

Analisis Sentimen Masyarakat terhadap Kenaikan Harga BBM Non-Subsidi Akibat Penutupan Selat Hormuz Menggunakan IndoBERT

Jaysen Stephanus ^{1,*}, Jonathan Tanujaya ² and Muhammad Rizky Pribadi ³

^{1,2,3}Program Studi Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang; jaysenstephanus_2327250042@mhs.mdp.ac.id; jonathantanujaya_2327250025@mhs.mdp.ac.id; rizky@mdp.ac.id

* Korespondensi: jaysenstephanus_2327250042@mhs.mdp.ac.id

Info Artikel:

Dikirim: 01 Mei 2026

Direvisi: 30 Mei 2026

Diterima: 06 Mei 2026

Abstract: Public discussions regarding the potential increase in non-subsidized fuel prices resulting from the closure of the Strait of Hormuz on the X platform between January 1, 2026, and May 17, 2026, were highly intensive and generated diverse public responses to the global economic impacts triggered by the geopolitical conflict between Iran and Israel. The primary issue addressed in this study is the growing public concern over the possibility of rising non-subsidized fuel prices, which may affect transportation costs, logistics distribution, and daily living expenses. This study aims to analyze public sentiment toward this issue using the IndoBERT deep learning model to obtain a more accurate understanding of public opinion trends. Data were collected through a scraping process on the X platform using keywords related to non-subsidized fuel and the Strait of Hormuz. The collected data were then processed through several preprocessing stages, including case folding, noise removal, tokenization, stopword removal, and stemming, before being classified into positive, neutral, and negative sentiment categories. Out of 412 analyzed tweets, negative sentiment emerged as the dominant category at 49.8%, followed by neutral sentiment at 48.5%, while positive sentiment accounted for only 1.7%. The findings indicate that the majority of the public expressed concern regarding the potential increase in non-subsidized fuel prices and its impact on economic conditions and household expenditures.

Keywords: Sentiment Analysis; Non-Subsidized Fuel; Deep Learning; IndoBERT; Natural Language Processing; Strait of Hormuz; X.

Intisari: Penyebaran opini masyarakat mengenai potensi kenaikan harga BBM non-subsidi akibat penutupan Selat Hormuz di platform X pada periode 1 Januari 2026 hingga 17 Mei 2026 berlangsung secara intens dan memunculkan beragam respons publik terhadap dampak ekonomi global yang dipicu konflik geopolitik Iran dan Israel. Permasalahan utama dalam penelitian ini adalah tingginya kekhawatiran masyarakat terhadap kemungkinan kenaikan harga BBM non-subsidi yang berdampak pada biaya transportasi, distribusi logistik, serta kebutuhan sehari-hari. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap isu tersebut menggunakan model deep learning IndoBERT guna memahami kecenderungan opini publik secara lebih akurat. Data dikumpulkan melalui proses scraping pada platform X menggunakan kata kunci terkait BBM non-subsidi dan Selat Hormuz, kemudian diproses melalui tahapan pre-processing seperti *case folding*, *noise removal*, *tokenization*, *stopword removal*, dan *stemming* sebelum diklasifikasikan ke dalam kategori sentimen positif, netral, dan negatif. Dari total 412 tweet yang dianalisis, sentimen negatif menjadi kategori dominan sebesar 49,8%, diikuti sentimen netral sebesar 48,5%, dan sentimen positif sebesar 1,7%. Hasil penelitian menunjukkan bahwa mayoritas

masyarakat memberikan respons kekhawatiran terhadap potensi kenaikan harga BBM non-subsidi serta dampaknya terhadap kondisi ekonomi masyarakat.

Kata Kunci: Analisis sentimen; BBM non-subsidi; *Deep Learning*; IndoBERT; *Natural Language Processing*; Selat Hormuz; X.

1. Pendahuluan

Memasnya konflik geopolitik antara Iran dan Israel telah memicu kekhawatiran mendalam terhadap stabilitas ekonomi global, khususnya pada sektor energi dan distribusi minyak dunia. Salah satu dampak paling krusial dari eskalasi ini adalah potensi penutupan Selat Hormuz yang menjadi jalur utama distribusi minyak dari kawasan Teluk Persia. Selat Hormuz memegang peran vital dalam perdagangan energi global karena mengalirkan sekitar 20% dari total konsumsi minyak dunia. Ketidakpastian jalur logistik ini menyebabkan harga minyak mentah dunia melonjak signifikan, yang pada gilirannya berpotensi memaksa penyesuaian harga Bahan Bakar Minyak (BBM) non-subsidi di Indonesia guna menekan beban anggaran negara. Dampak domino dari potensi kenaikan harga BBM tersebut memicu beragam respons dan opini dari masyarakat yang disampaikan secara masif melalui media sosial. Oleh karena itu, fenomena ini mendesak untuk diteliti lebih lanjut melalui pendekatan analisis sentimen dengan menerapkan model *deep learning* IndoBERT guna menangkap dinamika opini publik secara akurat [1].

Kenaikan harga Bahan Bakar Minyak (BBM) merupakan isu yang sangat sensitif di Indonesia karena berdampak langsung terhadap berbagai aspek kehidupan, mulai dari biaya transportasi, distribusi logistik, harga kebutuhan pokok, hingga tingkat inflasi nasional. Komoditas energi ini telah menjadi kebutuhan vital yang digunakan oleh hampir seluruh lapisan masyarakat, terutama pemilik kendaraan bermotor yang jumlahnya terus melonjak setiap tahun. Penyesuaian tarif tersebut, baik pada sektor subsidi maupun non-subsidi, seketika memicu kekhawatiran publik karena dapat menurunkan daya beli serta memperbesar beban ekonomi rumah tangga [2]. Dalam era digital saat ini, platform X menjadi salah satu ruang komunikasi publik yang paling aktif digunakan karena memungkinkan pengguna menyampaikan informasi dan pendapat secara cepat, terbuka, dan real-time. Tingginya intensitas interaksi pengguna pada platform X menjadikan media sosial ini sebagai wadah utama masyarakat dalam mengekspresikan pandangan terkait kenaikan harga BBM, sehingga data opini yang dihasilkan sangat relevan untuk dianalisis menggunakan metode analisis sentimen [3].

Penelitian terdahulu mengenai analisis sentimen terhadap isu kenaikan harga Bahan Bakar Minyak (BBM) di Indonesia telah dilakukan oleh Nurrahman dkk. menggunakan metode *Machine Learning* konvensional berupa *Support Vector Machine* (SVM) yang dikombinasikan dengan teknik penyeimbangan data *Synthetic Minority Oversampling Technique* (SMOTE) [4]. Hasil penelitian menunjukkan bahwa model SVM sebelum penerapan SMOTE memiliki ketimpangan performa klasifikasi, di mana akurasi pada kelas negatif mencapai 94%, sedangkan kemampuan mengenali kelas positif hanya sebesar 38% [4]. Setelah dilakukan penerapan SMOTE, performa klasifikasi meningkat dengan akurasi keseluruhan mencapai 84% [4]. Meskipun demikian, pendekatan berbasis *Machine Learning* tradisional seperti SVM masih memiliki keterbatasan dalam memahami konteks semantik pada teks media sosial karena sangat bergantung pada proses ekstraksi fitur manual (*feature engineering*). Kondisi tersebut menyebabkan model rentan mengalami misklasifikasi ketika menghadapi bahasa informal, singkatan, sarkasme, maupun variasi bahasa yang umum digunakan pengguna media sosial. Oleh karena itu, diperlukan pendekatan yang lebih adaptif dan kontekstual untuk meningkatkan performa analisis sentimen pada data X terkait isu kenaikan harga BBM non-subsidi.

Berdasarkan keterbatasan tersebut, pendekatan berbasis *deep learning* dengan arsitektur Transformer melalui metode transfer learning mulai banyak diterapkan untuk mengatasi permasalahan pemrosesan bahasa alami (*Natural Language Processing/NLP*) pada karakteristik teks media sosial [5]. Salah satu model yang dirancang untuk memahami karakteristik bahasa Indonesia adalah IndoBERT [5]. Sebagai pre-trained language model, IndoBERT memiliki kemampuan memahami konteks kalimat secara dua arah (bidirectional) melalui mekanisme self-attention sehingga mampu menangkap hubungan semantik antar kata dengan lebih baik [5], [6]. Berbeda dengan metode *Machine Learning* tradisional seperti SVM yang bergantung pada ekstraksi fitur manual, IndoBERT lebih adaptif dalam memahami bahasa informal, singkatan, maupun struktur kalimat tidak baku yang umum ditemukan pada media sosial [6]. Penelitian terdahulu menunjukkan bahwa model SVM berbasis TF-IDF memperoleh akurasi sebesar 83,29%, sedangkan model IndoBERT mampu mencapai akurasi sebesar 92,70% pada tugas klasifikasi sentimen [6]. Berdasarkan keunggulan tersebut, penelitian ini menggunakan varian IndoBERT untuk menganalisis sentimen masyarakat terhadap kenaikan harga BBM non-subsidi pada platform X.

Berdasarkan latar belakang krisis ekonomi geopolitik dan celah penelitian komputasi yang telah dipaparkan, penelitian ini bertujuan untuk mengklasifikasikan dan menganalisis dinamika sentimen masyarakat Indonesia di

platform X terhadap isu kenaikan harga BBM non-subsidi sebagai dampak dari potensi penutupan Selat Hormuz dengan menerapkan model *deep learning* IndoBERT. Kontribusi utama dari penelitian ini adalah menyajikan sebuah kerangka kerja pemetaan opini publik yang lebih akurat, objektif, dan *real-time* berbasis kecerdasan buatan tingkat lanjut tanpa bergantung pada teknik penyeimbangan data manual. Hasil analisis dari penelitian ini diharapkan dapat menjadi bahan rujukan teoretis dalam pengembangan model NLP berbahasa informal Indonesia, sekaligus menjadi instrumen evaluasi empiris bagi para pembuat kebijakan (*stakeholders*) dalam mengukur resiliensi psikologis serta dampak sosial-ekonomi masyarakat digital akibat fluktuasi energi global.

2. Tinjauan Pustaka

Bab ini berisi tinjauan pustaka yang digunakan sebagai dasar teori dalam penelitian ini, meliputi konsep NLP, analisis sentimen, media sosial, serta metode IndoBERT dan penelitian terdahulu yang relevan sebagai acuan dalam penyusunan penelitian.

2.1. Natural Language Processing (NLP)

Natural Language Processing (NLP) atau Pemrosesan Bahasa Alami merupakan cabang dari kecerdasan buatan (*Artificial Intelligence/AI*) yang berfokus pada interaksi antara manusia dan komputer melalui bahasa alami, baik dalam bentuk teks maupun suara. NLP memungkinkan komputer untuk memahami, menganalisis, menghasilkan, dan merespons bahasa manusia sehingga komunikasi antara manusia dan mesin menjadi lebih intuitif dan efisien. Perkembangan NLP mengalami kemajuan signifikan seiring penerapan metode *machine learning* dan *deep learning*, khususnya melalui arsitektur *Transformer* seperti BERT (*Bidirectional Encoder Representations from Transformers*) dan GPT (*Generative Pre-trained Transformer*) yang mampu memahami konteks kata dan frasa secara lebih mendalam menggunakan *contextual embeddings*. Dalam implementasinya, NLP mencakup berbagai komponen seperti *syntactic parsing*, *semantic analysis*, *Named Entity Recognition (NER)*, *sentiment analysis*, dan *text generation* yang dimanfaatkan pada berbagai aplikasi, seperti chatbot, asisten virtual, penerjemah bahasa otomatis, serta analisis opini di media sosial. Meskipun demikian, NLP masih menghadapi berbagai tantangan, seperti ambiguitas bahasa, bias data, keterbatasan bahasa lokal, serta kompleksitas pemahaman konteks, sehingga pengembangan teknologi ini terus dilakukan untuk meningkatkan kemampuan mesin dalam memahami bahasa manusia secara lebih alami dan akurat [7].

2.2. Analisis Sentimen

Analisis sentimen merupakan salah satu teknik dalam *Natural Language Processing (NLP)* yang digunakan untuk mengidentifikasi dan mengevaluasi opini seseorang terhadap suatu topik, produk, maupun layanan berdasarkan data berbentuk teks. Metode ini bertujuan untuk mengetahui kecenderungan sentimen yang muncul, baik berupa opini positif, negatif, maupun netral, sehingga dapat memberikan gambaran mengenai persepsi publik terhadap suatu objek tertentu. Analisis sentimen atau dapat disebut juga sebagai *opinion mining* merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk memperoleh informasi mengenai sentimen yang terkandung dalam suatu kalimat apakah berupa opini positif atau negatif [8]. Dalam penerapannya, analisis sentimen banyak dimanfaatkan pada media sosial, ulasan pelanggan, forum diskusi, serta survei kepuasan untuk membantu memahami tanggapan pengguna secara lebih efektif. Dengan adanya perkembangan teknologi NLP dan *machine learning*, proses pengolahan data teks dalam jumlah besar dapat dilakukan secara lebih cepat dan efisien sehingga menghasilkan informasi yang mendukung pengambilan keputusan.

2.3. X

X atau yang sebelumnya dikenal sebagai Twitter merupakan salah satu platform media sosial berbasis *microblogging* yang memungkinkan pengguna untuk berbagi informasi, opini, serta berinteraksi secara *real-time* melalui unggahan singkat, balasan, dan ruang komentar. Platform ini menjadi salah satu media komunikasi digital yang dinamis karena memungkinkan pengguna menyampaikan pendapat secara terbuka kepada audiens yang luas dan beragam. Selain digunakan sebagai sarana berbagi informasi dan hiburan, X juga berperan sebagai ruang diskusi publik yang sering digunakan untuk membahas isu sosial, politik, budaya, hingga fenomena yang sedang menjadi perhatian masyarakat [3]. Dalam praktiknya, ruang komentar pada platform X menjadi tempat berlangsungnya berbagai bentuk interaksi sosial, seperti penyampaian kritik, opini, dukungan, maupun resistansi sosial. Pengguna sering kali menggunakan berbagai strategi komunikasi melalui pemilihan diksi, simbol, emoji, maupun gaya bahasa tertentu untuk menyampaikan pendapat atau menanggapi perbedaan pandangan. Karakteristik komunikasi pada platform ini yang cepat, terbuka, dan interaktif menjadikan X sebagai

media yang relevan untuk menganalisis opini publik, pola komunikasi digital, serta perilaku pengguna media sosial [3].

2.4. IndoBERT

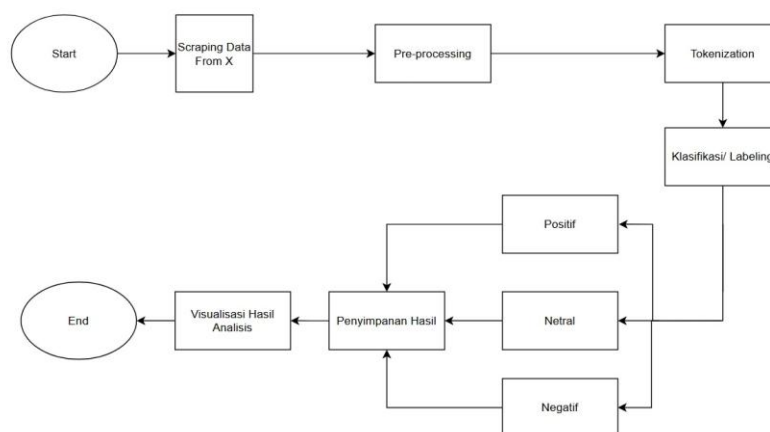
IndoBERT merupakan model *pre-trained language model* berbasis *Bidirectional Encoder Representations from Transformers* (BERT) yang dikembangkan secara khusus untuk pemrosesan bahasa Indonesia. Model ini dilatih menggunakan korpus teks berbahasa Indonesia dalam jumlah besar yang mencakup berbagai sumber, seperti berita daring, Wikipedia Bahasa Indonesia, forum diskusi, serta media sosial, sehingga mampu memahami karakteristik linguistik bahasa Indonesia dengan lebih baik [9]. IndoBERT menggunakan arsitektur *Transformer* yang memungkinkan pemahaman konteks kata secara dua arah (*bidirectional*), sehingga hubungan semantik antar kata dalam suatu kalimat dapat dipahami secara lebih mendalam dibandingkan metode tradisional [9]. Kemampuan tersebut menjadikan IndoBERT efektif dalam berbagai tugas NLP, seperti klasifikasi teks, deteksi ujaran kebencian, dan analisis sentimen, terutama pada data berbahasa Indonesia yang memiliki banyak variasi kata tidak baku, singkatan, maupun penggunaan bahasa informal di media sosial. Selain itu, IndoBERT dapat disesuaikan untuk tugas tertentu melalui proses *fine-tuning*, sehingga model tidak perlu dilatih dari awal dan dapat menghasilkan performa klasifikasi yang lebih optimal sesuai kebutuhan penelitian [9].

3. Metode dan Pembahasan

Penelitian ini menggunakan pendekatan *Deep Learning* berbasis arsitektur *Transformer*, yaitu IndoBERT (*Indonesian Bidirectional Encoder Representations from Transformers*) untuk melakukan klasifikasi sentimen masyarakat terhadap isu kenaikan harga BBM non-subsidi akibat potensi penutupan Selat Hormuz. Model IndoBERT dipilih karena memiliki kemampuan memahami konteks bahasa Indonesia secara dua arah (*bidirectional*) melalui mekanisme *self-attention*, sehingga mampu mengenali hubungan semantik antar kata secara lebih baik dibandingkan metode *machine learning* konvensional [9]. Selain itu, IndoBERT juga dinilai lebih adaptif dalam memahami karakteristik bahasa informal yang umum digunakan pada media sosial X, seperti singkatan, bahasa gaul, maupun variasi ekspresi emosional yang sering muncul pada opini masyarakat digital.

Dalam penelitian ini, tahapan penelitian dilakukan secara sistematis untuk memperoleh hasil klasifikasi sentimen yang akurat. Proses penelitian diawali dengan pengumpulan data opini publik pada platform X menggunakan teknik *data crawling* atau *web scraping* berdasarkan kata kunci yang relevan dengan isu kenaikan harga BBM non-subsidi akibat penutupan Selat Hormuz. Data yang diperoleh kemudian melalui tahap *pre-processing* untuk membersihkan teks dari unsur-unsur yang tidak relevan agar lebih terstruktur dan siap diproses oleh model. Selanjutnya dilakukan proses tokenisasi untuk mengubah teks ke dalam bentuk numerik yang dapat dipahami oleh IndoBERT. Tahap berikutnya adalah klasifikasi sentimen untuk memetakan opini masyarakat ke dalam kategori sentimen positif, netral, atau negatif. Proses akhir penelitian meliputi penyimpanan hasil analisis serta visualisasi distribusi sentimen guna memperoleh gambaran umum mengenai respons masyarakat terhadap isu tersebut.

Penelitian ini dilakukan melalui beberapa tahapan sistematis yang dimulai dari proses pengumpulan data opini masyarakat pada platform X, dilanjutkan dengan tahap *pre-processing*, tokenisasi, klasifikasi sentimen menggunakan model IndoBERT, hingga visualisasi hasil analisis. Alur tahapan penelitian secara keseluruhan ditunjukkan pada **Gambar 1**.



Gambar 1. Flowchart Alur Penelitian

3.1. Pengumpulan Data (Data Scraping)

Tahap awal penelitian dilakukan melalui proses pengumpulan data dari platform X menggunakan teknik *data scraping*. Teknik ini digunakan untuk memperoleh data opini masyarakat terkait isu kenaikan harga BBM non-subsidi akibat potensi penutupan Selat Hormuz. Pengambilan data dilakukan berdasarkan beberapa kata kunci yang relevan dengan topik penelitian, seperti “BBM naik”, “harga pertamax”, “Pertamax naik”, “Dexlite”, “Pertamina Dex”, “BBM non subsidi”, “Selat Hormuz”, “harga minyak dunia”, dan “minyak naik perang” guna memperoleh unggahan yang berkaitan dengan respons masyarakat terhadap isu tersebut. Data yang dikumpulkan berupa unggahan (*tweet*) masyarakat yang berisi opini, tanggapan, maupun diskusi terkait dampak ekonomi dari potensi kenaikan harga BBM non-subsidi.

Adapun atribut data yang diekstraksi dalam penelitian ini meliputi nama pengguna (*username*), isi unggahan (*tweet text*), serta waktu unggahan (*timestamp*). Untuk menjaga kualitas data, dilakukan proses seleksi dengan menghapus data duplikat, unggahan yang tidak relevan, serta data yang tidak memiliki isi teks yang dapat dianalisis. Seluruh data yang telah dikumpulkan kemudian disimpan dalam format *Comma Separated Values (.csv)* untuk mempermudah proses pengolahan data pada tahap selanjutnya menggunakan model IndoBERT.

3.2. Pra-pemrosesan Data (Pre-processing)

Tahap *pre-processing* dilakukan untuk membersihkan dan menyiapkan data teks hasil *scraping* agar lebih terstruktur sebelum diproses oleh model IndoBERT. Data mentah yang diperoleh dari platform X (Twitter) umumnya masih mengandung berbagai unsur yang tidak relevan (*noise*), seperti tautan (*URL*), simbol, emoji, tanda baca, maupun variasi penulisan kata tidak baku yang dapat memengaruhi hasil klasifikasi sentimen. Oleh karena itu, proses *pre-processing* diperlukan untuk meningkatkan kualitas data agar model dapat memahami konteks teks dengan lebih optimal.

Pada penelitian ini, tahapan *pre-processing* terdiri dari beberapa proses, yaitu *case folding*, *noise removal*, *tokenization*, *stopwords*, dan *stemming*. *Case folding* dilakukan dengan mengubah seluruh teks menjadi huruf kecil (*lowercase*) untuk menjaga konsistensi format data. *Noise removal* bertujuan menghapus karakter yang tidak diperlukan, seperti *mention*, *hashtag*, tautan (*URL*), emoji, angka, dan simbol tertentu. Selanjutnya, *tokenization* dilakukan untuk memecah kalimat menjadi unit-unit kata (*token*) agar mempermudah proses analisis. Tahap *stopwords* digunakan untuk menghilangkan kata-kata umum yang tidak memiliki kontribusi besar terhadap makna sentimen, sedangkan *stemming* dilakukan untuk mengubah kata berimbuhan menjadi kata dasar sehingga teks menjadi lebih konsisten dan mempermudah model dalam memahami konteks bahasa Indonesia.

3.3 Implementasi Model IndoBERT

Tahap implementasi model dilakukan dengan menggunakan IndoBERT sebagai model utama dalam proses klasifikasi sentimen. IndoBERT dipilih karena merupakan *pre-trained language model* berbasis arsitektur *Transformer* yang dirancang khusus untuk memahami karakteristik bahasa Indonesia. Model ini memiliki kemampuan memahami konteks kalimat secara dua arah (*bidirectional*) melalui mekanisme *self-attention*, sehingga mampu menangkap hubungan semantik antar kata dengan lebih baik, termasuk pada teks media sosial yang cenderung informal dan tidak terstruktur [9]. Penggunaan IndoBERT dinilai lebih efektif dibandingkan metode konvensional karena tidak memerlukan proses ekstraksi fitur manual (*feature engineering*) dan mampu memahami konteks bahasa secara lebih mendalam.

Sebelum dilakukan proses klasifikasi, data hasil *pre-processing* terlebih dahulu melalui tahap *tokenization* menggunakan *tokenizer* bawaan IndoBERT untuk mengubah teks menjadi representasi numerik (*token ID*) yang dapat dipahami model. Pada tahap ini diterapkan konfigurasi input berupa *padding* untuk menyeragamkan panjang data, *truncation* untuk memotong teks yang melebihi batas maksimum, serta *max length* guna menjaga efisiensi proses komputasi. Setelah proses tokenisasi selesai, data diproses oleh model IndoBERT untuk menghasilkan representasi konteks teks yang kemudian digunakan pada tahap klasifikasi sentimen terhadap opini masyarakat mengenai kenaikan harga BBM non-subsidi akibat potensi penutupan Selat Hormuz.

3.4 Klasifikasi Sentimen

Tahap klasifikasi sentimen dilakukan setelah data melewati proses *pre-processing* dan tokenisasi menggunakan model IndoBERT. Pada tahap ini, teks yang telah dikonversi ke dalam bentuk numerik diproses oleh model untuk mengidentifikasi kecenderungan opini masyarakat terhadap isu kenaikan harga BBM non-subsidi akibat potensi penutupan Selat Hormuz. Model IndoBERT akan menganalisis konteks keseluruhan

kalimat melalui mekanisme *self-attention* untuk memahami hubungan antar kata sehingga mampu memberikan prediksi sentimen secara lebih akurat [9].

Hasil prediksi model kemudian diklasifikasikan ke dalam tiga kategori utama, yaitu sentimen positif, netral, dan negatif. Proses penentuan label dilakukan berdasarkan nilai probabilitas yang dihasilkan model untuk setiap kategori sentimen, di mana kelas dengan probabilitas tertinggi dipilih sebagai hasil klasifikasi akhir. Hasil klasifikasi ini digunakan untuk mengetahui kecenderungan respons masyarakat terhadap potensi kenaikan harga BBM non-subsidi serta menjadi dasar dalam proses analisis dan interpretasi sentimen publik pada tahap berikutnya.

3.5 Visualisasi dan Analisis Hasil

Tahap akhir penelitian dilakukan dengan memvisualisasikan hasil klasifikasi sentimen yang diperoleh dari model IndoBERT untuk mempermudah proses interpretasi data. Hasil prediksi sentimen berupa kategori positif, netral, dan negatif digunakan untuk melihat kecenderungan opini masyarakat terhadap isu kenaikan harga BBM non-subsidi akibat potensi penutupan Selat Hormuz. Visualisasi data bertujuan untuk memberikan gambaran yang lebih jelas mengenai distribusi sentimen publik berdasarkan hasil klasifikasi model.

Pada penelitian ini, visualisasi hasil analisis dapat disajikan dalam bentuk diagram batang, diagram lingkaran, maupun grafik distribusi sentimen lainnya untuk menunjukkan perbandingan jumlah sentimen positif, netral, dan negatif yang muncul pada data opini masyarakat. Hasil visualisasi tersebut kemudian dianalisis untuk mengidentifikasi pola respons masyarakat serta mengetahui kecenderungan persepsi publik terhadap isu kenaikan harga BBM non-subsidi yang dipengaruhi oleh dinamika geopolitik global.

4. Hasil Penelitian

Penelitian ini berhasil mengimplementasikan model IndoBERT untuk melakukan klasifikasi sentimen masyarakat terhadap isu kenaikan harga BBM non-subsidi akibat potensi penutupan Selat Hormuz berdasarkan data opini pada platform X. Proses penelitian dilakukan melalui beberapa tahapan utama, yaitu pengumpulan data, *pre-processing*, tokenisasi, klasifikasi sentimen, serta visualisasi hasil analisis. Hasil penelitian menunjukkan bahwa model IndoBERT mampu melakukan klasifikasi sentimen dengan baik pada data teks berbahasa Indonesia yang bersifat informal dan dinamis.

4.1. Hasil Pengumpulan Data

Pengumpulan data dilakukan menggunakan teknik *scraping* pada platform X dalam periode 1 Januari 2026 hingga 17 Mei 2026. Proses *scraping* dilakukan menggunakan beberapa kata kunci yang berkaitan dengan isu kenaikan harga BBM non-subsidi dan penutupan Selat Hormuz. Berdasarkan proses tersebut, berhasil dikumpulkan sebanyak 412 data *tweet* yang selanjutnya digunakan dalam proses analisis sentimen menggunakan model IndoBERT. Tampilan data dapat dilihat di **Gambar 2**.

| conversation_id_str | created_at | favorite_count | full_text | id_str | image_url | in_reply_to_screen_name | lang | location | quote_count | reply_count | retweet_count |
|---------------------|---------------------|--------------------------------|---|---------------------|---|-------------------------|------|----------|-------------|-------------|---------------|
| 0 | 2054008506735640613 | Tue May 12 01:19:24 +0000 2026 | Bulyait Swasembada BBM: Tarif mobil online aja... | 2054008506735640613 | NaN | NaN | in | NaN | 0 | 0 | 0 |
| 1 | 2054005913435185305 | Tue May 12 01:09:06 +0000 2026 | Rupiah melemah bukan cuma angka. Arinya Semb... | 2054005916576686520 | NaN | mumtazmr | in | NaN | 0 | 1 | 0 |
| 2 | 2054004819892089120 | Tue May 12 01:04:45 +0000 2026 | Pengaruh BBM pada mesin AFR & timing ign... | 2054004819892089120 | https://pbs.twimg.com/media/HIFKcJzbUAEaf--.jpg | NaN | in | NaN | 1 | 4 | 3 |
| 3 | 205371295675458353 | Tue May 12 01:02:39 +0000 2026 | @tempodoto Fenomena aneh di Indonesia pemerin... | 2054004290461778313 | NaN | tempodoto | in | NaN | 0 | 0 | 0 |
| 4 | 2053993648526889207 | Tue May 12 00:52:01 +0000 2026 | Update Harga BBM Pertamina Shell hingga Vivo N... | 2053993648526889207 | NaN | NaN | in | NaN | 0 | 0 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 407 | 2053669440777224502 | Mon May 11 02:52:05 +0000 2026 | Kalau dunia sedang masuk fase perang minyak ma... | 2053669442668974233 | NaN | buildportosaham | in | NaN | 0 | 1 | 0 |
| 408 | 2053629971177066647 | Mon May 11 00:15:14 +0000 2026 | GM, Sabtu woles Minggu Cari Start Senin? Tru... | 2053629971177066647 | https://pbs.twimg.com/media/HH_1kHyBoAAWxWM.jpg | NaN | in | NaN | 1 | 13 | 4 |
| 409 | 2052673162974986269 | Sun May 10 00:07:30 +0000 2026 | @geloraco Ha ha ha... melipir mereka klo sdr b... | 2053265636261060609 | NaN | geloraco | in | NaN | 0 | 0 | 0 |
| 410 | 2052565068932591685 | Sat May 09 08:08:51 +0000 2026 | @arman_dhani Kik mnyak dunia lgi naik Km per... | 2053024384319963359 | NaN | arman_dhani | in | NaN | 0 | 0 | 0 |

Gambar 2. Data Analisis Sentimen

4.2. Hasil Pre-pemrosesan Data

Tahap *pre-processing* dilakukan untuk membersihkan data hasil *scraping* agar lebih terstruktur sebelum diproses menggunakan model IndoBERT. Proses ini meliputi *case folding*, *noise removal*, *tokenization*, *stopwords removal*, dan *stemming*. Pada tahap ini dilakukan penghapusan elemen yang tidak relevan seperti URL, *mention*, *hashtag*, *emoji*, tanda baca, serta karakter khusus lainnya yang dapat memengaruhi proses klasifikasi sentimen.

Selain itu, seluruh teks diubah menjadi huruf kecil dan kata-kata yang tidak memiliki makna penting dihilangkan agar data menjadi lebih konsisten. Hasil *pre-processing* menghasilkan data teks yang lebih bersih dan siap digunakan pada tahap klasifikasi sentimen.

4.3. Implementasi dan Pengujian Model IndoBERT

Pada tahap implementasi, data hasil pre-processing terlebih dahulu dilakukan tokenisasi menggunakan *tokenizer* bawaan IndoBERT untuk mengubah teks menjadi token ID yang dapat dipahami oleh model. Selanjutnya, proses klasifikasi dilakukan menggunakan model IndoBERT dengan konfigurasi *padding*, *truncation*, dan *max length* agar proses komputasi berjalan lebih optimal.

Model IndoBERT digunakan untuk memahami konteks kalimat secara dua arah (*bidirectional*) sehingga mampu mengenali hubungan semantik antar kata pada teks media sosial. Hasil implementasi menunjukkan bahwa model dapat digunakan dengan baik dalam proses klasifikasi sentimen terhadap opini masyarakat terkait kenaikan harga BBM non-subsidi.

4.4. Hasil Klasifikasi Sentimen

Pada tahap klasifikasi sentimen, data *tweet* yang telah melalui proses *pre-processing* diklasifikasikan menggunakan model IndoBERT ke dalam tiga kategori sentimen, yaitu sentimen positif, netral, dan negatif. Hasil klasifikasi menunjukkan adanya variasi opini masyarakat terhadap isu kenaikan harga BBM non-subsidi akibat potensi penutupan Selat Hormuz. Distribusi jumlah data pada masing-masing kategori sentimen dapat dilihat pada **Gambar 3**.

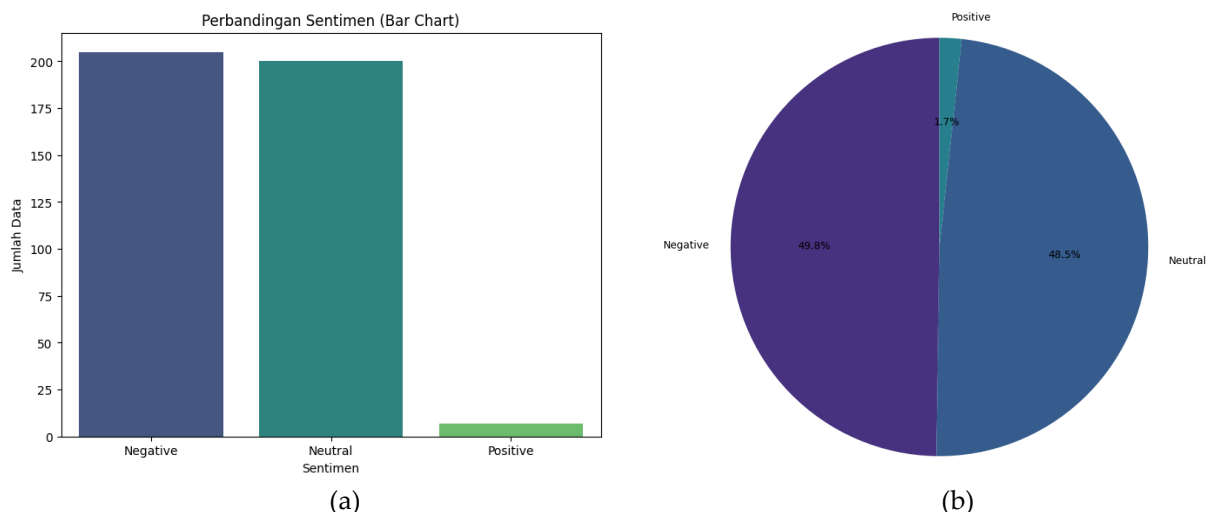
| | full_text | sentiment |
|-----|---|-----------|
| 0 | Bulsyit Swasembada BBM. Tarif mobil online aja... | Negative |
| 1 | Rupiah melemah bukan cuma angka. Artinya: Semb... | Negative |
| 2 | Pengaruh BBM pada mesin : AFR & timing ign... | Negative |
| 3 | @tempodotco Fenomena aneh di Indonesia pemerin... | Negative |
| 4 | Update Harga BBM Pertamina Shell hingga Vivo N... | Neutral |
| ... | ... | ... |
| 406 | @mctraningrum Harga minyak bakal terus naik. ... | Negative |
| 407 | Kalau dunia sedang masuk fase perang minyak ma... | Negative |
| 408 | GM . Sabtu woles Minggu Curi Start. Senin? Tru... | Neutral |
| 409 | @geloraco Ha ha ha .. melipir mereka klo sdh b... | Negative |
| 410 | @arman_dhani Kek minyak dunia lgi naik Krn per... | Negative |

Gambar 3. Hasil klasifikasi sentimen

4.5. Analisis Hasil Penelitian

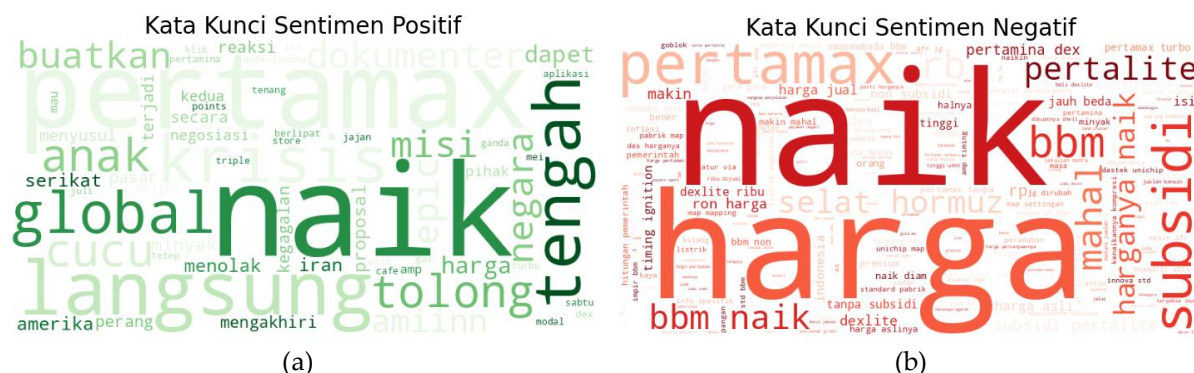
Hasil klasifikasi sentimen kemudian divisualisasikan untuk mempermudah proses analisis terhadap opini masyarakat mengenai kenaikan harga BBM non-subsidi akibat potensi penutupan Selat Hormuz. Berdasarkan hasil klasifikasi menggunakan model IndoBERT, diperoleh sentimen positif sebanyak 7 data (1,7%), sentimen netral sebanyak 200 data (48,5%), dan sentimen negatif sebanyak 205 data (49,8%).

Hasil tersebut menunjukkan bahwa sentimen negatif menjadi kategori yang paling dominan dibandingkan sentimen netral dan positif. Hal ini mengindikasikan bahwa sebagian besar masyarakat memberikan respons kekhawatiran terhadap potensi kenaikan harga BBM non-subsidi serta dampaknya terhadap kondisi ekonomi dan kebutuhan sehari-hari. Visualisasi distribusi sentimen dapat dilihat pada **Gambar 4**.



Gambar 4. Hasil distribusi klasifikasi sentimen
(a) Visualisasi dalam bentuk *bar chart*; (b) Visualisasi dalam bentuk *pie chart*

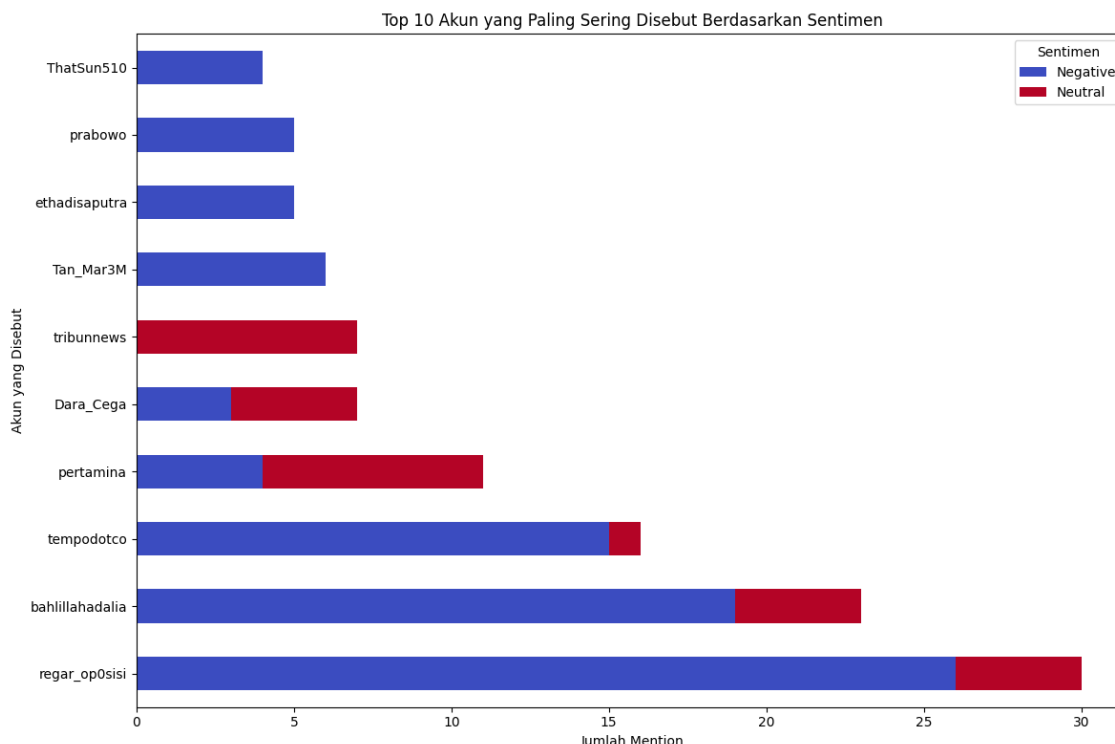
Selain visualisasi distribusi sentimen, penelitian ini juga menampilkan visualisasi *word cloud* untuk mengetahui kata-kata yang paling sering muncul pada data tweet masyarakat. Berdasarkan hasil *word cloud*, terdapat beberapa kata yang dominan muncul seperti “BBM”, “Pertamax”, “naik”, “harga”, dan “minyak”. Ukuran kata pada *word cloud* menunjukkan tingkat frekuensi kemunculan kata pada dataset. Hasil visualisasi *word cloud* dapat dilihat pada **Gambar 5**.



Gambar 5. Hasil perbandingan *word cloud* pengalaman positif dan negatif
(a) Visualisasi *word cloud* pengalaman yang positif; (b) Visualisasi *word cloud* pengalaman yang negatif

Penelitian ini juga menampilkan analisis top 10 akun yang paling sering disebut (*mentioned*) pada *tweet* berdasarkan kategori sentimen. Analisis dilakukan untuk mengetahui akun-akun yang paling banyak menjadi perhatian masyarakat dalam pembahasan isu kenaikan harga BBM non-subsidi akibat potensi penutupan Selat Hormuz.

Hasil analisis menunjukkan bahwa terdapat beberapa akun yang sering disebut oleh pengguna dalam *tweet* dengan sentimen positif, netral, maupun negatif. Perbedaan jumlah penyebutan pada masing-masing kategori sentimen menunjukkan adanya variasi respons masyarakat terhadap akun tertentu dalam diskusi terkait isu BBM non-subsidi. Visualisasi top 10 akun yang paling sering disebut berdasarkan sentimen dapat dilihat pada **Gambar 6**.



Gambar 6. Visualisasi top 10 akun yang paling sering disebut berdasarkan sentiment

Sebagai kesimpulan dari visualisasi hasil analisis, penelitian ini juga menampilkan word cloud keseluruhan untuk menggambarkan kata-kata yang paling dominan muncul pada seluruh data tweet. Visualisasi tersebut menunjukkan bahwa kata seperti “BBM”, “Pertamax”, “harga”, “naik”, dan “minyak” menjadi topik utama yang paling sering dibahas oleh masyarakat terkait isu kenaikan harga BBM non-subsidi akibat potensi penutupan Selat Hormuz. *Word cloud* keseluruhan dapat dilihat pada Gambar 7.



Gambar 7. Visualisasi *word cloud* dominan

5. Kesimpulan

Penelitian ini berhasil menerapkan model IndoBERT untuk melakukan klasifikasi sentimen masyarakat terhadap isu kenaikan harga BBM non-subsidi akibat potensi penutupan Selat Hormuz berdasarkan data opini pada platform X. Proses penelitian dilakukan melalui tahapan pengumpulan data, *pre-processing*, tokenisasi, klasifikasi sentimen, hingga visualisasi hasil analisis. Berdasarkan hasil klasifikasi terhadap 412 data *tweet*, diperoleh sentimen negatif sebesar 49,8%, sentimen netral sebesar 48,5%, dan sentimen positif sebesar 1,7%. Dominasi sentimen negatif menunjukkan bahwa mayoritas masyarakat merasa khawatir terhadap dampak ekonomi yang ditimbulkan oleh potensi kenaikan harga BBM non-subsidi, terutama terhadap biaya hidup dan stabilitas ekonomi rumah tangga. Selain itu, hasil *word cloud*

menunjukkan kata-kata dominan seperti “BBM”, “Pertamax”, “harga”, “naik”, dan “minyak” yang menjadi fokus utama pembahasan masyarakat. Dengan demikian, model IndoBERT terbukti mampu memahami konteks bahasa Indonesia pada media sosial dan memberikan hasil klasifikasi sentimen yang efektif terhadap opini publik. Penelitian selanjutnya diharapkan dapat menggunakan jumlah data yang lebih besar, menambahkan analisis temporal, serta membandingkan performa IndoBERT dengan model deep learning lainnya untuk memperoleh hasil yang lebih optimal.

Daftar Pustaka

- [1] A. J. Karso, “Iran-Israel Saling Serang Harga Minyak Mentah Merangkak Naik Dampaknya Terhadap Ekonomi Ri,” *SINERGI: Jurnal Riset Ilmiah*, vol. 2, no. 7, pp. 3073–3093, 2025, [Online]. Available: <https://doi.org/10.62335>
- [2] Badan Pusat Statistik Provinsi Jambi, “Harga BBM Terbang, Rakyat Tumbang,” <https://jambi.bps.go.id/id/news/2022/09/06/306/harga-bbm-terbang--rakyat-tumbang.html>.
- [3] I. Syapriani *et al.*, “Strategi Komunikasi Pengguna Media Sosial Dalam Mengelola Bahasa, Etika, Dan Resistansi Sosial Di Ruang Komentar Platform Digital X(Twitter),” *Jurnal Penelitian Nusantara*, vol. 1, pp. 360–368, Jun. 2025, doi: 10.59435/menulis.v1i6.348.
- [4] A. A. Nurrahman, M. Mauladi, and A. Rahman, “Analisis Sentimen Masyarakat terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Support Vector Machine dan SMOTE,” *sudo Jurnal Teknik Informatika*, vol. 4, no. 2, pp. 50–56, Jun. 2025, doi: 10.56211/sudo.v4i2.908.
- [5] N. P. D. Agustina, D. A. P. P. Tentriajaya, and I. G. N. L. Wijayakusuma, “Public Sentiment Analysis on Demonstration Actions Using IndoBERT Based on Transfer Learning,” 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [6] R. R. Cahya and A. Ichwani, “Komparasi Kinerja Model Support Vector Machine Dan Indonesian Bidirectional Encoder Representations From Transformers Dalam Klasifikasi Sentimen Ulasan Aplikasi Wondr By Bni,” 2026. [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [7] R. C. Tarumingkeng, *Natural Language Processing (NLP)*. Bogor: RUDYCT e-PRESS, 2024.
- [8] A. V. Sudiantoro and E. Zuliarso, “Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier,” *Dinamika Informatika*, vol. 10, no. 2, pp. 69–73, Oct. 2018.
- [9] D. K. Sumartha, “Implementasi Indobert Untuk Analisis Sentimen Opini Publik Terhadap Kebijakan Kenaikan Ukt Di Era Pemerintahan,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3S1, Oct. 2025, doi: 10.23960/jitet.v13i3S1.7880.