

Fine-Tuning Model IndoBERT untuk Analisis Sentimen Kenaikan Harga BBM Nonsubsidi pada Komentar Instagram

¹Hafizh Dzaky Ahya Gemilang, ²Ari Septiadi, ³Rama Dhitsaha Setya Negara, ⁴Iqrom Danang Putra, ⁵Pungkas Subarkah
^{1,2,3,4,5}Universitas Amikom Purwokerto
Banyumas, Indonesia

¹hdag304001@gmail.com, ²ariseptiadi323@gmail.com, ³ramadhitsaha98@gmail.com,
⁴iqromwaras441@gmail.com, ⁵subarkah@amikompurwokerto.ac.id

*Penulis Korespondensi

Diajukan : 04/07/2026

Diterima : 06/08/2026

Dipublikasi : 07/08/2026

ABSTRAK

Kenaikan harga Bahan Bakar Minyak (BBM) nonsubsidi memunculkan beragam tanggapan masyarakat yang banyak disampaikan melalui media sosial, termasuk Instagram. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap kenaikan harga BBM nonsubsidi menggunakan model IndoBERT. Data penelitian diperoleh melalui web scraping terhadap komentar pada dua unggahan Instagram Reels yang membahas kenaikan harga BBM nonsubsidi, sehingga diperoleh sebanyak 6.223 komentar. Data dipraproses, kemudian diberi label sentimen secara otomatis menggunakan Leksikon InSet ke dalam kategori positif, netral, dan negatif sebelum digunakan untuk proses fine-tuning model IndoBERT. Evaluasi model dilakukan menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Hasil pengujian menunjukkan model memperoleh accuracy sebesar 95,60%, dengan nilai precision, recall, dan F1-score yang tinggi pada seluruh kelas sentimen. Distribusi sentimen menunjukkan kategori netral mendominasi sebesar 35,78%, diikuti sentimen negatif sebesar 35,34% dan sentimen positif sebesar 28,88%. Hasil tersebut menunjukkan bahwa model IndoBERT mampu mempelajari pola label sentimen yang dihasilkan oleh Leksikon InSet pada dataset penelitian ini. Namun, karena pelabelan dilakukan secara otomatis dan data hanya berasal dari dua unggahan Instagram Reels, hasil penelitian ini memiliki keterbatasan dalam merepresentasikan sentimen masyarakat secara umum.

Kata Kunci: Analisis sentimen, BBM nonsubsidi, Fine-tuning, IndoBERT, Instagram

I. PENDAHULUAN

Ketegangan geopolitik antara Iran dan Israel sejak awal tahun 2024 menjadi ancaman terhadap stabilitas ekonomi global, khususnya pada sektor energi. Eskalasi konflik tersebut meningkatkan risiko gangguan pasokan minyak dunia dan mendorong kenaikan harga minyak mentah, yang berdampak pada meningkatnya biaya energi, biaya logistik, serta tekanan inflasi di berbagai negara, termasuk Indonesia (B. J. Karso, 2025). Kondisi tersebut turut memengaruhi harga Bahan Bakar Minyak (BBM) nonsubsidi yang dipengaruhi oleh fluktuasi harga minyak dunia (A. J. Karso, 2025).

Dampak konflik tidak hanya terbatas pada sektor energi, tetapi juga meluas pada aspek ekonomi global. Ketidakstabilan pasokan dan kenaikan harga energi memicu gangguan rantai pasok serta meningkatkan ketidakpastian ekonomi. Kondisi ini berpengaruh terhadap berbagai negara, termasuk Indonesia, yang rentan terhadap perubahan harga energi global (Nisa, 2025).

Dalam konteks tersebut, minyak bumi sebagai sumber energi utama memiliki peran penting dalam aktivitas ekonomi melalui produk turunannya berupa Bahan Bakar Minyak (BBM). BBM

digunakan secara luas pada sektor transportasi, industri, dan rumah tangga, sehingga perubahan harganya secara langsung memengaruhi stabilitas ekonomi Masyarakat .(Nurrahman dkk., 2025)

Di Indonesia, kenaikan harga BBM nonsubsidi dipengaruhi oleh fluktuasi harga minyak mentah dunia dan kebijakan energi pemerintah. Perubahan harga ini berdampak pada peningkatan biaya transportasi serta harga barang dan jasa, yang memicu inflasi dan menurunkan daya beli masyarakat. Kondisi tersebut menimbulkan berbagai respons publik terhadap kebijakan harga BBM (Putra dkk., 2025).

Kenaikan harga BBM nonsubsidi yang bersifat dinamis turut memengaruhi berbagai sektor ekonomi, baik secara langsung melalui peningkatan biaya operasional maupun secara tidak langsung melalui kenaikan biaya distribusi dan harga barang. Hal ini menunjukkan bahwa perubahan harga BBM nonsubsidi memiliki pengaruh signifikan terhadap aktivitas ekonomi dan menjadi isu penting dalam memahami dampaknya terhadap masyarakat (Riska dkk., 2024)

Instagram dipilih sebagai sumber data karena memiliki tingkat interaksi yang tinggi dan menjadi media bagi masyarakat untuk menyampaikan opini melalui kolom komentar. Oleh karena itu, komentar Instagram dapat dimanfaatkan sebagai sumber data dalam analisis sentimen terhadap respons masyarakat mengenai kenaikan harga BBM nonsubsidi (Alwi dkk., 2026). Dalam penelitian ini, data diperoleh dari komentar pada dua unggahan Instagram Reels yang membahas kenaikan harga BBM nonsubsidi. Kedua unggahan tersebut dipilih karena sesuai dengan fokus penelitian sehingga komentar yang diperoleh relevan untuk merepresentasikan respons masyarakat terhadap isu yang diteliti.

IndoBERT dipilih karena merupakan model berbasis Transformer yang dikembangkan khusus untuk bahasa Indonesia. Dibandingkan model Recurrent Neural Network (RNN), IndoBERT memiliki kemampuan yang lebih baik dalam memahami konteks semantik suatu kalimat, sehingga lebih efektif digunakan dalam tugas pemrosesan bahasa alami, termasuk analisis sentimen (Alfina & Muljono, 2026)

Penelitian terdahulu menunjukkan bahwa IndoBERT telah digunakan dalam berbagai penelitian analisis sentimen dan memberikan hasil yang baik. Namun, penelitian yang menganalisis sentimen masyarakat terhadap kenaikan harga BBM nonsubsidi menggunakan komentar Instagram Reels masih terbatas. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut dengan memanfaatkan data komentar Instagram serta pelabelan sentimen menggunakan Leksikon InSet sebelum proses *fine-tuning* IndoBERT.

Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kenaikan harga BBM nonsubsidi berdasarkan komentar Instagram menggunakan model fine-tuning IndoBERT. Penelitian ini diharapkan dapat memberikan gambaran mengenai sentimen masyarakat serta menjadi referensi bagi penerapan fine-tuning IndoBERT pada analisis sentimen media sosial dan penelitian selanjutnya.

II. STUDI LITERATUR

Penelitian Terdahulu

Penelitian oleh (Manoppo dkk., 2025) menganalisis sentimen publik terhadap kenaikan PPN 12% menggunakan model IndoBERT pada 2.581 data yang dikumpulkan dari platform X, Instagram, dan TikTok. Hasil penelitian menunjukkan model memperoleh accuracy sebesar 84,94% dengan dominasi sentimen negatif. Tujuan tersebut menunjukkan bahwa model IndoBERT memiliki performa yang baik dalam analisis sentimen terhadap isu kebijakan publik.

Penelitian oleh (Al-Kadzim dkk., 2025) menerapkan model IndoBERT untuk menganalisis sentimen publik terhadap konflik Palestina–Israel di media sosial X. Model memperoleh accuracy, precision, recall, dan F1-score masing-masing sebesar 0,73. Hasil penelitian menunjukkan bahwa IndoBERT memiliki performa yang baik dalam menganalisis sentimen pada isu publik di media sosial.

Penelitian oleh (Sumartha, 2025) mengimplementasikan model IndoBERT untuk menganalisis sentimen masyarakat terhadap kebijakan kenaikan UKT menggunakan 10.140 data hasil *web scraping* dari media sosial X. Hasil penelitian menunjukkan bahwa mayoritas opini masyarakat memiliki sentimen negatif. Temuan tersebut menunjukkan bahwa model IndoBERT memiliki performa yang baik dalam analisis sentimen terhadap isu kebijakan publik di media sosial.

Berdasarkan uraian sebelumnya, perbandingan penelitian terdahulu yang relevan dengan penelitian ini dapat dilihat pada Tabel 1.

Peneliti	Objek Penelitian	Metode	Dataset	Hasil Penelitian
Manoppo dkk., (2025)	Analisis sentimen publik terhadap kenaikan PPN 12%	IndoBERT	2.581 data dari X, Instagram, dan TikTok	Model memperoleh <i>accuracy</i> sebesar 84,94% dengan dominasi sentimen negatif, menunjukkan performa yang baik dalam analisis sentimen isu kebijakan publik.
Al-Kadzim dkk., (2025)	Analisis sentimen publik terhadap konflik Palestina-Israel	IndoBERT	Data dari media sosial X	Model memperoleh <i>accuracy</i> , <i>precision</i> , <i>recall</i> , dan <i>F1-score</i> masing-masing sebesar 0,73, serta mampu menganalisis sentimen masyarakat terhadap isu publik di media sosial.
Sumartha, (2025)	Analisis sentimen masyarakat terhadap kebijakan kenaikan UKT	IndoBERT	10.140 data hasil web scraping dari media sosial X	Mayoritas opini masyarakat memiliki sentimen negatif, menunjukkan performa yang baik dalam analisis sentimen terhadap isu kebijakan publik.
Penelitian Ini	Analisis sentimen masyarakat terhadap kenaikan harga BBM nonsubsidi pada komentar Instagram Reels	Leksikon InSet + IndoBERT	6.223 komentar Instagram Reels	Mengkaji sentimen masyarakat terhadap kenaikan harga BBM nonsubsidi dengan memanfaatkan pelabelan menggunakan Leksikon InSet sebelum proses <i>fine-tuning</i> model IndoBERT.

Berdasarkan uraian dan perbandingan penelitian terdahulu pada Tabel 1, dapat disimpulkan bahwa model IndoBERT telah banyak diterapkan dalam analisis sentimen pada berbagai isu publik di media sosial dengan hasil yang baik. Penelitian-penelitian tersebut menunjukkan bahwa IndoBERT memiliki potensi dalam mengklasifikasikan sentimen masyarakat terhadap berbagai topik, sehingga menjadi dasar yang relevan dalam penelitian ini.

BBM Nonsubsidi

Bahan Bakar Minyak (BBM) merupakan bahan bakar yang berasal dari hasil pengolahan minyak mentah dan digunakan sebagai sumber energi di berbagai sektor. Salah satu jenisnya adalah BBM nonsubsidi, yaitu bahan bakar yang tidak mendapatkan subsidi dari pemerintah sehingga harga jualnya ditentukan berdasarkan mekanisme pasar dan kebijakan badan usaha penyedia. Saat ini, distribusi BBM nonsubsidi tidak hanya dilakukan oleh Pertamina, tetapi juga oleh badan usaha swasta yang telah memperoleh izin, sehingga menciptakan persaingan dalam penyediaan BBM di Indonesia (Tambunan dkk., 2022).

Instagram

Instagram merupakan salah satu platform media sosial berbasis visual yang digunakan untuk berbagi foto, video, dan berbagai informasi. Platform ini menyediakan fitur interaktif, seperti unggahan foto dan video, filter, efek, serta kolom komentar yang memudahkan pengguna untuk berkomunikasi dan berbagi konten. Selain itu, Instagram banyak digunakan oleh kalangan muda dan mampu menjangkau pengguna secara luas, sehingga menjadi salah satu media yang efektif untuk menyebarkan informasi dan menyampaikan opini (Niken dkk., 2024).

Analisis sentimen

Analisis sentimen merupakan salah satu bidang dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi dan mengelompokkan opini atau emosi yang terkandung dalam suatu teks ke dalam kategori positif, netral, atau negatif. Analisis sentimen banyak dimanfaatkan untuk mengetahui pandangan masyarakat terhadap suatu topik berdasarkan data teks, sehingga dapat memberikan informasi yang bermanfaat sebagai bahan evaluasi dan pengambilan keputusan (Subarkah dkk., 2025).

Natural language processing (NLP)

Natural Language Processing (NLP) merupakan cabang dari *Artificial Intelligence* (AI) yang berfokus pada pengembangan model, sistem, dan algoritma agar komputer dapat memahami, mengolah, serta menganalisis bahasa manusia. NLP banyak diterapkan dalam berbagai tugas, seperti penerjemahan bahasa, peringkasan teks, *question answering*, dan analisis sentimen (Nuryadi dkk., 2025).

Bert

Bidirectional Encoder Representations from Transformers (BERT) merupakan model bahasa berbasis *Transformer* yang dikembangkan oleh *Google Research*. BERT mampu memahami konteks kata secara dua arah (*bidirectional*) dengan mempertimbangkan hubungan antar kata sebelum dan sesudahnya dalam sebuah kalimat. Kemampuan tersebut membuat BERT menghasilkan representasi bahasa yang lebih akurat sehingga banyak digunakan pada berbagai tugas *Natural Language Processing* (NLP) (Subarkah dkk., 2024).

IndoBERT

IndoBERT merupakan model BERT yang dikembangkan khusus menggunakan korpus bahasa Indonesia sehingga lebih mampu memahami karakteristik linguistik bahasa Indonesia dibandingkan model BERT standar (A'la, 2025). Selain itu, IndoBERT memiliki kemampuan memahami konteks semantik dua arah sehingga sesuai untuk menangani karakteristik bahasa pada media sosial, seperti penggunaan bahasa informal, singkatan, dan fenomena *code-mixing* yang umum ditemukan dalam komunikasi digital (Hakim dkk., 2026). Meskipun demikian, model berbasis *transformer* seperti IndoBERT masih memiliki keterbatasan dalam menangkap sinyal pragmatis yang tidak selalu diekspresikan secara eksplisit dalam bentuk leksikal, sehingga pemanfaatan fitur linguistik pragmatis dapat menjadi pelengkap untuk meningkatkan pemahaman terhadap makna implisit pada teks media sosial (Perdana dkk., 2026).

InSet Lexicon

InSet Lexicon (*Indonesian Sentiment Lexicon*) merupakan kamus leksikon berbahasa Indonesia yang digunakan dalam analisis sentimen. Setiap kata memiliki bobot polaritas dengan rentang nilai -5 hingga $+5$ yang digunakan untuk menghitung skor sentimen sehingga teks dapat diklasifikasikan ke dalam kategori positif, negatif, atau netral (Hasyati dkk., 2025).

Meskipun mudah diterapkan dan tidak memerlukan proses pelatihan model, InSet Lexicon memiliki keterbatasan dalam menangani makna kata yang bergantung pada konteks kalimat, termasuk nuansa kontekstual dan ekspresi non-literal (Rizkia dkk., 2025).

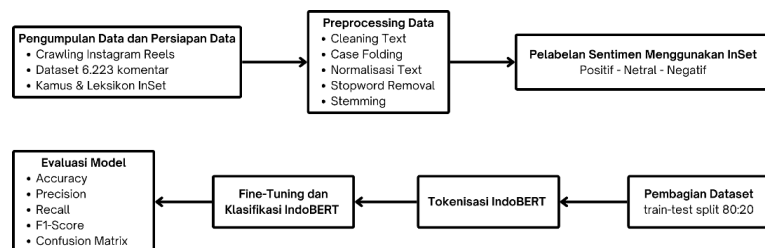
Fine-tuning

Fine-tuning merupakan proses penyesuaian model BERT yang telah melalui tahap pre-training agar dapat digunakan secara optimal pada tugas atau dataset tertentu. Pada tahap ini, model memanfaatkan pengetahuan yang telah dipelajari sebelumnya, kemudian disesuaikan dengan karakteristik data penelitian sehingga mampu meningkatkan performa dalam menyelesaikan tugas yang spesifik (Sofiana, 2024).

III. METODE

Penelitian ini dilakukan melalui beberapa tahapan, yaitu pengumpulan data komentar Instagram terkait kenaikan harga BBM nonsubsidi, prapemrosesan teks, pelabelan sentimen menggunakan leksikon InSet, fine-tuning model IndoBERT, dan evaluasi kinerja model.

Rangkaian tahapan penelitian tersebut disajikan dalam bentuk diagram alir pada Gambar 1.



Gambar 1. Metode penelitian

Pengumpulan Data

Pada penelitian ini, proses pengumpulan data dilakukan dari beberapa sumber. Adapun sumber data yang digunakan adalah sebagai berikut:

Dataset Komentar Instagram (*Scraping*)

Data komentar Instagram dikumpulkan menggunakan metode *web scraping* berbasis GraphQL yang diimplementasikan menggunakan Python dengan pustaka *httpx* dan autentikasi berbasis *cookies*. Proses *scraping* dilakukan pada tanggal 25 Mei 2026 terhadap dua unggahan Instagram Reels yang membahas kenaikan harga BBM nonsubsidi. Komentar diambil secara bertahap menggunakan mekanisme *pagination* GraphQL dan disimpan dalam format CSV. Hasil pengumpulan data dari kedua unggahan tersebut menghasilkan sebanyak 6.223 komentar yang digunakan sebagai dataset penelitian.

Seluruh komentar yang berhasil diperoleh dari proses *scraping* digunakan sebagai dataset penelitian tanpa dilakukan seleksi berdasarkan isi komentar pada tahap pengumpulan data. Tahap *preprocessing* selanjutnya dilakukan untuk membersihkan data sebelum proses pelabelan sentimen.

Leksikon sentimen InSet

Leksikon sentimen InSet digunakan sebagai acuan dalam pelabelan sentimen karena memuat kosakata bahasa Indonesia yang telah diklasifikasikan ke dalam kategori positif dan negatif. Leksikon ini membantu menentukan kecenderungan sentimen pada setiap komentar secara lebih konsisten. Pada penelitian ini, setiap komentar dipecah menjadi token, kemudian setiap token dicocokkan dengan kamus InSet untuk memperoleh bobot sentimen. Skor sentimen dihitung berdasarkan akumulasi bobot seluruh token pada komentar. Selanjutnya, komentar dengan skor lebih besar dari nol diberi label positif, skor lebih kecil dari nol diberi label negatif, sedangkan skor sama dengan nol diberi label netral.

Colloquial Indonesian Lexicon

Kamus normalisasi bahasa Indonesia tidak baku (*Colloquial Indonesian Lexicon*) digunakan untuk menormalisasi kata-kata tidak baku atau *slang* menjadi bentuk baku bahasa Indonesia. Proses normalisasi dilakukan sebelum pelabelan sentimen agar variasi penulisan kata dapat diseragamkan sehingga meningkatkan kecocokan kata dengan kamus InSet Lexicon dan mempermudah proses analisis pada tahap berikutnya.

Preprocessing Data

Tahap *preprocessing* dilakukan untuk membersihkan dan menyiapkan data teks sebelum proses analisis. Tahapan yang dilakukan meliputi *cleaning text*, *case folding*, normalisasi kata menggunakan kamus bahasa tidak baku, *stopword removal*, dan *stemming* menggunakan Sastrawi. Hasil dari proses ini berupa data yang lebih bersih dan terstruktur sehingga siap digunakan pada tahap pelabelan sentimen dan pemodelan (Khairani dkk., 2024).

Pelabelan Sentimen

Pelabelan sentimen dilakukan pada data yang telah melalui tahap *preprocessing*. Setiap komentar diberi skor berdasarkan bobot kata pada leksikon InSet, kemudian diklasifikasikan ke

dalam kategori positif, negatif, atau netral. Hasil pelabelan selanjutnya digunakan sebagai data latih pada proses pemodelan (Pratama dkk., 2025).

Pembagian Data

Dataset dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% menggunakan rasio 80:20. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk mengevaluasi performa model. Proses pelatihan menggunakan batch size sebesar 8 untuk mendukung proses fine-tuning model IndoBERT (Arsi dkk., 2025).

Tokenisasi IndoBERT

Data latih dan data uji kemudian ditokenisasi menggunakan tokenizer IndoBERT agar dapat diproses oleh model. Selanjutnya diterapkan padding dan truncation untuk menyeragamkan panjang data masukan sesuai kebutuhan model.

Fine-Tuning dan Klasifikasi Sentimen

Model IndoBERT yang telah melalui tahap *pre-training* kemudian di-*fine-tune* menggunakan data latih yang telah diberi label sentimen. Sebelum proses pelatihan, data ditokenisasi menggunakan *tokenizer* bawaan IndoBERT agar dapat diproses oleh model. Setelah proses *fine-tuning* selesai, model digunakan untuk mengklasifikasikan komentar ke dalam kategori sentimen positif, negatif, dan netral.

Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja IndoBERT dalam mengklasifikasikan sentimen komentar Instagram terkait kenaikan harga BBM nonsubsidi. Pengukuran performa dilakukan menggunakan accuracy, precision, recall, dan F1-score untuk menilai tingkat ketepatan klasifikasi pada setiap kategori sentimen. Selain itu, digunakan confusion matrix untuk memberikan gambaran mengenai hasil prediksi model terhadap data aktual, sehingga dapat diketahui kemampuan model dalam membedakan sentimen positif, negatif, dan netral.

IV. HASIL DAN PEMBAHASAN

Pengumpulan Data

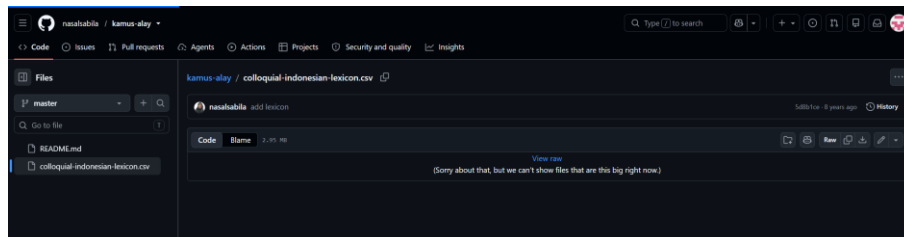
Proses pengumpulan data pada penelitian ini dilakukan dengan teknik *web scraping* terhadap komentar pengguna Instagram pada unggahan yang membahas kenaikan harga BBM nonsubsidi. Proses scraping diimplementasikan menggunakan bahasa pemrograman Python dengan memanfaatkan GraphQL serta autentikasi berbasis *cookies* untuk memperoleh data komentar secara bertahap (*pagination*). Hasil scraping menghasilkan sebanyak 6.223 komentar yang kemudian disimpan dalam format CSV. Dataset hasil *scraping* memuat beberapa atribut, namun untuk menjaga privasi pengguna hanya data komentar (*text*). Contoh dataset hasil scraping ditampilkan pada Tabel 1.

Tabel 1. Hasil Crawl Komentar Instagram

Text
Makanya shock banget liat harga BBM non subsidi up nya 🤔. Selain doa , gerak hemat Energi Hemat BBM, yuk bisaa mulai perjalanan yang prioritas aja.
Terlihat sepele tapi fatal ya, fenomena global sengaruh itu sama kondisi di Indonesia, meski Pertamina udah nahan harga biar ga naik, tetep sih kita juga kudu do something, misal kurangi pake kendaraan pribadi dan beralih ke kendaraan umum dan jalan kaki kalau jaraknya dekat. yuk lebih hemat pakai bbm.
Komoditas apapun itu, polanya sama. Kalo pasokan udah terganggu, sedangkan permintaan tetap dan malah tidak berkurang sama sekali. Maka udah bisa dipastikan bahwa harga BBM bakalan melonjak. Harga BBM karena faktor global. Yuk lebih hemat pakai BBM 🤔🤔🤔🤔
Banyak banget kenaikan harganya, yuk lebih hemat pakai BBM.

Akhirnya tercerahkan dengan informasi ini 🤔

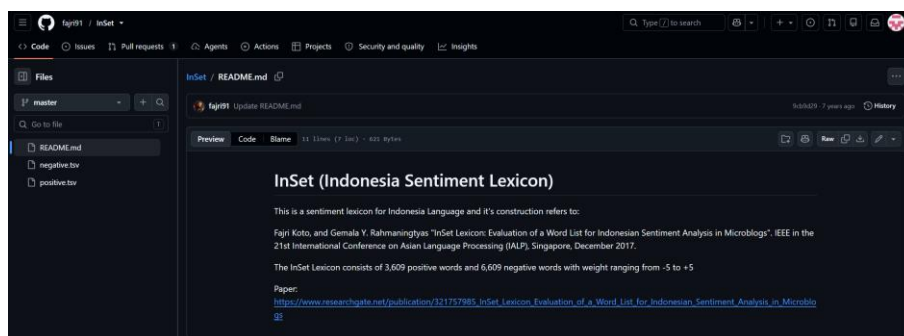
Berdasarkan Tabel 1, komentar yang diperoleh masih mengandung berbagai karakteristik bahasa media sosial, seperti penggunaan kata tidak baku, singkatan, bahasa gaul (*slang*), emoji, serta variasi penulisan yang tidak konsisten. Oleh karena itu, sebelum dilakukan proses preprocessing, penelitian ini memanfaatkan Colloquial Indonesian Lexicon sebagai kamus normalisasi untuk mengubah kata-kata tidak baku menjadi bentuk baku sehingga data menjadi lebih seragam dan siap diproses pada tahapan selanjutnya. Contoh isi kamus normalisasi yang digunakan ditampilkan pada Gambar 2.



Gambar 2. Github *colloquial indonesian lexicon*

Kamus normalisasi tersebut digunakan untuk memetakan kata-kata tidak baku menjadi bentuk baku sesuai kaidah bahasa Indonesia. Proses ini bertujuan mengurangi variasi penulisan pada data komentar sehingga dapat meningkatkan kualitas data sebelum dilakukan pelabelan sentimen.

Selanjutnya, penelitian ini menggunakan Leksikon Sentimen InSet sebagai acuan dalam proses pelabelan sentimen. Leksikon InSet merupakan kumpulan kosakata bahasa Indonesia yang telah diklasifikasikan ke dalam kategori sentimen positif dan negatif beserta nilai bobotnya. Leksikon ini digunakan untuk menentukan kecenderungan sentimen pada setiap komentar secara konsisten sebelum data digunakan dalam proses pelatihan model IndoBERT. Contoh isi leksikon InSet yang digunakan pada penelitian ini ditunjukkan pada Gambar 3.



Gambar 3. Github InSet

Preprocessing

Tahap preprocessing dilakukan untuk membersihkan dan menyeragamkan data komentar sebelum digunakan pada proses pelabelan sentimen dan pelatihan model IndoBERT. Tahapan yang dilakukan meliputi *cleaning text*, *case folding*, normalisasi menggunakan Colloquial Indonesian Lexicon, *stopword removal*, dan *stemming*. Contoh hasil perubahan data pada setiap tahapan ditampilkan pada Tabel 2.

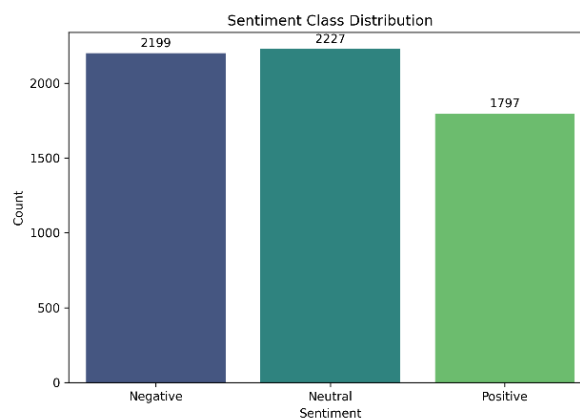
Tabel 2. Preprocessing

Original Text	Makanya shock banget liat harga BBM non subsidi up nya 🤔. Selain doa , gerak hemat Energi Hemat BBM, yuk bisaa mulai perjalanan yang prioritas aja.
Case Folding	makanya shock banget liat harga bbm non subsidi up nya 🤔.

	selain doa , gerak hemat energi hemat bbm, yuk bisaa mulai perjalanan yang prioritas aja.
Cleaning	makanya shock banget liat harga bbm non subsidi up nya selain doa gerak hemat energi hemat bbm yuk bisaa mulai perjalanan yang prioritas aja
Normalization	makanya shock banget lihat harga bbm non subsidi up nya selain doa gerak hemat energi hemat bbm yuk bisa mulai perjalanan yang prioritas saja
Stopword Removal	shock banget harga bbm non subsidi up doa gerak hemat energi hemat bbm yuk perjalanan prioritas
Stemming	shock banget harga bbm non subsidi up doa gerak hemat energi hemat bbm yuk jalan prioritas

Pelabelan Sentimen

Setelah melalui tahap preprocessing, setiap komentar diberi label sentimen menggunakan Leksikon InSet ke dalam tiga kategori, yaitu positif, negatif, dan netral. Distribusi hasil pelabelan ditunjukkan pada Gambar 4, yang menunjukkan bahwa sentimen netral memiliki jumlah data terbanyak, yaitu 2.227 komentar, diikuti sentimen negatif sebanyak 2.199 komentar, dan sentimen positif sebanyak 1.797 komentar. Hasil tersebut menunjukkan bahwa distribusi data pada setiap kelas relatif seimbang sehingga dapat digunakan sebagai data latih dalam proses fine-tuning model IndoBERT dan membantu mengurangi potensi bias model terhadap kelas sentimen tertentu.

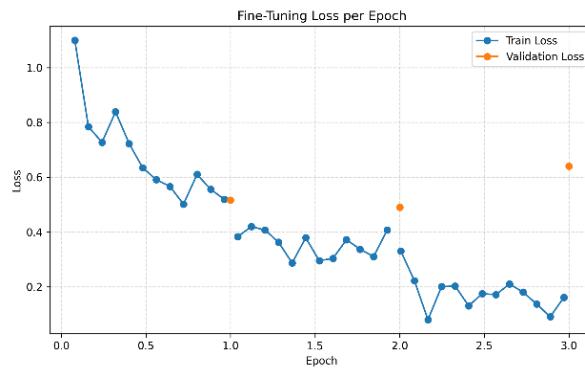


Gambar 4. Grafik Distribusi Sentimen

Fine-Tuning IndoBERT

Proses *fine-tuning* dilakukan menggunakan model IndoBERT Large (indobenchmark/indobert-large-p2) pada dataset hasil pelabelan sentimen yang telah dibagi menjadi data latih dan data uji dengan rasio 80:20. Sebelum proses pelatihan, data ditokenisasi menggunakan tokenizer IndoBERT dengan panjang maksimum 128 token. Parameter pelatihan yang digunakan meliputi 3 *epoch*, *batch size* sebesar 8, *weight decay* sebesar 0,01, dan *warmup ratio* sebesar 0,1.

Hasil proses *fine-tuning* ditunjukkan pada Gambar 5. Grafik menunjukkan bahwa nilai training loss mengalami penurunan selama proses pelatihan, yang mengindikasikan bahwa model semakin mampu mempelajari pola pada data latih. Sementara itu, nilai *validation loss* yang diperoleh dari proses evaluasi pada setiap akhir *epoch* mengalami perubahan antar *epoch*, dengan nilai terendah diperoleh pada *epoch* kedua sebelum meningkat pada *epoch* ketiga. Grafik tersebut menunjukkan perkembangan proses pelatihan model selama *fine-tuning*.



Gambar 5. Grafik performa epoch

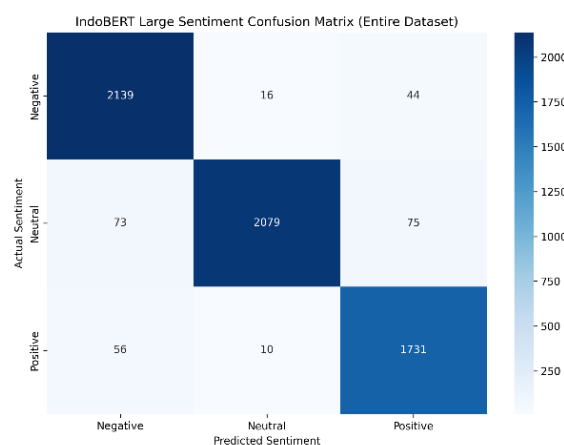
Evaluasi Model

Evaluasi model dilakukan untuk mengetahui performa IndoBERT dalam mengklasifikasikan sentimen komentar Instagram mengenai kenaikan harga BBM nonsubsidi. Pengujian dilakukan menggunakan data uji (*testing data*) yang telah dipisahkan pada tahap pembagian dataset. Performa model dievaluasi menggunakan beberapa metrik, yaitu accuracy, precision, recall, F1-score, dan confusion matrix.

Tabel 3. Tabel Classification Report

Kelas	Precision (%)	Recall (%)	F1-Score (%)	Support
Negatif	94,30	97,30	95,80	2.199
Netral	98,80	93,40	96,00	2.227
Positif	93,60	96,30	94,90	1.797
Accuracy			95,60	6.223
Macro Average	95,60	95,70	95,60	6.223
Weighted Average	95,70	95,60	95,60	6.223

Berdasarkan Tabel 3, model IndoBERT memperoleh accuracy sebesar 95,60%, yang menunjukkan bahwa model memiliki tingkat ketepatan yang sangat baik dalam mengklasifikasikan sentimen komentar Instagram mengenai kenaikan harga BBM nonsubsidi. Selain itu, nilai weighted average untuk precision, recall, dan F1-score masing-masing sebesar 95,70%, 95,60%, dan 95,60%, sehingga menunjukkan bahwa performa model relatif konsisten pada seluruh kelas sentimen.



Gambar 6. Confusion Matrix

Berdasarkan Gambar 6, sebagian besar data berhasil diklasifikasikan dengan benar, yaitu 2.139 data sentimen negatif, 2.079 data sentimen netral, dan 1.731 data sentimen positif. Kesalahan klasifikasi antar kelas relatif sedikit, sehingga menunjukkan bahwa model IndoBERT mampu membedakan ketiga kategori sentimen dengan baik.

V. KESIMPULAN

Penelitian ini berhasil menerapkan model IndoBERT untuk menganalisis sentimen komentar Instagram mengenai kenaikan harga BBM nonsubsidi. Data penelitian diperoleh melalui web scraping terhadap dua unggahan Instagram Reels sehingga diperoleh sebanyak 6.223 komentar. Hasil evaluasi menunjukkan bahwa model memperoleh accuracy sebesar 95,60% dengan nilai precision, recall, dan F1-score yang tinggi pada seluruh kategori sentimen. Distribusi sentimen didominasi oleh sentimen netral sebesar 35,78%, diikuti sentimen negatif sebesar 35,34% dan sentimen positif sebesar 28,88%. Hasil tersebut menunjukkan bahwa model IndoBERT memiliki kinerja yang baik dalam mengklasifikasikan sentimen pada dataset penelitian ini. Penelitian ini memiliki keterbatasan karena data hanya berasal dari dua unggahan Instagram Reels dan pelabelan sentimen dilakukan secara otomatis menggunakan Leksikon InSet tanpa validasi manual. Penelitian selanjutnya disarankan menggunakan data yang lebih beragam, melakukan validasi manual terhadap pelabelan sentimen, serta membandingkan performa IndoBERT dengan model klasifikasi lainnya.

VI. REFERENSI

- A'la, F. Y. (2025). Optimasi Klasifikasi Sentimen Ulasan Game Berbahasa Indonesia: IndoBERT dan SMOTE untuk Menangani Ketidakseimbangan Kelas. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 256–265. <https://doi.org/10.29408/edumatic.v9i1.29666>
- Alfina, L. M., & Muljono. (2026). Optimalisasi Akurasi dan Stabilitas Analisis Sentimen Ulasan E-Commerce Indonesia melalui Fine-Tuning Transformer IndoBERT. *Infotekmesin*, 17(01), 10–17. <https://doi.org/10.35970/infotekmesin.v17i1.3037>
- Al-Kadzim, M. G., Rasim, R., & Herbert, H. (2025). Analisis Perubahan Sentimen Publik di Media Sosial X terhadap Konflik Palestina-Israel Menggunakan Model IndoBERT. *Digital Transformation Technology*, 4(2), 1167–1174. <https://doi.org/10.47709/digitech.v4i2.5312>
- Alwi, H., Julkarnain, M., & Oktavia, S. A. (2026). Klasifikasi Sentimen Publik di Instagram Menggunakan Indobert Terhadap Respons DPR Pasca-Aksi '17+8 Tuntutan Rakyat. *Digital Transformation Technology*, 6(1), 34–40. <https://doi.org/10.47709/digitech.v6i1.7909>
- Arsi, P., Firmanda, R. A., Prayoga, I., & Subarkah, P. (2025). Opinion Mining on Spotify Music App Reviews Using Bidirectional LSTM and BERT. *Jurnal Informatika*, 12(2), 68–74. <https://doi.org/10.31294/inf.v12i2.25323>
- Hakim, L. N., Hana, F. M., & Wahyudin, W. C. (2026). Klasifikasi Komentar Toksik Berbahasa Indonesia di Media Sosial Berbasis Fine-Tuning IndoBERT. *JURIKOM (Jurnal Riset Komputer)*, 13(1), 202–209. <https://doi.org/10.30865/jurikom.v13i1.9449>
- Hasyiyati, Z., Qausar, H., Yumni, H., & Jannah, M. (2025). Ekonomi Aset Digital: Analisis Sentimen Masyarakat Berbasis Leksikon Terhadap Kebijakan Kripto di Indonesia. *TIN: Terapan Informatika Nusantara*, 6(1), 1–14. <https://doi.org/10.47065/tin.v6i1.7508>
- Karso, A. J. (2025). IRAN-ISRAEL SALING SERANG HARGA MINYAK MENTAH MERANGKAK NAIK DAMPAKNYA TERHADAP EKONOMI RI. *SINERGI: Jurnal Riset Ilmiah*, 2(7), 3073–3093. <https://doi.org/https://doi.org/10.62335>
- Karso, B. J. (2025). Perang Israel-Iran, Antisipasi Dampak Terhadap Perekonomian Indonesia. *Jurnal Ilmiah Multidisipin*, 3(7), 520–528. <https://doi.org/10.60126/jim.v3i7.1151>
- Khairani, U., Mutiawani, V., & Ahmadian, H. (2024). Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(4), 887–894. <https://doi.org/10.25126/jtiik.1148315>
- Manoppo, M. R., Kolang, I. C., Nur Fiat, D. N., Mawara, R. M. C., Sumarno, A. D. P., Yusupa, A., & Tarigan, V. (2025). ANALISIS SENTIMEN PUBLIK DI MEDIA SOSIAL TERHADAP KENAIKAN PPN 12% DI INDONESIA MENGGUNAKAN INDOBERT. *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, 4(2), 152–163. <https://doi.org/10.69916/jkbti.v4i2.322>
- Niken, I. P., Hamzah, A., & Fatkhiyah, E. (2024). ANALISIS SENTIMEN PADA MEDIA SOSIAL INSTAGRAM TERHADAP PERANG ISRAEL VS HAMAS MENGGUNAKAN METODE NAÏVE BAYES DAN SVM. *Jurnal SCRIPT*, 12(1), 30–38.

- Nisa, A. Z. C. (2025). Konflik Iran-Israel: Analisis Faktor-Faktor yang Mempengaruhi Dinamika Hubungan dan Dampaknya. *Syntax Literate*, 10(7), 8557–8568. <https://doi.org/https://doi.org/10.36418/syntax-literate.v10i7.60741>
- Nurrahman, A. A., Mauladi, M., & Rahman, A. (2025). Analisis Sentimen Masyarakat terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Support Vector Machine dan SMOTE. *sudo Jurnal Teknik Informatika*, 4(2), 50–56. <https://doi.org/10.56211/sudo.v4i2.908>
- Nuryadi, D., Metandi, F., Hadiwijaya, N. A., Rohman, M. Z., Hartanto, S., Syafrizal, A., & Yadie, E. (2025). FINE TUNING INDOBERT UNTUK ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI TIKET.COM DI GOOGLE PLAY STORE. *Jurnal Mahasiswa Teknik Informatika*, 9(2), 3577–3583. <https://doi.org/https://doi.org/10.36040/jati.v9i2.13204>
- Perdana, A., Sugandi, F., & Raihannicko, A. (2026). Evaluasi Lintas Domain Deteksi Sarkasme Bahasa Indonesia: Pendekatan Hybrid IndoBERT dan Fitur Pragmatis. *Jurnal Edukasi dan Penelitian Informatika*, 12(1), 75–82. <https://doi.org/https://doi.org/10.26418/jp.v12i1.102465>
- Pratama, M. R., Handayani, M. R., Yuniarti, W. D., & Umam, K. (2025). MODEL PREDIKSI SENTIMEN ULASAN MOBILE LEGENDS DI GOOGLE PLAY STORE DAN YOUTUBE BERBASIS PELABELAN OTOMATIS ROBERTA DAN KLASIFIKASI RANDOM FOREST. *JSiI | Jurnal Sistem Informasi*, 12(2), 39–45.
- Putra, K. A., Husein, Z., Putri, S. A., & Pangestoeti, W. (2025). Jurnal Riset Multidisiplin Edukasi DAMPAK KENAIKAN HARGA BBM TERHADAP INFLASI DAN DAYA BELI MASYARAKAT DI PROVINSI KEPULAUAN RIAU. *JURMIE*, 2(6), 1064–1078. <https://doi.org/https://doi.org/10.71282/jurmie.v2i6.592>
- Riska, T. J., Suraji, A., & Ophiyandri, T. (2024). Dampak Kenaikan Harga Solar Non Subsidi Terhadap Harga Satuan Pekerjaan. *RRJ*, 6(6), 2330–2333. <https://doi.org/10.38035/rrj.v6i6>
- Rizkia, A. S., Wufron, W., & Roji, F. F. (2025). Analisis Sentimen Coretax: Perbandingan Pelabelan Data Manual, Transformers-Based, dan Lexicon-Based pada Performa IndoBERT. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(3). <https://doi.org/10.57152/malcom.v5i3.2151>
- Sofiana, S. (2024). *KONSEP BERT PADA NATURAL LANGUAGE PROCESSING*. Eureka Media Aksara.
- Subarkah, P., Ikhsan, A. N., Anggraeni, E., & Sabaniyah, A. P. (2025). Sentiment Perspective of Government's Free Nutritious Meal Policy on Social Media X using Indo-BERT and Bi-LTSM. *Journal of Technology and Informatics (JoTI)*, 7(2), 201–210. <https://doi.org/10.37802/joti.v7i2.1065>
- Subarkah, P., Rozaq, H. A. A., Arsi, P., Sholikhatin, S. A., Riyanto, R., & Marcos, H. (2024). Implementation of Text Mining to Detect Emotions of Fuel Price Increase using BERT-LSTM Methods. *Gazi University Journal of Science*, 37(4), 1707–1716. <https://doi.org/10.35378/gujs.1424742>
- Sumartha, D. K. (2025). IMPLEMENTASI INDOBERT UNTUK ANALISIS SENTIMEN OPINI PUBLIK TERHADAP KEBIJAKAN KENAIKAN UKT DI ERA PEMERINTAHAN. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3S1), 867–875. <https://doi.org/10.23960/jitet.v13i3S1.7880>
- Tambunan, N., Aprilia, S., & Rahayu, N. P. (2022). STUDY LITERATURE: DAMPAK KENAIKAN BBM BAGI PEREKONOMIAN RAKYAT. *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*, 2(1), 329–336. <https://doi.org/10.54443/sibatik.v2i1.550>