

Prediksi Harga Emas Menggunakan Metode Autoregressive Integrated Moving Average dan Long Short-Term Memory

Muhammad Bryan, Dr. Anita Hidayati, S.Kom., M.Kom.

Teknik Informatika, Teknik Informatika dan Komputer, Politeknik Negeri Jakarta

Depok, Jawa Barat

Email: muhammad.bryan.tik22@mhsw.pnj.ac.id

Abstract - Gold is a volatile investment instrument that requires intensive monitoring. This study develops Goltraak, an AI-based gold price tracking and prediction system. Using the Waterfall method with a separated client-server architecture, the system is built with Flutter (frontend), FastAPI (backend), and implemented on Google Cloud Platform (GCP) with an automated MLOps pipeline. Goltraak compares the statistical ARIMA model and deep learning Univariate LSTM for 7-day price projections, and provides an interactive chatbot with a Retrieval-Augmented Generation (RAG) approach to present accurate latest price information. Black Box Testing of 11 functional scenarios achieved a 100% success rate. In the forecasting model tests, ARIMA showed superior and more stable performance with a Mean Absolute Percentage Error (MAPE) of 1.21% for buying scenarios and 1.43% for selling scenarios. Meanwhile, the LSTM model produced higher error rates, at 3.42% for buying and 6.13% for selling. User Acceptance Testing (UAT) involving 10 gold investment enthusiasts showed a 94% Overall Satisfaction rate. These results prove that Goltraak is highly feasible, reliable, and useful for the public in monitoring daily gold market trends.

Keywords: Mobile Application, ARIMA, Gold Price, LSTM, RAG

Abstrak - Emas merupakan instrumen investasi fluktuatif yang memerlukan pemantauan intensif. Penelitian ini mengembangkan Goltraak, sistem pelacakan dan prediksi harga emas berbasis kecerdasan buatan (Artificial Intelligence). Menggunakan metode Waterfall dengan arsitektur server-klien terpisah, sistem ini dibangun dengan Flutter (frontend), FastAPI (backend), serta diimplementasikan pada Google Cloud Platform (GCP) dengan pipeline MLOps terotomatisasi. Goltraak membandingkan model statistik ARIMA dan deep learning Univariate LSTM untuk proyeksi harga 7 hari ke depan, serta menyediakan chatbot interaktif berpendekatan Retrieval-Augmented Generation (RAG) untuk menyajikan informasi harga terkini secara akurat. Hasil pengujian Black Box Testing terhadap 11 skenario fungsional mencatatkan tingkat keberhasilan 100%. Pada pengujian model peramalan, ARIMA menunjukkan performa yang lebih unggul dan stabil dengan nilai Mean Absolute Percentage Error (MAPE) sebesar 1,21% untuk skenario beli dan 1,43% untuk skenario jual. Sementara itu, model LSTM menghasilkan tingkat galat yang lebih tinggi, yaitu sebesar 3,42% pada skenario beli dan 6,13% pada skenario jual. Pengujian melalui User Acceptance Testing (UAT) dilakukan dengan melibatkan 10 orang responden pegiat investasi emas, menunjukkan tingkat kepuasan pengguna (Overall Satisfaction) mencapai 94%. Hasil ini membuktikan bahwa Goltraak sangat layak, andal, dan bermanfaat bagi masyarakat dalam memantau tren pasar emas harian.

Kata Kunci: Aplikasi Mobile, ARIMA, Harga Emas, LSTM, RAG

I. PENDAHULUAN

Emas merupakan salah satu komoditas investasi yang memiliki peranan vital dalam sistem ekonomi global maupun nasional. Sebagai aset safe

haven, emas menjadi instrumen utama bagi masyarakat maupun bank sentral untuk melindungi nilai kekayaan dari inflasi serta ketidakpastian geopolitik. Namun, harga emas bersifat sangat dinamis dan dipengaruhi oleh variabel kompleks

yang menuntut adanya akses informasi cepat dan akurat.

Di Indonesia, meskipun informasi harga emas tersedia melalui berbagai platform digital, akses tersebut masih bersifat pasif. Kemajuan teknologi Large Language Model (LLM) menawarkan solusi melalui chatbot, namun menghadapi kendala knowledge cutoff dan risiko halusinasi. Untuk mengatasi ini, teknologi Retrieval-Augmented Generation (RAG) hadir sebagai solusi yang mengintegrasikan pengetahuan eksternal harga emas harian secara dinamis. Selain itu, diperlukan pemrosesan data historis untuk proyeksi harga menggunakan model AutoRegressive Integrated Moving Average (ARIMA) dan Long Short-Term Memory (LSTM).

Penelitian ini bertujuan membangun sistem Goltraak (chatbot cerdas) yang mengintegrasikan LLM dengan RAG dan model peramalan ARIMA serta LSTM sebagai pembanding, memberikan asisten virtual interaktif dan prediktif bagi masyarakat.

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan model Waterfall dengan langkah-langkah terstruktur dari analisis kebutuhan hingga pengujian.

Tahap pengumpulan data harga emas harian diperoleh melalui *automated web scraping* dari situs web resmi Pegadaian. Pra-pemrosesan data tersebut meliputi penanganan nilai kosong (*missing values*) menggunakan metode Interpolasi Linear untuk menangani data hari libur bursa, serta transformasi data menggunakan *Min-Max Scaling* ke dalam skala 0 hingga 1 yang dikhususkan untuk arsitektur LSTM. Dalam perancangan model prediksinya, metode ARIMA menerapkan pengujian stasioneritas menggunakan *differencing* orde satu ($d=1$) untuk menangani tren non-stasioner pada data historis emas. Sementara itu, pemodelan LSTM menggunakan pendekatan Univariate LSTM yang dikonfigurasi dengan metode *sliding window* menggunakan parameter *lookback* sepanjang 60 hari untuk mencapai target proyeksi *direct multi-step* sepanjang 7 hari ke depan. Secara arsitektur, keseluruhan sistem ini dibagi menjadi antarmuka

Frontend menggunakan Flutter/Dart dan pusat pemrosesan *Backend* menggunakan Python/FastAPI, dengan implementasi *cloud computing* pada Google Cloud Platform. Fitur *chatbot* pada sistem ini diintegrasikan dengan Gemini LLM yang kemampuannya ditingkatkan menggunakan *framework Retrieval-Augmented Generation (RAG)*.

Keandalan fitur *chatbot* dalam menghindari fenomena halusinasi informasi (*hallucination*) bertumpu pada optimalisasi rekayasa perintah (*prompt engineering*) yang diterapkan langsung pada sisi server *backend*. Framework *Retrieval-Augmented Generation (RAG)* bekerja dengan cara mengekstraksi tiga lapis pengetahuan eksternal secara dinamis dari database, yaitu catatan harga riil 30 hari terakhir, draf proyeksi 7 hari ke depan dari model statistik ARIMA, serta draf proyeksi dari jaringan saraf LSTM. Seluruh data tersebut kemudian dibungkus ke dalam sebuah cetak biru instruksi sistem (*system prompt*) yang sangat ketat sebelum dikirimkan menuju Gemini LLM. Struktur dan semantik instruksi baku yang ditanamkan pada sistem komputer dirancang dengan format kode sebagai berikut.

""You are Goltraak AI, a helpful and professional financial assistant specialized in gold prices in Indonesia.

You speak Indonesian by default unless asked otherwise. Do not use markdown bolding too aggressively. Be conversational but precise.

IMPORTANT: Never open your replies with a greeting such as "Halo", "Halo!", "Hai", or any similar salutation. Jump straight into the answer.

CRITICAL CONTEXT:

1. Hari ini adalah tanggal {current_date}. Semua perhitungan waktu ("hari ini", "kemarin", "besok") harus merujuk pada tanggal ini.

2. Semua data harga emas yang diberikan adalah dalam satuan "per 0.01 gram" (Tabungan Emas Pegadaian). Jangan pernah berasumsi harga ini adalah per 1 gram.

Here is the actual gold price data for the last 30 days:

{history_text}

Here is our ARIMA model's predicted gold price for the next 7 days:

{arima_text if arima_text else "No ARIMA predictions available right now."}

Here is our LSTM model's predicted gold price for the next 7 days:

{lstm_text if lstm_text else "No LSTM predictions available right now."}

Your task: Answer the user's question accurately using ONLY the data provided above. If they ask about a date not in the data, apologize and say you only have recent data and 7-day predictions.

""

Melalui penerapan instruksi sistem di atas, kecerdasan buatan dipaksa untuk mengabaikan salam pembuka bumbu percakapan biasa dan langsung berfokus pada penyelesaian inti kueri pengguna. Batasan kontekstual paling kritis terletak pada penegasan satuan metrik dasar, di mana model dilarang keras mengasumsikan angka nominal sebagai harga per 1 gram, melainkan wajib menerjemahkannya sebagai harga pecahan terkecil per 0,01gram sesuai regulasi Tabungan Emas PT Pegadaian. Mekanisme pengondisian kognitif berbasis RAG inilah yang melandasi tingginya validitas jawaban chatbot serta akurasi kesesuaian respons dengan kebutuhan riil para investor.

III. HASIL DAN PEMBAHASAN

Pengujian sistem dilakukan dalam beberapa skenario, yang mencakup fungsionalitas aplikasi, performa model prediksi, keandalan *chatbot*, hingga tingkat penerimaan pengguna. Pengujian fungsionalitas sistem dilakukan menggunakan metode *Black Box Testing*. Pengujian ini bertujuan untuk memastikan bahwa seluruh antarmuka dan alur logika pada aplikasi Goltraak berjalan sesuai dengan Spesifikasi Kebutuhan Perangkat Lunak (SKPL) tanpa memandang struktur kode di dalamnya. Pengujian difokuskan pada 11 skenario utama yang mencakup fitur autentikasi, penarikan harga emas harian, integrasi model prediksi, hingga respons *chatbot* RAG. Rincian skenario dan hasil pengujian fungsional disajikan pada Tabel 1.

Tabel 1 Skenario Black Box Testing

No.	Skenario Pengujian	Hasil yang Diharapkan	Status
1	Mendaftar dengan data baru yang valid	Sistem berhasil menyimpan data pengguna, menampilkan pesan sukses, dan mengarahkan ke halaman Login.	Sukses

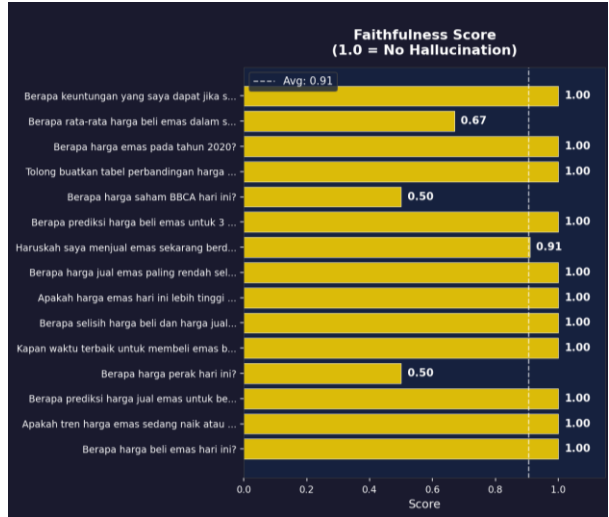
No.	Skenario Pengujian	Hasil yang Diharapkan	Status
2	Mendaftar dengan email yang sudah terdaftar	Sistem menolak proses registrasi dan menampilkan notifikasi error “Email sudah digunakan”.	Sukses
3	Login dengan kredensial yang valid	Sistem memvalidasi data, memberikan token akses, dan mengarahkan pengguna ke halaman Dashboard.	Sukses
4	Login dengan password yang salah	Sistem menolak akses dan menampilkan pesan error “Email atau password salah”.	Sukses
5	Menampilkan harga emas terkini	Sistem mengambil data dari API dan menampilkan harga emas hari ini beserta persentase perubahannya dengan benar.	Sukses
6	Menampilkan riwayat harga emas	Sistem menampilkan grafik garis (<i>line chart</i>) pergerakan harga emas sesuai dengan rentang waktu yang dipilih.	Sukses

No.	Skenario Pengujian	Hasil yang Diharapkan	Status
7	Menampilkan prediksi harga	Sistem menampilkan visualisasi grafik prediksi harga untuk beberapa hari ke depan berdasarkan model.	Sukses
8	Mengirimkan teks ke <i>chatbot</i>	Pesan tampil di layar obrolan, dan bot membalas dengan respons yang relevan sesuai konteks pertanyaan.	Sukses
9	Menerima <i>Push Notification</i>	Perangkat menerima notifikasi (FCM). Jika ditekan, pengguna akan diarahkan ke dalam aplikasi Goltraak.	Sukses
10	<i>Retrain Model</i> Prediksi	Sistem backend menarik data terbaru, menjalankan ulang proses training, dan mengembalikan status berhasil saat selesai.	Sukses
11	Mengakhiri sesi pengguna	Sistem menghapus token sesi pengguna dan mengarahkan	Sukses

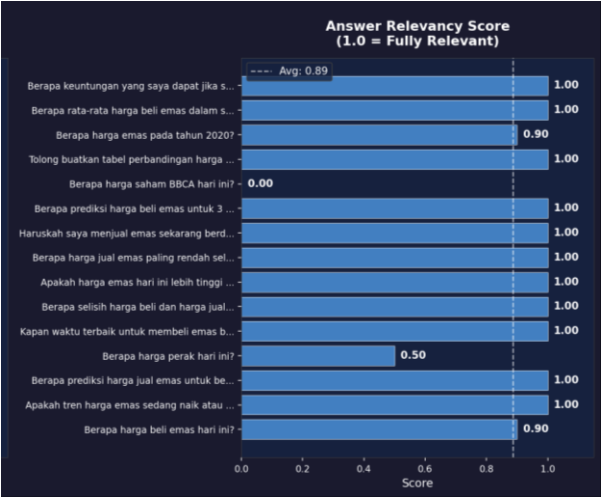
No.	Skenario Pengujian	Hasil yang Diharapkan	Status
		kembali ke halaman Login.	

Berdasarkan hasil pengujian pada Tabel 1, dapat disimpulkan bahwa keseluruhan 11 skenario pengujian berhasil dieksekusi dengan status valid atau tingkat keberhasilan 100%. Tidak ditemukan adanya *bug* atau kegagalan fungsional (*system crash*) saat pengguna berinteraksi dengan antarmuka aplikasi. Integrasi antara *frontend* Flutter dan *backend* FastAPI terbukti berjalan mulus, khususnya pada modul *chatbot* RAG yang mampu memanggil dan memproses data dari *database* secara *real-time* sesuai dengan *prompt* yang dikirimkan oleh pengguna.

Sementara itu, untuk pengujian *chatbot* RAG, evaluasi otomatis dilakukan menggunakan metrik RAGAS (*Retrieval Augmented Generation Assessment*). Seperti yang diilustrasikan pada Gambar 1 dan Gambar 2, pengujian ini mencatatkan tingkat *Faithfulness* sebesar 90,5% yang membuktikan bahwa sistem terhindar dari halusinasi data, serta mencapai tingkat *Answer Relevancy* sebesar 88,7%."



Gambar 1 Hasil Pengujian RAGAS Faithfulness Pada Chatbot



Gambar 2 Hasil Pengujian RAGAS Answer Relevancy Pada Chatbot

Selanjutnya, evaluasi kinerja model prediksi harga emas 7 hari ke depan diukur menggunakan metrik kesalahan *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE). Hasil perbandingan performa antara kedua model tersebut disajikan pada Tabel 2.

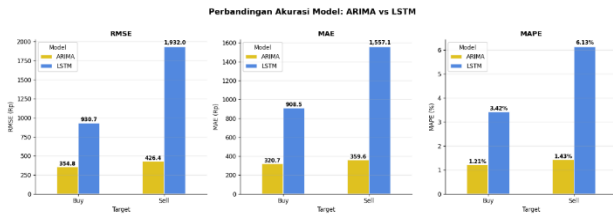
Tabel 2 Perbandingan Akurasi Model ARIMA dan LSTM

Model	Skenario	RMSE	MAE	MAPE
ARIMA	Beli	354.81	320.74	1.21%
ARIMA	Jual	426.44	359.64	1.43%
LSTM	Beli	930.66	908.52	3.42%
LSTM	Jual	1931.97	1557.06	6.13%

Berdasarkan hasil evaluasi pada tabel tersebut, model ARIMA menunjukkan tingkat galat yang lebih rendah dan konsisten dalam menangkap tren pergerakan harga yang linear. Sebaliknya, model LSTM mengalami tingkat *error* yang lebih tinggi. Hal ini dipengaruhi oleh sifat arsitektur *neural network* yang berusaha mengenali pola volatilitas ekstrem (non-linear), namun menjadi rentan terjebak *overfitting* akibat jumlah dataset historis yang masih cukup terbatas, yaitu sekitar 2.000 data observasi.

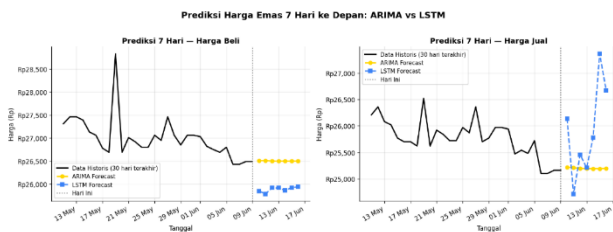
Sebagai pelengkap data dari Tabel 2, representasi visual dari komparasi tingkat kesalahan

(error) antara model ARIMA dan LSTM diilustrasikan secara lebih jelas melalui grafik batang pada Gambar 3.



Gambar 3 Perbandingan Akurasi Model ARIMA dan LSTM

Lebih lanjut, untuk mengevaluasi bagaimana garis prediksi masing-masing model merespons fluktuasi harga emas aktual di pasar, sistem melakukan proyeksi peramalan (*forecasting*) secara langsung selama 7 hari ke depan. Hasil plot kurva prediksi (*garis tren*) dari kedua arsitektur terhadap data harga riil disajikan pada Gambar 4.



Gambar 4 Perbandingan Hasil Prediksi Model ARIMA dan LSTM

Dari visualisasi garis peramalan pada Gambar 4, terlihat secara empiris pola perilaku dari masing-masing model. Model ARIMA cenderung membentuk garis proyeksi yang linear, stabil, dan konservatif. Sebaliknya, model LSTM berupaya keras mengidentifikasi dan mengikuti lonjakan volatilitas harga ekstrem secara non-linear. Namun, karena pendekatan *Deep Learning* membutuhkan asupan data historis dalam jumlah masif (puluhan hingga ratusan ribu baris), model LSTM mengalami deviasi yang lebih lebar ketika dihadapkan pada dataset Pegadaian yang baru berjumlah sekitar 2.000 observasi harian.

Khusus untuk memvalidasi pemodelan *Deep Learning*, evaluasi juga dilakukan terhadap kurva *training loss* untuk memastikan jaringan dilatih secara optimal.



Gambar 5 Training Loss Harga Beli Model LSTM



Gambar 6 Training Loss Harga Jual Model LSTM

Kurva *loss* pada Gambar 5 dan Gambar 6 menunjukkan proses konvergensi model selama tahapan *training* dan *validation*. Grafik tersebut membuktikan bahwa meskipun akurasi LSTM secara komparatif berada di bawah ARIMA pada kasus dataset ini, pelatihannya berjalan dengan stabil, penurunan *loss* terjadi secara gradual, dan model tidak mengalami indikasi *overfitting* yang parah pada akhir iterasi (*epochs*).

Evaluasi komprehensif pada tahap akhir melibatkan interaksi langsung dengan pengguna akhir (*end-user*) melalui metode *User Acceptance Testing* (UAT). Pengujian ini melibatkan 10 orang responden yang memiliki ketertarikan intensif dalam ekosistem investasi logam mulia. Pengumpulan opini dilakukan menggunakan kuesioner terstruktur berbasis skala Likert 5 poin untuk mengukur persepsi subjektif pengguna terhadap performa dan kegunaan aplikasi. Instrumen penilaian dibagi menjadi tiga dimensi utama yang merepresentasikan kualitas sistem secara menyeluruh, yaitu aspek fungsionalitas sistem, aspek kualitas antarmuka (*user interface*), serta dimensi keandalan kinerja asisten virtual. Rekapitulasi distribusi skor hasil pengisian kuesioner oleh para responden dirangkum secara mendetail pada Tabel 3.

Tabel 3 Hasil Evaluasi User Acceptance Testing Berdasarkan Aspek

No.	Aspek Evaluasi	Persentase Skor
1	Aspek Fungsionalitas	88.85%
2	Aspek Usability & Antarmuka	90.67%
3	Aspek Kinerja & Keandalan	93.00%
4	Kepuasan Pengguna	94.00%

Analisis data terhadap hasil rekapitulasi pada Tabel 3 memperlihatkan bahwa dimensi keandalan kinerja memperoleh kepuasan tertinggi dari pengguna. Hal ini mengindikasikan bahwa kemampuan kolaborasi antara kecepatan pemrosesan data di sisi *backend* dan ketepatan asisten virtual berpendekatan RAG dalam menyajikan data finansial harian telah memenuhi ekspektasi pengguna. Di sisi lain, indikator fungsionalitas sistem meskipun menempati posisi skor terendah secara komparatif, tetap berada pada koridor kategori sangat layak.

Tingginya nilai akumulasi total kepuasan pengguna yang menyentuh angka 94% mengonfirmasi secara empiris bahwa sistem yang dikembangkan tidak hanya andal dari sudut pandang arsitektur komputasi, melainkan juga memiliki tingkat keterpakaian (*usability*) yang tinggi untuk membantu masyarakat umum dalam memantau pergerakan pasar komoditas.

IV. SIMPULAN DAN SARAN

Pengembangan sistem Goltraak dengan arsitektur RAG berhasil menyajikan data harga emas dengan tingkat akurasi (*Faithfulness*) sebesar 90,5%. Pada pengujian prediksi untuk 7 hari ke depan, model ARIMA menunjukkan performa yang lebih unggul dibandingkan dengan model LSTM, yang dibuktikan

dengan perolehan nilai MAPE sebesar 1,21% pada skenario harga beli dan 1,43% pada skenario harga jual. Secara keseluruhan, sistem ini sangat layak untuk diimplementasikan dengan pencapaian tingkat kepuasan pengguna melalui uji UAT sebesar 94%.

Sebagai rekomendasi untuk pengembangan pada penelitian selanjutnya, disarankan untuk menguji dan menerapkan arsitektur Transformer yang dikhususkan untuk deret waktu, seperti TimeGPT, atau menggunakan pendekatan model *Hybrid*. Selain itu, fungsionalitas aplikasi ke depannya juga dapat diperkaya dengan mengaplikasikan fitur manajemen portofolio investasi serta peringatan pergerakan harga (*Price Alert*) yang terintegrasi secara langsung berbasis *push notification*.

V. REFERENSI

1. A. A. Adebisi, O. O. Adewumi, and C. K. Ayo, "Stock Price Prediction Using the ARIMA Model," in Proceedings of the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge: IEEE, 2014, pp. 105–111.
2. V. I. Kontopoulou et al., "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," Future Internet, vol. 15, no. 8, p. 255, 2023.
3. J. Ling and M. Afzaal, "Automatic question-answer pairs generation using pre-trained large language models in higher education," Computers and Education: Artificial Intelligence, vol. 6, p. 100252, 2024.
4. S. Vidivelli, M. Ramachandran, and A. Dharunbalaji, "Efficiency-Driven Custom Chatbot Development: Unleashing LangChain, RAG, and Performance-Optimized LLM Fusion," CMC, 2024.

5. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.