

# Public Sentiment Analysis of Gold and Bitcoin Investment Using IndoBERT

## [Analisis Sentimen Publik Tentang Investasi Emas dan Bitcoin Menggunakan IndoBERT]

Rachmad Aditya Alfaris<sup>1)</sup>, Mochamad Alfian Rosid<sup>2)</sup>, Ika Ratna Indah Astutik<sup>3)</sup>, Azmuri Wahyu Azinar<sup>4)</sup>

<sup>1,2,3,4)</sup> Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

\*Email Penulis Korespondensi: alfanrosid@umsida.ac.id

**Abstract.** *The development of Gold and Bitcoin investments has driven increased public opinion on social media, especially on the "Ngomongin Uang" financial education channel. The massive volume of comments and informal language makes manual sentiment analysis ineffective. This study aims to map public sentiment towards Gold and Bitcoin investments and evaluate IndoBERT's capability in classifying Indonesian comments. The research data consisted of 4,909 unique comments from YouTube and TikTok grouped into positive, neutral, and negative sentiments. The method utilized was the fine-tuning of the indobenchmark/indobert-base-p2 variant using class weight and topic insertion (prepend) strategies to overcome class imbalance. The results showed that the A4 configuration (5 epochs, class weight, prepend, and AdamW optimizer) yielded the best performance with 72.51% accuracy and a Test Macro F1-Score of 0.6983. This strategy improved the model's ability to recognize negative sentiment with a 0.6087 recall and demonstrated stable performance across Bitcoin and Gold topics.*

**Keywords** - Bitcoin; Class Weight; Fine-Tuning; IndoBERT; Sentiment Analysis

**Abstrak.** *Perkembangan investasi emas dan Bitcoin mendorong meningkatnya opini publik di media sosial, terutama pada kanal edukasi finansial "Ngomongin Uang". Banyaknya komentar serta penggunaan bahasa informal membuat analisis sentimen manual menjadi kurang efektif. Penelitian ini bertujuan memetakan sentimen publik terhadap investasi Emas dan Bitcoin serta mengevaluasi kemampuan IndoBERT dalam mengklasifikasikan komentar berbahasa Indonesia. Data penelitian berjumlah 4.909 komentar dari YouTube dan TikTok dengan label sentimen positif, netral, dan negatif. Metode yang digunakan adalah fine-tuning IndoBERT varian indobenchmark/indobert-base-p2 dengan strategi class weight dan penyisipan topik (prepend) untuk mengatasi ketidakseimbangan kelas. Hasil penelitian menunjukkan konfigurasi A4 (5 epochs, class weight, prepend, AdamW) menghasilkan performa terbaik dengan akurasi 72,51% dan Test Macro F1-Score 0,6983. Strategi tersebut meningkatkan kemampuan model mengenali sentimen negatif dengan recall 0,6087 serta menunjukkan stabilitas performa lintas topik.*

**Kata Kunci** - Analisis Sentimen; Bitcoin; Class Weight; Fine-Tuning; IndoBERT

## I. PENDAHULUAN

Dalam dinamika lanskap keuangan modern, emas konsisten menjadi instrumen yang diminati masyarakat sebagai sarana penyimpanan nilai [1]. Minat ini didorong oleh stabilitas karakteristik emas saat menghadapi ketidakpastian ekonomi, serta kemampuannya mempertahankan nilai jangka panjang dari pengaruh faktor makroekonomi seperti inflasi dan suku bunga [2], sehingga tetap diandalkan sebagai aset pelindung nilai dalam portofolio investasi. Bersamaan dengan kemajuan teknologi digital, Bitcoin muncul sebagai alternatif investasi terdesentralisasi dengan kelangkaan terprogram yang dibatasi maksimal 21 juta koin [3]. Kemiripan karakteristik kelangkaan ini memunculkan persepsi Bitcoin sebagai "emas digital" yang dalam beberapa studi dilaporkan berpotensi meningkatkan imbal hasil jangka panjang portofolio [4]. Namun, regulasi di Indonesia yang mengklasifikasikan Bitcoin murni sebagai komoditas aset kripto dan bukan alat pembayaran sah memicu dinamika persepsi masyarakat terkait tingkat keamanan serta legalitasnya [5]. Perkembangan kedua instrumen tersebut memicu diskursus masif di platform media sosial seperti YouTube dan TikTok, terutama melalui kanal edukasi finansial "Ngomongin Uang".

Tingginya partisipasi publik melalui volume komentar yang besar mencerminkan antusiasme masyarakat, namun sekaligus menghadirkan tantangan teknis dalam analisis manual akibat besarnya volume data serta penggunaan bahasa yang beragam dan informal. Keterbatasan pemrosesan manual ini menyebabkan ekstraksi opini publik berjalan lambat dan rentan terhadap subjektivitas, sehingga menghambat pemetaan sentimen masyarakat secara aktual dan menyeluruh. Sebagai rencana penyelesaian atas permasalahan tersebut, penelitian ini mengajukan pendekatan komputasi otomatis berbasis *deep learning* melalui arsitektur BERT yang terbukti andal membedah nuansa dan struktur bahasa yang kompleks [6], khususnya menggunakan model IndoBERT. Sebagai model representasi bahasa yang dikembangkan spesifik untuk bahasa Indonesia [7], IndoBERT memiliki keunggulan fungsional dalam

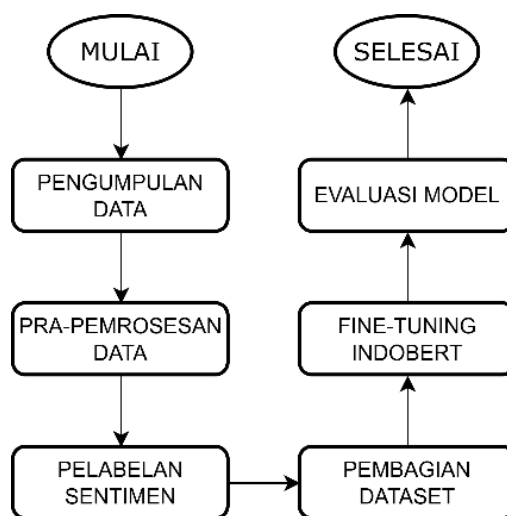
memahami konteks serta relasi antar kata, menjadikannya solusi paling relevan untuk mengatasi kompleksitas dan ambiguitas klasifikasi sentimen pada teks media sosial lokal.

Pendekatan IndoBERT secara arsitektur dikembangkan ke dalam tiga varian utama—yaitu Lite, Base, dan Large—berdasarkan kedalaman lapisannya [8]. Studi terdahulu membuktikan IndoBERT andal dalam menganalisis emosi teks berbahasa Indonesia [9], serta mampu menghasilkan klasifikasi yang konsisten meskipun dengan penyesuaian parameter yang minimal [10]. Secara empiris, IndoBERT berkinerja unggul pada *dataset* komentar karena kemampuannya menangani nuansa bahasa informal di media sosial [11]. Keunggulan pemahaman konteks linguistik lokal ini membuatnya lebih sesuai dan berkinerja lebih baik dibandingkan algoritma konvensional seperti *Naïve Bayes*, *K-NN*, *Decision Tree*, dan *Support Vector Machine* (SVM) [12] [13] [14]. Selain penyesuaian parameter, penggunaan skema partisi *dataset* dengan proporsi 80:10:10 juga telah terbukti dalam penelitian terkait mampu memfasilitasi fase pembelajaran model secara mendalam tanpa mengorbankan objektivitas evaluasi [9]. Implementasi IndoBERT melalui teknik *fine-tuning* dan eksplorasi berbagai *hyperparameter* terbukti krusial untuk mengoptimalkan performa model [15], meskipun efektivitasnya tetap dipengaruhi oleh tingkat ambiguitas data atau ketidakseimbangan distribusi kelas.

Meskipun superioritas IndoBERT telah terbukti pada berbagai literatur terdahulu, implementasinya secara komprehensif pada domain literasi finansial—khususnya polarisasi opini terkait komparasi Emas dan Bitcoin—masih sangat terbatas. Pemanfaatan representasi data opini dari kanal edukasi di YouTube dan TikTok juga belum banyak dieksplorasi. Maka dari itu, penelitian ini dilakukan guna menjawab celah literatur tersebut dengan mengekstraksi dan memetakan persepsi publik di kanal "Ngomongin Uang" secara sistematis. Tujuan utama dari penelitian ini adalah melakukan ekstraksi dan pemetaan sentimen masyarakat terhadap komparasi investasi Emas serta Bitcoin secara objektif. Lebih dari itu, studi ini juga bertujuan menakar ketahanan (*robustness*) model IndoBERT dalam menangani perbincangan finansial yang dinamis, terutama saat dihadapkan pada minimnya data di salah satu kelas sentimen.

## II. METODE

Tahapan penelitian ini diawali dengan pengumpulan komentar media sosial yang dilanjutkan dengan pra-pemrosesan teks agar siap dianalisis. Data tersebut kemudian diberi label sentimen dan dipartisi menjadi data latih serta data uji untuk keperluan pemodelan, sebagaimana dipaparkan pada alur Gambar 1.



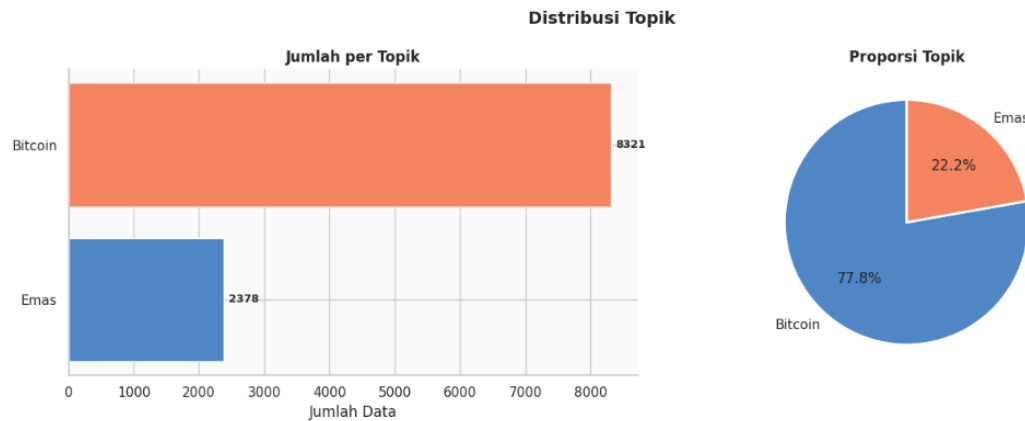
**Gambar 1.** Diagram tahapan penelitian

Tahap selanjutnya melibatkan proses *fine-tuning* IndoBERT untuk menyesuaikan model dengan karakteristik spesifik data investasi, diikuti dengan evaluasi kinerja secara sistematis guna mengukur tingkat akurasi klasifikasi sentimen.

### A. Pengumpulan Data

Data riset ini dikumpulkan dari dua platform media sosial yaitu YouTube dan TikTok, dengan berfokus pada video di kanal edukasi finansial “Ngomongin Uang” yang secara eksplisit membahas investasi Bitcoin dan emas. Ekstraksi data dilakukan secara otomatis menggunakan YouTube Data API v3 dan platform Apify untuk TikTok. Komentar yang dianalisis berasal dari berbagai periode waktu dengan rentang publikasi antara tahun 2020 hingga 2025. Mayoritas data berbahasa Indonesia karena bersumber dari kreator lokal.

Proses pengumpulan ini menghasilkan 11.302 komentar mentah yang kemudian disimpan dalam format CSV dan XLSX. Secara proporsi awal, diskusi terkait Bitcoin mendominasi dengan 8.321 komentar (77,8%) dibandingkan emas yang berjumlah 2.378 komentar (22,2%). Distribusi awal ini memberikan gambaran umum mengenai dominasi topik sebelum melalui tahapan pra-pemrosesan. Pemetaan distribusi awal ditampilkan pada di Gambar 2 serta Tabel 1.



**Gambar 2.** Distribusi label per topik

**Tabel 1.** Contoh data komentar

Topik	Tanggal	Komentar
Bitcoin	2025-02-16	admin udah beli bitcoin ?
Emas	2023-03-09	Semua bakal emas pada waktu nya

## B. Pra-pemrosesan Data

Proses pengolahan teks awal dilakukan melalui beberapa tahapan berurutan. Tahap deteksi dan translasi bahasa bertujuan memastikan keseragaman representasi data dengan mengidentifikasi dan menerjemahkan komentar berbahasa asing ke dalam bahasa Indonesia. Tahap case folding kemudian menyeragamkan seluruh teks menjadi huruf kecil (lowercase) guna menghindari misinterpretasi sistem. Selanjutnya, tahap Cleaning berfungsi membuang noise berupa karakter khusus, emoji, angka, tanda baca, serta tautan (URL).

Mengingat tingginya variasi bahasa di media sosial, tahap normalisasi krusial dilakukan untuk mengonversi istilah tidak standar (slang, singkatan, bahasa informal) menjadi bentuk baku yang seragam. Terakhir, Tokenisasi memecah untaian kalimat menjadi token-token kata individual, yang merupakan prasyarat mutlak bagi algoritma pemrosesan bahasa alami (NLP). Tabel 2 menampilkan lima tahapan dalam pra-pemrosesan data teks yang dilakukan supaya data siap digunakan secara lebih lanjut.

**Tabel 2.** Pengolahan data

Tahapan	Hasil Pemrosesan Teks
Data Awal	Nyesel dulu jual btc saya 14 coin masih harga 43 juta an ...hehehe Rizki flu...
Dateksi dan translasi bahasa	Nyesel dulu jual btc saya 14 koin masih harga 43 juta an ...hehehe Rizki flu...
Case folding	nyesel dulu jual btc saya 14 koin masih harga 43 juta an ...hehehe rizki flu...
Cleaning	nyesel dulu jual btc saya koin masih harga juta an hehehe rizki fluktuatif
Normalisasi	menyesal dulu jual bitcoin saya koin masih harga juta an hehehe rizki fluktu...
Tokenisasi	['menyesal', 'dulu', 'jual', 'bitcoin', 'saya', 'koin', 'masih', 'harga', 'j...

## C. Pelabelan Sentimen dan Pembagian Data

Pelabelan sentimen dilakukan secara manual dengan mengevaluasi konteks setiap komentar guna menghasilkan dataset pelatihan yang akurat dan konsisten. Data diklasifikasikan ke dalam tiga kategori: negatif (0) untuk kritik atau

keraguan, netral (1) untuk informasi objektif, positif (2) untuk dukungan. Contoh hasil pelabelan ini disajikan pada Tabel 3.

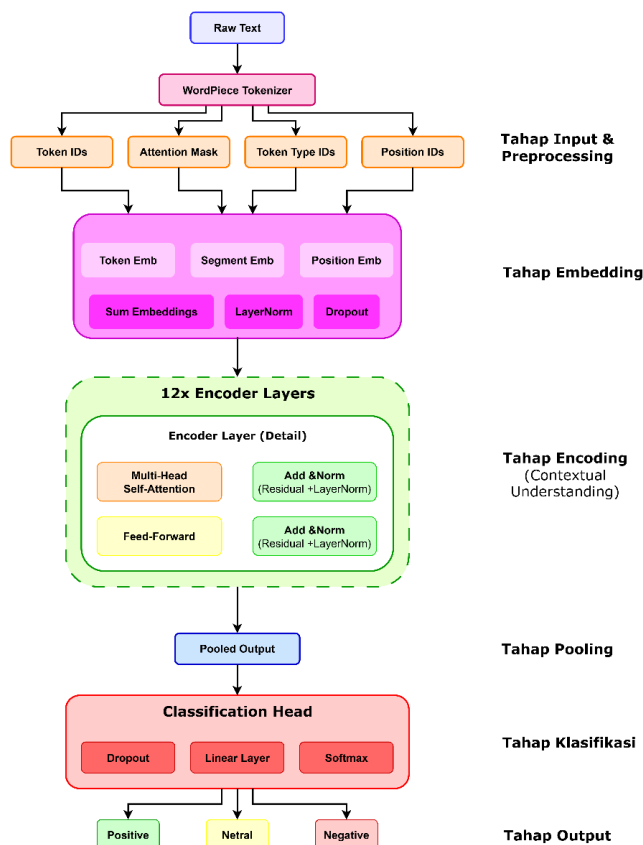
**Tabel 3.** Data dengan label

Komentar	Topik	Label
kedepannya emas akan semakin mahal karena banyaknya uang kertas yang beredar dan jumlah emas yang terbatas	Emas	2
mau bertanya adakah kemungkinan emas di masa depan akan tidak ada nilainya	Emas	1
emas pada akhirnya akan tergantikan dengan cryptocurrency	Emas	0
wakkk gila sih kalau benaran bitcoin segitu konsisten cicil bitcoin tiap bulan	Bitcoin	2
febuari ini bearish kelihatan dump kenapa min	Bitcoin	1
bitcoin akan nyungsep karena ada teknologi yang menyaingi blockchain	Bitcoin	0

Proses pelabelan ini dilakukan setelah tahapan pra-pemrosesan terhadap data mentah yang berjumlah 11.302 komentar. Setelah melalui proses *preprocessing* dan validasi tambahan (seperti penghapusan duplikat), diperoleh 5.068 komentar bersih. Jumlah *dataset* akhir yang digunakan dalam proses pelatihan menjadi 4.909 komentar. *Dataset* ini kemudian dipecah menggunakan rasio 80:10:10 (Data Latih : Validasi : Uji). Data latih difungsikan sebagai basis pembelajaran, data validasi untuk mencegah *overfitting*, sedangkan data uji digunakan untuk evaluasi akhir. Partisi ini menggunakan teknik *stratified random sampling* ganda (berdasarkan kombinasi label sentimen dan topik) guna menjaga konsistensi distribusi data.

#### D. Fine-tuning IndoBERT dan Evaluasi Model

Penelitian ini mengimplementasikan varian indobenchmark/indobert-base-p2 yang alur pemrosesannya diawali oleh WordPiece Tokenizer, dilanjutkan ke 12 lapisan Encoder (Multi-Head Self-Attention), hingga bermuara di Classification Head untuk memprediksi tiga polaritas sentimen. Representasi visual mengenai alur arsitektur model ini ditampilkan pada Gambar 3 sebagai berikut.



Gambar 3. Arsitektur model

Penulis penanggung jawab atau penulis korespondensi atau *corresponding author* harus ditandai dengan tanda *asterisk* diikuti tanda koma “\*”). Di bagian kiri bawah halaman pertama harus dituliskan tanda Penulis Korespondensi atau *Corresponding Author* dan dituliskan pula alamat emailnya (lihat contoh). Komunikasi tentang revisi artikel dan keputusan akhir hanya akan disampaikan melalui email penulis korespondensi.

Proses *fine-tuning* dieksplorasi melalui 10 eksperimen dalam dua skenario utama (Skenario A dan Skenario B) untuk menguji efektivitas *prepend* topik, *class weight*, fungsi optimasi, *dropout*, serta stabilitas generalisasi model. Tahap evaluasi selanjutnya diekstraksi melalui *confusion matrix* dan empat metrik klasifikasi utama. Penjabaran masing-masing metrik:

Akurasi (*Accuracy*): Mengukur tingkat ketepatan prediksi model secara keseluruhan terhadap seluruh data.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

(Persamaan 1)

Presisi (*Precision*): Menunjukkan ketepatan model dalam memprediksi suatu kelas tertentu, terutama dalam meminimalkan kesalahan positif palsu (false positive).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

(Persamaan 2)

*Recall*: Mengukur kemampuan model dalam mengidentifikasi seluruh data yang relevan, sehingga meminimalkan kesalahan negatif palsu (*false negative*).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

(Persamaan 3)

*F1-Score*: Nilai ini merupakan bentuk rata-rata harmonik dari metrik presisi dan recall, yang berfungsi untuk menghasilkan tolok ukur penilaian model yang lebih berimbang.

$$F1\text{-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

(Persamaan 4)

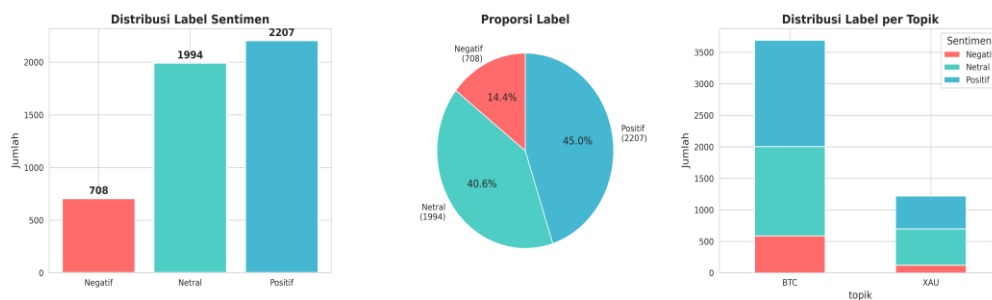
Evaluasi performa akhir pada data uji dianalisis menggunakan *Confusion Matrix* beserta perhitungan Akurasi (Persamaan 1), Presisi (Persamaan 2), Recall (Persamaan 3), dan F1-Score (Persamaan 4). Dalam studi yang memiliki sebaran kelas tidak seimbang, *Macro F1-Score* dinilai sebagai metrik yang paling adil karena memberikan porsi evaluasi yang setara tanpa memihak kelas yang jumlah datanya lebih besar. Di samping itu, pengujian juga dipecah berdasarkan topik (Bitcoin dan Emas) demi memverifikasi konsistensi logika prediksi model.

### III. HASIL DAN PEMBAHASAN

#### A. Karakteristik Dataset

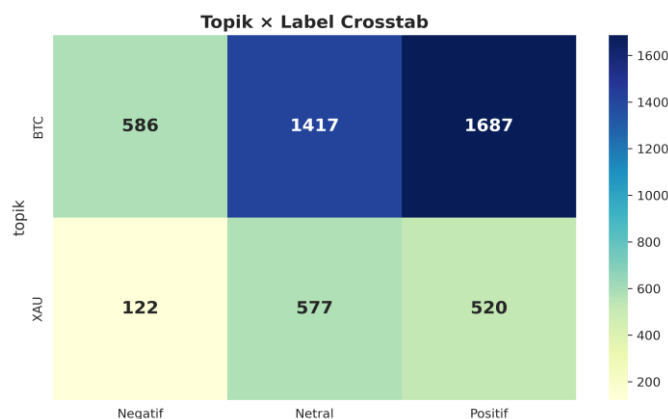
Analisis karakteristik dataset dilakukan untuk mengidentifikasi distribusi data sebelum pelatihan model. Dataset akhir yang digunakan terdiri dari 4.909 komentar yang telah melalui seluruh tahapan pra-pemrosesan, pelabelan, serta pembersihan data lanjutan. Dataset ini kemudian dianalisis untuk mengidentifikasi karakteristik distribusi topik dan sentimen sebelum digunakan dalam proses pelatihan model.

Karakteristik pertama yang dianalisis adalah distribusi topik dan label sentimen. Data menunjukkan dominasi topik Bitcoin (BTC) sebanyak 3.690 komentar (75,2%) dibandingkan topik Emas (XAU) sebanyak 1.219 komentar (24,8%). Dari segi polaritas sentimen, kelas Positif mendominasi dengan 2.207 komentar (45,0%), diikuti kelas Netral sebanyak 1.994 komentar (40,6%), dan kelas Negatif sebanyak 708 komentar (14,4%). Visualisasi distribusi ini disajikan pada Gambar 4.



Gambar 4. Distribusi label sentimen dan proporsi data

Untuk melihat korelasi antara instrumen investasi dengan polaritas sentimen, rincian tabulasi silang disajikan melalui *heatmap* pada Gambar 5. Data menunjukkan bahwa pada kedua topik, sentimen Positif konsisten mendominasi, sementara sentimen Negatif menjadi kelas minoritas dengan jumlah paling sedikit, terutama pada topik Emas yang hanya memiliki 122 komentar negatif. Rangkuman kuantitatif distribusi tersebut disajikan pada Tabel 4.



Gambar 5. Heatmap tabulasi silang topik dan label

**Tabel 4.** Distribusi label sentiment per topik

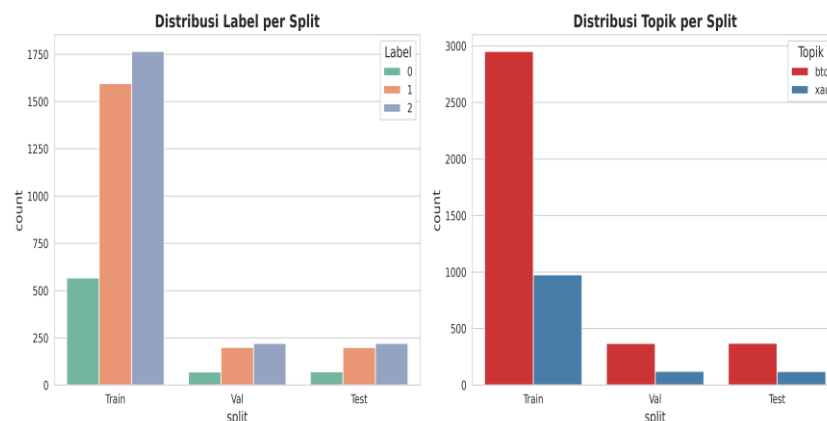
Topik	Negatif	Netral	Positif	Total
BTC	586	1.417	1.687	3.690
XAU	122	577	520	1.219
Total	708	1.994	2.207	4.909

Ketimpangan jumlah data antar kelas (*class imbalance*) diatasi dengan menerapkan teknik penyesuaian bobot kelas (*class weight*). Bobot yang dihitung secara proporsional menghasilkan nilai: Negatif = 2,3086; Netral = 0,8207; dan Positif = 0,7416. Dengan bobot kelas negatif yang dirancang tiga kali lebih besar, model komputasi akan "dipaksa" untuk lebih peka saat memproses fitur dari kelas minoritas tersebut. Gambaran terkait bobot kelas ada pada Gambar 6.

**Gambar 6.** Diagram batang bobot kelas

Analisis panjang token menunjukkan bahwa rata-rata panjang teks berada di bawah 128 token, baik pada skenario *plain text* maupun *prepend* topik. Distribusi ini menjustifikasi penggunaan parameter `max_length = 128` sebagai *default* tokenisasi, karena mayoritas teks sudah terakomodasi tanpa pemotongan signifikan.

Tahap akhir persiapan data melibatkan pembagian *dataset* menggunakan rasio 80:10:10 yang menghasilkan 3.927 data latih, 491 data validasi, dan 491 data uji. Proses stratifikasi ganda berdasarkan label sentimen dan topik diterapkan untuk memastikan proporsi distribusi yang representatif pada setiap *subset*. Visualisasi distribusi *split* disajikan pada Gambar 7. *Dataset* ini merepresentasikan data yang telah bebas dari duplikasi dan siap digunakan sebagai dasar evaluasi performa model secara objektif.

**Gambar 7.** Distrbusi label dan topik per split

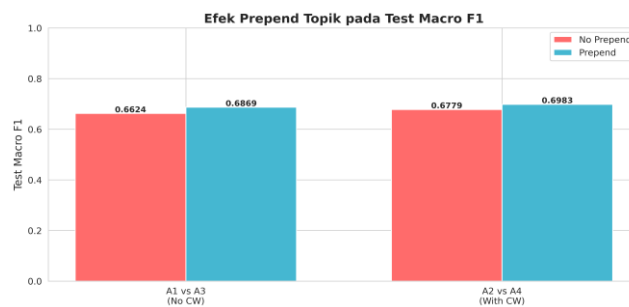
## B. Hasil Eksperimen

Tahapan fine-tuning model IndoBERT diimplementasikan melalui sepuluh konfigurasi eksperimen yang terbagi ke dalam Skenario A (5 epochs) dan Skenario B (20–50 epochs). Rangkuman performa seluruh konfigurasi disajikan pada Tabel 5.

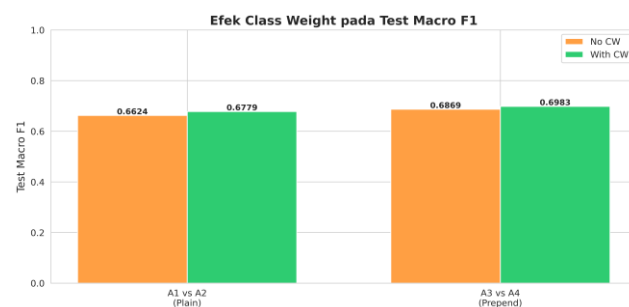
Kode	Skenario	Prepend	CW	Epoch	Test Accuracy	Test Macro F1	Waktu (menit)
A1	A	No	No	5	0,71	0,68	2,6
A2	A	No	Yes	5	0,70	0,67	2,4
A3	A	Yes	No	5	0,71	0,68	2,4
A4	A	Yes	Yes	5	0,72	0,69	2,4
A5	A	Yes	Yes	5	0,72	0,69	3,0
B1	B	Yes	Yes	20	0,70	0,68	9,5
B2	B	Yes	Yes	50	0,73	0,69	23,8
B3	B	Yes	No	20	0,70	0,67	9,6
B4	B	No	Yes	20	0,69	0,67	16,7
B5	B	Yes	Yes	50	0,70	0,67	42,5

Berdasarkan Tabel 5, terdapat beberapa temuan komparatif penting. Pertama, eksperimen A4 yang mengombinasikan *prepend* topik, *class weight*, dan *optimizer* AdamW berhasil menjadi model terbaik berdasarkan Validation Macro F1 tertinggi sebesar 0,7142, dengan Test Macro F1 sebesar 0,6983 dan akurasi 72,51%.

Kedua, penambahan label topik pada teks masukan (*prepend*) secara konsisten meningkatkan Test Macro F1, baik tanpa *class weight* (A1: 0,6843 → A3: 0,6869; peningkatan 0,38%) maupun dengan *class weight* (A2: 0,6779 → A4: 0,6983; peningkatan 3,01%). Efek sinergi terbesar terjadi ketika *prepend* topik dikombinasikan dengan *class weight*, yang juga terbukti meningkatkan F1-Score kelas Negatif dari 0,5694 (A1) menjadi 0,6087 (A4). Visualisasi efek *prepend* topik di Gambar 8 dan *class weight* pada Gambar 9.

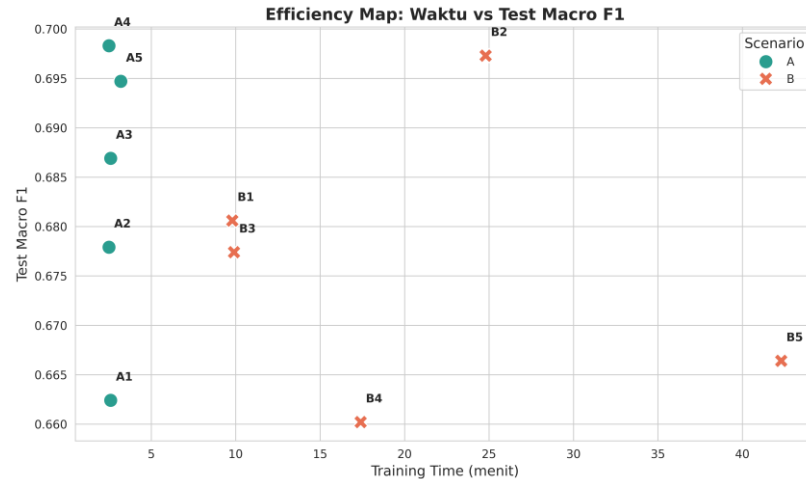


Gambar 8. Diagram batang prepend topik terhadap test macro f1



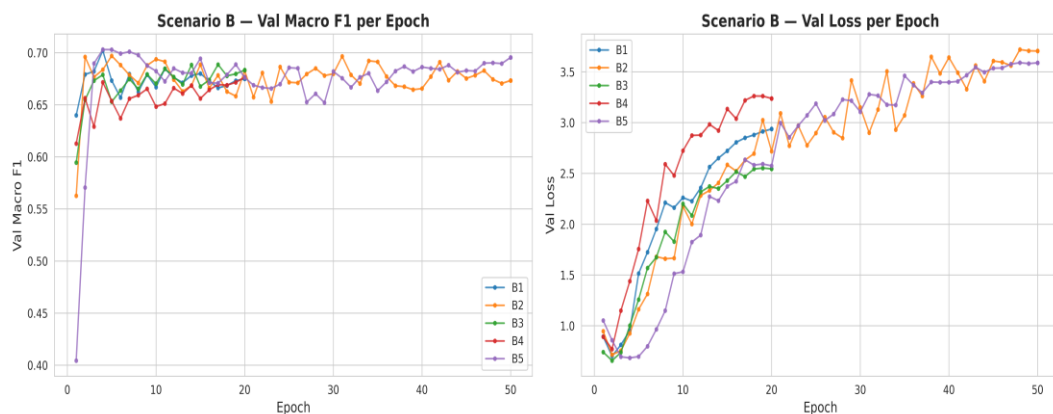
Gambar 9. Prepend topik dan class weight terhadap test macro f1

Ketiga, penambahan durasi pelatihan pada Skenario B (hingga 50 *epochs*) tidak menghasilkan peningkatan performa yang sebanding. Sebagai contoh, eksperimen B5 yang membutuhkan waktu 42,5 menit ( $17\times$  lebih lama dari A4) justru menghasilkan Macro F1 yang lebih rendah (0,6791 vs 0,6983). Hal ini mengindikasikan bahwa model IndoBERT cenderung mencapai titik jenuh lebih cepat pada *dataset* ini. Visualisasi perbandingan performa antara kedua skenario disajikan pada Gambar 10.



Gambar 10. Perbandingan test macro f1 skenario A dan skenario B

Pengujian lebih lanjut terhadap Skenario B membuktikan terjadinya fenomena *overfitting*. Nilai *training loss* terus menurun mendekati nol seiring bertambahnya *epoch*, namun *validation loss* melonjak tajam dari rata-rata 0,7 pada *epoch* awal menjadi 2,5–3,7 pada *epoch* akhir. Sebagai contoh, eksperimen B2 (50 *epochs*) mencatatkan *validation loss* akhir sebesar 3,7051, meskipun Test Macro F1-nya (0,6973) tidak berbeda signifikan dari A4 (0,6983) yang hanya dilatih selama 5 *epochs* dengan *validation loss* 1,2663. Bukti lonjakan *loss* ini divisualisasikan pada Gambar 11.



Gambar 11. Kurva validation macro f1 dan validation loss skenario B per epoch

Analisis efisiensi komputasi menunjukkan bahwa eksperimen A4 menempati posisi paling optimal: memberikan F1-Score tertinggi dengan durasi pelatihan hanya 2,4 menit. Temuan ini menegaskan bahwa konfigurasi parameter yang tepat pada siklus pelatihan pendek jauh lebih efektif dan efisien dibandingkan menambah jumlah *epoch* secara berlebihan.

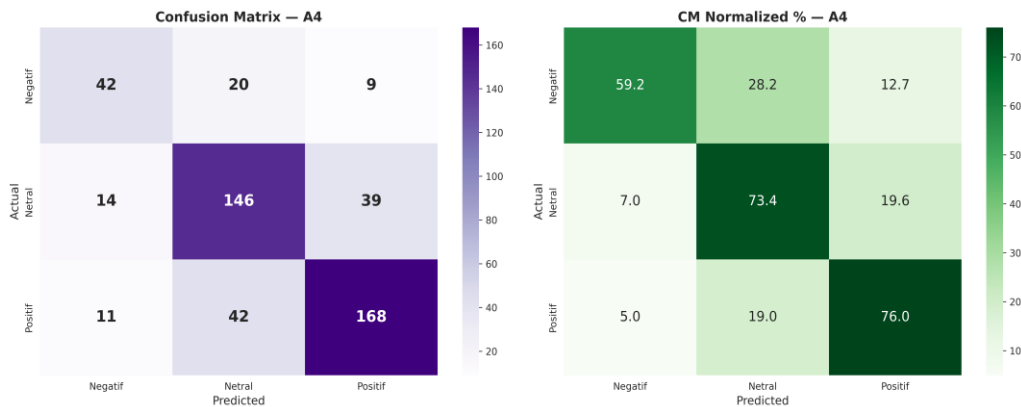
### C. Evaluasi Model Terbaik

Berdasarkan evaluasi komparatif pada tahap fine-tuning dan tinjauan efisiensi komputasi, konfigurasi eksperimen A4 secara definitif ditetapkan sebagai model terbaik. Saat dieksekusi pada data uji baru, model ini menghasilkan Akurasi sebesar 0,7251 dan Test Macro F1-Score 0,6983. Rincian performa klasifikasi per kelas disajikan pada Tabel 6.

**Tabel 6.** Laporan klasifikasi model A4 pada data uji

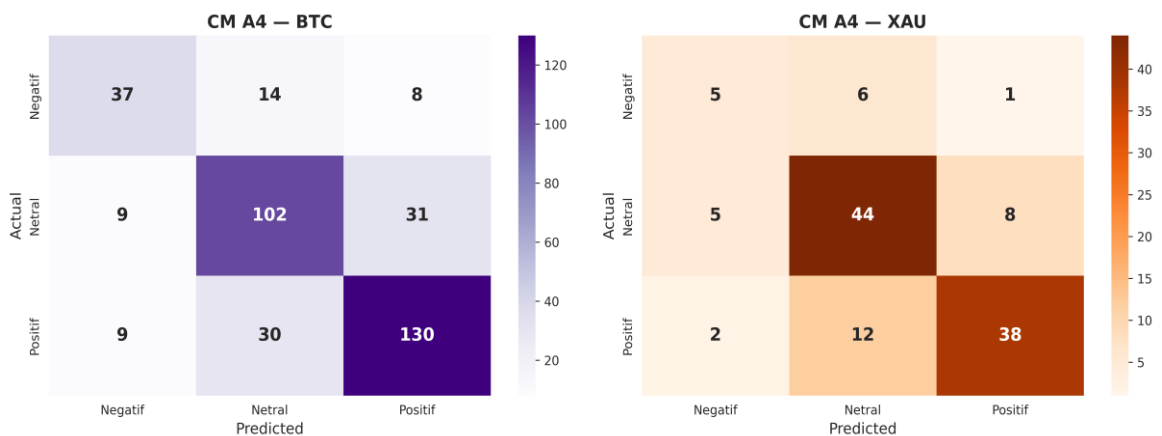
Kelas	Precision	Recall	F1-Score	Support
Negative (0)	0,62	0,59	0,60	71
Neutral (1)	0,70	0,73	0,71	199
Positive (2)	0,77	0,76	0,76	221
Akurasi			0,72	
Macro avg	0,70	0,69	0,69	491

Merujuk pada Tabel 6, model menunjukkan kemampuan klasifikasi yang andal pada kelas Positif dan Netral, dengan F1-Score masing-masing sebesar 0,7689 dan 0,7174. Pencapaian paling esensial terletak pada performa kelas Negatif: meskipun proporsi datanya sangat minim di dalam *dataset* (14,4%), model tetap mampu mengidentifikasinya dengan F1-Score sebesar 0,6087. Hal ini membuktikan secara empiris bahwa penetapan *class weight* berhasil mencegah model dari bias prediksi ke arah kelas mayoritas. Pemetaan distribusi prediksi digambarkan oleh matrik kebingungan di Gambar 12.



**Gambar 12.** Matrik kebingungan model A4 (absolut dan ternormalisasi)

Lebih lanjut, guna memvalidasi stabilitas model dan memastikan tidak adanya bias terhadap salah satu instrumen investasi, evaluasi performa dipartisi berdasarkan topik pembahasan. *Confusion matrix* per topik untuk BTC dan XAU disajikan pada Gambar 13, dengan rangkuman kuantitatif pada Tabel 7.



**Gambar 13.** Confusion matrix topik BTC dan XAU

Tabel 7. Perbandingan performa model A4 per topik

Topik	Sampel	Akurasi	Macro F1	Weighted F1
BTC	370	0,72	0,70	0,72
XAU	121	0,71	0,64	0,71
Selisih		0,0080	0,0676	0,0071

Secara komparatif, model A4 mendemonstrasikan performa yang stabil pada kedua instrumen. Metrik Akurasi antara topik BTC (0,7270) dan XAU (0,7190) hanya berselisih 0,80%. Pada Macro F1, terdapat selisih sebesar 6,76% yang utamanya disebabkan oleh keterbatasan jumlah data negatif pada topik Emas—hanya 12 sampel di data uji—bukan karena kelemahan arsitektur model. Secara Weighted F1, selisih hanya 0,71% yang mengonfirmasi bahwa model memiliki kemampuan generalisasi yang adil antar topik dan secara praktis andal untuk tugas analisis sentimen literasi finansial.

## VI. SIMPULAN

Penelitian ini berhasil membuktikan keandalan implementasi model IndoBERT versi indobenchmark/indobert-base-p2 untuk mendeteksi arah sentimen publik pada korpus teks keuangan kontemporer. Dari keseluruhan uji coba parameter, skema A4 yang memadukan teknik *prepend* topik, penyesuaian bobot kelas, dan mekanisme optimasi AdamW selama 5 *epochs* terbukti menyajikan kinerja yang paling efisien, menghasilkan Akurasi sebesar 72,51% serta nilai *Test Macro F1* sebesar 0,6983. Hasil ini memvalidasi bahwa integrasi penanda topik dan *class weight* mampu menajamkan kepekaan model dalam menangkap sinyal emosi negatif dari kelas minoritas. Sebaliknya, memaksakan proses pelatihan berjalan terlalu lama hingga 50 *epochs* terindikasi kuat memicu kendala *overfitting* yang ditandai dengan lonjakan tajam pada kurva *validation loss* tanpa adanya peningkatan performa yang berarti.

## REFERENSI

- [1] I. R. Johan and S. A. Azariani, "Persepsi Risiko, Financial Self-Efficacy dan Minat Investasi Emas pada Generasi Z," *J. Ilmu Kel. dan Konsum.*, vol. 18, no. 1, pp. 26–37, 2025, doi: 10.24156/jikk.2025.18.1.26.
- [2] D. N. Ahsanah, "Emas sebagai Instrumen Investasi Jangka Panjang," *Shar-E J. Kaji. Ekon. Huk. Syariah*, vol. 8, no. 1, pp. 177–187, 2022.
- [3] A. Wicaksono, C. E. Violita, and E. R. Kamila, "Bitcoin sebagai Instrumen Investasi yang Menguntungkan," *Greenomika*, vol. 4, no. 1, p. 44, 2022.
- [4] M. Hougan and D. Lawant, *Cryptoassets: The Guide to Bitcoin, Blockchain, and Cryptocurrency for Investment Professionals*. CFA Institute Research Foundation, 2021. [Online]. Available: <https://www.cfainstitute.org/research/foundation>
- [5] D. Anisa, T. Anggraini, and K. Tambunan, "Analisis Cryptocurrency Sebagai Alat Alternatif Berinvestasi di Indonesia," *Own. Ris. dan J. Akunt.*, vol. 7, no. 3, 2023, doi: 10.33395/owner.v7i3.1698.
- [6] M. A. Rosid, D. O. Siahaan, and A. Saikhu, "Sarcasm Detection in Indonesian-English Code-Mixed Text Using Multihead Attention-Based Convolutional and BiDirectional GRU," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3436107.
- [7] R. I. Perwira, V. A. Permadi, D. I. Purnamasari, and R. P. Agusdin, "DomainSpecific Fine-Tuning of IndoBERT for Aspect-Based Sentiment Analysis in Indonesian Travel User-Generated Content," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 11, no. 1, pp. 30–40, 2025, doi: 10.20473/jisebi.11.1.30-40.
- [8] A. N. Alfarobby and H. Irawan, "Analisis Sentimen Kepuasan Konsumen Pengguna Transportasi Online pada Ulasan Google Playstore Menggunakan Indobert dan Topic Modeling (Studi Kasus: Gojek dan Grab)," *eProceeding Manag.*, vol. 11, no. 1, p. 72, 2024.
- [9] S. Riyadi, L. K. Salsabila, C. Damarjati, and R. Abdul Karim, "Sentiment Analysis of YouTube Users on Blackpink Kpop Group Using IndoBERT," *INTENSIF J. Ilm. Penelit. dan Penerapan Teknol. Sist. Inf.*, vol. 8, no. 2, p. 233, 2024, doi: 10.29407/intensif.v8n2.22678.
- [10] S. Aras, M. Yusuf, R. Ruimassa, E. A. B. Wambrauw, and E. B. Palalangan, "Sentiment Analysis on Shopee Product Reviews Using IndoBERT," *J. Inf. Syst. Informatics*, vol. 6, no. 3, p. 1616, 2024, doi: 10.51519/journalisi.v6i3.814.

- [11] W. M. Baihaqi and A. Munandar, "Sentiment Analysis of Student Comment on the College Performance Evaluation Questionnaire Using Naïve Bayes and IndoBERT," *JUITA J. Inform.*, vol. 11, no. 2, pp. 213–220, 2023.
- [12] Nurhasiyah, R. Dwiyanaputra, S. I. Murpratiwi, and A. Aranta, "Analisis Sentimen Pengguna Platform Media Sosial X pada Topik Pemilihan Presiden 2024 Menggunakan Perbandingan Model Monolingual dan Multilingual BERT," *JATI J. Mhs. Tek. Inform.*, vol. 9, no. 1, p. 626, 2025.
- [13] A. Jazuli, Widowati, and R. Kusumaningrum, "Aspect-Based Sentiment Analysis on Student Reviews Using the Indo-Bert Base Model," *E3S Web Conf.*, vol. 448, 2023, doi: 10.1051/e3sconf/202344802004.
- [14] C. Apriyadi and Styawati, "Sentiment Analysis of Cyber Attacks in Bank Syariah Indonesia Using SVM and Indobert Method," *J. Tek. Inform.*, vol. 6, no. 2, pp. 819–838, 2025, doi: 10.52436/1.jutif.2025.6.2.2636.
- [15] O. A. Irmawan, I. Budi, A. B. Santoso, and P. K. Putra, "Improving Sentiment Analysis and Topic Extraction in Indonesian Travel App Reviews Through BERT Fine-Tuning," *J. Nas. Pendidik. Tek. Inform.*, vol. 13, no. 2, p. 359, 2024.

**Conflict of Interest Statement:**

*The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.*