

INTEGRASI INDOBERT DAN LONG SHORT-TERM MEMORY UNTUK KLASIFIKASI SENTIMEN
PASAR BITCOIN BERBASIS MEDIA SOSIAL XHelmi Septianto¹, Yusuf Danda Purwanto², Andi Riansyah³Program Studi Teknik Informatika, Fakultas Teknologi Industri,
Universitas Islam Sultan Agung semarang;helmiseptianto14@gmail.com; permata.arm@gmail.com; andi@unissula.ac.id* Korespondensi : helmiseptianto14@gmail.com

Abstract: *Bitcoin's price is notoriously volatile, and much of that volatility traces back to how investors feel about the market at any given moment, which makes an automatic system for reading Indonesian-language sentiment genuinely useful. Word-embedding approaches that treat each word in isolation tend to fall short here because they cannot hold onto sentence-level context or the way meaning unfolds across a sequence of informal, slang-heavy words. This study builds a three-class sentiment classifier (positive, neutral, negative) for the Bitcoin market by pairing IndoBERT, used purely as a contextual feature extractor, with a Long Short-Term Memory (LSTM) layer that learns the sequential pattern on top of it. The dataset came from crawling X (formerly Twitter) for Indonesian-language tweets between January and December 2024, yielding 3,579 raw posts that were cleaned and manually labeled down to 3,542 usable tweets with a mild class imbalance. Preprocessing covered cleaning, case folding, normalization of crypto slang and emoji, and a selective stopword removal that deliberately kept negation words intact. All twelve IndoBERT encoder layers were frozen, and only a 128-unit LSTM with 0.4 dropout sat on top, trained with Focal Loss and class weighting to keep the imbalance from skewing predictions, then evaluated on a split of 2,856 training, 339 validation, and 347 test tweets. The resulting hybrid model reached 90.20% accuracy, 91.13% macro precision, 90.33% macro recall, and a 90.58% macro F1-score, performing best on the negative class (F1-score 92.19%) while most of its remaining confusion involved positive tweets being read as neutral, largely because of ambiguous crypto slang. Overall, combining IndoBERT with LSTM produces a Bitcoin sentiment classifier accurate enough to serve as one input among several for investors weighing their next move.*

Keywords: *Bitcoin, IndoBERT, LSTM, X(Media Sosial), Sentimen Pasar.*

Intisari: Harga Bitcoin dikenal sangat fluktuatif, dan sebagian besar pergerakannya lebih banyak dipicu oleh suasana hati pasar ketimbang faktor fundamental, sehingga sistem yang mampu membaca sentimen berbahasa Indonesia secara otomatis menjadi sesuatu yang cukup dibutuhkan. Pendekatan embedding kata statis kurang memadai untuk kebutuhan ini karena tidak dapat menangkap konteks kalimat maupun urutan makna pada teks media sosial yang informal dan sarat istilah slang. Penelitian ini membangun sistem klasifikasi sentimen pasar Bitcoin tiga kelas (positif, netral, negatif) dengan menggabungkan IndoBERT yang difungsikan murni sebagai pengekstraksi fitur kontekstual, dengan lapisan Long Short-Term Memory (LSTM) yang mempelajari pola urutannya. Dataset diperoleh melalui crawling tweet

Article History

Received: September 2026

Reviewed: September 2026

Published: September 2026

Plagiarism Checker

No 237.u873247.fh8374

Copyright : Author

Publish by : Kohesi



This work is licensed
under a [Creative Commons
Attribution-
NonCommercial 4.0
International License](https://creativecommons.org/licenses/by-nc/4.0/)



berbahasa Indonesia di platform X selama Januari-Desember 2024, menghasilkan 3.579 tweet mentah yang setelah dibersihkan dan diberi label manual menyisakan 3.542 tweet dengan tingkat ketidakseimbangan kelas yang tergolong ringan. Tahap preprocessing mencakup pembersihan teks, case folding, normalisasi slang dan emoji khas kripto, serta stopword removal selektif yang secara sengaja tetap mempertahankan kata negasi. Seluruh dua belas lapisan encoder IndoBERT dibekukan, sementara hanya lapisan LSTM berdimensi 128 unit dengan dropout 0,4 yang dilatih menggunakan Focal Loss dan pembobotan kelas, lalu diuji pada pembagian data 2.856 data latih, 339 data validasi, dan 347 data uji. Model hybrid yang dihasilkan mencapai akurasi 90,20%, presisi makro 91,13%, recall makro 90,33%, dan F1-score makro 90,58%, dengan performa terbaik pada kelas negatif (F1-score 92,19%), sedangkan kekeliruan yang tersisa sebagian besar berupa tweet positif yang terbaca sebagai netral akibat ambiguitas istilah slang kripto. Secara keseluruhan, integrasi IndoBERT dan LSTM menghasilkan sistem klasifikasi sentimen pasar Bitcoin yang cukup akurat untuk dijadikan salah satu bahan pertimbangan investor dalam mengambil keputusan.

Kata Kunci: Bitcoin, IndoBERT, LSTM, X (Media Sosial), Sentimen Pasar.

1. PENDAHULUAN

Aset kripto, terutama Bitcoin, tumbuh pesat dan menjadi salah satu instrumen investasi digital yang paling banyak diminati, tidak terkecuali di Indonesia. Namun Bitcoin juga dikenal luas sebagai aset dengan volatilitas ekstrem, dan pergerakan harganya sering kali tidak sejalan dengan logika fundamental ekonomi konvensional, melainkan lebih peka terhadap sentimen pasar dan spekulasi para investornya [3]. Kondisi ini membuat keputusan berinvestasi di Bitcoin menyimpan risiko yang tidak kecil, khususnya bagi investor ritel yang porsinya mendominasi pasar.

Dominasi investor ritel inilah yang membuat pengaruh sentimen publik terhadap dinamika harga menjadi semakin terasa. Aktivitas jual-beli yang dilakukan investor individu -- yang keputusannya sering kali terbentuk dari informasi yang beredar di media sosial -- terbukti berkorelasi cukup kuat dengan volatilitas harga Bitcoin [11]. Artinya, opini yang berkembang di ruang digital bukan sekadar riuh-rendah percakapan, melainkan turut membentuk arah pergerakan pasar itu sendiri.

X (dahulu Twitter) menjadi salah satu tempat utama di mana opini dan informasi seputar Bitcoin beredar secara real-time. Volume percakapan dan pencarian terkait Bitcoin di platform-platform digital bahkan terbukti berhubungan erat dengan aktivitas transaksi dan volatilitas pasar [4]. Dengan kata lain, isi obrolan di media sosial dapat dibaca sebagai representasi sentimen kolektif investor yang berpotensi dijadikan salah satu indikator dalam pengambilan keputusan investasi.

Persoalannya, volume data yang besar serta karakteristik bahasa yang dipakai di media sosial -- bahasa tidak baku, slang kripto, emoji, ditambah noise berupa spam dan konten tidak relevan -- membuat analisis sentimen secara manual menjadi pekerjaan yang berat dan sulit diskalakan [2], [17]. Karena itu dibutuhkan pendekatan otomatis yang cukup andal untuk mengolah data teks semacam ini secara efektif.

Pendekatan berbasis Natural Language Processing (NLP) dengan memanfaatkan Deep Learning menjadi jalan keluar yang relevan. IndoBERT, sebagai model representasi bahasa berbasis Transformer yang dilatih khusus pada korpus Bahasa Indonesia, telah berulang kali terbukti unggul dalam memahami konteks semantik teks tidak terstruktur, termasuk untuk

keperluan deteksi emosi dan klasifikasi sentimen [1], [7]. Di sisi lain, Long Short-Term Memory (LSTM) sebagai varian Recurrent Neural Network dikenal cakap dalam menangkap ketergantungan sekuensial dan pola urutan kata, dan telah terbukti meningkatkan akurasi prediksi pada berbagai tugas berbasis teks maupun time-series [6], [16].

Menggabungkan kedua model tersebut menjadi arsitektur hybrid dinilai menjanjikan karena memungkinkan model memahami konteks bahasa secara mendalam sekaligus tetap peka terhadap pola urutan informasi dalam kalimat. Pendekatan hibrida semacam ini terbukti mampu mendongkrak akurasi analisis sentimen dan prediksi pasar kripto dibandingkan model tunggal [9], dan informasi sentimen yang dihasilkannya berpotensi menjadi dasar yang berguna dalam pengambilan keputusan investasi, sebagaimana ditunjukkan oleh penelitian yang memanfaatkan sentimen investor untuk memprediksi indikator keuangan [8].

Berangkat dari uraian tersebut, penelitian ini berfokus pada perancangan model klasifikasi sentimen pasar Bitcoin berbasis data media sosial X, dengan mengintegrasikan IndoBERT sebagai pengekstraksi fitur kontekstual dan LSTM sebagai pengolah pola sekuensialnya. Kombinasi ini diharapkan mampu mengklasifikasikan sentimen pasar Bitcoin ke dalam tiga kategori -- positif, negatif, dan netral -- secara lebih akurat dibandingkan pendekatan konvensional, sekaligus menghasilkan informasi yang dapat dijadikan bahan pertimbangan tambahan bagi investor ritel di tengah tingginya ketidakpastian pasar aset kripto.

2. METODE PENELITIAN

Penelitian ini menempuh serangkaian tahapan yang dimulai dari pengumpulan data, preprocessing, pelabelan dan pembagian data, perancangan arsitektur model hybrid IndoBERT-LSTM, konfigurasi pelatihan, hingga evaluasi performa model menggunakan beberapa metrik klasifikasi standar.

2.1. Pengumpulan Data

Data penelitian dikumpulkan melalui teknik crawling pada platform X menggunakan kombinasi kata kunci aset (\$BTC, Bitcoin, BTCIDR) dan kata kunci bermuatan sentimen (serok, cuan, rugi, jual, beli, naik, turun), dibatasi pada tweet berbahasa Indonesia periode 1 Januari hingga 31 Desember 2024. Tweet yang teridentifikasi sebagai spam, promosi bot, atau berita otomatis tanpa opini pengguna disingkirkan sejak tahap ini. Proses crawling menghasilkan 3.579 tweet mentah, yang setelah dibuang baris kosong dan duplikat persis menyisakan 3.577 baris untuk diproses lebih lanjut.

Table 1 Karakteristik Dataset Penelitian

Karakteristik	Keterangan
Sumber Data	Crawling platform X (Twitter)
Periode	Januari - Desember 2024
Jumlah Data Mentah	3.579 tweet
Jumlah Data Bersih	3.542 tweet
Atribut Utama	Teks tweet (full_text)

2.2. Preprocessing Data

Preprocessing untuk model Transformer seperti IndoBERT berbeda secara mendasar dari metode klasik. Model klasik (SVM, Naive Bayes) sangat bergantung pada stemming dan stopwords removal untuk menyederhanakan fitur, sedangkan IndoBERT adalah model kontekstual yang justru membutuhkan struktur kalimat utuh untuk memahami makna secara optimal. Stemming yang agresif berisiko menghilangkan nuansa gramatikal, dan penghapusan stopwords tanpa pengecualian dapat mengubah makna kalimat secara drastis -- misalnya frasa "tidak untung" akan berbalik makna apabila kata "tidak" ikut dihapus. Berdasarkan pertimbangan ini, preprocessing dalam penelitian ini difokuskan pada pembersihan minimal, normalisasi istilah kripto, dan tokenisasi khusus model, dengan tahapan sebagai berikut:

a. Cleaning: decoding entitas HTML, penghapusan mention (@username), indikator retweet (RT), tautan URL, serta simbol pagar (#) tanpa menghapus kata di belakangnya, dan penyeragaman spasi. Emoji sengaja dipertahankan karena membawa muatan sentimen yang informatif.

b. Case folding: seluruh teks diseragamkan menjadi huruf kecil.

c. Normalisasi slang dan emoji kripto: istilah tidak baku seperti "serok", "cuan", "hodl", "fomo", "akumulasi", "whale" diterjemahkan ke padanan kata baku menggunakan kamus normalisasi yang disusun khusus, begitu pula emoji seperti 🚀 dan 📉 diterjemahkan menjadi kata yang merepresentasikan maknanya.

d. Stopword removal selektif: hanya diterapkan pada versi teks untuk keperluan eksplorasi data (EDA), dengan kata negasi (tidak, bukan, jangan, belum, kurang, tapi, namun) dikecualikan dari daftar pembuangan. Teks yang menjadi masukan tokenizer IndoBERT tetap mempertahankan struktur kalimat utuh tanpa stopwords removal.

e. Penentuan panjang token maksimum: ditentukan secara data-driven dari persentil ke-95 distribusi panjang tweet setelah dibersihkan, sehingga tokenisasi tidak boros komputasi namun tetap tidak memotong informasi penting.



Gambar 1 Cloud Kosakata Dominan pada Tiap Kelas Sentimen

2.3. Pelabelan Data

Karena data mentah belum memiliki label sentimen, penelitian ini menerapkan pelabelan awal secara otomatis (pseudo-labeling) menggunakan model IndoBERT sentimen mdhugol/indonesia-bert-sentiment-classification, dilengkapi ambang batas netral dan bonus skor positif untuk mengoreksi kecenderungan model umum yang kerap salah membaca istilah khas kripto seperti "breakout" atau "akumulasi" sebagai sentimen negatif. Hasil pelabelan otomatis pada 3.542 tweet ini kemudian dikoreksi satu per satu secara manual oleh peneliti untuk memastikan akurasi label sebelum dipakai melatih model akhir.

Table 2 Distribusi Label Sentimen Hasil Koreksi Manual

Kelas Sentimen	Jumlah Tweet	Persentase
Positif	1.347	38,0%



Kelas Sentimen	Jumlah Tweet	Persentase
Netral	1.235	34,9%
Negatif	960	27,1%
Total	3.542	100%

Rasio antara kelas terbesar (positif) dan kelas terkecil (negatif) tercatat 1,40 : 1, tergolong ketidakseimbangan ringan sehingga tidak memerlukan penanganan agresif seperti oversampling maupun undersampling di tahap awal ini.

2.4. Pembagian Data

Pembagian dataset menggunakan pendekatan group-aware splitting: setiap tweet diberi signature dari enam kata pertama teks bersihnya (dengan angka dinormalisasi menjadi token khusus), lalu GroupShuffleSplit memastikan tweet dengan signature identik tidak tersebar ke lebih dari satu subset, sehingga potensi kebocoran data akibat tweet mirip atau redundan dapat dicegah. Dari 3.542 baris data, pembagian dilakukan dengan proporsi 80% data latih, 10% data validasi, dan 10% data uji, menghasilkan 2.856 baris data latih, 339 baris data validasi, dan 347 baris data uji.

2.5. Evaluasi Hasil Clustering

Model dibangun menggunakan kerangka kerja PyTorch dan pustaka Hugging Face Transformers, terdiri atas empat lapisan utama. Pertama, input layer mengonversi teks menjadi input_ids dan attention_mask melalui tokenizer indobenchmark/indobert-base-p1 berbasis algoritma WordPiece, dengan panjang token maksimum (MAX_LEN) 64 token. Kedua, lapisan IndoBERT menghasilkan vektor embedding kontekstual berdimensi 768 untuk setiap token, dan alih-alih hanya mengambil representasi token [CLS], seluruh urutan vektor keluaran (last_hidden_state) diteruskan ke lapisan berikutnya agar informasi kontekstual antarkata tetap terjaga. Ketiga, lapisan LSTM satu arah berdimensi 128 hidden units memproses urutan vektor tersebut, dengan keluarannya digabungkan melalui mean pooling dan max pooling yang memperhitungkan attention mask agar token padding tidak ikut tercampur, menghasilkan satu vektor berdimensi 256 yang melewati dropout 0,4. Keempat, lapisan fully-connected memproyeksikan vektor tersebut menjadi tiga keluaran kelas (negatif, netral, positif) melalui fungsi aktivasi Softmax.

Strategi pelatihan yang dipilih bukan melatih dari awal, melainkan fine-tuning terbatas: seluruh 12 lapisan encoder IndoBERT beserta lapisan embedding dibekukan sepenuhnya sehingga IndoBERT murni berperan sebagai feature extractor, sementara hanya lapisan LSTM dan fully-connected yang dilatih. Konfigurasi ini menyisakan 123.850.752 parameter beku dan hanya 590.592 parameter yang benar-benar dilatih. Keputusan ini diambil setelah percobaan dengan lebih sedikit lapisan dibekukan menunjukkan overfitting yang signifikan, di mana macro-F1 data latih melesat jauh melampaui macro-F1 validasi.

2.6. Konfigurasi dan Skenario Pelatihan

Pelatihan dilakukan hingga maksimum 30 epoch dengan mekanisme early stopping berdasarkan macro-F1 validasi (patience 4 epoch). Fungsi loss yang digunakan adalah Focal Loss (gamma = 2,0) yang dikombinasikan dengan bobot kelas hasil strategi balanced, dipilih karena secara eksplisit menekan kontribusi contoh yang sudah mudah diklasifikasikan dan memperbesar fokus pembelajaran pada contoh yang sulit. Optimizer yang dipakai adalah AdamW dengan learning rate 5e-4 untuk kelompok parameter head (LSTM dan fully-connected), weight decay 0,05, learning rate

scheduler linear dengan warmup 10% dari total langkah pelatihan, serta gradient clipping pada norma maksimum 1,0 untuk menjaga stabilitas pelatihan.

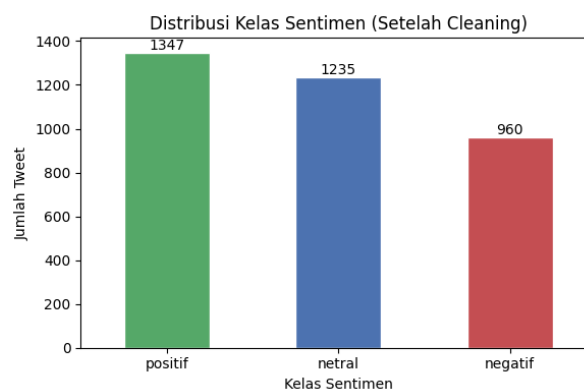
2.7. Evaluasi Model

Kinerja model diukur menggunakan Confusion Matrix beserta empat metrik turunannya, yaitu Accuracy, Precision, Recall, dan F1-Score, dengan pendekatan Macro-Average agar performa pada kelas minoritas tetap diperhitungkan secara setara dan tidak tertutupi oleh dominasi kelas mayoritas. Sebagai pelengkap, dilakukan pula analisis kurva Receiver Operating Characteristic (ROC) dengan pendekatan one-vs-rest untuk mengukur kemampuan model membedakan tiap kelas sentimen pada berbagai ambang batas keputusan.

3. HASIL DAN PEMBAHASAN

3.1. Karakteristik dan Distribusi Data

Dataset akhir yang dipakai dalam penelitian ini berjumlah 3.542 tweet berbahasa Indonesia bertopik Bitcoin, dengan distribusi kelas positif 1.347 tweet (38,0%), netral 1.235 tweet (34,9%), dan negatif 960 tweet (27,1%), sebagaimana telah dirangkum pada Tabel 2. Rasio 1,40 : 1 antara kelas terbesar dan terkecil ini tergolong ketidakseimbangan ringan sehingga cukup ditangani melalui pembobotan pada fungsi loss, tanpa perlu resampling data secara eksplisit.



Gambar 2 Distribusi Kelas Sentimen pada Data Bersih

Visualisasi word cloud pada Gambar 2 memperlihatkan pola kosakata yang cukup berbeda antar kelas. Kelas positif didominasi kata seperti "naik", "kenaikan", dan "aset", mencerminkan narasi optimisme kenaikan harga. Kelas netral banyak diisi kata-kata faktual seperti "market", "kripto", "beli", dan "turun" yang muncul berdampingan tanpa kecenderungan arah sentimen yang kuat, mengindikasikan konten yang sifatnya deskriptif. Sementara itu, kelas negatif didominasi kata "turun", "altcoin", "resistance", dan "panik" yang secara konsisten mencerminkan narasi penurunan harga dan kekhawatiran investor.

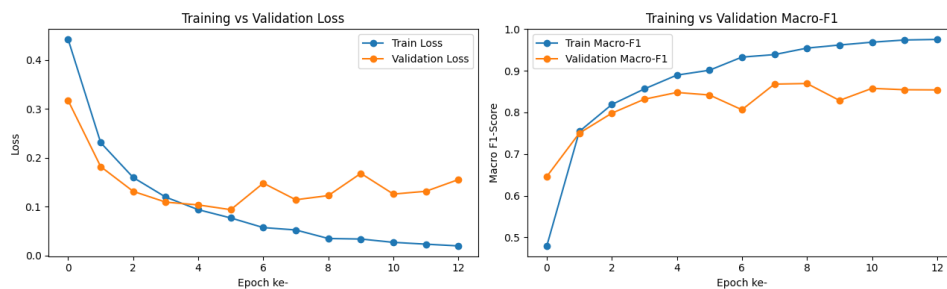
3.2. Penanganan Ketidakseimbangan Kelas

Pada data latih (2.856 baris), distribusi kelas tercatat positif 1.042 tweet, netral 984 tweet, dan negatif 830 tweet, dengan rasio kelas terbesar terhadap terkecil 1,26 : 1. Karena rasio ini masih tergolong ringan, penelitian ini secara sengaja tidak menerapkan oversampling maupun undersampling, mengingat oversampling acak berisiko membuat model menghafal baris yang diduplikasi persis, sementara undersampling akan membuang data asli yang jumlahnya sudah relatif terbatas.

Ketidakseimbangan kelas sepenuhnya ditangani melalui pembobotan pada fungsi loss, dengan bobot kelas negatif 1,1470, netral 0,9675, dan positif 0,9136 -- kelas dengan data lebih sedikit memperoleh bobot lebih besar agar kontribusinya terhadap loss total lebih ditekankan.

3.3. Proses Pelatihan Model

Selama pelatihan, model menunjukkan peningkatan performa yang konsisten pada beberapa epoch awal. Pada epoch 1, macro-F1 validasi tercatat 0,6461 dan terus meningkat hingga mencapai puncaknya pada epoch 9 dengan macro-F1 validasi 0,8694. Proses pelatihan berhenti otomatis (early stopping) pada epoch 13 setelah macro-F1 validasi tidak membaik selama 4 epoch berturut-turut, dan checkpoint model dengan performa validasi terbaik (epoch 9) dimuat kembali untuk tahap evaluasi akhir.

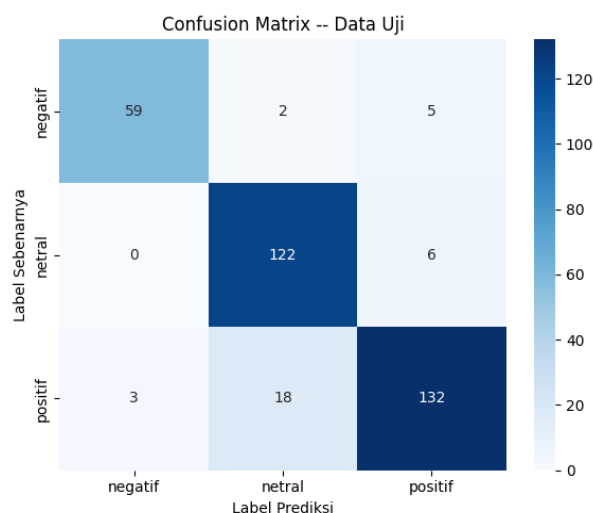


Gambar 3 Distribusi Kelas Sentimen pada Data Bersih

Gambar 3 memperlihatkan bahwa loss data latih menurun tajam dan konsisten dari 0,44 di epoch awal hingga mendekati 0,02 pada epoch terakhir, sedangkan loss data validasi hanya menurun hingga sekitar epoch 4-6 sebelum berfluktuasi dan cenderung naik kembali. Gap yang melebar antara loss latih dan loss validasi ini menjadi indikasi awal terjadinya overfitting seiring bertambahnya epoch, sehingga strategi penyimpanan checkpoint terbaik yang diterapkan pada penelitian ini menjadi krusial agar model yang dipakai untuk evaluasi akhir benar-benar berada pada titik generalisasi terbaiknya, bukan pada epoch terakhir yang berpotensi sudah overfit.

3.4. Evaluasi Kinerja Klasifikasi

Evaluasi akhir dilakukan pada 347 data uji yang sama sekali belum pernah dilihat model selama pelatihan maupun validasi.



Gambar 4 Confusion Matrix pada Data Uji

Berdasarkan Gambar 4, model berhasil memprediksi dengan benar 59 dari 66

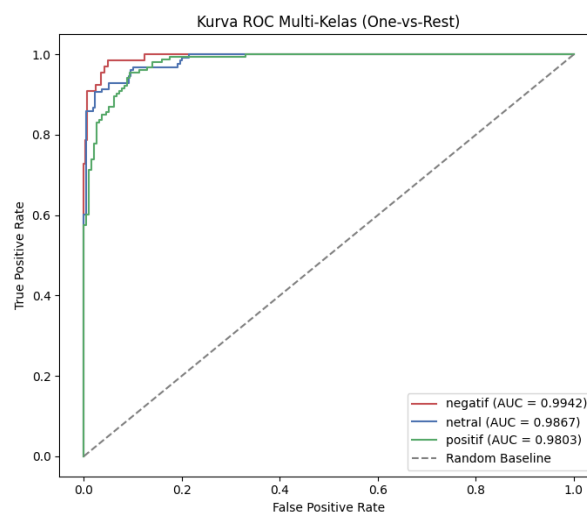
tweet negatif, 122 dari 128 tweet netral, dan 132 dari 153 tweet positif. Kesalahan terbanyak terjadi pada 18 tweet positif yang justru diprediksi sebagai netral, diikuti 6 tweet netral yang diprediksi positif, 5 tweet negatif diprediksi positif, 3 tweet positif diprediksi negatif, dan 2 tweet negatif diprediksi netral. Secara total, 34 dari 347 prediksi keliru, atau setara tingkat kesalahan 9,8%.

Table 3 Classification Report Model pada Data Uji

Kelas	Precision	Recall	F1-Score	Support
Negatif	0,9516	0,8939	0,9219	66
Netral	0,8592	0,9531	0,9037	128
Positif	0,9231	0,8627	0,8919	153
Akurasi			0,9020	347
Macro Avg	0,9113	0,9033	0,9058	347
Weighted Avg	0,9049	0,9020	0,9020	347

Model mencapai akurasi keseluruhan 90,20%, dengan rata-rata makro precision 91,13%, recall 90,33%, dan F1-score 90,58%. Kelas negatif mencatatkan precision tertinggi (95,16%), menunjukkan model jarang keliru menandai tweet lain sebagai negatif meski sebagian kecil tweet negatif justru lolos tidak terdeteksi. Kelas netral mencatatkan recall tertinggi (95,31%) namun precision relatif lebih rendah (85,92%), mengindikasikan model cenderung "menampung" sebagian tweet dari kelas lain -- khususnya positif -- ke dalam kelas netral. Adapun kelas positif mencatatkan recall paling rendah di antara ketiga kelas (86,27%), menegaskan bahwa kelas ini paling sering "diserap" oleh kelas netral, kemungkinan karena batas antara ungkapan optimisme yang implisit dengan pernyataan informasional yang netral memang cukup tipis pada teks berbahasa Indonesia.

3.5. Kurva ROC dan AUC



Gambar 5 Kurva ROC Multi-Kelas (One-vs-Rest)

Gambar 5 menunjukkan nilai Area Under Curve (AUC) yang tinggi pada ketiga kelas: negatif 0,9942, netral 0,9867, dan positif 0,9803. Nilai AUC yang mendekati sempurna ini mengonfirmasi bahwa model memiliki kemampuan diskriminasi antar kelas yang sangat baik, jauh melampaui baseline acak, meskipun pada praktiknya dengan threshold keputusan default 0,5 tetap ditemukan sejumlah kekeliruan

klasifikasi sebagaimana terlihat pada confusion matrix.

3.6. Analisis Kesalahan Klasifikasi

Analisis kesalahan secara kualitatif diarahkan khususnya pada kekeliruan silang antara kelas positif dan negatif, yang meski jumlahnya relatif kecil (8 dari 347 data uji) dinilai paling merugikan secara praktis dibandingkan kekeliruan yang melibatkan kelas netral misalnya tweet bernada optimis yang justru dibaca sebagai pesimis, atau sebaliknya. Pemeriksaan manual menunjukkan sebagian kesalahan terjadi pada kalimat dengan sentimen implisit atau sarkastik, maupun kalimat yang memuat kata kunci dari kedua polaritas sekaligus, misalnya narasi harga turun yang justru dibingkai sebagai peluang untuk membeli. Pengujian tambahan pada 30 kalimat buatan yang mencampurkan istilah baku dan slang kripto juga memperlihatkan pola serupa: kalimat "Investor besar atau whales mulai akumulasi Bitcoin lagi nih", yang secara konotasi sebenarnya cenderung bullish di kalangan komunitas kripto, justru diprediksi negatif dengan keyakinan rendah (56,8%), menandakan model belum sepenuhnya menangkap konotasi positif dari istilah "akumulasi" meski istilah tersebut telah dipetakan secara eksplisit dalam kamus normalisasi.

3.7. Implementasi Prototipe

Model yang telah dilatih dan dievaluasi kemudian diimplementasikan ke dalam dashboard interaktif berbasis Streamlit bernama "Dashboard Analisis Sentimen Bitcoin", yang menyediakan empat menu utama: Ringkasan, Prediksi Tunggal, Prediksi Batch, dan Evaluasi Model. Pada pengujian dengan tweet tunggal, kalimat "BTC kembali naik dan membuat investor semakin optimis" diberi label positif dengan keyakinan 78,7%, sedangkan kalimat berlawanan "Harga BTC kembali turun dan membuat investor panik" diberi label negatif dengan keyakinan 96,7%. Pada pengujian batch terhadap 300 tweet baru, dashboard mencatat distribusi hasil positif 31,3%, netral 44,0%, dan negatif 24,7%, dengan rata-rata confidence keseluruhan 87,3%. Hasil ini menunjukkan bahwa model yang dibangun tidak berhenti sebagai eksperimen di notebook, melainkan dapat diwujudkan menjadi alat bantu yang benar-benar bisa dipakai, baik untuk memeriksa satu tweet maupun sekumpulan data sekaligus.

4. KESIMPULAN

Penelitian ini berhasil mengintegrasikan IndoBERT dan Long Short-Term Memory untuk mengklasifikasikan sentimen pasar Bitcoin pada tweet berbahasa Indonesia. Preprocessing berlapis yang dirancang khusus untuk karakteristik teks media sosial meliputi cleaning, normalisasi slang dan emoji kripto, serta stopword removal selektif yang tetap mempertahankan kata negasi terbukti penting untuk mengurangi noise tanpa merusak struktur kalimat yang dibutuhkan IndoBERT. Ketidakseimbangan kelas yang tergolong ringan (rasio 1,26 : 1 pada data latih) berhasil ditangani cukup efektif melalui kombinasi class weight dan Focal Loss, tanpa perlu oversampling maupun undersampling. Strategi membekukan seluruh 12 lapisan encoder IndoBERT dan hanya melatih lapisan LSTM serta fully-connected di atasnya (590.592 dari sekitar 124,4 juta parameter) terbukti menekan risiko overfitting secara signifikan dibandingkan konfigurasi dengan lebih sedikit lapisan yang dibekukan.

Pada data uji, model hybrid IndoBERT-LSTM mencapai akurasi 90,20%, precision makro 91,13%, recall makro 90,33%, dan F1-score makro 90,58%, dengan nilai AUC di atas 0,98 pada ketiga kelas sentimen. Kesalahan klasifikasi yang tersisa didominasi oleh kekeliruan antara kelas positif dan netral, mengindikasikan bahwa tantangan utama model terletak pada teks dengan ekspresi optimisme implisit yang berbatasan tipis dengan pernyataan informasional netral, ketimbang kekeliruan silang antara sentimen positif dan negatif yang jumlahnya jauh



lebih kecil. Hasil ini menunjukkan bahwa integrasi IndoBERT dan LSTM menghasilkan sistem klasifikasi sentimen pasar Bitcoin yang cukup akurat dan dapat dijadikan salah satu informasi pendukung dalam pengambilan keputusan investasi, meski pengembangan lanjutan seperti memperbanyak data latih pada kelas minoritas, membuka sebagian lapisan encoder secara bertahap, serta menambahkan metode explainability seperti SHAP masih terbuka untuk meningkatkan akurasi dan transparansi model pada penelitian berikutnya.

DAFTAR PUSTAKA

- [1] L. Alexander, R. Gian, A. Ananta, and J. Harefa, "Evaluating IndoBERT for Indonesian Hoax News Detection: A Comparative Study with Ensemble and CNN-LSTM Models," *Procedia Computer Science*, vol. 269, pp. 1625-1633, 2025, doi: 10.1016/j.procs.2025.09.105.
- [2] M. Alfhi, "Online Gambling Promotion Detection in Indonesian YouTube Comments Using Semi-Supervised IndoBERT Classification," *Procedia Computer Science*, vol. 269, pp. 1269-1278, 2025, doi: 10.1016/j.procs.2025.09.068.
- [3] J. Angelo, "Coexistence of Bitcoin, Fiat Money, and Risk-Free Bonds," 2025.
- [4] S. Arsi, "Bitcoin Futures in the Google Search Arena," *Finance Research Letters*, vol. 88, 2026, doi: 10.1016/j.frl.2025.109178.
- [5] J. Beckmann, T. Geldner, and J. Wustefeld, "The Relevance of Media Sentiment for Small and Large Scale Bitcoin Investors," *Journal of International Financial Markets, Institutions & Money*, vol. 92, 2024, doi: 10.1016/j.intfin.2024.101963.
- [6] Z. Chen, T. Hsieh, and Z. Chen, "An Optimization Enabled Hierarchical Attention-Deep LSTM Model for Sentiment Analysis on Cloth Products from Customer Rating," *Data & Knowledge Engineering*, vol. 161, 2026, doi: 10.1016/j.datak.2025.102523.
- [7] W. Christian, D. Adamlu, A. Yu, and D. Suhartono, "Leveraging IndoBERT and DistilBERT for Indonesian Emotion Classification in E-Commerce Reviews," *Procedia Computer Science*, vol. 269, pp. 321-330, 2025, doi: 10.1016/j.procs.2025.08.284.
- [8] A. Dolaeva, U. Beliaeva, D. Grigoriev, and A. Semenov, "Analyzing and Forecasting P/E Ratios Using Investor Sentiment in Panel Data Regression and LSTM Models," *International Review of Economics and Finance*, vol. 98, 2025, doi: 10.1016/j.iref.2025.103840.
- [9] P. Han, H. Chen, A. Rasool, Q. Jiang, and M. Yang, "MFB: A Generalized Multimodal Fusion Approach for Bitcoin Price Prediction Using Time-Lagged Sentiment and Indicator Features," *Expert Systems With Applications*, vol. 261, 2025, doi: 10.1016/j.eswa.2024.125515.
- [10] K. John and J. Li, "Bitcoin Price Volatility: Effects of Retail Traders, Illegal Users, and Sentiment," *Journal of Corporate Finance*, vol. 94, 2025, doi: 10.1016/j.jcorpfin.2025.102837.
- [11] M. Li, V. Manahov, and J. Ashton, "A Note on the Relationship Between Bitcoin Price and Sentiment: New Evidence Obtained from a Cryptocurrency Heist," *North American Journal of Economics and Finance*, vol. 78, 2025, doi: 10.1016/j.najef.2025.102432.
- [12] U. B. Mahadevaswamy and P. Swathi, "Sentiment Analysis Using Bidirectional LSTM Network," *Procedia Computer Science*, vol. 218, pp. 45-56, 2023, doi: 10.1016/j.procs.2022.12.400.
- [13] R. Marti and A. Reimat, "Bitcoin and Global Stock Portfolio: Dynamics Across Short and Long-Term Horizons," *Finance Research Letters*, vol. 86, 2025, doi: 10.1016/j.frl.2025.108918.
- [14] C. S. K. A. Muhammad Dhafa Maulana, "Perbandingan IndoBERT dan Bi-LSTM dalam Mendeteksi Pelanggaran," *Sintech Journal*, vol. 8, no. 1, pp. 52-59, 2025.
- [15] E. Putra, R. Eko, R. Kurniawan, W. Caesarendra, P. Ugiana, and B. Pardamean, "Improved LSTM Hyperparameters Alongside Sentiment Walk-Forward Validation for Time Series Prediction," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 11, no. 1, 2025, doi: 10.1016/j.joitmc.2024.100458.



- [16] G. Singh, A. Saxena, R. Nahta, and Y. Kumar, "Hierarchical Attention for Aspect Extraction Using LSTM in Fine-Grained Sentiment Analysis and Evaluation," *Applied Soft Computing*, vol. 167, 2024, doi: 10.1016/j.asoc.2024.112408.
- [17] E. Supriyadi and P. Nurhuda, "Sentiment Analysis of TikTok User Comments on QRIS Adoption in Indonesia Using IndoBERT," *Procedia Computer Science*, vol. 269, pp. 121-130, 2025, doi: 10.1016/j.procs.2025.08.265.