

Perbandingan Kinerja Model ARIMA dan LSTM pada Multi-Horizon Forecasting Harga Emas dengan Evaluasi Mean Directional Accuracy

Muhamad Prasetyo Bayu Aji*, Aris Rakhmadi

Fakultas Komunikasi dan Informatika, Prodi Teknik Informatika, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia

Email: ^{1,*}l200220072@student.ums.ac.id, ²aris.rakhmadi@ums.ac.id

Email Penulis Korespondensi: l200220072@student.ums.ac.id

Submitted: 20/05/2026; Accepted: 30/06/2026; Published: 30/06/2026

Abstrak—Logam mulia khususnya emas merupakan salah satu instrumen investasi pelindung nilai yang sangat diminati masyarakat, namun fluktuasi harganya yang dinamis menghadirkan tantangan yang menjadikannya aset yang sulit diprediksi namun krusial bagi pengambilan keputusan investasi. Penelitian ini bertujuan memprediksi harga emas dengan membandingkan kinerja model peramalan *Autoregressive Integrated Moving Average* (ARIMA) dan *Long Short-Term Memory* (LSTM) melalui tiga skenario pengujian: *single-step*, *multi-step forecasting*, dan *rolling forecasting*. Penelitian ini menggunakan data historis harga emas selama 40 tahun yang diperoleh dari sumber publik Kaggle. Model ARIMA diimplementasikan pada data stasioner, sementara LSTM dioptimalkan dengan fitur tambahan *lag*, *volatility* dan *momentum*. Hasil eksperimen menunjukkan bahwa pada skenario *single-step*, kedua model menghasilkan akurasi yang ekuivalen dengan MAPE di bawah 1%. Pada skenario *multi-step*, LSTM unggul secara signifikan dengan MAPE 1,89% dibandingkan ARIMA 3,54%. Pada skenario *rolling*, LSTM kembali lebih baik dengan MAPE 1,83% berbanding 3,52% ARIMA. Sebaliknya, ARIMA secara konsisten mencatatkan nilai MDA yang lebih tinggi pada seluruh skenario, mencapai 57,30% pada *rolling forecast* dibandingkan LSTM sebesar 46,07%, mengindikasikan keunggulan ARIMA dalam mengidentifikasi arah tren. Penelitian ini menyimpulkan bahwa pendekatan LSTM lebih optimal untuk mencapai presisi nilai prediksi pada horizon menengah, sedangkan metode statistik ARIMA lebih andal untuk memproyeksikan arah pergerakan pasar secara akurat.

Kata Kunci: ARIMA; LSTM; Forecasting; Gold; MDA

Abstract—Precious metals, particularly gold, represent one of the most sought-after value-preserving investment instruments, yet their dynamic price fluctuations present significant challenges, making gold a difficult-to-predict yet crucial asset for investment decision-making. This study aims to forecast gold prices by comparing the performance of Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) forecasting models through three testing scenarios: single-step, multi-step forecasting, and rolling forecasting. This study utilizes 40 years of historical gold price data obtained from the public source Kaggle. The ARIMA model was implemented on stationary data, while LSTM was optimized with additional lag, volatility, and momentum features. Experimental results indicate that in the single-step scenario, both models produced equivalent accuracy with a MAPE below 1%. In the multi-step scenario, LSTM significantly outperformed ARIMA with a MAPE of 1.89% compared to 3.54%. In the rolling scenario, LSTM again performed better with a MAPE of 1.83% versus 3.52% for ARIMA. Conversely, ARIMA consistently recorded higher Mean Directional Accuracy (MDA) values across all scenarios, reaching 57.30% in the rolling forecast compared to LSTM's 46.07%, indicating ARIMA's advantage in identifying trend direction. This study concludes that the LSTM approach is more optimal for achieving numerical prediction precision over medium-term horizons, while the statistical ARIMA method is more reliable for accurately projecting market movement direction.

Keywords: ARIMA; LSTM; Forecasting; Gold; MDA

1. PENDAHULUAN

Investasi merupakan cara untuk mempertahankan nilai uang dari inflasi yang umum dilakukan oleh seorang individu atau perusahaan. Investasi sering dipilih karena memiliki potensi untuk memberikan keamanan dan keuntungan dalam waktu panjang secara signifikan. Investasi yang memberikan keamanan ini salah satunya yaitu logam mulia. Logam mulia juga dapat melindungi kekayaan para investor dari fluktuasi nilai mata uang dan inflasi sehingga menjadi instrument yang populer dan semakin diminati masyarakat pada umumnya [1].

Emas merupakan salah satu komoditas yang memiliki kecenderungan nilainya naik dalam kurun waktu tertentu dan merupakan instrument investasi yang stabil sehingga banyak digemari investor maupun masyarakat umum [2]. Harga emas dapat naik dan turun yang disebabkan oleh faktor faktor yang memengaruhinya seiring berjalannya waktu. Naik turunnya harga ini adalah fenomena yang disebut dengan fluktuasi yang menimbulkan tantangan bagi seorang investor maupun masyarakat umum untuk memutuskan kapan waktu yang tepat untuk menjual atau membeli emas [3]. Logam mulia seringkali dianggap sebagai pelindung terhadap inflasi dan krisis ekonomi. Dengan adanya fenomena fluktuasi tersebut sehingga diperlukan metode analisis yang mampu membantu memberikan prediksi harga emas secara lebih akurat.

Teknologi *machine learning* telah berkembang pesat dalam beberapa tahun terakhir. *Machine learning* menyediakan kemampuan yang baik dalam menganalisis data serta melakukan peramalan [4]. Metode analisis yang dapat dimanfaatkan untuk memprediksi fluktuasi harga emas adalah metode yang memanfaatkan *machine learning* serta analisis data deret waktu atau *time-series* untuk mengetahui bagaimana pergerakan dan arah harga emas di masa mendatang. Metode yang umumnya diterapkan pada kasus ini adalah *Autoregressive Integrated Moving Average* (ARIMA), yakni pendekatan statistik yang mampu mengidentifikasi pola linier dari data historis dan menghasilkan ramalan berdasarkan tren sebelumnya [5]. Karakteristik yang dimiliki ARIMA tersebut memiliki keterbatasan ketika menangkap pola non-linier dan fluktuasi tinggi yang kerap muncul pada pergerakan harga emas. Sedangkan model

berbasis *deep learning* sering kali menunjukkan kemampuan yang lebih baik dalam mengenali pola yang kompleks dibandingkan model statistik tradisional seperti ARIMA [6]. Kemajuan teknologi *deep learning* menawarkan alternatif yang lebih efisien dan akurat [7]. Untuk mengatasi kekurangan tersebut, model berbasis kecerdasan buatan seperti *Long Short-Term Memory* (LSTM) dapat digunakan. LSTM, sebagai salah satu jenis jaringan saraf berulang, yang memiliki kemampuan untuk melakukan pembelajaran pada ketergantungan jangka panjang, sehingga lebih adaptif dalam mendeteksi pola harga yang kompleks [8]. Oleh karena itu, perbandingan antara ARIMA dan LSTM menjadi penting untuk mengetahui model yang lebih efektif dalam menghasilkan prediksi harga emas yang akurat.

Beberapa penelitian sebelumnya telah melakukan penelitian dengan membandingkan kedua metode ini dengan hasil yang beragam tergantung dengan karakteristik data. Beberapa studi menemukan bahwa model dengan pendekatan statistik tradisional memiliki keunggulan pada kondisi tertentu. Seperti penelitian mengenai peramalan penjualan [9], prediksi curah hujan di Kalimantan tengah [10], serta jumlah pengunjung perpustakaan [11], memberikan hasil bahwa ARIMA cenderung lebih unggul dalam meminimalisir *error* ketika diterapkan pada data dengan tren linier, pola musiman yang teratur, dan tingkat kompleksitas yang relatif rendah. Sebaliknya model dengan pendekatan *deep learning* memiliki keunggulan pada data yang fluktuatif. Hal ini sejalan dengan penelitian yang membandingkan kedua model pada deret waktu dengan fluktuasi yang tinggi [12], serta penelitian pada peramalan harga emas dan komoditas [13], temuan ini juga didukung oleh studi peramalan harga emas harian dan *multi horizon* [14], yang menyimpulkan bahwa arsitektur *deep learning* seperti LSTM lebih unggul dalam menangkap volatilitas serta dependensi jangka panjang dibandingkan model konvensional.

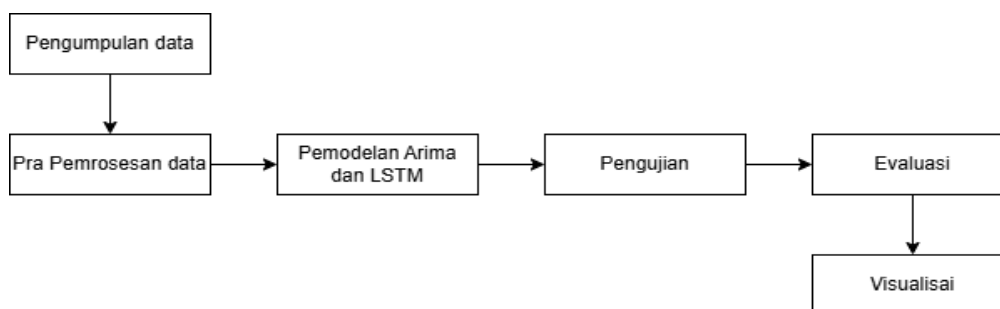
Meskipun literatur mengenai prediksi harga komoditas emas sudah beragam, terdapat kesenjangan penelitian (*Gap analysis*) yang signifikan. Mayoritas penelitian terdahulu [9] - [14] cenderung bergantung pada skenario *single-step forecasting* sebagai satu-satunya pengujian. Skenario ini hanya mengukur kemampuan model memprediksi satu hari ke depan menggunakan data aktual sebagai input di setiap langkah kondisi yang tidak merepresentasikan kebutuhan nyata investor yang seringkali memerlukan proyeksi harga untuk beberapa minggu ke depan tanpa ketersediaan data aktual harian. Penelitian terdahulu juga hanya mengandalkan metrik evaluasi kesalahan absolut yang mengukur seberapa jauh nilai prediksi dengan nilai aktual, namun tidak mengukur apakah model mampu mengidentifikasi arah pergerakan harga secara benar, Padahal dari sudut pandang praktisi identifikasi pergerakan arah lebih informatif karena secara langsung mencerminkan nilai ekonomi dan profil risiko dari strategi perdagangan dan juga kinerja statistik yang kuat tidak menjamin profitabilitas ekonomi [15].

Berdasarkan dua celah penelitian tersebut, penelitian ini memberikan dua kontribusi utama. Pertama, secara metodologis, penelitian ini merancang kerangka evaluasi komparatif yang mencakup tiga skenario forecasting diantaranya *single-step*, *multi-step*, dan *rolling forecast* sehingga konsistensi dan ketahanan kinerja model dapat dinilai pada berbagai kondisi horizon prediksi, bukan hanya pada kondisi ideal satu langkah. Kedua, secara evaluatif, penelitian ini mengintegrasikan MDA sebagai pelengkap metrik evaluasi RMSE MAE dan MAPE sehingga penilaian kinerja model mencakup dimensi akurasi nilai prediksi sekaligus ketepatan sinyal arah yang relevan untuk pengambilan keputusan investasi. Secara spesifik, tujuan utama dari penelitian ini mencakup tiga hal. Pertama, membangun model prediksi harga harian yang akurat untuk komoditas emas berbasis data historis. Kedua, melakukan analisis komparasi kinerja antara model ARIMA dan LSTM untuk menentukan metode mana yang paling tangguh melalui tiga skenario forecasting. Ketiga, mengevaluasi akurasi model menggunakan empat metrik (RMSE, MAE, MAPE, dan MDA) untuk menghasilkan rekomendasi pemilihan model paling efektif untuk kebutuhan prediksi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan komparatif untuk mengevaluasi kinerja dua model peramalan deret waktu, yaitu *Autoregressive Integrated Moving Average* (ARIMA) dan *Long Short-Term Memory* (LSTM), pada tiga skenario *forecasting* yang berbeda. Perbandingan dilakukan secara sistematis menggunakan empat metrik evaluasi sehingga penilaian mencakup akurasi nilai prediksi sekaligus ketepatan arah pergerakan harga. *Pipeline* penelitian meliputi enam tahapan utama: pengumpulan data, pra-pemrosesan, pembangunan model, pengujian skenario, evaluasi, dan visualisasi. Tahapan-tahapan tersebut digambarkan secara keseluruhan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Dataset yang digunakan pada penelitian ini merupakan data harga emas yang diperoleh dari sumber publik Kaggle (<https://www.kaggle.com/datasets/shreyanshdangi/gold-silver-price-vs-geopolitical-risk-19852025>). Periode data mencakup rentang waktu dari tahun 1985 hingga tahun 2025 dan berisi data historis harga emas harian yang terhitung dalam *business day* yang tidak termasuk akhir pekan. Variabel utama yang digunakan dalam penelitian ini adalah harga penutupan emas (*Gold Price*) karena dianggap mampu merepresentasikan nilai pasar pada akhir periode perdagangan.

2.3 Pra-pemrosesan Data

Tahapan pra pemrosesan data merupakan tahapan yang penting untuk dilakukan dengan tujuan meningkatkan kualitas data dan relevansinya sehingga data siap digunakan untuk analisis. Oleh karena itu serangkaian tahapan pra pemrosesan perlu dilakukan sebelum data digunakan untuk model.

a. Pembersihan data (*cleaning data*)

Data diperiksa terlebih dahulu untuk memastikan format dan konsistensinya. Variabel tanggal dikonversi ke format *datetime* dan digunakan sebagai indeks data. Nilai kosong yang muncul akibat proses perhitungan fitur dan transformasi data kemudian dihapus sehingga hanya data yang lengkap yang digunakan pada proses pelatihan model [16].

b. Transformasi data *return*

Setelah melalui tahap pembersihan data, harga penutupan aktual ditransformasikan menjadi bentuk *return*. Secara matematis perhitungan *return* dirumuskan pada persamaan (1) berikut :

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1)$$

Dimana R_t adalah nilai *return* pada waktu t , P_t adalah harga komoditas aktual pada waktu ke- t dan $P_t - 1$ adalah harga aktual pada waktu sebelumnya [17]. Dalam penelitian ini nilai *return* ditetapkan sebagai target utama yang akan diprediksi kedua model. Transformasi ini bertujuan untuk mengurangi unsur tren yang dominan pada pergerakan harga asli. Dengan hilangnya tren tersebut, fluktuasi data menjadi lebih stabil sehingga sangat membantu dalam meningkatkan tingkat stasioneritas deret waktu [18]. Dengan mengubah nilai harga absolut menjadi bentuk persentase *return* juga berfungsi untuk mengurangi perbedaan skala yang ekstrem antara periode awal dan akhir dataset. Penggunaan *return* memastikan bahwa algoritma dapat mengenali dan mempelajari pola fluktuasi secara konsisten selama 40 tahun dari rentang waktu yang digunakan, tanpa terdistorsi oleh perbedaan nominal harga yang jauh lebih besar di tahun-tahun terakhir. Setelah model menghasilkan prediksi dalam bentuk *return* yang telah dinormalisasi, dilakukan proses inverse transform menggunakan *StandardScaler* untuk mengembalikan nilai prediksi ke skala *return* asli. Selanjutnya, nilai *return* tersebut direkonstruksi menjadi harga prediksi secara kumulatif menggunakan harga terakhir sebagai nilai dasar. Proses ini dilakukan agar hasil prediksi berada pada skala harga aktual sehingga dapat dievaluasi menggunakan metrik RMSE, MAE, MAPE, dan MDA.

c. Feature engineering

Pemodelan *deep learning* melalui arsitektur LSTM multivariat memerlukan tahap rekayasa fitur (*feature engineering*) untuk meningkatkan daya tangkap model terhadap dinamika pasar yang kompleks. Rekayasa ini mengekstraksi serangkaian fitur tambahan, meliputi *lag return* historis untuk memberikan konteks autoregresif secara eksplisit [19]. *volatility* (volatilitas) yang berfungsi mengukur tingkat ketidakpastian pasar, serta fitur *momentum* yang mengkuantifikasi probabilitas pembalikan arah tren jangka pendek. [20]. Selain indikator turunan tersebut, model juga menggunakan *return* perak (*silver return*) sebagai prediktor eksternal. Penambahan fitur ini didasarkan pada penelitian terdahulu yang membuktikan adanya hubungan jangka panjang yang erat antara emas dan perak akibat pengaruh kondisi makroekonomi, serta tingginya interaksi saling memengaruhi antarlharga kedua logam tersebut secara dinamis dari waktu ke waktu [21]. Bersama dengan indeks waktu (*time idx*) yang merepresentasikan komponen tren secara implisit, penambahan fitur-fitur multivariat ini memberikan dimensi informasi yang jauh lebih luas. Hal ini memungkinkan lapisan memori LSTM memetakan pola non-linier tersembunyi yang tidak dapat dijelaskan apabila hanya menggunakan data *return* emas murni.

d. Normalisasi data

Tahap selanjutnya adalah normalisasi atau penskalaan data untuk meningkatkan akurasi model secara signifikan [22]. Metode Normalisasi yang digunakan adalah *Z-Score (Standard Scaler)*. Pendekatan ini mentransformasikan setiap titik data sehingga keseluruhan distribusi memiliki nilai rata-rata nol ($\mu = 0$) dan standar deviasi satu ($\sigma = 1$), dengan persamaan matematis (2) berikut :

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

Pemilihan metode *Standard Scaler* dilandasi oleh urgensi untuk mempertahankan integritas nilai negatif yang secara alamiah terdapat pada data *return* harga emas [23]. Data *return* merepresentasikan pergerakan arah harga, di

mana nilai positif mengindikasikan apresiasi harga dan nilai negatif mengindikasikan depresiasi harga dengan mempertahankan nilai positif dan negatif ini berguna untuk mendukung evaluasi model berbasis arah, seperti MDA.

e. Pembagian data

Tahap terakhir dalam pra-pemrosesan adalah pembagian data (*data splitting*) untuk data latih (*data training*) dan data uji (*data testing*). Penggunaan 80% dari keseluruhan data digunakan sebagai data latih yang bertujuan agar model mampu mempelajari pola, tren dan dependensi waktu dari data historis agar algoritma dapat mengenali karakteristik pergerakan data tersebut. Sementara itu 20% sisa data akan digunakan untuk data uji yang bertujuan untuk mengevaluasi kinerja model terhadap data baru yang belum pernah dilihat sebelumnya [24]. Sifat data deret waktu memiliki dependensi temporal yang kuat, sehingga pembagian data menjadi set pelatihan (*training set*) dan set pengujian (*testing set*) dilakukan secara sekuensial (berurutan berdasarkan urutan waktu) [20]. Pembagian data ini penting untuk memastikan bahwa model yang dibangun memiliki kinerja yang baik dan dapat diandalkan dalam meramalkan harga emas di masa mendatang.

2.4 Pembangunan Model

a. Model Autoregressive Integrated Moving Average (ARIMA)

ARIMA merupakan metode peramalan statistik yang memodelkan deret waktu berdasarkan hubungan historisnya. Model ini dinotasikan sebagai $ARIMA(p, d, q)$, yang dibentuk oleh tiga komponen utama: *Autoregressive* (p) yang merepresentasikan hubungan antara nilai masa kini dan nilai masa lalu, *Integrated* (d) menunjukkan proses diferensiasi untuk menghilangkan tren agar data menjadi stasioner, dan *Moving Average* (q) yang memperhitungkan hubungan antara nilai residual model saat ini dengan nilai residual sebelumnya [25]. Secara matematis, model ARIMA dirumuskan pada persamaan (3) berikut :

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (3)$$

Di mana Y_t adalah nilai data pada waktu ke- t , c adalah konstanta, ϕ adalah koefisien parameter AR, θ adalah koefisien parameter MA, dan ϵ_t merepresentasikan galat acak (*white noise*).

Pemodelan ARIMA mewajibkan himpunan data bersifat stasioner dengan nilai p -value dengan skor tertentu. Oleh karena itu, harga emas absolut ditransformasikan menjadi bentuk *return* ($Y_t = R_t$), yang kemudian ditetapkan sebagai target prediksi utama. Transformasi ini secara otomatis membuat parameter diferensiasi (d) bernilai nol. Uji *Augmented Dickey-Fuller* (ADF) kemudian dilakukan untuk memverifikasi secara statistik bahwa data *return* tersebut telah sepenuhnya stasioner sebelum proses pemodelan dilanjutkan [26].

Untuk mengidentifikasi kandidat parameter p dan q secara manual, dilakukan analisis visual pada plot fungsi korelasi. Plot *Autocorrelation Function* (ACF) diamati untuk melihat korelasi sisaan masa lalu guna menentukan ordo q . Sedangkan plot *Partial Autocorrelation Function* (PACF) diamati untuk melihat korelasi data masa lalu guna menentukan ordo p [27]. Selain melalui analisis plot, penentuan parameter optimal secara komputasi dilakukan menggunakan pendekatan *Auto ARIMA* dengan metode *grid search*. Metode ini mengevaluasi berbagai kombinasi parameter (p, d, q) berdasarkan kriteria minimalisasi nilai *Akaike Information Criterion* (AIC) untuk menjamin pemilihan model yang paling efisien dan akurat dengan ketentuan nilai yang semakin kecil berarti semakin benar atau baik modelnya [28], yang dirumuskan pada persamaan (4) berikut :

$$AIC = 2k - 2 \ln(L) \quad (4)$$

Di mana k adalah jumlah parameter di dalam model dan L adalah nilai *likelihood function*. Model dengan nilai AIC terendah dipilih karena memberikan keseimbangan terbaik antara kompleksitas parameter dan akurasi prediksi.

b. Model Long Short-Term Memory (LSTM)

LSTM merupakan arsitektur *deep learning* pengembangan dari *Recurrent Neural Network* (RNN) yang dirancang khusus untuk mengatasi permasalahan hilangnya gradien (*vanishing gradient*) pada pemrosesan data sekuensial jangka panjang [14]. Keunggulan LSTM terletak pada mekanisme tiga gerbang *forget gate*, *input gate*, dan *output gate* yang secara selektif mengatur aliran informasi pada *cell state*, memungkinkan model untuk mempertahankan atau membuang informasi historis sesuai relevansinya.

Pada penelitian ini, model LSTM dikembangkan menggunakan arsitektur bertingkat (*stacked LSTM*) dengan mengatur panjang jendela waktu pengamatan (*window size*) sebanyak 60 hari berturut-turut. *Window size* 60 dipilih untuk memberikan konteks historis yang cukup panjang bagi model LSTM dalam mempelajari pola tren dan volatilitas harga emas tanpa menghasilkan kompleksitas input yang berlebihan. Lapisan tersembunyi (*hidden layer*) pertama terdiri atas 64 unit neuron yang dikonfigurasi untuk meneruskan seluruh urutan informasi temporal (*return sequences=True*) menuju lapisan berikutnya. Selanjutnya, lapisan kedua beroperasi menggunakan 32 unit neuron. Untuk memitigasi risiko model terlalu menghafal data latih (*overfitting*), teknik regulasi berupa *Dropout* sebesar 0,2 (20%) disematkan pada setiap lapisan. Nilai ini dipilih karena data finansial memiliki *signal-to-noise ratio* yang rendah, sehingga *Dropout* yang terlalu besar berisiko membuang sinyal informasi yang relevan. Pada tahap akhir komputasi, lapisan keluaran (*output layer*) menggunakan satu neuron *Dense* standar untuk menghasilkan titik tunggal nilai prediksi *return* harga pada periode selanjutnya.

Proses pembaruan bobot jaringan selama fase pelatihan dieksekusi menggunakan algoritma optimasi *Adam* (*Adaptive Moment Estimation*) dengan tingkat pembelajaran (*learning rate*) yang ditetapkan secara spesifik sebesar 0,0005. Penggunaan *learning rate* yang lebih kecil dari nilai *default* (0,001) didasarkan pada karakteristik data keuangan yang sangat sensitif terhadap pembaruan bobot yang terlalu agresif, di mana nilai *learning rate* 0,0005 memberikan keseimbangan antara kecepatan konvergensi dan stabilitas pelatihan. Untuk mengevaluasi penyimpangan tebakan selama pelatihan, fungsi kerugian (*loss function*) yang digunakan adalah *Huber Loss*. Fungsi ini diimplementasikan karena karakteristik komputasinya yang mampu menyeimbangkan keunggulan *Mean Absolute Error* (MAE) dan *Mean Squared Error* (MSE), sehingga kinerja jaringan menjadi jauh lebih stabil dan tebal (*robust*) terhadap nilai-nilai pencilon ekstrem (*outliers*) yang umum terjadi pada data deret waktu finansial.

Proses pelatihan model dilakukan menggunakan maksimum 100 *epoch* dengan ukuran batch sebesar 32. Untuk mengurangi risiko *overfitting*, digunakan mekanisme *Early Stopping* dengan nilai *patience* sebesar 10 *epoch*. Pelatihan akan dihentikan secara otomatis apabila performa model pada data validasi tidak mengalami peningkatan selama 10 *epoch* berturut-turut.

2.5 Pengujian

Proses peramalan menggunakan beberapa pendekatan yaitu *single step forecasting*, *multi step forecasting*, dan juga *rolling forecast*, penggunaan ketiga skenario peramalan yang berbeda ditujukan untuk mencari tau pengaruh horizon waktu yang digunakan terhadap model [29]. Skenario yang pertama yaitu *Single step forecasting* merupakan metode peramalan jangka pendek satu langkah kedepan, dimana model hanya memprediksi nilai pada periode $t + 1$ seperti pada persamaan (5) berikut

$$\hat{y}_{t+1} = f(y_1, y_2, \dots, y_t) \quad (5)$$

Dengan merupakan nilai prediksi, $\hat{y}$ sebagai nilai aktual dan $f(\dots)$ merupakan fungsi model baik ARIMA atau LSTM. Sehingga model selalu mendapatkan pembaruan informasi pada setiap langkahnya [30]. Kemudian *multi step forecasting* merupakan metode peramalan jangka panjang dimana model menggunakan hasil prediksinya sendiri sebagai informasi untuk memprediksi horizon waktu h langkah kedepan seperti pada persamaan (6) berikut

$$\hat{y}_{t+h} = f(y_1, \dots, y_t, \hat{y}_{t+1}, \dots, \hat{y}_{t+h-1}) \quad (6)$$

Skenario terakhir *rolling forecast* yang merupakan pendekatan gabungan yang mencoba mengurangi kesalahan yang diperbuat pada *multi step forecast* yang menggulung dari waktu ke waktu. Dalam skenario ini model melakukan peramalan berulang dimana ukuran jendela data masukan selalu diperbarui dan digeser ke depan secara berkala menggunakan data aktual terbaru [31]. Skenario ini secara matematis menggeser rentang historis asal sebanyak k langkah kedepan, dengan persamaan (7) berikut

$$\hat{y}_{t+k+1} = F(y_{1+k}, y_{2+k}, \dots, y_{t+k}) \quad (7)$$

Melalui pendekatan ini, data lama yang berada di luar batas jendela ($y_1 \dots y_k$) akan dikeluarkan dari memori model, sementara data harga emas aktual yang paling baru ($y_{t+1} \dots y_{t+k}$) digunakan untuk memperbarui jendela sebelum proses peramalan pada blok k hari berikutnya dilakukan. Pembaruan jendela tersebut dilakukan setelah seluruh prediksi pada satu blok selesai dihasilkan, sehingga data aktual yang digunakan merupakan observasi dari periode yang telah berakhir. Dengan demikian, model tidak memanfaatkan informasi dari periode yang sedang diprediksi, sehingga mekanisme rolling tetap mempertahankan urutan temporal data dan terhindar dari data leakage. Skenario evaluasi berbasis jendela waktu (*window-based evaluation*) ini sangat krusial bagi model *deep learning* seperti LSTM untuk mengoreksi deviasi lintasan prediksinya (*trajectory correction*) berdasarkan informasi pasar terkini [32].

2.6 Evaluasi

Evaluasi performa model pada penelitian ini dilakukan menggunakan empat metrik pengukuran, yaitu *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), dan *Mean Directional Accuracy* (MDA). Penggunaan metrik evaluasi yang beragam ini bertujuan untuk memberikan penilaian yang lebih komprehensif terhadap kualitas prediksi [24], baik dari sisi tingkat besarnya kesalahan nominal maupun kemampuan algoritma dalam mengidentifikasi arah pergerakan harga. Sebelum menghitung metrik evaluasi, hasil prediksi dalam bentuk return terlebih dahulu direkonstruksi menjadi harga prediksi menggunakan harga terakhir sebagai nilai dasar sehingga evaluasi dilakukan pada skala harga aktual.

Untuk mempermudah penjelasan matematis dari setiap metrik, ditetapkan bahwa n adalah jumlah total observasi data, Y_t merepresentasikan nilai harga aktual pada periode ke- t , dan $\hat{y}_t$ adalah nilai harga hasil prediksi (*forecast*) dari model pada periode ke- t .

a. Root Mean Square Error (RMSE)

Secara matematis, perhitungan RMSE dirumuskan pada persamaan (8) berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2} \quad (8)$$

Metrik RMSE digunakan untuk mengukur rata-rata besarnya penyimpangan atau kesalahan prediksi. Keunikan dari metrik ini adalah kemampuannya memberikan penalti yang lebih besar terhadap *error* yang berukuran ekstrem (pencilan) akibat adanya proses kuadrat pada setiap selisih hasil prediksi dan data aktual. Semakin kecil nilai RMSE, semakin dekat pula hasil prediksi terhadap nilai aktual.

b. *Mean Absolute Error* (MAE)

Persamaan untuk metrik MAE dirumuskan pada persamaan (9) berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

MAE mengukur rata-rata selisih absolut antara nilai aktual dan nilai prediksi, tanpa memberikan bobot penalti tambahan pada *error* yang nilainya besar. Metrik pengukuran ini digunakan untuk menggambarkan besarnya kesalahan rata-rata dari prediksi model secara langsung, di mana *output* yang dihasilkan tetap berada dalam satuan skala yang sama dengan data aslinya.

c. *Mean Absolute Percentage Error* (MAPE)

Perhitungan tingkat kesalahan berbasis persentase menggunakan MAPE dirumuskan pada persamaan (10) berikut:

$$MAPE = \sum_{t=1}^n \frac{y_t - \hat{y}_t}{\hat{y}_t} \times 100\% \quad (10)$$

Metrik MAPE secara khusus digunakan untuk mengukur tingkat penyimpangan kesalahan prediksi dalam bentuk persentase ekuivalen. Transformasi nilai absolut menjadi persentase ini sangat memudahkan proses interpretasi serta perbandingan performa komparatif antar model yang berbeda. Semakin kecil persentase nilai MAPE yang dihasilkan, hal tersebut menunjukkan bahwa kinerja proyeksi model semakin akurat dan mendekati pergerakan nilai aktual.

d. *Mean Directional Accuracy* (MDA)

Rumus menghitung tingkat akurasi arah pergerakan dirumuskan pada persamaan (11) berikut :

$$MDA = \frac{1}{n} \sum_{t=1}^n a_t \quad (11)$$

dengan syarat kondisi fungsi indikator a_t pada persamaan (12) berikut:

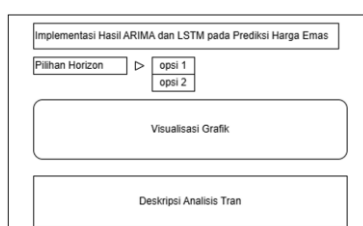
$$a_t = \begin{cases} 1, (f_{i+1} - y_i) * (y_{i+1} - y_i) \geq 0 \\ 0, otherwise \end{cases} \quad (12)$$

Berbeda secara fundamental dengan metrik RMSE, MAE, dan MAPE yang hanya berfokus pada kuantifikasi besaran nominal kesalahan prediksi, MDA digunakan untuk mengevaluasi ketepatan model dalam memprediksi sentimen arah pergerakan harga. MDA mengukur persentase kesesuaian antara arah perubahan harga aktual di pasar (naik atau turun) dengan arah perubahan harga dari hasil tebakan model [15]. Metrik ini dinilai sangat penting dalam konteks investasi finansial, karena kemampuan sebuah sistem dalam mengidentifikasi probabilitas keberlanjutan tren (kenaikan atau penurunan harga) sering kali memiliki relevansi manajerial dan taktis yang jauh lebih krusial dibandingkan sekadar menghasilkan angka proyeksi yang mendekati harga aktual semata [33].

Kombinasi keempat metrik tersebut digunakan secara terpadu untuk memperoleh gambaran empiris yang lebih menyeluruh mengenai performa model statistik (ARIMA) dan *deep learning* (LSTM) pada berbagai skenario *forecasting* di dalam penelitian ini, baik dari aspek akurasi presisi nilai maupun keandalan model dalam mengidentifikasi lintasan arah pergerakan harga emas.

2.7 Visualisasi

Hasil prediksi harga emas yang diperoleh dari model terbaik pada penelitian ini diimplementasikan dalam bentuk aplikasi web sederhana menggunakan Streamlit. Streamlit merupakan *framework* berbasis Python yang digunakan untuk membangun aplikasi web interaktif dengan proses pengembangan yang relatif sederhana. *Framework* ini dipilih karena mampu menyediakan antarmuka pengguna yang responsif serta mendukung visualisasi data dan grafik interaktif yang memudahkan pengguna dalam memahami hasil prediksi [34]. Aplikasi yang dikembangkan berfungsi untuk menampilkan hasil prediksi harga emas, menyajikan visualisasi pergerakan harga dalam bentuk grafik, serta memberikan analisis tren secara otomatis berdasarkan hasil prediksi yang dihasilkan model seperti pada Gambar 2.



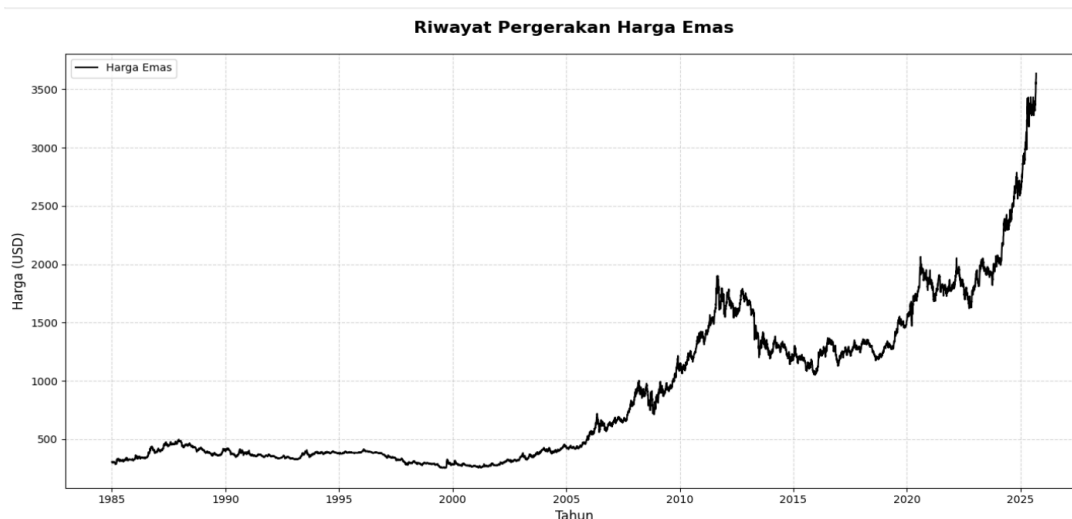
Gambar 2. Rancangan Antarmuka

Implementasi ini bertujuan untuk mempermudah pengguna dalam mengakses dan menginterpretasikan hasil penelitian melalui sebuah dashboard web interaktif.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Data

Penelitian menggunakan dataset harga emas harian dari tahun 1985 hingga tahun 2025 yang dapat dilihat pada grafik data historis harga emas Gambar 3 berikut



Gambar 3. Grafik Data Historis Harga Emas

Data harga emas menunjukkan pola pergerakan yang fluktuatif, ditandai dengan kenaikan yang cukup signifikan pada periode tahun 2010 dan volatilitas yang terjadi pada tahun berikutnya. Untuk mengetahui gambaran tingkat volatilitas harga emas secara numerik pada keseluruhan data akan disajikan pada Tabel 1.

Tabel 1. Statistik Deskriptif Data

NO	Statistik	Harga Emas (USD)
1	Jumlah data (count)	10.558
2	Rata-rata (mean)	887.21
3	Standar deviasi (std)	665.29
4	Nilai minimum (min)	252.55
5	Nilai tengah (Median)	462.86
6	Nilai maksimum (max)	3635.84

Berdasarkan hasil statistik deskriptif pada Tabel 1, data harga emas memiliki tingkat fluktuasi yang tinggi. Hal ini terlihat dari nilai standar deviasi sebesar 665,2 yang tergolong besar dibandingkan nilai rata-rata (*mean*) sebesar 887,2. Kondisi tersebut menunjukkan bahwa harga emas mengalami volatilitas yang cukup signifikan selama periode pengamatan.

Selain itu, distribusi data cenderung miring ke kanan (*right-skewed*). Hal ini dapat dilihat dari nilai rata rata yang lebih besar dibandingkan nilai tengah sebesar 462,8, serta nilai maksimum sebesar 3635,8 yang sangat jauh di atas rata-rata. Perbedaan yang cukup besar antara nilai minimum sebesar 252,5 dan nilai maksimum sebesar 3635,8 menunjukkan bahwa rentang data sangat luas dan terdapat beberapa nilai ekstrem yang memengaruhi distribusi data. Jumlah total data yang siap digunakan pada penelitian ini adalah 10.558 data. Dari data tersebut di bagi menjadi data latih dan data uji untuk model ARIMA dan LSTM. Untuk keperluan ini 80% dari keseluruhan data digunakan untuk data latih (*training data*) dan sisa 20% dari keseluruhan data digunakan sebagai data uji (*test data*).

3.2 Konfigurasi Model

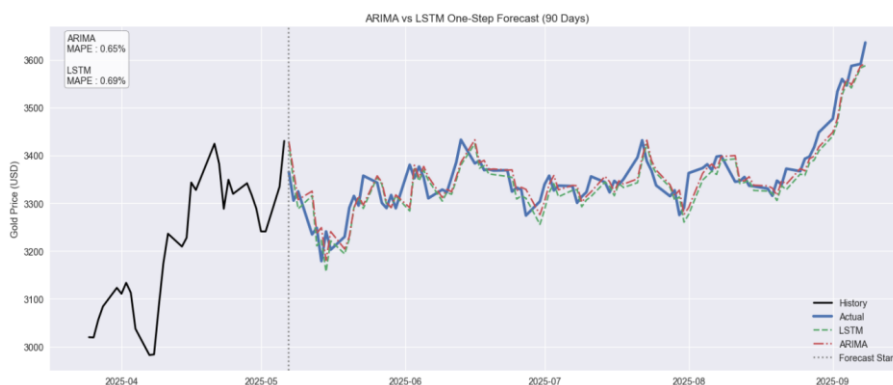
Transformasi harga absolut menjadi return harian menghasilkan deret waktu yang jauh lebih stasioner. Hal ini dikonfirmasi dengan hasil uji ADF yang menghasilkan nilai *p-value* sebesar 0,000 yang masih berada dibawah ambang batas 0.05. Stasioneritas ini menjadi prasyarat yang terpenuhi untuk pemodelan ARIMA. Proses *grid search* Auto ARIMA dengan kriteria minimalisasi AIC menghasilkan ARIMA (0,0,1) sebagai konfigurasi model terbaik. Parameter ini secara matematis merepresentasikan model *Moving Average* orde pertama MA(1), yang berarti prediksi *return* pada periode *t* hanya dipengaruhi oleh satu suku *error* masa lalu. Dengan parameter $d = 0$, model mengkonfirmasi bahwa transformasi *return* yang dilakukan pada tahap pra-pemrosesan telah secara efektif

menggantikan kebutuhan diferensiasi eksplisit. Sementara untuk model LSTM arsitektur dikonfigurasi menggunakan fitur multivariat yang mencakup *lag*, volatilitas, momentum, *silver return*, dan indeks waktu. Konfigurasi ini digunakan untuk meningkatkan kemampuan LSTM dalam menangkap hubungan non-linier tersembunyi pada data deret waktu finansial, sehingga model diharapkan bisa lebih efektif dalam peramalan harga.

Setelah proses pelatihan (*training*) pada arsitektur model ARIMA dan LSTM selesai dilaksanakan, tahap selanjutnya adalah mengevaluasi kinerja kedua model tersebut.

3.3 Pengujian

Pengujian ini dilakukan dengan menerapkan tiga skenario peramalan yang berbeda, yaitu *single-step forecasting*, *multi-step forecasting*, dan *rolling forecasting*. Pada skenario *single-step forecasting*, kedua model dievaluasi atas kemampuannya memprediksi *return* satu hari ke depan menggunakan data aktual terbaru sebagai input di setiap langkah. Hasil visualisasi perbandingan prediksi terhadap data aktual ditampilkan pada Gambar 4.

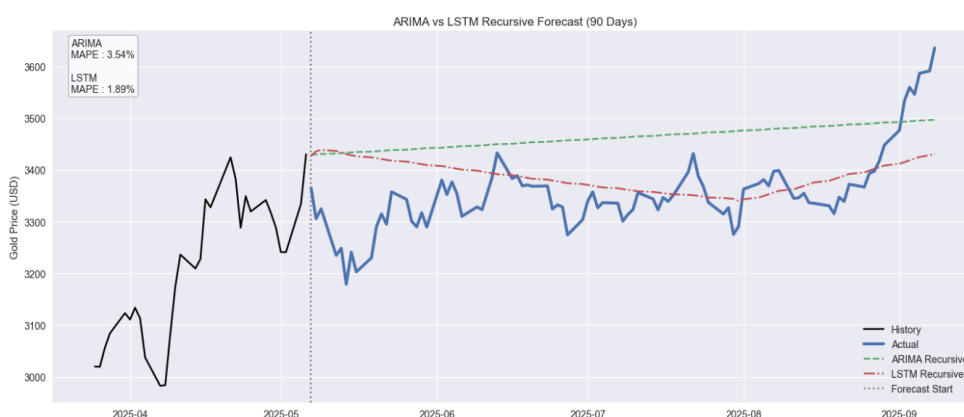


Gambar 4. Grafik Hasil Single-step Forecasting

Hasil menunjukkan perbandingan metrik evaluasi kedua model berada pada rentang akurasi yang ekuivalen dengan error yang rendah dimana hasil MAPE kurang dari 1% yang umumnya dikategorikan sebagai hasil prediksi dengan akurasi tinggi, yang berarti kedua model masih mampu untuk memprediksi dalam jangka waktu pendek atau satu langkah kedepan. Kesetaraan kinerja ini dapat dijelaskan melalui karakteristik data return harga emas pada prediksi satu langkah. Data return yang telah stasioner memiliki autokorelasi jangka pendek yang dapat ditangkap oleh komponen MA(1) pada ARIMA(0,0,1). Kondisi ini menunjukkan bahwa untuk data dengan dependensi temporal jangka pendek, model statistik sederhana dapat bersaing dengan arsitektur deep learning yang lebih kompleks. Di sisi lain, LSTM yang dirancang untuk menangkap pola jangka panjang belum dapat memanfaatkan keunggulan arsitekturnya secara penuh ketika hanya diminta memprediksi satu langkah ke depan. arsitekturnya secara penuh ketika hanya diminta memprediksi satu langkah ke depan.

Kemudian skenario kedua yaitu multi-step forecasting dengan horizon 90 hari yang memperlihatkan perbedaan kinerja yang signifikan. Dimana hasil visualisasi perbandingan prediksi ditampilkan pada Gambar 5.

Berdasarkan metrik evaluasi model LSTM menunjukkan kinerja yang lebih baik dalam meminimalkan tingkat kesalahan dibandingkan dengan model ARIMA. Hal ini dibuktikan dengan metrik evaluasi LSTM yang menghasilkan nilai MAPE yang sangat baik yaitu 1.89%. Sebagai perbandingan, model ARIMA menghasilkan nilai MAPE sebesar 3.54%. Mekanisme yang menjelaskan kesenjangan ini berpusat pada fenomena error propagation dalam peramalan rekursif. Pada skenario ini, hasil prediksi langkah $t + 1$ digunakan sebagai input untuk memprediksi Langkah $t + 2$, dan seterusnya hingga 90 hari. Setiap kesalahan prediksi pada langkah awal akan terakumulasi dan merambat ke langkah berikutnya.



Gambar 5. Grafik Hasil Skenario Multi-step Forecasting

Bagi ARIMA(0,0,1), propagasi ini sangat merusak karena model MA(1) hanya memiliki satu komponen memori: error satu langkah sebelumnya, sehingga dalam beberapa langkah prediksi ARIMA cenderung ke nilai rata-rata. Dampaknya terlihat pada Gambar 5 dimana prediksi ARIMA membentuk garis mendatar yang gagal mengikuti tren harga aktual, terutama saat terjadi lonjakan harga yang signifikan.

Sebaliknya LSTM menunjukkan ketahanan yang lebih baik terhadap error propagation. Mekanisme gating pada arsitektur LSTM khususnya forget gate dan input gate memungkinkan model untuk secara selektif mempertahankan informasi historis yang relevan dari cell state sambil mengintegrasikan sinyal baru dari setiap langkah prediksi. Kemampuan ini, dikombinasikan dengan fitur multivariat yang memungkinkan LSTM mempertahankan konsistensi tren prediksi meski input telah berganti dari data aktual ke estimasi.

Skenario yang ketiga yaitu *rolling forecast* dengan interval pembaruan data aktual setiap 7 hari yang dirancang untuk mengukur sejauh mana injeksi data aktual dapat memperbaiki akumulasi *error* yang terjadi pada *multi-step* rekursif. Hasil visualisasi perbandingan prediksi ditampilkan pada Gambar 6.



Gambar 6. Grafik Hasil Skenario Rolling Forecasting

Hasil menunjukkan peningkatan pada model LSTM yang mencatatkan hasil metrik evaluasi (MAPE 1.83%) yang lebih rendah dibandingkan scenario multi-step (MAPE 1.89%). ARIMA menghasilkan MAPE 3.52% sedikit lebih rendah dari skenario *multi step* (MAPE 3.54%). Hasil ini menjelaskan bahwa LSTM merespons pembaruan data aktual setiap 7 hari secara jauh lebih signifikan dibandingkan ARIMA. Hal ini terjadi karena setiap pembaruan jendela menghadirkan data aktual baru ke dalam *window* 60 hari, yang memungkinkan *cell state* LSTM untuk dikoreksi berdasarkan kondisi pasar terkini sebelum blok 7 hari berikutnya dimulai. Mekanisme koreksi lintasan (*trajectory correction*) ini secara berkala mereset akumulasi *error* yang telah terjadi, sehingga LSTM pada skenario *rolling* mampu mempertahankan kedekatan dengan tren aktual lebih baik daripada pada *multi-step* murni.

Sebaliknya ARIMA hanya bisa sedikit memanfaatkan pembaruan data aktual. Hal ini disebabkan oleh sifat struktural MA(1): model ini hanya memiliki satu parameter yang menghubungkan prediksi saat ini dengan *error* satu langkah sebelumnya. Ketika data aktual disuntikkan kembali setiap 7 hari, ARIMA mendapatkan titik informasi yang lebih akurat, namun keterbatasan orde model menghalanginya untuk memanfaatkan informasi dalam jendela yang lebih panjang secara efektif. Akibatnya, dalam 7 hari berikutnya, prediksi ARIMA kembali berkonvergensi ke rata-rata return historis secara cepat.

3.4 Komparasi Metrik Evaluasi

Setelah dilakukan pengujian seluruh hasil evaluasi pada ketiga skenario disajikan pada tabel 2 berikut

Tabel 2. Komparasi Keseluruhan Hasil Evaluasi

NO	Skenario	Model	RMSE	MAE	MAPE (%)	MDA (%)
1	Single-Step	ARIMA	18.8949	12.8245	0.6549	53.1769
2	Single-Step	LSTM	19.8642	13.6350	0.6925	46.7253
3	Multi-Step	ARIMA	126.0123	118.0261	3.5406	56.1798
4	Multi-Step	LSTM	85.2773	63.0708	1.8857	55.0562
5	Rolling	ARIMA	125.3095	117.2533	3.5173	57.3034
6	Rolling	LSTM	77.1543	61.7563	1.8291	46.0674

Tabel 2 menunjukkan perbandingan kinerja antara model ARIMA dan LSTM pada seluruh skenario pengujian, dimana ARIMA sedikit unggul pada skenario *single-step* dimana memiliki *error* yang lebih rendah jika dibandingkan dengan LSTM. Sebaliknya LSTM unggul pada skenario pengujian lainnya yaitu skenario *multi-step* dan *rolling forecast* dibuktikan dengan nilai metrik evaluasi tingkat kesalahan (RMSE, MAE, dan MAPE) yang lebih rendah. Dimana pada skenario *rolling* menghasilkan kompromi yang baik antara akurasi dan kemampuan adaptasi model. Pembaruan data aktual setiap tujuh hari mampu mengurangi akumulasi *error* yang umumnya terjadi pada *multi-step*

forecasting, sehingga menghasilkan performa yang lebih baik dibandingkan prediksi *multi-step* murni. Temuan ini menunjukkan bahwa frekuensi pembaruan data aktual memiliki peran penting dalam menjaga stabilitas prediksi pada horizon yang lebih panjang.

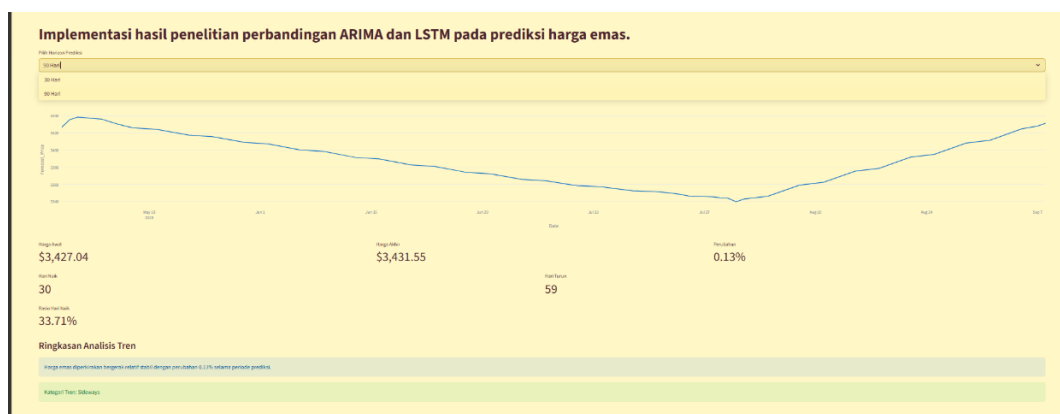
Fenomena di mana model dengan error lebih besar (ARIMA) justru lebih baik dalam mengidentifikasi arah pergerakan pasar ini dipengaruhi oleh beberapa hal. Dengan transformasi harga absolut menjadi data return yang stasioner menyebabkan data kehilangan tren jangka panjangnya dan bergerak secara fluktuatif di sekitar nilai rata-rata yang stabil. Karakteristik stasioneritas ini memunculkan pola mean-reverting berupa short-term reversal (pembalikan arah jangka pendek) yang lemah namun konsisten pada return harga emas, di mana return negatif pada hari ini memiliki kecenderungan untuk diikuti oleh return positif pada hari berikutnya. Karakteristik pembalikan ini berhasil ditangkap secara presisi oleh komponen MA(1) pada model terpilih ARIMA(0,0,1), yang memodelkan korelasi secara linier antara return saat ini dengan error pada satu langkah sebelumnya. Hal inilah yang membuat ARIMA jauh lebih sensitif terhadap perubahan arah secara instan.

Sebaliknya, rendahnya nilai MDA pada prediksi LSTM disebabkan oleh munculnya *lagging effect*, yaitu kondisi di mana kurva prediksi mengikuti pergerakan harga aktual namun dengan jeda temporal yang terlihat jelas secara visual. Fenomena ini merupakan konsekuensi struktural dari adanya *near-term dependency* pada data finansial, di mana harga hari ini merupakan prediktor terkuat untuk harga esok hari. Arsitektur LSTM dengan *window size* 60 hari secara mekanis memberikan bobot komputasi yang besar pada nilai historis terdekat untuk mengikuti momentum jangka pendek (*persistence effect*) [35]. Akibatnya, ketika terjadi pembalikan tren atau harga yang mendadak, jaringan LSTM memerlukan beberapa langkah transisi untuk menyesuaikan dan memperbarui bobot neuronnya. Kondisi inilah yang menyebabkan prediksi LSTM sangat optimal dalam mereduksi error (RMSE dan MAPE lebih kecil), namun kurang responsif dalam merespons titik balik perubahan arah jangka pendek.

Dari perspektif praktis investasi, temuan MDA ini memiliki relevansi yang tak kalah penting dari metrik *error* absolut. Seorang investor yang bertujuan membangun strategi *long-short* berbasis sinyal arah tren akan lebih diuntungkan oleh model dengan MDA yang lebih tinggi, meskipun RMSE-nya lebih besar. Dengan demikian, pemilihan model yang optimal tidak dapat diputuskan secara absolut, melainkan bergantung pada tujuan penggunaannya. Memanfaatkan LSTM untuk aplikasi yang membutuhkan akurasi nilai prediksi pada horizon menengah, dan ARIMA untuk aplikasi yang mengutamakan ketepatan sinyal arah.

3.5 Implementasi

Hasil prediksi harga emas yang diperoleh dari model terbaik pada penelitian ini diimplementasikan dalam bentuk aplikasi web sederhana menggunakan Streamlit. Aplikasi yang dikembangkan berfungsi untuk menampilkan hasil prediksi harga emas, menyajikan visualisasi pergerakan harga dalam bentuk grafik, serta memberikan analisis tren secara otomatis berdasarkan hasil prediksi yang dihasilkan model. Implementasi ini bertujuan untuk mempermudah pengguna dalam mengakses dan menginterpretasikan hasil penelitian melalui sebuah dashboard web interaktif.



Gambar 7. Antarmuka Visualisasi

Antarmuka aplikasi yang dikembangkan ditunjukkan pada Gambar 7. Aplikasi menyediakan fitur pemilihan horizon prediksi, visualisasi hasil prediksi, serta ringkasan analisis tren yang dapat digunakan sebagai pendukung interpretasi hasil peramalan.

3.6 Pembahasan

Hasil penelitian ini secara konsisten memperkuat literatur yang ada, sekaligus memperluas cakupannya melalui dimensi multi-skenario dan evaluasi MDA.

Pada domain data dengan volatilitas rendah dan pola linier, temuan penelitian ini sejalan dengan Khikmah dan Sofro [10] yang menemukan ARIMA lebih unggul pada prediksi curah hujan di Kalimantan Tengah, serta Umam dan Ardiansyah [11] yang menemukan ARIMA lebih efisien pada prediksi jumlah pengunjung perpustakaan. Penelitian-penelitian tersebut beroperasi pada data dengan kompleksitas struktural yang jauh lebih rendah dibandingkan harga emas, sehingga asumsi linearitas ARIMA dapat terpenuhi dengan baik. Dalam penelitian ini, ekuivalensi ARIMA dan

LSTM pada skenario *single-step* konsisten dengan pola tersebut: ketika horizon prediksi pendek dan data telah distasionerkan melalui transformasi *return*, model linier sederhana mampu bersaing.

Pada domain data finansial yang fluktuatif, Taslim dan Murwantara [12] menemukan bahwa LSTM lebih adaptif terhadap deret waktu dengan fluktuasi tinggi, temuan yang dikonfirmasi secara kuantitatif dalam penelitian ini melalui keunggulan LSTM pada MAPE *multi-step* (1,89% vs 3,54%) dan *rolling* (1,83% vs 3,52%). Madhika et al. [13] pada prediksi harga emas menggunakan ARIMA dan LSTM juga menemukan superioritas LSTM pada metrik error. Arham et al. [14] secara spesifik pada prediksi harga emas harian dan multi-horizon menemukan bahwa LSTM dan model hybrid ARIMA-LSTM mengungguli ARIMA pada horizon yang lebih panjang, yang konsisten dengan temuan skenario *multi-step* dan *rolling* dalam penelitian ini.

Dengan demikian, hasil penelitian ini mempertegas bahwa keberhasilan *forecasting* tidak hanya ditentukan oleh pemilihan algoritma, melainkan juga strategi peramalan yang digunakan. Jika studi sebelumnya sering kali hanya berfokus pada superioritas salah satu model, penelitian ini menunjukkan adanya alternatif yang signifikan, pemilihan horizon prediksi dan frekuensi pembaruan data terbukti memberikan pengaruh yang lebih krusial terhadap performa model dibandingkan sekadar pemilihan arsitektur model itu sendiri.

4. KESIMPULAN

Penelitian ini membandingkan kinerja model ARIMA dan LSTM dalam memprediksi harga emas menggunakan data historis selama 40 tahun melalui tiga skenario peramalan, yaitu *single-step*, *recursive*, dan *rolling forecasting*. Hasil penelitian menunjukkan bahwa kedua model memiliki performa yang ekuivalen dengan MAPE di bawah 1% pada skenario *single-step* menunjukkan bahwa model statistik linier sederhana masih kompetitif ketika horizon prediksi pendek dan data telah distasionerkan melalui transformasi *return*. Ketika horizon diperluas melalui skenario *multi-step* dan *rolling forecast*, LSTM secara konsisten unggul pada metrik kesalahan absolut dengan MAPE *multi-step* 1,89% berbanding 3,54% ARIMA, dan MAPE *rolling* 1,83% berbanding 3,52% ARIMA, membuktikan bahwa LSTM lebih tahan terhadap akumulasi *error propagation* pada prediksi jangka menengah. Meskipun demikian, hasil evaluasi menggunakan MDA menunjukkan bahwa ARIMA secara konsisten menghasilkan nilai yang lebih baik dibandingkan LSTM pada seluruh skenario pengujian. ARIMA menghasilkan MDA 53,18%, 56,18%, dan 57,30% pada ketiga skenario, sementara LSTM hanya mencapai 46,73%, 55,06%, dan 46,07%. Temuan ini mengindikasikan bahwa ARIMA lebih baik dalam mengidentifikasi arah pergerakan harga emas, sedangkan LSTM lebih unggul dalam memprediksi besaran nilai harga secara akurat. Secara keseluruhan, penelitian ini menunjukkan bahwa performa *forecasting* tidak hanya dipengaruhi oleh model yang digunakan, tetapi juga oleh strategi peramalan yang diterapkan. LSTM lebih sesuai digunakan ketika tujuan utama adalah memperoleh akurasi prediksi harga yang tinggi pada horizon yang lebih panjang, sedangkan ARIMA tetap relevan digunakan ketika fokus analisis berada pada identifikasi arah tren pergerakan pasar. Keterbatasan dalam penelitian ini terletak pada penggunaan model tunggal sehingga pada beberapa kondisi menunjukkan efek lag di mana model memerlukan waktu transisi dalam mengikuti perubahan harga yang terjadi secara mendadak. Selain itu, proses peramalan masih berfokus pada data historis harga dan beberapa indikator teknikal, sehingga belum sepenuhnya merepresentasikan faktor-faktor eksternal yang memengaruhi pergerakan harga emas. Oleh karena itu, penelitian selanjutnya dapat mengembangkan model *hybrid* atau memanfaatkan arsitektur *Transformer based* yang mampu memberi bobot adaptif pada seluruh rentang historis tanpa bias jarak, serta mempertimbangkan penggunaan variabel eksternal seperti indikator geopolitik, inflasi, suku bunga, dan nilai tukar guna meningkatkan kemampuan model dalam menangkap dinamika pasar yang lebih kompleks.

REFERENCES

- [1] A. Syauqi and I. Yusuf, "Analisis Investasi Syariah Melalui Logam Mulia (Di Toko Emas Melati Kebun Sayur Balikpapan)," *Bridg. J. Islam. Digit. Econ. Manag.*, vol. 1, no. 1, pp. 304–311, 2024, [Online]. Available: <https://journal.alshobar.or.id/index.php/bridging>
- [2] B. Heradhyaksa, "Implementasi Investasi Emas Syariah Perspektif Hukum Islam," *J. Huk. Ekon. Islam*, vol. 6, no. 1, pp. 35–51, 2022, [Online]. Available: <https://jhei.appheisi.or.id/index.php/jhei/article/download/111/60>
- [3] Z. S. Putri and N. Nur'aeni, "Analisis Fluktuasi Harga Emas dan Tingkat Inflasi terhadap Pendapatan Gadaai Emas Syariah di Bank Syariah Mandiri," *Indones. J. Econ. Manag.*, vol. 1, no. 3, pp. 489–498, 2021, doi: 10.35313/ijem.v1i3.3491.
- [4] H. Imaduddin, S. Agrippina, A. Yusuf, and M. S. Adhantoro, "A Robust Model for Early Detection of Chronic Kidney Disease Leveraging Machine Learning and Data Balancing Techniques," *Bull. Electr. Eng. Informatics*, vol. 15, no. 2, pp. 1442–1451, 2026, doi: 10.11591/eei.v15i2.11247.
- [5] A. Zulfahrizan, M. Alby Savana HSB, S. Putra Paskah Halawa, and F. Ramadhani, "Penerapan Model Hybrid Machine Learning Arima-Lstm (Long Short-Term Memory) Dalam Prediksi Harga Emas 5 Tahun Ke-Depan," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 4, pp. 5737–5741, 2025, doi: 10.36040/jati.v9i4.13914.
- [6] S. Nendi, P. P. Harjono, and H. Rizki, "Prediksi Penjualan Aerosol Menggunakan Algoritma ARIMA," *INSOLOGI J. Sains dan Teknol.*, vol. 4, no. 1, pp. 113–126, 2025, doi: 10.55123/insologi.v4i1.4868.
- [7] A. M. Zebada and E. W. Pamungkas, "Analysis of A Deep Learning Algorithm for Fracture Detection in X-Ray Images," *Int. J. Adv. Data Inf. Syst.*, vol. 6, no. 3, pp. 716–734, 2025, doi: 10.59395/ijadis.v6i3.1451.
- [8] F. Fauzi, S. Aulia, A. R. Syaifullah, and T. W. Utami, "Peramalan Harga Emas Menggunakan Pendekatan Long-Short Term Memory (LSTM)," *J. Edukasi dan Penelit. Inform.*, vol. 10, no. 2, p. 252, 2024, doi: 10.26418/jp.v10i2.78332.
- [9] I. G. T. Suryawan, K. N. Putra, P. M. Meliana, and I. G. I. Sudipa, "Performance Comparison of ARIMA, LSTM, and

- Prophet Methods in Sales Forecasting,” *Sink. J. dan Penelit. Tek. Inform.*, vol. 8, no. October, pp. 2410–2421, 2024, doi: 10.33395/sinkron.v8i4.14057.
- [10] K. N. Khikmah and A. Sofro, “Perbandingan Kinerja Model ARIMA dan LSTM untuk Peramalan Curah Hujan di Kalimantan Tengah,” *J. Ris. dan Apl. Mat.*, vol. 9, no. 2, pp. 199–207, 2025, [Online]. Available: journal.unesa.ac.id/index.php/jram
 - [11] K. Umam and Ardiansyah, “Perbandingan Metode ARIMA dan LSTM pada Prediksi Jumlah Pengunjung Perpustakaan,” *MIND (Multimedia Artif. Intell. Netw. Database) J.*, vol. 8, no. 2, pp. 119–129, 2023, doi: <https://doi.org/10.26760/mindjournal.v8i2.119-129>.
 - [12] D. G. Taslim and I. M. Murwantara, “Comparative Analysis of ARIMA and LSTM for Predicting Fluctuating Time Series Data,” *Bull. Electr. Eng. Informatics*, vol. 13, no. 3, pp. 1943–1951, 2024, doi: 10.11591/eei.v13i3.6034.
 - [13] Y. R. Madhika, Kusriani, and T. Hidayat, “Gold Price Prediction Using the ARIMA and LSTM Models,” *Sink. J. dan Penelit. Tek. Inform.*, vol. 7, no. 3, pp. 1255–1264, 2023, doi: 10.33395/sinkron.v8i3.12461.
 - [14] B. Arham, Sukasih, and R. Sani, “Prediksi Harga Emas Harian dan Multi-Horizon Menggunakan ARIMA, LSTM dan Model Hybrid ARIMA-LSTM,” *J. Sains Inform. Terap.*, vol. 5, no. 1, pp. 18–23, 2026, doi: 10.62357/joseamb.v2i2.
 - [15] C. Tian, Y. Bai, R. Tansuchat, and S. Sriboonchitta, “Nonlinear State Estimation with Deep Learning for Financial Forecasting : An EKF-LSTM Hybrid Approach with Cross-Market Evidence,” *Economies*, vol. 14, no. 5, pp. 1–31, 2026, doi: 10.3390/economies14050184.
 - [16] H. A. Nabila and E. W. Pamungkas, “Perbandingan Algoritma Machine Learning: SVM, Random Forest, dan XGBoost untuk Prediksi Stroke,” *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 1098–1110, 2025, doi: <https://doi.org/10.36341/rabit.v10i2.6444>.
 - [17] D. N. Fadhilah, K. Parmikanti, and B. N. Ruchjana, “Peramalan Return Saham Subsektor Perbankan Menggunakan Model ARIMA-GARCH,” *J. FOURIER*, vol. 13, no. 1, pp. 1–19, 2024, doi: 10.14421/fourier.2024.131.1-19.
 - [18] A. D. Milniadi and N. O. Adiwijaya, “Analisis Perbandingan Model Arima Dan Lstm Dalam Peramalan Harga Penutupan Saham (Studi Kasus : 6 Kriteria Kategori Saham Menurut Peter Lynch),” *Sibatik J.*, vol. 2, no. 6, 2023, doi: 10.54443/sibatik.v2i6.798.
 - [19] P. S. T. P. Purnama, “Optimizing Bitcoin Price Prediction with LSTM : A Comprehensive Study on Feature Engineering and the April 2024 Halving Impact,” *Elinvo (Electronics, Informatics, Vocat. Educ.)*, vol. 9, no. 1, pp. 165–177, 2024, doi: 10.21831/elinvo.v9i1.72518.
 - [20] C. Cheng, C. Li, and G. Weng, “An Improved LSTM-Based Approach for Stock Price Volatility Prediction with Feature Selection Optimization,” *Artif. Intell. Mach. Learn. Rev.*, pp. 1–15, 2023, doi: 10.69987/AIMLR.2023.40101.
 - [21] D. N. Rathnayake, O. Alele, and P. A. Louembe, “Empirical study on the volatility spillover effect of gold , silver and platinum prices,” *Financ. Stat. J.*, vol. 8, no. 1, pp. 1–19, 2025, doi: 10.24294/fsj9514.
 - [22] Y. S. Priyambudi, Zulfikar Setyo Nugroho, “An Implementation of Normalization to Predict Student Performance,” *AIP Conf. Proc.*, vol. 2926, no. 1, 2024, doi: <https://doi.org/10.1063/5.0182879>.
 - [23] Nilasari, R. E. Saputro, and G. Karyono, “Perbandingan Model LSTM dan Temporal Fusion Transformer untuk Prediksi Harga Emas,” *Build. Informatics, Technol. Sci.*, vol. 7, no. 4, pp. 2811–2820, 2026, doi: 10.47065/bits.v7i4.8204.
 - [24] Y. B. Caesar, K. Hadiono, and E. Ardianto, “Evaluasi Empiris Model ARIMA dan LSTM dalam Konteks Peramalan Penjualan Mobil Toyota,” *AITI J. Teknol. Inf.*, vol. 22, no. 2, pp. 221–235, 2025, doi: <https://doi.org/10.24246/aiti.v22i2.221-235>.
 - [25] A. L. Schaffer, T. A. Dobbins, and S.-A. Pearson, “Interrupted Time Series Analysis Using Autoregressive Integrated Moving Average (ARIMA) Models : a Guide for Evaluating Large-scale Health Interventions,” *BMC Med. Res. Methodol.*, vol. 21, no. 58, pp. 1–12, 2021, doi: <https://doi.org/10.1186/s12874-021-01235-8>.
 - [26] R. M. Sujarwo, “Penerapan Model Arima untuk Memprediksi Tren Harga TBS Sawit di Provinsi Jambi,” *Sci. J. Reflect. Econ. Accounting, Manag. Bus.*, vol. 8, no. 1, pp. 251–261, 2025, doi: 10.37481/sjr.v8i1.1040.
 - [27] H. Thamrin and M. Nu’man Normas, “Analyzing and Forecasting Admission Data using Time Series Model,” *JOIN (Jurnal Online Inform.)*, vol. 5, no. 1, pp. 35–44, 2020, doi: 10.15575/join.v5i1.546.
 - [28] M. N. A. Simanjuntak, R. Tipani, Z. M. Afriyanti, Indriyanto, and N. Hidayati, “Peramalan Jumlah Kedatangan Jalur Udara di Bandara Depati Amir Menggunakan Model Arima,” *J. Fraction*, vol. 3, no. 2, pp. 44–52, 2023, doi: 10.33019/fraction.v3i2.45.
 - [29] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice (3rd ed)*. Melbourne, Australia: OTexts, 2021. [Online]. Available: OTexts.com/fpp3
 - [30] W. N. W. M. Din and N. Shaadan, “Application of Functional Time Series Model in Forecasting Monthly Diurnal Maximum API Curves : A Comparison between Multi- Step Ahead and Iterative One-Step Ahead Approach,” *Malaysian J. Fundam. Appl. Sci.*, vol. 18, no. 2, pp. 124–137, 2022, doi: 10.11113/mjfas.v18n1.2435.
 - [31] M. N. Ulwan, Danardono, and A. Azwar, “Peramalan Penerimaan Iuran Program Jaminan Kehilangan Pekerjaan BPJS Ketenagakerjaan Menggunakan Metode Hybrid Prophet -BiLSTM,” *J. Jamsostek*, vol. 3, no. 2, pp. 174–201, 2025, doi: 10.61626/jamsostek.v3i2.123.
 - [32] H. Hewamalage, C. Bergmeir, and K. Bandara, “Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions,” *Int. J. Forecast.*, vol. 37, no. 1, pp. 388–427, 2021, doi: 10.48550/arXiv.1909.00590.
 - [33] M. F. Dixon, I. Halperin, and P. Bilokon, *Machine Learning in Finance*. Cham, Switzerland: Springer, 2020. doi: 10.1007/978-3-030-41068-1.
 - [34] S. Somiseti and G. P. Manla, “Multi-Model Ensemble Approach for Stock Price Forecasting : An Interactive Dashboard Implementation Using ARIMA , Prophet , and LSTM Models with Advanced Trading Strategy Integration,” Vellore Institute of Technology, Amaravathi (VIT-AP University), 2025. doi: <https://doi.org/10.36227/techrxiv.175322805.53639971/v1>.
 - [35] M. E. Putra and T. Oktavia, “Comparative Analysis of Temporal Fusion Transformer and Long Short- Term Memory Architecture Resilience in Predicting Solana Price Volatility Across Different Market Phases,” *J. Tek. Inform.*, vol. 7, no. 3, pp. 2935–2943, 2026, doi: 10.52436/1.jutif.2026.7.3.5894.