Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 4, August 2026, Page. 3130-3145
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.4.5610

Retrieval-Augmented Large Language Model Using FAISS for Stock Recommendation on LQ45 Index Based on Technical and Fundamental Analysis

Murtiyoso*[1], Imam Tahyudin[2], Berlilana[3]

[123]Magister of Computer Science, Faculty of Computer Science, Universitas Amikom Purwokerto, Indonesia

Email: [1]moertiyoso@gmail.ac.id

Received : Jan 5, 2026; Revised : Jan 5, 2026; Accepted : Jan 19, 2026; Published : Aug 14, 2026

Abstract

The analysis of Indonesian stock markets is inherently complex due to market volatility, heterogeneous financial indicators, and the need to integrate both technical and fundamental information. Traditional analytical approaches often struggle to provide transparent and consistent investment recommendations, while Large Language Models (LLMs) may suffer from hallucination when applied to structured financial prediction tasks. This study proposes a Retrieval-Augmented Generation (RAG) framework based on the LLaMA model and FAISS to enhance the reliability of stock investment recommendations for LQ45-listed companies. Historical OHLCV data and fundamental financial reports are collected from Yahoo Finance and preprocessed into a structured knowledge base using embedding representations. Relevant information is retrieved using a top-k similarity search mechanism and integrated into a structured prompt engineering scheme to generate BUY, HOLD, or SELL recommendations. Experimental results demonstrate that the proposed approach achieves an F1-score of 0.76 for classification performance, with regression-based error metrics of MAE 299.38 and MAPE 10.06%. In addition, text-based evaluation using ROUGE yields a high score of 0.9906, indicating strong alignment between generated explanations and reference analyses. The findings suggest that the RAG framework significantly reduces hallucination by grounding LLM outputs in retrieved financial data, thereby improving the interpretability and robustness of AI-driven stock analysis. This research contributes to the field of informatics by providing a practical and extensible RAG-based framework for structured financial prediction tasks.

Keywords : *FAISS, Investment Recommendation, Large Language Model, LQ45 Stocks, Retrieval-Augmented Generation, Stock Analysis*

This work is an open access article licensed under a Creative Commons Attribution 4.0 International License.

1. PENDAHULUAN

Pemilihan saham sebagai instrumen investasi merupakan keputusan yang kompleks dan penuh tantangan karena dipengaruhi oleh berbagai faktor, seperti tingkat risiko, potensi imbal hasil, serta dinamika pasar yang terus berubah. Saham dikenal sebagai instrumen investasi dengan tingkat volatilitas yang tinggi, di mana pergerakan harga saham dipengaruhi oleh banyak faktor. Harga saham tentu tidak hanya dipengaruhi oleh kinerja internal perusahaan, masih ada faktor lainnya, seperti faktor eksternal seperti kondisi makroekonomi dan sentimen pasar. Seiring dengan meningkatnya jumlah investor di pasar modal Indonesia dalam beberapa tahun terakhir, kebutuhan akan sistem pendukung keputusan yang mampu membantu investor dalam memilih saham yang optimal menjadi semakin penting. [1][2]

Beberapa metode sudah digunakan untuk memahami dan memproyeksikan pergerakan harga saham, seperti metode analisis fundamental dan metode analisis teknikal. Analisis fundamental menilai prospek emiten melalui kondisi makro, kinerja perusahaan, serta laporan keuangan; sedangkan analisis teknikal memanfaatkan dinamika historis harga dan volume untuk membaca kecenderungan pasar. Kendati umum dipakai, kedua pendekatan ini sering menghadapi kendala saat berhadapan dengan

volatilitas tinggi, pola nonlinier, dan arus data pasar yang bising, sehingga pemisahan sinyal yang benar-benar relevan dari noise menjadi tantangan utama.

Perkembangan Large Language Model (LLM), membuka peluang untuk menghasilkan penjelasan analitis yang lebih fleksibel dan kontekstual. Namun pada domain finansial, LLM masih menunjukkan keterbatasan pada penalaran kuantitatif dan sintesis lintas sumber—padahal kedua aspek tersebut krusial karena materi finansial sering memadukan teks dengan angka, tabel, serta struktur dokumen yang kompleks. [3][4]

Studi terbaru memperlihatkan bahwa penguatan retrieval memang penting untuk tugas finansial, tetapi fokus riset RAG masih dominan pada data tekstual. FinRAGBench-V menyoroti bahwa banyak riset RAG finansial berorientasi pada teks dan berpotensi kehilangan insight penting ketika informasi dari tabel/grafik tidak terakomodasi dengan baik; selain itu, model juga masih kesulitan dalam penalaran numerik pada konteks tabel/grafik. Sementara itu, FinDoc-RAG menunjukkan bahwa dokumen finansial memiliki kandungan numerik yang padat dan menuntut quantitative reasoning serta sintesis lintas dokumen; performa sistem RAG cenderung turun signifikan ketika berpindah dari ekstraksi fakta ke sintesis multi-dokumen.[5][6][7][8]

Pada konteks saham, gap riset menjadi lebih spesifik. Banyak pendekatan LLM untuk analisis pasar masih bertumpu pada data tidak terstruktur seperti berita atau media sosial, sehingga pemanfaatan data numerik terstruktur dan pola deret waktu historis sering belum menjadi fokus utama. FinSrag (2025) menegaskan bahwa tantangan peramalan saham adalah mengekstraksi pola halus namun berdampak dari deret waktu yang bising, sementara metode retrieval konvensional sering kesulitan menangkap ketergantungan temporal dan konteks yang relevan untuk prediksi yang stabil. Berdasarkan celah tersebut, penelitian ini bertujuan mengembangkan sistem rekomendasi saham berbasis RAG yang memanfaatkan data numerik terstruktur (CSV) untuk menghasilkan keputusan BUY/HOLD/SELL dan target harga yang data-grounded pada saham LQ45. Sistem mengintegrasikan FAISS untuk retrieval dan LLM (Llama) untuk generasi rekomendasi. Untuk efisiensi adaptasi model, penelitian ini menerapkan Parameter-Efficient Fine-Tuning melalui LoRA, sehingga penyesuaian perilaku model dapat dilakukan dengan biaya komputasi yang lebih rendah dibandingkan pelatihan ulang penuh.[9][10][11]

2. METODE

Pada bab metode ini dijelaskan tentang rancangan yang dilakukan selama penelitian, data yang diambil beserta sumbernya, tahapan praproses, konstruksi knowledge base, mekanisme retrieval berbasis FAISS, strategi prompting, serta metode evaluasi untuk menilai kinerja sistem pemilihan saham berbasis Retrieval-Augmented Generation (RAG). Fokus utama metodologi adalah memastikan keluaran LLM ter-grounding pada konteks historis (teknikal) dan fundamental yang relevan untuk saham dan periode yang dianalisis.

2.1. Large Language Model

Large Language Model (LLM) merupakan model pembelajaran mesin berskala besar dan dibangun dengan arsitektur dari model Transformer untuk memproses dan menghasilkan teks, serta mengikuti instruksi pengguna berdasarkan konteks yang diberikan. Pada penelitian ini, LLM berperan sebagai komponen generatif yang menghasilkan keputusan rekomendasi investasi (BUY/HOLD/SELL) dan penjelasan analitis yang dirujuk dari konteks hasil retrieval. Model yang digunakan berasal dari keluarga Llama 4 yang mengadopsi arsitektur Mixture of Experts (MoE) dan mendukung pemrosesan multimodal. Agar klaim teknis konsisten dengan implementasi, perlu ditegaskan varian model yang dipakai: pada layanan serving (misalnya Groq), Llama 4 Maverick Instruct memiliki context window 131.072 token. console.groq.com Sementara klaim konteks hingga 10 juta token lebih tepat dikaitkan

Vol. 7, No. 4, August 2026, Page. 3130-3145
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.4.5610

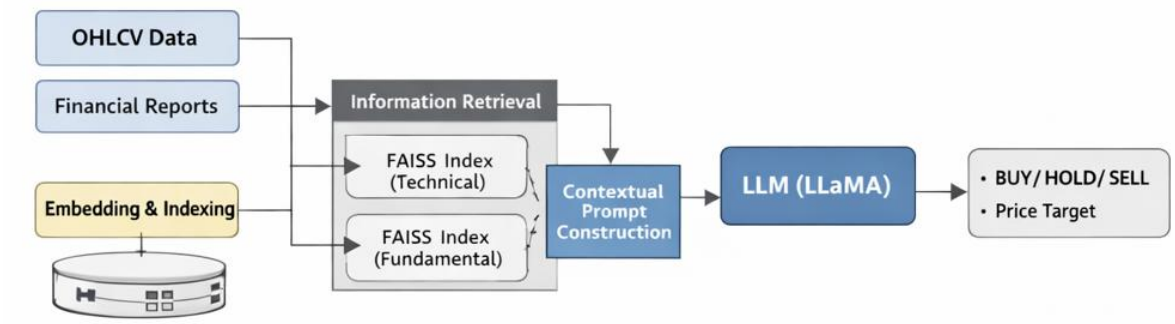
pada varian tertentu (misalnya Llama 4 Scout pada kanal yang mampu menangani konteks sangat panjang). [12][13]

Untuk efisiensi adaptasi pada tugas rekomendasi saham, penelitian ini mengadopsi Parameter-Efficient Fine-Tuning (PEFT) agar penyesuaian model dapat dilakukan dengan kebutuhan komputasi lebih rendah dibanding full fine-tuning. Survei PEFT menekankan bahwa PEFT menyesuaikan model untuk tugas/domain spesifik dengan meminimalkan jumlah parameter tambahan dan sumber daya komputasi yang diperlukan.Teknik yang digunakan adalah LoRA, yang membekukan bobot utama dan melatih parameter tambahan berukuran kecil pada lapisan-lapisan yang ada.[14]

2.2. Retrieval Augmented Generation

Retrieval Augmented Generation adalah menggabungkan dua komponen utama: modul retriever dan modul generator. Modul retriever bertugas untuk mencari dan mengambil informasi yang paling relevan dari knowledge base, sementara modul generator (biasanya berupa Large Language Model/LLM) menggunakan informasi tersebut untuk menghasilkan jawaban atau teks yang lebih terinformasi dan akurat. Dalam penelitian ini, RAG digunakan untuk memastikan bahwa rekomendasi investasi yang dihasilkan Llama ter-grounding pada data historis harga saham dan laporan keuangan perusahaan, mengurangi risiko hallucination yang dapat terjadi pada model generatif. [15][16]

Secara umum gambar 2 menjelaskan mengenati teknik Retrieval Augmented Generation. Dengan memanfaatkan teknik similarity search, modul retriever memilih data yang relevan dan memberikan konteks yang diperlukan bagi LLM untuk menghasilkan analisis yang lebih berbasis bukti. Pendekatan ini memungkinkan model untuk menghasilkan rekomendasi yang lebih konsisten dan transparan, berfokus pada data teknikal dan fundamental yang relevan. Dengan demikian, RAG meningkatkan akurasi dan kredibilitas rekomendasi investasi, serta memastikan bahwa keputusan yang dihasilkan selalu didasarkan pada data yang dapat diverifikasi.



Gambar 1. Alur Retrieval Augmented Generation

2.3. FAISS (Retrieval)

Knowledge base pada penelitian ini dibangun dari dua sumber data terstruktur berformat CSV, yaitu (i) data historis perdagangan saham (fitur OHLCV beserta indikator teknikal) dan (ii) data fundamental emiten dari laporan keuangan. Setiap potongan informasi (knowledge chunk) dikonversi menjadi teks terstruktur, kemudian dipetakan ke representasi vektor menggunakan model embedding. Selanjutnya, seluruh vektor diindeks menggunakan FAISS (Facebook AI Similarity Search) untuk memungkinkan pencarian konteks relevan secara efisien pada ruang berdimensi tinggi melalui mekanisme nearest neighbor search.[17][18]

Pada tahap retrieval, sistem membentuk query embedding dari kebutuhan analisis (misalnya kondisi teknikal terbaru dan/atau ringkasan fundamental emiten), lalu mengambil Top-k chunk dengan jarak paling dekat terhadap query. Parameter k berfungsi sebagai hiperparameter yang mengontrol

jumlah konteks yang dimasukkan ke prompt: nilai k yang terlalu kecil dapat menyebabkan konteks penting terlewat, sedangkan k yang terlalu besar berpotensi menambah noise dan memperpanjang konteks, sehingga menurunkan efisiensi dan konsistensi keluaran model. Untuk menjaga relevansi konteks dan mengurangi pencampuran informasi yang tidak sejenis, penelitian ini menerapkan dua indeks FAISS terpisah, masing-masing untuk knowledge base teknikal (harga/indikator) dan fundamental (laporan keuangan). Dengan pemisahan indeks ini, retrieval dapat dilakukan secara lebih terarah sesuai kebutuhan analisis, dan konteks yang diberikan ke LLM menjadi lebih terstruktur.

2.4. PEFT / LoRa

Untuk menyesuaikan model LLM pada tugas rekomendasi saham tanpa biaya komputasi yang tinggi, penelitian ini mempertimbangkan Parameter-Efficient Fine-Tuning (PEFT). Pendekatan PEFT berfokus pada pelatihan sejumlah kecil parameter tambahan sehingga adaptasi model dapat dilakukan lebih efisien dibanding full fine-tuning, sekaligus lebih praktis untuk eksperimen berulang pada domain finansial yang dinamis. Metode PEFT yang digunakan adalah LoRA (Low-Rank Adaptation), yaitu teknik yang mempertahankan bobot utama model tetap (frozen) dan menambahkan komponen adaptasi berukuran kecil pada lapisan tertentu. Dengan cara ini, penyesuaian perilaku model—misalnya konsistensi keputusan BUY/HOLD/SELL dan gaya penjelasan berbasis konteks—dapat dicapai dengan kebutuhan memori dan waktu pelatihan yang lebih rendah.

Setelah kerangka konseptual ditetapkan, penelitian ini mengimplementasikan pipeline end-to-end untuk menghasilkan rekomendasi saham berbasis RAG pada saham-saham LQ45. Pipeline dimulai dari pengumpulan dan praproses data, dilanjutkan dengan konstruksi knowledge base serta pengindeksan untuk mendukung retrieval berbasis FAISS. Konteks hasil retrieval kemudian digunakan dalam tahap generasi rekomendasi melalui strategi prompting pada LLM, dengan opsi adaptasi perilaku model menggunakan PEFT (LoRA). Kinerja sistem dievaluasi secara kuantitatif dan kualitatif menggunakan metrik klasifikasi, prediksi numerik, dan kualitas narasi.[19][20]

2.5. Data Collection

Penelitian ini menggunakan data harga saham yang berasal dari Yahoo Finance dan mencakup saham-saham yang tergabung dalam indeks LQ45. Data historis harga saham harian dikumpulkan dalam format CSV dengan atribut Open, High, Low, Close, dan Volume (OHLCV) untuk periode 1 Januari 2021 hingga 30 September 2025. Selain data teknikal, data fundamental perusahaan berupa laporan keuangan tahunan (misalnya pendapatan, laba bersih, total aset, dan rasio keuangan utama) juga dikumpulkan dalam format CSV untuk memperkaya konteks analisis.[21]

2.6. Preprocessing Data

Tahapan preprocessing dilakukan pada data harga maupun data fundamental untuk memastikan konsistensi dan kesiapan data sebelum dikonstruksi menjadi knowledge base. Pertama, dilakukan pembersihan data untuk memastikan kelengkapan kolom yang dibutuhkan serta menangani nilai hilang (missing values) dan nilai tak hingga (infinity). Nilai hilang diimputasi menggunakan rata-rata kolom terkait untuk menjaga stabilitas perhitungan fitur.

Kedua, dilakukan penyelarasan tanggal karena kalender perdagangan dapat berbeda antar emiten. Penyelarasan ini memastikan setiap jendela waktu analisis memiliki panjang observasi yang konsisten, sehingga proses chunking dan evaluasi antar saham dapat dilakukan secara adil. Ketiga, dilakukan transformasi fitur teknikal dengan menghitung indikator seperti SMA, EMA, RSI, MACD, Bollinger Bands, ATR, dan VWAP untuk memperkaya representasi dinamika harga. Terakhir, hari tanpa transaksi (misalnya libur bursa) diperlakukan sebagai non-trading days sesuai kebutuhan pipeline, agar tidak mengganggu pembentukan jendela waktu.

2.7. Konstruksi Knowledge Base

Konstruksi knowledge base bertujuan mengubah data mentah menjadi unit pengetahuan terstruktur (knowledge chunks) yang dapat ditelusuri dan diambil kembali saat inferensi. Pada tahap ini, sistem membentuk konteks teknikal (misalnya ringkasan harga terkini, return, volatilitas, drawdown, dan indikator) serta konteks fundamental (misalnya laba bersih dan rasio utama) untuk memastikan LLM menerima informasi yang relevan terhadap saham dan periode analisis. Fungsi build_context digunakan untuk menyusun hasil retrieval teknikal dan fundamental menjadi konteks terstruktur yang siap dipakai pada tahap generasi rekomendasi. [22][23]

2.8. Chunking Berbasis Time Window

Data yang telah dipraproses dipecah menjadi unit-unit kecil berbasis *rolling time window* (misalnya 30 atau 60 hari). Setiap chunk merepresentasikan satu unit pengetahuan yang memuat ringkasan statistik harga (return rata-rata, volatilitas, drawdown) dan, jika tersedia, ringkasan indikator fundamental pada periode yang sama. Strategi ini memungkinkan pengelolaan konteks historis secara terstruktur dan konsisten, serta memudahkan pemetaan konteks terhadap horizon evaluasi.[24][25]

2.9. Embedding

Setiap knowledge chunk dikonversi ke representasi vektor menggunakan model sentence-transformers/all-MiniLM-L6-v2. [10] Proses embedding ini memungkinkan sistem melakukan pencarian kesamaan semantik terhadap kueri yang menggambarkan kondisi teknikal dan/atau fundamental suatu saham pada waktu tertentu. Pada implementasi, encoding teks dihasilkan dalam format float32 agar kompatibel dengan mesin indeks vektor dan efisien saat komputasi similarity search.

2.10. Indexing FAISS

Embedding yang dihasilkan kemudian diindeks menggunakan FAISS melalui struktur IndexFlatL2 (exact nearest neighbor search berbasis jarak Euclidean/L2) agar retrieval dapat dilakukan secara deterministik dan ringan tanpa pelatihan ulang indeks. Untuk menjaga pemisahan konteks dan meningkatkan presisi retrieval, penelitian ini membangun dua indeks terpisah, yaitu indeks teknikal (berbasis harga/indikator) dan indeks fundamental (berbasis laporan keuangan). Selain indeks, disimpan pula metadata (misalnya ticker, rentang tanggal, dan fitur terkait) sehingga indeks dapat digunakan kembali tanpa membangun ulang pada eksekusi berikutnya. [17]

2.11. Retrieval (FAISS Top-k)

Tahap retrieval bertujuan mengambil konteks paling relevan dari vector index berdasarkan kueri yang dibentuk oleh sistem atau pengguna, misalnya “kondisi teknikal terbaru” atau “kondisi fundamental terbaru” untuk ticker tertentu. Proses retrieval dilakukan dengan mengubah query text menjadi embedding, kemudian menjalankan pencarian nearest neighbor pada indeks terkait untuk memperoleh kandidat konteks teratas berdasarkan kedekatan jarak L2.

Parameter Top-k ditetapkan eksplisit pada modul retrieval (misalnya top_k = 5) sehingga sistem selalu mengambil sejumlah konteks teratas yang paling dekat secara semantik, dan nilainya dapat disesuaikan pada eksperimen tanpa mengubah desain pipeline. Untuk evaluasi per emiten, retrieval dilakukan secara lebih longgar terlebih dahulu (misalnya mengambil lebih banyak kandidat), kemudian disaring berdasarkan ticker agar konteks yang digunakan benar-benar spesifik pada saham yang sedang dianalisis. Strategi ini menjaga retrieval tetap efisien namun lebih presisi, sekaligus mengurangi risiko tercampurnya konteks lintas emiten yang dapat mengganggu koherensi narasi maupun keputusan rekomendasi.

2.12. Generation (Prompting + RAG)

Pada tahap generasi, pendekatan Retrieval-Augmented Generation (RAG) digunakan untuk mengintegrasikan konteks hasil retrieval dengan kemampuan generatif LLM. Konteks Top-k yang diperoleh dari FAISS disusun menjadi prompt terstruktur yang memuat ringkasan teknikal dan fundamental, lalu diberikan sebagai masukan ke model Llama untuk menghasilkan rekomendasi investasi BUY, HOLD, atau SELL beserta narasi analitis pendukung. Dengan demikian, RAG diperlakukan sebagai mekanisme integrasi antara retrieval dan generasi yang memastikan keluaran model tetap grounded pada evidence yang ditarik dari knowledge base. Prompt engineering diterapkan untuk meningkatkan konsistensi dan relevansi keluaran. Desain prompt disusun agar informasi teknikal dan fundamental disajikan jelas dan fokus, sehingga model menerima konteks yang tepat untuk menghasilkan keputusan dan alasan rekomendasi. Struktur prompt juga dioptimalkan agar ringkas, menghindari konteks yang tidak relevan, serta mengurangi risiko hallucination dengan memastikan narasi merujuk pada data yang tersedia. Dengan strategi prompting yang tepat, keluaran model diharapkan lebih stabil, dapat dipertanggungjawabkan, dan selaras dengan konteks historis serta fundamental yang digunakan dalam sistem. [26][27]

Dengan penerapan prompt engineering yang tepat, Llama mampu menghasilkan rekomendasi investasi (BUY/HOLD/SELL) yang lebih konsisten dan relevan. Proses ini membantu model menghasilkan rekomendasi yang didukung oleh data historis yang valid, seperti harga saham dan laporan keuangan perusahaan, yang memastikan keputusan investasi yang lebih terinformasi dan berbasis bukti.

2.13. Evaluasi

Evaluasi kinerja sistem dilakukan dari tiga aspek: (1) klasifikasi keputusan investasi (BUY/HOLD/SELL), (2) prediksi numerik (harga/target), dan (3) kualitas narasi analitis. Untuk menguji efektivitas sistem, keluaran model dibandingkan dengan *ground truth* yang dibentuk dari realisasi return pada horizon evaluasi tertentu. Label *ground truth* ditetapkan sebagai BUY jika return melebihi ambang positif T_{buy}, SELL jika return berada di bawah ambang negatif T_{sell}, dan HOLD untuk kondisi selain keduanya. Nilai T_{buy} dan T_{sell} ditentukan secara empiris untuk menjaga keseimbangan kelas dan mengurangi bias akibat pergerakan ekstrem.

2.13.1. Evaluasi Klasifikasi

Kinerja klasifikasi BUY/HOLD/SELL diukur menggunakan Precision, Recall, dan F1-score. F1-score dirumuskan sebagai:

$$\mathrm{F1-score} = 2x\ \frac{(Precisson\ x\ Recall)}{(Precisson+Recall)} \qquad (1)$$

Di mana: Precision mengukur proporsi prediksi positif yang benar (misalnya, prediksi BUY yang benar-benar harga saham naik). Recall mengukur seberapa banyak kasus positif yang berhasil diprediksi oleh model.

Pada penelitian ini, F1-score digunakan untuk mengevaluasi kinerja klasifikasi BUY/HOLD/SELL, karena metrik ini mempertimbangkan keseimbangan antara precision dan recall, yang sangat penting ketika data memiliki distribusi kelas yang tidak seimbang, seperti pada pasar saham.Penggunaan F1-score untuk mengukur keseimbangan antara precision dan recall dalam klasifikasi keputusan investasi (buy/hold/sell). Pada penelitian ini, label ground-truth dibentuk dari return realisasi pada horizon evaluasi: BUY jika return > +T_buy, SELL jika return < −T_sell, dan HOLD selain itu. Nilai T_buy dan T_sell mengikuti konfigurasi run eksperimen terakhir dan seluruh label dinormalisasi menjadi {BUY, HOLD, SELL}. Pada penelitian ini, label *ground-truth* BUY,

HOLD, dan SELL ditentukan berdasarkan return realisasi pada horizon evaluasi tertentu. Keputusan BUY diberikan jika return melebihi ambang positif (T_buy), SELL jika return berada di bawah ambang negatif (T_sell), dan HOLD untuk kondisi selain keduanya. Nilai ambang T_buy dan T_sell ditentukan secara empiris berdasarkan distribusi return historis untuk menjaga keseimbangan antar kelas dan menghindari dominasi satu kelas tertentu. Pendekatan ini bertujuan memastikan proses evaluasi mencerminkan kondisi pasar yang realistis serta mengurangi bias yang dapat timbul akibat pergerakan harga ekstrem.

2.13.2. Evaluasi Prediksi Harga

Akurasi prediksi harga saham dievaluasi menggunakan Mean Absolute Error (MAE) dan Mean Absolute Percentage Error (MAPE):

$$\text{MAE} = \frac{1}{n} \sum_{i-1}^{n} | \text{yi} - \text{y\^i} | \quad (2)$$

yi adalah nilai harga saham aktual (nyata).
$\text{y\^}^{\text{i}}$ adalah nilai harga saham yang diprediksi oleh model.
n adalah jumlah data (periode waktu yang diamati).

$$\text{MAPE} = \frac{1}{n} \sum_{i-1}^{n} \left| \frac{\text{yi} - \text{y\^i}}{\text{yi}} \right| x\ 100\% \quad (3)$$

yi adalah nilai harga saham aktual.
$\text{y\^}^{\text{i}}$ adalah nilai harga saham yang diprediksi oleh model.
n adalah jumlah data.

MAPE memberikan gambaran yang jelas tentang akurasi prediksi dalam bentuk persentase, dan lebih mudah untuk dibandingkan antara data dengan skala yang berbeda. MAPE yang lebih rendah menunjukkan prediksi yang lebih akurat.

2.13.3. Evaluasi Narasi

Kualitas narasi analitis yang dihasilkan LLM dievaluasi menggunakan metrik ROUGE untuk mengukur kesesuaian leksikal dan struktur kalimat terhadap teks referensi profesional. Metrik ROUGE digunakan untuk mengukur kesesuaian leksikal dan struktur kalimat antara narasi yang dihasilkan model dan teks referensi. ROUGE-N mengukur kecocokan n-gram, untuk ROUGE-L digunakan untuk mengevaluasi kesamaan urutan kata yang paling panjang (Longest Common Subsequence). Evaluasi ini memiliki tujuan untuk menilai sejauh mana model mampu menghasilkan narasi yang dihasilkan Large Language model dan selaras dengan gaya penulisan dan terminologi yang lazim digunakan dalam analisis pasar saham profesional, bukan untuk menilai kreativitas bahasa atau kebenaran faktual secara langsung.[16]

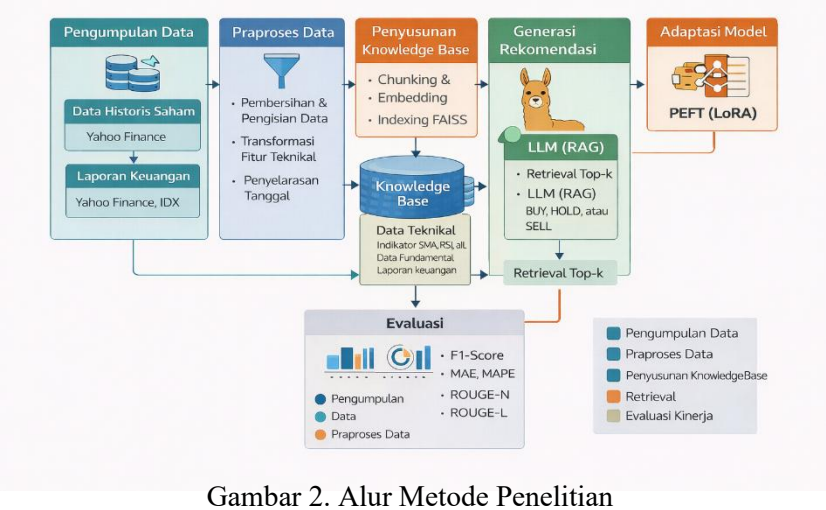
$$\text{ROUGE-N} = \frac{\sum \text{n-gramCount of Overlap n-grams}}{\sum \text{n-gramCount of Reference n-grams}} \quad (4)$$

ROUGE-N menghitung kesesuaian antara n-gram (urutan kata sepanjang n) yang ada dalam kalimat yang dihasilkan dan kalimat referensi. N-gram ini bisa berupa unigram (1 kata), bigram (2 kata), trigram (3 kata), dan seterusnya. ROUGE-N digunakan untuk mengukur seberapa banyak kata-kata yang serupa antara kalimat yang dihasilkan oleh model dan kalimat referensi.
Mengukur persentase kecocokan n-gram antara hasil yang dihasilkan model dan kalimat referensi.

$$\text{ROUGE-L} = \frac{length\ of\ reference\ sentence}{LCS(generated, reference)} \quad (5)$$

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 4, August 2026, Page. 3130-3145
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.4.5610

ROUGE-L mengukur kesesuaian urutan kata yang panjang (Longest Common Subsequence, LCS). LCS adalah subsekuens terpanjang yang muncul dalam kedua kalimat (kalimat hasil dan kalimat referensi). ROUGE-L digunakan untuk menilai kesesuaian urutan kata, bukan hanya kecocokan n-gram.



Gambar 2. Alur Metode Penelitian

Gambar 2 di atas menggambarkan alur proses keseluruhan yang digunakan dalam penelitian ini, dimulai dari Pengumpulan Data hingga Evaluasi Kinerja. Setiap tahapan dalam proses ini diwakili oleh blok yang terhubung dengan panah, menunjukkan hubungan dan urutan antar tahapan. Diagram ini memberikan gambaran visual yang jelas mengenai bagaimana data dikumpulkan, diproses, dan digunakan untuk menghasilkan rekomendasi saham berbasis Retrieval-Augmented Generation (RAG).

Secara rinci, proses dimulai dengan pengumpulan data saham yang mencakup data teknikal dan fundamental, dilanjutkan dengan preprocessing data untuk memastikan kualitas dan konsistensi. Kemudian, data yang telah diproses digunakan untuk membangun knowledge base, yang selanjutnya digunakan dalam tahap retrieval untuk mencari informasi relevan. Hasil dari retrieval ini digunakan dalam proses generasi rekomendasi melalui teknik prompting pada model LLM, dengan tambahan PEFT (LoRA) untuk adaptasi model agar lebih efisien dalam menghasilkan keputusan BUY, HOLD, atau SELL. Terakhir, sistem dievaluasi menggunakan metrik yang relevan untuk mengukur kinerjanya dalam memberikan rekomendasi yang akurat dan dapat diandalkan.

3. HASIL

Pada pembahasan hasil, dataset yang digunakan dalam penelitian ini terdiri dari data harga saham yang masuk ke dalam indeks LQ45 yang diperoleh dari Yahoo Finance, kolom yang digunakan pada penelitian ini adalah harga *open, high, low, close, volume* (OHLCV) serta data lainnya yang merupakan fundamental perusahaan. Data dikumpulkan dalam format CSV dan selanjutnya melalui tahap pra-pemrosesan yang meliputi pembersihan nilai hilang, normalisasi numerik, serta segmentasi data menjadi *knowledge chunks* untuk membentuk basis pengetahuan terstruktur. Tahapan ini bertujuan memastikan bahwa data numerik dapat direpresentasikan secara konsisten sebelum dilakukan embedding dan proses retrieval. Distribusi harga saham menunjukkan variasi volatilitas antar emiten, yang menegaskan kompleksitas prediksi pada pasar saham Indonesia.

Data harga saham periode 01 Januari 2020–30 September 2025 digunakan sebagai basis informasi model. Evaluasi dilakukan secara out-of-sample dengan memanfaatkan data harga aktual periode 1 Oktober–30 November 2025 untuk menghitung return realisasi, sehingga sistem rekomendasi dapat dinilai kinerjanya melalui metrik F1-score (klasifikasi BUY/HOLD/SELL) serta MAE dan MAPE untuk deviasi prediksi harga terhadap harga pasar.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 4, August 2026, Page. 3130-3145
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.4.5610

Tabel 1. Contoh Data Saham TLKM

Date	Open	High	Low	Close	Volume
2020-01-02	3.970	4.000	3900	3910	52.094.000
2021-05-18	3190	3190	3150	3180	77.310.400
2022-07-01	4120	4190	4090	4170	68.415.100
2023-09-07	3780	3790	3730	3760	63.054.500
2024-11-15	2500	2580	2500	2540	103.680.400
2025-09-25	3160	3200	3110	3120	37.627.500

Data yang kedua adalah berupa laporan keuangan dari saham TLKM, mulai dari laba, beban operastional dan lainnya. Seperti dimuat pada Tabel 2 yang merupakan laporan keuangan dari saham TLKM.

Tabel 2. Contoh Laporan Keuangan TLKM

Komponen	2024-12-31	2023-12-31	2022-12-31	2021-12-31
Tax Effect Of Unusual Items	69,776,086,623	-168,620,927,671	-1,478,046,115,838	767,426,781,590
Tax Rate For Calcs	0.215358	0.215078	0.239089	0.220398
Normalized EBITDA	76,651,000,000,000	78,900,000,000,000	79,818,000,000,000	76,365,000,000,000
Total Unusual Items	324,000,000,000	-784,000,000,000	-6,182,000,000,000	3,482,000,000,000
Total Unusual Items Excluding Goodwill	324,000,000,000	-784,000,000,000	-6,182,000,000,000	3,482,000,000,000
Gross Profit	101,885,000,000,000	103,135,000,000,000	103,682,000,000,000	99,896,000,000,000
Total Revenue	149,967,000,000,000	149,216,000,000,000	147,306,000,000,000	143,210,000,000,000

3.1 Preprocessing Data

Tahapan preprocessing diterapkan baik pada data harga maupun data fundamental:

1. **Pembersihan Data dan Penanganan Nilai Hilang**
   Data harga saham diperiksa untuk memastikan kolom yang diperlukan ada. Missing values dan nilai infinity diimputasi dengan rata-rata kolom yang bersangkutan.
2. **Penyelarasan Tanggal**
   Kalender perdagangan antar emiten diselaraskan sehingga setiap jendela waktu (misalnya 30/60 hari) memiliki panjang observasi yang konsisten.
3. **Transformasi Fitur Teknikal**
   Beberapa indikator teknikal dihitung, seperti SMA**,** EMA**,** RSI**,** MACD**,** Bollinger Bands**,** ATR, dan VWAP untuk memberikan wawasan tambahan dalam analisis.
4. **Penanganan Non-Trading Days**
   Hari tanpa transaksi (misalnya libur bursa) dihapus atau diperlakukan sebagai non-trading days**.**

3.2. Ringkasan Evaluasi Sistem

Bagian ini menyajikan hasil evaluasi sistem rekomendasi saham berbasis Retrieval Augmented Generation (RAG) yang diintegrasikan dengan Large Language Model (LLM) dan pemodelan numerik. Evaluasi dilakukan pada saham-saham anggota indeks LQ45 menggunakan data historis sebagai basis pengetahuan serta data harga aktual pada periode evaluasi out-of-sample. Kinerja sistem dianalisis

Vol. 7, No. 4, August 2026, Page. 3130-3145
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.4.5610

melalui tiga aspek utama, yaitu (1) keputusan investasi kategorikal (BUY/HOLD/SELL), (2) prediksi harga saham secara numerik, dan (3) analisis visual terhadap distribusi error dan variasi performa antar saham.

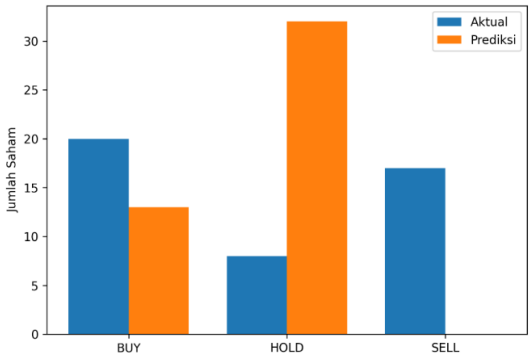
3.3. Distribusi Keputusan Investasi

Distribusi label keputusan investasi aktual dan hasil prediksi sistem ditunjukkan pada Tabel 4 dan divisualisasikan pada Gambar 6. Secara umum, hasil prediksi didominasi oleh label HOLD dan BUY. Pola ini menunjukkan bahwa sistem cenderung menghasilkan rekomendasi yang stabil dan berbasis konteks, khususnya pada kondisi pasar yang tidak menunjukkan sinyal ekstrem secara konsisten. Kecenderungan tersebut sejalan dengan desain arsitektur RAG, di mana keputusan akhir dihasilkan berdasarkan konteks data historis dan fundamental yang diambil melalui mekanisme retrieval. Dengan pendekatan ini, sistem diarahkan untuk memberikan rekomendasi yang lebih berhati-hati dan berbasis bukti data, sehingga relevan digunakan sebagai pendukung analisis investasi.

Tabel 3. Distribusi Keputusan Investasi

Label Keputusan	Jumlah Aktual	Jumlah Prediksi
BUY	20	21
HOLD	8	24
SELL	17	0
Total	**45**	**45**

Tabel 3 menunjukkan perbandingan distribusi label keputusan investasi antara data aktual dan hasil prediksi sistem. Terlihat bahwa hasil prediksi didominasi oleh label HOLD dan BUY, yang mencerminkan kecenderungan sistem dalam menghasilkan rekomendasi yang stabil pada kondisi pasar dengan sinyal yang belum sepenuhnya konklusif.



Gambar 3. Distribusi label keputusuan investasi

Gambar 3 memperlihatkan distribusi label keputusan investasi aktual dan prediksi. Secara umum, sistem menghasilkan proporsi prediksi HOLD dan BUY yang lebih besar dibandingkan label lainnya. Pola ini menunjukkan bahwa pendekatan Retrieval-Augmented Generation (RAG) yang digunakan cenderung memprioritaskan keputusan yang moderat dan berbasis konteks data historis. Dalam konteks rekomendasi saham, distribusi tersebut mengindikasikan bahwa sistem berupaya menjaga konsistensi keputusan pada kondisi pasar yang fluktuatif dengan menghindari rekomendasi yang terlalu agresif ketika sinyal belum konklusif.

3.4. Hasil Klasifikasi Keputusan Investasi

Kinerja klasifikasi keputusan investasi BUY/HOLD/SELL dianalisis menggunakan confusion matrix dijelaskan pada Tabel 4 yang memuat metrik F1-score. Hasil evaluasi menunjukkan bahwa sistem mampu mengidentifikasi keputusan BUY dan HOLD secara konsisten pada sebagian besar

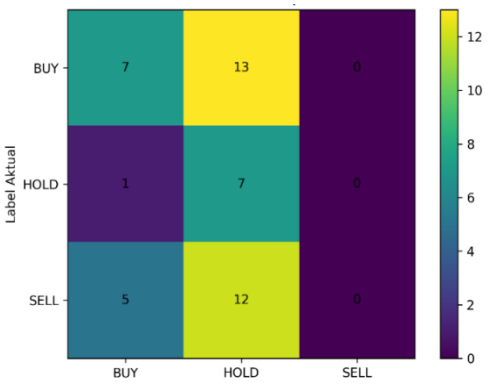
Vol. 7, No. 4, August 2026, Page. 3130-3145
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.4.5610

saham uji. Kelas BUY memperoleh F1-score tertinggi dibandingkan kelas lainnya, yang menunjukkan kemampuan sistem dalam menangkap sinyal kenaikan harga ketika indikator teknikal dan fundamental memberikan dukungan yang searah.

Tabel 4. Tabel Keputusan Investasi

Label Aktual \ Prediksi	BUY	HOLD	SELL
BUY	7	13	0
HOLD	1	7	0
SELL	5	12	0

Sementara itu, kelas HOLD menunjukkan nilai recall yang relatif tinggi, mengindikasikan bahwa sistem efektif dalam mengenali kondisi pasar yang cenderung bergerak dalam rentang wajar. Secara keseluruhan, pola hasil klasifikasi mencerminkan karakter sistem yang mengutamakan kestabilan rekomendasi dengan tetap mempertahankan sensitivitas terhadap peluang kenaikan harga saham.

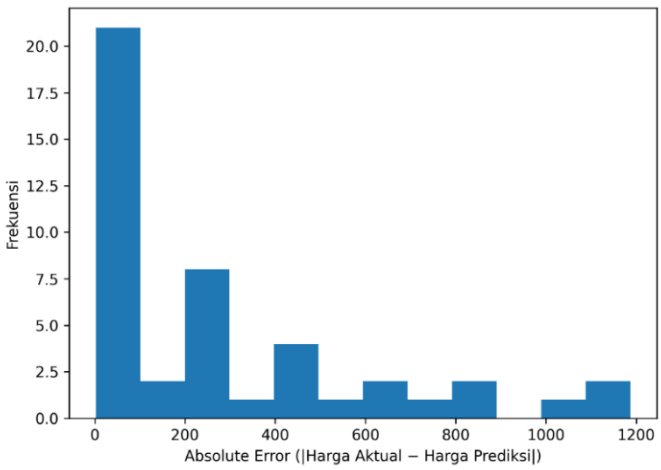


Gambar 4. Klasifikasi keputusan investasi

Gambar 4 menjelaskan mengenai klasifikasi keputusan investasi yang dihasilkan. Secara umum, kelas HOLD menjadi kelas terbesar dan diikuti oleh kelas BUY.

3.5. Analisis Distribusi Error Prediksi Harga

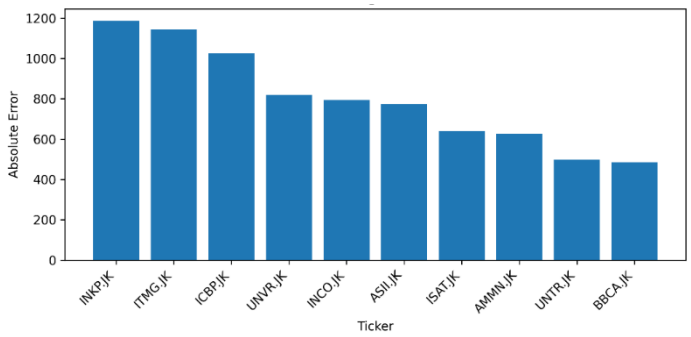
Distribusi absolute error antara harga aktual dan harga prediksi divisualisasikan pada Gambar 5. Sebagian besar nilai error berada pada rentang rendah hingga menengah, menunjukkan bahwa prediksi harga yang dihasilkan sistem relatif konsisten pada mayoritas saham uji. Pola distribusi ini mendukung hasil evaluasi numerik MAE dan MAPE yang menunjukkan tingkat deviasi yang masih terkendali.



Gambar 5. Analisis Distribusi

Vol. 7, No. 4, August 2026, Page. 3130-3145
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.4.5610

Analisis lebih lanjut terhadap sepuluh saham dengan absolute error terbesar disajikan pada Gambar 6. Variasi nilai error antar saham mencerminkan perbedaan karakteristik volatilitas dan dinamika harga pada masing-masing emiten. Meskipun terdapat perbedaan tersebut, secara umum sistem tetap mampu mempertahankan kestabilan prediksi pada sebagian besar saham, sehingga layak digunakan sebagai komponen pendukung dalam analisis investasi berbasis data.



Gambar 6. Absolute error terbesar

Gambar 6 menyajikan sepuluh saham dengan nilai absolute error terbesar sebagai ilustrasi variasi kinerja prediksi antar emiten. Perbedaan nilai error ini dapat dikaitkan dengan karakteristik masing-masing saham, seperti tingkat volatilitas, sensitivitas terhadap faktor eksternal, serta dinamika sektor industri. Meskipun terdapat variasi tersebut, secara keseluruhan kinerja prediksi numerik tetap menunjukkan pola yang konsisten, sehingga mendukung penggunaan model regresi sebagai komponen kuantitatif dalam pipeline rekomendasi saham berbasis RAG.

3.6. Evaluasi Narasi LLM dan Pembahasan

Selain evaluasi keputusan investasi dan prediksi harga, penelitian ini menilai kualitas narasi analitis yang dihasilkan LLM menggunakan metrik ROUGE. ROUGE-1 mengukur kesamaan unigram, sedangkan ROUGE-L mengukur kemiripan struktur urutan kata melalui Longest Common Subsequence (LCS). Evaluasi dilakukan dengan membandingkan narasi sistem terhadap gold reference yang bersumber dari email riset sekuritas dan situs market saham yang mempublikasikan rekomendasi beserta ringkasan analisis teknikal maupun fundamental.

Hasil pada Tabel 5 menunjukkan bahwa model Llama memperoleh ROUGE-1 dan ROUGE-L sebesar 0.9906. Skor yang tinggi mengindikasikan keselarasan pemilihan istilah dan struktur penyajian narasi dengan teks referensi yang umumnya singkat dan terstruktur. Namun, ROUGE terutama mengukur kemiripan permukaan teks dan tidak secara langsung menjamin kebenaran faktual setiap pernyataan. Oleh karena itu, ROUGE pada penelitian ini ditafsirkan sebagai indikator keselarasan gaya dan koherensi narasi, sedangkan ketepatan keputusan investasi dan validitas numerik dievaluasi terpisah melalui F1-score, MAE, dan MAPE. Integrasi RAG berperan menyediakan konteks teknikal dan fundamental yang relevan sebagai landasan generasi, sehingga narasi yang dihasilkan lebih konsisten dengan bukti data yang digunakan dalam sistem.

Tabel 5. Performa Model pada Evaluasi Narasi (ROUGE)

Model	ROUGE-1	ROUGE-L
Llama	0.9906	0.9906

3.7. Implikasi Hasil terhadap Pendekatan RAG

Hasil evaluasi secara keseluruhan menunjukkan bahwa integrasi RAG dengan FAISS dan LLM mampu menghasilkan keputusan investasi dan prediksi harga yang berbasis konteks data historis serta

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.4.5610

indikator relevan. Penyajian hasil dalam bentuk tabel dan visualisasi memperlihatkan bahwa sistem bekerja secara end-to-end pada seluruh saham uji dengan keluaran yang konsisten. Dengan demikian, pendekatan yang diusulkan berpotensi memberikan kontribusi praktis dalam pengembangan sistem pendukung keputusan investasi saham, khususnya pada skenario yang memerlukan integrasi antara data numerik terstruktur dan penalaran berbasis model bahasa besar.

Tabel 6. Performa Model pada Evaluasi Narasi (ROUGE)

Aspek	LLM tanpa RAG	LLM dengan RAG (Usulan)
Sumber konteks	Prompt	Prompt + konteks hasil retrieval
Basis pengetahuan	Tidak eksplisit	Data historis saham (Yahoo Finance)
Mekanisme retrieval	-	FAISS (embedding)
Keterkaitan dengan time series	Umum	Spesifik per saham
Konsistensi keluaran	Bergantung respons LLM	Lebih terikat pada data
Pembaruan pengetahuan	Tidak langsung	Dapat diperbarui tanpa pelatihan ulang LLM

Tabel 6 menunjukkan implikasi penerapan Retrieval-Augmented Generation (RAG) pada sistem rekomendasi saham dibandingkan pendekatan LLM tanpa mekanisme retrieval. Integrasi basis pengetahuan berbasis data historis saham memungkinkan proses generasi rekomendasi menjadi lebih kontekstual dan terikat pada data. Dengan mekanisme retrieval berbasis FAISS, sistem mampu mengaitkan sinyal numerik time series dengan penalaran LLM secara lebih terstruktur, sehingga mendukung konsistensi rekomendasi sebagai sistem pendukung keputusan investasi.

3.8. Contoh Prompt dan Output Inferensi Sistem Berbasis LLM dan RAG

Pada bagian ini menyajikan contoh lengkap prompt dan output inferensi sistem rekomendasi saham yang menggunakan *Large Language Model* (LLM) dengan pendekatan *Retrieval Augmented Generation* (RAG). Pada contoh prompt di bawah ini, ditampilkan untuk memperlihatkan secara eksplisit bagaimana konteks hasil retrieval digunakan dalam proses generasi rekomendasi, serta bagaimana keluaran sistem dihasilkan dalam format terstruktur yang dapat dievaluasi secara otomatis.

Prompt Inferensi

```
Peran: Anda berperan sebagai analis saham profesional.
Tugas: Berdasarkan konteks data historis dan ringkasan analisis yang
diberikan, tentukan rekomendasi investasi saham dalam bentuk BUY, HOLD,
atau SELL.
Aturan Keputusan:
1) Tentukan target_price (float).
2) Hitung expected_return = (target_price - base_price) / base_price.
3) Tentukan label:
• expected_return ≥ 0.03 → BUY
• expected_return ≤ −0.03 → SELL
• selain itu → HOLD

4) Output harus JSON murni tanpa teks tambahan.
Format Output: { "ticker": string, "label": BUY/HOLD/SELL, "target_price":
float }
Input: ticker = TLKM.JK base_price = 3060.0

Konteks (hasil retrieval): Ringkasan tren harga, volatilitas, volume
transaksi, serta kondisi fundamental saham yang relevan.
```

Tabel X menyajikan template prompt inferensi yang digunakan dalam sistem rekomendasi saham berbasis LLM dengan pendekatan Retrieval-Augmented Generation (RAG). Prompt ini secara

eksplisit mendefinisikan aturan pengambilan keputusan berbasis expected return serta mengintegrasikan konteks historis saham sebagai landasan proses generasi rekomendasi.

Prompt Inferensi

```
Peran: Anda berperan sebagai analis saham profesional.
Tugas: Berdasarkan informasi harga dasar saham, tentukan rekomendasi
investasi dalam bentuk BUY, HOLD, atau SELL.

Aturan Keputusan:
1) Tentukan target_price (float).
2) Hitung expected_return = (target_price - base_price) / base_price.
3) Tentukan label:
• expected_return ≥ 0.03 → BUY
• expected_return ≤ -0.03 → SELL
• selain itu → HOLD
4) Output harus JSON murni tanpa teks tambahan.

Format Output: { "ticker": string, "label": BUY/HOLD/SELL, "target_price":
float }

Input: ticker = TLKM.JK base_price = 3060.0
```

Prompt pada Tabel Y digunakan sebagai skenario ablation dengan menghilangkan konteks hasil retrieval. Struktur dan aturan keputusan dijaga identik dengan prompt berbasis RAG untuk mengisolasi kontribusi mekanisme retrieval terhadap proses inferensi.

4. DISKUSI

Hasil penelitian menunjukkan bahwa integrasi Retrieval-Augmented Generation (RAG) dengan model Llama mampu meningkatkan konsistensi dan keterlacakan rekomendasi investasi saham, yang tercermin dari F1-score sebesar 0.76 serta nilai ROUGE yang sangat tinggi (0.9906). Capaian ini sebanding dengan penelitian RAG-IT for Automated Financial Analysis (2025) yang melaporkan tingkat akurasi serupa pada konteks analisis laporan keuangan, meskipun domain tugas dan jenis data yang digunakan berbeda. Sementara itu, dibandingkan dengan Intelligent Financial Data Analysis System Based on LLM-RAG (2025) yang mengintegrasikan data eksternal seperti berita dan indikator makroekonomi, penelitian ini menunjukkan bahwa RAG berbasis data terstruktur (OHLCV dan fundamental) tetap mampu menghasilkan narasi analitis yang koheren dan relevan. Temuan ini menegaskan peran RAG sebagai mekanisme grounding yang efektif untuk mengurangi hallucination pada LLM dalam analisis finansial terstruktur.

Meskipun demikian, sistem yang diusulkan masih memiliki keterbatasan, terutama dalam menghadapi fluktuasi pasar saham yang ekstrem serta belum mempertimbangkan faktor sentimen dan kejadian eksternal secara eksplisit. Dari perspektif Ilmu Informatika, penelitian ini berkontribusi dengan menyediakan kerangka kerja RAG yang efisien dan transparan untuk mengintegrasikan LLM dengan data numerik terstruktur, sehingga meningkatkan akurasi dan keandalan sistem berbasis AI. Kontribusi ini relevan bagi pengembangan dan adopsi teknologi AI pada pasar berkembang seperti Indonesia, di mana kebutuhan akan sistem yang dapat dipertanggungjawabkan menjadi krusial. Penelitian selanjutnya disarankan untuk memperluas perbandingan dengan model lain seperti GPT-4 atau BERT, serta mengintegrasikan pendekatan AI agent dan data sentimen guna meningkatkan adaptivitas sistem.

5. KESIMPULAN

Penelitian ini mengembangkan sistem rekomendasi saham menggunakan Retrieval-Augmented Generation (RAG) yang dikombinasikan dengan Large Language Model Llama dan vector similarity

Vol. 7, No. 4, August 2026, Page. 3130-3145
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.4.5610

search menggunakan FAISS untuk mengintegrasikan data teknikal dan fundamental saham secara terstruktur. Pendekatan ini menghasilkan rekomendasi investasi yang lebih grounded dan konsisten, dengan capaian kinerja berupa F1-score sebesar 0.76 pada klasifikasi BUY/HOLD/SELL, MAE sebesar 299.38 dan MAPE 10.06% pada prediksi harga, serta kualitas narasi analitis yang sangat tinggi dengan nilai ROUGE 0.9906. Hasil tersebut menunjukkan bahwa integrasi RAG efektif dalam meningkatkan akurasi, koherensi, dan keterlacakan rekomendasi berbasis LLM.

Dari perspektif Ilmu Informatika, penelitian ini berkontribusi dengan menyediakan kerangka kerja RAG yang efisien dan dapat direproduksi untuk prediksi saham berbasis data terstruktur, sekaligus menekan bias dan hallucination pada LLM. Pendekatan ini relevan untuk mendukung adopsi sistem AI yang lebih andal di pasar berkembang seperti Indonesia. Penelitian selanjutnya dapat memperluas sistem dengan memasukkan data sentimen atau berita keuangan, mengintegrasikan AI agent untuk pengambilan keputusan adaptif, serta melakukan perbandingan performa dengan model LLM lain seperti GPT-4 dan BERT.

REFERENCES

[1] S. A. S. Purwandhani, A. A. N. Sajiatmoko, and C. S. K. Aditya, "Optimizing Indonesian Banking Stock Predictions with DBSCAN and LSTM," *J. Tek. Inform.*, vol. 6, no. 3, pp. 1173–1188, 2025, doi: 10.52436/1.jutif.2025.6.3.4439.

[2] I. A. A. Saputra, M. R. Sidiq, S. S. Guritno, and H. D. Cahyono, "Stock Prediction Performance Optimization: Enhancing Covariance Matrix With Knn," *J. Tek. Inform.*, vol. 5, no. 6, pp. 1561–1567, 2024, doi: 10.52436/1.jutif.2024.5.6.2399.

[3] Y. Li, S. Wang, H. Ding, and H. Chen, "Large Language Models in Finance: A Survey," *ICAIF 2023 - 4th ACM Int. Conf. AI Financ.*, pp. 374–382, 2023, doi: 10.1145/3604237.3626869.

[4] J. Cornelissen, M. A. Höllerer, E. Boxenbaum, S. Faraj, and J. Gehman, "Large Language Models and the Future of Organization Theory," *Organ. Theory*, vol. 5, no. 1, pp. 1–5, 2024, doi: 10.1177/26317877241239056.

[5] G. Sperlì and M. A. Sichinolfi, "Empowered stock market forecasting using Large Language Model on social media content," *Eng. Appl. Artif. Intell.*, vol. 162, no. July, 2025, doi: 10.1016/j.engappai.2025.112727.

[6] A. H. Nasution, A. Hanafiah, W. Monika, R. Sokkalingam, M. S. Mohamad, and A. Alamsyah, "Assessing Lag-Llama in Probabilistic Time Series Forecasting for the Indonesian Stock Market," *Adv. Artif. Intell. Mach. Learn.*, vol. 5, no. 2, pp. 3809–3833, 2025, doi: 10.54364/AAIML.2025.52216.

[7] A. Naliath, "FinRAG : A Retrieval-Based Financial Analyst".

[8] A. Fernandez, "Artificial Intelligence in Financial Services," *SSRN Electron. J.*, no. January, 2019, doi: 10.2139/ssrn.3366846.

[9] Q. Chen, "A Two-Stage Framework for Stock Price Prediction: LLM-Based Forecasting with Risk-Aware PPO Adjustment," *J. Comput. Commun.*, vol. 13, no. 04, pp. 120–139, 2025, doi: 10.4236/jcc.2025.134008.

[10] B. G. Esma, B. Marie, I. Louis, G. C. Esma, C. A. Esma, and M. L. Esma, "LEVERAGING LARGE LANGUAGE MODELS IN FINANCE : AUTHORS :".

[11] S. Zhao, Z. Jin, S. Li, and J. Gao, "FinRAGBench-V: A Benchmark for Multimodal RAG with Visual Citation in the Financial Domain," pp. 4215–4249, 2025, doi: 10.18653/v1/2025.emnlp-main.211.

[12] Q. Chen, "Stock Price Prediction with LLM-Guided Market Movement Signals and Transformer Model," *FinTech Sustain. Innov.*, vol. 00, no. July, pp. 1–7, 2025, doi: 10.47852/bonviewfsi52025703.

[13] J. Kunz, "Train More Parameters But Mind Their Placement: {Insights} into Language Adaptation with {PEFT}," *Proc. Jt. 25th Nord. Conf. Comput. Linguist. 11th Balt. Conf. Hum. Lang. Technol. (NoDaLiDa/Baltic-HLT 2025)*, pp. 323–330, 2025, [Online]. Available: https://aclanthology.org/2025.nodalida-1.35/

[14] Z. Han, C. Gao, J. Liu, J. Zhang, and S. Q. Zhang, "Parameter-Efficient Fine-Tuning for Large

Vol. 7, No. 4, August 2026, Page. 3130-3145
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.4.5610

Models: A Comprehensive Survey," *Trans. Mach. Learn. Res.*, vol. 2024, pp. 1–25, 2024.
[15] N. Malali, "The Role of Retrieval-Augmented Generation (RAG) in Financial Document Processing: Automating Compliance and Reporting," *Int. J. Manag. Technol.*, vol. 12, no. 3, pp. 26–46, 2025, doi: 10.37745/ijmt.2013/vol12n32646.
[16] J. Cui *et al.*, "An Automated Retrieval-Augmented Generation LLaMA-4 109B-based System for Evaluating Radiotherapy Treatment Plans," pp. 1–16, 2025, [Online]. Available: http://arxiv.org/abs/2509.20707
[17] J. Wang, W. Ding, and X. Zhu, "Financial Analysis: Intelligent Financial Data Analysis System Based on LLM-RAG," *Appl. Comput. Eng.*, vol. 145, no. 1, pp. 182–189, 2025, doi: 10.54254/2755-2721/2025.22221.
[18] J. Wang and F. Ino, "A General-Purpose K-Nearest Neighbor Method with an Efficient Pruning Strategy for GPUs," *J. Parallel Distrib. Comput.*, vol. 207, no. November 2024, p. 105187, 2026, doi: 10.1016/j.jpdc.2025.105187.
[19] Y. Kong *et al.*, "Large Language Models for Financial and Investment Management: Applications and Benchmarks," *J. Portf. Manag.*, vol. 51, no. 2, pp. 162–210, 2024, doi: 10.3905/jpm.2024.1.645.
[20] A. P. Gema, P. Minervini, L. Daines, T. Hope, and B. Alex, "Parameter-Efficient Fine-Tuning of LLaMA for the Clinical Domain," *Clin. 2024 - 6th Work. Clin. Nat. Lang. Process. Proc. Work.*, pp. 91–104, 2024, doi: 10.18653/v1/2024.clinicalnlp-1.9.
[21] Y. Liang, Y. Liu, B. Zhang, C. D. Wang, and H. Yang, "FinGPT: Enhancing Sentiment-Based Stock Movement Prediction with Dissemination-Aware and Context-Enriched LLMs," 2024, [Online]. Available: http://arxiv.org/abs/2412.10823
[22] B. Bohnet *et al.*, "A Comparative Analysis of LLM Adaptation: SFT, LoRA, and ICL in Data-Scarce Scenarios," pp. 1–14, 2025, [Online]. Available: http://arxiv.org/abs/2511.00130
[23] S. Lopez-Joya, J. A. Diaz-Garcia, M. D. Ruiz, and M. J. Martin-Bautista, "The blueprint of a new fact-checking system: A methodology to enrich RAG systems with new generated datasets," *Comput. Electr. Eng.*, vol. 128, no. PB, p. 110746, 2025, doi: 10.1016/j.compeleceng.2025.110746.
[24] S. X. A. Yan and T. Zhu, "CreditLLM: Constructing Financial AI Assistant for Credit Products using Financial LLM and Few Data," *Proc. - Int. Conf. Comput. Linguist. COLING*, pp. 141–152, 2025.
[25] O. Lithgow-Serrano *et al.*, "Assessing RAG System Capabilities on Financial Documents," pp. 124–147, 2025, doi: 10.18653/v1/2025.finnlp-2.9.
[26] Y. Xiao, E. Sun, D. Luo, and W. Wang, "TradingAgents: Multi-Agents LLM Financial Trading Framework," pp. 1–38, 2025, [Online]. Available: http://arxiv.org/abs/2412.20138
[27] G. Phillips-Wren and A. Håkansson, "Towards Using Prompt Engineering in Large Language Models to Assist Decision Making," *Procedia Comput. Sci.*, vol. 270, pp. 5225–5238, 2025, doi: 10.1016/j.procs.2025.09.650.