JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1295-1305
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1346

Evaluasi Leakage-Aware dan Imbalance-Sensitive pada BiLSTM dan Machine Learning Klasik untuk Klasifikasi Arah Pergerakan Harga Emas ANTAM

Juni Ismail[1*], Randi Sumitro[2], Juliana Rotua Pasaribu[3], Elida Madona Siburian[4], Renovand Mikael Situmorang[5]

[1]Teknik Komputer, Politeknik Bisnis Indonesia, Pematangsiantar, Sumatera Utara, Indonesia
[2]Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia
[3,4,5]Sekretari, Politeknik Bisnis Indonesia, Pematangsiantar, Sumatera Utara, Indonesia
Email: [1*]juniismailll@gmail.com, [2]randisumitro2@gmail.com, [3]juliana280780@gmail.com, [4]madonaely356@gmail.com, [5]renovandsitumorang@gmail.com
(*Email Corresponding Author: juniismailll@gmail.com)
Received: 27 Juni 2026 | Revision: 30 Juni 2026 | Accepted: 1 Juli 2026

Abstrak

Emas ANTAM merupakan instrumen lindung nilai yang banyak diminati masyarakat Indonesia, namun penentuan momentum transaksi terkendala oleh pergerakan harga yang fluktuatif dan tidak linier. Sejumlah studi terdahulu melaporkan akurasi prediksi yang nyaris sempurna, tetapi capaian tersebut kerap bersumber dari prosedur evaluasi yang rentan terhadap kebocoran data (data leakage) sehingga tidak mencerminkan kemampuan generalisasi yang sebenarnya. Penelitian ini menyusun kerangka evaluasi yang sadar-kebocoran (leakage-aware) sekaligus sensitif terhadap ketidakseimbangan kelas (imbalance-sensitive) untuk tugas klasifikasi arah pergerakan harga emas ANTAM. Data harga harian periode 2010–2025 sebanyak 5.751 sampel diolah menjadi 14 fitur teknikal—mencakup log-return berjeda, volatilitas, momentum, rasio rerata bergerak, dan RSI—kemudian dilabeli berdasarkan arah forward return lima hari. Model Bidirectional Long Short-Term Memory (BiLSTM) dibandingkan dengan Random Forest, Decision Tree, dan baseline kelas mayoritas melalui validasi walk-forward lima lipatan disertai pemangkasan (purging) serta penskalaan yang difit hanya pada data latih. Kinerja diukur menggunakan Balanced Accuracy, Macro-F1, Matthews Correlation Coefficient (MCC), ROC-AUC, dan PR-AUC. Hasil menunjukkan seluruh model klasifikasi mengungguli baseline mayoritas; Decision Tree memperoleh Macro-F1 tertinggi sebesar 0,534, diikuti Random Forest 0,510 dan BiLSTM 0,497, dengan MCC terbaik 0,074. Temuan ini mengindikasikan keterprediksian arah harga emas yang terbatas namun nyata, sekaligus menegaskan bahwa prosedur evaluasi yang ketat menghasilkan estimasi kinerja yang jauh lebih konservatif dan kredibel dibandingkan klaim akurasi tinggi pada penelitian terdahulu.

Kata Kunci: Emas ANTAM, Klasifikasi Arah, BiLSTM, Leakage-Aware, Imbalance-Sensitive

Abstract

ANTAM gold is a widely used hedging instrument among Indonesian investors, yet determining the right moment to transact remains difficult because of its volatile and non-linear price movements. Several prior studies have reported near-perfect predictive accuracy; however, such results frequently stem from evaluation procedures that are prone to data leakage and therefore do not reflect genuine generalization ability. This study develops a leakage-aware and imbalance-sensitive evaluation framework for classifying the directional movement of ANTAM gold prices. Daily price data from 2010 to 2025 (5,751 samples) are transformed into 14 technical features—comprising lagged log-returns, volatility, momentum, moving-average ratios, and RSI—and labelled according to the sign of the five-day forward return. A Bidirectional Long Short-Term Memory (BiLSTM) model is benchmarked against Random Forest, Decision Tree, and a majority-class baseline using five-fold walk-forward validation with purging and train-only feature scaling. Performance is assessed through Balanced Accuracy, Macro-F1, the Matthews Correlation Coefficient (MCC), ROC-AUC, and PR-AUC. All classifiers outperform the majority baseline, with Decision Tree attaining the highest Macro-F1 of 0.534, followed by Random Forest (0.510) and BiLSTM (0.497), and a best MCC of 0.074. These findings indicate limited but real directional predictability and confirm that rigorous evaluation yields markedly more conservative and credible performance estimates than the inflated accuracies claimed in earlier work.

Keywords: *ANTAM Gold, Directional Classification, BiLSTM, Leakage-Aware, Imbalance-Sensitive*

1. PENDAHULUAN

Emas menempati posisi istimewa dalam lanskap investasi masyarakat Indonesia karena dipandang sebagai aset yang relatif aman serta mampu mempertahankan daya beli ketika kondisi makroekonomi bergejolak. Di tingkat nasional, harga jual emas batangan produksi PT Aneka Tambang Tbk (ANTAM) berfungsi sebagai acuan utama yang diikuti oleh pasar ritel maupun investor perorangan. Sepanjang satu dekade terakhir, harga emas ANTAM menunjukkan tren menaik yang kuat, namun pada skala harian pergerakannya tetap berfluktuasi dan dipengaruhi oleh banyak faktor yang saling berkaitan, mulai dari harga acuan internasional, nilai tukar Rupiah terhadap Dolar Amerika Serikat, hingga sentimen pasar global. Dinamika semacam ini menyulitkan pelaku pasar untuk menentukan waktu yang tepat dalam membeli atau menjual, sehingga muncul kebutuhan akan model prediktif yang dapat membantu pengambilan keputusan secara lebih terukur.

This work is licensed under a Creative Commons Attribution 4.0 International License.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1295-1305
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1346

Persoalan mendasar dalam memprediksi harga emas terletak pada karakteristik datanya yang tidak linier, berderau (noisy), dan mendekati proses acak. Pendekatan statistik konvensional yang hanya mengandalkan rerata bergerak atau ekstrapolasi tren sederhana sering gagal menangkap pola ketergantungan antarvariabel yang kompleks. Keterbatasan ini mendorong adopsi metode pembelajaran mesin (machine learning) dan pembelajaran mendalam (deep learning) yang secara empiris lebih mampu memodelkan hubungan non-linier pada deret waktu keuangan. Dalam ranah peramalan harga emas, arsitektur berbasis Long Short-Term Memory (LSTM) dan variannya, termasuk Bidirectional LSTM (BiLSTM) serta model hibrida CNN-LSTM, telah banyak digunakan untuk menangkap dependensi temporal dan dilaporkan memberikan akurasi yang menjanjikan [2], [3], [4], [5].

Meskipun demikian, mayoritas kajian terdahulu menempatkan permasalahan ini sebagai tugas regresi yang meramalkan nilai harga secara numerik, dan banyak di antaranya melaporkan koefisien determinasi atau akurasi yang sangat tinggi. Bagi pelaku transaksi, informasi mengenai arah pergerakan—apakah harga akan naik atau turun pada horizon tertentu—justru lebih bersifat operasional dibandingkan estimasi nilai absolut. Sayangnya, klaim kinerja yang nyaris sempurna pada sebagian literatur patut dicermati secara kritis. Pada konteks deret waktu keuangan, capaian tersebut kerap merupakan artefak dari prosedur evaluasi yang keliru, bukan cerminan kemampuan generalisasi yang sahih.

Beberapa sumber kesalahan metodologis yang lazim ditemukan meliputi: pembagian data secara acak tanpa memperhatikan urutan waktu sehingga model memanfaatkan informasi masa depan (look-ahead bias); penskalaan atau normalisasi yang dihitung atas keseluruhan data sebelum pemisahan latih–uji; serta ketidaksesuaian metrik, yakni penggunaan metrik regresi untuk mengevaluasi keluaran yang sejatinya bersifat klasifikasi. Sebagai ilustrasi, sebuah studi prediksi harga emas di Indonesia melaporkan nilai Area Under Curve (AUC) hingga 0,987 [1]; capaian setinggi itu sukar dipertanggungjawabkan untuk tugas penentuan arah harga komoditas yang, menurut hipotesis pasar efisien, mendekati gerak acak [16]. Tanpa pengendalian kebocoran data (data leakage) dan tanpa metrik yang sesuai, angka tersebut berpotensi menyesatkan dan tidak dapat direproduksi pada kondisi nyata [10], [11].

Selain isu kebocoran, aspek ketidakseimbangan kelas (class imbalance) jarang ditangani secara eksplisit. Pada tugas klasifikasi arah, proporsi hari naik dan turun dapat tidak setara, sehingga metrik akurasi tunggal mudah memberikan kesan kinerja yang menyesatkan. Evaluasi yang adil menuntut penggunaan metrik yang peka terhadap ketidakseimbangan seperti Balanced Accuracy, Macro-F1, dan Matthews Correlation Coefficient (MCC) [12], [13], [14], [15]. Hingga kini, belum banyak penelitian yang secara khusus mengevaluasi klasifikasi arah harga emas ANTAM dengan kerangka yang sekaligus sadar-kebocoran (leakage-aware) dan sensitif terhadap ketidakseimbangan (imbalance-sensitive). Celah inilah yang menjadi fokus penelitian ini.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan membangun dan mengevaluasi secara ketat model klasifikasi arah pergerakan harga emas ANTAM dengan membandingkan pendekatan deep learning BiLSTM terhadap algoritma machine learning klasik, yaitu Random Forest dan Decision Tree, serta baseline kelas mayoritas. Seluruh model diuji menggunakan validasi walk-forward yang menghormati urutan waktu, dilengkapi pemangkasan (purging) pada batas latih–uji dan penskalaan yang difit hanya pada data latih, lalu dievaluasi dengan beragam metrik yang sesuai untuk data tidak seimbang. Kontribusi penelitian ini ada tiga. Pertama, menyajikan kerangka evaluasi leakage-aware dan imbalance-sensitive untuk klasifikasi arah harga emas ANTAM yang dapat direproduksi. Kedua, memberikan tolok ukur (benchmark) jujur yang membandingkan deep learning sekuensial dengan model klasik berbasis fitur teknikal. Ketiga, menyuguhkan analisis kepentingan fitur yang mengungkap indikator paling informatif bagi penentuan arah harga. Sisa artikel disusun sebagai berikut: Bagian 2 memaparkan metodologi penelitian, Bagian 3 menyajikan hasil dan pembahasan, dan Bagian 4 menutup dengan kesimpulan.

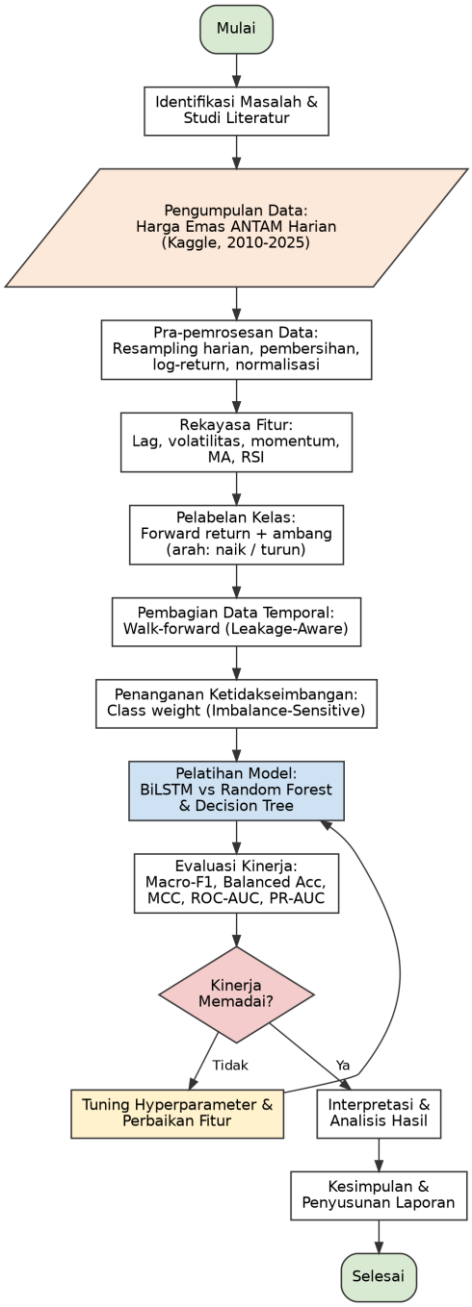
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan sistematis yang dirancang untuk menjamin kesahihan model sekaligus mencegah kebocoran informasi antar-tahap. Alur kerja keseluruhan disajikan pada Gambar 1. Proses diawali dengan identifikasi masalah dan studi literatur, dilanjutkan dengan pengumpulan data harga emas ANTAM, pra-pemrosesan, rekayasa fitur, pelabelan kelas, pembagian data secara temporal, penanganan ketidakseimbangan, pelatihan model, serta evaluasi. Apabila kinerja belum memadai, dilakukan penyetelan ulang hyperparameter dan perbaikan fitur; sebaliknya, jika telah memadai, proses berlanjut pada interpretasi hasil dan penarikan kesimpulan.

Copyright © 2026 **Author**, Page 1296
This work is licensed under a Creative Commons Attribution 4.0 International License.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1295-1305
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1346



Gambar 1. Diagram alir tahapan penelitian

Setiap tahap pada Gambar 1 dirancang agar tidak ada informasi dari periode uji yang merembes ke proses pelatihan. Penskalaan fitur, penghitungan bobot kelas, dan pemilihan model seluruhnya didasarkan semata pada data latih di tiap lipatan, sehingga estimasi kinerja yang dihasilkan bersifat konservatif dan mendekati kondisi penerapan sesungguhnya.

2.2 Pengumpulan Data

Data yang digunakan merupakan data sekunder berupa harga jual harian emas batangan ANTAM ukuran satu gram dalam satuan Rupiah, yang dihimpun melalui repositori publik Kaggle. Rentang waktu data mencakup periode 4 Januari 2010 hingga 20 November 2025. Berkas mentah memuat 4.893 baris pencatatan dengan tiga atribut, yakni stempel waktu, harga, dan tanggal. Karena pencatatan dapat terjadi lebih dari satu kali dalam sehari, data diringkas menjadi deret harian dengan mengambil kuotasi terakhir pada setiap tanggal.

2.3 Pra-pemrosesan Data

Deret harga harian dibentuk melalui resampling harian, kemudian celah pada tanggal non-perdagangan diisi menggunakan nilai terakhir yang tersedia (forward fill) agar deret menjadi teratur. Setelah tahap ini diperoleh 5.800 titik harga harian. Untuk menstabilkan ragam dan menormalkan skala, harga ditransformasikan ke dalam log-return sebagaimana Persamaan (1), dengan P_t menyatakan harga pada hari ke-t.

Copyright © 2026 **Author**, Page 1297
This work is licensed under a Creative Commons Attribution 4.0 International License.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1295-1305
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1346

$$r_t = ln(P_t) - ln(P_{t-1}) \quad (1)$$

2.4 Rekayasa Fitur dan Pelabelan

Dari deret harga harian dibangun 14 fitur teknikal yang seluruhnya bersifat kausal, yakni hanya memanfaatkan informasi masa lampau. Fitur tersebut meliputi log-return terkini beserta tiga jeda waktunya (ret_lag1–ret_lag3), volatilitas berupa simpangan baku log-return pada jendela 5, 10, dan 20 hari (vol5, vol10, vol20), momentum harga pada jendela yang sama (mom5, mom10, mom20), rasio harga terhadap rerata bergerak (maratio5, maratio10, maratio20), serta Relative Strength Index dengan periode 14 hari (RSI) yang dihitung melalui Persamaan (2). Pada persamaan tersebut, RS merupakan rasio rerata kenaikan terhadap rerata penurunan dalam periode pengamatan.

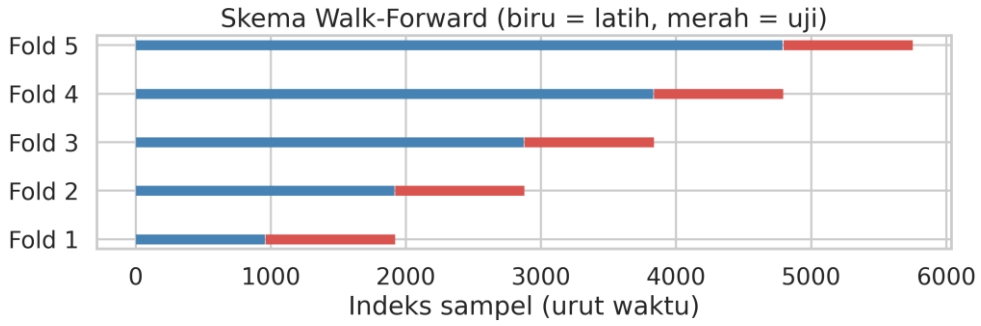
$$RSI = 100 - 100 / (1 + RS), \;\; RS = rerata\_gain / rerata\_loss \quad (2)$$

Label kelas ditetapkan berdasarkan arah forward return pada horizon lima hari sebagaimana Persamaan (3). Sampel diberi label 1 (naik) apabila harga lima hari ke depan lebih tinggi daripada harga saat ini, dan 0 (turun) untuk kondisi sebaliknya. Untuk pelatihan BiLSTM, fitur disusun menjadi sekuens dengan panjang jendela 30 hari, sedangkan untuk Random Forest dan Decision Tree digunakan vektor fitur pada titik waktu terkini.

$$y_t = 1 \; jika \; [ln(P_{t+h}) - ln(P_t)] > 0, \; selainnya \; 0 \quad (3)$$

2.5 Validasi Temporal yang Sadar-Kebocoran (Leakage-Aware)

Untuk menghindari look-ahead bias, pembagian data dilakukan dengan strategi walk-forward berbasis jendela mengembang (expanding window) sebanyak lima lipatan, sebagaimana diilustrasikan pada Gambar 2. Pada setiap lipatan, model dilatih menggunakan seluruh data hingga titik waktu tertentu, lalu diuji pada periode sesudahnya yang belum pernah dilihat. Tiga prosedur diterapkan untuk menutup celah kebocoran: (a) pemangkasan (purging) sepanjang horizon label pada batas latih–uji guna mencegah tumpang-tindih informasi forward return; (b) penghitungan parameter penskalaan StandardScaler hanya pada data latih; dan (c) penetapan bobot kelas semata berdasarkan distribusi data latih. Prosedur ini konsisten dengan praktik evaluasi deret waktu yang dianjurkan dalam literatur [10], [11].



Gambar 2. Skema validasi walk-forward (biru: data latih, merah: data uji)

2.6 Penanganan Ketidakseimbangan Kelas (Imbalance-Sensitive)

Distribusi kelas pada tugas ini tidak selalu setara sehingga model berpotensi bias terhadap kelas mayoritas. Untuk mengatasinya digunakan pembobotan kelas seimbang (balanced class weight) yang memberi penalti lebih besar pada kesalahan klasifikasi kelas minoritas. Pendekatan pembobotan dipilih alih-alih penyintesisan sampel seperti SMOTE, karena penyintesisan pada data deret waktu berisiko menimbulkan kebocoran temporal [12], [17].

2.7 Arsitektur Model

Model utama yang diusulkan adalah BiLSTM yang memproses sekuens fitur dua arah sehingga mampu menangkap konteks temporal maju dan mundur [4], [5]. Arsitekturnya terdiri atas satu lapis Bidirectional LSTM berkapasitas 48 unit, diikuti lapisan dropout 0,3, lapisan padat 32 neuron beraktivasi ReLU, dropout 0,2, dan neuron keluaran tunggal beraktivasi sigmoid. Pelatihan menggunakan pengoptimal Adam dengan laju pembelajaran 0,001, fungsi galat binary cross-entropy, serta mekanisme early stopping dan penurunan laju pembelajaran adaptif. Sebagai pembanding digunakan Random Forest dengan 300 pohon [6], Decision Tree dengan kedalaman maksimum 8, dan baseline kelas mayoritas yang selalu memprediksi kelas terbanyak pada data latih.

2.8 Metrik Evaluasi

Mengingat tugas ini bersifat klasifikasi pada data yang berpotensi tidak seimbang, kinerja model tidak diukur dengan akurasi semata, melainkan dengan lima metrik komplementer: Balanced Accuracy (Persamaan 4), Macro-F1 (Persamaan 5), MCC (Persamaan 6), ROC-AUC, dan PR-AUC. Prediksi luar-lipatan (out-of-fold) dari seluruh lipatan dikumpulkan untuk menghitung metrik gabungan, sedangkan ragam antar-lipatan dianalisis melalui Macro-F1 per lipatan.

This work is licensed under a Creative Commons Attribution 4.0 International License.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1295-1305
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1346

$$Balanced\ Accuracy = (TPR + TNR) / 2 \quad (4)$$

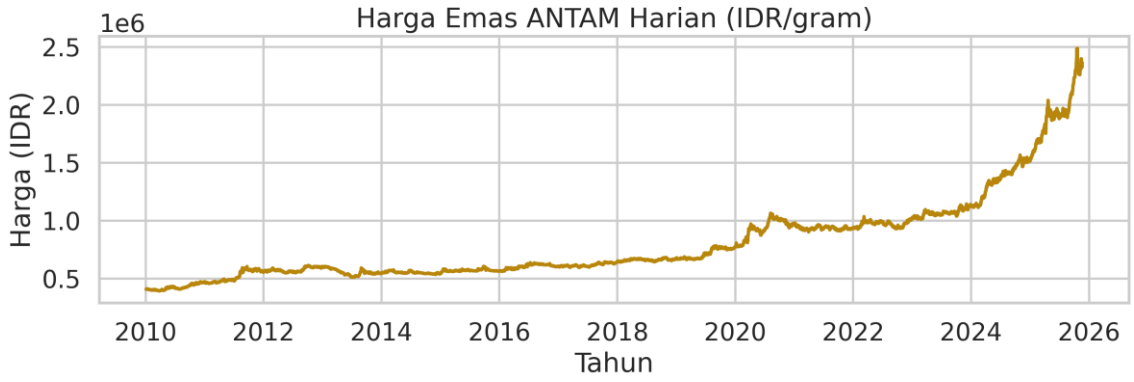
$$Macro\text{-}F1 = (F1_{naik} + F1_{turun}) / 2 \quad (5)$$

$$MCC = (TP \cdot TN - FP \cdot FN) / \sqrt{((TP+FP)(TP+FN)(TN+FP)(TN+FN))} \quad (6)$$

3. HASIL DAN PEMBAHASAN

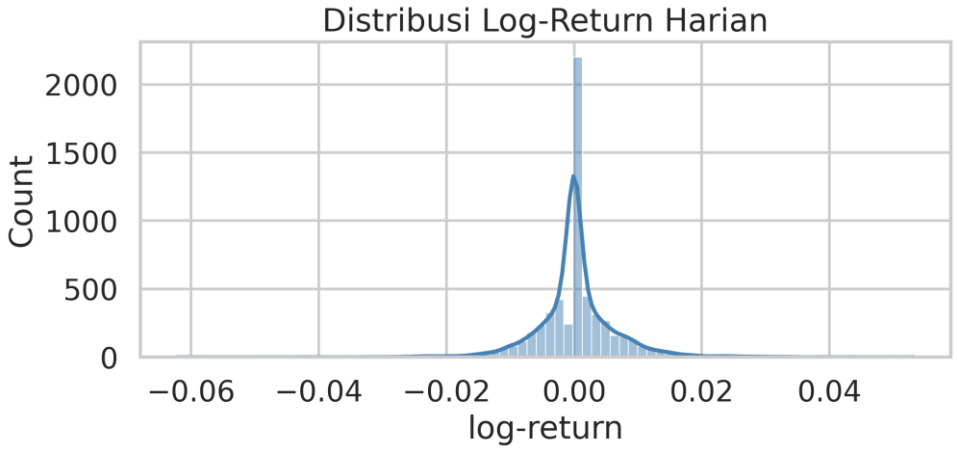
3.1 Karakteristik Data

Gambar 3 memperlihatkan lintasan harga emas ANTAM harian sepanjang 2010–2025. Terlihat tren menaik yang kuat, khususnya percepatan tajam pada periode 2024–2025, yang mengonfirmasi peran emas sebagai aset lindung nilai pada masa ketidakpastian ekonomi. Meskipun demikian, di balik tren jangka panjang tersebut terdapat fluktuasi harian yang substansial.



Gambar 3. Lintasan harga emas ANTAM harian (IDR/gram), 2010–2025

Distribusi log-return harian pada Gambar 4 memusat di sekitar nol dengan puncak yang tajam dan ekor yang relatif gemuk (leptokurtik), pola yang khas pada deret imbal hasil keuangan. Karakteristik ini menyiratkan bahwa sebagian besar perubahan harian berskala kecil, namun lonjakan ekstrem tetap terjadi sesekali sehingga menyulitkan prediksi arah secara konsisten.

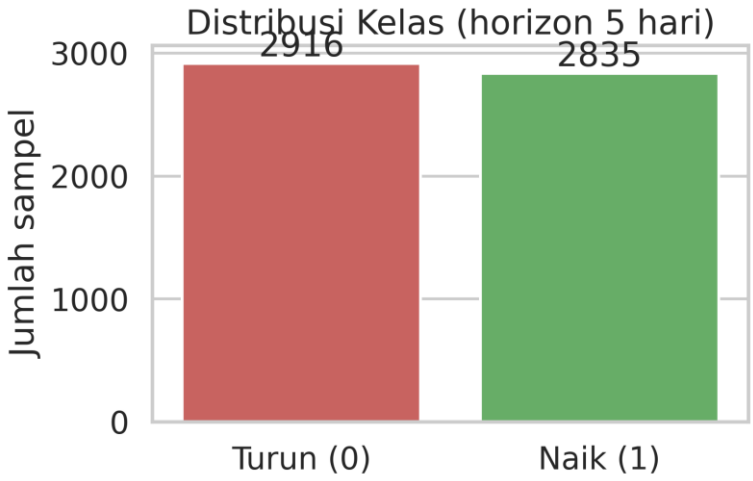


Gambar 4. Distribusi log-return harian

Gambar 5 menampilkan distribusi kelas pada horizon lima hari. Dari 5.751 sampel, sebanyak 2.916 berlabel turun dan 2.835 berlabel naik. Distribusi yang relatif berimbang ini tetap menuntut metrik yang peka ketidakseimbangan agar evaluasi tidak bias, terlebih pada masing-masing lipatan walk-forward proporsi kelas dapat bervariasi.

Copyright © 2026 **Author**, Page 1299
This work is licensed under a Creative Commons Attribution 4.0 International License.

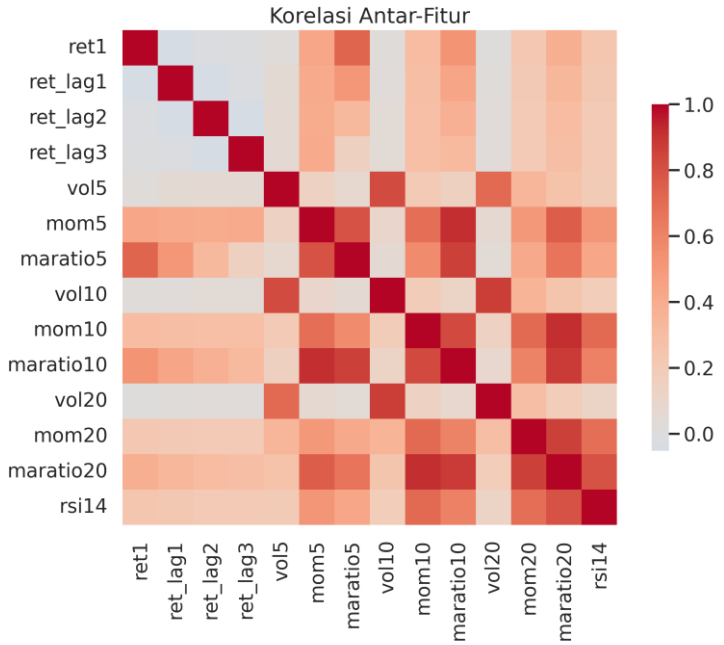
JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1295-1305
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1346



Gambar 5. Distribusi kelas pada horizon lima hari

3.2 Analisis Fitur

Matriks korelasi antar-fitur pada Gambar 6 menunjukkan bahwa fitur log-return dan jedanya hampir tidak berkorelasi satu sama lain, sedangkan kelompok volatilitas, momentum, dan rasio rerata bergerak menampakkan korelasi sedang hingga kuat pada jendela yang berdekatan. Pola ini wajar karena fitur-fitur tersebut diturunkan dari jendela waktu yang saling tumpang-tindih, dan menegaskan perlunya model yang tangguh terhadap multikolinearitas seperti metode ensembel.

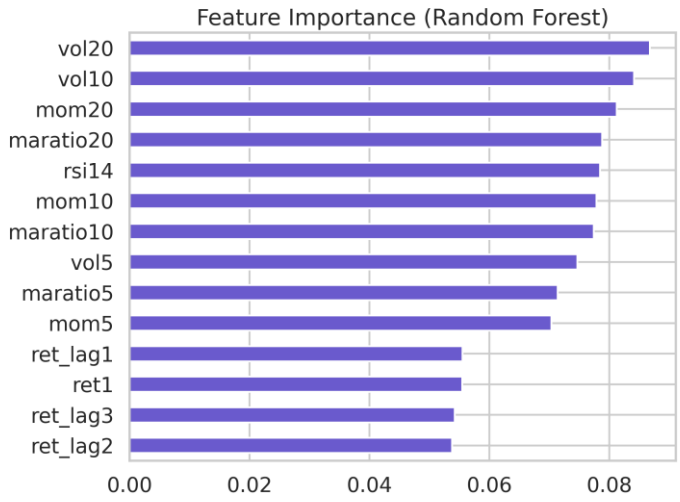


Gambar 6. Matriks korelasi antar-fitur

Peringkat kepentingan fitur dari Random Forest pada Gambar 7 menempatkan volatilitas jendela panjang (vol20 dan vol10) sebagai prediktor paling informatif, disusul momentum dan rasio rerata bergerak jendela 20 hari serta RSI. Sebaliknya, log-return tunggal dan jedanya memberi kontribusi paling kecil. Temuan ini mengindikasikan bahwa konteks volatilitas dan momentum jangka menengah lebih berdaya prediksi dibandingkan pergerakan harian tunggal yang cenderung berderau.

Copyright © 2026 **Author**, Page 1300
This work is licensed under a Creative Commons Attribution 4.0 International License.

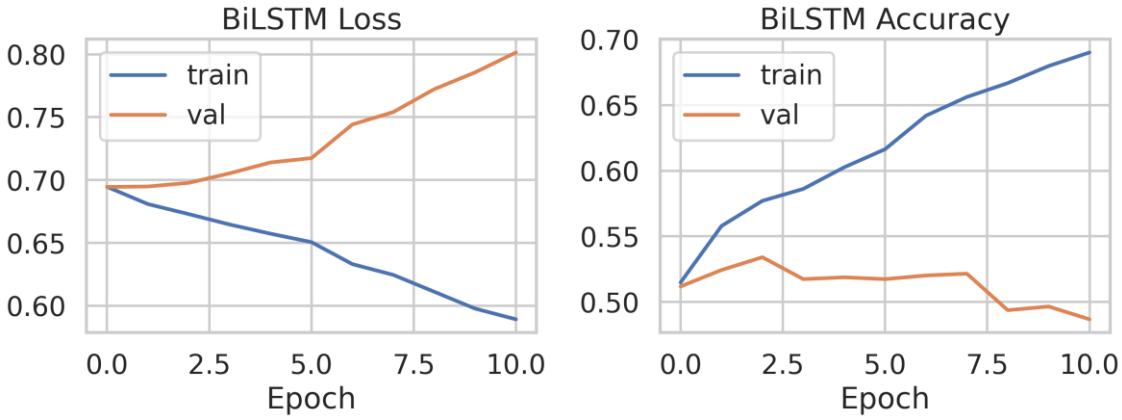
JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1295-1305
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1346



Gambar 7. Peringkat kepentingan fitur (Random Forest)

3.3 Dinamika Pelatihan

Gambar 8 menyajikan kurva galat dan akurasi BiLSTM pada salah satu lipatan. Galat pelatihan menurun secara konsisten, sementara galat validasi cenderung mendatar dan kemudian menaik, sehingga mekanisme early stopping mengembalikan bobot terbaik. Pola ini merupakan indikasi keterbatasan sinyal yang dapat dipelajari dari data, bukan kegagalan implementasi; pada tugas penentuan arah harga keuangan, jarak antara kinerja pelatihan dan validasi memang lazim menyempit pada tingkat yang rendah.



Gambar 8. Kurva galat dan akurasi pelatihan BiLSTM

3.4 Kinerja Komparatif Model

Tabel 1 merangkum kinerja luar-lipatan keempat model. Decision Tree memperoleh Macro-F1 tertinggi (0,534) dan MCC terbaik (0,074), diikuti Random Forest (Macro-F1 0,510) dan BiLSTM (0,497). Keempat model klasifikasi secara konsisten mengungguli baseline kelas mayoritas yang hanya mencapai Macro-F1 0,405 dengan MCC negatif, yang menandakan bahwa fitur teknikal yang dirancang memang memuat sinyal informatif, sekalipun terbatas.

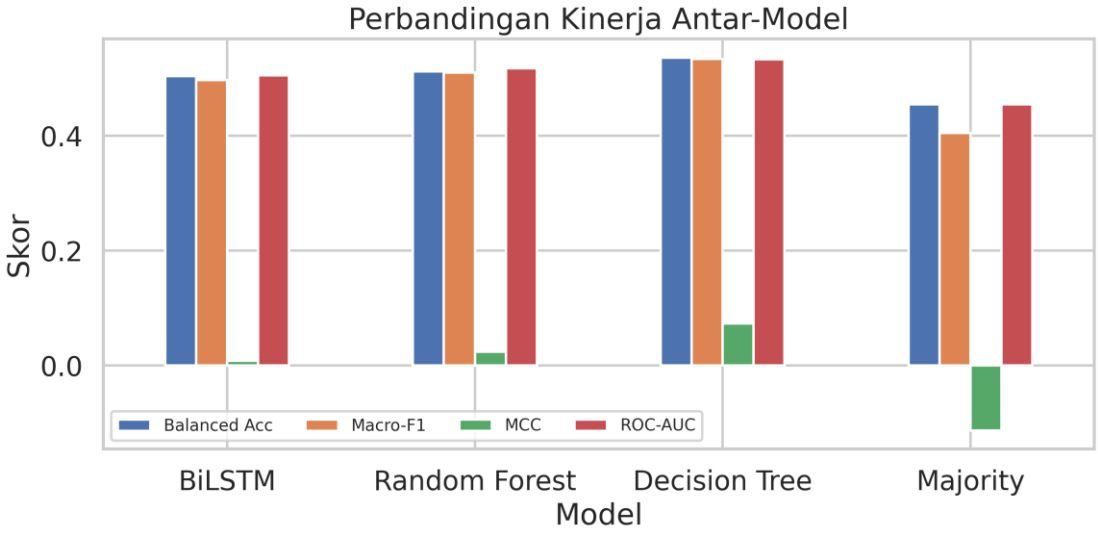
Tabel 1. Perbandingan kinerja model (rerata luar-lipatan)

Model	Acc.	Bal. Acc.	Macro-F1	MCC	ROC-AUC	PR-AUC
BiLSTM	0,501	0,504	0,497	0,008	0,505	0,499
Random Forest	0,514	0,512	0,510	0,024	0,517	0,493
Decision Tree	0,539	0,536	0,534	0,074	0,533	0,499
Majority	0,464	0,455	0,405	-0,113	0,455	0,468

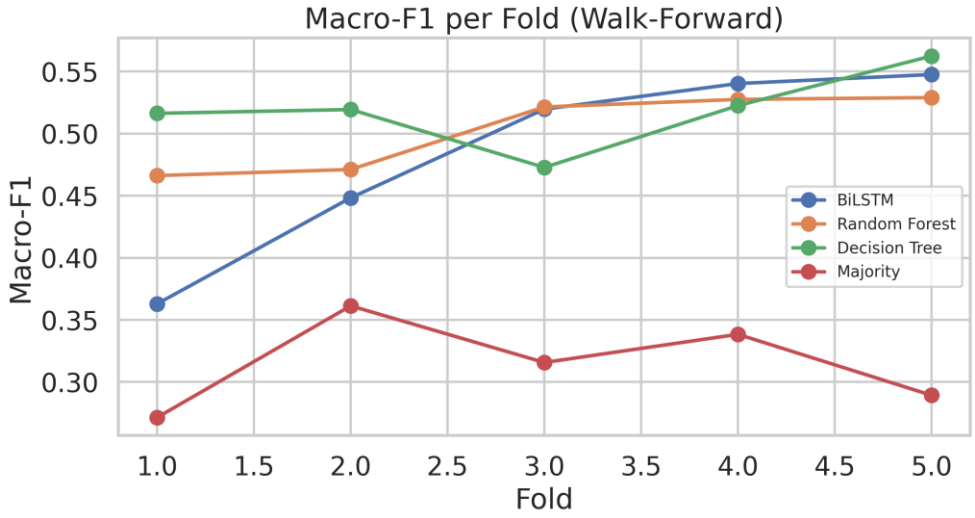
Copyright © 2026 **Author**, Page 1301
This work is licensed under a Creative Commons Attribution 4.0 International License.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1295-1305
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1346

Untuk menelaah stabilitas, Gambar 9 membandingkan metrik utama antar-model, sedangkan Gambar 10 menampilkan Macro-F1 pada setiap lipatan. Tampak bahwa kinerja BiLSTM meningkat seiring bertambahnya data latih—dari sekitar 0,36 pada lipatan pertama menjadi 0,55 pada lipatan terakhir—yang memperlihatkan sifat khas model deep learning yang haus data. Pada lipatan-lipatan akhir, BiLSTM bahkan setara dengan model klasik, namun secara agregat lima lipatan, model klasik tetap sedikit lebih unggul.



Gambar 9. Perbandingan metrik kinerja antar-model



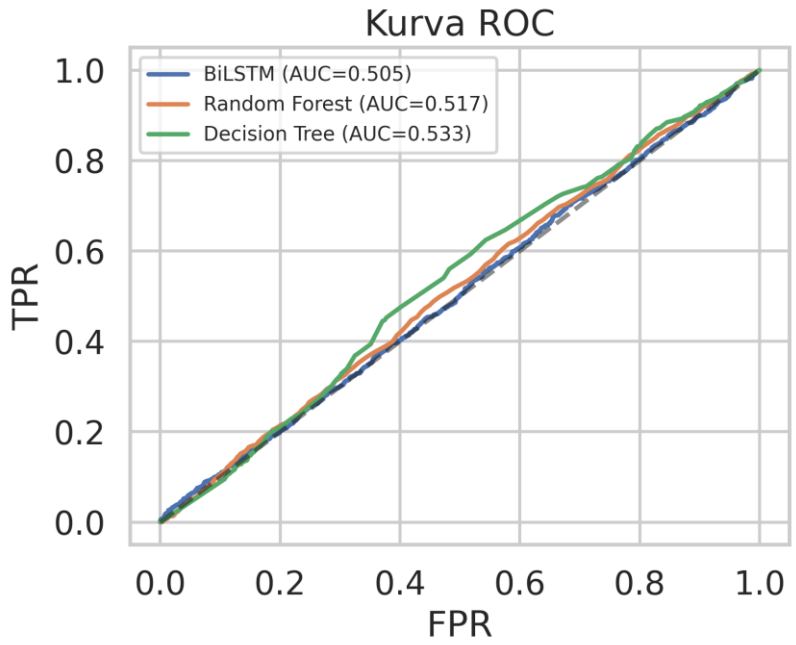
Gambar 10. Macro-F1 setiap lipatan pada validasi walk-forward

3.5 Analisis Daya Diskriminasi

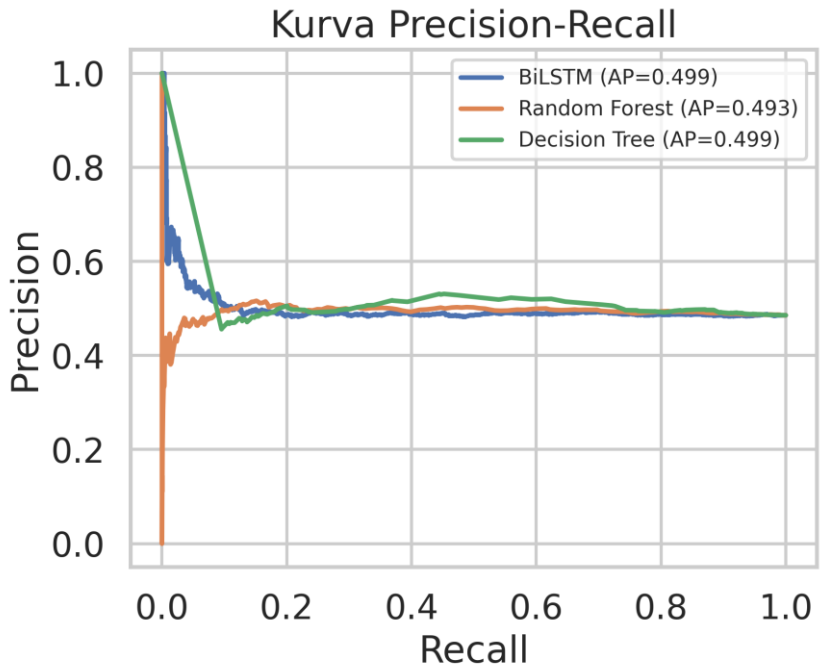
Kurva ROC pada Gambar 11 dan kurva Precision-Recall pada Gambar 12 mengonfirmasi temuan tabel: seluruh model berada hanya sedikit di atas garis acak, dengan ROC-AUC pada kisaran 0,505–0,533. Decision Tree kembali memimpin tipis. Matriks kebingungan model terbaik (Decision Tree) pada Gambar 13 memperlihatkan kemampuan mengenali kelas turun yang lebih baik (1.542 benar) dibandingkan kelas naik (1.039 benar), sebuah kecenderungan yang konsisten dengan sifat data dan pembobotan kelas yang diterapkan.

Copyright © 2026 **Author**, Page 1302
This work is licensed under a Creative Commons Attribution 4.0 International License.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1295-1305
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1346



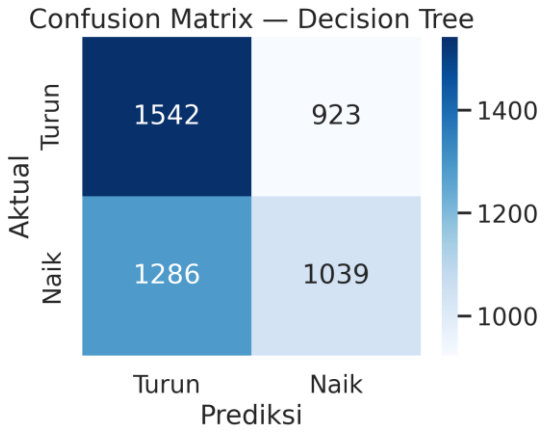
Gambar 11. Kurva ROC perbandingan model



Gambar 12. Kurva Precision-Recall perbandingan model

Copyright © 2026 **Author**, Page 1303
This work is licensed under a Creative Commons Attribution 4.0 International License.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1295-1305
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1346



Gambar 13. Matriks kebingungan model terbaik (Decision Tree)

3.6 Pembahasan

Secara keseluruhan, hasil penelitian ini menyuguhkan gambaran yang jujur sekaligus instruktif. Keterprediksian arah harga emas ANTAM ternyata terbatas: kinerja terbaik hanya mencapai Macro-F1 0,534 dan MCC 0,074. Angka ini sejalan dengan hipotesis pasar efisien yang menyatakan bahwa harga aset menyerap informasi dengan cepat sehingga komponen pergerakan harian sebagian besar bersifat acak [16]. Yang patut digarisbawahi, capaian yang tampak moderat ini justru merupakan estimasi yang kredibel karena diperoleh tanpa kebocoran data, berbeda dengan klaim AUC mendekati sempurna pada sebagian literatur terdahulu yang dievaluasi tanpa menghormati urutan waktu [1].

Fakta bahwa Decision Tree dan Random Forest sedikit mengungguli BiLSTM bukanlah anomali. Pada volume data harian yang relatif terbatas (kurang dari 6.000 sampel) dan ruang fitur yang telah direkayasa dengan indikator teknikal, model klasik berbasis pohon cenderung kompetitif, sedangkan deep learning sekuensial baru menunjukkan keunggulannya ketika tersedia data dan dimensi masukan yang jauh lebih besar. Hal ini diperkuat oleh tren peningkatan kinerja BiLSTM seiring bertambahnya data latih pada Gambar 10. Dari sisi fitur, dominasi volatilitas dan momentum jendela menengah memberi petunjuk praktis bahwa rezim volatilitas lebih informatif daripada perubahan harga harian tunggal.

Implikasi praktis temuan ini adalah perlunya kehati-hatian dalam menafsirkan model prediktif harga emas: model dapat memberikan keunggulan tipis namun nyata di atas tebakan acak, sehingga lebih tepat diposisikan sebagai alat bantu, bukan penentu tunggal keputusan. Penelitian ini memiliki keterbatasan, antara lain penggunaan fitur yang semata bersumber dari harga tanpa variabel makroekonomi eksternal seperti nilai tukar dan harga emas dunia, serta horizon pelabelan tunggal. Pengembangan lanjutan dapat mengeksplorasi fitur multivariat, horizon majemuk, arsitektur hibrida, maupun perumusan tiga kelas (naik, turun, sideways) dengan tetap mempertahankan disiplin evaluasi yang sama.

4. KESIMPULAN

Penelitian ini membangun kerangka evaluasi leakage-aware dan imbalance-sensitive untuk klasifikasi arah pergerakan harga emas ANTAM, serta membandingkan BiLSTM dengan Random Forest, Decision Tree, dan baseline kelas mayoritas melalui validasi walk-forward yang menghormati urutan waktu. Hasil menunjukkan seluruh model klasifikasi mengungguli baseline mayoritas, dengan Decision Tree memperoleh kinerja terbaik (Macro-F1 0,534; MCC 0,074), disusul Random Forest dan BiLSTM. Capaian yang moderat namun kredibel ini menegaskan bahwa keterprediksian arah harga emas bersifat terbatas, dan bahwa prosedur evaluasi yang ketat menghasilkan estimasi kinerja yang jauh lebih dapat dipertanggungjawabkan dibandingkan klaim akurasi tinggi pada penelitian terdahulu yang rentan kebocoran data. Fitur volatilitas dan momentum jangka menengah teridentifikasi sebagai prediktor paling informatif. Penelitian selanjutnya disarankan mengintegrasikan variabel makroekonomi eksternal, menelaah horizon prediksi majemuk, dan menguji arsitektur hibrida dengan tetap menjaga kedisiplinan evaluasi yang sama agar hasilnya tetap sahih dan reprodusibel.

REFERENCES

[1] N. S. U. Purba, S. S. S. Sidabutar, W. L. Simbolon, F. I. Sitohang, S. M. Samosir, and J. T. Hardinata, “Prediksi Harga Emas di Indonesia Menggunakan Machine Learning,” Jurnal Komputer Teknologi Informasi Sistem Komputer (JUKTISI), vol. 5, no. 1, pp. 999–1011, 2026, doi: 10.62712/juktisi.v5i1.1261.

[2] Y. Liang, Y. Lin, and Q. Lu, “Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM,” Expert Systems with Applications, vol. 206, art. 117847, 2022, doi: 10.1016/j.eswa.2022.117847.

Copyright © 2026 **Author**, Page 1304
This work is licensed under a Creative Commons Attribution 4.0 International License.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1295-1305
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1346

[3] P. K. Sarangi, R. Verma, S. Inder, and N. Mittal, "Machine learning based hybrid model for gold price prediction in India," in Proc. 2021 9th Int. Conf. Reliability, Infocom Technologies and Optimization (ICRITO), 2021, pp. 1–5, doi: 10.1109/ICRITO51393.2021.9596391.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[5] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.

[6] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[7] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. 3rd Int. Conf. Learning Representations (ICLR), 2015, doi: 10.48550/arXiv.1412.6980.

[8] F. Chollet, Deep Learning with Python. Shelter Island, NY, USA: Manning Publications, 2017.

[9] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," Information Sciences, vol. 191, pp. 192–213, 2012, doi: 10.1016/j.ins.2011.12.028.

[10] M. López de Prado, Advances in Financial Machine Learning. Hoboken, NJ, USA: Wiley, 2018.

[11] H. He and E. A. Garcia, "Learning from imbalanced data," IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263–1284, 2009, doi: 10.1109/TKDE.2008.239.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[13] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, art. 6, 2020, doi: 10.1186/s12864-019-6413-7.

[14] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," PLOS ONE, vol. 10, no. 3, art. e0118432, 2015, doi: 10.1371/journal.pone.0118432.

[15] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010.

[16] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," The Journal of Finance, vol. 25, no. 2, pp. 383–417, 1970, doi: 10.2307/2325486.

[17] J. W. Wilder, New Concepts in Technical Trading Systems. Greensboro, NC, USA: Trend Research, 1978.

[18] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," Biochimica et Biophysica Acta (BBA) - Protein Structure, vol. 405, no. 2, pp. 442–451, 1975, doi: 10.1016/0005-2795(75)90109-9.

Copyright © 2026 **Author**, Page 1305
This work is licensed under a Creative Commons Attribution 4.0 International License.