

Analisis Sentimen Komentar Youtube terhadap Kondisi Bursa Saham Indonesia akibat Isu Pengunduran Serempak Dewan BEI Menggunakan IndoBERT

Felix Gunawan ¹, Daffa Yudha Musyaffa and Muhammad Rizky Pribadi ^{3,*}

^{1,2,3}Program Studi Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang;
felixgunawan_2327250047@mhs.mdp.ac.id;daffayudhamusyaffa_2327250064@mhs.mdp.ac.id;
rizky@mdp.ac.id

* Korespondensi : felixgunawan_2327250047@mhs.mdp.ac.id

Info Artikel:

Dikirim: 05 Mei 2026

Direvisi: 25 Mei 2026

Diterima: 06 Juni 2026

Abstract: Social media platforms such as YouTube have long served as a primary forum for discussion among retail investor communities in Indonesia. This study aims to analyze public sentiment in order to understand perception trends and the digital psychology of capital market participants regarding the issue of the simultaneous resignation of the board members of the Indonesia Stock Exchange (IDX). The study employs the IndoBERT (Bidirectional Encoder Representations from Transformers for the Indonesian language) deep learning architecture through a fine-tuning process on a dataset of YouTube comments. The textual corpus was preprocessed by removing noise, normalizing stock market slang vocabulary, tokenizing text, and automatically classifying comments into three sentiment polarities: positive, neutral, and negative. The analysis was further complemented by dominant keyword extraction using Word Cloud visualization and word frequency trend mapping to identify psychological factors influencing market opinions. The proposed model successfully classified the semantic complexity of informal language in an objective manner. The visualization results revealed that communication dynamics were predominantly dominated by negative sentiment (57.5%), indicating widespread concern and a decline in public confidence regarding capital market stability following the structural crisis. This study demonstrates the effectiveness of a locally developed transformer-based model as a digital market psychology extraction tool, supporting the automation of real-time investment decision-making processes.

Keywords: Stock Exchange; Data Crawling; IndoBERT; YouTube Comments; Natural Language Processing.

Intisari: Platform media sosial seperti YouTube sudah lama menjadi wadah utama diskusi bagi komunitas investor ritel di Indonesia. Penelitian ini bertujuan menganalisis sentimen publik untuk memahami tren persepsi dan psikologi digital para pelaku pasar modal terhadap isu pengunduran diri serempak dewan Bursa Efek Indonesia (BEI). Penelitian ini menerapkan arsitektur *Deep Learning* IndoBERT (*Bidirectional Encoder Representations from Transformers* untuk bahasa Indonesia) melalui proses *fine-tuning* pada dataset komentar YouTube. Korpus data tekstual dibersihkan dari derau (*noise*), dinormalisasi dari kosakata *slang* pasar modal, ditokenisasi, dan diklasifikasikan secara otomatis ke dalam tiga polaritas sentimen: positif, netral, dan negatif. Tahap analisis dilanjutkan dengan ekstraksi kata kunci dominan berbasis *Word Cloud* serta pemetaan tren frekuensi kata untuk mengidentifikasi variabel psikologis yang menggerakkan opini pasar. Model berhasil mengklasifikasikan kompleksitas semantik bahasa informal secara objektif. Hasil visualisasi menunjukkan bahwa dinamika komunikasi didominasi secara mutlak oleh sentimen

negatif (57,5%), mencerminkan adanya keresahan massal dan penurunan tingkat kepercayaan terhadap stabilitas pasar modal akibat krisis struktural tersebut. Penelitian ini membuktikan efektivitas model *transformer* lokal sebagai instrumen ekstraksi psikologi pasar digital guna mendukung otomatisasi pengambilan keputusan investasi secara *real-time*.

Kata Kunci: Bursa Saham; Crawling Data; IndoBERT; Komentar YouTube; Pengolahan Bahasa Alami.

1. Pendahuluan

Perkembangan pasar modal Indonesia dalam beberapa tahun terakhir menunjukkan lonjakan partisipasi investor ritel yang sangat signifikan [1]. Dinamika pergerakan Indeks Harga Saham Gabungan (IHSG) saat ini tidak lagi eksklusif dipengaruhi oleh faktor makroekonomi fundamental, tetapi juga sangat rentan terhadap sentimen publik yang menyebar luas melalui media sosial [2]. Salah satu platform rujukan utama bagi investor ritel dalam mencari informasi, analisis, dan rekomendasi saham adalah YouTube [3]. Kolom komentar pada platform ini telah bertransformasi menjadi ruang diskusi publik yang memuat kekayaan opini, spekulasi, dan reaksi emosional terhadap kondisi pasar [4]. Opini publik yang tidak terstruktur ini sering kali mencerminkan psikologi pasar yang berpotensi memicu volatilitas harga saham [5]. Oleh karena itu, ekstraksi sentimen secara otomatis menjadi krusial untuk memetakan tren persepsi pasar. Analisis sentimen, sebagai salah satu cabang krusial dalam *Natural Language Processing* (NLP), memungkinkan pengklasifikasian teks ke dalam polaritas tertentu untuk mengukur secara kuantitatif kecenderungan emosional dan ekspektasi pelaku pasar [1], [6].

Meskipun demikian, pemrosesan teks berbahasa Indonesia pada platform media sosial memiliki tantangan komputasional tersendiri [4]. Komentar pengguna YouTube umumnya bersifat informal, dipenuhi dengan bahasa gaul (*slang*), akronim spesifik pasar modal, serta pencampuran kode (*code-mixing*) yang sulit dipahami secara akurat oleh algoritma konvensional. Penelitian terdahulu menunjukkan bahwa model berbasis arsitektur *Recurrent Neural Network* [7](RNN) maupun *Long Short-Term Memory* (LSTM) sering kali mengalami keterbatasan dalam menangkap semantik kontekstual dari struktur kalimat yang kompleks dan ambigu pada ulasan publik [2], [8]. Untuk mengatasi degradasi performa tersebut, pemanfaatan *pre-trained language model* berbasis arsitektur *Transformer*, khususnya IndoBERT, menjadi solusi yang sangat relevan [9]. IndoBERT dilatih secara spesifik menggunakan korpus teks bahasa Indonesia berskala masif, sehingga menguasai pemahaman sintaksis, kosakata, dan konteks lokal yang berkinerja jauh lebih unggul dibandingkan arsitektur pendahulunya [5], [10]. Secara empiris, implementasi IndoBERT terbukti konsisten mengungguli model tradisional dalam menangkap sinyal sentimen yang halus, bahkan pada kondisi kelas data yang sangat tidak seimbang (*imbalanced data*) [4], [11].

Meskipun arsitektur IndoBERT terbukti efisien dalam berbagai tugas pemrosesan bahasa alami, evaluasi dan penerapannya yang difokuskan secara spesifik pada pengolahan teks informal terkait pasar modal Indonesia di YouTube masih sangat terbatas. Mayoritas literatur belum secara komprehensif mengintegrasikan klasifikasi sentimen dengan ekstraksi fitur berbasis kata kunci (*keywords*) yang dapat merepresentasikan secara langsung alasan di balik polaritas opini investor ritel. Berdasarkan *research gap* tersebut, penelitian ini bertujuan untuk mengimplementasikan dan melakukan *fine-tuning* pada model IndoBERT guna mengklasifikasikan sentimen komentar YouTube ke dalam tiga polaritas: positif, negatif, dan netral. Selain berfokus pada klasifikasi, penelitian ini berkontribusi dalam mengekstraksi kata-kata kunci utama yang mendasari pembentukan opini publik. Melalui pemetaan distribusi polaritas dan evaluasi performa model terhadap karakteristik teks informal pasar modal, hasil penelitian ini diharapkan dapat menyajikan representasi psikologi pasar secara *real-time* yang valid, guna mendukung pengambilan keputusan komplementer dalam analisis fundamental dan teknikal.

2. Tinjauan Pustaka

Bagian ini menyajikan kajian terhadap berbagai literatur pendukung dan konsep teoretis yang relevan dari sumber pustaka terpercaya guna memberikan landasan akademis terhadap metode, pendekatan, serta metrik evaluasi yang digunakan dalam penelitian ini.

2.1. Analisis Sentimen

Analisis sentimen, atau dikenal sebagai *opinion mining*, merupakan cabang dari *Natural Language Processing* (NLP) yang berfokus pada ekstraksi, identifikasi, dan komputasi orientasi afektif serta kecenderungan emosional yang terkandung dalam data teks [7]. Dalam domain keuangan modern, analisis sentimen telah bergeser dari

sekadar pengenalan polaritas teks umum menjadi instrumen prediktif yang krusial untuk menangkap psikologi pasar [1]. Penelitian terdahulu mengonfirmasi bahwa pergerakan instrumen generasi baru di pasar modal tidak hanya didorong oleh data fundamental makroekonomi, tetapi juga oleh akumulasi ekspektasi psikologis yang terefleksi dari opini publik [6]. Hubungan empiris ini menjadikan klasifikasi sentimen ke dalam kategori positif, negatif, dan netral sebagai variabel penting dalam mengukur anomali pasar. Dalam konteks penelitian ini, analisis sentimen diterapkan secara spesifik untuk membedakan polaritas pandangan investor ritel terhadap dinamika pergerakan harga saham, di mana akurasi pemetaan persepsi tersebut menjadi fondasi utama dalam memvalidasi pergeseran sentimen pasar secara *real-time*.

2.2. YouTube sebagai Sumber Data

Transformasi digital telah mengubah lanskap penyebaran informasi keuangan, di mana platform media sosial seperti YouTube kini menjadi episentrum interaksi dan rujukan utama bagi investor ritel dalam mencari analisis serta rekomendasi saham [3]. Fenomena ini menghasilkan volume data opini yang masif melalui fitur interaktif seperti kolom komentar [4]. Eksplorasi data berskala besar ini dimungkinkan melalui teknik *crawling* data, sebuah proses otomatisasi pengambilan dokumen tekstual langsung dari antarmuka web [4]. Kendati demikian, pemanfaatan data hasil *crawling* dari YouTube menghadirkan tantangan komputasional yang signifikan karena sifatnya yang tidak terstruktur dan memiliki noise tinggi [2]. Dalam penelitian pasar modal Indonesia, data tekstual ini menjadi cerminan langsung dari sentimen akar rumput (*grassroots sentiment*). Kompleksitas pola bahasa alami yang dinamis dalam diskusi video saham menuntut adopsi model komputasi canggih yang mampu mengekstrak sinyal pasar tersembunyi di balik tumpukan data non-konvensional tersebut.

2.3. BERT dan IndoBERT

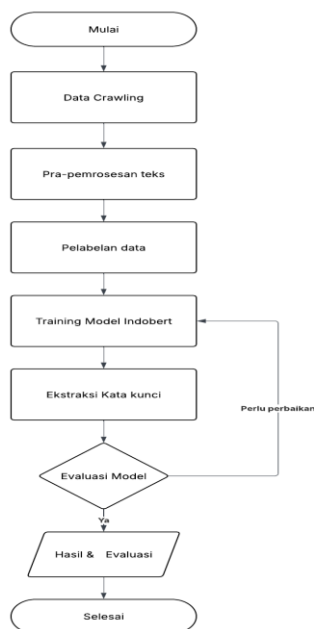
Arsitektur *Bidirectional Encoder Representations from Transformers* (BERT) membawa paradigma baru dalam representasi bahasa alami melalui mekanisme *self-attention* dua arah (*bidirectional*). Mekanisme ini memungkinkan model untuk mengodekan konteks suatu kata berdasarkan posisi sebelum dan sesudahnya secara simultan, mengatasi keterbatasan inheren pada model sekuensial tradisional seperti *Recurrent Neural Network* (RNN) dan *Long Short-Term Memory* (LSTM) yang kerap mengalami kehilangan informasi pada konteks kalimat panjang (*vanishing gradient*) [2], [8]. Sebagai bentuk adaptasi lokal, IndoBERT dikembangkan sebagai *pre-trained language model* yang dilatih secara khusus menggunakan korpus bahasa Indonesia berskala masif [9]. Penelitian oleh N. A. dkk. [1] menunjukkan bahwa IndoBERT memiliki efikasi yang jauh lebih superior dalam menangkap ambiguitas, sintaksis, dan nuansa semantik bahasa lokal jika dibandingkan dengan model multibahasa standar [5], [10]. Melalui tahapan *fine-tuning*, parameter internal IndoBERT dioptimalkan kembali menggunakan dataset spesifik pasar modal Indonesia, sehingga model dapat mengenali entitas dan bobot emosional dari diskusi finansial secara presisi [4], [11].

2.4. Pra-pemrosesan dan Normalisasi Slang

Karakteristik utama dari data tekstual media sosial, khususnya kolom komentar YouTube terkait bursa efek, adalah tingginya penggunaan bahasa informal, pencampuran kode (*code-mixing*), serta akronim spesifik [4]. Investor ritel Indonesia sering kali menggunakan istilah leksikal khusus seperti "cuan", "sangkut", atau "serok", di samping singkatan reguler pasar modal seperti "IHSG" dan "emiten" [7]. Kondisi ini menyebabkan tahap pra-pemrosesan dan normalisasi slang menjadi fase yang mutlak diperlukan sebelum data diumpukan ke model permodelan bahasa. Tanpa adanya normalisasi, variasi ejaan tidak baku dan istilah non-semantik akan mengonstruksi token baru yang tidak dikenali (*out-of-vocabulary*), sehingga mendegradasi performa ekstraksi fitur dari arsitektur Transformer. Proses koreksi teks informal menjadi bentuk baku esensial untuk memastikan bahwa representasi vektor yang dihasilkan oleh IndoBERT tetap fokus pada ekstraksi inti sentimen, sekaligus meminimalkan kesalahan interpretasi makna selama fase pelatihan model.

3. Metode dan Pembahasan

Penelitian ini menggunakan model *Deep Learning* berbasis arsitektur Transformer, yaitu IndoBERT. Secara spesifik, model dioptimasi menggunakan metode *fine-tuning* pada dataset komentar keuangan berbahasa Indonesia [10]. Pemilihan model ini didasarkan pada keunggulannya dalam memahami konteks bahasa alami (NLP) informal media sosial secara dua arah. Selain itu, penelitian ini memanfaatkan akselerasi perangkat keras menggunakan teknologi CUDA melalui GPU untuk meningkatkan kecepatan komputasi selama proses inferensi model [9]. Implementasi *GPU acceleration* memungkinkan pemrosesan ribuan komentar secara paralel dan lebih efisien dibandingkan penggunaan CPU biasa. Seluruh alur tahapan eksperimen penelitian ini disajikan secara sistematis pada Gambar 1.



Gambar 1. Flowchart Alur Penelitian

3.1. Data Crawling / Scraping

Tahap awal penelitian dimulai dengan akuisisi data dari platform YouTube melalui kombinasi teknik data crawling dan web scraping. Web scraping merupakan proses pengumpulan data secara otomatis dari situs web menggunakan program perangkat lunak atau bot [11]. Data yang dikumpulkan berfokus pada diskusi publik dalam kolom komentar video analisis pasar modal Indonesia terkait krisis kepengurusan dewan BEI untuk memperoleh basis data opini investor ritel. Proses pengambilan data dilakukan menggunakan parameter pencarian tertentu guna memastikan relevansi dataset yang diperoleh. Atribut utama yang diekstraksi meliputi identitas pengguna (*author*), isi komentar (*comment*), jumlah tanda suka (*likes*), serta waktu publikasi komentar (*published_at*). Data mentah yang berhasil dikumpulkan kemudian disimpan dalam format *Comma Separated Values* (.csv) untuk digunakan pada tahapan pemrosesan selanjutnya.

3.2. Pre-Processing (Pra-Pemrosesan Data)

Pre-processing merupakan tahapan krusial dalam NLP yang bertujuan untuk membersihkan teks mentah untuk mengekstraksi fitur-fitur penting dan meningkatkan relevansi kata sebelum diproses oleh model. Langkah-langkah pembersihan yang dilakukan meliputi:

- *Case Folding*: Mengonversi seluruh huruf kapital menjadi huruf kecil secara seragam.
- *Filtering* : Mengeliminasi tautan URL, karakter non-alfabet, simbol, angka, serta emoji menggunakan Ekspresi Reguler (Regex).
- *Slang Word Normalization* : Mengubah kata-kata tidak baku pasar modal (seperti "cuan" menjadi "untung", "sangkut" menjadi "tertahan") menjadi kosakata baku agar mudah dikenali oleh model [13].
- *Stopword Elimination*: Menyaring kata-kata umum (seperti "yang", "di", "dari") yang dominan namun tidak membawa nilai analitis terhadap muatan sentimen [10].

3.3. Tokenisasi dan Konfigurasi Input

Sebelum memasuki fase klasifikasi, data teks harus dikonversi ke dalam bentuk numerik yang dapat dipahami oleh model melalui proses tokenisasi. Tokenisasi merupakan tahapan pemrosesan teks yang bertujuan untuk memecah kalimat panjang menjadi unit-unit terkecil yang disebut token, untuk memfasilitasi ekstraksi fitur dan pengenalan struktur kalimat dalam analisis sentimen. Pada tahap ini, teks dipecah berdasarkan kosakata model IndoBERT *pre-trained* [10]. Konfigurasi input mencakup pengaktifan fitur *truncation* otomatis dengan batasan panjang maksimal (*max length*) ditetapkan sebesar 512 token untuk menjaga efisiensi memori

GPU [9]. Terakhir, dilakukan *tensor conversion* untuk mengubah daftar ID token numerik menjadi objek Tensor PyTorch agar siap diproses di lingkungan CUDA secara paralel.

3.4. Klasifikasi Sentimen

Tahap klasifikasi sentimen dilakukan dengan memasukkan representasi tensor hasil tokenisasi ke dalam model IndoBERT untuk mengekstraksi fitur semantik kontekstual dari setiap teks [1]. Model kemudian menghasilkan nilai prediksi awal (logits) pada masing-masing kelas sentimen. Proses penentuan label akhir dilakukan melalui beberapa tahapan, yaitu penggunaan fungsi softmax untuk mengubah nilai logits menjadi distribusi probabilitas dengan total nilai sebesar 1.0, dilanjutkan dengan proses argmax untuk memilih probabilitas tertinggi sebagai hasil prediksi akhir. Selanjutnya, indeks hasil prediksi dipetakan ke dalam tiga kategori polaritas utama, yaitu positif, netral, dan negatif.

3.5. Penyimpanan dan Output Analisis

Hasil proses klasifikasi yang berupa label sentimen kemudian diintegrasikan kembali ke dalam dataframe utama sebagai atribut baru bernama sentiment. Pada tahap akhir, seluruh data hasil analisis diserialisasikan dan disimpan ke dalam berkas baru dengan format .csv untuk memudahkan proses pengolahan lanjutan. Dataset hasil klasifikasi tersebut selanjutnya dimanfaatkan sebagai sumber data utama dalam proses visualisasi, seperti analisis distribusi polaritas sentimen, grafik frekuensi, serta pembentukan visualisasi kata kunci dominan menggunakan metode *Word Cloud* [11].

4. Hasil Penelitian

Bagian ini menyajikan hasil implementasi serta evaluasi model deep learning berbasis arsitektur IndoBERT yang digunakan untuk menganalisis sentimen opini investor pasar modal Indonesia terkait krisis kepemimpinan direksi BEI.

4.1. Tahap Data Crawling / Scraping

Implementasi akuisisi data pada platform YouTube berhasil merekam korpus data mentah berjumlah 1.000 baris rekaman ulasan diskusi komunitas investor digital. Seluruh korpus data direkam secara *real-time* berdasarkan video yang memiliki interaksi tinggi seputar isu pengunduran diri massal dewan BEI. Struktur data mentah yang berhasil diekstraksi disajikan pada sampel log Gambar 2.

index	author	comment	likes	published_at
0	@nurhadi_official	Kalo ekonoki hancur kenapa IMF aiap goontor dana pinjaman... masa IMF mau kasihbpinjaman ke negara bangkrut..mikir nyo..gak usah koar koar	0	2026-05-19T01:26:18Z
1	@muktiwibowo4492	♥MANTAP URAIANNYA BUNG.. SANGAT BERMANFAAT BUAT KAMI♥	1	2026-05-18T08:33:16Z
2	@sunarto12narto89	Terlalu banyak maling dan begal di semua lini yg merugikan negara dan rakyat Hukum harus tegas dan berpihak pada kebenaran dan kejujuran bukan manipulasi dan penipuan	0	2026-05-17T10:38:02Z
3	@bedjosoedarso701	Iya bener sih jadi kayak kasino bursa saham negri ini. Bandar bisa mainin harga saham mana yang akan naik dan mana yang akan turun setiap harinya.	0	2026-05-16T13:28:10Z
4	@bedjosoedarso701	kalimat sakti bang Pandji di pakai,"menurut keyakinan saya". 🤔🤔🤔	0	2026-05-16T13:22:50Z
5	@KoyokCabe	Bangkai bertahun tahun ditutupi ... Mulai insider, gorengan, rigging dst .. yg mmang tuh bandar2 ... Mereka dah simpen tilepan Wang di Cayman, virgin island, panama ... Paham kan ?	0	2026-05-14T00:24:07Z
6	@kangdeden1571	Sekarang 11 mei 2026 , terbukti semua	1	2026-05-11T13:00:14Z
7	@perfectgive9607	Mustahil bisa berubah, bukan pesimis tapi fakta, orang konoha cuma haram sama babi tapi uang korupsi dan pelicin harum, yggw heran mereka yg suka korupsi dan terima pelicin yg buat hancur ekonomi di semua sektor, klo indonesia bubar apa rupiah yg mereka kumpulkan masih ada harganya	1	2026-05-08T16:12:42Z

Gambar 2. Hasil Crawling Basis Data Mentah dari YouTube Finansial

4.2. Tahap Pre-Processing (Pra-Pemrosesan Data)

Melalui tahapan pembersihan karakter non-alfabet, penghapusan derau (noise), serta normalisasi istilah informal dan slang pasar modal, setiap komentar berhasil ditransformasikan ke dalam representasi teks yang lebih terstruktur pada atribut *cleaned_text*. Proses ini bertujuan meningkatkan kualitas fitur masukan agar lebih optimal dalam tahap klasifikasi sentimen. Contoh hasil pra-pemrosesan teks yang telah direduksi dari elemen-elemen tidak relevan ditunjukkan pada Gambar 3.

index	comment	cleaned_text
0	Wah wah. Pandai. Jadi. Presiden. Saja	pandai presiden saja
1	cintailah cutloss kalian 🍷	cintailah cutloss kalian
2	Salut dengan media ini... Dibalik gencar nya berita soal MBG .. ada informasi serius yang rakyat di paksa fokus soal pemberitaan MBG...	salut media dibalik gencar berita soal mbg informasi serius rakyat paksa fokus soal pemberitaan mbg
3	Sangat mudah Krn genderuwo srimulyani cabut, mereka panik gak mau kena kambing hitam atas carut marutnya keuangan negara	sangat mudah genderuwo srimulyani cabut panik mau kena kambing hitam atas carut marutnya keuangan negara
4	bener-bener udah kejadian sekarang yang di menit 3:12 pertanggal hari ini 30 April 2025 kata raymond bakalan terjadi di 2-3 bulan lagi ehheh beneran he predict the future bro 🍷	bener bener kejadian pertanggal april kata raymond bakalan terjadi lagi ehheh beneran he predict the future
5	Bro Raymond selalu memberikan informasi yang inspiratif ❤️	raymond selalu memberikan informasi inspiratif
6	Biar remuk redam hancur luluh sama sama elit dan rakyatnya mati kelaparan harta elit tidak berguna lagi	remuk redam hancur luluh elit rakyatnya mati kelaparan harta elit berguna lagi
7	Biar remuk redam hancur luluh sama sama elit dan rakyatnya mati kelaparan harta elit tidak berguna lagi	remuk redam hancur luluh elit rakyatnya mati kelaparan harta elit berguna lagi

Gambar 3. Proses Cleaned Text

Selanjutnya dilakukan tahap pembersihan data dengan menghapus baris yang tidak memiliki isi teks maupun mengandung nilai null guna menjaga stabilitas proses komputasi pada tahap klasifikasi sentimen. Hasil proses cleaning menunjukkan bahwa jumlah data tetap konsisten dengan dataset hasil crawling, sehingga mengindikasikan tidak ditemukannya data kosong pada kumpulan komentar YouTube yang digunakan. Tampilan dataset setelah melalui proses pembersihan data ditunjukkan pada Gambar 4.

index	cleaned_text
0	pandai presiden saja
1	cintailah cutloss kalian
2	salut media dibalik gencar berita soal mbg informasi serius rakyat paksa fokus soal pemberitaan mbg
3	sangat mudah genderuwo srimulyani cabut panik mau kena kambing hitam atas carut marutnya keuangan negara
4	bener bener kejadian pertanggal april kata raymond bakalan terjadi lagi ehheh beneran he predict the future
5	raymond selalu memberikan informasi inspiratif
6	remuk redam hancur luluh elit rakyatnya mati kelaparan harta elit berguna lagi

Gambar 4. Hasil pembersihan data

Tahap terakhir pada proses pra-pemrosesan adalah stopwords customization, yaitu penyaringan kata-kata umum, istilah informal, serta frasa yang terlalu dominan namun tidak memiliki kontribusi signifikan terhadap polaritas sentimen. Proses ini dilakukan secara dinamis untuk meningkatkan relevansi fitur teks yang digunakan dalam analisis sentimen komentar pasar modal. Karena mekanisme penyaringan diterapkan langsung selama proses pemrosesan dan visualisasi data, hasilnya tidak ditampilkan dalam bentuk tabel statis, melainkan tercermin pada kualitas hasil klasifikasi sentimen dan visualisasi kata kunci yang dihasilkan pada tahap akhir penelitian.

4.3. Hasil Tokenisasi dan Klasifikasi Sentimen

Model IndoBERT digunakan untuk mengekstraksi representasi semantik kontekstual dari atribut cleaned_text guna mengidentifikasi kecenderungan sentimen dan orientasi psikologis investor ritel terhadap kondisi pasar modal. Proses klasifikasi dilakukan dengan memanfaatkan pemahaman kontekstual model terhadap pola bahasa informal pada komentar YouTube, sehingga setiap teks dapat dipetakan ke dalam kategori sentimen positif, netral, atau negatif. Tampilan hasil keluaran klasifikasi sentimen dari arsitektur model IndoBERT ditunjukkan pada Gambar 5.

index	cleaned_text	sentiment
0	pandai presiden saja	negative
1	cintailah cutloss kalian	positive
2	salut media dibalik gencar berita soal mbg informasi serius rakyat paksa fokus soal pemberitaan mbg	negative
3	sangat mudah genderuwo srimulyani cabut panik mau kena kambing hitam atas carut marutnya keuangan negara	negative
4	bener bener kejadian pertanggal april kata raymond bakalan terjadi lagi ehheh beneran he predict the future	negative
5	raymond selalu memberikan informasi inspiratif	positive
6	remuk redam hancur luluh elit rakyatnya mati kelaparan harta elit berguna lagi	negative
7	remuk redam hancur luluh elit rakyatnya mati kelaparan harta elit berguna lagi	negative

Gambar 5. Proses Sentiment

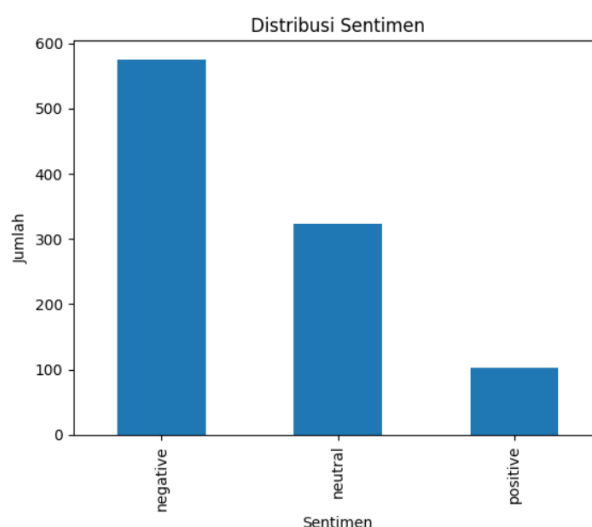
4.4. Distribusi Frekuensi Sentimen Kuantitatif

Berdasarkan hasil prediksi yang dihasilkan oleh model IndoBERT, dilakukan perhitungan frekuensi akumulatif pada setiap kategori polaritas sentimen secara terstruktur. Proses ini bertujuan untuk mengetahui distribusi dominasi sentimen positif, netral, dan negatif pada opini investor ritel terhadap isu pasar modal yang dianalisis. Representasi hasil komputasi kuantitatif terhadap distribusi polaritas sentimen tersebut ditampilkan pada Gambar 6.

sentiment	
negative	575
neutral	323
positive	102

Gambar 6. Perhitungan Jumlah *Sentiment*

Berdasarkan hasil prediksi yang dihasilkan oleh model IndoBERT, dilakukan perhitungan frekuensi akumulatif pada setiap kategori polaritas sentimen secara terstruktur. Proses ini bertujuan untuk mengetahui distribusi dominasi sentimen positif, netral, dan negatif pada opini investor ritel terhadap isu pasar modal yang dianalisis. Representasi hasil komputasi kuantitatif terhadap distribusi polaritas sentimen tersebut ditampilkan pada Gambar 7.



Gambar 7. Diagram Distribusi *Sentiment*

Hasil analisis visualisasi menunjukkan adanya dominasi mutlak dari sentimen negatif sebesar 57,5% (575 ulasan), mencerminkan adanya respons kecemasan masif dari pelaku pasar ritel terhadap stabilitas pasar modal Indonesia.

4.5. Ekstraksi Kata Kunci Dominan (*Word Cloud*)

Tahap akhir dari alur penelitian ini adalah melakukan penyimpanan dan agregasi statistik terhadap berkas keluaran model. Berdasarkan hasil perhitungan jumlah total sentimen pada 1.000 rekaman komentar, polaritas sentimen negatif mendominasi ruang diskusi publik secara mutlak dengan persentase sebesar 57,5% (575 ulasan), diikuti oleh sentimen netral sebesar 32,3% (323 ulasan), sedangkan sentimen positif menjadi bagian terkecil dengan perolehan hanya 10,2% (102 ulasan). Tingginya angka polaritas negatif ini merefleksikan adanya keresahan massal, ketidakpastian regulasi, serta penurunan tingkat kepercayaan secara drastis dari investor ritel terhadap stabilitas fungsional bursa efek pasca-isu pengunduran diri serempak dewan direksi BEI.

Analisis kualitatif dipertajam menggunakan teknik ekstraksi kata kunci berbasis frekuensi kemunculan kata pada visualisasi *Word Cloud* untuk memetakan perbedaan fokus psikologis antar-klaster emosi secara presisi. Representasi pemetaan kata dominan disajikan pada Gambar 8.

- [8] W. M. Munandar, "Sentiment Analysis of Student Comment on the College Performance Evaluation Questionnaire Using Naïve Bayes and IndoBERT," 2023. [Online]. Available: <https://github.com/notfound313/sentimen->
- [9] M. G. Al-Kadzim, R. Rasim, and H. Herbert, "Analisis Perubahan Sentimen Publik di Media Sosial X terhadap Konflik Palestina-Israel Menggunakan Model IndoBERT," *Digital Transformation Technology*, vol. 4, no. 2, pp. 1167–1174, Jan. 2025, doi: 10.47709/digitech.v4i2.5312.
- [10] U. Fadhila Dina Puspita, A. Pramita Widyasari, P. Studi Informatika, and S. Tinggi Teknologi Ronggolawe, "Program Studi Teknik Informatika," 2026.
- [11] S. Maratus Sholihah, Afril Efan Pajri, and Ita Aristia Sa'ida, "Analisis Sentimen Komentar iPhone 17 pada Platform YouTube Menggunakan IndoBERT dan Support Vector Machine," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 7, no. 3, pp. 933–945, Mar. 2026, doi: 10.30865/json.v7i3.9507.