



Analisis sentimen terhadap pemerintahan Prabowo–Gibran menggunakan IndoBERT dan LDA

Sahrial Ihsani Ishak¹, Okma Arnilia², Tri Widodo³, I Gusti Nyoman Agung Bisma Tatwa⁴

¹Universitas Dian Nusantara, Bekasi, Indonesia

²Universitas Islam Negeri Siber Syekh Nurjati Cirebon, Cirebon, Indonesia

³Universitas Teknokrat Indonesia, Lampung, Indonesia

⁴Institut Pertanian Bogor, Bogor, Indonesia

Riwayat Artikel:

Diterima 15 Oktober 2025

Direvisi 18 Oktober 2025

Disetujui 28 Oktober 2025

Kata Kunci:

Analisis sentimen

IndoBERT

LDA

Media online

Prabowo–Gibran

ABSTRACT. This study analyzes public perception of the Prabowo–Gibran administration through online news coverage using a Natural Language Processing (NLP) approach. Data were collected from credible news portals, including Antara News and Detik, totaling 195 articles. The analysis was conducted in two stages: first, IndoBERT was used to classify news sentiment into positive, negative, and neutral; second, Latent Dirichlet Allocation (LDA) was applied to identify the main topics driving coverage. Sentiment results showed that most topics, particularly those related to the economy, public policy, and governance, were dominated by negative sentiment (80%), while positive sentiment accounted for 15.9% and neutral sentiment for 4.1%. These findings indicate public criticism and concern regarding the effectiveness of policies and economic stability. The combined IndoBERT and LDA approach proved effective in providing a comprehensive understanding of public opinion dynamics in the digital era. It can serve as a consideration for the government in formulating more responsive and transparent communication strategies.

ABSTRAK. Penelitian ini menganalisis persepsi publik terhadap kepemimpinan Prabowo–Gibran melalui pemberitaan media online menggunakan pendekatan Natural Language Processing (NLP). Data dikumpulkan dari portal berita kredibel seperti Antara News dan Detik dengan total 195 artikel. Analisis dilakukan dalam dua tahap: pertama, IndoBERT digunakan untuk mengklasifikasikan sentimen berita menjadi positif, negatif, dan netral; kedua, *Latent Dirichlet Allocation* (LDA) diterapkan untuk mengidentifikasi topik utama yang mendominasi pemberitaan. Hasil sentimen menunjukkan bahwa sebagian besar topik, terutama terkait ekonomi, kebijakan publik, dan pemerintahan, didominasi oleh sentimen negatif (80%), sedangkan sentimen positif tercatat 15,9% dan netral 4,1%. Temuan ini mengindikasikan adanya kritik dan keprihatinan publik terhadap efektivitas kebijakan dan stabilitas ekonomi. Hasil menunjukkan bahwa sebagian besar topik, terutama terkait ekonomi, kebijakan publik, dan pemerintahan, didominasi oleh sentimen negatif. Temuan ini mengindikasikan adanya kritik dan keprihatinan publik terhadap efektivitas kebijakan dan stabilitas ekonomi. Pendekatan kombinatif IndoBERT dan LDA terbukti efektif dalam memberikan pemahaman komprehensif mengenai dinamika opini publik di era digital, serta dapat menjadi bahan pertimbangan bagi pemerintah dalam merumuskan strategi komunikasi yang lebih responsif dan transparan.

This is an open-access article under the CC-BY-SA license.



Penulis Korespondensi:

Sahrial Ihsani Ishak,

Universitas Dian Nusantara,

Jl. Mini 3, Kelurahan Bambu Apus, Cipayung, Jakarta Timur, 13890, Indonesia.

Email: sahrial.ihsani.ishak@dosen.undira.ac.id

PENDAHULUAN

Perkembangan teknologi informasi yang masif telah mengubah cara masyarakat dalam memperoleh dan menafsirkan suatu informasi (Feldman, 2013). Salah satu platform yang menjadi sumber utama

informasi adalah media online (Liu, 2015). Saat ini, media online berperan dalam membentuk persepsi masyarakat terhadap kebijakan politik (Feldman, 2013). Pasca demonstrasi yang dilakukan oleh berbagai elemen masyarakat di akhir Agustus yang lalu, kepemimpinan Prabowo menjadi sorotan masyarakat. Pemberitaan media berpotensi untuk memengaruhi citra dan opini publik secara signifikan (Liu, 2015).

Analisis sentimen merupakan alat yang tepat dalam memahami bagaimana persepsi publik tersebut terbentuk dari pemberitaan media online (Liu, 2015). Setiap berita dapat dikategorikan sebagai positif, negatif, dan netral yang dapat mengidentifikasi kecenderungan opini publik terhadap pemerintahan Prabowo–Gibran. Selain itu, analisis tidak hanya memberikan gambaran kuantitatif tentang persepsi publik, tetapi juga bisa membantu mencari faktor-faktor apa saja yang memengaruhi opini (Liu, 2015).

Natural Language Processing (NLP) dapat melakukan analisis tersebut secara efektif yang mampu memproses dan memahami teks secara otomatis (Jurafsky & Martin, 2025). NLP memungkinkan identifikasi sentimen dari berbagai pemberitaan, sehingga pola opini publik dapat dianalisis secara lebih sistematis (Feldman, 2013). Dengan bantuan NLP, peneliti dapat menangkap konteks kata, hubungan antarfrasa, serta nuansa bahasa yang sering kali sulit diinterpretasikan secara manual (Jurafsky & Martin, 2025).

Perkembangan *Natural Language Processing* (NLP) modern telah menghadirkan model berbasis *deep learning*, seperti BERT (*Bidirectional Encoder Representations from Transformers*) dan varian khusus untuk bahasa Indonesia, yaitu IndoBERT (Devlin *et al.*, 2019; Koto *et al.*, 2020). Berbeda dengan metode berbasis leksikon yang mengandalkan daftar kata sentimen, IndoBERT mampu memahami konteks kata dalam kalimat secara lebih mendalam, termasuk menangkap ironi, sarkasme, dan penggunaan kata yang ambigu (Devlin *et al.*, 2019). Hal ini menjadikan analisis sentimen lebih akurat dan relevan, terutama untuk pemberitaan politik yang kompleks (Mandhasiya *et al.*, 2022).

Studi oleh Manoppo *et al.* (2025) menunjukkan bahwa IndoBERT dapat mencapai akurasi hingga 84,94% dalam menganalisis sentimen publik terhadap kebijakan kenaikan PPN 12% di Indonesia, dengan distribusi sentimen dominan negatif. Penelitian ini mengumpulkan 2.581 sampel data dari platform media sosial seperti X (Twitter), Instagram, dan TikTok, dan menunjukkan efektivitas IndoBERT dalam tugas klasifikasi sentimen berbahasa Indonesia pada data media sosial.

Selain itu, penelitian oleh (Ramadhan *et al.*, 2025) membandingkan kinerja model *Convolutional Neural Network* (CNN) dan IndoBERT dalam mengklasifikasikan sentimen berita politik berbahasa Indonesia. Hasilnya menunjukkan bahwa IndoBERT unggul dengan akurasi 92,93%, terutama dalam klasifikasi sentimen netral, sementara CNN memiliki akurasi 89,13% namun kesulitan dalam mengklasifikasikan sentimen netral.

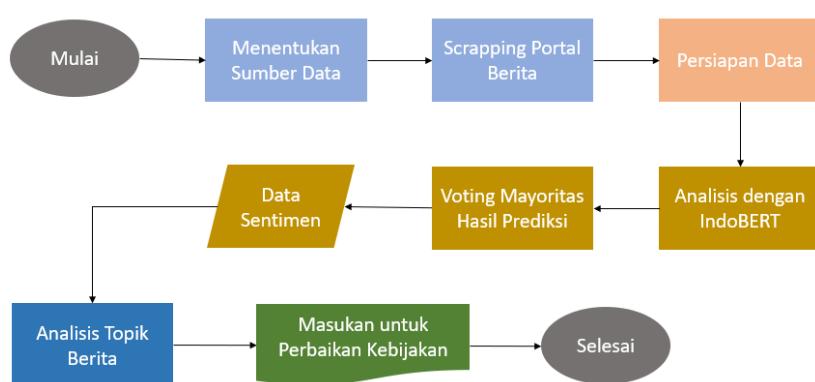
Penelitian terdahulu juga menunjukkan potensi penggabungan analisis sentimen berbasis IndoBERT dengan *topic modeling* seperti *Latent Dirichlet Allocation* (LDA) untuk memahami isu utama yang muncul dalam wacana publik (Naury *et al.*, 2021; Puspita *et al.*, 2024; Rakhmawati *et al.*, 2024). Pendekatan ini memungkinkan identifikasi topik utama seperti ekonomi, kebijakan publik, sosial, dan pendidikan yang beririsan dengan persepsi publik terhadap kinerja pemerintah.

Berdasarkan uraian tersebut, penelitian ini diarahkan untuk menganalisis persepsi publik terhadap kepemimpinan Prabowo–Gibran melalui media online dengan menggunakan pendekatan *Natural Language Processing* (NLP). Pada tahap pertama, IndoBERT digunakan untuk memprediksi atau mengklasifikasikan sentimen setiap berita menjadi kategori positif, negatif, atau netral. Penggunaan IndoBERT dipilih karena spesifikasinya pada korpus bahasa Indonesia, sehingga memungkinkan model memahami struktur dan pola dalam suatu sentimen (Syafutra & Kusrini, 2025). Selanjutnya, hasil klasifikasi tersebut dianalisis menggunakan *Latent Dirichlet Allocation* (LDA) untuk mengklasterkan, meringkas dan menghubungkan data dalam jumlah data yang nantinya dapat

digunakan dalam analisis topik dalam sentimen. Pendekatan kombinatif antara IndoBERT dan LDA ini diharapkan dapat memberikan wawasan yang lebih komprehensif mengenai dinamika opini publik di era digital (Mandhasiya *et al.*, 2022).

METODE

Penelitian yang dilakukan dapat dilihat pada Gambar 1. Penelitian ini dimulai dengan menentukan sumber data yaitu dari portal berita Tempo, Detik dan Antara News yang merupakan tiga portal berita di Indonesia yang bereputasi dan faktual. Berita yang ada dilakukan *scraping* untuk tahun 2025 menggunakan *library* BeautifulSoup di Python. Setelah itu data yang ada mengalami proses pembersihan data sehingga menghasilkan teks bersih yang siap dianalisis. Tahapan selanjutnya adalah menganalisis sentimen menggunakan model pretrained IndoBERT. Terakhir adalah tahapan analisis per topik menggunakan *Latent Dirichlet Allocation* (LDA) yang diperoleh dari berita sehingga hasilnya dapat digunakan sebagai bahan evaluasi terhadap fokus pemberitaan dan opini publik terkait kepemimpinan Prabowo-Gibran. Berikut adalah tahapan lengkap dari penelitian.



Gambar 1. Diagram alir penelitian

a. Sumber Data

Penelitian ini menggunakan berita online dari tiga portal berita utama di Indonesia, yaitu Antara, Detik, dan Tempo. Data yang dikumpulkan meliputi judul berita, url, isi konten, dan portal sumber. Pemilihan portal berita ini didasarkan pada reputasi dan cakupan informasi yang luas (Dewan Pers, 2025), sehingga data yang diperoleh dapat mewakili opini publik terhadap kepemimpinan Prabowo-Gibran.

b. Pengumpulan Data

Pengumpulan data dilakukan melalui proses *web scraping* menggunakan Python dengan *library* BeautifulSoup. Setiap berita yang relevan dikumpulkan dan disimpan dalam format *DataFrame*, dengan kolom title, url, content, dan portal. Kemudian, hasil proses *scraping* tidak selalu sesuai dengan yang diharapkan, karena beberapa konten ada yang tidak dapat diakses, sehingga membuat konten kosong atau tidak lengkap (Foerderer, 2023; Ulloa *et al.*, 2025). Oleh karena itu, diperlukan proses pembersihan data untuk menghapus baris-baris yang berisi nilai kosong agar data yang digunakan dalam analisis tetap valid dan representatif.

c. Persiapan Data

Tahapan persiapan data dimulai dengan pembersihan teks (*text cleaning*) untuk menghapus karakter non-alfabet, angka, dan simbol. Selanjutnya dilakukan *case folding* untuk mengubah semua teks menjadi huruf kecil, serta *stopword removal* menggunakan daftar *stopword* dalam bahasa Indonesia seperti "yang", "dan", "di", "ke", "dari", "untuk" dan lain sebagainya. Setiap

kata kemudian diubah menjadi bentuk dasarnya melalui *stemming* menggunakan *library* Sastrawi. Persiapan data ini bertujuan untuk menghasilkan teks bersih yang siap dianalisis.

d. Analisis Sentimen

Analisis sentimen dilakukan menggunakan model IndoBERT, yaitu versi BERT yang telah dilatih khusus untuk bahasa Indonesia dan berbasis *deep learning*. Teks berita yang telah dibersihkan melalui proses *preprocessing* diinput ke model untuk diklasifikasikan menjadi sentimen positif, negatif, atau netral. Model IndoBERT memiliki batas maksimal input 512 token per batch, oleh karena itu teks yang lebih panjang dibagi menjadi beberapa *chunk* dengan panjang maksimal 512 token per *chunk* atau potongan. Pendekatan ini memungkinkan model menangkap konteks kata secara *bidirectional*, sehingga lebih efektif dalam menganalisis nuansa bahasa Indonesia yang kompleks, termasuk kata ambigu, ironi, atau sarkasme, dibandingkan metode berbasis leksikon tradisional (Devlin *et al.*, 2019; Koto *et al.*, 2020). IndoBERT disini digunakan untuk memprediksi sentimen terhadap konten berita.

e. Analisis Topik

Setelah teks berita dibersihkan dan diklasifikasikan menggunakan model IndoBERT menjadi tiga kategori sentimen (positif, negatif, dan netral), analisis topik dilakukan secara terpisah untuk masing-masing kelompok sentimen. Pendekatan ini bertujuan untuk mengidentifikasi tema dominan yang muncul dalam setiap jenis sentimen, sehingga memberikan pemahaman yang lebih mendalam mengenai topik apa yang paling sering dikaitkan dengan opini publik positif, negatif, maupun netral terhadap suatu isu (Blei *et al.*, 2003). Setiap topik terdiri dari beberapa kata dominan dengan bobot kontribusi tertinggi, yang kemudian dijadikan dasar untuk pemberian label topik secara manual sesuai konteks berita. Topik yang dibahas yaitu tentang Ekonomi, Gizi, Pendidikan, Korupsi Pejabat, Peristiwa, Kegiatan Nasional, Transportasi, Kriminalitas, Lingkungan dan Konflik Global.

HASIL DAN DISKUSI

Pengambilan Data

Penelitian ini menggunakan pustaka BeautifulSoup dari bahasa pemrograman Python untuk melakukan proses *web scraping* terhadap portal berita daring. Pengumpulan data dilakukan pada tiga portal berita nasional yang memiliki kredibilitas tinggi, yaitu Antara, Detik, dan Tempo, dengan periode pengambilan data dari Januari hingga Oktober 2025.

Proses pengumpulan data difokuskan pada berita yang relevan dengan isu pemerintahan baru Indonesia. Relevansi data ditentukan berdasarkan pencarian menggunakan kata kunci “Pemerintahan Prabowo-Gibran” yang dimasukkan pada kolom pencarian masing-masing portal. Pemilihan kata kunci ini didasarkan pada fokus penelitian, yaitu analisis opini publik dan kecenderungan pemberitaan terkait pemerintahan baru. Penerapan kriteria ini memastikan bahwa hanya artikel yang berkaitan langsung dengan konteks penelitian yang dikumpulkan. Proses scraping menghasilkan 300 artikel berita dari tiga portal berita, sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Total artikel yang berhasil diperoleh

No.	Portal Berita	Jumlah Artikel
1	Antara	100 Artikel
2	Detik	100 Artikel
3	Tempo	100 Artikel
Total		300 Artikel

Dari hasil *scraping* awal, ditemukan bahwa sebagian data memiliki nilai kosong (*null*) pada kolom judul atau konten berita. Nilai null ini muncul akibat ketidakkonsistenan struktur HTML antar portal, seperti perbedaan tag *<div>* atau *<p>* yang menyimpan isi artikel, berita yang telah dihapus oleh portal, maupun halaman yang tidak dapat diakses penuh oleh sistem *scraping*.

Persiapan Data

Tahapan ini merupakan proses pembersihan dan normalisasi teks (*text preprocessing*) sebelum analisis dilakukan. Proses dimulai dengan pembuatan *stemmer* menggunakan pustaka Sastrawi yang berfungsi untuk mengubah setiap kata menjadi bentuk dasarnya agar memiliki keseragaman makna. Selanjutnya, dilakukan pemanggilan daftar *stopword* dari file CSV yang berisi kata-kata umum dalam bahasa Indonesia yang tidak memiliki kontribusi penting terhadap konteks analisis. Daftar *stopword* tersebut kemudian diperluas dengan *stopword* tambahan secara manual, seperti kata penghubung dan kata ganti (contohnya: yang, dengan, untuk, ini, itu, dan sebagainya). Selain itu, ditambahkan pula *stopword* khusus media online seperti “*editor*”, “*reporter*”, “*kompas*”, “*detik*”, “*antara*” dan nama penulis tertentu, untuk memastikan teks yang dianalisis murni berisi isi berita, bukan informasi teknis atau metadata media. Hasil sebelum dibersihkan dapat dilihat pada Tabel 2.

Tabel 2. Artikel sebelum dibersihkan

No.	Portal Berita	Judul	Konten
1	Detik	Ruang Perbaikan untuk Program Makan Bergizi Gr..	Membayangkan sebuah negara di mana semua anak ...
2	Detik	Detik-detik Negeri 'Tongkat Malaikat' Mengayun ...	Tawuran bersenjata 'tongkat malaikat'di Pebayu...
3	Detik	Catat! Daftar Hari Libur Nasional dan Cuti Ber...	Pemerintah menetapkan libur nasional dan cuti ...
...
195	Antara	Kementerian ATR dapat pinjaman Rp5,72 triliun ...	Jakarta (ANTARA) - Kementerian Agraria dan Tat...

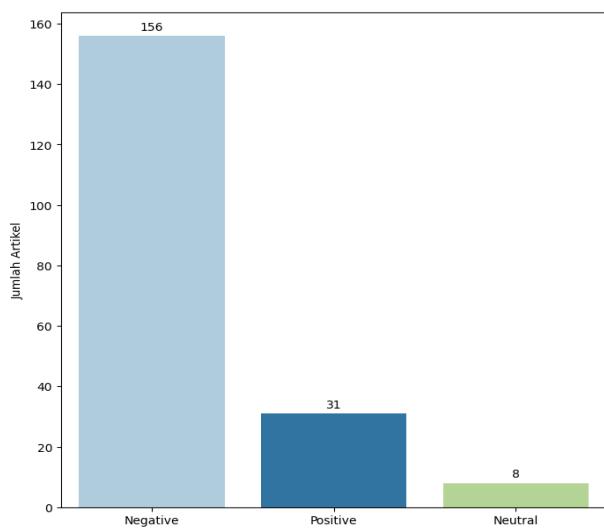
Setelah daftar *stopword* disiapkan seperti kata 'yang', 'dengan', 'untuk', 'ini', selanjutnya dilakukan fungsi pembersihan teks yang mencakup penghapusan footer berita, tautan, angka, tanda baca, serta spasi berlebih. Teks yang tersisa kemudian dipecah menjadi token (kata-kata terpisah), dihapus kata-kata yang termasuk dalam daftar *stopword*, dan akhirnya dilakukan stemming agar setiap kata berubah menjadi bentuk dasar. Tabel 3 menunjukkan sampel data hasil persiapan data. Hasilnya akan dianalisis menggunakan IndoBERT. Data dari portal Tempo tidak dapat dibersihkan karena meskipun URL tersedia, konten beritanya tidak ada atau gagal diambil saat scraping, sehingga teks untuk analisis sentimen tidak tersedia.

Analisis Sentimen

Hasil analisis sentimen terhadap keseluruhan artikel berita menunjukkan bahwa pemberitaan didominasi oleh sentimen negatif (Gambar 2). Berdasarkan distribusi data, terdapat 156 artikel dengan sentimen negatif, 31 artikel dengan sentimen positif, dan hanya 8 artikel dengan sentimen netral. Dominasi berita negatif ini menunjukkan bahwa sebagian besar konten media cenderung menyoroti isu-isu yang bersifat kritis, seperti kebijakan pemerintah, dinamika politik, maupun permasalahan sosial yang sedang terjadi. Porsi besar sentimen negatif juga mengindikasikan bahwa pemberitaan media lebih berfokus pada aspek kontroversial yang memiliki daya tarik tinggi bagi pembaca, sejalan dengan karakteristik media daring yang sering menonjolkan isu-isu sensitif untuk meningkatkan interaksi publik.

Tabel 3. Hasil pembersihan data

Portal Berita	Judul	Konten	Hasil Pembersihan
Detik	Ruang Perbaikan untuk Program Makan Bergizi Gratis	Membayangkan sebuah negara di mana semua anak bisa menikmati makanan bergizi...	bayang negara anak nikmat makan gizi gratis program makan gizi gratis gagas prabowo subianto ...
Detik	Detik-detik Negeri 'Tongkat Malaikat' Mengayun hingga Jatuh Korban Jiwa	Tawuran bersenjata 'tongkat malaikat' di Pebayuran...	tawur senjata tongkat malaikat bayur kabupaten bekasi makan korban jiwa remaja tewas akibat sabet ...
Detik	Catat! Daftar Hari Libur Nasional dan Cuti Bersama 2025	Pemerintah menetapkan libur nasional dan cuti bersama tahun 2025 sebanyak 27 hari...	perintah tetap libur nasional cuti daftar libur nasional cuti tetap surat putus skb menteri ...

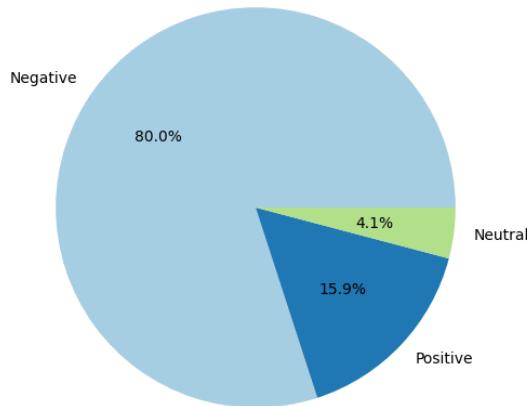


Gambar 2. Visualisasi hasil sentimen menggunakan IndoBERT

Sementara itu, berita dengan sentimen positif hanya mencakup sebagian kecil dari keseluruhan data. Artikel-artikel positif biasanya berisi laporan mengenai capaian kebijakan, program sosial, atau kegiatan yang memberikan dampak baik bagi masyarakat. Namun, rendahnya proporsi berita positif memperlihatkan bahwa narasi pembangunan dan keberhasilan pemerintah belum terlalu mendapat porsi pemberitaan yang seimbang. Di sisi lain, artikel dengan sentimen netral berjumlah paling sedikit, hanya sekitar 4 persen dari total berita. Konten dalam kategori ini umumnya bersifat informatif dan objektif, seperti pengumuman resmi atau laporan administratif yang tidak menunjukkan kecenderungan emosional tertentu. Proporsi jumlah sentimen dapat dilihat pada Gambar 3.

Berdasarkan visualisasi *Word Cloud* (Gambar 4) untuk masing-masing sentimen, dapat dilihat bahwa sentimen positif didominasi kata-kata yang berkaitan dengan pembangunan, ekonomi, dan dukungan terhadap program pemerintah seperti "BUMN", "negara", "tingkat", "kerja", dan "saham", menunjukkan bahwa sebagian berita memfokuskan pada capaian atau aspek positif pemerintahan. Sentimen netral cenderung menampilkan kata-kata umum atau informatif seperti "program", "ekonomi", "sekolah", "makan", dan "tahun", yang menunjukkan berita bersifat faktual tanpa penilaian emosional yang kuat. Sementara itu, sentimen negatif didominasi kata-kata seperti

“pendapatan”, “masyarakat”, dan “korban” yang mengindikasikan perhatian terhadap isu masalah, keluhan publik, atau kontroversi, sehingga berita dengan sentimen ini menyoroti dampak negatif atau kritik terhadap kebijakan dan situasi tertentu. Secara keseluruhan, pola kata ini merepresentasikan kecenderungan model untuk membedakan konten berita berdasarkan nuansa positif, netral, dan negatif di korpus berita yang dianalisis.

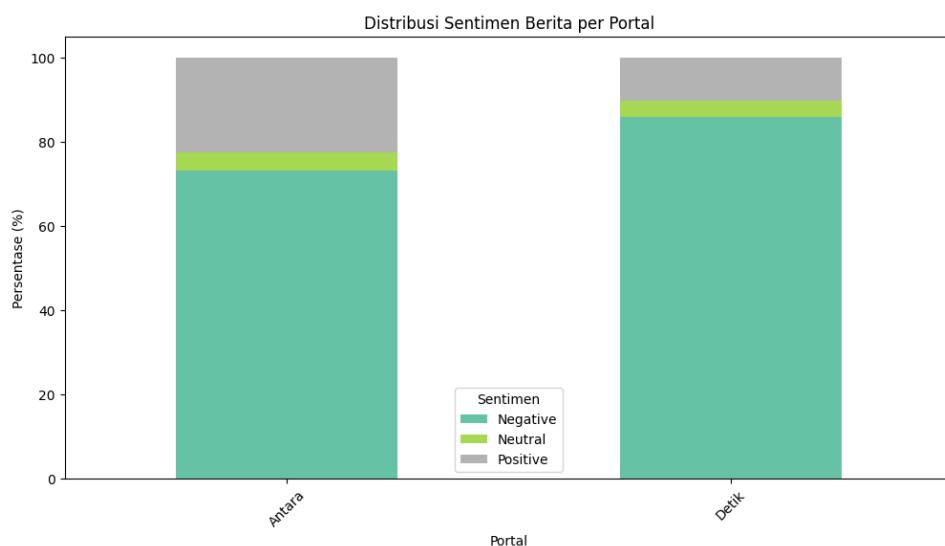


Gambar 3. Visualisasi hasil persentase sentimen berita



Gambar 4. *Word Cloud* sentimen berita

Berdasarkan analisis sentimen terhadap berita dari dua portal, terlihat bahwa mayoritas berita di portal Antara bersifat negatif sebesar 73,03%, diikuti sentimen positif 22,47% dan netral 4,49%. Sementara itu, portal Detik memiliki proporsi berita negatif yang lebih tinggi, yakni 85,85%, dengan sentimen positif hanya 10,38% dan netral 3,77%. Hal ini menunjukkan bahwa kedua portal cenderung menghadirkan berita dengan sentimen negatif sebagai dominan, meskipun portal Antara relatif lebih seimbang dengan proporsi positif yang lebih besar dibandingkan Detik (Gambar 5).



Gambar 5. Word Cloud sentimen berita

Analisis Topik Pembahasan

Hasil analisis sentimen terhadap kumpulan berita yang dikumpulkan dari berbagai portal kemudian dikelompokkan lebih lanjut berdasarkan tema atau topik utama pembahasan. Pengelompokan ini dilakukan menggunakan metode *Latent Dirichlet Allocation* (LDA), yaitu salah satu teknik pemodelan topik (*topic modeling*) yang secara otomatis mengidentifikasi pola kata dalam teks untuk menemukan kelompok isu yang memiliki kemiripan konteks semantik.

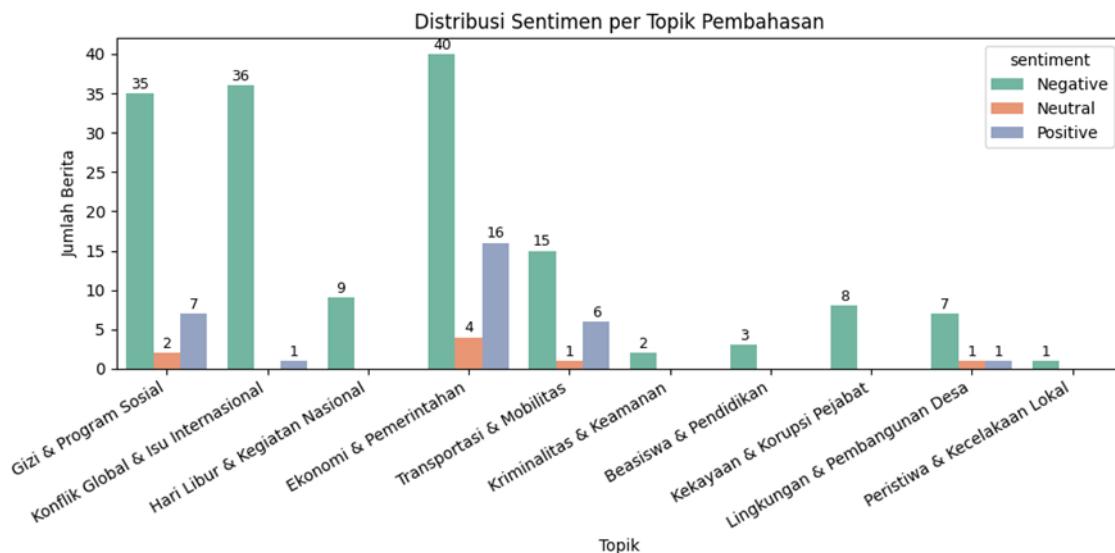
Dalam penelitian ini, model LDA diterapkan pada data teks yang telah melalui tahap pra-pemrosesan (seperti pembersihan teks, penghapusan *stopword*, dan pembentukan n-gram). Sampel hasil pengolahan n-gram dapat dilihat pada Tabel 4. Parameter jumlah topik ($k = 10$) ditentukan untuk menghasilkan sepuluh kelompok tema utama berita yang paling representatif. Setiap topik terdiri dari sejumlah kata dominan yang menjadi ciri khas isu yang dibahas di dalamnya. Gambar 4 menunjukkan analisis sentimen per topik pembahasan.

Tabel 4. Total artikel yang berhasil diperoleh

No.	N-Gram	Frequency
1	rp triliun	75
2	makan gizi	65
3	gencat senjata	15
4	korban tewas	16

Berdasarkan hasil analisis distribusi sentimen per topik pemberitaan, terlihat adanya variasi persepsi publik yang cukup tajam terhadap berbagai isu nasional. Topik “Ekonomi & Pemerintahan” mencatat jumlah berita negatif tertinggi (40 artikel), mencerminkan dominasi narasi kritis terhadap kebijakan ekonomi, aktivitas pemerintahan, serta dinamika politik. Hal ini menunjukkan bahwa masyarakat memiliki tingkat sensitivitas tinggi terhadap isu ekonomi dan kinerja pemerintah, terutama terkait kebijakan fiskal, kenaikan harga, dan stabilitas politik (Gambar 6). Topik “Gizi & Program Sosial” juga menunjukkan dominasi sentimen negatif (35 artikel) meski disertai beberapa narasi positif, seperti liputan mengenai program makan bergizi gratis. Temuan ini mengindikasikan adanya apresiasi terhadap kebijakan pro-rakyat, namun juga menunjukkan bahwa implementasi program sosial masih menghadapi kritik dalam hal efektivitas dan pemerataan manfaat. Di sisi lain, topik

“Konflik Global & Isu Internasional” didominasi sentimen negatif (36 artikel), menggambarkan keprihatinan publik terhadap isu kemanusiaan dan geopolitik global. Sementara itu, “Transportasi & Mobilitas” menunjukkan keseimbangan antara berita positif (16 artikel) dan negatif (15 artikel), menandakan bahwa sektor ini dipersepsi relatif stabil dan memiliki kemajuan yang diakui publik.



Gambar 6. Distribusi sepuluh topik utama terhadap kategori sentimen berita.

Berdasarkan hasil ini, pemerintahan Prabowo–Gibran perlu memperkuat komunikasi publik yang transparan, responsif, dan berbasis solusi, terutama dalam bidang ekonomi dan kebijakan sosial yang menjadi fokus perhatian masyarakat. Pemerintah dapat mengoptimalkan strategi komunikasi digital dan media engagement untuk menyeimbangkan persepsi publik terhadap kebijakan yang dijalankan. Selain itu, diperlukan evaluasi terhadap pelaksanaan program sosial agar manfaatnya lebih merata dan tepat sasaran. Dalam konteks isu global dan keamanan, penting bagi pemerintah untuk menampilkan peran diplomasi yang aktif dan empatik, guna menumbuhkan citra positif di mata publik. Sementara pada sektor transportasi, momentum persepsi positif sebaiknya dijaga melalui percepatan pembangunan infrastruktur dan pelayanan publik yang lebih inklusif. Pendekatan berbasis transparansi, partisipasi masyarakat, dan konsistensi kebijakan akan membantu meningkatkan kepercayaan publik terhadap pemerintahan yang baru.

KESIMPULAN

Penelitian ini menunjukkan bahwa analisis sentimen dan topik berita media online menggunakan pendekatan *Natural Language Processing* (NLP) dengan model IndoBERT serta metode *Latent Dirichlet Allocation* (LDA) berhasil mengelompokkan berita ke dalam sepuluh topik utama, seperti ekonomi, pemerintahan, sosial, pendidikan, dan konflik global. Hasil analisis memperlihatkan sebagian besar topik didominasi oleh sentimen negatif, terutama pada isu ekonomi, kebijakan publik, dan pemerintahan, yang mencerminkan kritik dan keprihatinan publik terhadap kinerja pemerintahan Prabowo–Gibran. Berdasarkan temuan tersebut, pemerintah diharapkan dapat menggunakan hasil analisis ini sebagai bahan evaluasi untuk memperkuat komunikasi publik, meningkatkan transparansi, dan menyampaikan narasi kebijakan yang lebih jelas dan faktual. Penelitian ini memiliki keterbatasan, yakni cakupan data hanya dari tiga portal berita utama di Indonesia (Antara, Detik, dan Tempo) dan periode pengumpulan data terbatas selama 10 bulan, sehingga representasi opini publik secara keseluruhan mungkin tidak sepenuhnya mencerminkan dinamika di media lain atau periode waktu lebih panjang.

Untuk penelitian lanjutan, disarankan agar cakupan data diperluas dengan melibatkan lebih banyak portal berita, media alternatif, dan platform media sosial untuk memperoleh representasi opini publik yang lebih lengkap. Selain itu, penelitian selanjutnya dapat mengeksplorasi *fine-tuning* model IndoBERT menggunakan dataset sentimen politik Indonesia yang lebih besar dan beragam, sehingga akurasi klasifikasi sentimen dapat meningkat, khususnya untuk menangkap nuansa bahasa kompleks seperti ironi, sarkasme, atau ambiguitas dalam pemberitaan politik. Pendekatan kombinatif antara analisis sentimen berbasis IndoBERT dan *topic modeling* seperti LDA tetap relevan untuk memberikan wawasan komprehensif mengenai opini publik dan dinamika topik utama di era digital.

REFERENSI

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data. *Journal of Machine Learning Research*, 3, 993–1022. <https://doi.org/10.1016/B978-0-12-411519-4.00006-9>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm), 4171–4186.
- Dewan Pers. (2025). *Daftar Perusahaan Pers Terverifikasi Administrasi dan Faktual*. Diakses pada 27 Oktober 2025.
- Feldman, R. (2013). Techniques and applications for sentiment analysis. *Commun. ACM*, 56(4), 82–89. <https://doi.org/10.1145/2436256.2436274>
- Foerderer, J. (2023, August 7). *Should we trust web-scraped data?* Working paper, Technical University of Munich. <https://doi.org/10.48550/arXiv.2308.02231>.
- Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* (3rd ed.). <https://web.stanford.edu/~jurafsky/slp3>. Diakses pada 27 Oktober 2025.
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, 757–770. <https://doi.org/10.18653/v1/2020.coling-main.66>
- Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press. <https://doi.org/DOI: 10.1017/CBO9781139084789>
- Mandhasiya, D. G., Murfi, H., Bustamam, A., & Anki, P. (2022). Evaluation of Machine Learning Performance Based on BERT Data Representation with LSTM Model to Conduct Sentiment Analysis in Indonesian for Predicting Voices of Social Media Users in the 2024 Indonesia Presidential Election. *2022 5th International Conference on Information and Communications Technology (ICOIACT)*, 441–446. <https://doi.org/10.1109/ICOIACT55506.2022.9972206>
- Manoppo, M. R., Kolang, I. C., Fiat, D. N., Michelly, R., & Mawara, C. (2025). *Analisis Sentimen Publik Di Media Sosial Terhadap Kenaikan Ppn 12 % Di Indonesia Menggunakan Indobert Analysis of Public Sentiment on Social Media Towards the 12 % Ppn Increase in Indonesia Using Indobert*. 4(2), 152–163. <https://repository.uinjkt.ac.id/dspace/handle/123456789/76873>. Diakses pada 27 Oktober 2025.
- Naury, C., Fudholi, D. H., & Hidayatullah, A. F. (2021). Topic Modelling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM. *Jurnal Media Informatika Budidarma*, 5(1), 24. <https://doi.org/10.30865/mib.v5i1.2556>
- Puspita, E., Shiddieq, D. F., & Roji, F. F. (2024). Topic Modeling on Online News Media Using Latent Dirichlet Allocation (Case Study Somethinc Brand). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 481–489. <https://doi.org/10.30865/mib.v5i1.2556>
- Rakhmawati, N. A., Cisatra, A., Ansori, D. D. M., Akmal, D. N. F. A., & Ramadhani, S. (2024). Identifikasi Topik Hangat di Media Berita Menggunakan Latent Dirichlet Allocation. *Journal of Information Engineering and Educational Technology*, 8(1), 14–17. <https://doi.org/10.26740/jieet.v8n1.p14-17>
- Ramadhan, C., Atina, V., & Permatasari, H. (2025). Analisis Perbandingan Model CNN dan IndoBERT Dalam Sentimen Berita Politik Indonesia. 110–118. <https://doi.org/10.47701/v1r9ka69>

Syafutra, A. D., & Kusrini, K. (2025). Analisis Sentimen Omnibus Law di Twitter dengan Machine Learning dan Teknik Resampling. *Jambura Journal of Informatics*, 1(1), 22–35.
<https://doi.org/10.37905/jji.v1i1.30935>

Ulloa, R., Mangold, F., Schmidt, F., Gilzbach, J., & Stier, S. (2025). Beyond time delays: how web scraping distorts measures of online news consumption. *Communication Methods and Measures*, 19(3), 179–200.
<https://doi.org/10.1080/19312458.2025.2482538>