

BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian-penelitian sebelumnya telah memberikan kontribusi terkait prediksi menggunakan metode *hybrid*. Oleh karena itu, tabel berikut menyajikan perbandingan penelitian terdahulu yang mencakup jenis data, metode yang digunakan, kinerja model, serta keterbatasan masing-masing penelitian sebagai dasar pengembangan penelitian ini.



DIGITAL REPOSITORY UNIVERSITAS JEMBER

Tabel 2.1 Analisis Car Perhitungan

No	Peneliti	Dataset	Metode	Hasil	Gap
1	(Kasemset dkk., 2023)	Harga bulanan bawang merah di Thailand Utara (Jan 2014 – Des 2020) dengan variabel suhu dan curah hujan.	Perbandingan model ARIMA, <i>Holt-Winters</i> , LSTM, dan <i>Hybrid</i> ARIMA-LSTM.	<i>Hybrid</i> ARIMA-LSTM merupakan model terbaik dengan nilai RMSE 10,275 Baht dan MAE 8,512 Baht.	Data yang digunakan berbeda dan belum menggunakan metode STL Decomposition.
2	(Ray dkk., 2022)	Indeks harga bulanan kacang-kacangan (Gram, Moong, Urad) di India periode April 1994 – Maret 2021 (324 data).	<i>Hybrid</i> ARIMA-LSTM dengan seleksi lag menggunakan Random Forest, dibandingkan terhadap ARIMA, GARCH, dan LSTM standar.	Model hibrida yang diusulkan unggul dengan peningkatan akurasi 33–71% untuk RMSE dan 33–77% untuk MAPE dibandingkan model statistik tradisional.	Fokus utama pada data univariat; belum mengintegrasikan variabel eksternal.
3	(D. Xu dkk., 2021)	Indeks SPEI dari 613 stasiun di Tiongkok (1980–2019) yang dihitung dari data curah hujan dan suhu bulanan.	ARIMA, SVR, LSTM, ARIMA-SVR, LS-SVR, dan <i>Hybrid</i> ARIMA-LSTM.	Model ARIMA-LSTM memiliki akurasi tertinggi pada skala 6, 12, dan 24 bulan, menjadikannya model yang paling sesuai untuk prediksi kekeringan jangka panjang di Tiongkok	Data yang digunakan berbeda dan belum menerapkan metode STL Decomposition.
4	(Dias dkk., 2025)	Data harian GPP (Gross Primary Productivity) yang dihitung dari data satelit MERRA-2 dan MODIS (MOD15A2H)	Model Hibrida STL-ARIMA-ES-LSTM: Menggabungkan dekomposisi STL dengan ARIMA (untuk tren), Exponential Smoothing (untuk musiman), dan LSTM (untuk residu).	Model hibrida jauh lebih akurat dibanding LSTM mandiri dengan nilai RMSE 1,69 gC/m ² /hari, MAE 1,35, dan R ² 0,39	Data yang digunakan berbeda. Dan tidak menggunakan data faktor eksternal.

Berdasarkan Tabel 2.1, penelitian terdahulu menunjukkan bahwa model *hybrid* mampu meningkatkan akurasi peramalan data deret waktu dibandingkan model tunggal karena dapat mengombinasikan keunggulan metode statistik dan *deep learning*. Beberapa penelitian juga telah mengembangkan pendekatan *Hybrid* STL-ARIMA-ES-LSTM dengan memanfaatkan dekomposisi STL untuk memisahkan komponen tren, musiman, dan residual sebelum proses pemodelan. Namun, masih terdapat beberapa kesenjangan, yaitu sebagian besar penelitian menggunakan objek penelitian yang berbeda dengan harga bawang merah di Provinsi Jawa Timur, beberapa penelitian belum menerapkan dekomposisi STL, serta penelitian yang telah menggunakan *Hybrid* STL-ARIMA-ES-LSTM belum mengintegrasikan variabel eksternal, seperti suhu dan kelembapan. Oleh karena itu, penelitian ini mengusulkan model *Hybrid* ARIMA-ES-LSTM berbasis STL dengan mengintegrasikan variabel suhu dan kelembapan pada model LSTM untuk menghasilkan prediksi harga bawang merah di Provinsi Jawa Timur yang lebih akurat.

2.2 Harga Bawang Merah

Bawang merah merupakan salah satu komoditas hortikultura strategis yang memiliki karakteristik harga sangat fluktuatif dan bersifat musiman (Haryati dkk., 2024). Di Indonesia, volatilitas harga bawang merah tergolong tinggi dan bersifat persisten dalam jangka panjang, di mana wilayah Jawa Timur mencatatkan tingkat volatilitas tertinggi dibandingkan wilayah sentra lainnya seperti DKI Jakarta, Jawa Tengah, dan Jawa Barat (Putri dkk., 2023). Perbedaan harga antar wilayah dan waktu ini dipicu oleh adanya kesenjangan (gap) antara penawaran dan permintaan, terutama saat memasuki masa bukan panen di mana ketersediaan bawang merah berkurang karena sifatnya yang tidak tahan lama untuk disimpan dalam jangka panjang (Wibowo & Surbakti, 2023).

Harga bawang merah pada periode sebelumnya terbukti menjadi prediktor penting bagi harga di periode berikutnya, dan faktor lain seperti harga komoditas komplementer seperti bawang putih serta respons kebijakan impor ketika produksi domestik turun turut memperkuat volatilitas harga bawang merah di pasar domestik

(Maggellang dkk., 2021). Faktor alam seperti cuaca ekstrem, curah hujan yang tidak menentu, serta gangguan hama dan penyakit menjadi determinan utama yang dapat memicu gagal panen dan kelangkaan pasokan (Putri dkk., 2023). Fluktuasi harga bawang merah yang tinggi dan sulit diprediksi memberikan dampak luas terhadap kesejahteraan petani serta stabilitas pangan konsumen (Wibowo & Surbakti, 2023). Dampak nyata pada komoditas bawang merah terlihat dari terganggunya keputusan produksi di sisi produsen serta penurunan konsumsi di sisi masyarakat (Putri dkk., 2023). Oleh karena itu, diperlukan analisis dan prediksi yang akurat sebagai dasar perumusan kebijakan stabilisasi pasar dan minimalisasi risiko ekonomi di masa depan.

Fluktuasi harga bawang merah di Jawa Timur selama periode tahun 2013 - 2025 dapat dilihat pada Gambar 2.1.



Gambar 2.1 Grafik Harga Bawang Merah Tahun 2013 - 2025 (data yang sudah diolah dari SISKAPERBAPO Jawa Timur)

2.3 Data Deret Waktu

Data Deret Waktu adalah data yang disusun dalam urutan waktu agar pola historisnya dapat dianalisis, khususnya saat digunakan dalam pemodelan dan prediksi. Tiap titik data mencerminkan nilai pengamatan variabel pada titik waktu tertentu, dan data ini dapat dipakai untuk mengidentifikasi komponen seperti tren, musiman dan fluktuasi acak yang berguna dalam peramalan (Fatima & Rahimi,

2024). Analisis *time series* memiliki beberapa jenis pola data yang umum, antara lain (Shumway & Stoo, 2017):

- a. Tren: Merupakan arah pergerakan dasar atau perkembangan mendasar dari data dalam jangka panjang. Tren dapat berupa linier, eksponensial, atau parabola.
- b. Musiman: Menangkap variasi atau pola yang berulang secara teratur dalam periode waktu tertentu, seperti siklus tahunan yang dipengaruhi oleh faktor alam (cuaca) maupun aktivitas manusia.
- c. Residual: Merupakan fluktuasi acak atau sisa ketidakteraturan yang tetap ada dalam data setelah pengaruh tren dan musiman diperhitungkan.

2.4 Seasonal-Trend decomposition using Loess (STL)

Seasonal-Trend decomposition using Loess (STL) merupakan sebuah teknik statistik yang digunakan untuk memisahkan data deret waktu menjadi tiga komponen utama, yaitu komponen musiman (*seasonal*), tren (*trend*), dan sisa (*residual*) (Gordan dkk., 2024). Persamaan STL adalah sebagai berikut:

$$Y_t = T_t + S_t + R_t$$

Metode STL memiliki keunggulan signifikan dalam hal fleksibilitas, karena memungkinkan komponen musiman untuk berubah seiring berjalannya waktu dan tidak mengasumsikan pola musiman yang konstan atau monoton sepanjang periode data (Gordan dkk., 2024). Selain itu, STL bersifat sangat kokoh terhadap pencilan (*robust to outliers*) dibandingkan dengan metode polinomial lokal tradisional maupun metode dekomposisi klasik (Dias dkk., 2025).

2.5 Autoregressive Integrated Moving Average (ARIMA)

Model ARIMA merupakan metodologi yang digunakan untuk membangun model deret waktu dari observasi univariat yang terdiri atas tiga komponen utama, yaitu istilah *autoregressive* (AR), *integrated* (I), dan *moving average* (MA) (Ray dkk., 2022). Secara teknis, model ini dinyatakan melalui notasi standar (p, d, q) yang mencerminkan struktur parameter di dalamnya. Parameter p merujuk pada

komponen *Autoregressive* (AR) yang mengekspresikan hubungan autokorelasi antara observasi masa lalu dengan observasi saat ini. Komponen *Integrated* (I) yang disimbolkan dengan parameter d menunjukkan tingkat pembedaan (differencing) yang dilakukan pada observasi untuk mengubah data non-stasioner menjadi stasioner. Sementara itu, parameter q mewakili komponen *Moving Average* (MA) yang menggambarkan struktur autokorelasi dari residu atau galat yang dihasilkan oleh model. Secara matematis, model ARIMA (p, d, q) dirumuskan sebagai berikut:

$$\Delta^d y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

Dalam aplikasinya, model ARIMA sangat efektif untuk menangkap pola temporal yang bersifat linier dalam data deret waktu (Mahmud dkk., 2025). Namun, sebagai model linier, ARIMA memiliki keterbatasan dalam meramalkan data yang mengandung volatilitas tinggi atau hubungan nonlinier yang kompleks (D. Xu dkk., 2021). Untuk menentukan nilai parameter p dan q yang optimal, peneliti melakukan pemeriksaan terhadap plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) (Kasemset dkk., 2023).

2.6 Exponential Smoothing

Exponential Smoothing (ES) merupakan metode peramalan deret waktu yang memberikan bobot lebih besar pada data terbaru dibandingkan data sebelumnya, sehingga lebih responsif terhadap perubahan pola data (W. Xu dkk., 2022).

Metode ES memiliki beberapa pengembangan, yaitu *Single Exponential Smoothing* untuk data tanpa tren dan musiman, *Double Exponential Smoothing* (*Holt*) untuk data dengan tren, serta *Triple Exponential Smoothing* (*Holt-Winters*) untuk data dengan tren dan musiman (Dias dkk., 2025). Metode *Holt-Winters* tersedia dalam dua bentuk, yaitu aditif dan multiplikatif, yang disesuaikan dengan karakteristik pola musiman pada data. Model aditif digunakan ketika pola musiman bersifat konstan, sedangkan model multiplikatif digunakan ketika pola musiman bersifat proporsional terhadap level data.

Secara matematis, formulasi *Holt-Winters Additive* dinyatakan sebagai berikut:

Persamaan Level:

$$L_t = \alpha(Y_t - S_{\{t-m\}}) + (1 - \alpha)(L_{\{t-1\}} + T_{\{t-1\}})$$

Persamaan Trend:

$$T_t = \beta(L_t - L_{\{t-1\}}) + (1 - \beta)T_{\{t-1\}}$$

Persamaan Musiman:

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{\{t-m\}}$$

Persamaan Peramalan:

$$\hat{Y}_{\{t+h\}} = L_t + hT_t + S_{\{t+h-m\}}$$

2.7 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) merupakan variasi *deep learning* dan *recurrent neural networks* (RNN) yang dirancang untuk mempertahankan informasi secara efektif dalam periode yang diperpanjang guna mengidentifikasi pola jangka panjang serta ketergantungan kompleks dalam tren deret waktu (*time series*) (Kasemset dkk., 2023). Modul memori LSTM terdiri dari unit penyimpanan dan tiga komponen komputasi yang disebut *input gate*, *forget gate*, dan *output gate* yang masing-masing berfungsi untuk mengontrol proses pembacaan, penulisan, dan pengaturan ulang data sel memori (D. Xu dkk., 2021). Dalam proses pemodelan menggunakan *Long Short-Term Memory (LSTM)*, pengelolaan informasi di dalam *cell state* dilakukan melalui beberapa gerbang (*gate*). Secara umum, mekanisme tersebut dapat dijelaskan sebagai berikut (Zhang dkk., 2024):

1. Forget Gate

Forget gate berfungsi untuk menentukan informasi apa saja dari *cell state* sebelumnya yang perlu dipertahankan atau dihilangkan. Gerbang ini menerima *hidden state* pada waktu sebelumnya (h_{t-1}) dan input saat ini

(x_t) , kemudian memprosesnya menggunakan fungsi aktivasi sigmoid (σ) yang menghasilkan nilai antara 0 dan 1. Nilai keluaran ini selanjutnya dikalikan dengan *cell state* sebelumnya (C_{t-1}) untuk menentukan informasi yang tetap disimpan. Persamaan forget gate ditunjukkan pada Persamaan (1).

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

2. Input Gate

Input gate berperan dalam menentukan bagian dari *cell state* yang akan diperbarui menggunakan input saat ini. Gerbang ini terdiri dari dua fungsi aktivasi, yaitu *sigmoid* dan *tanh*. Fungsi sigmoid menghasilkan vektor dengan rentang nilai antara 0 hingga 1 yang menunjukkan tingkat pembaruan informasi, sedangkan fungsi tanh menghasilkan vektor dengan nilai antara -1 hingga 1 yang merepresentasikan kandidat informasi baru yang akan disimpan. Perhitungan *input gate* dinyatakan pada Persamaan (2) dan Persamaan (3).

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

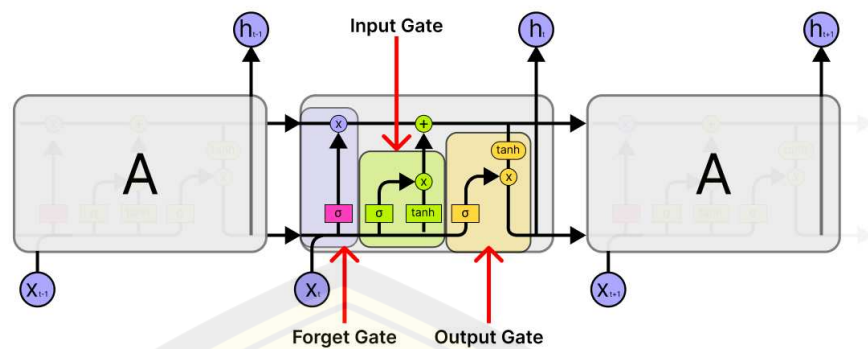
$$\tilde{C}_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \quad (3)$$

3. Output Gate

Output gate berfungsi untuk menentukan bagaimana *hidden state* pada waktu saat ini dihasilkan berdasarkan *cell state*. Gerbang ini memanfaatkan fungsi aktivasi *sigmoid* dan *tanh* untuk menghasilkan *hidden state* yang baru. Persamaan output gate dan perhitungan *hidden state* ditunjukkan pada Persamaan (4) dan Persamaan (5).

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t * \tanh(C_t) \quad (5)$$



Gambar 2.2 Arsitektur LSTM

2.8 Evaluasi Model

2.8.1 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) adalah metrik evaluasi yang diperoleh dengan menghitung akar dari nilai rata-rata kuadrat perbedaan antara hasil prediksi dan data aktual (Chicco dkk., 2021). Pengambilan akar kuadrat ini bertujuan untuk mengembalikan satuan ukuran kesalahan ke dalam satuan asli data yang dianalisis, sehingga hasil evaluasi model menjadi lebih mudah diinterpretasikan.

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

RMSE berguna untuk mengevaluasi kemampuan model dalam menangani fluktuasi harga pada data *time series*, karena metrik ini sensitif terhadap kesalahan prediksi yang besar. Dengan demikian, RMSE tidak hanya mencerminkan tingkat akurasi rata-rata model, tetapi juga menunjukkan sejauh mana model mampu mengendalikan kesalahan ketika terjadi lonjakan harga.

2.8.2 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah metrik evaluasi pada model regresi yang menghitung rata-rata dari selisih absolut antara nilai prediksi dan nilai aktual. Nilai MAE sebesar 0 menandakan hasil prediksi yang sepenuhnya akurat,

sedangkan nilai yang semakin besar menunjukkan performa model yang semakin kurang baik (Chicco dkk., 2021). Rumus perhitungan MAE sebagai berikut:

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

MAE digunakan untuk menghitung rata-rata selisih absolut antara harga bawang merah aktual dengan harga yang diprediksi. Nilai MAE yang kecil menunjukkan bahwa variasi antara data aktual dan data hasil ramalan sangat tipis, yang berarti model memiliki akurasi yang tinggi.

2.8.3 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) merupakan ukuran penilaian kinerja model yang menyatakan tingkat kesalahan prediksi dalam satuan persentase. Dengan demikian, metrik ini lebih berfokus pada besarnya kesalahan secara relatif dibandingkan pada nilai absolut dalam permasalahan regresi (Chicco dkk., 2021). Oleh karena itu, MAPE sering digunakan ketika perubahan relatif dianggap lebih berpengaruh terhadap kinerja model.

Secara matematis, MAPE dihitung sebagai rata-rata dari nilai absolut selisih antara nilai aktual dan nilai prediksi yang dinormalisasi terhadap nilai aktual, kemudian dinyatakan dalam bentuk persentase. Rumus MAPE ditunjukkan pada persamaan berikut:

$$MAPE = \left(\frac{100\%}{n}\right) \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Penggunaan MAPE memberikan gambaran tingkat kesalahan model dalam bentuk persentase, sehingga menyajikan ukuran kesalahan relatif yang lebih intuitif untuk memahami besarnya penyimpangan antara hasil prediksi dan harga aktual bawang merah.