E-ISSN: 3025-0994
Halaman 7476-7490
Volume 4 Nomor 3 Tahun 2026

Perbandingan XGBoost dan LSTM untuk Prediksi Rentang Harga Penutupan Saham BBRI, ANTM dan ASII

Auliya Afifah Adnan Hakim[1], Adha Mashur Sajiah[2], Ilham Julian Efendi[3]

[1,2] Program Studi Teknik Informatika, Fakultas Teknik, Universitas Halu Oleo
[3] Program Studi Ilmu Komputer, Fakultas MIPA, Universitas Halu Oleo
e-mail: auliyaafifah24@gmail.com

Abstrak

Prediksi harga saham harian umumnya menghasilkan satu nilai titik sehingga tidak menggambarkan ketidakpastian pergerakan harga, padahal pelaku pasar membutuhkan gambaran rentang untuk mengelola risiko. Penelitian ini bertujuan membandingkan kinerja XGBoost dan Long Short-Term Memory (LSTM) dalam memprediksi titik tengah serta rentang harga penutupan harian saham BBRI, ASII, dan ANTM. Data historis 2020 sampai 2025 dibentuk menjadi empat belas fitur relatif, lalu dibagi secara kronologis menjadi data *training*, *validation*, dan *testing* agar informasi masa depan tidak bocor ke dalam proses pelatihan. Ketiga kuantil, yaitu kuantil bawah, tengah, dan atas, dipelajari secara terpisah melalui pendekatan regresi kuantil. Evaluasi dilakukan dengan akar galat kuadrat rata-rata, galat persentase absolut rata-rata, koefisien determinasi, ketepatan arah, cakupan rentang, lebar rentang rata-rata, *interval score*, waktu komputasi, dan ukuran model. Ketahanan model juga diuji pada kondisi *bullish*, *bearish*, dan *sideways*, kemudian diverifikasi melalui *forward test* 121 hari perdagangan tahun 2026 tanpa pelatihan ulang. Pada *historical test*, XGBoost memberikan RMSE terendah untuk seluruh saham dan unggul pada tujuh dari sembilan kombinasi kondisi pasar. Pada *forward test*, XGBoost tetap unggul untuk BBRI dan ASII, sedangkan LSTM sedikit lebih baik untuk ANTM serta menghasilkan *coverage* lebih tinggi pada ketiga saham. Hasil ini menunjukkan bahwa XGBoost lebih konsisten untuk prediksi titik, sementara LSTM lebih baik dalam mencakup harga aktual pada rentang prediksi forward.

Kata Kunci: XGBoost, LSTM, Harga Saham, Regresi Kuantil, Kondisi Pasar

Abstract

Daily stock price prediction usually produces a single point value that does not represent the uncertainty of price movement, whereas market participants need a price range to manage risk. This study aimed to compare XGBoost and Long Short-Term Memory (LSTM) in predicting the midpoint and the range of daily closing prices of BBRI, ASII, and ANTM. Historical data from 2020 to 2025 were transformed into fourteen relative features and then split chronologically into training, validation, and testing partitions so that future information did not leak into the training process. Three quantiles, namely the lower, median, and upper quantiles, were learned separately through a quantile regression approach. The evaluation

used root mean squared error, mean absolute percentage error, the coefficient of determination, directional accuracy, coverage, mean interval width, interval score, computation time, and model size. Model robustness was also examined under bullish, bearish, and sideways conditions, and was then verified through a 121-trading-day forward test in 2026 without retraining. In the historical test, XGBoost produces the lowest error for all stocks and outperforms LSTM in seven of nine market condition combinations. In the forward test, XGBoost remains superior for BBRI and ASII, whereas LSTM is slightly better for ANTM and yields higher coverage for all three stocks. These results indicate that XGBoost is more consistent for point prediction, while LSTM is more reliable in covering the actual prices within the forward prediction interval.

Keywords: *XGBoost, LSTM, Stock Price, Quantile Regression, Market Condition*

LATAR BELAKANG

Harga saham membentuk runtun waktu yang dipengaruhi tren, volatilitas, dan perubahan kondisi pasar. Pola tersebut membuat satu angka prediksi belum cukup untuk menjelaskan ketidakpastian harga berikutnya. Rentang prediksi memberi batas bawah, nilai tengah, dan batas atas yang dapat dibaca bersama-sama. *Quantile regression* menyediakan ketiga keluaran itu melalui q10, q50, dan q90 (Koenker & Bassett, 1978).

XGBoost membangun kumpulan pohon secara bertahap untuk memperbaiki kesalahan model sebelumnya dan memakai regularisasi untuk mengendalikan kompleksitas (Chen & Guestrin, 2016). LSTM memproses urutan melalui mekanisme gerbang sehingga dapat mempertahankan informasi temporal yang relevan (Hochreiter & Schmidhuber, 1997). Penelitian terdahulu menunjukkan hasil yang tidak selalu sama: keunggulan XGBoost atau LSTM berubah menurut saham, panjang data, fitur, serta rancangan pengujian (Pratama & Banowosari, 2024; Selayanti dkk., 2025; Aiyegbeni & Li, 2024; Li, 2023; Oukhouya dkk., 2024).

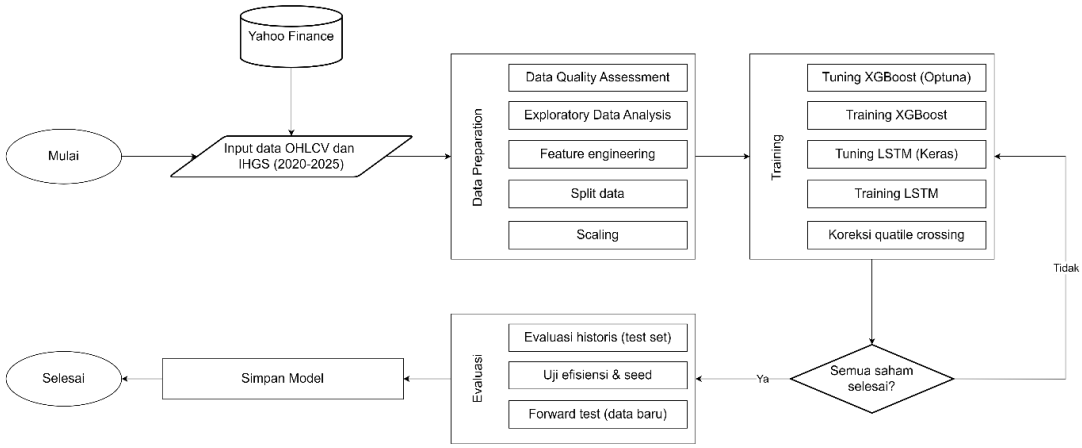
Penelitian ini memakai rancangan perbandingan yang sama untuk kedua algoritma. Keduanya menerima empat belas fitur, mempelajari target yang sama, dan diuji pada tanggal yang sama. Evaluasi tidak berhenti pada nilai kesalahan agregat, tetapi juga memeriksa rentang prediksi, kondisi *bullish*, *bearish*, dan *sideways*, biaya komputasi, serta *forward test* tanpa pelatihan ulang. Tujuannya adalah menentukan model yang lebih konsisten dalam memprediksi rentang harga penutupan BBRI, ASII, dan ANTM serta melihat ketahanannya ketika kondisi pasar berubah.

Pratama dan Banowosari membandingkan XGBoost dan LSTM pada saham BMRI, sedangkan Aiyegbeni dan Li membandingkan keduanya bersama ARIMA untuk prediksi saham Google (Pratama & Banowosari, 2024; Aiyegbeni & Li, 2024). Selayanti dkk. memakai model gabungan LSTM-XGBoost pada BBRI (Selayanti dkk., 2025). Penelitian lain juga menemukan bahwa hasil dapat berubah ketika ukuran data, pasar, dan skema *backtesting* berbeda (Li, 2023; Oukhouya dkk., 2024). Penelitian ini tidak membangun model gabungan; perbandingan dilakukan secara langsung agar sumber perbedaan kinerja tetap mudah dilacak.

Kontribusi penelitian terletak pada tiga bagian. Pertama, fitur harga absolut direkonstruksi menjadi bentuk relatif untuk mengurangi *target proxying*; pentingnya *feature engineering* dalam prediksi saham juga dilaporkan pada penelitian sebelumnya (Yun dkk.,

2021; Mostafavi & Hooman, 2025; Zeng dkk., 2023). Kedua, hasil agregat diuraikan menurut kondisi pasar sehingga kelemahan model tidak tertutup oleh satu nilai rata-rata (Liu dkk., 2023; Oksak dkk., 2025). Ketiga, model yang sama diuji kembali pada data tahun 2026 tanpa *tuning* maupun *training* ulang.

METODE



Gambar 1. *Pipeline* penelitian *machine learning*

Gambar 1 merangkum alur penelitian dari pengumpulan data sampai evaluasi. Tahap pertama mengambil data harian ketiga saham beserta IHSG, lalu membentuk empat belas fitur relatif dan target *return* penutupan D+1. Data dibagi secara kronologis menjadi *training*, *validation*, dan *testing*, dengan *scaler* yang hanya dihitung dari partisi *training*. Tahap berikutnya melatih tiga model kuantil terpisah, yaitu q10, q50, dan q90, untuk XGBoost maupun LSTM memakai konfigurasi hasil *tuning*. Keluaran kedua model dievaluasi pada *historical test*, diuraikan menurut kondisi pasar, kemudian diverifikasi melalui *forward test* tahun 2026 tanpa pelatihan ulang. Urutan yang sama dipakai untuk kedua algoritma sehingga perbedaan kinerja dapat ditelusuri pada algoritmanya, bukan pada perbedaan perlakuan data.

Data dan Rancangan Pengujian

Data harian BBRI, ASII, ANTM, dan IHSG diperoleh melalui yFinance dengan penyesuaian aksi korporasi (Yun dkk., 2021). Periode pembentukan model berlangsung dari 2 Januari 2020 sampai 30 Desember 2025. Setiap saham memiliki 1.447 observasi mentah dan tidak ditemukan nilai hilang maupun tanggal duplikat. Perhitungan indikator bergulir mengurangi data siap model menjadi 1.427 observasi.

Tabel 1. Pembagian data secara kronologis

Saham	Partisi	Awal	Akhir	N
BBRI	*Training*	2020-01-30	2023-08-10	849
BBRI	*Validation*	2023-08-11	2024-03-22	149
BBRI	*Testing*	2024-03-25	2025-12-29	429
ASII	*Training*	2020-01-30	2023-08-10	849
ASII	*Validation*	2023-08-11	2024-03-22	149
ASII	*Testing*	2024-03-25	2025-12-29	429
ANTM	*Training*	2020-01-30	2023-08-10	849
ANTM	*Validation*	2023-08-11	2024-03-22	149
ANTM	*Testing*	2024-03-25	2025-12-29	429

Pembagian data *training*, *validation*, dan *testing* secara kronologis disajikan pada Tabel 1. Susunan waktu dipertahankan agar informasi masa depan tidak masuk ke tahap *training*, sedangkan *scaler* hanya dihitung dari partisi *training*. *Forward test* memakai 121 hari perdagangan dari 5 Januari sampai 8 Juli 2026 dan menjalankan model tersimpan tanpa *refit*. Rancangan ini mengikuti prinsip evaluasi runtun waktu yang memisahkan data pengembangan dan data yang benar-benar belum dilihat model (Makridakis dkk., 2022; Kumbure dkk., 2022).

Pembagian acak tidak digunakan karena dapat menempatkan observasi yang lebih baru pada *training* dan observasi yang lebih lama pada *testing*. Susunan seperti itu menghasilkan evaluasi yang tidak sesuai dengan urutan penggunaan model. Partisi *validation* dipakai untuk pemilihan parameter dan penghentian *training*, sedangkan partisi *testing* baru dibuka setelah konfigurasi akhir ditetapkan.

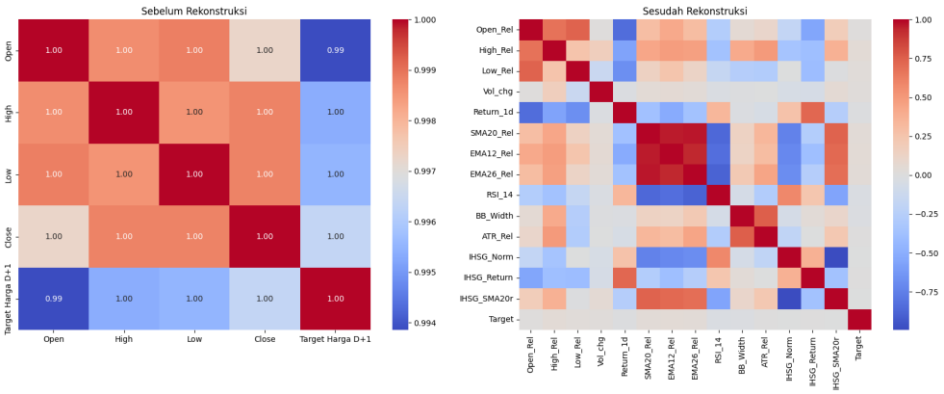
Kondisi pasar ditentukan dari *return* kumulatif 20 hari menggunakan ambang +3 persen untuk *bullish*, -3 persen untuk *bearish*, dan nilai di antaranya untuk *sideways*. Label dihitung hanya dari data yang tersedia sampai tanggal observasi, kemudian dipakai untuk mengelompokkan hasil *testing*. Label tersebut tidak menjadi target tambahan dan tidak mengubah model yang telah dilatih; evaluasi terpisah digunakan untuk melihat perubahan kinerja ketika rezim pasar berbeda (Liu dkk., 2023; Oksak dkk., 2025).

Pembentukan Fitur dan Target

Empat belas masukan yang digunakan ialah Open_Rel, High_Rel, Low_Rel, Vol_chg, *Return*_1d, SMA20_Rel, EMA12_Rel, EMA26_Rel, RSI_14, BB_Width, ATR_Rel, IHSG_Norm, IHSG_*Return*, dan IHSG_SMA20r. Variabel berskala harga diubah menjadi rasio relatif agar model tidak sekadar menyalin level harga hari ini. Target adalah *return* penutupan D+1, kemudian prediksi *return* dikembalikan ke satuan harga. Pemakaian indikator teknikal

dan transformasi fitur sejalan dengan temuan bahwa kualitas masukan berpengaruh kuat terhadap prediksi saham (Yun dkk., 2021; Mostafavi & Hooman, 2025; Zeng dkk., 2023; Lin dkk., 2021).

Open_Rel, High_Rel, dan Low_Rel menyatakan posisi harga pembukaan, tertinggi, dan terendah terhadap harga penutupan. *Return*_1d dan Vol_chg menangkap perubahan harian. SMA20_Rel, EMA12_Rel, dan EMA26_Rel mewakili kecenderungan tren, sedangkan RSI_14, BB_Width, dan ATR_Rel mewakili momentum serta volatilitas (Mostafavi & Hooman, 2025; Lin dkk., 2021). Tiga fitur IHSG memberi konteks pasar yang sama kepada setiap saham. Seluruh fitur dihitung dengan data historis sampai hari t untuk memprediksi penutupan hari perdagangan berikutnya.



Gambar 2. Korelasi fitur terhadap target sebelum dan sesudah rekonstruksi

Gambar 2 memperlihatkan bahwa korelasi hampir sempurna pada fitur harga absolut turun setelah rekonstruksi. Perubahan ini tidak berarti fitur menjadi tidak berguna. Model masih dapat mempelajari gabungan pola antarfitur, tetapi risiko *target proxying* dari level harga yang hampir sama menjadi lebih kecil.

Tuning dan Training

XGBoost dituning menggunakan Optuna dengan 150 *trial* dan TimeSeriesSplit lima *fold* (Akiba dkk., 2019). LSTM menggunakan *sequence* 30 hari, 64 unit, *dropout* 0,3, *batch size* 32, *optimizer* Adam dengan *learning rate* 0,001, dan maksimum 50 *epoch*. *Early stopping* mengembalikan bobot pada *validation loss* terendah. Setiap algoritma melatih tiga model kuantil terpisah untuk q10, q50, dan q90. Konfigurasi utama kedua model dirangkum pada Tabel 2.

Ruang pencarian XGBoost meliputi jumlah pohon, kedalaman, *learning rate*, subsample, colsample_bytree, serta parameter regularisasi. Setiap *trial* dievaluasi memakai *pinball loss* pada validasi silang berbasis waktu. Untuk LSTM, arsitektur dibuat tetap agar variasi yang diuji tidak terlalu besar dan prosesnya dapat dijelaskan kembali. Pencarian

parameter dan *feature engineering* perlu dinilai sebagai satu rangkaian karena keduanya dapat mengubah hasil model secara nyata (Yun dkk., 2021; Zeng dkk., 2023; Xu dkk., 2022).

Keadilan perbandingan dijaga pada data, fitur, target, kuantil, dan tanggal evaluasi. Namun, proses *tuning* kedua algoritma tidak identik karena bentuk modelnya berbeda. Optuna mencari parameter XGBoost, sedangkan LSTM memakai konfigurasi jaringan yang telah ditetapkan dan *early stopping*. Karena itu waktu *tuning* dibaca sebagai biaya konfigurasi penelitian, bukan peringkat umum kecepatan algoritma. Alur lengkap penelitian ditunjukkan pada Gambar 1.

Tabel 2. Konfigurasi utama model

Model	Komponen	Nilai
XGBoost	Kuantil	q10, q50, q90
XGBoost	Pencarian	Optuna, 150 *trial*
XGBoost	Validasi	TimeSeriesSplit, 5 *fold*
LSTM	*Sequence*	30 hari
LSTM	Arsitektur	64 unit, *dropout* 0,3
LSTM	*Training*	*batch* 32, maks. 50 *epoch*
LSTM	*Optimizer*	Adam, lr 0,001

Metrik Evaluasi

Prediksi titik q50 dinilai dengan MAE, RMSE, MAPE, R^2, dan *directional accuracy*. Rentang q10 sampai q90 dinilai menggunakan *coverage*, *mean width*, dan *interval score*. RMSE memberi penalti lebih besar pada kesalahan yang jauh dari harga aktual, sedangkan MAE dan MAPE memberi ukuran kesalahan absolut dan relatif (Karunasingha, 2022; Chicco dkk., 2021). *Coverage* nominal rentang q10-q90 adalah 80 persen, tetapi nilainya harus dibaca bersama lebar dan *interval score* (Gneiting & Raftery, 2007).

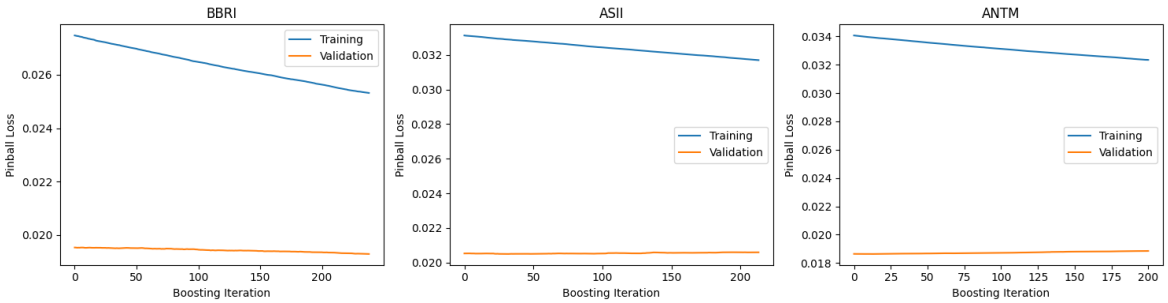
R^2 menunjukkan proporsi variasi harga pada periode uji yang dapat diikuti model dan perlu dibaca bersama metrik kesalahan lain (Chicco dkk., 2021). *Directional accuracy* membandingkan arah perubahan prediksi dengan arah perubahan aktual. Metrik arah dibaca terpisah dari RMSE karena prediksi yang dekat secara harga belum tentu benar dalam menebak arah harian (Aiyegbeni & Li, 2024; Yun dkk., 2021).

Coverage tinggi belum otomatis berarti rentang lebih baik karena rentang yang sangat lebar dapat mencakup lebih banyak harga aktual tetapi kurang informatif. *Mean width* mengukur lebar rata-rata, sedangkan *interval score* menambahkan penalti ketika nilai aktual berada di luar batas. Ketiga ukuran tersebut dibaca bersama mengikuti prinsip evaluasi prediksi probabilistik (Gneiting & Raftery, 2007).

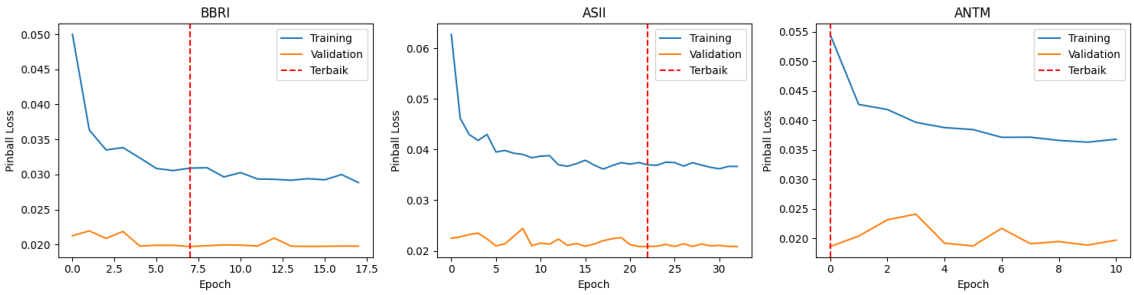
HASIL DAN PEMBAHASAN

Proses Training

Kurva *training* dipakai untuk memeriksa perubahan *pinball loss* pada data *training* dan *validation*. Perubahan *loss* XGBoost ditunjukkan pada Gambar 3, sedangkan riwayat LSTM ditunjukkan pada Gambar 4. *Early stopping* menghasilkan *epoch* terbaik yang berbeda antar saham, sehingga bobot akhir diambil dari *validation loss* terendah dan bukan selalu dari *epoch* terakhir.



Gambar 3. *Learning curve* XGBoost untuk model q50



Gambar 4. *Learning curve* LSTM untuk model q50

Berdasarkan Gambar 3 dan Gambar 4, *loss training* kedua model menurun selama proses pembelajaran, sedangkan *validation loss* LSTM lebih berfluktuasi. Kurva tersebut hanya menunjukkan proses optimasi; kemampuan generalisasi tetap ditentukan menggunakan *testing* dan *forward test* yang tidak dipakai untuk mengatur bobot model (Makridakis dkk., 2022; Kumbure dkk., 2022).

Pada XGBoost, selisih *loss training* dan *validation* tidak membesar secara tiba-tiba. Pada LSTM, *validation loss* mencapai titik terbaik pada *epoch* yang berbeda untuk setiap saham. Pola tersebut mendukung penggunaan *early stopping* karena lama *training* yang sama tidak otomatis sesuai untuk seluruh data.

Hasil Historical Test

Tabel 3. Perbandingan *historical test* dan *baseline* naif

Saham	Model	RMSE	*Baseline*	Selisih (%)	DA (%)
BBRI	XGBoost	78.59	78.04	-0.72	46.85
BBRI	LSTM	79.15	78.04	-1.43	47.32
ASII	XGBoost	82.37	82.38	0.02	45.45
ASII	LSTM	83.79	82.38	-1.71	42.19
ANTM	XGBoost	63.67	63.85	0.28	45.69
ANTM	LSTM	64.81	63.85	-1.50	46.62

Berdasarkan Tabel 3, XGBoost menghasilkan RMSE lebih rendah daripada LSTM pada ketiga saham. Namun, perbaikannya terhadap *baseline* naif yang menetapkan harga besok sama dengan harga hari ini hanya muncul secara tipis pada ASII dan ANTM, sedangkan BBRI masih sedikit lebih buruk. Temuan ini membatasi klaim penelitian: model mempelajari pola, tetapi belum memberikan peningkatan besar atas persistensi harga harian. *Baseline* sederhana penting dipertahankan agar model yang kompleks tidak dinilai hanya dari kecocokannya terhadap level harga (Makridakis dkk., 2022; Kumbure dkk., 2022).

Selisih RMSE XGBoost dan LSTM relatif kecil pada seluruh saham. Nilai R^2 keduanya tinggi karena level harga pada periode *testing* dapat diikuti dengan baik, tetapi hasil tersebut tidak bertentangan dengan *baseline*. Harga harian mempunyai persistensi kuat sehingga model dapat mengikuti level harga tanpa selalu menghasilkan perbaikan besar untuk prediksi satu hari ke depan (Chicco dkk., 2021; Makridakis dkk., 2022).

Directional accuracy kedua model masih berada di bawah 50 persen. Hasil ini tidak menunjukkan keunggulan arah yang stabil, sehingga keluaran tidak ditafsirkan sebagai sinyal beli atau jual. Keterbatasan arah harian juga dilaporkan pada perbandingan algoritma prediksi saham lain (Aiyegbeni & Li, 2024; Yun dkk., 2021).

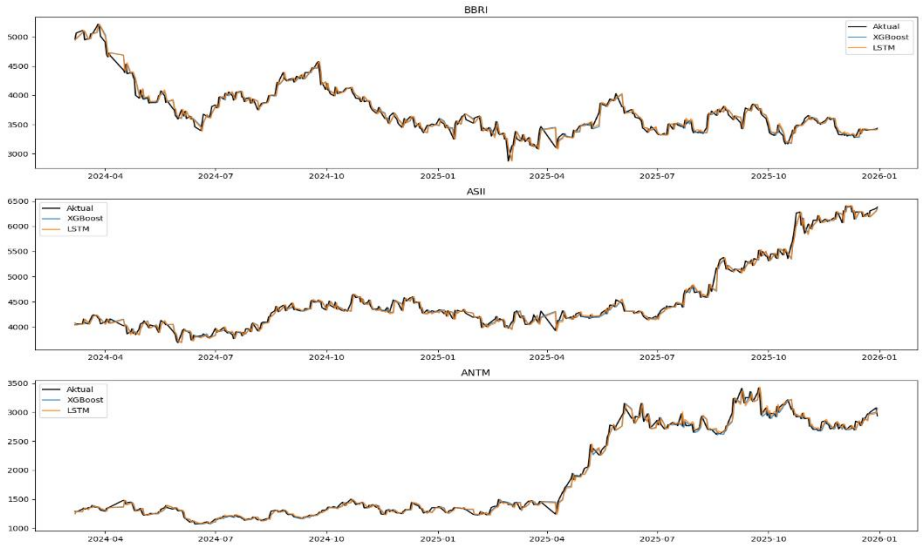
Tabel 4. Evaluasi rentang pada *historical test*

Saham	Model	*Coverage* (%)	Lebar	*Interval score*
BBRI	XGBoost	69.00	146.47	297.74
BBRI	LSTM	73.43	159.86	290.01
ASII	XGBoost	75.99	177.01	293.13
ASII	LSTM	74.59	181.54	292.87

Saham	Model	*Coverage* (%)	Lebar	*Interval score*
ANTM	XGBoost	76.46	126.40	205.55
ANTM	LSTM	75.06	119.16	209.36

Berdasarkan Tabel 4, model dengan *coverage* lebih tinggi berbeda menurut saham. LSTM memperoleh *interval score* lebih rendah pada BBRI dan ASII, sedangkan XGBoost lebih baik pada ANTM. Seluruh *coverage* masih berada di bawah nilai nominal 80 persen, sehingga kualitas rentang harus dibaca dari gabungan *coverage*, *mean width*, dan *interval score* (Gneiting & Raftery, 2007).

Perbedaan *coverage* antarmodel relatif kecil, sementara lebar rentangnya tidak selalu bergerak searah dengan *coverage*. Temuan ini menunjukkan bahwa tidak ada satu model yang menang pada seluruh ukuran rentang. Pergerakan harga aktual dan prediksi q50 sepanjang *historical test* ditunjukkan pada Gambar 5.



Gambar 5. Harga aktual dan prediksi q50 pada *historical test*

Ketahanan pada Kondisi Pasar

Tabel 5. Kinerja model pada kondisi pasar

Saham	Kondisi	Model	RMSE	DA (%)	N
BBRI	*Bullish*	XGBoost	72.35	46.72	137
BBRI	*Bullish*	LSTM	71.33	48.91	137

Saham	Kondisi	Model	RMSE	DA (%)	N
BBRI	*Bearish*	XGBoost	80.21	50.26	193
BBRI	*Bearish*	LSTM	81.86	47.67	193
BBRI	*Sideways*	XGBoost	83.56	40.40	99
BBRI	*Sideways*	LSTM	83.94	44.44	99
ASII	*Bullish*	XGBoost	94.49	48.17	164
ASII	*Bullish*	LSTM	95.00	40.24	164
ASII	*Bearish*	XGBoost	78.73	35.21	71
ASII	*Bearish*	LSTM	84.50	33.80	71
ASII	*Sideways*	XGBoost	72.02	46.91	194
ASII	*Sideways*	LSTM	72.67	46.91	194
ANTM	*Bullish*	XGBoost	75.03	48.45	194
ANTM	*Bullish*	LSTM	77.42	41.75	194
ANTM	*Bearish*	XGBoost	48.25	41.96	143
ANTM	*Bearish*	LSTM	47.01	48.95	143
ANTM	*Sideways*	XGBoost	58.43	45.65	92
ANTM	*Sideways*	LSTM	59.25	53.26	92

Berdasarkan Tabel 5, XGBoost memiliki RMSE terendah pada tujuh dari sembilan kombinasi saham dan kondisi pasar. LSTM hanya unggul pada BBRI *bullish* dan ANTM *bearish*. Hasil ini menunjukkan bahwa pemenang agregat tidak selalu menang pada setiap kondisi, sedangkan *directional accuracy* yang berkisar di sekitar tebakan acak tidak mendukung penggunaan model sebagai penentu arah tunggal (Aiyegbeni & Li, 2024; Oksak dkk., 2025).

Besarnya error berubah menurut saham dan kondisi. Jumlah observasi setiap kondisi juga tidak seimbang, sehingga hasil per kelompok bersifat deskriptif dan tidak dapat dianggap sebagai peringkat pasti untuk seluruh periode mendatang. Temuan yang berubah antarrezim sejalan dengan penelitian yang menunjukkan bahwa karakter volatilitas memengaruhi kinerja model dan manfaat *feature engineering* (Raudys & Goldstein, 2022; Oksak dkk., 2025).

Secara umum, XGBoost lebih stabil karena hanya kalah pada dua kondisi, sedangkan LSTM tetap relevan ketika pola sekuensial tertentu muncul. Temuan ini menjawab pertanyaan ketahanan model: hasil agregat XGBoost lebih konsisten, tetapi pergantian kondisi dapat mengubah model dengan RMSE terendah.

Waktu Komputasi dan Ukuran Model

Tabel 6. Waktu komputasi dan ukuran model

Saham	Model	*Tuning* (s)	*Training* (s)	Inferensi (ms)	Ukuran (MB)
BBRI	XGBoost	866.1	1.8	0.013	3.08
BBRI	LSTM	42.9	20.3	0.823	0.78
ASII	XGBoost	787.2	1.3	0.008	2.97
ASII	LSTM	55.2	23.7	0.819	0.78
ANTM	XGBoost	456.2	0.6	0.007	1.18
ANTM	LSTM	38.0	14.5	0.824	0.78

Berdasarkan Tabel 6, *tuning* XGBoost membutuhkan waktu lebih lama karena menjalankan 150 *trial*, sedangkan *training* final dan inferensinya lebih singkat. LSTM mempunyai ukuran berkas lebih kecil, tetapi inferensinya lebih lambat. Angka waktu tidak diperlakukan sebagai sifat mutlak algoritma karena bergantung pada perangkat keras, implementasi, dan ruang pencarian (Akiba dkk., 2019; Zeng dkk., 2023).

Kedua model tetap cukup cepat untuk prediksi harian. Perbedaan waktu baru menjadi penting ketika jumlah saham atau frekuensi prediksi meningkat. Ukuran XGBoost berubah antar saham karena jumlah dan struktur pohon hasil *tuning* berbeda, sementara arsitektur LSTM dibuat sama.

Forward Test

Tabel 7. Evaluasi *forward test* tahun 2026

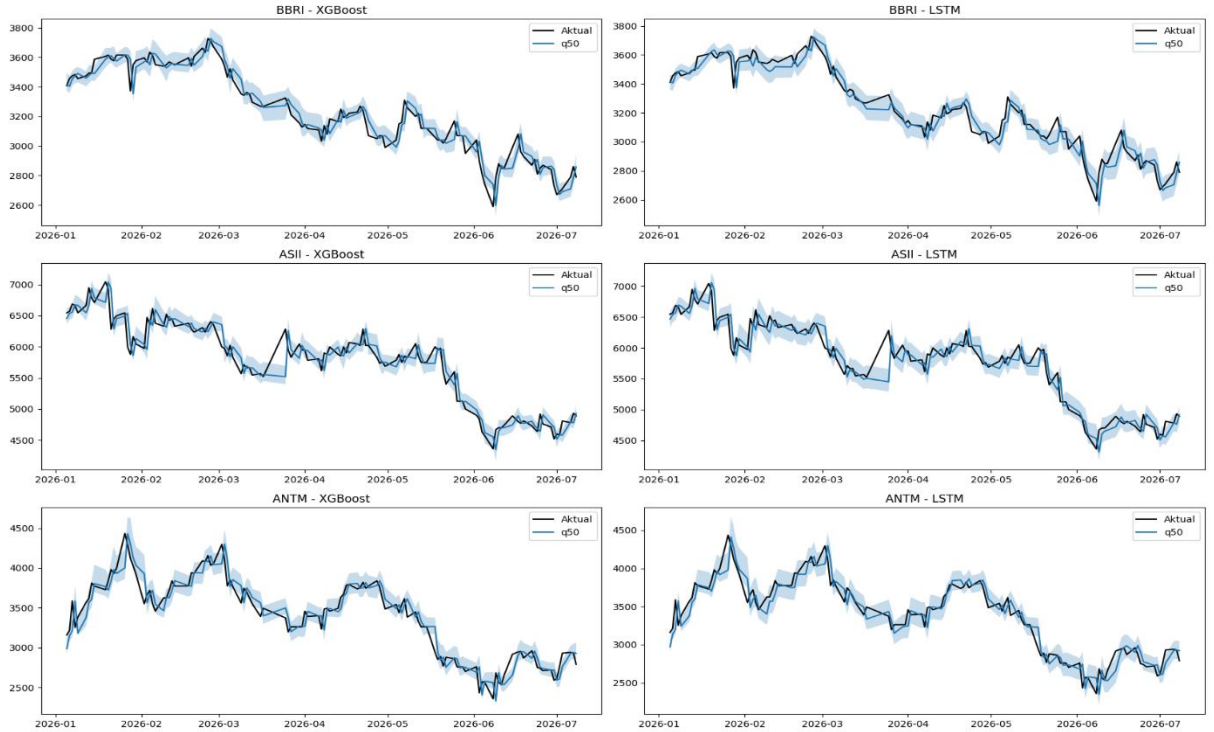
Saham	Model	RMSE	MAPE (%)	R^2	DA (%)	*Coverage* (%)	*Interval score*
BBRI	XGBoost	68.19	1.64	0.943	49.59	66.12	270.25
BBRI	LSTM	70.86	1.71	0.939	52.89	71.90	253.04
ASII	XGBoost	186.83	2.31	0.918	53.72	61.16	747.74
ASII	LSTM	190.30	2.34	0.915	50.41	69.42	710.52
ANTM	XGBoost	142.64	3.25	0.913	46.28	64.46	550.62
ANTM	LSTM	140.74	3.19	0.915	48.76	70.25	513.79

Berdasarkan Tabel 7, XGBoost mempertahankan RMSE lebih rendah untuk BBRI dan ASII, sedangkan LSTM sedikit lebih baik pada ANTM. Berbeda dari evaluasi titik, LSTM memperoleh *coverage* dan *interval score* lebih baik pada ketiga saham. Dengan demikian, pemilihan model bergantung pada tujuan: XGBoost lebih konsisten untuk titik tengah, sementara LSTM lebih baik ketika prioritasnya adalah cakupan rentang pada data baru.

Perubahan error dari *historical test* ke *forward test* paling terlihat pada ASII dan ANTM. Hasil ini menunjukkan bahwa kinerja periode pengembangan tidak langsung berpindah ke data 2026. Evaluasi pada periode berikutnya diperlukan untuk mengungkap perubahan distribusi yang tidak terlihat pada satu partisi *testing* (Makridakis dkk., 2022; Kumbure dkk., 2022).

Interval score LSTM lebih rendah pada seluruh saham forward, tetapi keunggulan tersebut disertai rentang yang umumnya lebih lebar. Karena itu LSTM lebih menarik ketika cakupan menjadi prioritas, sedangkan XGBoost tetap lebih konsisten untuk estimasi titik pada dua dari tiga saham.

Berdasarkan Gambar 6, kedua rentang mengikuti arah harga secara umum, tetapi harga aktual lebih sering keluar dari batas pada periode forward. Penurunan *coverage* menunjukkan perubahan pola antara data 2020-2025 dan kondisi pasar 2026, sehingga evaluasi satu periode saja belum cukup (Oksak dkk., 2025; Kumbure dkk., 2022).



Gambar 6. Rentang prediksi pada *forward test*

Pembahasan

Hasil penelitian memberi jawaban yang berbeda menurut jenis keluaran. Untuk q50, XGBoost lebih konsisten karena unggul pada seluruh *historical test* dan dua dari tiga *forward test*. Untuk rentang q10 sampai q90, LSTM lebih baik pada *forward test* karena *coverage* lebih tinggi dan *interval score* lebih rendah. Karena itu istilah model terbaik harus selalu disertai konteks keluaran yang dibandingkan (Pratama & Banowosari, 2024; Selayanti dkk., 2025; Aiyegbeni & Li, 2024; Li, 2023; Oukhouya dkk., 2024).

Keunggulan XGBoost pada prediksi titik dapat dikaitkan dengan bentuk masukan yang telah merangkum tren, momentum, volatilitas, dan konteks IHSG pada satu baris. Struktur seperti ini sesuai dengan model pohon yang mempelajari interaksi nonlinier antarfitur (Chen & Guestrin, 2016; Yun dkk., 2021; Mostafavi & Hooman, 2025). LSTM menerima informasi yang sama dalam urutan 30 hari, tetapi jumlah *sequence training* tetap terbatas dibandingkan kapasitas model sekuensial (Hochreiter & Schmidhuber, 1997; Li, 2023).

LSTM menghasilkan rentang forward yang lebih mampu mencakup harga aktual. Urutan 30 hari dapat membantu model menangkap perubahan volatilitas yang tidak seluruhnya tampak pada satu baris fitur, sebagaimana dilaporkan pada penelitian LSTM dan XGBoost untuk data finansial lain (Oukhouya dkk., 2024; Raudys & Goldstein, 2022; Sathiyapriya dkk., 2023). Namun, *coverage* masih berada di bawah target nominal 80 persen sehingga rentang belum dapat disebut terkalibrasi sempurna. Penelitian ini tidak menerapkan kalibrasi setelah *training*.

Baseline naif menjadi pemeriksaan penting karena harga penutupan satu hari ke depan sangat dekat dengan harga hari ini. Tanpa *baseline*, R^2 yang tinggi dapat terlihat sangat meyakinkan meskipun peningkatan RMSE model kompleks sangat kecil. Karena itu kontribusi model lebih tepat dibaca bersama kemampuan membentuk rentang dan menjaga kinerja pada beberapa kondisi (Chicco dkk., 2021; Makridakis dkk., 2022).

Perbedaan hasil historical dan *forward test* menunjukkan risiko memilih model dari satu periode. Peningkatan error pada ASII dan ANTM menandakan adanya pola baru yang tidak sepenuhnya terwakili dalam data *training*. Penyimpanan model dan pelaksanaan forward tanpa *refit* membuat perubahan kinerja tersebut terlihat apa adanya (Oksak dkk., 2025; Kumbure dkk., 2022).

Secara praktis, prediksi q50 memberikan perkiraan pusat, sedangkan q10 dan q90 memberi batas ketidakpastian empiris. Ketika rentang melebar, ketidakpastian model juga meningkat. Karena hasil tidak memuat biaya transaksi, likuiditas, maupun batas risiko investor, keluaran tidak dapat langsung diterjemahkan menjadi keputusan beli atau jual (Koenker & Bassett, 1978; Gneiting & Raftery, 2007).

SIMPULAN

Dengan rancangan pengujian yang sama, XGBoost lebih konsisten dalam memprediksi titik tengah harga penutupan. Model ini memiliki RMSE historis terendah pada BBRI, ASII, dan ANTM serta unggul pada tujuh dari sembilan kombinasi kondisi pasar. Pada *forward test*, XGBoost tetap terbaik untuk BBRI dan ASII, sedangkan LSTM lebih baik untuk ANTM. Untuk rentang prediksi, LSTM menghasilkan *coverage* dan *interval score* forward yang lebih baik pada ketiga saham. Keunggulan terhadap *baseline* naif tetap kecil, sehingga hasil penelitian tidak mendukung penggunaan prediksi sebagai satu-satunya dasar keputusan

investasi. Penelitian berikutnya perlu menguji periode forward yang lebih panjang, saham yang lebih beragam, dan fitur eksternal yang benar-benar tersedia sebelum tanggal prediksi.

DAFTAR PUSTAKA

Aiyegbeni, G. & Li, Y. 2024. A Comparative Analysis Of LSTM, ARIMA, Xgboost Algorithms In Predicting Stock Price Direction. Engineering And Technology Journal. Vol. 9, No. 8, Hlm. 4978-4986. Https://Doi.Org/10.47191/Etj/V9i08.50

Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. 2019. Optuna: A Next-Generation Hyperparameter Optimization Framework. Proceedings Of The 25th ACM SIGKDD International Conference On Knowledge Discovery And Data Mining. Hlm. 2623-2631. Https://Doi.Org/10.1145/3292500.3330701

Chen, T. & Guestrin, C. 2016. Xgboost: A Scalable Tree Boosting System. Proceedings Of The 22nd ACM SIGKDD International Conference On Knowledge Discovery And Data Mining. Hlm. 785-794. Https://Doi.Org/10.1145/2939672.2939785

Chicco, D., Warrens, M. J. & Jurman, G. 2021. The Coefficient Of Determination R-Squared Is More Informative Than SMAPE, MAE, MAPE, MSE And RMSE In Regression Analysis Evaluation. Peerj Computer Science. Vol. 7, Art. E623. Https://Doi.Org/10.7717/Peerj-Cs.623

Gneiting, T. & Raftery, A. E. 2007. Strictly Proper Scoring Rules, Prediction, And Estimation. Journal Of The American Statistical Association. Vol. 102, No. 477, Hlm. 359-378. Https://Doi.Org/10.1198/016214506000001437

Hochreiter, S. & Schmidhuber, J. 1997. Long Short-Term Memory. Neural Computation. Vol. 9, No. 8, Hlm. 1735-1780. Https://Doi.Org/10.1162/Neco.1997.9.8.1735

Karunasingha, D. S. K. 2022. Root Mean Square Error Or Mean Absolute Error? Use Their Ratio As Well. Information Sciences. Vol. 585, Hlm. 609-629. Https://Doi.Org/10.1016/J.Ins.2021.11.036

Koenker, R. & Bassett, G. 1978. Regression Quantiles. Econometrica. Vol. 46, No. 1, Hlm. 33-50. Https://Doi.Org/10.2307/1913643

Kumbure, M. M., Lohrmann, C., Luukka, P. & Porras, J. 2022. Machine Learning Techniques And Data For Stock Market Forecasting: A Literature Review. Expert Systems With Applications. Vol. 197, Art. 116659. Https://Doi.Org/10.1016/J.Eswa.2022.116659

Li, Z. 2023. Comparison Of Xgboost And LSTM Models For Stock Price Prediction. Advances In Economics, Management And Political Sciences. Vol. 61, Hlm. 147-155. Https://Doi.Org/10.54254/2754-1169/61/20231181

Lin, Y. H., Liu, S. C., Yang, H. J. & Wu, H. R. 2021. Stock Trend Prediction Using Candlestick Charting And Ensemble Machine Learning Techniques With A Novelty Feature Engineering Scheme. IEEE Access. Vol. 9, Hlm. 101433-101446. Https://Doi.Org/10.1109/ACCESS.2021.3096825

Liu, F., Umair, M. & Gao, J. 2023. Assessing Oil Price Volatility Co-Movement With Stock Market Volatility Through Quantile Regression Approach. Resources Policy. Vol. 81, Art. 103375. Https://Doi.Org/10.1016/J.Resourpol.2023.103375

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. 2022. M5 Accuracy Competition: Results, Findings, And Conclusions. International Journal Of Forecasting. Vol. 38, No. 4, Hlm. 1346-1364. Https://Doi.Org/10.1016/J.Ijforecast.2021.11.013

Mostafavi, S. M. & Hooman, A. R. 2025. Key Technical Indicators For Stock Market Prediction. Machine Learning With Applications. Vol. 20, Art. 100631. Https://Doi.Org/10.1016/J.Mlwa.2025.100631

Oksak, Y., Buyukkor, Y. & Saritas, T. 2025. Wavelet-Enhanced Multimodel Framework For Stock Market Forecasting: A Comprehensive Analysis Across Market Regimes. Borsa Istanbul Review. Art. 100771. Https://Doi.Org/10.1016/J.Bir.2025.100771

Oukhouya, H., Kadiri, H., El Himdi, K. & Guerbaz, R. 2024. Forecasting International Stock Market Trends: Xgboost, LSTM, LSTM-Xgboost, And Backtesting Xgboost Models. Statistics, Optimization And Information Computing. Vol. 12, No. 1, Hlm. 200-209. Https://Doi.Org/10.19139/Soic-2310-5070-1822

Pratama, B. & Banowosari, L. Y. 2024. Perbandingan Metode Extreme Gradient Boosting (Xgboost) Dengan Long Short-Term Memory (LSTM) Untuk Prediksi Saham PT Bank Mandiri Tbk. (BMRI). Journal Of Economic, Business And Accounting. Vol. 7, No. 3, Hlm. 5631-5636. Https://Doi.Org/10.31539/Costing.V7i3.9473

Raudys, A. & Goldstein, E. 2022. Forecasting Detrended Volatility Risk And Financial Price Series Using LSTM Neural Networks And Xgboost Regressor. Journal Of Risk And Financial Management. Vol. 15, No. 12, Art. 602. Https://Doi.Org/10.3390/Jrfm15120602

Sathiyapriya, K., Vankadara, S., Babu, K. S. & Muralidharan, M. 2023. Performance Comparison Of LSTM And Xgboost For Ether Price Prediction From Spam-Filtered Tweets. Proceedings Of The International Conference On Intelligent Systems For Communication, Iot And Security. Hlm. 650-655. Https://Doi.Org/10.1109/Iciscois56541.2023.10100425

Selayanti, N., Putri, D. A., Trimono & Idhom, M. 2025. Prediksi Harga Penutupan Saham BBRI Dengan Model Hybrid LSTM-Xgboost. Informatika: Jurnal Teknik Informatika Dan Multimedia. Vol. 5, No. 1, Hlm. 52-64. Https://Doi.Org/10.51903/Informatika.V5i1.1011

Xu, J., He, J., Gu, J., Wu, H., Wang, L., Zhu, Y., Wang, T., He, X. & Zhou, Z. 2022. Financial Time Series Prediction Based On Xgboost And Generative Adversarial Networks. International Journal Of Circuits, Systems And Signal Processing. Vol. 16, Hlm. 637-645. Https://Doi.Org/10.46300/9106.2022.16.79

Yun, K. K., Yoon, S. W. & Won, D. 2021. Prediction Of Stock Price Direction Using A Hybrid GA-Xgboost Algorithm With A Three-Stage Feature Engineering Process. Expert Systems With Applications. Vol. 186, Art. 115716. Https://Doi.Org/10.1016/J.Eswa.2021.115716

Zeng, X., Cai, J., Liang, C. & Yuan, C. 2023. Prediction Of Stock Price Movement Using An Improved NSGA-II-RF Algorithm With A Three-Stage Feature Engineering Process. PLOS ONE. Vol. 18, No. 6, Art. E0287754. Https://Doi.Org/10.1371/Journal.Pone.0287754