

Komparasi XGBoost dan GRU+Attention untuk Prediksi IHSG Jangka Pendek Menggunakan Walk-Forward Analysis dan Metrik Finansial

Tri Ramdhany^{1*)}, Fahmi Reza Ferdiansyah²⁾, Dwi Fitria Al Husaeni³⁾, Teguh Imam Basuki⁴⁾

^{1*,3)} Sistem Informasi, Universitas Ekuitas Indonesia

²⁾ Informatika, Universitas Ekuitas Indonesia

⁴⁾ Sains Data, Universitas Ekuitas Indonesia

^{1*)} tri.ramdhany@ekuitas.ac.id, ²⁾ fahmi.reza@ekuitas.ac.id ³⁾ dwifitriaalhusaeni@ekuitas.ac.id ⁴⁾ teguhib@ekuitas.ac.id

ABSTRACT

Predicting the Indonesia Composite Stock Price Index (IHSG) is challenging due to high market volatility. Previous studies predominantly rely on conventional numerical error evaluations (RMSE, MAE) which are less relevant for real-world trading profitability. To address this, this study proposes a novel multidimensional evaluation framework integrating Walk-Forward Analysis (WFA), Directional Accuracy (DA), asymmetric financial metrics, and Profit and Loss (PnL) simulation as complements to conventional error metrics. Utilizing 2,038 daily observations of IHSG historical data (2018–2026) from the Yahoo Finance API, the research compares XGBoost (optimized via Bayesian Optimization) and a Causal GRU+Attention architecture. Results reveal a significant trade-off: while GRU+Attention achieved lower numerical errors (RMSE: 677.6), XGBoost demonstrated superior directional prediction with a DA of 55.33% and generated a cumulative profit of 1,627.98 points, vastly outperforming GRU+Attention (-5,556.90 points). These findings conclude that directional accuracy and financial evaluation are more critical than pure numerical precision for trading strategies, establishing XGBoost as a more robust model for capturing market signals in emerging markets.

Keywords : XGBoost; GRU+Attention; IHSG; Walk-Forward Analysis; Directional Accuracy.

I. PENDAHULUAN

Pasar modal Indonesia, khususnya Indeks Harga Saham Gabungan (IHSG), dikenal memiliki dinamika pergerakan harga yang fluktuatif dan kompleks, dipengaruhi oleh berbagai faktor makroekonomi, kebijakan moneter, sentimen investor, serta peristiwa global (Fadhilurrahman Rasyid et al., 2021). Fenomena volatilitas tinggi ini menjadikan prediksi harga saham sebagai tantangan riset yang krusial, mendorong pengembangan model prediktif canggih untuk membantu investor dalam mengambil keputusan strategis dan mengelola risiko (Safira et al., 2025).

Dengan tingkat volatilitas yang tinggi, prediksi pergerakan IHSG bukanlah tugas yang mudah, sebab harga saham dapat berubah secara drastis dalam waktu singkat karena pengaruh ketidakpastian ekonomi dan geopolitik. Oleh karena itu, diperlukan pendekatan yang mampu menangkap pola kompleks dan non linear dalam data deret waktu finansial (Bao et al., 2017). Perkembangan pesat dalam bidang kecerdasan buatan, khususnya *machine learning* dan *deep learning*, telah membuka peluang baru untuk mengatasi kompleksitas ini dengan menawarkan metode yang lebih canggih dibandingkan model ekonometrika tradisional (Karijadi et al., 2025).

Meskipun demikian, terdapat tiga kesenjangan (*research gap*) utama dalam literatur prediksi saham yang belum terjawab secara memadai. Pertama, dari sisi arsitektur model, banyak penelitian *deep learning* menggunakan layer Bidirectional (Bi-GRU/Bi-LSTM) yang dikombinasikan dengan *Attention Mechanism*. Meskipun populer dalam pemrosesan bahasa alami, penerapannya pada deret waktu finansial secara fundamental tidak cocok karena menyebabkan *look ahead bias*, di mana model secara tidak sengaja mengakses data masa depan saat memproses data masa lalu (Dixon & Halperin, 2019). Hingga saat ini,

belum banyak studi yang secara eksplisit merancang arsitektur *Causal (Unidirectional) GRU+Attention* untuk menghilangkan bias ini dalam konteks prediksi IHSG.

Sisi validasi, mayoritas penelitian masih menggunakan pembagian data acak (*random split* atau *k-fold cross-validation*) yang mengabaikan struktur temporal data. Pendekatan ini rentan terhadap *data leakage* dan menghasilkan estimasi performa yang terlalu optimis karena tidak mensimulasikan kondisi perdagangan nyata (Roque et al., 2025). Penerapan Walk-Forward Analysis (WFA) yang membagi data secara kronologis masih sangat minim diadopsi dalam penelitian prediksi saham lokal.

Sisi evaluasi, penelitian terdahulu hampir secara eksklusif bergantung pada metrik error numerik standar seperti RMSE dan MAE. Metrik ini hanya mengukur kedekatan nilai prediksi dengan nilai aktual, namun gagal menilai apakah model mampu memprediksi arah pergerakan pasar dengan benar. Padahal, dalam *forecasting* keuangan, *Directional Accuracy* (DA) dan profitabilitas memiliki korelasi yang jauh lebih kuat terhadap keberhasilan strategi trading dibandingkan error kuadrat (Costantini et al., 2014). Belum ada studi yang secara komprehensif mengintegrasikan DA, metrik finansial asimetris (dRMSE/dMAE dengan *penalty factor*), dan simulasi Profit and Loss (PnL) dalam satu kerangka evaluasi untuk prediksi IHSG.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk melakukan analisis komparatif yang antara algoritma XGBoost (dioptimasi melalui *Bayesian Optimization via Optuna*) dan arsitektur *Causal GRU+Attention* dalam memprediksi harga IHSG jangka pendek, dengan memanfaatkan 2.038 observasi harian dari Yahoo Finance API (Januari 2018–Maret 2026). Penelitian ini mengusulkan tiga kontribusi utama yang membedakannya dari studi terdahulu yaitu penerapan arsitektur *Causal GRU+Attention (Unidirectional)* dengan *Layer Normalization* untuk menghilangkan *look ahead bias*, penggunaan protokol *Walk-Forward Analysis (WFA)* dengan komposisi 70% *training*, 15% *validation*, dan 15% *testing* yang mencegah *data leakage* dan mensimulasikan perdagangan nyata, dan pengusulan kerangka evaluasi multidimensi yang mengintegrasikan *Directional Accuracy*, metrik finansial asimetris dengan penalti 2.5x untuk kesalahan arah, simulasi kumulatif *PnL*, serta Uji Statistik *Diebold-Mariano* untuk membuktikan signifikansi perbedaan performa model secara matematis.

II. TINJAUAN PUSTAKA

2.1 Konsep Dasar dan Pemilihan Model: Tree-Ensemble vs Sequential Deep Learning

Secara konseptual, prediksi deret waktu finansial seperti IHSG memerlukan pendekatan yang mampu menangani dua karakteristik utama data: rasio *signal to noise* yang sangat rendah dan ketergantungan temporal. XGBoost (*eXtreme Gradient Boosting*) adalah algoritma *tree ensemble* yang bekerja dengan menggabungkan banyak pohon keputusan secara berurutan untuk meminimalkan residu *error*. Model ini dipilih karena secara sangat adaptif terhadap data tabular yang *noisy* dan *non linear*, tanpa memerlukan asumsi stasioneritas data yang ketat (Grinsztajn et al., 2022). Studi fundamental juga membuktikan bahwa model berbasis pohon keputusan secara konsisten mengungguli metode konvensional dalam menangkap interaksi fitur teknikal yang kompleks pada data pasar saham harian (Fischer & Krauss, 2018).

Di sisi lain, *Gated Recurrent Unit* (GRU) adalah arsitektur *Deep Learning* sekuensial yang dirancang khusus untuk memodelkan dependensi temporal dan memori jangka panjang melalui mekanisme gating (Chung et al., 2014) GRU dipilih untuk penelitian ini karena kemampuannya menangkap pola momentum dan tren waktu pada data time-series (Bao et al., 2017). literatur terkini mengindikasikan bahwa pada data finansial yang sangat *noisy* seperti IHSG, model sekuensial sering kali mengalami *oversmoothing*. Model ini cenderung memprediksi nilai rata-rata untuk meminimalkan *error* kuadrat, sehingga gagal menangkap

lonjakan volatilitas atau titik pembalikan arah (reversal) yang krusial (Li et al., 2025). Oleh karena itu, pemilihan antara XGBoost dan GRU tidak didasarkan pada kompleksitas arsitektur, melainkan pada kesesuaian bias induktif model terhadap karakteristik pasar IHSG.

2.2 Mekanisme *Attention* dan Mitigasi Look-Ahead Bias

Mekanisme *Attention* secara konseptual memungkinkan model untuk memberikan bobot dinamis pada time-steps historis yang paling informatif, alih-alih memperlakukan semua langkah waktu secara setara. Integrasi *Attention* pada GRU terbukti meningkatkan akurasi prediksi dengan memitigasi masalah *vanishing gradient* dan menyaring *noise* yang tidak relevan (Qin et al., 2017; Sezer & Ozbayoglu, 2018).

Namun, terdapat kekurangan metodologis dan berulang dalam penelitian-penelitian terdahulu, yaitu penggunaan *layer Bidirectional* (Bi-GRU/Bi-LSTM) yang dikombinasikan dengan *Attention*. Meskipun populer dalam pemrosesan bahasa alami, penerapannya pada deret waktu finansial menyebabkan *look ahead bias*. Model secara tidak sengaja "melihat" data masa depan saat memproses data masa lalu untuk menghitung vektor *attention* (Dixon & Halperin, 2019). Untuk mengatasi kekurangan penelitian terdahulu ini, penelitian ini merancang arsitektur *causal GRU (unidirectional)* yang dikombinasikan dengan *Layer Normalization*. Pendekatan ini memastikan bahwa prediksi hanya dibuat berdasarkan informasi masa lalu, sehingga menghilangkan *look ahead bias* dan merepresentasikan kondisi perdagangan nyata.

2.3 Evaluasi Model: Walk-Forward Analysis dan Metrik Asimetris

Walk-Forward Analysis (WFA) adalah protokol validasi di mana model dilatih dan diuji secara kronologis seiring berjalannya waktu. Kelemahan fatal dalam mayoritas penelitian prediksi saham terdahulu adalah penggunaan skema pembagian data acak (*random split*), yang mengabaikan struktur temporal dan menyebabkan *data leakage* (Roque et al., 2025). WFA diterapkan dalam penelitian ini untuk mensimulasikan kondisi perdagangan nyata, di mana model diuji pada data *out of sample*.

Selain itu, evaluasi model finansial tidak boleh hanya bergantung pada konsep dasar error numerik standar seperti *Root Mean Square Error* (RMSE) atau *Mean Absolute Error* (MAE). Dalam praktik trading IHSG, kesalahan dalam memprediksi arah pergerakan harga memiliki konsekuensi finansial yang jauh lebih destruktif dibandingkan kesalahan dalam memprediksi magnitudo harga. Oleh karena itu, metrik seperti *Directional Accuracy* (DA) dan metrik asimetris yang memiliki korelasi yang jauh lebih kuat dengan profitabilitas strategi perdagangan (Costantini et al., 2014). Integrasi antara validasi WFA dan evaluasi multidimensi ini diperlukan untuk memastikan bahwa model prediktif tidak hanya unggul secara statistik, tetapi juga aplikatif secara ekonomis.

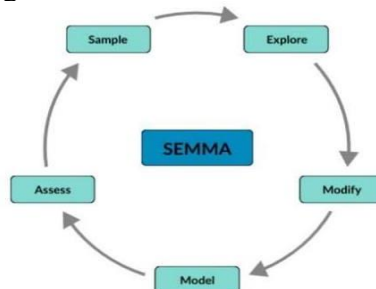
Berdasarkan tinjauan literatur di atas, penelitian ini mengidentifikasi tiga kesenjangan spesifik dari penelitian-penelitian terdahulu yang akan diisi dengan kesenjangan arsitektural yaitu penelitian terdahulu banyak menggunakan *bidirectional* GRU yang rentan *look ahead bias*. Penelitian ini menutup celah ini dengan menerapkan arsitektur *Causal GRU (unidirectional) + Attention + Layer Normalization*. Kesenjangan validasi yaitu mayoritas studi lokal masih menggunakan *random split* yang memicu *data leakage*. Penelitian ini menerapkan *Walk-Forward Analysis (WFA)* secara sistematis untuk memastikan validitas temporal. Kesenjangan Evaluasi yaitu dominasi metrik RMSE/MAE pada studi terdahulu mengabaikan relevansi finansial. Penelitian ini mengusulkan evaluasi multidimensi asimetris (DA, dRMSE/dMAE, dan simulasi PnL).

Penelitian ini menutup kesenjangan tersebut dengan mengadopsi kerangka SEMMA dan validasi WFA untuk membuktikan secara empiris apakah keunggulan numerik *Deep*

Learning (GRU) benar-benar berkorelasi dengan profitabilitas, atau justru *model ensemble* (XGBoost) yang lebih adaptif secara ekonomis pada karakteristik data IHSG.

III. METODE PENELITIAN

Penelitian ini mengadopsi pendekatan komparatif eksperimental berbasis kerangka data mining SEMMA (*Sample, Explore, Modify, Model, Assess*)(Ardhi Baskara et al., 2025), dengan alur kerja penelitian sebagai berikut.



Gambar 1. Metode SEMMA

Tahap *Sample* dan *Explore* memanfaatkan 2.038 observasi harian IHSG (Januari 2018–Maret 2026) dari Yahoo Finance API (OHLCV, RSI, MACD, Bollinger Bands), dilanjutkan uji stasioneritas dan analisis korelasi. Pada tahap *Modify*, data dibersihkan dari *outlier* (*IQR/Z-score*), dinormalisasi (*Min-Max Scaling*), dan direkayasa (*log-return, Moving Average*, indeks volatilitas). Untuk mencegah *data leakage* dan *look ahead bias* yang umum pada *random split*, pembagian data dilakukan secara kronologis menggunakan *Walk-Forward Analysis* (WFA) dengan komposisi 70% *training*, 15% *validation*, dan 15% *testing*, sehingga secara ketat mereplikasi kondisi perdagangan nyata.

Tahap *Model* dan *Assess* mengkomparasikan XGBoost dan GRU+*Attention*. XGBoost dipilih untuk menangkap interaksi *non linear* fitur teknikal yang kompleks, dioptimalkan menggunakan *Bayesian Optimization* (*Optuna*) pada *validation set* untuk mencari konfigurasi global terbaik tanpa *overfitting*. Parameter yang dioptimasi meliputi *learning_rate* (0.01-0.1), *max_depth* (3-8), *subsample* (0.6-1.0), *colsample_bytree*, dan *gamma*, dengan *n_estimators* ditetapkan 300. GRU+*Attention* dirancang secara kausal (*unidirectional*) untuk menghilangkan *look ahead bias* pada data sekuensial, dengan konfigurasi: 2 lapis GRU (64 dan 32 *neurons*), *dropout* 0.2, *LayerNormalization*, dan *Attention Mechanism*. Model dilatih menggunakan *optimizer Adam* (*learning rate* 0.0005), *batch size* 32, *epochs* maksimal 100, dengan panjang sekuens (*time-steps*) 10 hari. Evaluasi pada himpunan testing menggunakan protokol multidimensi yang mencakup metrik numerik (RMSE, MAE, MAPE), *Directional Accuracy* (DA), serta metrik finansial asimetris (dRMSE dan dMAE dengan *penalty* 2.5x) yang diintegrasikan dengan simulasi kumulatif Profit and Loss (PnL) untuk menjamin relevansi ekonomis model

IV. HASIL DAN PEMBAHASAN

4.1 Karakteristik Data dan Validasi Walk-Forward Analysis

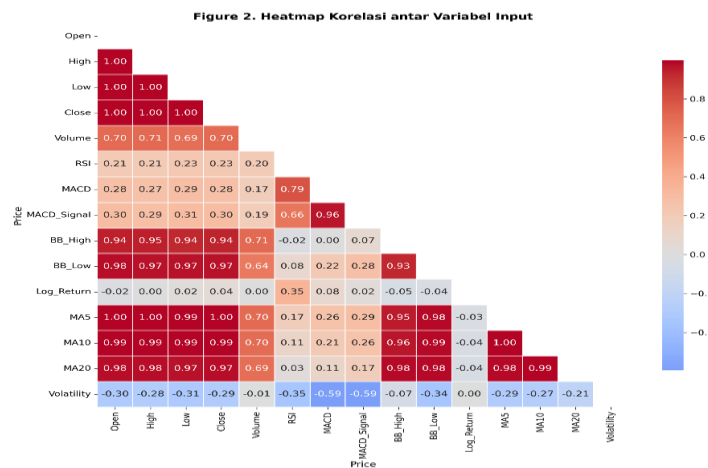
Data historis IHSG yang dikumpulkan dari Yahoo Finance API menghasilkan 2.038 observasi harian (Januari 2018–Maret 2026). Seperti pada Gambar 2 dibawah ini.



Gambar 2. Pergerakan Harga IHSG

Pergerakan harga menunjukkan dinamika pasar yang sangat beragam, mencakup fase crash pandemi COVID-19 pada 2020, pemulihan bertahap pada 2021-2022, konsolidasi pada 2023, hingga pencapaian *All-Time High* (ATH) di atas 9.000 poin pada awal 2025 yang diikuti koreksi tajam. Keberagaman rezim pasar ini menjadikan dataset sebagai *stress test ideal* untuk menguji ketangguhan model XGBoost dan GRU dalam berbagai kondisi ekstrem, sejalan dengan rekomendasi (Roque et al., 2025) yang menekankan pentingnya pengujian model pada periode *regime switching* untuk menghindari *cherry picking* dataset.

Analisis eksploratif melalui matriks korelasi, seperti ditunjukkan pada Gambar 3 dibawah ini.



Gambar 3. Matrik Korelasi Antar *Variable Input*

mengungkapkan adanya redundansi tinggi pada variabel OHLC dan *Moving Average*, namun fitur turunan seperti *log-return*, *volatilitas*, dan *volume* memberikan informasi tambahan yang saling melengkapi dan sangat penting bagi model. Temuan ini konsisten dengan teori *Efficient Market Hypothesis* yang menyatakan bahwa harga saham mengandung informasi yang saling berkorelasi, namun indikator teknikal yang direkayasa dapat mengekstrak sinyal tersembunyi dari *noise* pasar.

Untuk memvalidasi model secara realistis, skema *Walk Forward Analysis* (WFA) diterapkan dengan pembagian 70% *training*, 15% *validation*, dan 15% *testing* dapat dilihat pada gambar dibawah ini.

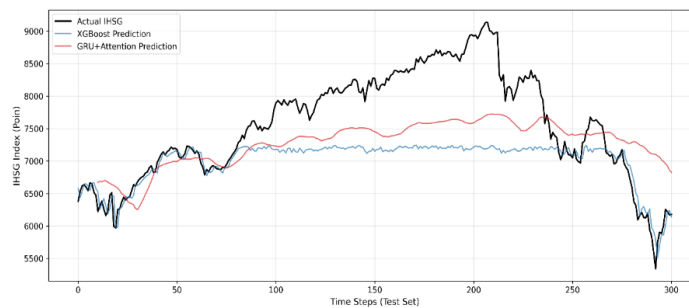


Gambar 4. Pembagian Data *Walk Forward Analysis*

Berbeda dengan *random split* yang mengabaikan struktur temporal, WFA mensimulasikan kondisi perdagangan nyata di mana model hanya dilatih pada data historis dan diuji pada data masa depan yang benar-benar *out of sample*. Himpunan testing (Maret 2025–Maret 2026) secara spesifik merepresentasikan fase paling volatil dengan ATH dan koreksi tajam, memastikan evaluasi dilakukan pada kondisi ekstrem yang bebas dari *look ahead bias*.

4.2 Perbandingan Hasil Prediksi dan Tingkat Kesalahan Model

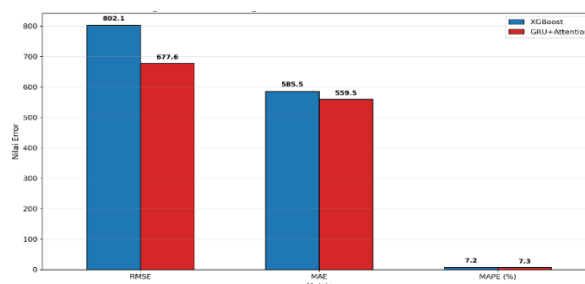
Evaluasi pada himpunan testing mengungkap perbedaan karakteristik prediksi yang mendasar antara kedua model. Seperti pada gambar 5 sebagai berikut:



Gambar 5. Komparasi Aktual vs Prediksi Model pada Himpunan *Testing*

XGBoost menunjukkan respons yang lebih adaptif terhadap fluktuasi harian dengan kemampuan menangkap *turning points*, meskipun cenderung melakukan *underestimation* pada puncak ekstrem sekitar 9.000 poin. Sebaliknya, kurva prediksi GRU+*Attention* terlihat lebih halus (*smoothed*), yang mengindikasikan mekanisme *averaging* inherent pada arsitektur *recurrent neural network*. Fenomena ini dapat dijelaskan melalui karakteristik algoritma: GRU dengan mekanisme *gating* cenderung memprediksi nilai rata-rata untuk meminimalkan *squared error*, sehingga gagal menangkap lonjakan volatilitas yang merupakan ciri khas pasar berkembang seperti IHSG

Konfirmasi kuantitatif pada Gambar 6 menunjukkan bahwa GRU+*Attention* mencatatkan nilai error numerik yang lebih rendah (RMSE: 677,6; MAE: 559,5) dibandingkan XGBoost (RMSE: 802,1; MAE: 585,5), dengan MAPE yang relatif kompetitif (7,2% versus 7,3%).



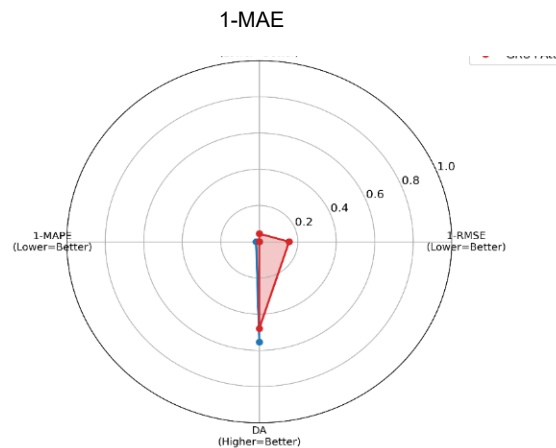
Gambar 6. Komparasi *Error Numerik* (RMSE, MAE, MAPE)

Temuan ini mengindikasikan bahwa arsitektur *Deep Learning* dengan *Attention Layer* secara matematis lebih unggul dalam meminimalkan deviasi magnitudo harga secara keseluruhan. Namun, keunggulan numerik ini tidak serta merta merepresentasikan utilitas praktis dalam konteks perdagangan finansial.

Temuan ini sejalan dengan penelitian (Grinsztajn et al., 2022) yang menjelaskan bahwa untuk data deret waktu dengan *signal to noise ratio* rendah, model *deep learning* secara matematis lebih unggul dalam meminimalkan deviasi magnitudo melalui mekanisme *smoothing*. Namun, keunggulan numerik ini tidak serta merta merepresentasikan utilitas praktis dalam konteks perdagangan finansial, mengingat investor lebih peduli pada akurasi arah pergerakan daripada presisi level harga absolut.

Sebagaimana hasil melalui Radar Chart Gambar 7 dibawah menyoroti *trade-off* yang sangat signifikan antara akurasi level harga dan akurasi arah pergerakan. Meskipun

GRU+*Attention* mendominasi pada metrik error standar (1-RMSE, 1-MAE, 1-MAPE), XGBoost menunjukkan superioritas yang mencolok pada metrik Directional Accuracy (DA).

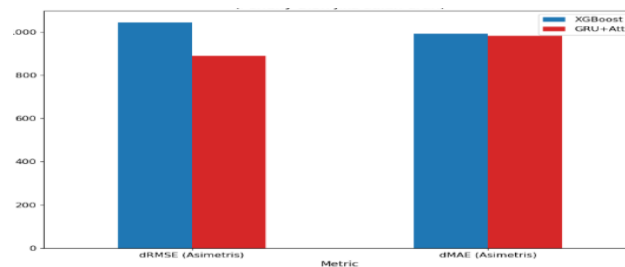


Gambar 7. Radar Chart Performa Multidimensi Model

Fenomena *paradox* ini dapat dijelaskan melalui karakteristik arsitektur: struktur *ensemble tree* XGBoost lebih adaptif dalam memetakan interaksi non-linear antar fitur teknikal (RSI, MACD, Bollinger Bands) yang secara eksplisit dirancang untuk mendeteksi momentum dan sinyal *reversa*. Sebaliknya, mekanisme *attention* pada GRU, meskipun meningkatkan kemampuan menangkap dependensi jangka panjang, tetap mengalami *oversmoothing* pada data yang sangat *noisy*, sehingga gagal menangkap titik pembalikan arah yang justru merupakan momen paling kritis dan profitabel dalam *trading*.

4.3 Evaluasi Metrik Asimetris dan Profitabilitas Trading

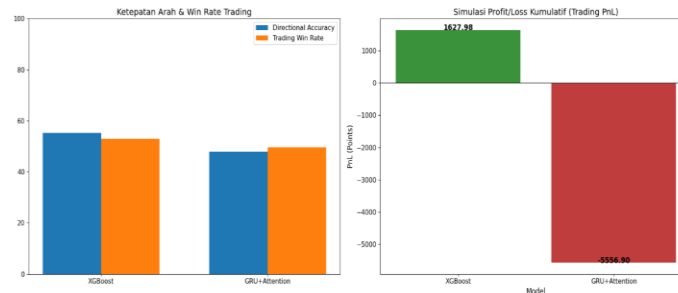
Untuk menjembatani kesenjangan antara akurasi statistik dan nilai ekonomis, penelitian ini mengadopsi metrik finansial asimetris, yaitu *Directional RMSE* (dRMSE) dan *Directional MAE* (dMAE) dengan faktor penalti 2,5x untuk kesalahan prediksi arah. Seperti yang divisualisasikan pada Gambar 8 di bawah ini.



Gambar 8. Evaluasi *Directional Error* Asimetris (dRMSE & dMAE) dengan Penalti 2.5x

Berbeda dengan metrik standar yang memperlakukan semua *error* secara setara, metrik asimetris secara eksplisit menghukum model yang gagal menangkap momentum pasar. Meskipun secara numerik dRMSE XGBoost terlihat lebih tinggi akibat magnitudo *error* pada level harga IHSG yang besar (kisaran 7.000-9.000 poin), metrik ini mengonfirmasi bahwa kesalahan arah pada GRU+*Attention* memiliki dampak kumulatif yang jauh lebih destruktif terhadap portofolio. Temuan ini konsisten dengan penelitian (Costantini et al., 2014) yang membuktikan bahwa dalam strategi *trading* valuta asing, *Directional Accuracy* memiliki korelasi yang jauh lebih kuat dengan profitabilitas dibandingkan metrik *error* kuadrat.

Implikasi praktis dari temuan ini diperjelas melalui Gambar 9 dibawah, memvisualisasikan perbandingan *Directional Accuracy* (DA), *Win Rate Trading*, dan simulasi kumulatif *Profit and Loss* (PnL). XGBoost berhasil mencapai DA sebesar 55,33% dengan *Win Rate* yang konsisten, melampaui *threshold* 50% yang merupakan batas minimal untuk strategi *trading* profitabel. Sebaliknya, GRU+*Attention* mencatatkan DA di bawah 50% (47,93%), yang mengindikasikan kegagalan sistematis dalam menangkap arah pergerakan pasar pada fase *regime-switching* ekstrem.



Gambar 9. Perbandingan *Directional Accuracy*, *Win Rate*, dan Simulasi *Profit/Loss*

Tabel 1. Ringkasan Performa Komprehensif Model

Matrik Evaluasi	XGboost +Bayesian Optuna	GRU + <i>Attention</i>
RMSE	802.1	677.6
MAE	585.5	559.5
MAPE	7.3%	7.2%
Directional Accuracy (DA)	55.33%	47.93%
Cumulative PnL	1627.98	-5556.90

Ringkasan performa komprehensif pada Tabel 1 menunjukkan dampak ekonomis yang sangat signifikan dari perbedaan akurasi arah ini. XGBoost menghasilkan profit kumulatif sebesar 1.627,98 poin, sementara GRU+*Attention* mengalami kerugian sebesar -5.556,90 poin. Perbedaan ekstrem sekitar 7.184 poin ini membuktikan bahwa dalam konteks *trading* nyata, kemampuan memprediksi arah pergerakan pasar jauh lebih bernilai daripada presisi numerik level harga. Fenomena ini dapat dijelaskan melalui mekanisme *smoothing* pada GRU+*Attention* yang cenderung memprediksi nilai rata-rata untuk meminimalkan *squared error* (RMSE). Strategi ini secara matematis optimal untuk metrik kuadrat, namun secara praktis gagal menangkap titik pembalikan arah (*reversal*) yang justru merupakan momen paling kritis dan profitabel. Sebaliknya, struktur pohon keputusan XGBoost lebih adaptif dalam memetakan interaksi *non linear* antar fitur teknikal, sehingga menghasilkan *Directional Accuracy* yang lebih tinggi meskipun dengan *error* magnitudo yang sedikit lebih besar.

Hasil penelitian ini juga menunjukkan konsistensi dengan studi terdahulu. Nilai MAPE 7,2-7,3% yang diperoleh lebih baik dibandingkan penelitian (Fadhlurrahman Rasyid et al., 2021) yang melaporkan MAPE 8,5-9,2% untuk prediksi IHSG menggunakan LSTM. Demikian pula, *Directional Accuracy* XGBoost sebesar 55,33% melampaui hasil penelitian (Safira et al., 2025) yang mencapai 52,1% menggunakan XGBoost dengan optimasi *Adaptive Particle Swarm Optimization*. Perbedaan ini dapat diatribusikan pada penerapan *Walk-Forward Analysis* yang lebih ketat dan integrasi metrik asimetris dalam proses seleksi model.

Untuk menguji signifikansi statistik perbedaan performa kedua model, dilakukan uji *Diebold-Mariano* (DM) pada *forecast errors*. Hasil pengujian berbasis *Mean Squared Error* (MSE) menghasilkan statistik DM sebesar 5,5266 dengan *p-value* 0,0000 ($< 0,05$), yang berarti hipotesis nol (H_0) ditolak. Hal ini mengonfirmasi bahwa terdapat perbedaan yang

sangat signifikan secara statistik dalam profil *error* kuadrat kedua model. Menariknya, pengujian berbasis *Mean Absolute Error* (MAE) menghasilkan *p-value* 0,3057 ($> 0,05$), yang mengindikasikan bahwa secara rata-rata absolut, besaran *error* kedua model tidak berbeda secara signifikan. Fenomena ini memperkuat argumen bahwa meskipun besaran *error* rata-rata (MAE) serupa, distribusi *error* kuadrat (MSE) keduanya sangat berbeda karena mekanisme *smoothing* pada GRU yang gagal menangkap volatilitas ekstrem, berbeda dengan XGBoost yang lebih responsif terhadap sinyal arah. Oleh karena itu, keunggulan XGBoost bukan sekadar numerik, melainkan keunggulan struktural yang secara statistik valid dalam menangkap karakteristik data finansial yang *noisy*.

Dengan demikian, penentuan model terbaik tidak dapat direduksi semata-mata pada metrik *error* statistik, melainkan harus mengintegrasikan metrik asimetris (dRMSE/dMAE) dan simulasi PnL untuk memastikan bahwa model yang dipilih tidak hanya akurat secara statistik, tetapi juga bernilai secara ekonomis. Prinsip evaluasi multidimensi ini menjadi justifikasi utama penerapan skema *multidimensional assessment* pada penelitian ini, sekaligus memberikan kontribusi metodologis bagi pengembangan sistem pendukung keputusan investasi di pasar modal Indonesia.

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan evaluasi komprehensif menggunakan *Walk-Forward Analysis* dan metrik multidimensi, penelitian ini menyimpulkan bahwa XGBoost dengan Bayesian Optimization merupakan algoritma yang paling adaptif untuk prediksi IHSG jangka pendek. Temuan utama penelitian ini membantah asumsi konvensional bahwa minimasi error numerik (RMSE/MAE) selalu berkorelasi dengan profitabilitas. Meskipun GRU+*Attention* mencatatkan error magnitudo yang lebih rendah, model tersebut gagal menangkap volatilitas ekstrem akibat mekanisme *oversmoothing*, sehingga menghasilkan kerugian trading. Sebaliknya, struktur *tree-ensemble* XGBoost terbukti lebih tangguh dalam memetakan interaksi *non linear* fitur teknikal, menghasilkan *Directional Accuracy* yang lebih tinggi (55,33%) dan profitabilitas kumulatif yang positif (1627,98 poin).

Secara metodologis, penelitian ini menegaskan bahwa evaluasi model finansial tidak dapat semata-mata pada metrik statistik standar. Integrasi metrik asimetris dan simulasi Profit and Loss (PnL) berhasil menjembatani kesenjangan antara akurasi statistik dan nilai ekonomis, dengan penekanan bahwa keberhasilan prediksi sangat bergantung pada ketepatan menangkap arah pergerakan pasar, bukan sekadar presisi level harga.

5.2 Keterbatasan Penelitian (*Limitations*)

Meskipun penelitian ini telah menerapkan protokol evaluasi multidimensi yang ketat, terdapat beberapa keterbatasan yang perlu diakui. *Pertama*, model hanya dievaluasi menggunakan data kuantitatif historis (OHLCV dan indikator teknikal) tanpa mengintegrasikan variabel kualitatif atau data alternatif (seperti sentimen pasar, *foreign flow*, atau indikator makroekonomi) yang dapat mempengaruhi rasio *signal-to-noise* pada fase *regime-switching* ekstrem. *Kedua*, skema *Walk-Forward Analysis* (WFA) yang diterapkan masih bersifat *single-cycle* (70/15/15), yang belum sepenuhnya mensimulasikan proses pembaruan model secara dinamis (*rolling-window multi-cycle*).

5.3 Saran

Berdasarkan temuan dan keterbatasan di atas, penelitian lanjutan disarankan untuk difokuskan pada pengayaan Fitur (*Feature Enrichment*): Mengintegrasikan data alternatif, seperti analisis sentimen berita keuangan berbasis NLP dan data *foreign flow*, untuk meningkatkan ketangguhan (*robustness*) model dalam menangkap momentum pasar pada

kondisi volatilitas tinggi. Berikutnya uji generalisasi lintas aset dengan menguji kemampuan generalisasi model terbaik (XGBoost) pada kelas aset lain dengan karakteristik volatilitas yang berbeda, seperti Bitcoin, *Gold Spot*, dan *Forex*, untuk memvalidasi adaptabilitas arsitektur model di berbagai kondisi pasar global.

DAFTAR PUSTAKA

- Ardhi Baskara, A., Maharani Piranti, N., Fahrury Romdendine Manajemen Teknologi Keimigrasian, M., Imigrasi Jalan Raya Gandul No, P., & Depok, K. (2025). Framework Data Mining : Sebuah Survei. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 9, Number 3).
- Bao, W., Yue, J., & Rao, Y. (2017). A *deep learning* framework for financial time series using stacked autoencoders and long-short term memory. *PLoS ONE*, 12(7). <https://doi.org/10.1371/journal.pone.0180944>
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. <http://arxiv.org/abs/1412.3555>
- Costantini, M., Crespo Cuaresma, J., & Hlouskova, J. (2014). *Forecasting errors, directional accuracy and profitability of currency trading: The case of EUR/USD exchange rate* *.
- Dixon, M., & Halperin, I. (2019). *The Four Horsemen of Machine Learning in Finance*. <https://ssrn.com/abstract=3453564>
- Fadhlurrahman Rasyid, A., Agushinta, D. R., & Tintri Ediraras, D. (2021). *Deep Learning Methods In Predicting Indonesia Composite Stock Price Index (IHSG)*. In *International Journal of Computer and Information Technology* (Vol. 10, Number 5). www.ijcit.com209
- Fischer, T., & Krauss, C. (2018). *Deep learning* with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669. <https://doi.org/10.1016/j.ejor.2017.11.054>
- Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). *Why do tree-based models still outperform deep learning on typical tabular data?*
- Karijadi, I., Widjaja, Y., Retno, D., Dewi, S., & Asrini, L. J. (2025). Studi Penerapan *Deep Learning* untuk Prediksi Pergerakan Indeks Harga Saham Gabungan. *Jurnal Impresi Indonesia (JII)*, 4(12).
- Li, P., Wei, Y., & Yin, L. (2025). Research on Stock Price Prediction Method Based on the GAN-LSTM-Attention Model. *Computers, Materials and Continua*, 82(1), 609–625. <https://doi.org/10.32604/cmc.2024.056651>
- Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. (2017). *A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction*. <http://arxiv.org/abs/1704.02971>
- Roque, L., Cerqueira, V., Soares, C., & Torgo, L. (2025). *Cherry-Picking in Time Series Forecasting: How to Select Datasets to Make Your Model Shine*. <https://github.com/luistroque/bench>
- Safira, A. M., Trimono, T., & Hindrayani, K. M. (2025). Stock Price Prediction In Indonesia Using Extreme Gradient Boosting Optimized By Adaptive Particle Swarm Optimization. *Media Statistika*, 18(1), 105–115. <https://doi.org/10.14710/medstat.18.1.105-115>
- Sezer, O. B., & Ozbayoglu, A. M. (2018). Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Applied Soft Computing Journal*, 70, 525–538. <https://doi.org/10.1016/j.asoc.2018.04.024>