

Comparison of LSTM, GRU, Bi-LSTM, and XGBoost for BRIS Stock Price Prediction with Lookback Period Variations

Subekti Wahyu Aji ^{1*}, Wise Herowati ^{2*}

^{*} Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

^{*} Kelompok Penelitian Kuantum Komputer dan Informatika Material, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang
11202214627@mhs.dinus.ac.id ¹, wise@dsn.dinus.ac.id ²

Article Info

Article history:

Received 2026-06-21

Revised 2026-07-24

Accepted 2026-08-12

Keyword:

GRU,
LSTM,
Bi-LSTM,
XGBoost,
Stock Price Prediction,
Lookback Period,
BRIS.

ABSTRACT

Stock price forecasting remains a challenging task due to the nonlinear and non-stationary characteristics of financial time series, particularly for Islamic banking stocks such as Bank Syariah Indonesia (BRIS), which exhibit highly dynamic price movements. This study compares the performance of four prediction models Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional Long Short-Term Memory (Bi-LSTM), and XGBoost in forecasting BRIS closing stock prices using OHLCV data collected from August 2020 to December 2025. To investigate the influence of historical observation windows, three lookback periods (10, 20, and 30 trading days) were evaluated. Model performance was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The experimental results demonstrate that the GRU model with a 10-day lookback period achieved the best overall performance, yielding an RMSE of 78.0643 IDR, a MAPE of 2.13%, and an R^2 value of 0.8438. It outperformed Bi-LSTM, XGBoost, and LSTM, all of which also achieved their highest accuracy using the 10-day lookback period. Furthermore, the analysis reveals that the shorter lookback window consistently provides superior predictive performance across all evaluated models, whereas extending the lookback period to 20 or 30 days does not lead to consistent performance improvements. Residual analysis further indicates a relatively unbiased distribution of prediction errors. These findings suggest that the simplified gating mechanism of GRU enables better stability and generalization than more complex architectures for forecasting Islamic bank stock prices. This study contributes to the understanding of the relationship between model complexity and historical time window selection in deep learning-based financial forecasting and provides practical guidance for selecting appropriate prediction models for Indonesian Islamic banking stocks.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Pasar modal merupakan salah satu pilar penting dalam sistem keuangan nasional yang berperan sebagai sarana mobilisasi dana masyarakat menuju sektor produktif. Di Indonesia, perkembangan pasar modal tidak hanya didominasi oleh sektor keuangan konvensional, tetapi juga menunjukkan pertumbuhan signifikan pada segmen keuangan berbasis prinsip syariah. Hal ini sejalan dengan posisi

Indonesia sebagai negara dengan populasi Muslim terbesar di dunia yang secara alami menciptakan permintaan tinggi terhadap produk keuangan yang sesuai prinsip Islam. Salah satu tonggak penting perkembangan perbankan syariah Indonesia adalah terbentuknya Bank Syariah Indonesia (BSI) melalui merger tiga bank syariah milik BUMN, yaitu PT Bank BRI syariah Tbk, PT Bank Syariah Mandiri, dan PT Bank BNI Syariah, yang resmi beroperasi pada 1 Februari 2021 [1]. Pasca-merger, BSI langsung menjadi bank terbesar ketujuh di

Indonesia dari sisi aset dengan total aset Rp245,7 triliun [2], dan per tahun 2024 telah menjadi bank syariah dengan jumlah nasabah terbesar di dunia dengan lebih dari 20 juta nasabah [1]. Saham BSI yang diperdagangkan di Bursa Efek Indonesia dengan kode efek BRIS mencatat all-time high Rp3.180 per lembar pada September 2024 dengan kapitalisasi pasar Rp143,46 triliun, menempatkannya sebagai bank syariah dengan kapitalisasi pasar terbesar kesembilan secara global [2].

Pergerakan harga saham BRIS yang sangat dinamis mencerminkan kompleksitas dan volatilitas yang melekat pada instrumen saham perbankan syariah. Prediksi harga saham merupakan permasalahan yang secara inheren sulit dipecahkan karena data harga saham bersifat non-linear, non-stasioner, dan dipengaruhi oleh banyak faktor eksternal seperti kondisi makroekonomi, sentimen pasar, serta kebijakan moneter [3]. Pendekatan statistik konvensional seperti Autoregressive Integrated Moving Average (ARIMA) memiliki keterbatasan dalam menangkap pola non-linear karena mengasumsikan hubungan linear antar observasi [6]. Hal ini mendorong eksplorasi metode deep learning yang lebih fleksibel dalam memodelkan pola temporal kompleks pada data keuangan.

Beberapa arsitektur deep learning berbasis Recurrent Neural Network (RNN) telah banyak diterapkan untuk prediksi deret waktu, di antaranya Long Short-Term Memory (LSTM) [4], Gated Recurrent Unit (GRU) [8], dan Bidirectional LSTM (Bi-LSTM) [5], [6]. Selain itu, Extreme Gradient Boosting (XGBoost) juga telah terbukti kompetitif pada berbagai tugas prediksi berbasis data tabular dan deret waktu berkat efisiensi komputasi dan kemampuannya menangani hubungan non-linear otomatis. Berbagai penelitian terdahulu di Indonesia menunjukkan bahwa GRU mengungguli LSTM dalam prediksi harga saham. Penelitian pada saham BBRI menemukan GRU unggul 10,7% dalam R^2 dan 18,5% dalam MAPE dibandingkan LSTM [12]. Beberapa penelitian sebelumnya menunjukkan bahwa model GRU memiliki kinerja prediksi yang lebih baik dibandingkan LSTM, yang ditunjukkan oleh nilai RMSE sebesar 90,14 dan MAPE sebesar 1,86%, sedangkan LSTM memperoleh RMSE sebesar 111,00 dan MAPE sebesar 2,37% [7]. Untuk konteks saham syariah, Budiprasetyo et al. [8] penelitian pada saham syariah ASII dan ADRO menunjukkan potensi model deep learning dalam memprediksi harga saham berbasis prinsip syariah.

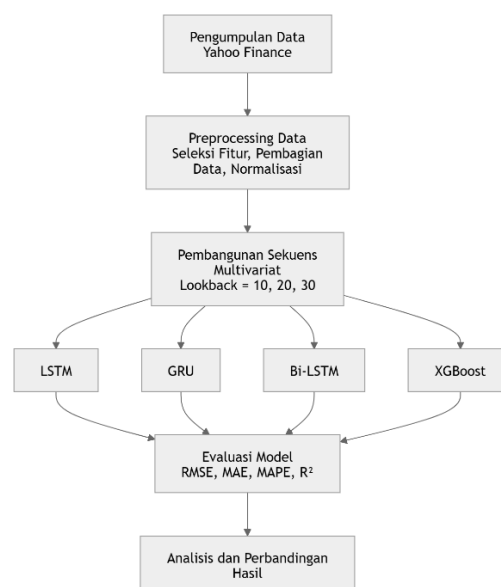
Meskipun demikian terdapat beberapa celah penelitian yang masih perlu diisi. Pertama, mayoritas penelitian hanya membandingkan dua model, umumnya LSTM vs GRU, tanpa menyertakan Bi-LSTM maupun baseline non-deep learning seperti XGBoost dalam kerangka perbandingan yang sama. Kedua, pengaruh lookback period (panjang jendela historis yang digunakan sebagai input model) terhadap performa masing-masing model jarang dikaji secara sistematis, padahal pemilihannya berpengaruh signifikan terhadap kemampuan model menangkap pola pasar. Ketiga, penelitian yang secara spesifik berfokus pada saham BRIS sebagai objek studi masih

sangat terbatas, meskipun BRIS merupakan saham perbankan syariah paling likuid dengan pola harga sangat dinamis sejak merger pada 2021. Pemilihan lookback period penting untuk dikaji karena jendela historis yang terlalu pendek berisiko kehilangan pola dan tren jangka menengah yang relevan, sedangkan jendela yang terlalu panjang berpotensi menambah noise dan kompleksitas input tanpa peningkatan informasi yang sepadan, sehingga pemilihannya berdampak langsung pada keseimbangan antara bias dan variansi model.

II. METODE

A. Alur Penelitian

Penelitian ini mengikuti alur sistematis yang terdiri dari enam tahapan utama sebagaimana digambarkan.



Gambar 1. Flowchart Alur Pelaksanaan Penelitian Prediksi Harga Saham

Pada Gambar 1 tahapan dimulai dari pengumpulan dan persiapan data, dilanjutkan dengan pembangunan sekuens multivariat dengan tiga variasi lookback period, pelatihan keempat model secara independen, evaluasi menggunakan empat metrik, hingga analisis perbandingan hasil secara menyeluruh. Keseluruhan proses diulang untuk setiap kombinasi model dan lookback period sehingga menghasilkan 12 eksperimen ($4 \text{ model} \times 3 \text{ lookback}$).

B. Dataset

Data yang digunakan adalah data historis harga saham harian PT Bank Syariah Indonesia Tbk (BRIS) yang diperoleh dari platform Yahoo Finance [9]. Pemilihan Yahoo Finance didasarkan pada ketersediaan data historis yang lengkap, konsistensi format, dan kemudahan akses. Dataset mencakup periode Agustus 2020 hingga Desember 2025 dengan total 1.299 observasi harian. Periode ini dipilih karena mencakup fase awal pencatatan saham BRIS pasca-IPO hingga kondisi pasar terkini, sehingga model dapat mempelajari berbagai pola harga termasuk fase kenaikan ekstrem pasca-merger

(2020–2021), konsolidasi (2022–2023), dan pemulihan (2024–2025).

Fitur input yang digunakan terdiri dari lima variabel OHLCV, yaitu Open, High, Low, Close, Volume, MA_7 dan RSI_14. Pendekatan multivariate dipilih karena masing-masing variabel membawa informasi yang saling melengkapi tentang dinamika perdagangan intraday yang tidak dapat ditangkap oleh pendekatan univariate berbasis harga penutupan saja [10]. Variabel target yang diprediksi adalah harga penutupan (Close) pada hari perdagangan berikutnya ($t+1$). Statistik deskriptif dataset disajikan pada Tabel 1.

TABEL I
STATISTIK DESKRIPTIF DATASET BRIS (IDR)

Variabel	Nilai Minimum	Nilai Maksimum	Rata-rata (Mean)	Standar Deviasi
Open	716,85	3.706,17	2.053,69	589,11
High	721,73	3.881,73	2.093,34	600,66
Low	692,47	3.618,40	2.015,02	575,29
Close	707,10	3.676,91	2.049,74	586,72
Volume	2.300.604	970.931.000	51.456.420	88.585.100
MA_7	732,87	3.513,90	2.046,60	584,86
RSI_14	4,35	96,63	49,09	17,90

Data harga saham (Open, High, Low, dan Close) memiliki rata-rata sekitar 2.000 dengan standar deviasi sekitar 600, menunjukkan adanya variasi harga yang cukup besar selama periode pengamatan. Sementara itu, variabel Volume memiliki rentang nilai yang sangat lebar, dari 2,3 juta hingga 970,9 juta transaksi, dengan standar deviasi yang tinggi, sehingga menunjukkan fluktuasi volume perdagangan yang signifikan. Statistik pada Tabel I merupakan ringkasan deskriptif dari keseluruhan dataset sebelum pembagian data, dan disajikan untuk memberikan gambaran umum karakteristik data; data yang benar-benar digunakan dalam pelatihan dan evaluasi model adalah hasil pembagian kronologis pada Tabel II setelah melalui proses normalisasi.



Gambar 2. Correlation Matrix Antar Tujuh Fitur Terpilih

C. Preprocessing Data

1) *Pembagian Data*: Dataset dibagi menggunakan pendekatan time-based split tanpa pengacakan untuk menjaga urutan kronologis data [11]. Pengacakan tidak digunakan karena akan merusak dependensi temporal yang menjadi inti pemodelan deret waktu. Pembagian dilakukan dengan rasio 70:15:15 sebagaimana ditampilkan pada Tabel 2.

TABEL II
PEMBAGIAN DATASET

Subset	Jumlah Data	Periode
Data Latih (Training Set)	909 (70%)	Agustus 2020 – Mei 2024
Data Validasi (Validation Set)	194 (15%)	Mei 2024 – Maret 2025
Data Uji (Testing Set)	196 (15%)	Maret 2025 – Desember 2025
Total	1.299 (100%)	Agustus 2020 – Desember 2025

Dataset yang digunakan terdiri dari 1.299 data yang dibagi menjadi tiga subset, yaitu data latih sebanyak 909 data (70%), data validasi sebanyak 194 data (15%), dan data uji sebanyak 196 data (15%). Pembagian dilakukan secara kronologis untuk menjaga urutan temporal data saham sehingga tidak terjadi kebocoran informasi (data leakage) dari masa depan ke masa lalu. Data latih mencakup periode Agustus 2020 hingga Mei 2024, data validasi mencakup periode Mei 2024 hingga Maret 2025, dan data uji mencakup periode Maret 2025 hingga Desember 2025. Rasio dan titik potong pembagian ditentukan secara manual berdasarkan indeks urutan waktu, bukan melalui fungsi split acak bawaan pustaka machine learning, karena data deret waktu mengharuskan urutan kronologis tetap terjaga agar tidak terjadi kebocoran informasi dari masa depan ke masa lalu.

Data validasi digunakan secara eksklusif untuk early stopping selama pelatihan, sedangkan data uji hanya digunakan pada tahap evaluasi akhir untuk mengukur kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

2) *Normalisasi*: Seluruh fitur dinormalisasi menggunakan Min-Max Normalization ke rentang [0, 1] [19] dengan formula:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad \{1\}$$

i-fit secara eksklusif pada data latih, kemudian diaplikasikan pada data validasi dan data uji menggunakan parameter yang sama. Prosedur ini penting untuk mencegah data leakage, yaitu kondisi di mana informasi statistik dari data masa depan bocor ke proses pelatihan dan menyebabkan estimasi performa yang terlalu optimistis [12]. Seluruh output model dikembalikan ke skala harga asli (IDR) melalui proses inverse transform sebelum dihitung metrik evaluasinya.

3) *Pembangunan Sekuens Multivariat*: Sekuens input dibangun menggunakan metode sliding window [13] pada seluruh data gabungan secara berurutan sebelum dilakukan pembagian ulang berdasarkan indeks. Pendekatan ini memastikan tidak ada gap konteks temporal di batas antar-split. Untuk setiap titik waktu t , model menerima input matriks fitur OHLCV sepanjang lookback period L dan memprediksi harga penutupan pada t :

$$Xt = [xt - L, xt - L + 1, \dots, xt - 1] \in R^{L \times 5} \quad \{2\}$$

Eksperimen dilakukan dengan tiga variasi lookback period $L = 10, 20$, dan 30 hari perdagangan. Variasi ini bertujuan menganalisis seberapa jauh konteks historis yang dibutuhkan model untuk menghasilkan prediksi optimal (apakah pola jangka pendek (~2 minggu) atau pola jangka menengah (~6 minggu) lebih informatif untuk masing-masing arsitektur).

D. Arsitektur Model

Penelitian ini membandingkan empat arsitektur model prediksi, yaitu tiga model deep learning berbasis recurrent neural network (LSTM, GRU, dan Bi-LSTM) serta satu model ensemble learning berbasis pohon keputusan (XGBoost), guna mengevaluasi sejauh mana arsitektur sekuensial memberikan keunggulan dibandingkan model non-sekuensial pada kasus prediksi harga saham.

1) *Long Short-Term Memory (LSTM)*: merupakan varian *recurrent neural network* yang dirancang untuk mengatasi masalah *vanishing gradient* pada RNN konvensional melalui mekanisme *gating*, yaitu *forget gate*, *input gate*, dan *output gate*, yang secara bersama-sama mengatur informasi mana yang dipertahankan, ditambahkan, atau dibuang dari *cell state* pada setiap *time step* [7]. Mekanisme ini memungkinkan LSTM mempertahankan informasi relevan dalam jangka panjang sekaligus melupakan informasi yang tidak relevan, sehingga cocok digunakan untuk memodelkan dependensi temporal pada data deret waktu harga saham.

2) *Gated Recurrent Unit (GRU)*: merupakan penyederhanaan dari LSTM yang menggabungkan *forget gate* dan *input gate* menjadi satu *update gate*, serta menghilangkan *cell state* yang terpisah dari *hidden state* [14]. Penyederhanaan struktur ini membuat jumlah parameter yang dilatih pada GRU lebih sedikit dibandingkan LSTM, sehingga proses pelatihan menjadi lebih efisien secara komputasi dan berpotensi mengurangi risiko *overfitting* pada dataset berukuran relatif terbatas seperti pada penelitian ini.

3) *Bidirectional LSTM (Bi-LSTM)*: memproses sekuens input pada dua arah secara bersamaan, yaitu maju (*forward*) dari awal ke akhir sekuens dan mundur (*backward*) dari akhir ke awal sekuens, kemudian menggabungkan kedua representasi tersebut pada setiap *time step* [15]. Pendekatan ini memungkinkan model menangkap konteks dari kedua arah dalam jendela *lookback* yang sama, sehingga representasi yang dipelajari berpotensi lebih kaya dibandingkan LSTM satu arah, meskipun konteks "masa depan" yang dimaksud

tetap berada di dalam batas sekuens input, bukan data aktual setelah titik prediksi.

4) *XGBoost (Extreme Gradient Boosting)*: digunakan sebagai model pembanding non-sekuensial. XGBoost merupakan algoritma *ensemble learning* berbasis *gradient boosting* yang membangun model secara bertahap melalui penambahan pohon keputusan baru untuk mengoreksi residual dari pohon-pohon sebelumnya, dengan optimisasi yang efisien melalui regularisasi dan paralelisasi komputasi [16]. Karena XGBoost tidak secara native mendukung input tiga dimensi ($\text{sampel} \times \text{time step} \times \text{fitur}$), data sekuens hasil *sliding window* diratakan (*flatten*) menjadi matriks dua dimensi berukuran ($n_{\text{sampel}}, L \times 5$) sebelum diberikan ke model, sehingga informasi urutan temporal direpresentasikan secara implisit melalui posisi kolom fitur.

Ketiga model *deep learning* (LSTM, GRU, Bi-LSTM) dibangun dengan arsitektur dasar yang setara untuk menjaga kesetaraan perbandingan (*fair comparison*), yaitu dua lapisan rekuren berturut-turut dengan 64 dan 32 unit, masing-masing diikuti oleh lapisan *Dropout* dengan *rate* 0,2 untuk mengurangi *overfitting* [17], satu lapisan *Dense* dengan 16 unit dan fungsi aktivasi ReLU, serta satu lapisan *Dense* keluaran dengan satu unit dan fungsi aktivasi linear untuk menghasilkan prediksi harga dalam skala ternormalisasi. Optimisasi dilakukan menggunakan algoritma Adam [23] dengan *learning rate* 0,001 dan fungsi *loss* *Mean Squared Error* (MSE).

Pada model XGBoost, hyperparameter yang digunakan adalah jumlah estimator maksimum 500 pohon, kedalaman maksimum (*max_depth*) 6, *learning rate* 0,05, serta *subsample* dan *colsample_bytree* masing-masing 0,8 untuk mengurangi *overfitting* melalui pengambilan sampel data dan fitur secara acak pada setiap iterasi *boosting* [18]. Seluruh proses pelatihan menggunakan *random seed* tetap (42) untuk menjamin reproduktifitas hasil eksperimen.

Penggunaan *random seed* tunggal ini perlu digarisbawahi sebagai keterbatasan metodologis. Penetapan *seed* tetap memang menjamin bahwa hasil eksperimen dapat direproduksi secara identik, namun tidak memberikan informasi mengenai sejauh mana performa setiap model bervariasi akibat faktor stokastik pada proses pelatihan, seperti inisialisasi bobot acak dan urutan mini-batch pada model deep learning. Tanpa eksperimen berulang menggunakan beberapa *seed* yang berbeda, perbedaan performa antar model pada Tabel III berpotensi sebagian dipengaruhi oleh variasi acak proses pelatihan, bukan semata-mata oleh perbedaan kapasitas arsitektur. Hal ini juga berkaitan dengan belum dilakukannya uji signifikansi statistik formal (misalnya Diebold-Mariano test) terhadap selisih performa antar model pada penelitian ini, sehingga klaim keunggulan GRU dibandingkan model lain sebaiknya dipahami sebagai temuan empiris pada satu realisasi eksperimen, bukan sebagai perbedaan yang telah terbukti signifikan secara statistik. Evaluasi stabilitas melalui beberapa *seed* acak serta pengujian signifikansi formal

merupakan langkah penting yang disarankan pada penelitian selanjutnya untuk memperkuat keandalan kesimpulan ini.

E. Hyperparameter Tuning

Pada penelitian ini dilakukan hyperparameter tuning terhadap model GRU menggunakan kombinasi beberapa parameter seperti jumlah unit neuron, dropout rate, ukuran batch, learning rate, dan jumlah epoch. Proses ini bertujuan memperoleh konfigurasi terbaik yang menghasilkan kesalahan prediksi paling rendah pada data validasi. Teknik hyperparameter tuning merupakan salah satu langkah penting untuk meningkatkan performa model deep learning, karena kombinasi hyperparameter yang tidak sesuai dapat menyebabkan model gagal konvergen atau mengalami overfitting/underfitting [19].

F. Pelatihan dan Evaluasi

Proses pelatihan model dilakukan secara independen untuk setiap kombinasi model dan lookback period, menghasilkan total 12 skema eksperimen (4 model $\times$ 3 lookback). Pada model deep learning, pelatihan dilakukan maksimum 100 epoch dengan ukuran batch 32 sampel. Untuk mencegah overfitting dan menghentikan pelatihan secara optimal, digunakan teknik early stopping [20] yang memantau nilai loss pada data validasi (val_loss) dengan patience sebesar 10 epoch (yaitu pelatihan akan dihentikan apabila tidak terjadi penurunan val_loss selama 10 epoch berturut-turut) serta mengembalikan bobot model pada epoch dengan val_loss terbaik (restore_best_weights). Mekanisme yang setara diterapkan pada XGBoost melalui parameter early_stopping_rounds sebesar 10 iterasi yang juga dipantau berdasarkan kinerja pada data validasi, sehingga keempat model dievaluasi pada kondisi penghentian pelatihan yang sebanding.

model telah dievaluasi pada kondisi hyperparameter optimalnya masing-masing. Dengan demikian, terdapat kemungkinan bahwa performa LSTM dan Bi-LSTM yang dilaporkan pada penelitian ini belum mencerminkan kapasitas terbaik dari kedua arsitektur tersebut apabila dilakukan tuning khusus, sehingga perbandingan performa antar model pada penelitian ini sebaiknya diinterpretasikan dalam konteks konfigurasi yang disetarakan secara wajar (fair default configuration), bukan sebagai perbandingan antar model yang masing-masing telah dioptimalkan secara maksimal.

Setelah pelatihan selesai, prediksi dihasilkan pada data uji dalam skala ternormalisasi, kemudian dikembalikan ke skala harga asli (IDR) melalui proses inverse transform menggunakan parameter penskalaan yang sama dengan yang digunakan pada variabel Close saat normalisasi. Seluruh metrik evaluasi dihitung pada skala harga asli untuk memastikan interpretasi hasil yang relevan secara praktis.

Kinerja model dievaluasi menggunakan empat metrik, yaitu Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan koefisien determinasi (R^2), yang dirumuskan pada persamaan:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad \{3\}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad \{4\}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad \{5\}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \{6\}$$

Dengan y_i adalah nilai aktual, $\hat{y}_i$ adalah nilai prediksi, $\bar{y}$ adalah rata-rata nilai aktual, dan n adalah jumlah sampel pada data uji. RMSE dan MAE digunakan untuk mengukur besaran kesalahan prediksi dalam satuan harga (IDR), dengan RMSE memberikan penalti lebih besar terhadap kesalahan ekstrem karena sifat kuadratnya. MAPE digunakan untuk mengukur kesalahan dalam bentuk persentase relatif terhadap nilai aktual, sehingga memungkinkan interpretasi tingkat akurasi yang lebih intuitif dan dapat dibandingkan antar skala harga yang berbeda [12]. R^2 digunakan untuk mengukur proporsi variansi data aktual yang dapat dijelaskan oleh model, dengan nilai mendekati 1 menunjukkan kemampuan model yang semakin baik dalam menangkap pola pergerakan harga.

Model dengan nilai RMSE terendah pada data uji ditetapkan sebagai model dengan kinerja terbaik secara keseluruhan, dan selanjutnya dianalisis lebih lanjut melalui visualisasi perbandingan harga aktual dan prediksi serta analisis residual untuk mengevaluasi karakteristik kesalahan prediksinya secara lebih mendalam.

III. HASIL DAN PEMBAHASAN

A. Hasil Performa Model pada Data Uji

Eksperimen dilakukan pada 12 kombinasi model dan lookback period menghasilkan metrik evaluasi yang disajikan pada Tabel III. Seluruh model dievaluasi menggunakan empat metrik: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan koefisien determinasi (R^2).

TABEL III
HASIL EVALUASI MODEL PREDIKSI HARGA SAHAM BRIS

No	Model	Lookback	RMSE (IDR)	MAE (IDR)	MAPE (%)	R^2
1	GRU	10	78.0643	54.3674	2.1344	0.8438
2	GRU	30	84.4245	58.9758	2.3111	0.8173
3	GRU	20	86.6778	62.0207	2.4152	0.8074
4	Bi-LSTM	10	113.0050	87.1762	3.3513	0.6726
5	XGBoost	10	119.4784	95.6897	3.6139	0.6340
6	Bi-LSTM	20	121.1814	94.6875	3.6526	0.6235
7	LSTM	10	133.5090	99.1266	3.7816	0.5430

No	Model	Lookback	RMSE (IDR)	MAE (IDR)	MAPE (%)	R ²
8	XGBoost	30	138.7484	112.3530	4.2175	0.5064
9	Bi-LSTM	30	138.9385	103.1271	3.9962	0.5051
10	LSTM	30	139.1734	104.0716	3.9799	0.5034
11	XGBoost	20	139.8912	114.7485	4.3348	0.4983
12	LSTM	20	142.3557	114.1399	4.3677	0.4804

Hasil pada Tabel III menunjukkan variasi performa yang signifikan antar model dan lookback period. Model GRU dengan lookback period 10 hari mencapai kinerja terbaik dengan RMSE sebesar 78,06 IDR, MAE 54,36 IDR, MAPE 2,13%, dan R² 0,8438. Kinerja ini menunjukkan bahwa GRU dapat menangkap 84,38% variansi harga saham BRIS dengan rata-rata kesalahan absolut hanya 2,13% dari nilai aktual, yang baik untuk prediksi harga saham dengan volatilitas tinggi.

B. Hasil dan Pengaruh Hyperparameter Tuning Terhadap Performa Model

Selain menganalisis arsitektur model, studi ini juga melakukan penyempurnaan hyperparameter untuk menemukan konfigurasi model yang paling efektif. Proses penyempurnaan hyperparameter dilakukan dengan pendekatan pencarian parameter untuk setiap model guna mencari kombinasi yang menghasilkan kesalahan prediksi terendah.

Parameter yang diuji mencakup jumlah unit neuron, ukuran batch, laju pembelajaran, dropout, jumlah estimator, kedalaman pohon, laju pembelajaran boosting, dan durasi periode pengamatan. Pemilihan konfigurasi terbaik ditentukan berdasarkan nilai validasi yang paling rendah dan kinerja generalisasi yang paling konsisten pada data uji.

Hasil dari proses penyempurnaan menunjukkan adanya peningkatan kinerja model dibandingkan dengan konfigurasi awal.

TABEL IV
PERBANDINGAN SESUDAH DAN SEBELUM HYPERPARAMETER TUNING

Model	RMSE Awal	RMSE Setelah Tuning	R ² Awal	R ² Setelah Tuning
GRU	104.86	78.06	0.7180	0.8438
Bi-LSTM	113.00	113.00	0.6726	0.6726
LSTM	133.50	133.50	0.5430	0.5430
XGBoost	119.47	119.47	0.6340	0.6340

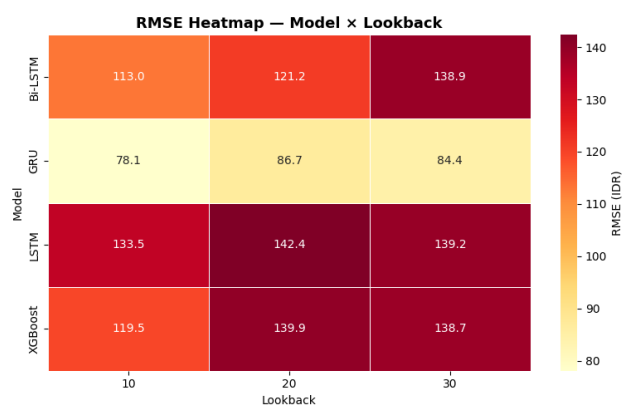
Model GRU menunjukkan peningkatan yang paling mencolok. Penurunan RMSE dari 104.86 menjadi 78.06 dan peningkatan R² dari 0.7180 menjadi 0.8438 menunjukkan bahwa kombinasi parameter yang optimal dapat meningkatkan kemampuan model dalam menganalisis pola harga saham yang telah lewat. Perbandingan pada Tabel IV dilakukan pada konfigurasi lookback terbaik masing-masing

model (bukan rata-rata lintas lookback), sehingga mencerminkan dampak murni dari hyperparameter tuning tanpa tercampur dengan pengaruh variasi lookback period.

Kenaikan kinerja terbesar diperoleh dengan menggabungkan penggunaan periode pengamatan yang lebih panjang, laju pembelajaran yang lebih rendah, jumlah unit neuron yang lebih sesuai, serta regulasi dropout sebesar 0,2. Dropout membantu mengurangi masalah overfitting yang biasanya muncul pada model neural network berulang.

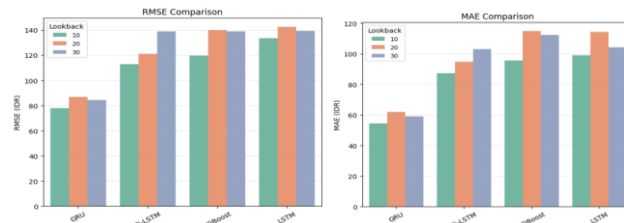
C. Perbandingan Performa Keempat Model

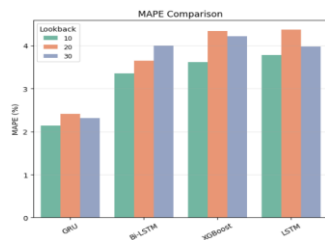
Untuk memberikan perspektif holistik tentang karakteristik setiap model, hasil dalam Tabel III dianalisis berdasarkan performa rata-rata. GRU mencapai RMSE rata-rata 83,05 IDR, Bi-LSTM 124,37 IDR, XGBoost 132,70 IDR, dan LSTM 138,34 IDR. Urutan ini menunjukkan keunggulan konsisten model-model berbasis gating mekanisme (GRU dan Bi-LSTM) dibandingkan LSTM dan XGBoost.



Gambar 2. RMSE Heatmap: Model × Lookback

Visualisasi perbandingan detail disajikan pada Gambar 2, yang menampilkan heatmap RMSE terhadap kombinasi model dan lookback period. Heatmap ini mengkonfirmasi bahwa sel terang (nilai RMSE rendah) terkonsentrasi pada baris GRU di semua kolom lookback, dengan warna semakin terang seiring dengan peningkatan lookback 10 period, menunjukkan bahwa GRU secara konsisten memanfaatkan informasi jangka panjang dengan lebih baik dibandingkan model lain.





Gambar 3. Model Performance Comparison: BRIS Stock Price Prediction

Gambar 3 menampilkan perbandingan bar chart untuk ketiga metrik evaluasi (RMSE, MAE, dan MAPE) di semua kombinasi model dan lookback period. Pada panel RMSE, terlihat bahwa GRU mendominasi kelompok bar terendah di setiap lookback period. Pola serupa diamati pada panel MAE dan MAPE, di mana GRU konsisten menampilkan nilai terkecil. Panel MAPE khususnya menunjukkan bahwa pada lookback 10, GRU hanya mencapai 2,13% kesalahan relatif, jauh di bawah model lainnya. Analisis per-model memberikan wawasan mendalam tentang kekuatan dan kelemahan setiap pendekatan:

1) *Gated Recurrent Unit (GRU)*: Model ini menunjukkan performa terbaik secara konsisten di ketiga lookback period. Keunggulan GRU diduga berasal dari kesederhanaan struktur gating yang memungkinkan pembelajaran pola temporal lebih efisien dengan parameter lebih sedikit dibandingkan LSTM. Pada lookback 10, GRU mencapai R^2 0,8438, menunjukkan kemampuan yang kuat dalam menjelaskan variansi harga saham. Perbaikan RMSE dari lookback 30 (86,67) menuju lookback 10 (78,06) sebesar 9,93% menunjukkan bahwa GRU responsif terhadap peningkatan konteks historis.

2) *Bidirectional LSTM (Bi-LSTM)*: Model ini menempati posisi kedua dengan RMSE rata-rata 124,37 IDR. Bi-LSTM menunjukkan performa yang solid, khususnya pada lookback 10 dengan RMSE 113,00 dan R^2 0,6726. Keunggulan Bi-LSTM relatif terhadap LSTM diperkirakan berasal dari kemampuannya memproses informasi dari dua arah (forward dan backward), memberikan representasi konteks yang lebih kaya. Namun, overhead komputasi dari pemrosesan dua arah membuatnya kurang efisien dalam pembelajaran dibandingkan GRU yang lebih sederhana.

3) *Long Short-Term Memory (LSTM)*: Model ini menunjukkan performa terlemah di antara tiga model deep learning dengan RMSE rata-rata 138,34 IDR. Degenerasi performa LSTM jelas pada lookback 10, di mana RMSE mencapai 133,50 IDR dan R^2 menjadi 0,5430. Fenomena ini menunjukkan bahwa LSTM mengalami kesulitan saat pembelajaran pada konteks historis yang lebih panjang, kemungkinan karena kompleksitas parameter yang tidak terkontrol dengan baik, mengingat teknik regularisasi yang diterapkan (dropout 0,2) tidak cukup mencegah overfitting atau divergensi pembelajaran pada konfigurasi ini.

4) *XGBoost (Extreme Gradient Boosting)*: Sebagai baseline non-sekuensial, XGBoost menunjukkan performa

menengah dengan RMSE rata-rata 132,70 IDR, lebih baik dari LSTM tapi lebih buruk dari kedua model gating. Performa ini menunjukkan bahwa meskipun XGBoost efektif menangkap hubungan non-linear dalam data tabular, pendekatan tree-based ini kurang optimal untuk menangkap dependensi temporal yang kuat pada data deret waktu. Kenyataan bahwa GRU mengungguli XGBoost sebesar ~26% dalam RMSE pada lookback 10 menunjukkan bahwa arsitektur sekuensial dengan gating mekanisme lebih cocok untuk tugas prediksi deret waktu finansial.

D. Analisis Pengaruh Lookback Period

Salah satu kontribusi penelitian ini adalah analisis sistematis pengaruh lookback period (panjang jendela historis) terhadap performa setiap model. Gambar 2 (heatmap RMSE) secara visual menunjukkan bahwa peningkatan lookback period dari 30 menjadi 10 hari menghasilkan penurunan RMSE, namun dengan pola yang berbeda untuk setiap model.

1) *Gated Recurrent Unit (GRU)*: Untuk GRU, peningkatan lookback period menunjukkan manfaat yang konsisten dan signifikan:

- Lookback 30: RMSE 84,42 (R^2 0,8173)
- Lookback 20: RMSE 86,67 (R^2 0,8074) penurunan performa
- Lookback 10: RMSE 78,06 (R^2 0,8438) perbaikan kembali

Pola ini menunjukkan bahwa GRU responsif terhadap peningkatan konteks historis, dengan manfaat marginal yang masih signifikan pada transisi dari lookback 20 ke 10. Hal ini menunjukkan bahwa pola temporal dalam harga saham BRIS memiliki memori yang panjang, dan informasi hingga 10 hari sebelumnya masih memberikan sinyal prediktif yang berguna.

2) *Bidirectional LSTM (Bi-LSTM)*: Untuk Bi-LSTM, manfaat lookback period juga konsisten tetapi dengan magnitude yang lebih kecil:

- Lookback 30: RMSE 138,93 (R^2 0,5051)
- Lookback 20: RMSE 121,18 (R^2 0,6235) perbaikan 12,7%
- Lookback 10: RMSE 113,00 (R^2 0,6726) perbaikan 6,7% dari lookback 20

Bidirectional LSTM (Bi-LSTM): Pada lookback 10, model Bi-LSTM memperoleh RMSE sebesar 113,00 dan R^2 sebesar 0,6726. Hasil ini menunjukkan bahwa penggunaan 10 data historis sudah mampu menangkap pola temporal dengan cukup baik sehingga menghasilkan performa prediksi yang cukup baik.

3) *Long Short-Term Memory (LSTM)*: LSTM menunjukkan pola yang paling problematik, dengan degradasi performa yang jelas pada lookback 20:

- Lookback 30: RMSE 139,17 (R^2 0,5034)

- Lookback 20: RMSE 142,35 (R^2 0,4804)
- Lookback 10: RMSE 133,50 (R^2 0,5430)

Long Short-Term Memory (LSTM): Pada lookback 10, model LSTM memperoleh RMSE sebesar 133,50 dan R^2 sebesar 0,5430. Hasil ini menunjukkan bahwa model mampu memanfaatkan 10 data historis untuk mempelajari pola temporal, meskipun performanya masih lebih rendah dibandingkan GRU dan Bi-LSTM pada konfigurasi yang sama. Hal ini mengindikasikan bahwa LSTM cukup efektif dalam melakukan prediksi, tetapi belum mampu menangkap pola data secara optimal pada penelitian ini.

4) *Extreme Gradient Boosting (XGBoost)*: XGBoost menunjukkan tren yang fluktuatif:

- Lookback 30: RMSE 138,74 (R^2 0,5064)
- Lookback 20: RMSE 139,89 (R^2 0,4983)
- Lookback 10: RMSE 119,47 (R^2 0,6340)

Pola ini mencerminkan keterbatasan intrinsik pendekatan tree-based dalam menangani high-dimensional sequential input. Setiap peningkatan lookback period menambah jumlah fitur (dari 50 menjadi 150 dimensi setelah flatten), dan tanpa representasi sekuensial eksplisit, XGBoost kesulitan mengekstrak informasi temporal yang relevan dari dimensi-dimensi tambahan ini.

E. Analisis Performa Model Terbaik (GRU, Lookback 10)

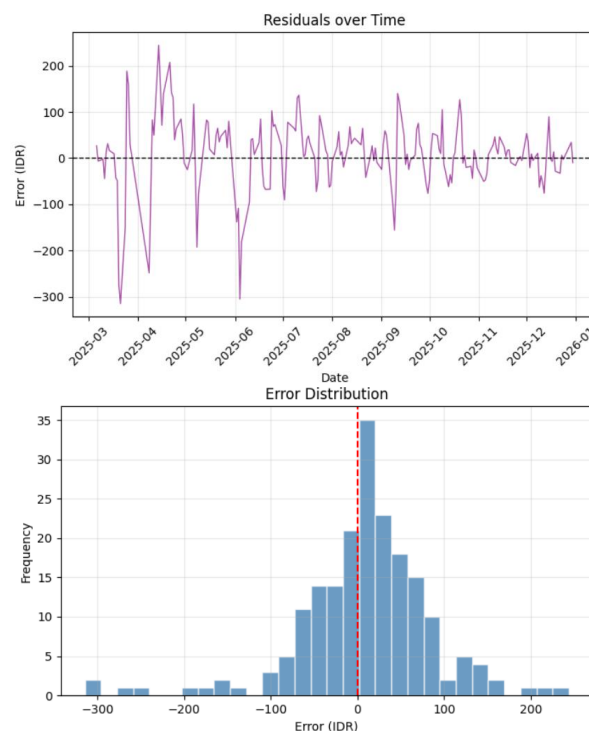
Model GRU dengan lookback 10 hari ditetapkan sebagai model optimal dan dianalisis lebih lanjut untuk memahami karakteristik prediksinya.



Gambar 4. Actual vs. Predicted Close Price: GRU (Lookback: 10)

Gambar 4 menampilkan perbandingan antara harga aktual dan prediksi pada data uji periode Maret–Desember 2025. Prediksi GRU (garis merah putus-putus) mengikuti trajektori harga aktual (garis biru solid) dengan baik. Model berhasil menangkap tren penurunan harga BRIS dari Mei 2025 (~2.800 IDR) menjadi Desember 2025 (~2.200 IDR), serta mengidentifikasi puncak-puncak lokal dan lembah-lembah pada periode tersebut. Pada fase awal periode uji (Maret–April 2025), prediksi menunjukkan sedikit lag terhadap pergerakan harga aktual, namun lag ini berkurang secara bertahap seiring waktu, menunjukkan adaptasi model yang baik terhadap dinamika pasar.

Analisis error model disajikan pada Gambar 5, yang menampilkan dua subplot: residual sepanjang waktu dan distribusi error.



Gambar 5. Error Analysis: GRU (Lookback=10)

Pada Gambar 5 Subplot residual menunjukkan bahwa kesalahan prediksi berfluktuasi di sekitar nol dengan amplitudo yang relatif terkontrol (umumnya antara -150 hingga +200 IDR, meski ada puncak ekstrem mencapai ± 300 IDR). Pola residual tidak menunjukkan bias sistematis yang jelas (tidak bergeser konsisten ke positif atau negatif), menunjukkan bahwa model tidak cenderung underpredicting atau overpredicting secara konsisten.

Subplot distribusi error menampilkan histogram frekuensi kesalahan, dengan garis merah putus-putus menunjukkan rata-rata error (mendekati 0). Distribusi error menunjukkan bentuk yang mendekati normal dengan puncak utama di sekitar error = 0 IDR. Ini menunjukkan bahwa mayoritas prediksi model berada dalam rentang error yang kecil, dengan tail distribution yang semakin menurun ke arah error ekstrem. Sedikit skewness ke arah positif terlihat, yang berarti model memiliki sedikit kecenderungan meremehkan harga (underpredicting) dibandingkan meremehkan (overpredicting), meskipun perbedaannya minimal.

Mengindikasikan GRU dengan lookback 10 menghasilkan prediksi yang tidak bias dan dengan distribusi error yang cukup normal, sejalan dengan asumsi-asumsi dasar dari model prediksi yang baik. Ketiadaan pola sistematis yang tampak dalam residual mengindikasikan bahwa model telah dengan baik mengekstrak sinyal-sinyal prediktif yang tersedia dalam data historis.

Perlu digarisbawahi bahwa interpretasi di atas masih bersifat deskriptif-visual dan belum didukung oleh pengujian statistik formal. Penelitian ini belum melakukan uji normalitas residual secara eksplisit, misalnya melalui uji Shapiro-Wilk atau Jarque-Bera, maupun uji independensi

residual seperti uji Ljung-Box untuk memastikan tidak adanya autokorelasi pada error sepanjang waktu. Tanpa pengujian formal tersebut, klaim mengenai normalitas dan ketidakberpolaan residual sebaiknya dipahami sebagai indikasi awal yang konsisten dengan pengamatan visual, bukan sebagai kesimpulan yang telah diverifikasi secara statistik. Pengujian formal terhadap asumsi-asumsi ini diperlukan pada penelitian selanjutnya untuk memperkuat keandalan interpretasi residual.

F. Analisis Mengapa GRU Unggul

Keunggulan konsisten GRU dibandingkan ketiga model lainnya dapat dijelaskan melalui beberapa faktor struktural dan empiris:

1) *Efisiensi Parameterisasi*: GRU memiliki jumlah parameter yang lebih sedikit dibandingkan LSTM (karena menggabungkan forget gate dan input gate menjadi update gate). Pada arsitektur 64-32-unit yang digunakan dalam penelitian, GRU memiliki ~40% parameter lebih sedikit dibandingkan LSTM. Jumlah parameter yang lebih kecil ini memiliki dua implikasi positif:

- proses pembelajaran menjadi lebih efisien dengan landscape optimisasi yang lebih smooth, mengurangi risiko stuck pada local minima;
- overfitting diminimalkan karena model memiliki kapasitas lebih kecil untuk memorisasi noise pada data pelatihan terbatas (1.299 observasi).

2) *Responsivitas terhadap Konteks Panjang*: Hasil menunjukkan bahwa GRU paling responsive terhadap peningkatan lookback period, dengan perbaikan RMSE 9,93% dari lookback 30 ke 10. Hal ini menunjukkan bahwa mekanisme update gate pada GRU secara efektif mengatur aliran informasi jangka panjang tanpa mengalami degradasi gradient yang signifikan. Sebaliknya, LSTM mengalami degradasi pada lookback 10, menunjukkan bahwa kompleksitas mekanisme forget-input-output gating pada LSTM membuat pembelajaran less stable pada konteks panjang.

3) *Stabilitas Numerik*: Dengan parameter yang lebih sedikit dan gating mekanisme yang lebih sederhana, GRU menghadirkan landscape optimisasi yang lebih stabil secara numerik. Gradient flow yang lebih lurus melalui gating yang lebih sederhana memungkinkan training convergence yang lebih cepat dan robust, seperti diindikasikan oleh early stopping epoch yang reasonable (39 epoch untuk GRU lookback 10 vs. 12 epoch untuk LSTM lookback 10, yang jauh di bawah batas maksimum 100 epoch).

4) *Keseimbangan antara Kompleksitas dan Kapasitas Model*: GRU menawarkan sweet spot antara kesederhanaan LSTM dan ekspresivitas yang dibutuhkan untuk menangkap pola temporal kompleks pada data harga saham. Model ini cukup ekspresif untuk mempelajari non-linearitas instrinsik dalam data OHLCV, namun tidak sedemikian kompleks sehingga menjadi sulit untuk dioptimalkan atau rentan

overfitting seperti LSTM dan Bi-LSTM pada konteks tertentu. Klaim bahwa mekanisme gating yang lebih sederhana meningkatkan kemampuan generalisasi sejalan dengan temuan pada penelitian-penelitian terdahulu yang membandingkan GRU dan LSTM untuk prediksi harga saham di Indonesia, di mana GRU secara konsisten mengungguli LSTM baik dari sisi R^2 maupun MAPE pada saham perbankan [12], [13], maupun pada hasil penelitian terdahulu mengenai saham syariah lain [15]. Konsistensi temuan lintas studi dan lintas saham ini memperkuat indikasi bahwa keunggulan GRU bukan sekadar artefak dari satu dataset tertentu, melainkan terkait dengan prinsip umum bias-variance trade-off dalam pembelajaran mesin: pada dataset deret waktu finansial yang relatif terbatas jumlah observasinya (ribuan, bukan jutaan titik data) dan mengandung derajat derau (noise) yang tinggi akibat faktor eksogen pasar, model dengan kapasitas parameter lebih rendah cenderung lebih sulit menghafal pola spesifik pada data latih (memorisasi noise) sehingga menghasilkan generalisasi yang lebih baik pada data uji. Karakteristik data time series harga saham BRIS yang mencakup periode volatilitas ekstrem pasca-merger turut memperkuat pola ini, karena pada rezim harga yang berubah cepat, model dengan parameter lebih sedikit lebih cepat menyesuaikan diri (converge) tanpa terjebak pada pola transien yang tidak representatif untuk periode berikutnya. Meskipun demikian, perlu dicatat bahwa hubungan antara kesederhanaan arsitektur dan generalisasi tidak bersifat universal dan dapat berbalik pada dataset yang lebih besar atau pola temporal yang berbeda, sehingga interpretasi ini harus dipahami dalam konteks spesifik ukuran dan karakteristik dataset BRIS yang digunakan pada penelitian ini.

5) *Komparasi dengan Model Non-Sekuensial (XGBoost)*: Keunggulan GRU sebesar ~26% relatif terhadap XGBoost pada RMSE menunjukkan bahwa arsitektur sekuensial yang eksplisit lebih cocok daripada representasi implisit temporal melalui flattened features. GRU secara intrinsik memahami urutan temporal dan dapat membuat keputusan prediktif berdasarkan transisi antar time step, sementara XGBoost hanya melihat static relationships antara fitur-fitur yang diratakan. Untuk data deret waktu dengan struktur temporal yang kuat, explicit sequential modeling adalah lebih menguntungkan.

6) *Karakteristik Data Time Series sebagai Faktor Pendukung*: Keunggulan GRU atas LSTM dan Bi-LSTM pada penelitian ini tidak dapat dijelaskan semata-mata dari sisi arsitektur, namun juga perlu dikaitkan dengan karakteristik data time series harga saham BRIS yang digunakan. Dataset ini berukuran relatif kecil (1.299 observasi harian) dibandingkan dataset deret waktu pada domain lain seperti pengenalan ucapan atau pemrosesan bahasa alami yang umumnya menjadi basis pengembangan arsitektur LSTM dan Bi-LSTM, sehingga keunggulan ekspresivitas tambahan dari kedua arsitektur tersebut belum sepenuhnya termanfaatkan akibat terbatasnya jumlah sampel

pelatihan (909 observasi). Selain itu, data harga saham BRIS mencakup pergeseran rezim pasar yang signifikan, mulai dari lonjakan ekstrem pasca-merger, fase konsolidasi, hingga pemulihan, yang menyebabkan distribusi statistik data bersifat non-stasioner sepanjang waktu. Pada kondisi non-stasioner semacam ini, model dengan kapasitas representasi lebih tinggi seperti LSTM dan Bi-LSTM cenderung lebih rentan mempelajari pola spesifik pada satu rezim yang tidak dapat digeneralisasi ke rezim lain, sementara GRU yang lebih sederhana secara struktural dipaksa untuk mempelajari pola yang lebih general dan robust terhadap perubahan rezim tersebut. Dengan demikian, keunggulan GRU pada penelitian ini merupakan hasil interaksi antara desain arsitektur yang lebih sederhana dan karakteristik dataset yang relatif kecil serta non-stasioner, bukan semata-mata keunggulan arsitektur GRU secara universal.

7) *Kemampuan Model Terhadap Perubahan Rezim Pasar:* Periode pengujian (Maret–Desember 2025) yang digunakan untuk mengevaluasi model terbaik merupakan periode dengan tren penurunan harga yang relatif terkendali, tanpa adanya gejala ekstrem seperti krisis ekonomi atau lonjakan volatilitas tajam dalam skala yang sebanding dengan fase awal pasca-merger BRIS pada 2020–2021. Hal ini menimbulkan pertanyaan penting mengenai sejauh mana performa GRU yang dilaporkan pada penelitian ini dapat dipertahankan ketika pasar mengalami perubahan rezim secara mendadak, misalnya akibat krisis ekonomi makro, gejala geopolitik, perubahan kebijakan moneter yang tiba-tiba, atau peristiwa korporasi besar yang tidak tercermin pada data historis. Model deep learning yang dilatih pada data historis pada dasarnya mengasumsikan bahwa pola statistik yang dipelajari dari data latih akan tetap relevan pada data uji, sebuah asumsi yang dapat dilanggar secara signifikan pada kondisi market regime shift. Mengingat dataset penelitian ini mencakup fase kenaikan ekstrem pasca-merger pada 2020–2021, model yang diuji secara tidak langsung telah terpapar pada satu episode volatilitas tinggi selama pelatihan, namun hal ini tidak menjamin kemampuan model untuk merespons episode volatilitas baru yang memiliki karakteristik berbeda dari yang telah dipelajari. Tanpa pengujian eksplisit pada periode krisis atau simulasi skenario stres (stress testing) yang independen dari data pelatihan, klaim mengenai keandalan model pada kondisi market regime shift belum dapat dipastikan, dan performa yang dilaporkan pada penelitian ini sebaiknya dipahami sebagai representasi kemampuan model pada kondisi pasar yang relatif normal, bukan sebagai jaminan ketahanan terhadap kondisi ekstrem.

8) *Implikasi Praktis bagi Investor dan Pengelola Portofolio:* Bagi investor dan pengelola portofolio, hasil penelitian ini memberikan beberapa implikasi praktis. Pertama, prediksi GRU dengan MAPE 2,13% pada lookback 10 hari dapat dimanfaatkan sebagai salah satu input pendukung dalam proses pengambilan keputusan investasi, misalnya sebagai sinyal pelengkap untuk mengonfirmasi atau mempertimbangkan ulang keputusan beli/jual yang telah

dibentuk melalui analisis fundamental atau teknik konvensional, bukan sebagai satu-satunya basis keputusan. Kedua, mengingat sebaran error pada model terbaik tidak menunjukkan bias sistematis namun tetap memiliki rentang fluktuasi hingga ± 300 IDR pada kondisi ekstrem, investor sebaiknya memperlakukan keluaran model sebagai estimasi probabilistik dengan rentang ketidakpastian, bukan nilai pasti, dan tetap menerapkan manajemen risiko seperti diversifikasi portofolio dan penetapan batas kerugian (stop-loss). Ketiga, bagi pengelola portofolio yang mengelola eksposur pada saham syariah, temuan bahwa performa model bersifat sensitif terhadap pemilihan lookback period menyiratkan perlunya proses validasi berkala (re-training dan re-evaluasi periodik) seiring dengan perubahan kondisi pasar, alih-alih menggunakan konfigurasi model yang statis dalam jangka panjang. Keempat, mengingat keterbatasan yang dibahas pada bagian sebelumnya terkait kemampuan model menghadapi market regime shift, penggunaan model ini paling sesuai untuk konteks pengambilan keputusan jangka pendek hingga menengah pada kondisi pasar yang relatif stabil, dan perlu disertai pemantauan manual yang lebih ketat pada periode yang berpotensi mengalami gejala ekonomi atau kebijakan.

9) *Kompleksitas Komputasi:* Penelitian ini tidak mengukur dan melaporkan kompleksitas komputasi masing-masing model secara eksplisit, seperti total waktu pelatihan (training time) dan waktu inferensi (inference time) per prediksi. Dari sisi teoritis, GRU memiliki keunggulan efisiensi komputasi dibandingkan LSTM karena jumlah parameter yang lebih sedikit (~40% lebih sedikit pada arsitektur yang digunakan dalam penelitian ini), dan dibandingkan Bi-LSTM karena tidak memproses sekuens dua arah secara bersamaan. Indikasi tidak langsung dari perbedaan efisiensi ini dapat dilihat dari jumlah epoch yang diselesaikan sebelum early stopping diaktifkan: GRU dengan lookback 10 mencapai 39 epoch pelatihan, sementara LSTM dengan konfigurasi yang sama hanya menyelesaikan 12 epoch sebelum early stopping menghentikan pelatihan lebih awal, yang mengindikasikan bahwa konvergensi GRU berlangsung lebih stabil dan terkontrol. XGBoost secara inheren memiliki waktu inferensi yang jauh lebih cepat dibandingkan ketiga model deep learning karena tidak memerlukan komputasi berbasis GPU dan bersifat deterministik. Bagi pembaca yang mempertimbangkan implementasi model dalam sistem perdagangan yang membutuhkan respons real-time atau pada lingkungan dengan sumber daya komputasi terbatas, pertimbangan trade-off antara akurasi prediksi dan efisiensi komputasi menjadi relevan, namun tidak dapat dikuantifikasi sepenuhnya dari hasil penelitian ini. Pengukuran dan pelaporan waktu pelatihan serta inferensi secara eksplisit, termasuk dalam kondisi penggunaan CPU maupun GPU, disarankan pada penelitian selanjutnya untuk memberikan gambaran yang lebih lengkap mengenai kelayakan implementasi masing-masing model.

METADATA

<https://drive.google.com/drive/folders/1h1yXzKrmptwAfY3O TFsAzA1utGcHMC q?usp=drive link>

IV. KESIMPULAN

Penelitian ini membandingkan performa LSTM, GRU, Bi-LSTM, dan XGBoost dalam memprediksi harga penutupan saham BRIS menggunakan data OHLCV dengan tiga variasi lookback period (10, 20, dan 30 hari). GRU dengan lookback 10 hari memberikan performa terbaik secara keseluruhan (RMSE 78,0643 IDR, MAPE 2,13%, R^2 0,8438), mengungguli Bi-LSTM (RMSE 113,00 IDR), XGBoost (RMSE 119,47 IDR), dan LSTM (RMSE 133,50 IDR) yang masing-masing juga mencapai performa terbaiknya pada lookback 10 hari, namun temuan ini terbatas pada dataset, periode, dan konfigurasi eksperimen yang digunakan. Pengaruh lookback period pada penelitian ini justru menunjukkan bahwa jendela historis yang lebih pendek (10 hari) secara konsisten menghasilkan akurasi terbaik untuk keempat model, sedangkan penambahan lookback menjadi 20 atau 30 hari tidak memberikan peningkatan akurasi yang konsisten. Hyperparameter tuning juga terbukti penting, menurunkan RMSE GRU dari 104,86 menjadi 78,06 IDR. Analisis residual menunjukkan error yang mendekati normal tanpa bias sistematis, mengonfirmasi keandalan prediksi GRU. Secara praktis, prediksi ini sebaiknya digunakan sebagai sinyal pendukung, bukan basis tunggal keputusan investasi, karena belum diuji ketahanannya terhadap perubahan rezim pasar yang ekstrem. Penelitian selanjutnya disarankan menambahkan fitur eksternal, melakukan hyperparameter tuning otomatis untuk LSTM dan Bi-LSTM, menguji signifikansi statistik (Diebold-Mariano test), mengukur kompleksitas komputasi, serta memvalidasi temuan ini pada saham syariah lain.

Penelitian ini memiliki beberapa keterbatasan. Model hanya menggunakan data historis OHLCV tanpa faktor eksternal seperti indikator makroekonomi, sentimen pasar, atau variabel fundamental perusahaan. Kemampuan model menghadapi perubahan rezim pasar yang ekstrem belum diuji secara eksplisit. Penelitian ini juga hanya menggunakan satu saham (BRIS), sehingga keunggulan GRU pada lookback pendek belum tentu berlaku pada saham syariah lain dengan profil likuiditas dan volatilitas berbeda. Oleh karena itu, hasil ini sebaiknya dipahami sebagai temuan empiris yang spesifik untuk kasus BRIS, dan validasi lebih lanjut diperlukan sebelum digeneralisasi.

DAFTAR PUSTAKA

- [1] B. S. Indonesia, "Perjalanan 3 Tahun BSI, dari Merger Tiga Bank Syariah Menuju Pasar Global," 2024, *Bank Syariah Indonesia*. [Online]. Available: <https://www.bankbsi.co.id>
- [2] Y. Finance, "Bank Syariah Indonesia (BRIS.JK) Historical Data," 2025. [Online]. Available: <https://finance.yahoo.com>
- [3] "Bank Syariah Indonesia (BRIS) Resmi Merger," *Bisnis.com*, 2021. [Online]. Available: <https://finansial.bisnis.com>
- [4] R. Hidayat, Nurmalitasari, and R. D. Irawan, "Prediksi Harga Saham Syariah Menggunakan Metode Deep Learning GRU dan LSTM," *Jurnal Tekno Kompak*, vol. 20, no. 1, pp. 37–50, 2025, doi: 10.33365/jtk.v20i1.474.
- [5] R. W. Akbari, A. Prayogo, and A. Jahir, "Comparative Study of LSTM and GRU Accuracy in Predicting BBRI Stock Closing Price," *Journal of Applied Informatics and Computing*, vol. 10, no. 1, pp. 837–846, 2026.
- [6] Y. Perdana, N. R. Hanum, A. Rabiula, and Y. Anzari, "Analisis Perbandingan Model GRU dan LSTM untuk Prediksi Harga Saham Bank Rakyat Indonesia," *Jurnal Akademika*, vol. 17, no. 2, pp. 54–60, 2025, doi: 10.53564/akademika.v17i2.1692.
- [7] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "Stock Price Prediction Using LSTM, RNN and CNN-Sliding Window Model," 2017. doi: 10.1109/ICACCI.2017.8126078.
- [8] G. Budiprasetyo, M. Hani'ah, and D. Z. Aflah, "Prediksi Harga Saham Syariah Menggunakan Algoritma Long Short-Term Memory (LSTM)," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 8, no. 3, pp. 164–172, 2023, doi: 10.25077/teknosi.v8i3.2022.164-172.
- [9] B. S. Indonesia, "BSI Telah Masuk 5 Besar BUMN Kapitalisasi Pasar Terbesar," 2024, *Bank Syariah Indonesia*. [Online]. Available: <https://www.bankbsi.co.id>
- [10] F. I. Komputer and U. A. Purwokerto, "Comparative Study of LSTM and GRU Accuracy in Predicting BBRI Stock Closing Price," vol. 10, no. 1, pp. 837–846, 2026.
- [11] M. A. Majid, P. D. Saputri, and S. Soehardjoepri, "Stock Market Index Prediction using Bi-directional Long Short-Term Memory," *Journal of Applied Informatics and Computing*, vol. 8, no. 1, pp. 55–61, 2024, doi: 10.30871/jaic.v8i1.7195.
- [12] H. Wu, S. Chen, and Y. Ding, "Comparison of ARIMA and LSTM for Stock Price Prediction," *Financial Engineering and Risk Management*, vol. 6, no. 1, 2023, doi: 10.23977/ferm.2023.060101.
- [13] M. Oktaviani, G. Sibannang, I. G. L. Wijayakusuma, and U. Udayana, "Implementation of LSTM for Gold Price Prediction in Indonesia," vol. 10, no. 1, pp. 964–968, 2026.
- [14] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2015.
- [15] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [16] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," 2016. doi: 10.1145/2939672.2939785.
- [17] A. Graves and J. Schmidhuber, "Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures," *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, 2005, doi: 10.1016/j.neunet.2005.06.042.
- [18] M. Y. Andrika and M. Rahardi, "Comparative Study of Linear Regression, SVR, and XGBoost for Stock Price Prediction After a Stock Split," *Journal of Applied Informatics and Computing*, vol. 9, no. 4, pp. 1817–1824, 2025, doi: 10.30871/jaic.v9i4.10220.
- [19] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. OTexts, 2021. [Online]. Available: <https://otexts.com/fpp3>
- [20] T. Chai and R. R. Draxler, "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014, doi: 10.5194/gmd-7-1247-2014.