

Prediksi Harga Saham Sektor Energi di BEI Menggunakan Model Multimodal LSTM dengan Integrasi Fitur Numerik dan Sentimen Berita Pasar Modal di Indonesia

Naufal Lathifan Yumna ¹, Ghufron ²

¹ Universitas Islam Sultan Agung ; 32602200110@std.unissula.ac.id

² Universitas Islam Sultan Agung ; ghufron@unissula.ac.id

* Korespondensi: 32602200110@std.unissula.ac.id

Info Artikel:

Dikirim: 21 Mei 2026

Direvisi: 30 Mei 2026

Diterima: 05 Juni 2026

Abstrack: The movement of energy sector stock prices exhibits high volatility, which is influenced by historical data and news sentiment. This research evaluates the performance of stock price prediction by integrating numerical data and economic news texts using a multimodal Long Short-Term Memory (LSTM) architecture. The data utilized includes the stock prices of ADRO, PGAS, and INDY from the 2021 to 2026 period, alongside Indonesian-language economic news. Sentiment extraction from the news texts was conducted using the IndoBERT model. The results indicate that the IndoBERT model achieved an accuracy and F1-score of 83%. The evaluation of the unimodal model (historical data only) yielded a Mean Absolute Percentage Error (MAPE) of 3.62% for ADRO, 3.17% for PGAS, and 5.90% for INDY. Meanwhile, the multimodal model, which combined numerical and sentiment features, resulted in a MAPE of 4.00% (ADRO), 5.46% (PGAS), and 8.47% (INDY). In conclusion, the unimodal LSTM model proved to be effective; however, the integration of sentiment features in the multimodal scheme did not provide a significant improvement in accuracy due to the highly volatile nature of the stocks.

Keywords: Stock Price Prediction; LSTM; Multimodal; Sentiment Analysis; IndoBERT

Intisari: Pergerakan harga saham sektor energi memiliki volatilitas tinggi yang dipengaruhi oleh data historis dan sentimen berita. Penelitian ini mengevaluasi performa prediksi harga saham dengan mengintegrasikan data numerik dan teks berita ekonomi menggunakan arsitektur *Long Short-Term Memory* (LSTM) multimodal. Data yang digunakan meliputi harga saham ADRO, PGAS, dan INDY dari periode 2021 hingga 2026 beserta berita ekonomi berbahasa Indonesia. Ekstraksi sentimen teks berita menggunakan model *IndoBERT*. Hasil penelitian menunjukkan bahwa model *IndoBERT* memiliki akurasi dan *F1-score* sebesar 83%. Evaluasi pada model unimodal (hanya data historis) menghasilkan *Mean Absolute Percentage Error* (MAPE) sebesar 3,62% untuk ADRO, 3,17% untuk PGAS, dan 5,90% untuk INDY. Sementara itu, model multimodal yang menggabungkan fitur numerik dan sentimen menghasilkan MAPE sebesar 4,00% (ADRO), 5,46% (PGAS), dan 8,47% (INDY). Kesimpulannya, model LSTM unimodal terbukti efektif, namun integrasi fitur sentimen dalam skema multimodal belum memberikan peningkatan akurasi yang signifikan akibat karakteristik volatilitas saham yang tinggi.

Kata Kunci: prediksi harga saham; LSTM; multimodal; analisis sentimen; IndoBERT

1. Pendahuluan

Pemanfaatan sistem berbasis *Artificial Intelligence* (AI) dalam konteks finansial telah mengubah paradigma analisis pasar modal dari pendekatan konvensional menuju pemodelan prediktif berbasis data yang lebih adaptif. Di Indonesia, urgensi penerapan teknologi ini sangat terasa pada sektor energi, yang merupakan salah satu pilar ekonomi nasional namun memiliki karakteristik fluktuasi harga yang sangat tinggi. Berdasarkan catatan Bursa Efek Indonesia (BEI), indeks sektor energi (*IDX Energy*) menunjukkan volatilitas yang signifikan, dengan lonjakan harian sebesar 2,45% pada pertengahan Februari 2026 sebagai respons terhadap dinamika harga komoditas global. Sebagaimana dikemukakan dalam penelitian [1], pergerakan harga saham emiten energi sering kali mengalami perubahan drastis dalam waktu singkat sehingga investor memerlukan instrumen prediksi berbasis komputasi cerdas untuk meminimalkan risiko investasi.

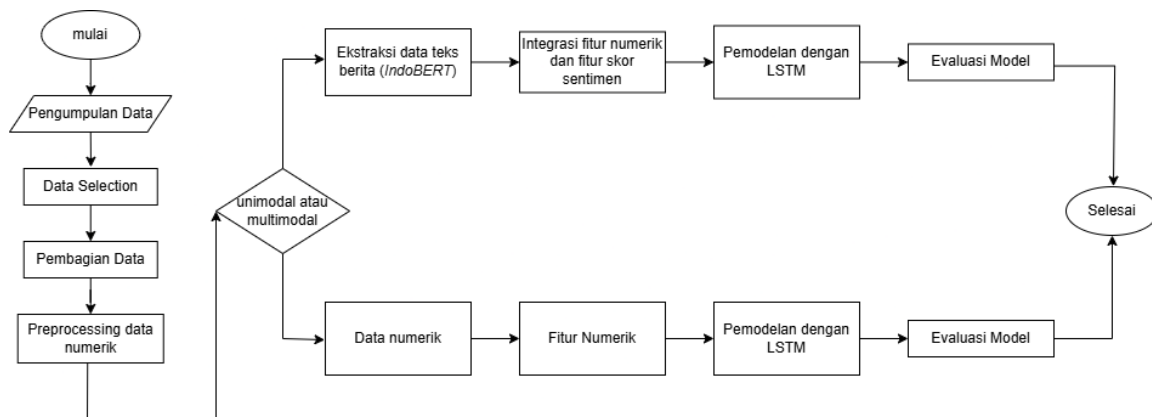
Secara teoritis, pergerakan harga saham tidak hanya terbentuk dari pola transaksi historis semata, melainkan juga merupakan refleksi dari berbagai informasi eksternal yang beredar di publik. Menurut [2], variabel makroekonomi dan isu global memiliki dampak signifikan terhadap volatilitas pasar saham. Informasi yang disampaikan melalui berita ekonomi berperan memengaruhi sentimen pasar yang selanjutnya mendorong aktivitas jual beli saham. Sentimen positif cenderung meningkatkan optimisme investor, sedangkan sentimen negatif dapat memicu respons pasar yang berlawanan [3]. Oleh karena itu, mengandalkan data numerik historis saja (unimodal) sering kali tidak cukup untuk menangkap sinyal pergerakan harga yang dipicu oleh faktor kualitatif.

Guna mengolah karakteristik *time-series* pada data harga saham, arsitektur *Long Short-Term Memory* (LSTM) telah terbukti menjadi pendekatan *deep learning* yang paling efektif. Berdasarkan kajian [4] dan [5], LSTM memiliki keunggulan mekanisme *gating* yang mampu mereduksi kendala *vanishing gradient* yang lazim ditemukan pada *Recurrent Neural Network* (RNN) konvensional. Di sisi lain, untuk menangani data tekstual berupa berita ekonomi berbahasa Indonesia, model *IndoBERT* menawarkan kemampuan pemahaman konteks yang superior. Mengacu pada penelitian [6], *IndoBERT* sebagai model *pre-trained* berbasis *Transformer* mampu mengekstraksi fitur semantik dan mendeteksi arah sentimen pasar dengan akurasi tinggi karena dilatih pada korpus data Bahasa Indonesia yang masif.

Meskipun LSTM dan *IndoBERT* masing-masing telah terbukti andal, terdapat kesenjangan penelitian yang signifikan dalam integrasi keduanya untuk pasar modal Indonesia. Mayoritas penelitian terdahulu masih berjalan secara terpisah, ada yang fokus memprediksi harga menggunakan LSTM [5], atau fokus klasifikasi sentimen menggunakan *IndoBERT* tanpa dikaitkan langsung dengan prediksi harga [7]. Penelitian hibrida seperti [8] pun belum menyentuh studi kasus spesifik pada emiten sektor energi yang sangat *news-sensitive*. Merespons tantangan tersebut, penelitian ini mengimplementasikan pendekatan Multimodal yang mengintegrasikan fitur numerik dan fitur sentimen *IndoBERT* ke dalam arsitektur LSTM melalui strategi *early fusion* [9], dengan objek studi tiga emiten sektor energi: ADRO, PGAS, dan INDY).

2. Metode Penelitian

Bagian ini memaparkan kerangka metodologi penelitian yang diawali dengan akuisisi dan seleksi data harga saham serta berita ekonomi. Data tersebut kemudian didistribusikan secara sekuensial menjadi data latih, validasi, dan uji guna menjaga karakteristik *time-series*. Setelah melalui tahap *preprocessing* dan *feature engineering*, dibangun dua skenario pemodelan menggunakan arsitektur LSTM yang kemudian dievaluasi dan dibandingkan kinerjanya: skenario unimodal (hanya berbasis fitur teknikal harga saham) dan multimodal (kombinasi fitur teknikal dengan fitur sentimen berita menggunakan *IndoBERT*). Seluruh rangkaian prosedur metodologi ini diilustrasikan secara komprehensif pada diagram alir di Gambar 1.



Gambar 1. Tahapan Metode Penelitian

2.1. Pengumpulan dan Seleksi Data

Pada Data yang digunakan dalam studi ini merupakan data sekunder dari dua modalitas berbeda pada rentang waktu Januari 2021 hingga Maret 2026. Data numerik berupa data historis saham harian dari tiga emiten sektor energi yang terdaftar di BEI, yaitu PT Alamtri Resources Indonesia Tbk (ADRO), PT Perusahaan Gas Negara Tbk (PGAS), dan PT Indika Energy Tbk (INDY), yang diperoleh melalui *Yahoo Finance* dengan atribut *Open*, *High*, *Low*, *Close*, dan *Volume*. Pemilihan ketiga saham didasarkan pada tingkat likuiditas tinggi, kapitalisasi pasar besar, dan keikutsertaan dalam indeks IDX80 dan LQ45 [2]. Masing-masing saham memiliki 1.260 rekaman data sehingga total data numerik berjumlah 3.780.

Data tekstual berupa berita ekonomi dikumpulkan dari portal CNBC Indonesia, Kontan, dan Detik.com menggunakan metode *web scraping* berbasis *library BeautifulSoup* dan *Requests* [6]. Penggunaan lebih dari satu portal dilakukan untuk menghindari bias redaksional. Seleksi berita dilakukan menggunakan filter kata kunci nama emiten pada judul maupun isi berita. Total data berita yang terkumpul berjumlah 2.981 artikel: 1.551 berita ADRO, 835 berita PGAS, dan 595 berita INDY.

2.2. Preprocessing Data

Preprocessing data numerik mencakup tiga tahap: (1) standardisasi format tanggal ke DD-MM-YYYY dan pengurutan kronologis; (2) pengecekan *missing value*, hasil menunjukkan tidak ada nilai yang hilang; (3) identifikasi duplikasi, tidak ditemukan duplikasi sehingga data siap diproses. Pembagian data dilakukan secara kronologis dengan rasio 80% *training*, 10% *validasi*, dan 10% *testing* tanpa *shuffling* untuk mempertahankan struktur temporal [10]. Normalisasi menggunakan *Min-Max Scaling* dengan rentang [0,1] untuk fitur teknikal dan [-1,1] untuk fitur sentimen guna mempertahankan polaritas nilainya, dihitung berdasarkan data *training* untuk mencegah data *leakage*.

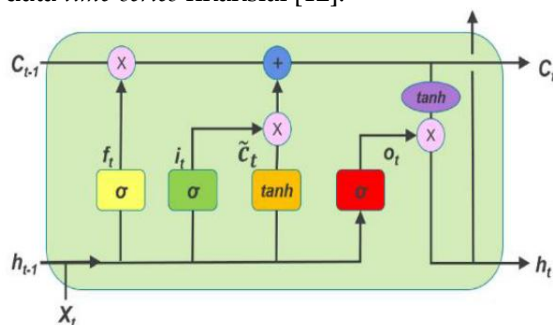
Feature engineering menghasilkan sembilan fitur teknikal: *log_close* (transformasi logaritmik harga penutupan), *Volume*, *Return*, *Moving Average 7* dan *14* hari (MA7, MA14), *Relative Strength Index (RSI)*, *Moving Average Convergence Divergence (MACD)* beserta *signal line*, dan *Bollinger Bands Width* [11]. Pembentukan *sequence* menggunakan teknik *sliding window* dengan ukuran *window* 21 hari untuk ADRO dan PGAS, serta 30 hari untuk INDY sesuai karakteristik volatilitasnya yang lebih tinggi. Pada model multimodal, ditambahkan satu fitur sentimen harian (*daily_sentiment*) sehingga total fitur menjadi 10.

Preprocessing data tekstual mencakup: pelabelan sentimen manual ke tiga kategori (Positif, Negatif, Netral) berdasarkan perspektif investor; pembersihan teks meliputi *lowercase*, penghapusan tanda baca, simbol, URL, dan angka; serta penggabungan judul dan isi berita untuk membentuk konteks yang lebih utuh sebelum tokenisasi dengan *tokenizer IndoBERT* berbasis *WordPiece* (maksimum 512 token).

2.3. Arsitektur Model

2.3.1. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) adalah varian dari *Recurrent Neural Network (RNN)* yang dirancang oleh Hochreiter dan Schmidhuber [4] untuk mengatasi masalah *vanishing gradient* yang muncul saat melatih RNN standar pada data sekuensial yang panjang. Berbeda dengan RNN konvensional, LSTM dilengkapi dengan mekanisme *gating* yang terdiri atas tiga buah gerbang (*gates*) serta dua jalur memori, yakni *cell state (C)* dan *hidden state (h)*. Mekanisme ini memungkinkan LSTM untuk secara selektif mengingat (*remember*) atau melupakan (*forget*) informasi dari *timestep* sebelumnya, sehingga sangat efektif dalam memodelkan dependensi jangka panjang pada data *time-series* finansial [12].



Gambar 2. Skema model LSTM [5]

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

Gambar 2 menunjukkan skema model LSTM, pada setiap *timestep* t , LSTM memproses tiga masukan utama: vektor fitur saat ini (x_t), *hidden state* sebelumnya (h_{t-1}), dan *cell state* sebelumnya (c_{t-1}). Aliran informasi diatur melalui mekanisme gerbang (*gates*) dengan formulasi berikut [4]:

1. *Forget Gate* untuk menentukan informasi dari memori sebelumnya yang akan dipertahankan atau dibuang.

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

2. *Input Gate* dan Kandidat untuk menyeleksi dan membentuk nilai informasi baru (\bar{C}_t) yang berpotensi ditambahkan ke dalam memori.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\bar{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

3. *Cell State* untuk memperbarui memori jangka panjang dengan membuang informasi tidak relevan dan menambahkan informasi baru yang telah diseleksi.

$$C_t = f_t * C_{t-1} + i_t * \bar{C}_t \quad (4)$$

4. Gerbang Keluaran dan *Hidden State*: Menentukan porsi *cell state* yang akan dikeluarkan sebagai *output* sekaligus memori jangka pendek (h_t).

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = O_t * \tanh(C_t) \quad (6)$$

2.3.2. IndoBERT dan *Fine-tuning* untuk klasifikasi sentimen

IndoBERT adalah model bahasa berbasis arsitektur *Bidirectional Encoder Representations from Transformers* (*BERT*) yang dilatih secara khusus menggunakan korpus data Bahasa Indonesia [6]. Model yang digunakan dalam penelitian ini adalah *indobenchmark/indobert-base-p2*, yang dilatih pada dataset Indo4B berisi sekitar 220 juta kata dari berbagai sumber teks Bahasa Indonesia. *IndoBERT* mengadopsi representasi teks bidireksional, sehingga mampu memahami konteks semantik suatu kata dari arah kiri maupun kanan dalam sebuah kalimat, berbeda dari model unidireksional konvensional.

Proses analisis sentimen menggunakan *IndoBERT* dalam penelitian ini terdiri atas empat tahap utama:

1. *Tokenisasi*, Teks berita dikonversi menjadi token menggunakan *tokenizer IndoBERT* berbasis algoritma *WordPiece*. Token khusus [CLS] disisipkan di awal dan [SEP] di akhir setiap teks. Panjang maksimum token ditetapkan 512 token sesuai batas arsitektur *BERT*.
2. *Encoding*, Token direpresentasikan sebagai tiga jenis *embedding* yang dijumlahkan: *Token Embedding* (representasi setiap *subword*), *Segment Embedding* (membedakan segmen kalimat A dan B), dan *Positional Embedding* (menyandikan posisi token dalam urutan).
3. *Transformer Encoder*, Representasi *embedding* diproses oleh 12 lapisan *Transformer encoder* bertumpuk. Setiap lapisan menerapkan mekanisme *Multi-Head Self-Attention* dan *Feed-Forward Network* untuk menghasilkan representasi kontekstual [13].
4. *Klasifikasi*, Representasi token [CLS] dari lapisan terakhir digunakan sebagai input *classification head* berupa lapisan *linear* yang menghasilkan probabilitas untuk tiga kelas sentimen (Positif, Negatif, Neutral).

Komponen inti *IndoBERT* adalah mekanisme *Self-Attention* yang memungkinkan model untuk memperhitungkan hubungan antar-token dalam satu kalimat secara bersamaan. Untuk setiap token, dibentuk tiga vektor yaitu *Query* (Q), *Key* (K), dan *Value* (V) yang diperoleh dari perkalian matriks bobot yang dipelajari. Skor perhatian dihitung sebagai:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

di mana d_k adalah dimensi vektor *Key*. Pembagian dengan $\sqrt{d_k}$ dilakukan untuk menstabilkan gradien selama pelatihan. Nilai *softmax* menghasilkan bobot perhatian yang menggambarkan seberapa besar perhatian yang diberikan setiap token terhadap token lainnya dalam konteks yang sama [14].

Setelah proses *encoding* selesai, representasi vektor token [CLS] yang merangkum konteks keseluruhan teks digunakan sebagai input *classification head*. Probabilitas setiap kelas sentimen dihitung menggunakan fungsi *softmax* [15]:

$$P(y = i|x) = \frac{e^{z_i}}{\sum_{j=1}^3 e^{z_j}} \quad (8)$$

Sentiment score akhir untuk setiap berita dihitung sebagai selisih probabilitas kelas Positif dan Negatif :

$$sentiment_{score} = P(positif) - P(negatif) \quad (9)$$

Nilai *Sscore* berada dalam rentang [-1, 1], di mana nilai mendekati +1 menunjukkan sentimen sangat positif dan nilai mendekati -1 menunjukkan sentimen sangat negatif [16]. Agregasi sentimen harian dilakukan dengan menghitung rata-rata seluruh *Score* berita yang dipublikasikan pada tanggal yang sama. Nilai 0 (netral) digunakan sebagai pengisi pada hari perdagangan yang tidak memiliki pemberitaan.

2.4. Evaluasi Model

Evaluasi kinerja model menggunakan tiga metrik [17]: *Root Mean Square Error* (RMSE) untuk mengukur rata-rata besarnya galat prediksi dengan bobot lebih besar pada kesalahan besar; *Mean Absolute Error* (MAE) untuk mengukur penyimpangan absolut dalam satuan Rupiah; dan *Mean Absolute Percentage Error* (MAPE) sebagai metrik utama untuk membandingkan akurasi antar saham secara persentase. Semakin rendah nilai ketiga metrik, semakin baik performa model.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (10)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (11)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (12)$$

Dalam perhitungan tersebut, y_i merupakan nilai aktual harga saham, \hat{y}_i adalah nilai prediksi, dan n mewakili total data uji.

3. Hasil dan Pembahasan

3.1. Hasil Fine-tuning IndoBERT

Proses *fine-tuning* IndoBERT dilakukan selama 4 epoch dengan konfigurasi: *batch size* 16, *learning rate* 2e-5, *weight decay* 0,05, *LR scheduler cosine*, dan *CrossEntropyLoss* terbobot. Model mencapai performa puncaknya pada *epoch* ke-4. Tabel 1 menyajikan hasil evaluasi model IndoBERT pada data validasi (596 sampel).

Tabel 1. Hasil Evaluasi *fine tuning* model IndoBERT

Kelas / Rata-rata	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0,81	0,85	0,83	196
Netral	0,68	0,73	0,70	63
Positif	0,88	0,83	0,85	337
<i>Accuracy</i>			0,83	596
<i>Macro Avg</i>	0,79	0,81	0,80	596
<i>Weighted Avg</i>	0,83	0,83	0,83	596

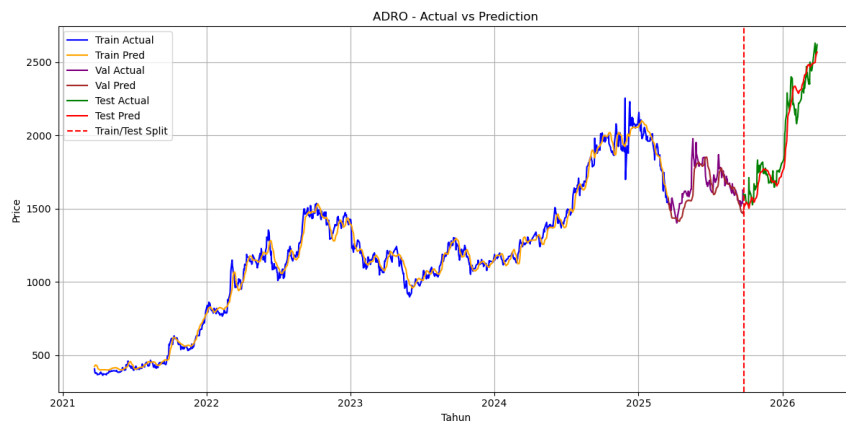
Model *IndoBERT* terbukti memiliki performa solid dalam mengklasifikasikan sentimen Positif dan Negatif, dengan *F1-Score* masing-masing 0,85 dan 0,83. Kategori Netral mencatatkan *F1-Score* terendah (0,70) sebagai dampak dari keterbatasan jumlah sampel. Secara keseluruhan, model mencapai akurasi dan *weighted F1-Score* sebesar 83%, yang menunjukkan kualitas memadai untuk digunakan sebagai komponen analisis sentimen dalam pipeline multi-modal. Model ini mengungguli pendekatan konvensional seperti TF-IDF dalam menerjemahkan konteks teks berita berbahasa Indonesia [6].

3.2. Hasil Evaluasi Model Unimodal

Model LSTM unimodal dilatih menggunakan 9 fitur teknikal dari data historis harga saham. Tabel 2 menyajikan hasil evaluasi pada data *testing*.

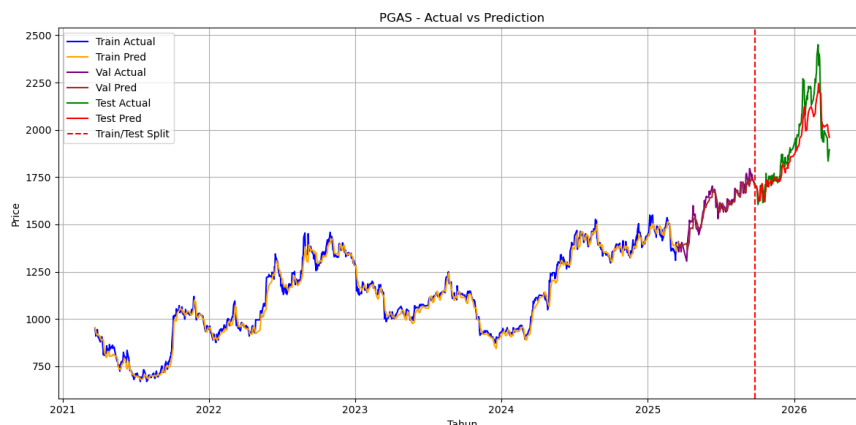
Tabel 2. Hasil Evaluasi Model LSTM Unimodal

Saham	RMSE	MAE	MAPE (%)
ADRO (PT Alamtri Resources Indonesia Tbk)	93,52	71,87	3,62
PGAS (PT Perusahaan Gas Negara Tbk)	84,85	64,66	3,17
INDY (PT Indika Energy Tbk)	241,03	176,88	5,90



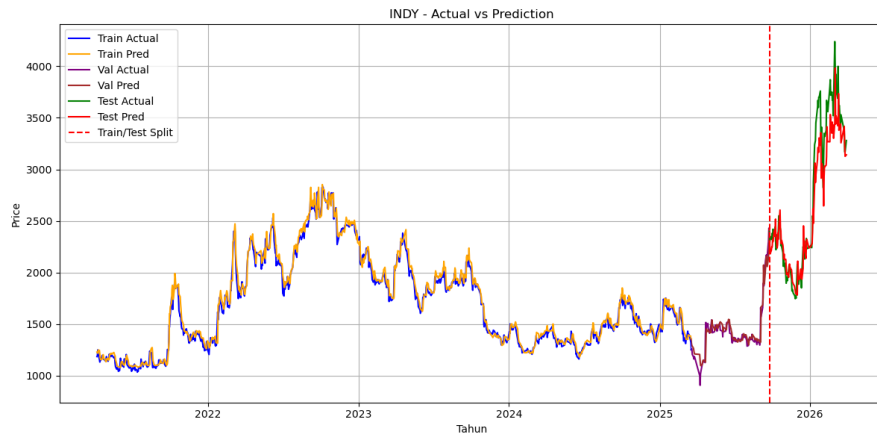
Gambar 3. Grafik prediksi dengan harga aktual saham ADRO (unimodal)

Gambar 3 memperlihatkan bahwa prediksi model unimodal pada saham ADRO mampu mengikuti arah tren harga aktual dengan cukup baik. Model mencatatkan MAPE sebesar 3,62% dengan MAE 71,87, menunjukkan deviasi harga prediksi yang relatif kecil terhadap harga penutupan aktual. Model terlihat sedikit terlambat merespons (*lag*) pada titik-titik pembalikan tren yang tajam, namun secara keseluruhan pola prediksi selaras dengan pergerakan harga nyata.



Gambar 4. Grafik prediksi dengan harga aktual saham PGAS (unimodal)

Gambar 4 menunjukkan bahwa saham PGAS menghasilkan performa terbaik pada model unimodal dengan MAPE 3,17% dan MAE 64,66. Garis prediksi terlihat sangat berdekatan dengan garis harga aktual, mengindikasikan bahwa pola pergerakan PGAS yang cenderung lebih stabil dibanding emiten lain dalam periode pengujian lebih mudah dimodelkan oleh LSTM berbasis fitur teknikal.



Gambar 5. Grafik prediksi dengan harga aktual saham INDY (unimodal)

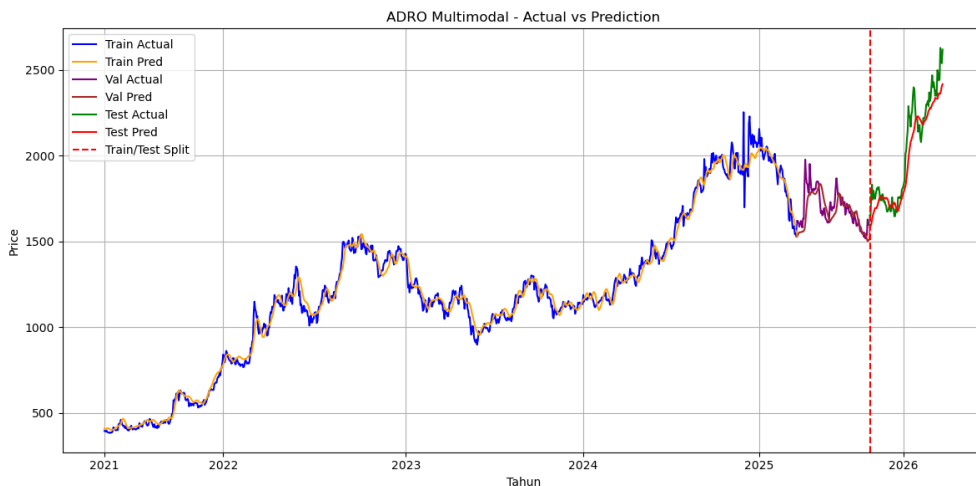
Gambar 5 menampilkan hasil prediksi INDY yang memiliki nilai error tertinggi dalam kelompok unimodal (MAPE 5,90%; RMSE 241,03). Hal ini mencerminkan karakteristik harga INDY yang lebih volatil dan memiliki kisaran nominal lebih tinggi. Meski demikian, model masih mampu menangkap arah tren secara umum, dan nilai MAPE di bawah 10% tetap mengindikasikan kategori akurasi yang baik [17].

3.3. Hasil Evaluasi Model Multimodal

Model LSTM multimodal menggunakan 10 fitur dengan arsitektur dan konfigurasi *hyperparameter* identik terhadap model unimodal. Tabel 3 menyajikan hasil evaluasinya.

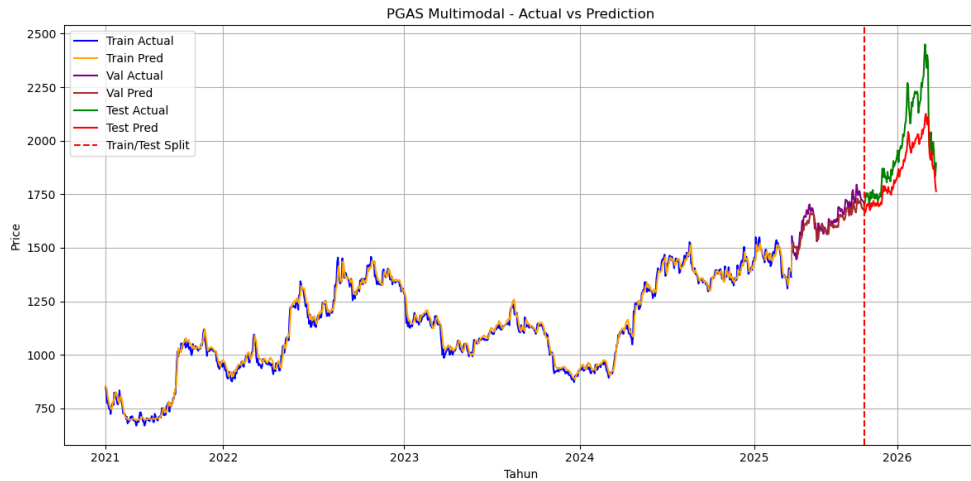
Tabel 3. Hasil Evaluasi Model LSTM multimodal

Saham	RMSE	MAE	MAPE (%)
ADRO (PT Alamtri Resources Indonesia Tbk)	109,06	84,04	4,00
PGAS (PT Perusahaan Gas Negara Tbk)	140,55	113,82	5,46
INDY (PT Indika Energy Tbk)	354,71	270,95	8,47



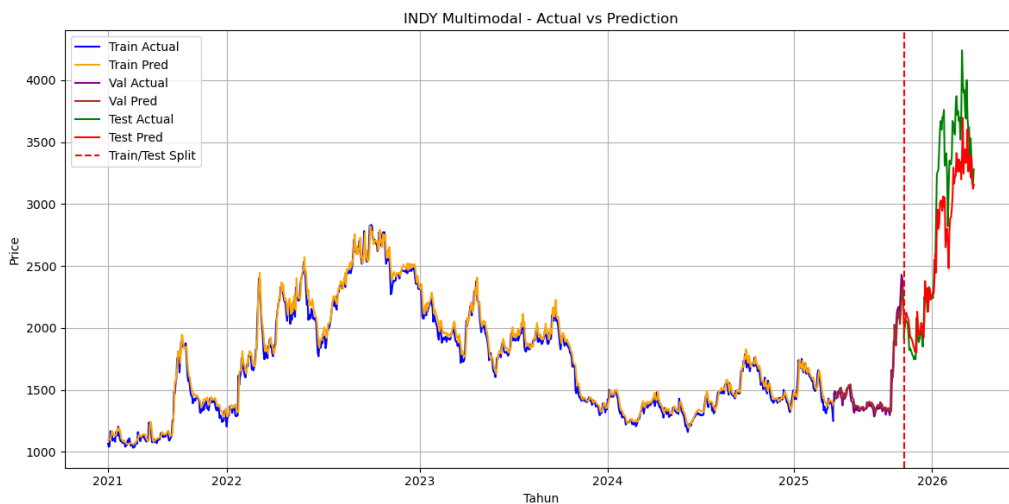
Gambar 6. Grafik prediksi dengan harga aktual saham ADRO (multimodal)

Gambar 6 memperlihatkan bahwa model multimodal pada saham ADRO mencatatkan MAPE 4,00%, sedikit lebih tinggi dibanding versi unimodal (3,62%). Secara visual, pola prediksi masih mengikuti tren aktual dengan baik, namun terdapat beberapa titik di mana penambahan fitur sentimen tampak memperkenalkan *noise* tambahan, terutama pada periode dengan *volume* berita tinggi namun pergerakan harga yang tidak selaras dengan sentimen dominan.



Gambar 7. Grafik prediksi dengan harga aktual saham PGAS (multimodal)

Gambar 7 menunjukkan penurunan performa yang lebih signifikan pada PGAS dengan MAPE multimodal 5,46% dibandingkan unimodal 3,17%. Garis prediksi cenderung memiliki deviasi lebih besar di beberapa segmen, mengindikasikan bahwa fitur sentimen berita PGAS yang berjumlah lebih sedikit (835 artikel) dibanding ADRO (1.551 artikel) belum cukup representatif untuk memberikan sinyal prediksi yang konsisten.



Gambar 8. Grafik prediksi dengan harga aktual saham INDY (multimodal)

Gambar 8 menampilkan bahwa INDY multimodal mencatatkan MAPE tertinggi sebesar 8,47% (RMSE 354,71), meningkat dari 5,90% pada unimodal. Dengan jumlah data berita yang paling sedikit (595 artikel), representasi sentimen harian INDY mengandung banyak nilai pengisi (0/netral) yang berpotensi menambah *noise* pada input model. Meski demikian, model masih mampu menangkap arah tren umum, dan nilai MAPE masih berada dalam batas kategori akurasi yang dapat diterima.

3.4. Perbandingan dan Pembahasan

Berdasarkan perbandingan Tabel 2 dan Tabel 3, model multimodal belum mampu mengungguli performa model unimodal pada semua emiten. Peningkatan MAPE dari unimodal ke multimodal mencapai 0,38 poin untuk ADRO, 2,29 poin untuk PGAS, dan 2,57 poin untuk INDY. Kondisi ini mengindikasikan bahwa sentimen berita yang

diintegrasikan melalui *early fusion* belum memberikan sinyal yang konsisten untuk meningkatkan akurasi prediksi harga saham sektor energi.

Temuan ini relevan dengan hasil penelitian [18] yang menunjukkan bahwa penambahan fitur sentimen dapat meningkatkan kinerja model pada kondisi tertentu, namun efektivitasnya bergantung pada kualitas dan kuantitas data tekstual. Rendahnya kontribusi sentimen kemungkinan disebabkan oleh beberapa faktor: (1) adanya lag antara publikasi berita dan respons pasar sehingga sentimen hari ini belum tentu tercermin pada harga hari yang sama; (2) saham sektor energi sangat dipengaruhi faktor makroekonomi global seperti harga komoditas, nilai tukar, dan geopolitik yang tidak tercakup dalam data berita lokal [17]; (3) ketidakseimbangan jumlah data berita antar emiten, terutama pada INDY, yang dapat memengaruhi kualitas representasi sentimen harian.

4. Kesimpulan

Penelitian ini berhasil mengimplementasikan dan membandingkan dua arsitektur prediksi harga saham sektor energi: LSTM unimodal berbasis data teknikal historis dan LSTM multimodal yang mengintegrasikan analisis sentimen berita berbahasa Indonesia menggunakan IndoBERT. Beberapa simpulan yang dapat ditarik adalah: Model LSTM unimodal terbukti efektif dalam memprediksi harga saham berbasis data historis dengan MAPE sebesar 3,62% (ADRO), 3,17% (PGAS), dan 5,90% (INDY), menunjukkan akurasi yang baik meskipun berbanding terbalik dengan tingkat volatilitas aset. Model IndoBERT berhasil melakukan klasifikasi sentimen berita ke dalam tiga kategori dengan akurasi dan F1-Score sebesar 83%, serta menghasilkan sentiment score numerik dalam rentang -1 hingga 1 yang merepresentasikan tingkat keyakinan model. Integrasi *sentiment score* melalui pendekatan Multimodal LSTM berhasil diimplementasikan, namun belum mampu mengungguli performa model unimodal (MAPE: 4,00% ADRO; 5,46% PGAS; 8,47% INDY), mengindikasikan bahwa sentimen berita lokal belum memberikan pengaruh signifikan pada saham sektor energi yang sangat dipengaruhi faktor makroekonomi global.

Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi arsitektur alternatif seperti *Bidirectional LSTM* atau *Transformer*, menambahkan variabel makroekonomi eksternal (harga komoditas, nilai tukar, suku bunga), memperluas sumber data berita, serta mengembangkan sistem dengan integrasi API data real-time agar prediksi dapat diperbarui secara otomatis.

Daftar Pustaka

- [1] D. I. Puteri, "Implementasi Long Short Term Memory (LSTM) dan Bidirectional Long Short Term Memory (BiLSTM) Dalam Prediksi Harga Saham Syariah," *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 11, no. 1, pp. 35–43, 2023, doi: 10.34312/euler.v11i1.19791.
- [2] A. S. Sanfa and B. B. Tjandrasa, "Pengaruh faktor-faktor makroekonomi pada dinamika Indeks Harga Saham Gabungan (IHSG)," *J. Manaj. Bisnis dan Kewirausahaan*, vol. 8, no. 6, pp. 1507–1521, 2024, doi: 10.24912/jmbk.v8i6.32167.
- [3] H. A. Utami, E. Herwiyanti, and I. Suparlinah, "Pengaruh Sentimen Pasar Terhadap Pergerakan Harga Saham Pada Perusahaan Industri Manufaktur Pada Bursa Efek Indonesia," *Bisnis dan Akunt.*, vol. 26, no. 1, pp. 25–34, 2024, [Online]. Available: www.idx.co.id.
- [4] R. Julian and M. R. Pribadi, "Peramalan Harga Saham Pertambangan Pada Bursa Efek Indonesia (BEI) Menggunakan Long Short Term Memory (LSTM)," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 3, pp. 1570–1580, 2021, doi: 10.35957/jatisi.v8i3.1159.
- [5] E. Pramudya, D. Retnoningsih, and D. Ruswanti, "Implementasi Metode LSTM untuk Prediksi Harga Saham PT Indofood CBP Sukses Makmur TBK," *J. Nas. Teknol. Inf. dan Apl.*, vol. 3, no. 4, pp. 933–940, 2025, [Online]. Available: <https://ejournal2.unud.ac.id/index.php/jnatia/article/view/186>
- [6] F. R. Arnanda and A. 'Azizah N. Rahmah, "Deteksi Pergerakan IHSG Berdasarkan Berita Daring Menggunakan Model Deep Learning Berbasis Transformer," *Semin. Nas. Off. Stat.*, vol. 2024, no. 1, pp. 439–448, 2024, doi: 10.34123/semnasoffstat.v2024i1.1951.
- [7] Y. Asri, D. Kuswardani, W. N. Suliyanti, Y. O. Manullang, and A. R. Ansyari, "Sentiment analysis based on Indonesian language lexicon and IndoBERT on user reviews PLN mobile application," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 38, no. 1, p. 677, 2025, doi: 10.11591/ijeecs.v38.i1.pp677-688.

- [8] R. Hidayat and W. Gata, "Pemanfaatan IndoBERT Untuk Analisis Sentimen Ulasan Aplikasi Depok Single Window," *Inf. Syst. Educ. Prof. J. Inf. Syst.*, vol. 10, no. 1, p. 13, 2025, doi: 10.51211/isbi.v10i1.3334. 263
264
- [9] S. Yacine, B. Abdenour, and A. Mohamed, "Early , intermediate and late fusion strategies for robust deep learning-based multimodal action recognition," 2021. 265
266
- [10] E. M. Rizqiyani, N. Y. Setiawan, and M. C. Saputra, "Studi Komparatif Algoritma Long Short Term Memory (Lstm) Dan Gated Recurrent Unit (Gru) Dalam Peramalan Harga Saham Pt. Fast Food Indonesia Tbk (Fast)," vol. 9, no. 11, pp. 2548–964, 2025, [Online]. Available: <http://j-ptiik.ub.ac.id> 267
268
269
- [11] A. N. H. Ika Apriliani, "ANALISIS TEKNIKAL SEBAGAI DASAR PENGAMBILAN KEPUTUSAN DALAM JUAL BELI SAHAM PADA JII-30 PERIODE," vol. 3, pp. 100–114, 2023. 270
271
- [12] G. Budiprasetyo, M. Hani'ah, and D. Z. Aflah, "Prediksi Harga Saham Syariah Menggunakan Algoritma Long Short-Term Memory (LSTM)," *J. Nas. Teknol. dan Sist. Inf.*, vol. 8, no. 3, pp. 164–172, 2023, doi: 10.25077/teknosi.v8i3.2022.164-172. 272
273
274
- [13] W. J. Kusoema, I. Ibrahim, P. Studi, T. Informatika, K. Bandung, and J. Barat, "Analisis Sentimen dalam Kasus Korupsi PT . Pertamina menggunakan Metode indoBERT dan RCNN Sentiment Analysis on the PT Pertamina Corruption Case using IndoBERT and RCNN Methods," vol. 14, pp. 2246–2257, 2025. 275
276
277
- [14] F. Putra, Ferdiansyah Permana. Dewi, Mukti Ratna. Hibatullah, "Analisis Sentimen Pasar melalui Berita Finansial untuke Prediksi Harga Saham PT Bank Rakyat Indonesia Tbk," *IJAI (Indonesian J. Appl. Informatics)*, vol. 9, p. 14, 2024. 278
279
280
- [15] Z. Li and C. Yang, "A Comparative Sentiment Analysis of Airline Customer Reviews Using Bidirectional Encoder Representations from Transformers (BERT) and Its Variants," 2024. 281
282
- [16] N. Afrianto, "Prediksi Saham dengan Bidirectional LSTM dan Analisis Sentimen," 2022. 283
- [17] N. Anggraini *et al.*, "PREDIKSI HARGA SAHAM SEKTOR ENERGI MENGGUNAKAN METODE SPATIAL TEMPORAL ATTENTION-BASED CONVOLUTIONAL NETWORK BERDASARKAN DATA TEKS DAN NUMERIK," vol. 9, no. 3, pp. 3872–3880, 2025. 284
285
286
- [18] A. Heiden, "Applying LSTM for Stock Price Prediction with Sentiment Analysis," 2021. 287
288
289