



PERBANDINGAN KINERJA LSTM, RANDOM FOREST, DAN XGBOOST DALAM MEMPREDIKSI HARGA PENUTUPAN INDEKS HARGA SAHAM GABUNGAN (IHSG) BERBASIS DATA HISTORIS

Mohammad Faiz Rakhman¹⁾, Cendra Devayana Putra²⁾

¹⁾ Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia

Email: mohammadfaiz.22113@mhs.unesa.ac.id

²⁾ Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia

Email: putracendra@mhs.unesa.ac.id

Abstract

This study compares the performance of three machine learning algorithms, Long Short-Term Memory (LSTM), Random Forest, and Extreme Gradient Boosting (XGBoost), in predicting the next-day closing price of Indonesia's Composite Stock Price Index (IHSG) as a baseline before further feature engineering is applied in a broader ongoing study. Daily historical price data (Open, High, Low, Close, Volume) covering January 2015 to early 2026 were collected from Yahoo Finance. Two feature representations were compared: the raw 5-dimensional OHLCV attributes, and a 64-dimensional temporal representation extracted from the same OHLCV data using a Bidirectional LSTM (Bi-LSTM) encoder. Each representation was evaluated using LSTM, Random Forest, and XGBoost, with performance measured by Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) on a chronological 80:20 train-test split. The results show that the raw OHLCV representation combined with LSTM achieved the best performance (RMSE = 196.11, MAE = 164.14, MAPE = 2.18%), outperforming Random Forest and XGBoost on the same representation (MAPE 3.78% and 3.80%). Encoding the OHLCV data into a 64-dimensional Bi-LSTM representation without any external signal reduced accuracy across all three algorithms, most notably for LSTM (MAPE rising to 7.13%), indicating that unsupervised temporal encoding discards useful absolute price information when no additional predictive feature is introduced. These findings establish a validated baseline configuration and evaluation pipeline for IHSG closing-price prediction, intended as the foundation for a subsequent study that integrates external textual features into the same experimental framework.

Keywords: IHSG; Stock Price Prediction; LSTM; Random Forest; XGBoost.

Abstrak

Penelitian ini membandingkan kinerja tiga algoritma machine learning, yaitu Long Short-Term Memory (LSTM), Random Forest, dan Extreme Gradient Boosting (XGBoost), dalam memprediksi harga penutupan Indeks Harga Saham Gabungan (IHSG) satu hari ke depan sebagai model dasar (baseline) sebelum rekayasa fitur lebih lanjut diterapkan pada penelitian yang lebih luas yang sedang berjalan. Data historis harian (Open, High, Low, Close, Volume) periode Januari 2015 hingga awal 2026 dikumpulkan dari Yahoo Finance. Dua representasi fitur dibandingkan, yaitu 5 atribut OHLCV mentah dan representasi temporal 64 dimensi hasil ekstraksi menggunakan encoder Bidirectional LSTM (Bi-LSTM) dari data OHLCV yang sama. Setiap representasi diuji menggunakan LSTM, Random Forest, dan XGBoost, dengan evaluasi kinerja memakai Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan Mean Absolute Percentage Error (MAPE) pada skema pembagian data kronologis 80:20. Hasil penelitian menunjukkan bahwa representasi OHLCV mentah yang dipadukan dengan LSTM menghasilkan kinerja terbaik (RMSE = 196,11, MAE = 164,14, MAPE = 2,18%), mengungguli Random Forest dan XGBoost pada representasi yang sama (MAPE 3,78% dan 3,80%). Reduksi dimensi data OHLCV menjadi representasi 64 dimensi Bi-LSTM tanpa sinyal eksternal justru menurunkan akurasi pada ketiga algoritma, terutama pada LSTM (MAPE naik menjadi 7,13%), yang mengindikasikan bahwa proses encoding temporal tanpa label (unsupervised) menghilangkan sebagian informasi harga absolut yang bermanfaat apabila belum disertai fitur prediktif tambahan. Temuan ini menetapkan konfigurasi baseline dan alur evaluasi yang tervalidasi untuk prediksi harga penutupan IHSG, yang menjadi fondasi bagi penelitian lanjutan yang mengintegrasikan fitur tekstual eksternal ke dalam kerangka eksperimen yang sama.

Kata Kunci: IHSG; Prediksi Harga Saham; LSTM; Random Forest; XGBoost.



PENDAHULUAN

Pasar modal merupakan salah satu indikator penting dalam mengukur kinerja perekonomian suatu negara, di mana Indeks Harga Saham Gabungan (IHSG) menjadi indikator utama yang mencerminkan kinerja pasar modal Indonesia (P3BMS UMA, 2025). Jumlah investor pasar modal Indonesia tercatat mencapai 20,32 juta Single Investor Identification (SID) per akhir Desember 2025, tumbuh 37 persen dibandingkan tahun sebelumnya (Heriyanto, 2025). Pertumbuhan ini turut mendorong kebutuhan akan model prediksi harga saham yang lebih akurat dan berbasis data.

Berbagai penelitian terdahulu telah mengkaji prediksi pergerakan harga saham IHSG menggunakan pendekatan machine learning, di antaranya perbandingan Nonlinear Auto-Regressive Exogenous Model (NARX) Neural Network dengan Long Short-Term Memory (LSTM) (hamzah & Hartono, 2023), serta pengujian Random Forest dan XGBoost pada peramalan harga saham dengan hasil yang bervariasi bergantung pada karakteristik data (Edi Ismanto et al., 2026). LSTM dipilih karena kemampuannya menangkap dependensi temporal jangka panjang pada data deret waktu (Hochreiter & Schmidhuber, 1997), sedangkan Random Forest (Breiman, 2001) dan XGBoost (C. CHEN, 2016) dipilih sebagai pembanding berbasis pohon keputusan yang telah terbukti kompetitif pada berbagai kasus peramalan harga saham (Jange, 2022).

Penelitian lain juga telah mengeksplorasi arsitektur yang lebih kompleks untuk prediksi IHSG, seperti kombinasi CNN-CA-LSTM yang diintegrasikan dengan aplikasi web (Indra Jaya et al., 2026) dan pemanfaatan LSTM berbasis data historis untuk memprediksi harga penutupan (Farhan Fadhila et al., 2025). Selain pendekatan berbasis data numerik historis, sejumlah studi turut mengeksplorasi sumber data non-numerik seperti sentimen media sosial sebagai variabel tambahan (Christophorus Bintang Saputra & Koesrindartoto, 2024). Keberagaman pendekatan ini menunjukkan bahwa belum terdapat konsensus tunggal mengenai representasi fitur dan algoritma yang paling optimal untuk kasus prediksi IHSG, sehingga pengujian baseline yang sistematis dan terukur tetap diperlukan sebagai titik pijak sebelum menambahkan kompleksitas fitur lebih lanjut.

Berdasarkan uraian tersebut, rumusan masalah pada penelitian tahap ini adalah bagaimana perbandingan kinerja algoritma LSTM, Random Forest, dan XGBoost dalam memprediksi harga penutupan IHSG pada representasi data OHLCV mentah dibandingkan dengan representasi 64 dimensi hasil ekstraksi Bi-LSTM. Tujuan penelitian ini adalah membangun dan mengevaluasi konfigurasi baseline yang optimal, sebagai dasar metodologis bagi tahap penelitian lanjutan yang akan mengintegrasikan sumber data eksternal ke dalam kerangka eksperimen yang sama.

Sebelum menambahkan fitur eksternal apa pun ke dalam model, diperlukan sebuah konfigurasi dasar (baseline) yang teruji secara menyeluruh, baik dari sisi representasi fitur maupun performa algoritma, sebagai titik acuan pembanding. Artikel ini menyajikan bagian awal dari penelitian yang lebih besar, dengan berfokus pada

perbandingan kinerja tiga algoritma, yaitu LSTM, Random Forest, dan XGBoost, dalam memprediksi harga penutupan IHSG menggunakan dua representasi data historis OHLCV: representasi mentah (5 fitur) dan representasi 64 dimensi hasil ekstraksi Bi-LSTM. Hasil dari tahap baseline ini akan menjadi acuan pembanding bagi tahap penelitian selanjutnya yang mengintegrasikan sumber data eksternal ke dalam kerangka eksperimen yang sama.

Penelitian tahap ini dibatasi pada objek Indeks Harga Saham Gabungan (kode indeks JKSE) dengan data historis harian periode 1 Januari 2015 hingga awal 2026 yang bersumber dari Yahoo Finance. Variabel target yang diprediksi dibatasi pada harga penutupan (Close) untuk periode satu hari ke depan (short-term prediction), dengan algoritma yang diuji dibatasi pada LSTM, Random Forest, dan XGBoost, serta evaluasi menggunakan metrik RMSE, MAE, dan MAPE.

LANDASAN TEORI

2.1 Indeks Harga Saham Gabungan (IHSG)

IHSG merupakan indeks yang menggambarkan pergerakan harga seluruh saham yang tercatat di Bursa Efek Indonesia (BEI) dan menjadi acuan utama untuk menilai kondisi pasar modal secara agregat. Pergerakan IHSG dipengaruhi oleh berbagai faktor, baik internal perusahaan maupun eksternal seperti kondisi makroekonomi dan kebijakan pemerintah (Aditya & Arafat, 2025), sehingga karakteristik data historisnya bersifat non-linear dan fluktuatif sebagaimana umumnya data deret waktu keuangan.

2.2 Long Short-Term Memory (LSTM)

LSTM adalah pengembangan dari Recurrent Neural Network (RNN) yang dirancang untuk mengatasi masalah vanishing gradient pada pemrosesan sekuens panjang (Hochreiter & Schmidhuber, 1997). Kemampuan ini diperoleh melalui struktur cell state dan tiga mekanisme gerbang, yaitu forget gate, input gate, dan output gate, yang secara selektif mengatur informasi mana yang dipertahankan atau dibuang sepanjang urutan waktu. Karakteristik ini menjadikan LSTM sesuai untuk menangkap pola tren dan momentum pada data historis harga saham.

2.3 Bidirectional LSTM (Bi-LSTM)

Bi-LSTM merupakan pengembangan dari LSTM yang memproses data sekuensial secara dua arah, yaitu dari awal ke akhir (forward) dan dari akhir ke awal (backward), sehingga representasi pada setiap titik waktu memuat konteks dari masa lalu maupun masa depan dalam jendela pengamatan yang sama (Schuster & Paliwal, 1997). Pada penelitian tahap ini, Bi-LSTM digunakan sebagai ekstraktor fitur (feature extractor) untuk mentransformasi data OHLCV mentah berdimensi 5 menjadi representasi temporal berdimensi 64, dengan menggabungkan hidden state arah maju dan mundur yang masing-masing berdimensi 32. Representasi ini bertujuan menangkap pola momentum dan volatilitas pasar secara lebih kaya dibandingkan nilai numerik mentahnya, dan kinerjanya



diuji secara langsung terhadap representasi mentah pada penelitian ini.

2.4 Random Forest

Random Forest merupakan algoritma ensemble learning yang membangun sejumlah pohon keputusan secara independen menggunakan subset data dan fitur yang diacak (bootstrap aggregating), kemudian menggabungkan hasil prediksi seluruh pohon melalui rata-rata (Breiman, 2001). Pendekatan ini efektif menekan risiko overfitting yang umum terjadi pada model pohon tunggal, sekaligus tetap mampu menangkap hubungan non-linear antar fitur.

2.5 Extreme Gradient Boosting (XGBoost)

XGBoost adalah algoritma ensemble berbasis boosting yang membangun pohon keputusan secara sekuensial, di mana setiap pohon baru dilatih untuk memperbaiki kesalahan residual dari pohon sebelumnya (C. CHEN, 2016). Algoritma ini dilengkapi mekanisme regularisasi untuk mengontrol kompleksitas model, sehingga umumnya menghasilkan akurasi yang kompetitif dengan waktu komputasi yang relatif efisien dibandingkan model ensemble lain.

2.6 Metrik Evaluasi

Evaluasi kinerja model prediksi pada penelitian ini menggunakan tiga metrik. RMSE mengukur akar rata-rata kuadrat selisih antara nilai prediksi dan aktual, serta memberi penalti lebih besar terhadap kesalahan besar. MAE mengukur rata-rata selisih absolut sehingga lebih tahan terhadap outlier. MAPE menyatakan kesalahan dalam bentuk persentase yang bersifat scale-independent, sehingga memudahkan perbandingan kinerja antar model maupun antar representasi fitur (Hyndman & Koehler, 2006). Semakin kecil nilai ketiga metrik ini, semakin baik kinerja model prediksi yang dihasilkan.

METODE PENELITIAN

3.1 Desain Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimental komputasional (Putra et al., 2025), yang bertujuan membandingkan kinerja tiga algoritma machine learning pada dua representasi fitur data historis saham IHSG.

3.2 Pengumpulan Data

Data historis saham IHSG dikumpulkan melalui scraping menggunakan pustaka Python yfinance dengan kode indeks ^JKSE, mencakup periode 1 Januari 2015 hingga awal tahun 2026 dengan frekuensi harian. Data terdiri dari 2.661 baris dengan atribut Open, High, Low, Close, dan Volume, dengan harga penutupan (Close) ditetapkan sebagai variabel target.

3.3 Representasi Fitur

Dua representasi fitur dibandingkan pada penelitian ini. Representasi pertama menggunakan 5 atribut OHLCV mentah tanpa proses encoding apa pun sebagai baseline. Representasi kedua menggunakan 64 dimensi fitur hasil

ekstraksi jaringan Bi-LSTM, yang memproses data OHLCV melalui sliding window 10 hari perdagangan untuk menangkap pola temporal dan momentum pasar. Output akhir Bi-LSTM diperoleh dari penggabungan hidden state arah maju dan mundur, masing-masing berdimensi 32, sehingga menghasilkan representasi akhir berdimensi 64.

3.4 Arsitektur Model

Model LSTM disusun dalam arsitektur stacked dua lapisan tersembunyi, dengan lapisan pertama sebanyak 64 unit dan lapisan kedua 32 unit, dilengkapi dropout 20% pada setiap lapisan untuk mencegah overfitting, dilanjutkan lapisan dense 16 neuron beraktivasi ReLU dan lapisan output regresi tunggal. Pelatihan menggunakan optimador Adam dengan fungsi kerugian Mean Squared Error (MSE). Model Random Forest dikonfigurasi dengan 300 pohon keputusan ($n_estimators = 300$) tanpa batas kedalaman maksimal ($max_depth = None$). Model XGBoost dikonfigurasi dengan 500 tahap boosting ($n_estimators = 500$), learning rate 0,05, kedalaman maksimal pohon 5, subsample 0,8, dan $colsample_bytree = 0,8$.

3.5 Skema Pembagian Data dan Evaluasi

Data dibagi secara kronologis dengan porsi 20% data terakhir sebagai data uji, agar integritas deret waktu tetap terjaga. Model Random Forest dan XGBoost diselaraskan pemotongan datanya (look back 10 hari) agar setara dengan LSTM. Normalisasi fitur menggunakan MinMax Scaler yang di-fit hanya pada data latih. Evaluasi kinerja model menggunakan tiga metrik, yaitu Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan Mean Absolute Percentage Error (MAPE) (Hyndman & Koehler, 2006), yang dihitung setelah nilai prediksi dikembalikan ke skala harga IHSG aktual.

3.6 Pra-pemrosesan Data

Data historis IHSG hasil scraping diperiksa terlebih dahulu untuk menangani nilai kosong (missing value) yang ditemukan pada dataset mentah, menggunakan metode forward-filling agar kontinuitas deret waktu tetap terjaga. Setelah data bersih, dilakukan pergeseran variabel target (Close) sebanyak satu baris ke atas untuk menyesuaikan skema prediksi satu hari ke depan, sehingga fitur pada waktu t dibandingkan dengan target harga pada waktu $t+1$ guna mencegah kebocoran data (data leakage).

3.7 Perangkat dan Lingkungan Eksperimen

Seluruh eksperimen pada penelitian tahap ini diimplementasikan menggunakan bahasa pemrograman Python, dengan pustaka yfinance untuk pengumpulan data, TensorFlow/Keras untuk pembangunan arsitektur LSTM dan Bi-LSTM, serta scikit-learn dan XGBoost untuk pemodelan Random Forest dan XGBoost. Seluruh proses pelatihan dan pengujian model dijalankan pada lingkungan komputasi berbasis cloud untuk mempercepat proses pelatihan model deep learning.

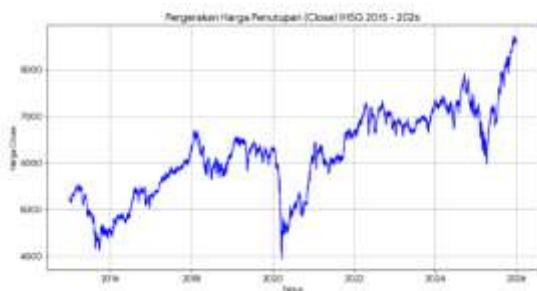


HASIL DAN PEMBAHASAN

4.1 Deskripsi Data

Proses scraping menghasilkan dataset mentah sebanyak 2.662 baris data historis IHSG dengan 6 atribut, yaitu Date, Open, High, Low, Close, dan Volume. Secara statistik deskriptif, variabel target (Close) mencatatkan nilai terendah sebesar 3.937,63 pada 26 Maret 2020 dan nilai tertinggi mencapai 8.710,70 pada 9 Desember 2025, dengan rata-rata pergerakan harga penutupan pada level 6.184,97 sepanjang periode observasi. Ditemukan 1 baris data dengan nilai kosong (missing value) yang ditangani pada tahap pra-pemrosesan sebelum data dibagi menjadi data latih dan data uji.

Secara visual, pergerakan harga penutupan IHSG selama periode observasi menunjukkan tren kenaikan jangka panjang yang diselingi oleh periode volatilitas tinggi, seperti penurunan tajam pada awal tahun 2020 yang bertepatan dengan awal pandemi, sebagaimana ditunjukkan pada Gambar 1. Karakteristik non-stasioner dan fluktuatif inilah yang menjadi alasan utama pemilihan skema pembagian data kronologis serta normalisasi MinMax Scaler yang di-fit hanya pada data latih, guna menghindari kebocoran informasi dari data uji ke proses pelatihan model.



Gambar 1. Pergerakan Harga Penutupan (Close) IHSG Periode 2015–2026

Sumber: Hasil Pengolahan Data (2026)

4.2 Hasil pada Representasi OHLCV Mentah

Pada representasi OHLCV mentah (5 fitur), ketiga algoritma dilatih murni menggunakan atribut numerik asli dari data historis saham. Hasil evaluasi disajikan pada Tabel 1.

Tabel 1. Hasil Evaluasi Model pada Representasi OHLCV Mentah

Model	RMSE	MAE	MAPE (%)
LSTM	196,11	164,14	2,18
Random Forest	482,74	298,75	3,78
XGBoost	476,59	299,51	3,80

Sumber: Hasil Pengolahan Data (2026)

LSTM menunjukkan keunggulan yang dominan pada representasi data mentah dengan MAPE terendah sebesar 2,18%, sedangkan Random Forest dan XGBoost menghasilkan akurasi yang hampir setara pada kisaran MAPE 3,78% hingga 3,80%.

4.3 Hasil pada Representasi OHLCV 64 Dimensi (Bi-LSTM)

Pada representasi kedua, ketiga model dilatih menggunakan 64 dimensi fitur hasil ekstraksi Bi-LSTM. Hasil evaluasi disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Model pada Representasi OHLCV 64 Dimensi

Model	RMSE	MAE	MAPE (%)
LSTM	638,63	542,36	7,13
Random Forest	507,94	327,76	4,17
XGBoost	565,33	391,63	5,02

Sumber: Hasil Pengolahan Data (2026)

Secara komparatif, representasi fitur yang telah di-encode secara mandiri (tanpa fitur eksternal) menghasilkan performa yang lebih rendah dibandingkan data mentah. MAPE pada LSTM melonjak signifikan menjadi 7,13%, sementara Random Forest dan XGBoost masing-masing mencatatkan MAPE 4,17% dan 5,02%. Hal ini mengindikasikan bahwa proses encoding unsupervised menyebabkan hilangnya sebagian sinyal detail harga saham absolut apabila tidak disertai penambahan informasi prediktif baru.

4.4 Perbandingan Antar Representasi

Tabel 3. Rekapitulasi Perbandingan MAPE (%) Antar Representasi Fitur

Model	OHLCV Mentah	OHLCV 64 Dimensi
LSTM	2,18	7,13
Random Forest	3,78	4,17
XGBoost	3,80	5,02

Sumber: Hasil Pengolahan Data (2026)

Berdasarkan keseluruhan eksperimen, kombinasi terbaik secara mutlak dicapai oleh LSTM pada representasi OHLCV mentah dengan RMSE 196,11 dan MAPE 2,18%. Data harga saham mentah terbukti memiliki kapabilitas sinyal murni (pure signal) yang lebih unggul dibandingkan data yang telah di-encode tanpa disertai fitur eksternal tambahan. Pada model berbasis pohon keputusan (Random Forest dan XGBoost), penurunan performa akibat encoding relatif lebih kecil dibandingkan LSTM, yang mengindikasikan bahwa model berbasis pohon lebih tahan terhadap perubahan representasi fitur pada kasus ini. Temuan pada tahap baseline ini menjadi dasar pertimbangan metodologis bahwa representasi fitur hasil encoding baru relevan digunakan apabila disertai dengan penambahan sumber informasi prediktif baru, yang menjadi fokus pada tahap penelitian selanjutnya.

Jika ditinjau dari sisi magnitude kesalahan absolut, selisih RMSE antara representasi mentah dan representasi 64 dimensi pada LSTM mencapai lebih dari 440 poin indeks (196,11 berbanding 638,63), sebuah selisih yang cukup besar mengingat rata-rata harga penutupan IHSG selama



periode observasi berada pada level 6.184,97. Hal ini menegaskan bahwa bagi kebutuhan prediksi jangka pendek dengan sumber data tunggal, representasi mentah tetap menjadi pilihan yang lebih aman secara praktis dibandingkan representasi hasil encoding, kecuali representasi tersebut disertai dengan penambahan sumber informasi baru yang mampu mengompensasi hilangnya sinyal absolut.

Perbedaan pola penurunan performa antara LSTM dan model berbasis pohon (Random Forest dan XGBoost) juga mengungkap karakteristik mendasar dari masing-masing pendekatan. LSTM sebagai model sekuensial sangat bergantung pada informasi urutan dan skala nilai mentah untuk mempelajari pola temporal secara langsung, sehingga kehilangan sebagian informasi absolut akibat proses encoding berdampak lebih besar terhadap akurasinya. Sebaliknya, Random Forest dan XGBoost memproses data dalam bentuk tabular flat yang telah direkayasa melalui teknik lag, sehingga keduanya relatif lebih stabil terhadap perubahan representasi fitur meskipun tetap mengalami penurunan akurasi. Karakteristik ini memberikan pertimbangan penting dalam pemilihan algoritma pada tahap penelitian lanjutan, khususnya ketika representasi fitur hasil encoding akan digabungkan dengan sumber data eksternal.

Perlu dicatat bahwa hasil pada tahap baseline ini dievaluasi pada satu kali pembagian data latih dan data uji secara kronologis (single split), sehingga belum mencerminkan variasi kinerja model pada berbagai kondisi rezim pasar (bullish, bearish, maupun sideways) secara terpisah. Selain itu, proses encoding Bi-LSTM pada representasi 64 dimensi bersifat unsupervised, yaitu dilatih tanpa memanfaatkan label harga secara langsung, sehingga representasi yang dihasilkan belum tentu optimal untuk tugas prediksi harga penutupan. Kedua hal ini menjadi pertimbangan penting dalam menginterpretasikan hasil perbandingan pada penelitian tahap ini.

Temuan ini juga memiliki implikasi praktis bagi pengembangan sistem prediksi harga saham berbasis data historis. Bagi praktisi maupun peneliti yang hendak membangun model prediksi IHSG dengan sumber data tunggal (single-source), penggunaan data OHLCV mentah bersama arsitektur LSTM stacked terbukti menjadi pilihan yang paling efisien tanpa memerlukan tahap encoding tambahan. Sebaliknya, apabila representasi fitur hasil encoding hendak digunakan, misalnya untuk menyeimbangkan dimensi dengan sumber data eksternal pada tahap penelitian selanjutnya, maka penurunan sinyal murni akibat proses encoding perlu dikompensasi dengan penambahan informasi prediktif baru agar akurasi model tetap terjaga atau meningkat.

KESIMPULAN

Penelitian tahap baseline ini berhasil membandingkan kinerja LSTM, Random Forest, dan XGBoost dalam memprediksi harga penutupan IHSG menggunakan dua representasi data historis OHLCV. Kombinasi LSTM dengan representasi data mentah menghasilkan kinerja prediksi terbaik (RMSE = 196,11,

MAE = 164,14, MAPE = 2,18%), mengungguli Random Forest dan XGBoost pada representasi yang sama. Reduksi dimensi OHLCV menjadi representasi 64 dimensi melalui Bi-LSTM tanpa disertai fitur eksternal justru menurunkan akurasi pada ketiga algoritma, terutama pada LSTM. Temuan ini menetapkan konfigurasi baseline dan alur evaluasi yang tervalidasi, yang menjadi fondasi metodologis bagi tahap penelitian lanjutan untuk mengintegrasikan sumber data eksternal ke dalam kerangka eksperimen yang sama guna meningkatkan akurasi prediksi harga saham IHSG.

Penelitian tahap selanjutnya disarankan untuk mengintegrasikan sumber data eksternal, seperti representasi teks regulasi pemerintah atau indikator makroekonomi lain, ke dalam kerangka eksperimen yang sama guna menguji apakah penambahan informasi tersebut dapat mengompensasi penurunan sinyal akibat proses encoding fitur OHLCV. Selain itu, pengujian pada rentang waktu look back yang lebih bervariasi dan validasi menggunakan skema walk-forward juga disarankan untuk memperkuat generalisasi model pada kondisi pasar yang berbeda-beda.

UCAPAN TERIMA KASIH

Terima kasih kepada bapak Cendra atas bimbingan yang diberikan selama proses penelitian ini, serta kepada Program Studi S1 Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya atas dukungan fasilitas dan lingkungan akademik yang mendukung penyelesaian penelitian.

DAFTAR PUSTAKA

- Aditya, & Arafat, F. (2025). Pengaruh Tingkat Suku Bunga Bank Indonesia, Jumlah Uang Beredar dan Tingkat Inflasi Terhadap Indeks Harga Saham Gabungan di Indonesia. *JEMSI (Jurnal Ekonomi, Manajemen, Dan Akuntansi)*, 11(2), 858–868. <https://doi.org/10.35870/jemsi.v11i2.3933>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- C. CHEN, T.; G. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California.
- Christophorus Bintang Saputra, & Koesrindartoto, D. P. (2024). Pemanfaatan Analisis Sentimen Youtube untuk Prediksi Harga Saham: Studi pada Investor Retail Indonesia. *Jurnal Manajemen*, 21(1). <https://doi.org/10.25170/jm.v21i1.5184>
- Edi Ismanto, Ahmad Gunawan Dalimunthe, Muhammad Iqbal, & Fauza Addinunnisa. (2026). Analisis Perbandingan Model Machine Learning dan Deep Learning untuk Peramalan Harga Saham. *Jurnal Ilmu Komputer Dan Teknik Informatika*, 2(1), 131–139. <https://doi.org/10.64803/juikti.v2i1.114>
- Farhan Fadhila, M., Septian Anwar, R., & Riyandi, A. (2025). *Journal of Innovative and Creativity Prediksi*



- Harga Penutupan Indeks Harga Gabungan (IHSG) Menggunakan Algoritma Long Short-Term Memory Berbasis Data Historis. In *Journal of Innovative and Creativity* (Vol. 5, Number 3).
- hamzah, B., & Hartono, J. (2023). ANALISIS PREDIKSI INDEKS HARGA SAHAM GABUNGAN (IHSG) MENGGUNAKAN METODE NONLINEAR AUTO-REGRESSIVE EXOGENOUS MODEL (NARX) NEURAL NETWORK DAN METODE LONG SHORT-TERM MEMORY (LSTM). <http://etd.repository.ugm.ac.id/>
- Heriyanto, M. (2025, December 30). Jumlah investor pasar modal RI capai 20,32 juta SID di akhir 2025. *ANTARA News*. <https://www.antaranews.com/berita/5327392/jumlah-investor-pasar-modal-ri-capai-2032-juta-sid-di-akhir-2025>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short Term Memory. *Neural Computation*. *Neural Computation*, 9(8).
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4). <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Indra Jaya, K., Mulyanti Suhartini, A., Jariyatin, N., & Septiani, D. (2026). PREDIKSI INDEKS HARGA SAHAM GABUNGAN INDONESIA MENGGUNAKAN MODEL CNN-CA-LSTM DAN APLIKASI WEBSITE STREAMLIT. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 10, Number 1).
- Jange, B. (2022). Prediksi Harga Saham Bank BCA Menggunakan XGBoost. *ARBITRASE: Journal of Economics and Accounting*, 3(2). <https://doi.org/10.47065/arbitrase.v3i2.495>
- P3BMS UMA. (2025, March 19). Pengaruh Anjloknya IHSG terhadap Perekonomian dan Pasar Modal Indonesia. P3BMS Universitas Medan Area. <https://p3bms.uma.ac.id/pengaruh-anjloknya-ihsg-terhadap-perekonomian-dan-pasar-modal-indonesia/>
- Putra, R. D., Murtadho, M. A., Isnaini, M., & Afgani, M. W. (2025). Design Penelitian Kuantitatif: Pengertian dan Macam-macam Jenisnya. *EDU SOCIETY: JURNAL PENDIDIKAN, ILMU SOSIAL DAN PENGABDIAN KEPADA MASYARAKAT*, 5(3). <https://doi.org/10.56832/edu.v5i3.1805>
- Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11). <https://doi.org/10.1109/78.650093>