memperkirakan kejadian yang akan datang (Kulyal & Saxena, 2022). Prediksi umumnya dilakukan dengan memanfaatkan metode statistik maupun pembelajaran mesin yang mampu mempelajari hubungan antar variabel dalam data (Nurchahyo et al., 2023a). Metode ini digunakan untuk menghasilkan estimasi yang lebih akurat dengan mempertimbangkan berbagai faktor yang memengaruhi data (Jiabul Hoque et al., 2024). Pada data deret waktu, metode prediksi mampu mengenali tren, pola musiman, dan fluktuasi dari waktu ke waktu (Rahman et al., 2023). Oleh karena itu, prediksi berperan penting dalam mengurangi ketidakpastian dan mendukung perencanaan yang lebih efektif (Nurchahyo et al., 2023).

2.5 Time-series

Time-series adalah kumpulan data yang dicatat atau diukur secara berurutan berdasarkan waktu tertentu, seperti harian, mingguan, bulanan, atau tahunan (DQLab, 2025). Dalam analisis *time-series*, data historis digunakan untuk mengidentifikasi pola seperti tren, musiman, dan fluktuasi yang terjadi dalam suatu periode. *Time-series* banyak digunakan dalam berbagai bidang, termasuk pertanian, karena mampu merepresentasikan perubahan data produksi secara dinamis (Rahman et al., 2023).

2.6 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) merupakan metode *deep learning* berbasis *Recurrent Neural Network* (RNN) yang dirancang untuk memproses data deret waktu (*time-series*) (Khairunisa & Hendikawati, 2024). LSTM bekerja melalui empat komponen utama, yaitu *forget gate*, *input gate*, *cell state*, dan *output gate*, yang berfungsi mengatur aliran informasi selama proses pembelajaran (Smagulova & James, 2019).

1. *Forget gate* berfungsi menentukan informasi pada *cell state* sebelumnya yang dipertahankan atau dihapus. Nilai keluaran fungsi *sigmoid* berkisar antara 0 hingga 1, di mana nilai mendekati 0 menghapus informasi, sedangkan nilai mendekati 1 mempertahankannya dalam memori.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

2. *Input gate* berfungsi menentukan informasi baru yang ditambahkan ke dalam *cell state*. Nilai i_t mengontrol banyaknya informasi yang disimpan, sedangkan $\tilde{C}_t$ merupakan kandidat nilai baru yang dihasilkan dari input saat ini dan *hidden state* sebelumnya menggunakan fungsi *tanh*.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

Input gate menentukan informasi yang diteruskan sebagai *hidden state*. Nilai o_t mengontrol bagian *cell state* yang diteruskan ke output sehingga menghasilkan *hidden state* h_t yang merepresentasikan kondisi data saat ini.

3. *Cell state* diperbarui dengan menggabungkan informasi lama yang dipertahankan ($f_t \cdot C_{t-1}$) dan informasi baru ($i_t \cdot \tilde{C}_t$). Mekanisme ini memungkinkan LSTM mempertahankan informasi penting sekaligus memperbarui memori berdasarkan data baru.

$$C = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4)$$

4. *Output gate* berfungsi mengontrol bagian *cell state* yang diteruskan sebagai *hidden state*. Nilai o_t menentukan informasi yang dikeluarkan ke output sehingga menghasilkan *hidden state* h_t yang merepresentasikan kondisi data saat ini.

Rumus

output

gate:

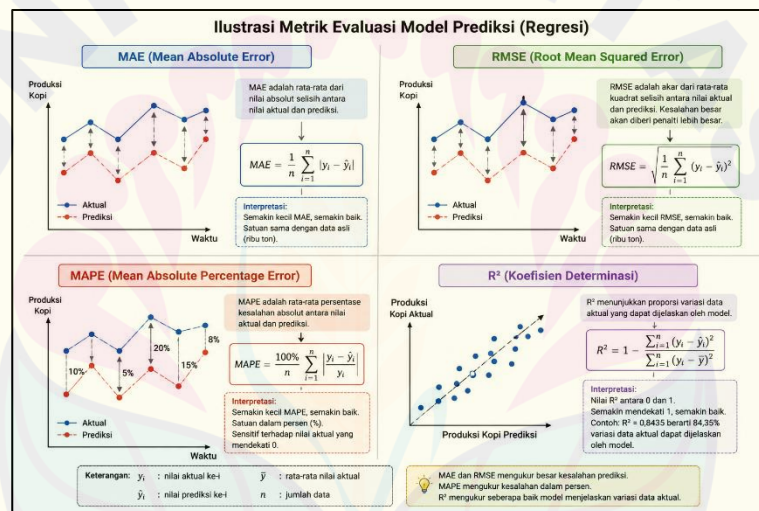
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

Rumus *hidden state*:

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

2.7 Evaluasi

Evaluasi model merupakan tahap penting dalam program untuk mengukur kinerja model dalam melakukan prediksi berdasarkan data uji (Jiabil Hoque et al., 2024). Proses evaluasi dilakukan untuk memastikan bahwa model tidak hanya memiliki akurasi tinggi pada data pelatihan, tetapi juga mampu melakukan generalisasi dengan baik pada data baru (Swamy et al., 2025). Beberapa metrik evaluasi yang umum digunakan dalam prediksi hasil produksi antara lain :



Gambar 2. Ilustrasi Metrik Evaluasi

1. Mean Absolute Error (MAE)

MAE merupakan metrik yang mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual. Rumus MAE dapat dituliskan sebagai:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

MAE memberikan interpretasi yang mudah karena hasilnya berada dalam satuan yang sama dengan data asli (Swamy et al., 2025).

2. Mean Squared Error (MSE)

MSE mengukur rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual . Rumus MSE adalah:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

MSE memberikan penalti lebih besar terhadap kesalahan yang besar sehingga sensitif terhadap *outlier* (T. Zhang et al., 2026).

3. Root Mean Squared Error (RMSE)

RMSE merupakan akar kuadrat dari MSE yang digunakan untuk mengembalikan satuan *error* ke skala asli data (T. Zhang et al., 2026). Rumus RMSE adalah:

$$RMSE = \sqrt{MSE} \quad (9)$$

RMSE sering digunakan karena memberikan keseimbangan antara interpretasi yang mudah dan sensitivitas terhadap kesalahan besar (Martello et al., 2022b).

4. Koefisien Determinasi (R^2)

R^2 digunakan untuk mengukur seberapa besar variasi data aktual yang dapat dijelaskan oleh model (Martello et al., 2022b). Rumus R^2 adalah:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (10)$$

Nilai R^2 mendekati 1 menunjukkan bahwa model memiliki kemampuan prediksi yang sangat baik (Jiabul Hoque et al., 2024).

2.8 SHAP (SHapley Additive exPlanations)

SHAP (SHapley Additive exPlanations) merupakan metode *Explainable Artificial Intelligence (XAI)* yang digunakan untuk menjelaskan kontribusi setiap fitur terhadap hasil prediksi model *machine learning* yang bersifat *black-box* (Nurchahyo et al., 2023b). Metode ini berasal dari konsep nilai *Shapley* dalam teori permainan kooperatif yang digunakan untuk menghitung kontribusi masing-masing variabel terhadap *output model* (Ed-Daoudi et al., 2025). SHAP mampu memberikan interpretasi secara global (keseluruhan model) dan lokal (per data individu), sehingga hubungan antar variabel dapat dipahami dengan lebih jelas

(Nurcahyo et al., 2023). Secara matematis, nilai SHAP untuk suatu fitur i dirumuskan sebagai:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (11)$$

dimana:

- ϕ_i = nilai SHAP untuk fitur ke- i
- S = subset fitur yang tidak mengandung fitur i
- N = himpunan seluruh fitur
- $f(S)$ = output model dengan subset fitur S

Rumus tersebut menunjukkan bahwa kontribusi setiap fitur dihitung berdasarkan rata-rata *marginal contribution* terhadap seluruh kombinasi fitur lainnya (Ed-Daoudi et al., 2025). SHAP juga mengikuti model aditif yang dirumuskan sebagai:

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i \quad (12)$$

dimana:

- $f(x)$ = hasil prediksi model
- ϕ_0 = nilai dasar (*baseline*)
- ϕ_i = kontribusi masing-masing fitur

Model ini memungkinkan interpretasi global dan lokal terhadap prediksi model (Nurcahyo et al., 2023). Berdasarkan persamaan tersebut, nilai SHAP dihitung dengan membandingkan perubahan hasil prediksi ketika suatu fitur dimasukkan ke dalam berbagai kombinasi fitur lainnya. Dengan cara ini, setiap fitur memperoleh nilai kontribusi rata-rata yang adil terhadap hasil prediksi model. Fitur dengan nilai *mean absolute SHAP* yang lebih besar diinterpretasikan memiliki pengaruh yang lebih besar terhadap prediksi model.