

PREDIKSI HARGA BITCOIN MENGGUNAKAN MODEL *HYBRID* LSTM DAN DISTILBERT DENGAN PENDEKATAN *LATE FUSION*

BITCOIN PRICE PREDICTION USING LSTM AND DISTILBERT HYBRID MODEL WITH A LATE FUSION APPROACH

Achmad Rizal¹, Intan Dzikria², Trio Adhi Pamungkas³

E-mail: ¹⁾ m11315816@mail.ntust.edu.tw, ²⁾ intandzikria@untag-sby.ac.id,
³⁾ trioadhipamungkas04@gmail.com

¹Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology

²Sistem dan Teknologi Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas, Universitas 17 Agustus 1945 Surabaya

³Teknik Informatika, Fakultas Teknologi Elektro dan Informatika Cerdas, Universitas 17 Agustus 1945 Surabaya

Abstrak

Bitcoin merupakan aset kripto dengan tingkat volatilitas yang tinggi, sehingga memunculkan tantangan dalam memprediksi pergerakan harganya secara akurat. Penelitian ini bertujuan untuk mengembangkan model prediksi harga Bitcoin mingguan dengan mengintegrasikan data historis harga dan data sentimen pasar menggunakan pendekatan *late fusion*. Metode yang digunakan mengombinasikan model *Long Short-Term Memory* (LSTM) untuk menganalisis pola data *time series* dengan model *Large Language Model* (LLM) berbasis DistilBERT untuk mengekstraksi sentimen dari berita dan media sosial. Data yang digunakan terdiri dari data historis harga Bitcoin periode 2022–2026 serta data teks sebanyak 19.468 berita yang diolah menjadi skor sentimen harian. Model LSTM dilatih untuk menghasilkan prediksi harga berdasarkan pola historis, sedangkan DistilBERT digunakan untuk klasifikasi sentimen tanpa proses *fine-tuning*. Kedua output model kemudian digabungkan menggunakan metode *weighted late fusion* dengan parameter bobot optimal yang ditentukan melalui *grid search*. Hasil penelitian menunjukkan bahwa model *late fusion* mampu meningkatkan akurasi prediksi dibandingkan dengan model LSTM tunggal, dengan nilai RMSE sebesar 0,0386, MAE sebesar 0,0289, dan koefisien determinasi (R^2) sebesar 0,9612. Selain itu, model mampu menghasilkan prediksi harga Bitcoin untuk 7 hari ke depan dengan tren kenaikan yang konsisten. Dengan demikian, integrasi data numerik dan data sentimen melalui pendekatan *late fusion* terbukti efektif dalam meningkatkan performa prediksi harga Bitcoin. Model yang diusulkan diharapkan dapat menjadi alternatif metode prediksi yang lebih akurat dan komprehensif bagi investor dan peneliti di bidang prediksi finansial.

Kata kunci: *Bitcoin, LSTM, LLM, Late Fusion, Prediksi Harga.*

Abstract

Bitcoin is a Cryptocurrency asset characterized by high volatility, which poses significant challenges in accurately predicting its price movements. This study aims to develop a weekly Bitcoin price prediction model by integrating historical price data and market sentiment using a late fusion approach. The proposed method combines a Long Short-Term Memory (LSTM) model for time series analysis with a Large Language Model (LLM), specifically DistilBERT, for extracting sentiment from news and social media data. The dataset consists of historical

Bitcoin price data from 2022 to 2026 and 19,468 textual data entries processed into daily sentiment scores. The LSTM model is trained to capture temporal patterns in historical data, while DistilBERT is utilized for sentiment classification without fine-tuning. The outputs of both models are then integrated using a weighted late fusion mechanism, with the optimal weight parameter determined through a grid search process. The results show that the late fusion model outperforms the standalone LSTM model, achieving an RMSE of 0.0386, MAE of 0.0289, and an R^2 score of 0.9612. Furthermore, the model successfully generates a 7-day ahead Bitcoin price prediction with a consistent upward trend. In conclusion, integrating numerical and sentiment data through a late fusion approach effectively improves Bitcoin price prediction performance. The proposed model is expected to serve as a more accurate and comprehensive alternative for investors and researchers in financial forecasting.

Keywords: *Bitcoin, LSTM, LLM, Late Fusion, Price Prediction.*

1. PENDAHULUAN

Cryptocurrency merupakan aset digital yang menggunakan teknologi blockchain untuk melakukan transaksi secara aman dan terdesentralisasi, independen dari bank sentral [1]. Bitcoin, sebagai *cryptocurrency* dengan kapitalisasi pasar terbesar, telah menarik perhatian luas dari investor, pedagang, dan peneliti di seluruh dunia. Di Indonesia, pertumbuhan aktivitas pasar kripto menunjukkan tren yang signifikan, dengan nilai transaksi aset kripto nasional mencapai Rp35,61 triliun pada April 2025 dan jumlah investor mencapai 14,16 juta orang pada periode yang sama [2], [3]. Sepanjang Januari hingga Juli 2025, total transaksi kripto telah mencapai Rp276,45 triliun dengan jumlah investor mencapai 16,5 juta orang pada Juli 2025 [4], [5]. Volatilitas harga Bitcoin yang tinggi menciptakan ketidakpastian bagi para pelaku pasar dan mendorong kebutuhan akan metode prediksi yang akurat untuk mendukung pengambilan keputusan investasi [6]. Metode tradisional seperti *AutoRegressive Integrated Moving average* (ARIMA) sering kali kurang mampu menangkap pola kompleks dari data time series yang bersifat nonlinier dan sangat fluktuatif [7]. Oleh karena itu, diperlukan pendekatan berbasis *deep learning* yang lebih adaptif dan mampu menangkap dinamika pasar secara lebih kontekstual [8].

Meskipun berbagai penelitian telah memanfaatkan LSTM untuk prediksi harga Bitcoin dengan data historis [9] dan LLM untuk analisis sentimen [10] secara terpisah, penelitian yang menggabungkan keduanya dalam satu kerangka terintegrasi masih jarang ditemukan [11]. Kondisi ini menunjukkan adanya kesenjangan penelitian dalam integrasi kedua pendekatan tersebut. Beberapa penelitian sebelumnya menunjukkan bahwa pendekatan *fusion* efektif dalam meningkatkan akurasi prediksi harga aset keuangan. [12] Chen et al. mengembangkan *deep fusion* model yang mengombinasikan data harga berbasis LSTM dengan berita pasar berbasis BERT melalui multimodal *deep fusion* terbukti meningkatkan akurasi prediksi dibandingkan dengan model tunggal. [13] Zhang et al. juga menunjukkan bahwa metode *fusion* yang menggabungkan sentimen dari teks dengan data pasar mampu menghasilkan prediksi Bitcoin yang lebih stabil dan akurat.

Berdasarkan temuan tersebut, penelitian ini mengusulkan pendekatan *late fusion* dengan mengintegrasikan LSTM untuk menganalisis pola harga Bitcoin berdasarkan data historis time series dan LLM DistilBERT untuk mengekstraksi sentimen dari berita dan media sosial terkait Bitcoin. Kedua keluaran model tersebut digabungkan melalui mekanisme *weighted fusion* dengan bobot optimal yang ditentukan secara empiris melalui *grid search*. Integrasi ini diharapkan mampu menghasilkan prediksi harga Bitcoin mingguan yang lebih akurat, komprehensif, dan bermanfaat bagi investor maupun pemangku kepentingan lainnya.

2. KAJIAN PUSTAKA

2.1. Large Language Model (LLM)

Large Language Model (LLM) merupakan model kecerdasan buatan berbasis *deep learning* yang dilatih menggunakan data teks berukuran sangat besar sehingga mampu mengenali pola bahasa, memahami konteks, dan melakukan klasifikasi sentimen secara efisien [14]. LLM dibangun di atas arsitektur *Transformer* yang memanfaatkan mekanisme *self-attention* untuk menangkap hubungan antar kata dalam kalimat [12]. Dalam konteks prediksi harga *cryptocurrency*, LLM terbukti mampu menangkap perubahan sentimen pasar secara *real-time* dari teks berita dan media sosial, suatu aspek yang sulit ditangkap oleh model numerik tradisional [15]. Salah satu varian LLM yang relevan adalah DistilBERT, model terkompresi dari BERT melalui teknik *knowledge distillation* dengan 6 *layer encoder* dan 66 juta parameter, yang mampu mempertahankan 97% kemampuan BERT dengan kecepatan inferensi 60% lebih tinggi [16].

2.2 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) merupakan pengembangan dari *Recurrent Neural Network* (RNN) yang dirancang untuk mengatasi masalah *vanishing gradient* dan *exploding gradient* pada RNN konvensional [17]. LSTM mempertahankan informasi jangka panjang melalui tiga gerbang utama yaitu *forget gate*, *input gate*, dan *output gate* yang dikendalikan fungsi aktivasi sigmoid. Keunggulan utama LSTM terletak pada kemampuannya menangkap pola nonlinier pada data *time series* dengan volatilitas tinggi, menjadikannya sangat relevan untuk prediksi harga Bitcoin [8]. [18] Pratama dan Utama membuktikan bahwa LSTM efektif dalam memodelkan ketergantungan jangka panjang pada data harga Bitcoin, menghasilkan tingkat kesalahan yang lebih rendah dibandingkan dengan metode konvensional seperti ARIMA. [17] Rizkilloh dan Widiyanesti juga membuktikan bahwa model LSTM memberikan hasil prediksi lebih akurat dengan nilai RMSE lebih rendah dibandingkan dengan metode tradisional pada data *time series* keuangan.

2.3 DistilBERT

DistilBERT merupakan varian terkompresi dari BERT yang dikembangkan melalui teknik *knowledge distillation*, di mana pengetahuan dari model BERT besar (*teacher model*) ditransfer ke model yang lebih ringan (*student model*) [19]. DistilBERT memiliki 6 *layer encoder* dan sekitar 66 juta parameter, mampu mempertahankan 97% kemampuan pemahaman bahasa alami BERT dengan kecepatan inferensi 60% lebih tinggi dan pengurangan parameter sebesar 40% [19]. Kemampuan ini menjadikan DistilBERT sangat efisien untuk klasifikasi sentimen pada skala besar tanpa mengorbankan akurasi. [20] Fajri et al. membuktikan bahwa DistilBERT menghasilkan tingkat akurasi yang baik dan secara signifikan mengungguli BERT pada dataset yang sama. Dalam penelitian ini, DistilBERT dimanfaatkan tanpa *fine-tuning* sebagai model klasifikasi sentimen dari teks berita dan media sosial terkait Bitcoin, menghasilkan label sentimen per teks yang kemudian diagregasi menjadi skor sentimen harian.

2.4 Late Fusion

Late fusion merupakan pendekatan integrasi multimodal yang menggabungkan keluaran dari beberapa model yang telah diproses secara independen pada tahap akhir pengambilan keputusan [21]. Berbeda dengan *early fusion* yang menggabungkan fitur mentah di awal, *late fusion* memungkinkan setiap model untuk mengoptimalkan representasinya masing-masing sebelum digabungkan, sehingga menghasilkan prediksi yang lebih robust [21]. [12] mengimplementasikan arsitektur *late fusion* yang menggabungkan LSTM dan FinBERT melalui *concatenation* dan *fully connected layer* untuk prediksi harga saham, terbukti meningkatkan akurasi dibandingkan model tunggal. [13] Juga menunjukkan bahwa *late fusion* yang mengintegrasikan data harga historis dengan sentimen pasar menghasilkan prediksi Bitcoin yang lebih stabil dan akurat. Dalam penelitian ini, mekanisme *weighted fusion* digunakan untuk

menggabungkan output prediksi LSTM dan skor sentimen harian DistilBERT sebagaimana ditunjukkan pada Persamaan (1), di mana $\hat{y}$ adalah prediksi harga Bitcoin akhir, α adalah bobot *fusion* optimal sebesar 0,985 yang ditentukan secara empiris melalui *grid search*, $\hat{y}_{LSTM}$ adalah output prediksi model LSTM dalam skala ternormalisasi (0–1), dan $S_{sentiment}$ adalah skor sentimen harian ternormalisasi hasil agregasi klasifikasi DistilBERT.

$$\hat{y} = \alpha \times \hat{y}_{LSTM} + (1 - \alpha) \times S_{sentiment} \quad (1)$$

3. METODOLOGI

Penelitian ini menggunakan dua sumber data utama yang saling melengkapi. Data historis harga Bitcoin diperoleh dari Yahoo Finance untuk periode 1 Januari 2022 hingga 22 April 2026, terdiri dari 1.573 baris data harian sebelum preprocessing yang mencakup atribut *open*, *high*, *low*, *close*, dan *volume*. Data teks berita dan media sosial sebanyak 19.468 baris dikumpulkan menggunakan API GDELT (*Global Database of Events, Language, and Tone*) dan News API untuk periode yang sama, mencakup atribut tanggal, teks, dan label sentimen. Pembagian data dilakukan secara kronologis dengan rasio 80:20 untuk mempertahankan konsistensi temporal.

3.1 Pemrosesan Data LSTM

Data historis harga Bitcoin diproses melalui beberapa tahapan. Pertama, dilakukan *feature engineering* dengan menambahkan indikator teknikal berupa MA7 (*moving average* 7 hari), MA14 (*moving average* 14 hari), dan *returns* (persentase perubahan harga). Kemudian data dinormalisasi menggunakan *Min-Max Scaler* ke rentang 0–1 menggunakan Persamaan (2), dimana x adalah nilai asli, $\min(x)$ adalah nilai minimum, dan $\max(x)$ adalah nilai maksimum [17].

$$x_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

Data historis harga Bitcoin kemudian diubah ke format *sequence* menggunakan *sliding window* dengan panjang 30 hari, di mana 30 hari berturut-turut digunakan sebagai input untuk memprediksi harga pada hari ke-31. Arsitektur LSTM yang digunakan terdiri dari tiga lapisan bertingkat dengan masing-masing 128, 64, dan 32 unit, dropout 0,2 pada setiap lapisan, fungsi aktivasi ReLU, optimizer Adam dengan learning rate 0,001, maksimal 100 epoch dengan EarlyStopping patience=10, serta satu lapisan Dense sebagai keluaran untuk menghasilkan prediksi harga dalam skala ternormalisasi (0–1).

3.2 Pemrosesan Data DistilBERT

Proses pemrosesan data DistilBERT meliputi beberapa tahapan: (1) Pembersihan data teks meliputi *lowercase*, hapus duplikat, dropna, dan *strip whitespace*; (2) Tokenisasi menggunakan AutoTokenizer DistilBERT berbasis subword dengan panjang maksimal 128 token, menambahkan token khusus [CLS] dan [SEP]; (3) Klasifikasi sentimen menghasilkan label per teks: 0 (negatif), 1 (netral), dan 2 (positif); (4) Agregasi sentimen harian dengan menghitung nilai rata-rata (*mean*), jumlah data (*count*), dan standar deviasi (*std*) per hari; serta (5) Normalisasi skor sentimen ke rentang 0–1 menggunakan formula $S = (\text{mean} - 1) / 2$ agar sebanding dengan output prediksi LSTM.

3.3 Metrik Evaluasi

Evaluasi performa model dilakukan menggunakan tiga metrik utama, yaitu *root mean square error* (RMSE), *mean absolute error* (MAE), dan koefisien determinasi R^2 . Ketiga metrik ini digunakan secara komparatif untuk mengevaluasi peningkatan kinerja antara model LSTM saja dengan model *Late fusion* yang diusulkan [22].

RMSE mengukur tingkat penyimpangan antara nilai aktual dan nilai prediksi, dengan memberikan penalti lebih besar terhadap kesalahan yang lebih besar karena menggunakan kuadrat selisih. Persamaan 3 menunjukkan formula RMSE, dimana $\hat{Y}_i$ adalah nilai prediksi, dan n adalah jumlah data [18]. Standar nilai RMSE berkisar antara 0 hingga 1 pada data yang telah dinormalisasi, dengan nilai di bawah 0,05 mengindikasikan bahwa model tergolong *good fit* dan memiliki tingkat kesalahan prediksi yang sangat rendah [18].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

MAE mengukur rata-rata besar kesalahan absolut antara nilai aktual dan nilai prediksi secara linier, yang dihitung menggunakan Persamaan (4). Standar nilai MAE pada data yang telah dinormalisasi berkisar antara 0 hingga 1, dengan nilai di bawah 0,05 mengindikasikan bahwa model memiliki tingkat kesalahan absolut yang sangat rendah dan tergolong *good fit*. Koefisien determinasi R^2 mengukur seberapa baik model mampu menjelaskan variabilitas data, dengan nilai mendekati 1 menunjukkan akurasi tinggi, seperti yang ditunjukkan pada Persamaan (5). Nilai R^2 dikatakan baik apabila bernilai di atas 0,75, yang berarti model mampu menjelaskan lebih dari 75% variabilitas data. Nilai R^2 di atas 0,90 dikategorikan sangat baik dan menunjukkan bahwa model memiliki kemampuan prediksi yang sangat tinggi terhadap pola data yang diuji

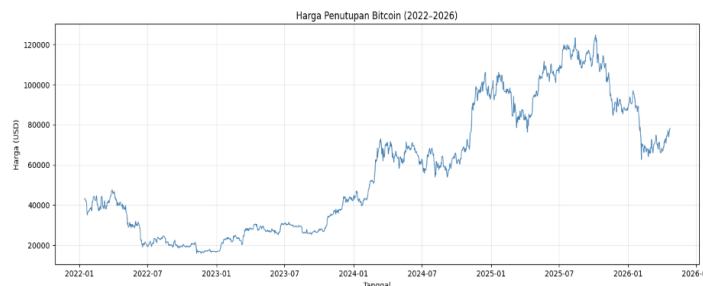
$$MAE = \frac{1}{n} \sum_{i=1}^n |Y'_i - Y_i| \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (5)$$

4. HASIL DAN PEMBAHASAN

4.1 Hasil Pelatihan Model Data

Dataset harga Bitcoin terdiri dari 1.560 data harian setelah *preprocessing* dengan periode 1 Januari 2022 hingga 22 April 2026. Statistik deskriptif data harga penutupan (*close price*) menunjukkan nilai rata-rata sebesar \$57.677, nilai minimum sebesar \$15.787, dan nilai maksimum sebesar \$124.752. Standar deviasi sebesar \$31.785 mencerminkan tingkat volatilitas yang sangat tinggi pada harga Bitcoin selama periode tersebut.



Berdasarkan grafik harga penutupan Bitcoin pada Gambar 1, terdapat tiga fase utama pergerakan harga: (1) penurunan signifikan pada periode 2022 hingga pertengahan 2023 dari kisaran \$40.000 hingga titik terendah sekitar \$15.000; (2) tren kenaikan konsisten pada 2023 hingga awal 2025 yang mencapai puncak di kisaran \$124.000; dan (3) koreksi pada akhir periode ke kisaran \$75.000–\$78.000.

Dataset sentimen terdiri dari 19.468 data teks dengan rentang periode 1 Januari 2022 hingga 22 April 2026, mencakup 996 hari unik dengan rata-rata 17 teks per hari. Distribusi label sentimen menunjukkan: label netral (1) mendominasi dengan 8.795 data (45,2%), diikuti label negatif (0) sebanyak 5.465 data (28,1%), dan label positif (2) sebanyak 5.207 data (26,7%). Distribusi yang lebih merata dibandingkan dengan dataset sentimen konvensional mengindikasikan adanya dinamika opini pasar yang beragam. Peningkatan frekuensi pemberitaan yang signifikan mulai pertengahan 2024 berkorelasi dengan kenaikan harga Bitcoin yang tajam pada periode yang sama, mengindikasikan hubungan antara volume pemberitaan dan pergerakan harga.

4.2 Hasil Pelatihan Model LSTM

Proses pelatihan model LSTM berjalan selama 43 epoch dari maksimal 100 yang ditetapkan. Mekanisme *EarlyStopping* menghentikan pelatihan secara otomatis karena *validation loss* tidak mengalami perbaikan selama 10 *epoch* berturut-turut. Model terbaik diperoleh pada *epoch* ke-33 dengan *validation loss* sebesar 0,001168, *training loss* akhir sebesar 0,001230, dan *validation loss* akhir sebesar 0,002109. Gambar 2 menunjukkan perbandingan antara *training loss* dan *validation loss*, di mana *training loss* menurun secara konsisten dan stabil, sementara *validation loss* berfluktuasi pada *epoch* 1–25, namun secara umum mengikuti tren penurunan. Fluktuasi ini mencerminkan volatilitas inheren dari data harga Bitcoin.



Gambar 2. Grafik Training vs Validations loss

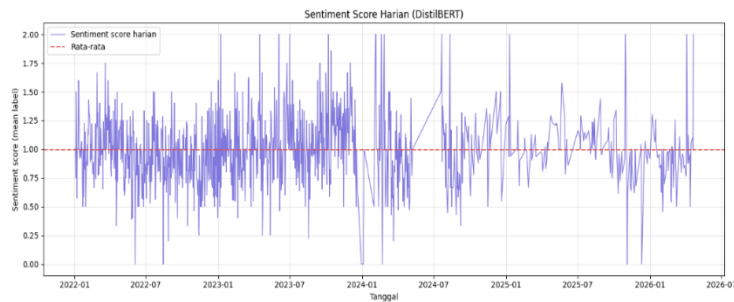
Tabel 1. Evaluasi Model LSTM

Metrik	Nilai	Interpretasi
RMSE	0,0400	Nilai di bawah 0,05 mengindikasikan model <i>good fit</i>
MAE	0,0300	Nilai di bawah 0,05 mengindikasikan model <i>good fit</i>
R ²	0,9565	Model mampu menjelaskan 95,65% variabilitas harga Bitcoin

Hasil evaluasi pada Tabel 1 menunjukkan bahwa model LSTM tergolong *good fit* dengan nilai RMSE sebesar 0,0400 dan MAE sebesar 0,0300 yang keduanya berada di bawah ambang batas 0,05. Nilai R² sebesar 0,9565 mengindikasikan model mampu menjelaskan 95,65% variabilitas harga Bitcoin, yang tergolong sangat baik untuk prediksi aset kripto dengan volatilitas tinggi. Model juga mampu mengikuti tren pergerakan harga dengan baik, termasuk fase penurunan tajam dan fase pemulihan, sejalan dengan penelitian sebelumnya yang menyatakan bahwa LSTM dengan arsitektur bertingkat mampu menangkap dependensi jangka panjang pada data deret waktu keuangan secara lebih efektif.

4.3 Hasil Pemrosesan DistilBERT

Proses klasifikasi sentimen menggunakan model DistilBERT pretrained tanpa *fine-tuning* berhasil mengklasifikasikan 19.467 data teks dari 996 hari unik. Skor sentimen harian memiliki nilai rata-rata 1,0334, minimum 0,0, dan maksimum 2,0, sesuai dengan rentang label yang digunakan (0, 1, 2).



Gambar 3. Grafik Sentiment Score harian

Grafik Sentiment Score harian pada Gambar 3 menunjukkan fluktuasi di sekitar nilai 1,0 (netral) sepanjang periode 2022–2026, dengan beberapa lonjakan skor ke arah positif (mendekati 2,0) maupun penurunan ke arah negatif (mendekati 0,0) yang terjadi sesekali, terutama pada periode 2022–2023 yang bertepatan dengan fase penurunan harga Bitcoin yang signifikan. Mulai pertengahan 2024, fluktuasi skor sentimen terlihat lebih stabil sejalan dengan stabilisasi volume pemberitaan.

4.4 Hasil Late Fusion

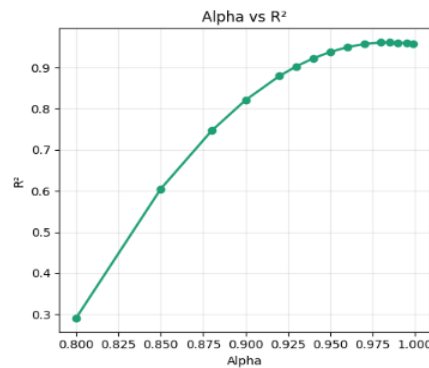
Proses *grid search* dilakukan pada rentang $\alpha = 0,80$ hingga 0,999 untuk menentukan bobot optimal. Setiap nilai α dievaluasi menggunakan ketiga metrik (R^2 , RMSE, MAE) pada data *test* yang terdiri dari 244 observasi. Nilai α optimal yang menghasilkan R^2 tertinggi dipilih sebagai bobot penggabungan akhir, seperti yang ditunjukkan pada Tabel 2. Berdasarkan hasil *grid search*, nilai α optimal adalah 0,985 yang menghasilkan R^2 tertinggi sebesar 0,9612 dengan RMSE 0,0386 dan MAE 0,0289. Nilai R^2 sebesar 0,9612 menunjukkan bahwa model Late Fusion mampu menjelaskan 96,12% variabilitas harga Bitcoin, yang tergolong sangat baik karena melampaui ambang batas 0,90. Nilai RMSE dan MAE yang berada di bawah 0,05 mengindikasikan model tergolong *good fit*.

Tabel 2. Hasil *Grid search* Alpha

Alpha (α)	RMSE	MAE	R^2
0,800	0,1649	0,1546	0,2910
0,900	0,0829	0,0720	0,8207
0,950	0,0488	0,0393	0,9378
0,970	0,0406	0,0315	0,9571
0,980	0,0388	0,0293	0,9608
0,985 *	0,0386	0,0289	0,9612 *
0,990	0,0389	0,0288	0,9606
0,999	0,0406	0,0302	0,9571

* Nilai optimal

Gambar 4 menunjukkan bahwa peningkatan nilai α dari 0,800 ke 0,985 secara konsisten meningkatkan performa model, namun setelah melewati titik optimal tersebut performa mulai menurun. Hal ini mengindikasikan bahwa kontribusi *sentiment score* yang terlalu besar justru mengurangi akurasi prediksi, sehingga diperlukan keseimbangan yang tepat antara informasi harga historis dan informasi sentimen.



Gambar 4. Grafik Alpha vs R²

Seperti yang ditunjukkan pada Tabel 3, hasil penelitian ini menunjukkan bahwa model Late Fusion menghasilkan performa lebih baik dibandingkan dengan LSTM pada semua metrik evaluasi. RMSE turun dari 0,0400 menjadi 0,0386 (turun 3,5%), MAE turun dari 0,0300 menjadi 0,0289 (turun 3,7%), dan R² meningkat dari 0,9565 menjadi 0,9612. Nilai RMSE sebesar 0,0386 dan MAE sebesar 0,0289 keduanya berada di bawah ambang batas 0,05 yang mengindikasikan bahwa model Late Fusion tergolong good fit, dengan tingkat kesalahan prediksi yang lebih rendah dibandingkan model LSTM yang memiliki RMSE 0,0400 dan MAE 0,0300.

Peningkatan ini menunjukkan bahwa penambahan informasi sentimen harian dari DistilBERT melalui mekanisme weighted fusion memberikan kontribusi positif terhadap akurasi prediksi. Nilai bobot optimal $\alpha = 0,985$ menunjukkan kontribusi LSTM yang dominan (98,5%) dibandingkan sentiment score (1,5%), mengindikasikan bahwa data historis harga tetap menjadi faktor penentu utama, sementara sentimen berperan sebagai faktor penyesuaian minor yang tetap memberikan kontribusi positif. Temuan ini konsisten dengan yang menunjukkan bahwa integrasi informasi sentimen melalui mekanisme late fusion mampu meningkatkan akurasi prediksi harga Bitcoin.

Tabel 3. Peningkatan Penggunaan Model

Model	RMSE	MAE	R²
LSTM	0,0400	0,0300	0,9565
Late fusion ($\alpha=0,985$)	0,0386	0,0289	0,9612
Peningkatan	-3,5%	-3,7%	+0,0047

4.5 Pembahasan

Hasil penelitian menunjukkan bahwa pendekatan late fusion dengan mekanisme weighted fusion terbukti lebih efektif dibandingkan model LSTM tunggal. Keberhasilan ini didorong oleh kemampuan DistilBERT menangkap sinyal sentimen pasar yang tidak tercermin dalam data harga historis, serta mekanisme grid search yang memungkinkan penentuan bobot optimal secara sistematis.

Dominasi bobot LSTM ($\alpha=0,985$) dalam model *fusion* sejalan dengan temuan [20] yang menyatakan bahwa sentimen pasar dari media sosial berkontribusi sebagai faktor pelengkap,

bukan pengganti, dalam prediksi harga *cryptocurrency*. Distribusi sentimen yang lebih merata pada dataset penelitian ini (45,2% netral, 28,1% negatif, 26,7% positif) berbeda dari dataset sentimen konvensional yang biasanya didominasi oleh label netral, mengindikasikan bahwa data berita yang dikumpulkan mencerminkan dinamika pasar yang lebih beragam. Dibandingkan dengan penelitian sebelumnya, hasil penelitian ini menunjukkan performa yang kompetitif. Chen et al. [12] melaporkan peningkatan akurasi melalui deep fusion LSTM-BERT pada data saham, sementara penelitian ini memperoleh $R^2 = 0,9612$ pada data Bitcoin yang memiliki volatilitas lebih tinggi. Zhang et al. [13] menggunakan arsitektur SentiStack berbasis LSTM dengan fusion sentimen dan memperoleh hasil yang stabil, konsisten dengan temuan penelitian ini bahwa sentimen berperan sebagai faktor penyesuaian yang memperkuat prediksi harga historis. Perbedaan utama penelitian ini terletak pada penggunaan DistilBERT tanpa fine-tuning sebagai ekstraktor sentimen, yang terbukti efisien secara komputasi namun tetap mampu memberikan kontribusi positif terhadap akurasi prediksi akhir.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan pendekatan late fusion dengan mengintegrasikan model LSTM dan DistilBERT untuk prediksi harga Bitcoin mingguan. Beberapa kesimpulan utama yang diperoleh adalah sebagai berikut. Pertama, model LSTM dengan arsitektur tiga lapisan bertingkat (128-64-32 unit) dan *sliding window* 30 hari berhasil memprediksi harga Bitcoin dengan $R^2 = 0,9565$, RMSE = 0,0400, dan MAE = 0,0300, menunjukkan kemampuan yang sangat baik dalam menangkap pola temporal data harga Bitcoin. Kedua, mekanisme *weighted late fusion* dengan bobot optimal $\alpha = 0,985$ yang ditentukan melalui *grid search* pada rentang 0,80–0,999 terbukti meningkatkan performa prediksi dengan $R^2 = 0,9612$, RMSE = 0,0386, dan MAE = 0,0289. Ketiga, model *Late fusion* menghasilkan peningkatan pada semua metrik evaluasi dibandingkan dengan LSTM saja, dengan penurunan RMSE sebesar 3,5% dan MAE sebesar 3,7%, membuktikan bahwa integrasi informasi sentimen melalui mekanisme *weighted fusion* memberikan nilai tambah terhadap akurasi prediksi. Keempat, prediksi harga Bitcoin untuk 7 hari ke depan (23–29 April 2026) menunjukkan tren kenaikan bertahap dari \$77.030 hingga \$78.249, konsisten dengan tren pasar aktual pada periode tersebut.

Keterbatasan penelitian meliputi pola prediksi yang cenderung monoton akibat metode autoregressive, kontribusi sentimen yang relatif kecil karena dominasi label netral, serta tidak dipertimbangkannya faktor makroekonomi eksternal. Untuk penelitian selanjutnya, disarankan mengeksplorasi model sentimen berbasis fine-tuning seperti FinBERT, menambahkan fitur makroekonomi, serta menguji pendekatan early fusion atau hybrid fusion sebagai pembanding.

Secara akademis, penelitian ini berkontribusi dalam menunjukkan efektivitas pendekatan *weighted late fusion* sebagai alternatif integrasi model LSTM dan DistilBERT untuk prediksi harga aset kripto. Secara praktis, hasil penelitian ini dapat dimanfaatkan oleh investor dan analis pasar sebagai referensi dalam pengambilan keputusan berbasis kombinasi data historis dan sentimen pasar. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi penggunaan model sentimen yang telah di-fine-tuning pada domain keuangan seperti FinBERT, mempertimbangkan faktor makroekonomi eksternal sebagai fitur tambahan, serta menguji pendekatan early fusion atau hybrid fusion sebagai perbandingan terhadap late fusion yang diusulkan dalam penelitian ini.

6. DAFTAR RUJUKAN

- [1] Fegiyanto, R., Hermawan, A. and Ardiani, F., 2024. Prediksi Harga Crypto dengan Algoritma Jaringan Saraf Tiruan. *Manajemen Informatika dan Komunikasi*, 5(3),

- pp.2265–2275.
- [2] Antara, 2025. Nilai transaksi aset kripto Rp35,61 triliun per April 2025. [Online] Available at: <https://www.antaranews.com/berita/4873889/nilai-transaksi-aset-kripto-rp3561-triliun-per-april-2025> [Accessed 2 October 2025].
 - [3] Andrianto, R., 2025. Jumlah Investor Kripto RI Tembus 13,71 Juta, Transaksi Rp 32,45 T. [Online] Available at: <https://www.cnbcindonesia.com/market/20250509144634-17-632460/jumlah-investor-kripto-ri-tembus-1371-juta-transaksi-rp-3245-t> [Accessed 2 October 2025].
 - [4] Arini, S.C., 2025. Nilai Transaksi Kripto RI Melesat Tembus Rp 276,45 T. [Online] Available at: <https://finance.detik.com/fintech/d-8095917/nilai-transaksi-kripto-ri-melesat-tembus-rp-276-45-t> [Accessed 2 October 2025].
 - [5] Aprilia, Z., 2025. OJK Percepat Peralihan Wewenang dari Bappebti. [Online] Available at: <https://www.cnbcindonesia.com/market/20250811162856-17-657001/ojk-percepat-peralihan-wewenang-dari-bappebti> [Accessed 2 October 2025].
 - [6] Nugraha, D.M. and Ariatmanto, D., 2025. Meningkatkan Akurasi Prediksi Harga Bitcoin dengan Algoritma GRU-LSTM Hibrida. *Jurnal Buffer Informatika*, 11(April), pp.16–26.
 - [7] Raju, S.M. and Tarif, A.M., 2020. Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis. *arXiv preprint*, arXiv:2006.14473. [Online] Available at: <https://arxiv.org/abs/2006.14473> [Accessed 2 October 2025].
 - [8] Gurgul, V., Lessmann, S. and Härdle, W.K., 2025. Deep learning and NLP in cryptocurrency forecasting: Integrating financial, blockchain, and social media data. *International Journal of Forecasting*, 41(4), pp.1666–1695..
 - [9] Htay, H.S., Ghahremani, M. and Shiaeles, S., 2025. Enhancing Bitcoin Price Prediction with Deep Learning: Integrating Social Media Sentiment and Historical Data. *Applied Sciences*, 15(3), pp.1–21..
 - [10] Merdiansah, R., Siska and Ridha, A.A., 2024. Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, 7(1), pp.221–228.
 - [11] Vallarino, D., 2025. An AI-Enhanced Forecasting Framework: Integrating LSTM and Transformer-Based Sentiment for Stock Price Prediction. *Journal of Economic Analysis*, 4(3), pp.1–15.
 - [12] Chen, P., Boukouvalas, Z. and Corizzo, R., 2024. *A deep fusion model for stock market prediction with news headlines and time series data*. Vol. 36, No. 34. London: Springer.
 - [13] Zhang, Z., Jiang, C. and Lu, M., 2025. Fusion of Sentiment and Market Signals for Bitcoin Forecasting: A SentiStack Network Based on a Stacking LSTM Architecture. *Preprint*, pp.1–29.
 - [14] Nada, A.S. and Haviana, S.F.C., 2025. Implementasi Zero-Shot Learning Untuk Prediksi Solusi Dari Kalimat Masalah Pada Artikel Ilmiah Menggunakan Large Language Models (LLM). *Jurnal Transistor Elektro dan Informatika (TRANSISTOR EI)*, 7(1), pp.28–36.
 - [15] Natzir, S.M. and Jatiprasetya, H., 2025. Prediksi harga cryptocurrency xlm menggunakan metode deep learning lstm dan gru. *HOAQ: Jurnal Teknologi Informasi*, 16(1), pp.49–58.
 - [16] Hussain, M. et al., 2025. Optimised knowledge distillation for efficient social media emotion recognition using DistilBERT. *Journal of Computer Networks, Architecture and*

- High Performance Computing*, 7(2), pp.1–16.
- [17] Rizkilloh, M.F. and Widiyanesti, S., 2022. Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM). *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(1), pp.25–31.
- [18] Pratama, M.L. and Utama, H., 2023. Pendekatan Deep Learning Menggunakan Metode LSTM untuk Prediksi Harga Bitcoin. [Online] Available at: <https://share.google/0LkqF5rnRFkEloRt> [Accessed 2 October 2025].
- [19] Putri, M., Sutanto, T.E. and Inna, S., 2023. Studi Empiris Model BERT dan DistilBERT: Analisis Sentimen pada Pemilihan Presiden Indonesia. *Indonesian Journal of Computer Science*, 12(1), pp.2972–2980.
- [20] Fajri, F., Tutuko, B. and Sukemi, S., 2022. Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter. *Jurnal*, 8(2), pp.71–80.
- [21] Pandeya, Y.R. and Lee, J., 2021. Deep learning-based late fusion of multimodal information for emotion classification of music video. *Multimedia Tools and Applications*, pp.2887–2905.
- [22] Hodson, T.O., 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, (2), pp.5481–5487.
- [23] Ramadhan, F.A. and Nathasia, N.D., 2025. Prediksi Harga Dan Kinerja Aset Bitcoin Menggunakan Algoritma Long Short-Term Memory. *Jurnal FASILKOM*, 15(1), pp.68–76.