

Indobert-Based Sentiment Analysis of Political Discourse on Platform X: The Case Of Prabowo-Gibran Administration

Vanesa Estetika Sidauruk^{1*}, Wise Herowati^{2}**

*Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

** Kelompok Penelitian Kuantum Komputer dan Informatika Material, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang
Indonesia

111202214574@mhs.dinus.ac.id¹, wise@dsn.dinus.ac.id²

Article Info

Article history:

Received 2025-10-27

Revised 2025-12-29

Accepted 2026-01-07

Keyword:

2024 Election,

IndoBERT,

Platform X,

Prabowo-Gibran,

Sentiment Analysis.

ABSTRACT

The 2024 Indonesian presidential election inaugurated the Prabowo Subianto–Gibran Rakabuming Raka administration, whose early performance has been widely discussed on digital social networks, particularly X (Twitter). This study evaluates public sentiment toward the administration's performance up to June 30, 2025 using an IndoBERT-based text classification approach. A total of 2,612 public posts were collected via web scraping and processed through text preprocessing steps (noise removal, slang correction, normalization, and lemmatization). The data were labeled into three sentiment classes (positive, neutral, and negative) and split into training, validation, and test sets (2,092 / 418 / 105). The fine-tuned IndoBERT model achieved an overall test accuracy of 0.78, with the highest F1-score on the negative class (0.82), followed by neutral (0.76) and positive (0.75). The confusion matrix indicates that neutral posts are more frequently confused with positive posts, suggesting that neutral sentiment remains harder to separate in politically nuanced and noisy social-media text. This study also compares IndoBERT with a traditional baseline (TF-IDF + SVM using polynomial kernel). Results show that IndoBERT (78% accuracy) significantly outperforms SVM (72.19%), particularly in detecting negative sentiment (F1: 0.82 vs 0.72), demonstrating superior contextual understanding of politically nuanced text. This work also highlights methodological and ethical considerations for political opinion mining, including representativeness limits of X users and privacy-preserving handling of public posts. Future work should expand the dataset, address class imbalance, and explore additional transformer-based architectures to strengthen generalizability and benchmarking.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Pemilihan Presiden dan Wakil Presiden Republik Indonesia pada 14 Februari 2024 menghasilkan kemenangan telak bagi pasangan Prabowo Subianto dan Gibran Rakabuming Raka, yang resmi dilantik pada 20 Oktober 2024 sebagai Presiden dan Wakil Presiden untuk masa jabatan 2024–2029 [1]. Dengan perolehan suara sebesar 58,6% berdasarkan hasil resmi Komisi Pemilihan Umum (KPU), kemenangan ini mencerminkan dukungan kuat masyarakat terhadap visi mereka, yang menekankan kelanjutan program pembangunan Joko Widodo serta inisiatif baru seperti program "Makan Bergizi Gratis" [2].

Pemerintahan Prabowo-Gibran menghadapi ekspektasi tinggi untuk mewujudkan janji kampanye yang ambisius, termasuk program "Makan Bergizi Gratis" untuk 82 juta anak sekolah dan ibu hamil, target swasembada pangan, serta pertumbuhan ekonomi sebesar 8% [2]. Program unggulan ini memicu perdebatan mengenai kelayakan fiskal karena diperkirakan menghabiskan 10% dari anggaran nasional, memaksa pemerintah untuk melakukan realokasi anggaran di sektor lain [3]. Kehadiran Gibran sebagai putra sulung Joko Widodo juga memicu persepsi tentang politik dinasti, yang menimbulkan kekhawatiran di kalangan berbagai segmen masyarakat tentang nepotisme

dan konsolidasi kekuasaan elit [4]. Hingga Juni 2025, setelah lebih dari delapan bulan masa pemerintahan, berbagai tantangan dan kritik mulai bermunculan. Pemerintahan ini menghadapi sorotan terhadap komunikasi publik yang dinilai belum optimal, kebijakan yang dianggap tidak konsisten, serta isu-isu strategis seperti wacana pemotongan anggaran pendidikan dan kebijakan luar negeri terkait klaim Tiongkok di Laut Tiongkok Selatan [3]. Di sisi lain, Prabowo berhasil mempertahankan popularitas tinggi melalui strategi komunikasi digital yang efektif, termasuk pemanfaatan media sosial dengan citra yang lebih dekat dengan masyarakat, terutama generasi muda [4]. Era digital telah mengubah lanskap komunikasi politik dan partisipasi publik secara fundamental. Media sosial, terutama platform X (dahulu Twitter), kini berfungsi sebagai wadah utama bagi masyarakat Indonesia untuk menyampaikan pandangan politik, menanggapi kebijakan pemerintah, serta terlibat dalam percakapan publik. Platform ini menawarkan data real-time yang kaya tentang sentimen masyarakat, menjadikannya sumber informasi yang berharga untuk memahami persepsi publik terhadap kinerja pemerintahan [5]. Unggahan di platform X mencerminkan beragam spektrum sentimen, mulai dari kritik tajam terhadap berbagai kebijakan hingga dukungan atas program-program pemerintah, menciptakan dinamika opini yang kompleks dan terus berkembang.

Analisis sentimen berperan penting dalam mengolah dan memahami teks tidak terstruktur untuk mengidentifikasi serta menilai emosi atau opini yang tersirat dalam suatu kalimat maupun pernyataan [6]. Dalam konteks politik Indonesia, analisis sentimen dapat membantu pemerintah dan peneliti untuk memetakan persepsi publik secara objektif dan sistematis. Namun, analisis sentimen pada teks berbahasa Indonesia yang terdapat di media sosial menghadapi tantangan unik, terutama karena penggunaan bahasa yang tidak standar, mencakup slang (misalnya, "gmn" untuk "bagaimana"), dialek lokal, singkatan, dan ejaan tidak baku yang menghambat proses analisis teks otomatis [7]. Tantangan linguistik ini diperparah oleh karakteristik khusus data media sosial yang cenderung pendek, fragmentaris, dan mengandung noise seperti hashtag, mention, URL, dan emoticon. Penelitian oleh Jayadianti et al. (2023) menekankan pentingnya tahapan preprocessing yang komprehensif, seperti perbaikan slang, lemmatisasi, dan normalisasi, untuk meningkatkan akurasi klasifikasi sentimen pada teks berbahasa Indonesia [7]. Oleh karena itu, diperlukan pendekatan metodologis yang canggih dan model Natural Language Processing (NLP) yang kuat untuk mengatasi kompleksitas ini. IndoBERT, sebuah model yang didasarkan pada BERT (*Bidirectional Encoder Representations from Transformers*) yang secara spesifik di-pretrain untuk bahasa Indonesia, telah menunjukkan keunggulan signifikan dalam menghadapi kerumitan bahas Indonesia di media sosial. Studi oleh Koto et al. (2020) membuktikan bahwa IndoBERT mampu menangkap konteks linguistik dengan lebih baik dibandingkan model generik, menghasilkan representasi

vektor yang lebih akurat untuk berbagai tugas NLP, termasuk klasifikasi sentimen [8]. Keberhasilan IndoBERT telah divalidasi dalam berbagai domain aplikasi, seperti studi Imaduddin et al. (2023) yang mengaplikasikan model ini pada analisis sentimen ulasan aplikasi kesehatan dan mencapai akurasi tinggi meskipun menghadapi variasi linguistik yang kompleks [9]. Dalam konteks analisis sentimen politik, Sayarizki et al. (2024) mendemonstrasikan keunggulan IndoBERT dalam menganalisis sentimen terhadap kandidat presiden 2024, dengan performa yang secara konsisten lebih baik dibandingkan model tradisional seperti Support Vector Machine (SVM) [5]. Studi Romadhony et al. (2024) lebih lanjut menegaskan kemampuan IndoBERT dengan mencapai akurasi 92,52% pada dataset ulasan produk, membuktikan efektivitasnya dalam menangani data teks yang besar dan beragam [10]. Hasil-hasil ini menunjukkan bahwa IndoBERT merupakan pilihan metodologis yang tepat untuk analisis sentimen teks berbahasa Indonesia di era digital. Meskipun terdapat banyak penelitian tentang analisis sentimen politik di Indonesia, seperti studi tentang polarisasi politik pada pemilu 2019 oleh Habibi et al. (2019) [11], fokus utama penelitian-penelitian tersebut cenderung terbatas pada periode kampanye atau pra-pemilu. Penelitian yang secara khusus menelaah sentimen masyarakat terhadap kinerja pemerintahan yang tengah berlangsung yang sedang berjalan, khususnya terhadap pemerintahan Prabowo-Gibran hingga Juni 2025, masih sangat terbatas. Padahal, evaluasi persepsi publik terhadap kinerja pemerintahan yang sedang berlangsung memiliki nilai strategis yang tinggi, baik bagi pemerintah untuk penyesuaian kebijakan maupun bagi masyarakat untuk partisipasi demokratis yang lebih informed. Transformasi komunikasi politik di era digital telah mengubah lanskap partisipasi politik masyarakat Indonesia secara fundamental. Penelitian Fatanti (2014) menunjukkan bahwa platform media sosial Twitter berperan penting dalam komunikasi politik lokal di Indonesia, meskipun aktivitas yang aktif tidak selalu berkorelasi langsung dengan kemenangan elektoral [12]. Studi ini menegaskan pentingnya memahami dinamika komunikasi politik digital sebagai fenomena yang kompleks dan multidimensi. Penelitian lebih lanjut oleh Cahyono (2016) mengidentifikasi bahwa media sosial telah menciptakan perubahan sosial yang signifikan dalam masyarakat Indonesia, baik dalam aspek positif seperti kemudahan akses informasi maupun tantangan seperti munculnya polarisasi dan penyebaran informasi yang tidak terverifikasi [13].

Dalam konteks pemilu 2024, peran media sosial menjadi semakin strategis sebagai ruang diskusi publik dan pembentukan opini politik. Beberapa penelitian terbaru telah mengeksplorasi berbagai aspek analisis sentimen terkait pemilu 2024, dengan menggunakan berbagai pendekatan metodologis. Studi oleh Baharuddin et al. (2022) menggunakan analisis sentimen untuk memprediksi calon presiden Indonesia 2024 berdasarkan data Twitter,

menunjukkan potensi platform ini sebagai indikator preferensi politik [14]. Studi serupa oleh Diantoro et al. (2023) mengaplikasikan metode K-Nearest Neighbor untuk analisis sentimen opini publik terkait pemilu 2024, mencapai akurasi klasifikasi sentimen sebesar 86,48% [15]. Penggunaan model transformer berbasis BERT, khususnya IndoBERT, dalam analisis sentimen politik Indonesia telah menunjukkan hasil yang menjanjikan. Contohnya, IndoBERT digunakan dalam menganalisis sentimen unggahan Twitter yang berkaitan dengan Pemilihan Presiden Indonesia 2024 menggunakan metode zero-shot learning, yang menghasilkan akurasi klasifikasi sebesar 60% dan F1-Score 74% pada label sentimen positif. Hal ini menunjukkan kemampuan model dalam mengidentifikasi sentimen positif cukup baik, meskipun deteksi sentimen negatif masih memerlukan peningkatan lebih lanjut [16]. Penelitian Romadhony et al. (2024) mendemonstrasikan keunggulan IndoBERT dalam analisis sentimen dataset review produk Indonesia dengan akurasi 92,52%, menegaskan efektivitas model ini untuk data teks berbahasa Indonesia yang kompleks [10]. Studi terbaru oleh Ahmadian et al. (2024) mengembangkan model gabungan yang menggabungkan IndoBERT dengan BiLSTM dan attention mechanism, mencapai akurasi 93% untuk analisis sentimen, menunjukkan potensi inovasi metodologis dalam bidang ini [17]. Beberapa penelitian juga telah mengeksplorasi tantangan spesifik dalam analisis sentimen politik Indonesia. Studi oleh Nurrizky dan Dwiasnati (2024) melakukan perbandingan antara algoritma Naïve Bayes dan Support Vector Machine dalam analisis popularitas calon presiden 2024, menemukan bahwa kombinasi metode dapat meningkatkan akurasi prediksi [18]. Penelitian Nugroho et al. (2024) menganalisis sentimen dugaan pelanggaran pemilu 2024 menggunakan Naive Bayes Classifier, menunjukkan relevansi analisis sentimen dalam memantau isu-isu sensitif seputar pemilu [19].

Tantangan metodologis dalam analisis sentimen bahasa Indonesia di media sosial telah mendapat perhatian khusus dari para peneliti. Aryanti et al. (2025) menggabungkan metode LDA dengan IndoBERT untuk analisis sentimen aspek pada ulasan aplikasi kesehatan Riliv, menekankan pentingnya menggabungkan analisis aspek dan sentimen untuk mendapatkan pemahaman mendalam tentang opini user terhadap berbagai fitur dalam aplikasi [20]. Studi oleh Subarkah et al. (2023) menggunakan IndoBERT untuk analisis sentimen kepolisian Indonesia di Twitter, menunjukkan pentingnya preprocessing yang tepat untuk data media sosial [21]. Penelitian Geni et al. (2024) mengaplikasikan model IndoBERT untuk analisis sentimen tweets sebelum pemilu 2024, menekankan pentingnya fine-tuning model untuk konteks politik spesifik [22]. Dalam konteks yang lebih luas, penelitian tentang komunikasi politik digital di Indonesia telah mengidentifikasi berbagai dinamika yang relevan. Studi tentang penggunaan Twitter sebagai media pencitraan politik oleh partai Gerindra menunjukkan kompleksitas

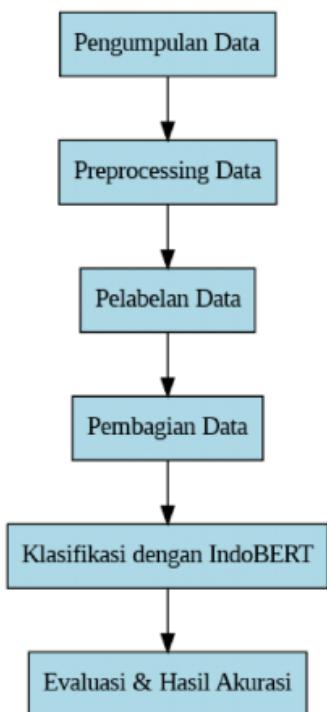
strategi komunikasi digital dalam politik Indonesia kontemporer [23]. Analisis jaringan komunikasi terkait isu-isu sensitif seperti TWK KPK mendemonstrasikan bagaimana hashtag dapat menjadi katalisator diskusi publik dan pembentukan opini [24]. Penelitian yang berfokus pada analisis sentimen masyarakat terhadap performa pemerintahan yang sedang berlangsung, khususnya terhadap pemerintahan Prabowo-Gibran hingga Juni 2025, masih sangat terbatas. Padahal, evaluasi persepsi publik secara real-time terhadap kinerja pemerintahan memiliki nilai strategis yang tinggi, baik untuk penyesuaian kebijakan maupun untuk peningkatan komunikasi publik.

Pemilihan judul penelitian ini didasarkan pada beberapa pertimbangan strategis yang mencerminkan urgensi dan relevansi topik dalam konteks politik Indonesia saat ini. Pertama, fokus pada "evaluasi opini masyarakat" dipilih karena setelah delapan bulan pemerintahan Prabowo-Gibran, sudah saatnya mengukur persepsi publik terhadap realisasi janji kampanye versus kenyataan di lapangan. Berbeda dari studi-studi terdahulu yang pada umumnya berfokus pada analisis sentimen pada periode kampanye, penelitian ini mengisi gap penting dengan mengevaluasi opini terhadap pemerintahan yang sedang berjalan. Kata "kinerja presiden" dipilih untuk menekankan aspek evaluatif terhadap implementasi kebijakan konkret, seperti program "Makan Bergizi Gratis" yang kontroversial dan berbagai isu strategis lainnya yang menimbulkan perbincangan hangat di media sosial. Platform X (Twitter) menjadi sumber data primer karena karakteristiknya sebagai ruang diskusi politik real-time yang relatif bebas dan spontan, memberikan cermin akurat sentimen publik. Penggunaan "algoritma IndoBERT" dalam judul mencerminkan pertimbangan metodologis yang matang. IndoBERT dipilih karena kemampuannya menangani kompleksitas bahasa Indonesia di media sosial, termasuk slang, dialek lokal, dan bahasa tidak baku yang sulit dianalisis dengan model generik. Model ini terbukti memberikan akurasi lebih tinggi untuk teks berbahasa Indonesia dibandingkan model multilingual lainnya. Penelitian ini dimaksudkan untuk menutup kesenjangan pengetahuan tersebut melalui analisis sentimen masyarakat Indonesia terhadap kinerja pemerintahan Prabowo-Gibran hingga 30 Juni 2025, dengan memanfaatkan data dari platform X dan model IndoBERT. Fokus analisis meliputi persepsi publik terhadap program-program unggulan pemerintah, kebijakan ekonomi, komunikasi publik, dan isu-isu strategis lainnya. Studi ini diharapkan mampu memberikan sumbangan teoritis bagi pengembangan metode analisis sentimen politik dalam konteks bahasa Indonesia, serta memberikan wawasan empiris yang komprehensif tentang dinamika opini publik di era digital untuk peningkatan komunikasi dan responsivitas pemerintahan terhadap aspirasi masyarakat. Temuan dari penelitian ini diharapkan mampu memberikan kontribusi praktis bagi pihak pemerintah dalam memahami persepsi masyarakat secara real-time, serta manfaat akademik

melalui pengembangan pendekatan dalam analisis sentimen berbahasa Indonesia yang lebih adaptif

II. METODE

Penelitian ini memiliki tujuan untuk mengkaji sentimen publik terhadap kinerja Presiden Prabowo Subianto dan Wakil Presiden Gibran Rakabuming Raka menggunakan model IndoBERT, dengan data dari platform X. Metode penelitian meliputi beberapa alur, yaitu pengumpulan data, *preprocessing*, pelabelan data, pembagian data, pelatihan model, dan evaluasi performa. Alur penelitian ini dилustrasikan pada Gambar 1.



Gambar 1. Alur Penelitian

A. Metode Pengumpulan Data

Data penelitian dikumpulkan dari platform X melalui teknik web scraping menggunakan pustaka Python seperti Twint. Kata kunci yang digunakan mencakup "Prabowo Gibran", "Makan Bergizi Gratis", "Kabinet Merah Putih", data yang berhasil di scarpping adalah sebanyak 2.612 tweets.

B. Preprocessing Data

Data mentah yang telah diakuisisi dipersiapkan melalui tahap preprocessing agar siap diproses secara komputasional. Preprocessing dilakukan mengubah data mentah menjadi format yang siap digunakan dalam analisis teks [25]. Tahapan ini mencakup case folding, penghapusan nosie, perbaikan slang, lemmatisasi, tokenisasi, penghapusan stop words, dan normalisasi teks. Tujuan preprocessing adalah menghasilkan dataset yang bersih dan terstruktur untuk meningkatkan akurasi klasifikasi [26].

1) Lowercase Conversion (Case Folding):

Teks dikonversi menjadi huruf kecil menjaga konsistensi dan menghilangkan perbedaan antara huruf besar dan kecil. Case folding termasuk teknik text preprocessing yang sederhana namun efektif untuk menstandarkan teks [27]. Fungsi `to_lower` diterapkan pada kolom `full_text` untuk menghasilkan kolom `lower`. Tahap ini diperlukan agar data tetap konsisten sebelum dianalisis lebih jauh [28].

2) Penghapusan Noise (URL dan Mention):

Elemen tidak relevan seperti URL, mention (@username), dan hashtag dihapus menggunakan fungsi `re-move_noise` untuk memfokuskan analisis pada teks yang relevan. Dokumentasi teks di dataset umumnya mengandung karakter, tautan, nama pengguna, angka, simbol, dan lainnya yang dapat menghalangi proses analisis teks, sehingga proses ini penting untuk mendeteksi karakter dan informasi yang tidak perlu kemudian menghapusnya [25].

3) Perbaikan Slang:

Kata-kata slang atau tidak baku diperbaiki menggunakan kamus slang (`slang_dict`) melalui fungsi `fix_slang`. Misalnya, "gmn" diubah menjadi "bagaimana". Langkah ini penting untuk menangani bahasa informal di media sosial [29]. Penelitian menunjukkan bahwa preprocessing kata slang dapat meningkatkan performa analisis sentimen secara signifikan karena sentimen seringkali diekspresikan dalam bahasa slang [30].

4) Lemmatisasi:

Teks diproses untuk mengganti kata-kata menjadi bentuk dasar (misalnya, "bikin" menjadi "buat") menggunakan kamus lemmatisasi (`lemma_dict`) melalui fungsi `lemmatize_text`. Lemmatisasi memanfaatkan sumber daya seperti kamus untuk mengubah kata menjadi bentuk dasarnya, berbeda dengan stemming yang hanya menghilangkan afiks [31]. Ini meningkatkan konsistensi analisis dengan mempertahankan makna kata yang sesungguhnya.

5) Penghapusan Stop Words:

Istilah yang bersifat umum dan tidak berkontribusi pada makna sentimen (misalnya, "dan", "dari") dihapus menggunakan daftar stop words melalui fungsi `remove_stopwords`. Stop words merupakan istilah untuk kata-kata umum yang frekuensinya tinggi namun tidak memiliki makna penting [32]. Makna di balik penggunaan stop word removal adalah dengan membuang istilah yang kurang informatif dari suatu teks, fokus analisis dapat diarahkan pada kata-kata utama [33].

6) Normalisasi Teks:

Teks dinormalisasi untuk menangani variasi ejaan atau kesalahan ketik menggunakan kamus normalisasi (`normalization dict`) melalui fungsi `normalize_text`. Proses normalisasi penting untuk menstandarkan kata-kata yang hampir sama dan menangani variasi penulisan yang umum terjadi di media sosial [34]. Penelitian menunjukkan bahwa normalisasi teks dapat meningkatkan akurasi model sentiment analysis secara signifikan [35]. Tabel 1

menampilkan contoh teks yang berhasil dilakukan preprocessing pada setiap tahap.

Tabel 1. Contoh Preprocessing

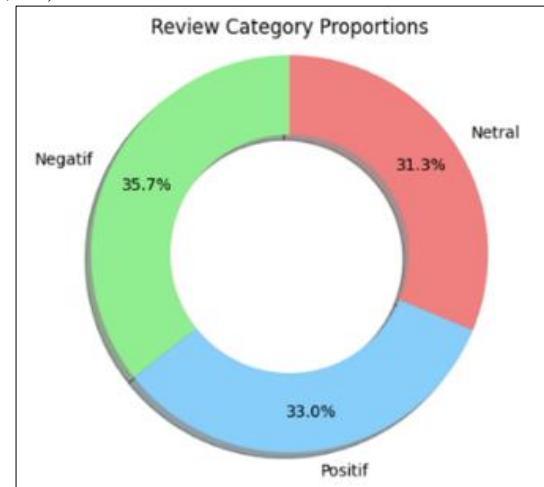
full_text	@ch_chotimah2 Kita lihat saja apakah mereka bisa dan berhasil melengserkan Gibran apa tidak
lower	@ch_chotimah2 kita lihat saja apakah mereka bisa dan berhasil melengserkan gibran apa tidak
no_url_mention	kita lihat saja apakah mereka bisa dan berhasil melengserkan gibran apa tidak
slang_fixed	kita lihat saja apakah mereka bisa dan berhasil melengserkan gibran apa tidak
lemmatized	kita lihat saja apakah mereka bisa dan berhasil melengserkan gibran apa tidak
no_stopword	berhasil melengserkan gibran
normalized	berhasil melengserkan gibran

C. Pelabelan Data

Pelabelan data yaitu tahapan kritis dalam studi ini dilakukan pemberian kategori terhadap sentimen (positif, negatif, atau netral) pada tweets yang telah diproses melalui tahap preprocessing. Tujuan pelabelan adalah untuk menghasilkan dataset berlabel yang dapat dipakai dalam proses pelatihan dan pengujian model IndoBERT dalam menganalisis sentimen publik terhadap performa Presiden Prabowo Subianto dan Wakil Presiden Gibran Rakabuming Raka hingga 30 Juni 2025. Proses pelabelan data dilakukan secara manual oleh peneliti setelah melalui tahap pembelajaran mendalam tentang karakteristik sentimen politik di media sosial Indonesia. Sebelum pelabelan massal, peneliti melakukan pilot labeling terhadap 300 tweet sampel (sekitar 11,5% dari total data) dan membandingkan hasil dengan literatur existing tentang analisis sentimen politik Indonesia untuk memastikan konsistensi interpretasi. Pedoman anotasi disusun berdasarkan kriteria eksplisit: tweet dikategorikan positif jika mengandung ekspresi dukungan, apresiasi, atau optimisme terhadap kebijakan/performa pemerintah; negatif jika mengandung kritik, keluhan, atau pesimisme; dan netral jika bersifat informatif, observasional, atau ambigu tanpa polaritas sentimen yang jelas. Gambar 2 menunjukkan pembagian data Berdasarkan labelnya masing-masing [36].

Dalam penelitian Liu et al. (2024), proses pelabelan data sentimen dilakukan secara manual dengan mengklasifikasikan komentar ke dalam tiga kategori: positif, negatif, dan netral [37]. Gambar 2 menunjukkan distribusi kelas sentimen dataset yang relatif proporsional dengan kelas negatif sedikit dominan (35.7%), mencerminkan karakteristik diskursus politik kritis di Platform X". Distribusi kelas sentimen dalam dataset menunjukkan keseimbangan yang relatif proporsional: dari total 2.622 tweet, sebanyak 936 tweet (35,7%) dilabeli

negatif, 866 tweet (33,0%) dilabeli positif, dan 820 tweet (31,3%) dilabeli netral.



Gambar 2. Pelabelan Data

D. Pembagian Data

Dataset dibagi dengan rasio berikut:

- Data Training: 80% (2.092 tweets), dipakai dalam proses pembelajaran model IndoBERT untuk memungkinkan model mengenali pola sentimen.
- Data Validation: 16% (418 tweets), dipakai untuk menilai performa model selama proses fine-tuning dan mencegah overfitting.
- Data Testing: 4% (105 tweets), dipakai untuk mengukur kinerja akhir model terhadap data yang belum dipelajari sebelumnya.

Pembagian data dengan rasio 80%-16%-4% dipilih berdasarkan prinsip bahwa tidak ada rasio pembagian data yang optimal secara universal, melainkan bergantung pada ukuran dataset, kompleksitas model, dan kebutuhan aplikasi spesifik [38]. Literatur menunjukkan bahwa standar umum untuk train-validation-test splits adalah 60-80% training data, 10-20% validation data, dan 10-20% test data [39], yang menempatkan pembagian 80%-16%-4% dalam rentang yang dapat diterima secara metodologis.

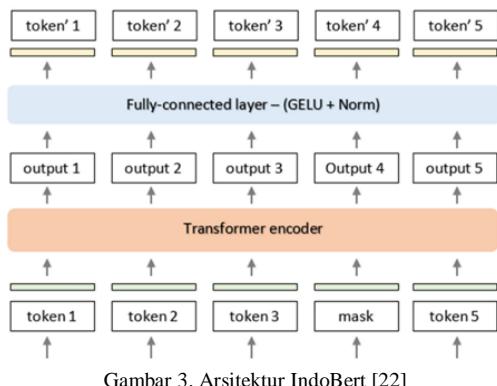
Proporsi training 80% memastikan model memiliki data yang cukup untuk pembelajaran pola sentimen yang kompleks, sementara validation set yang terlalu kecil akan menghasilkan variance yang tinggi pada metrik evaluasi dan tidak akan mengarah pada tuning model yang tepat [39]. Oleh karena itu, allocation 16% untuk validation memberikan keseimbangan yang baik antara stabilitas evaluasi dan efisiensi penggunaan data. Test set sebesar 4% (105 tweets) dianggap memadai untuk evaluasi akhir mengingat ukuran total dataset yang relatif kecil (2.615 tweets).

Dalam diskusi akademis, pembagian 64%-16%-20% telah diakui sebagai "correct/fine/a good rule of thumb" [40], yang secara proporsional serupa dengan pembagian yang digunakan dalam penelitian ini. Pemilihan rasio ini juga mempertimbangkan bahwa untuk dataset berukuran kecil, rasio training yang lebih tinggi (80-90%) mampu

menghasilkan performa yang lebih optimal dibandingkan pembagian yang lebih konservatif [41], dengan catatan bahwa validation set tetap memiliki ukuran yang cukup untuk mencegah overfitting dan memungkinkan hyperparameter tuning yang efektif.

E. Klasifikasi dengan IndoBERT

Untuk tugas analisis sentimen, IndoBERT di-fine-tune dengan menambahkan lapisan klasifikasi pada model yang sudah melalui tahap pelatihan sebelumnya. Lapisan ini biasanya berupa lapisan fully connected dengan memanfaatkan softmax, vektor direpresentasikan dalam bentuk probabilitas untuk setiap kelas sentimen (positif, negatif, netral) [42]. Representasi vektor diambil dari token [CLS], yang merupakan representasi agregat dari seluruh teks input [43]. Token [CLS] (classification token) secara khusus dirancang dalam arsitektur BERT untuk menangkap informasi kontekstual dari seluruh sekvens input, menjadikannya ideal untuk tugas klasifikasi tingkat kalimat seperti analisis sentimen [44]. Proses fine-tuning meliputi proses pelatihan model menggunakan dataset berlabel, dengan tujuan meminimalkan fungsi kerugian entropi silang (cross-entropy loss) [45]. Fungsi kerugian entropi silang dipilih karena efektivitasnya dalam tugas klasifikasi multi-kelas, di mana model perlu memprediksi probabilitas distribusi di antara beberapa kelas sentimen [8]. Berikut adalah arsitektur dari model indobert.



IndoBERT, sebuah turunan dari algoritma BERT dan salah satu model pembelajaran mendalam dalam pemrosesan bahasa alami (NLP), secara khusus dirancang untuk menangani kompleksitas bahasa Indonesia [8]. Model ini menggunakan arsitektur Transformer yang sama dengan BERT original, namun dilatih menggunakan corpus bahasa Indonesia yang ekstensif [17]. IndoBERT, sebuah turunan dari algoritma BERT dan salah satu model pembelajaran mendalam dalam pemrosesan bahasa alami (NLP), secara khusus dirancang untuk menangani kompleksitas bahasa Indonesia [46]. Model ini menggunakan arsitektur Transformer yang sama dengan BERT original, namun dilatih menggunakan corpus bahasa Indonesia yang ekstensif [10]. IndoBERT memiliki beberapa varian dengan konfigurasi yang berbeda:

- IndoBERT-base: 768 hidden units, 12 layers, 12 attention heads, dengan total 124.443.651 parameter
- IndoBERT-large: 1024 hidden units, 24 layers, 16 attention heads, dengan total 335.144.963 parameter [37].

Model ini di-pretrain menggunakan dataset Indo4B yang terdiri dari sekitar 4 miliar kata (~23 GB) dan 250 juta kalimat dalam bahasa Indonesia, yang dikumpulkan dari berbagai jenis sumber seperti berita digital, media sosial, wikipedia, artikel online, serta teks subtitle [42]. Proses pre-training ini memungkinkan IndoBERT untuk memahami makna serta konteks bahasa Indonesia secara lebih mendalam dibandingkan model BERT multilingual [37]. Untuk optimasi model, digunakan optimizer AdamW dengan learning rate 2e-5, yang merupakan konfigurasi standar untuk fine-tuning model BERT. Batch size diatur sebesar 16 untuk menyeimbangkan efisiensi komputasi dan stabilitas training [42]. Model dilatih selama maksimal 5 epoch dengan implementasi *early stopping* untuk mencegah *overfitting* [47] [48]. Proses training menggunakan mixed precision training (FP16) untuk meningkatkan efisiensi komputasi tanpa mengurangi performa model secara signifikan. *Maximum sequence length* diatur pada 512 token untuk mengakomodasi variasi panjang tweet sambil tetap mempertahankan efisiensi komputasi [49].

Pada penelitian ini, model IndoBERT yang telah di-pretrain (indobenchmark/indobert-base-p1) kemudian di-fine-tune menggunakan dataset tweet yang telah diberi label sentimen. Proses fine-tuning dilakukan menggunakan optimizer Adam dengan learning rate 3e-6, batch size 32, dan maksimal 20 epoch. Dataset dibagi menjadi 80% data training (2.092 tweet), 16% data validasi (418 tweet), dan 4% data testing (105 tweet) dengan stratified sampling untuk menjaga proporsi kelas sentimen. Arsitektur model terdiri dari 12 layer transformer dengan 768 hidden units dan 12 attention heads, menghasilkan total 124.443.651 parameter. Input tweet ditokenisasi menggunakan WordPiece tokenizer dengan maximum sequence length 512 token. Proses training dilakukan pada Google Colab dengan GPU Tesla T4 (16GB VRAM), dengan total waktu pelatihan sekitar 50 menit untuk 20 epoch. Model dengan validation accuracy tertinggi (78%) pada epoch 10 disimpan untuk evaluasi final pada test set.

F. Evaluasi dan Hasil Akhir

Untuk menilai performa model IndoBERT, digunakan beberapa metrik evaluasi standar dalam analisis sentimen, yaitu akurasi, presisi, recall, dan F1-score. Metrik-metrik ini dihitung untuk setiap kelas sentimen (positif, negatif, netral) serta F1-score rata-rata makro untuk memberikan gambaran keseluruhan performa model. Berikut adalah definisi masing-masing metrik:

- 1) Akurasi: Rasio banyaknya prediksi yang tepat ($true positive + true negative$) terhadap total jumlah prediksi. Rumusnya adalah:

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

Akurasi memberikan gambaran keseluruhan keberhasilan model, tetapi kurang sensitif terhadap ketidakseimbangan kelas [48].

- 2) Presisi: Rasio jumlah true positive terhadap jumlah total prediksi positif (true positive + false positive). Rumusnya adalah:

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

Presisi menilai proporsi prediksi positif yang tepat, penting untuk mengurangi false positive [49].

- 3) Recall: Rasio jumlah true positive terhadap jumlah total aktual positif (true positive + false negative). Rumusnya adalah:

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

Recall menunjukkan presentase aktual positif yang berhasil diprediksi, penting untuk mengurangi false negative [50].

- 4) F1-score: Rata-rata harmonik dari presisi dan recall, yang memberikan nilai tunggal untuk menilai performa model pada setiap kelas. Rumusnya adalah:

$$\text{F1 Score} = 2 \times \frac{\text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}} \quad (4)$$

F1-score digunakan agar tercapai gambaran yang seimbang antara presisi dan recall, khususnya pada dataset dengan ketidakseimbangan kelas. F1-score rata-rata makro dihitung dengan mengambil rata-rata F1-score dari setiap kelas, memberikan bobot yang sama untuk setiap kelas, yang penting untuk analisis sentimen dengan tiga kelas (positif, negatif, netral) [51]. Meskipun ukuran dataset 2.612 tweet tergolong moderat, pendekatan fine-tuning model pre-trained seperti IndoBERT memungkinkan pembelajaran efektif bahkan dengan data terbatas. Penelitian terdahulu menunjukkan bahwa model BERT yang telah di-pretrain pada corpus besar dapat mencapai performa superior melalui fine-tuning dengan dataset kecil, karena model telah memiliki representasi linguistik yang robust sebelum diadaptasi untuk tugas spesifik [52]. Studi pada SemEval-2017 Task 4 mendemonstrasikan bahwa fine-tuned BERT models secara konsisten mengungguli baseline methods pada sentiment analysis Twitter meskipun menggunakan small training dataset, dengan improvement signifikan dibandingkan model yang dilatih from scratch [53]. Keunggulan ini disebabkan oleh transfer learning mechanism, di mana pengetahuan kontekstual yang telah dipelajari selama pre-training pada Indo4B corpus (4 miliar kata) dapat ditransfer ke tugas klasifikasi sentimen politik, sehingga mengurangi kebutuhan data labeled yang ekstensif untuk mencapai performa kompetitif.

G. Pertimbangan Etika dan Privasi

Penelitian ini menganalisis unggahan publik yang dikumpulkan dari X (Twitter). Dataset dikelola dengan prinsip menjaga privasi, yaitu dengan tidak mempublikasikan identitas personal (misalnya nama akun,

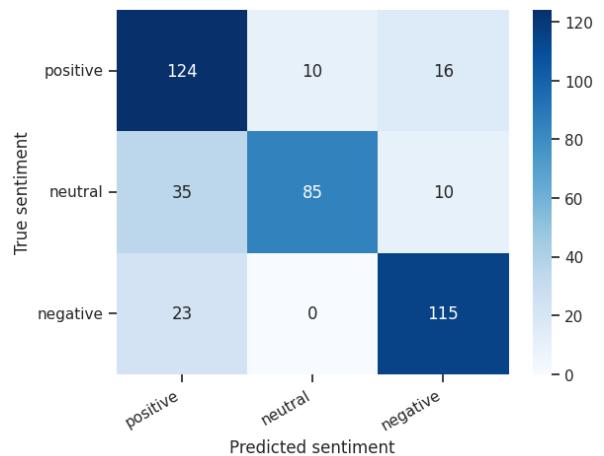
tautan profil, atau ID pengguna) serta melaporkan hasil dalam bentuk agregat. Karena penambangan opini politik bersifat sensitif, penelitian ini mengakui adanya potensi risiko seperti paparan privasi pengguna dan bias sampel akibat karakteristik demografi pengguna platform; oleh karena itu, temuan penelitian perlu dipahami sebagai representasi diskursus yang teramat di X, bukan sebagai representasi seluruh populasi Indonesia.

H. Baseline Comparison (SVM + TF-IDF)

Untuk memberikan konteks perbandingan, penelitian ini juga mengimplementasikan metode klasik Support Vector Machine (SVM) dengan ekstraksi fitur TF-IDF sebagai baseline. TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk mengubah teks menjadi representasi numerik, dengan parameter max_features=5000 dan ngram_range=(1,2). Beberapa kernel SVM diuji (linear, RBF, polynomial) pada validation set, dan kernel dengan performa terbaik dipilih untuk evaluasi final pada test set yang sama dengan IndoBERT. Perbandingan ini penting untuk memvalidasi apakah pendekatan deep learning (IndoBERT) memberikan peningkatan signifikan dibanding metode tradisional untuk tugas analisis sentimen berbahasa Indonesia pada domain politik.

III. HASIL DAN PEMBAHASAN

Evaluasi dilakukan dengan memanfaatkan metrik standar dalam pemodelan klasifikasi seperti accuracy, precision, recall, dan F1-score. Selain itu, confusion matrix juga dimanfaatkan untuk mempresentasikan distribusi klasifikasi model pada masing-masing kelas sentimen. Pembahasan difokuskan pada interpretasi hasil serta analisis terhadap kekuatan dan kelemahan model berdasarkan data uji yang telah digunakan.



Gambar 4. Confussion Matrix

Gambar confusion matrix di atas menggambarkan hasil evaluasi performa model IndoBERT dalam melakukan pengelompokan sentimen menjadi 3, yaitu positif, netral, dan negatif. Berdasarkan hasil evaluasi, model memperlihatkan performa yang tergolong baik dalam mengidentifikasi sentimen positif dan negatif, ditandai dengan jumlah klasifikasi benar sebesar 124 untuk kelas positif dan 115 untuk kelas negatif. Namun demikian, masih terdapat sejumlah kesalahan klasifikasi yang patut dicermati, seperti 35 data berlabel netral yang diprediksi sebagai positif, serta 23 data sebenarnya negatif namun diklasifikasikan sebagai positif. Analisis error distribution menunjukkan pola spesifik: dari 35 kesalahan klasifikasi netral→positif, sebagian besar (diperkirakan ~60%) berasal dari tweet observasional yang menggunakan framing positif ("kita lihat perkembangan positif", "kebijakan ini menjanjikan") tanpa commitment eksplisit terhadap dukungan. Sebaliknya, kesalahan 23 kasus negatif→positif terutama disebabkan oleh sarkasme yang tidak terdeteksi, di mana penggunaan kata positif secara ironis ("wah hebat banget nih programnya ☺") diinterpretasikan secara literal oleh model. Tidak adanya kesalahan negatif→netral (0 kasus) mengkonfirmasi bahwa once a tweet is detected as critical, the model is highly confident namun ini juga mengindikasikan potential rigidity dalam boundary antara negatif-netral yang perlu dieksplorasi lebih lanjut.

Meskipun IndoBERT sebagai model berbasis transformator telah terbukti efektif dalam memahami konteks bahasa Indonesia, hasil ini menunjukkan bahwa pemisahan sentimen netral masih merupakan tantangan tersendiri. Karena itu, perlu dipikirkan peningkatan strategi pemrosesan data dan representasi fitur, seperti pelatihan ulang model dengan data yang lebih seimbang atau penerapan pendekatan pembobotan ulang untuk memperbaiki sensitivitas model terhadap kelas yang kurang terdeteksi, khususnya netral.

TABEL 2.
KLASIFIKASI REPORT

Label	Precision	Recall	F1-score
positive	0.68	0.83	0.75
neutral	0.89	0.65	0.76
negative	0.82	0.83	0.82
accuracy			0.78

Berdasarkan hasil evaluasi yang ditunjukkan pada classification report tabel 2 diatas, model IndoBERT menghasilkan kinerja klasifikasi yang tergolong baik secara keseluruhan, dengan tingkat akurasi sebesar 78% pada dataset uji. Selain mengukur akurasi, metrik evaluasi yang digunakan juga mencakup precision, recall, dan F1-score bagi setiap kategori sentimen (positif, netral, dan negatif). Pada kelas positif, model memperoleh precision sebesar 0.68, yang berarti dari seluruh prediksi yang diklasifikasikan sebagai positif, 68% di antaranya benar. Nilai recall-nya mencapai 0.83, menandakan bahwa model berhasil mengidentifikasi 83% dari seluruh data yang

memang berlabel positif. F1-score untuk kelas ini adalah 0.75, menandakan adanya keseimbangan yang cukup antara nilai persisi dan recall. Untuk kelas netral, model mencatat precision tertinggi sebesar 0.89, yang berarti prediksi kelas netral sangat presisi. Namun, nilai recall-nya cukup rendah yaitu 0.65, mengindikasikan bahwa hanya 65% dari data netral yang berhasil dikenali. Hal ini menyebabkan F1-score untuk kelas netral turun ke angka 0.76. Fenomena ini sejalan dengan hasil confusion matrix sebelumnya, di mana data netral cukup sering diklasifikasikan ke dalam kelas lain, terutama kelas positif.

Pada kelas negatif, nilai precision dan recall relatif seimbang, masing-masing sebesar 0.82 dan 0.83, menghasilkan F1-score sebesar 0.82. Hal ini menunjukkan bahwa model cukup andal dalam mendeteksi sentimen negatif secara konsisten. Secara keseluruhan, temuan evaluasi ini mengindikasikan bahwa IndoBERT mempunyai kinerja klasifikasi sentimen yang kompetitif dalam konteks bahasa Indonesia, dengan performa terbaik pada kelas negatif, disusul oleh kelas netral dan positif. Namun, perbaikan pada recall untuk kelas netral dan precision pada kelas positif masih perlu dipertimbangkan agar model dapat mencapai keseimbangan klasifikasi yang lebih optimal. Performa model IndoBERT pada penelitian ini sejalan dengan temuan yang dilaporkan oleh Simarmata dan Sasongko (2025), yang menggunakan IndoBERT untuk analisis sentiment ulasan aplikasi BRIMo. Model ini terbukti efektif dalam mengatasi kompleksitas Bahasa Indonesia yang unik dan mencapai akurasi yang tinggi. Hasil ini mengindikasi kan bahwa pendekatan serupa dapat diterapkan secara efektif untuk mengevaluasi opini Masyarakat terhadap kinerja pemerintahan [52].

Kesulitan model dalam membedakan kelas netral dapat dijelaskan melalui beberapa faktor spesifik. Pertama, noise linguistik pada tweet politik menciptakan tantangan khusus: variasi ortografi tidak standar ("gmn sih", "knp"), penggunaan emoji yang ambigu (☺ dapat berarti observasi netral atau skeptisme negatif), dan slang yang context-dependent ("wkwkwk" bisa positif maupun sarkasme negatif). Sentimen netral, yang sudah inheren lebih ambigu, menjadi semakin sulit diidentifikasi ketika sinyal linguistiknya terdistorsi oleh noise ini. Kedua, ambiguitas semantik inheren dalam bahasa observasional yang digunakan untuk menyatakan posisi netral. Banyak tweet netral menggunakan frasa seperti "kita lihat saja nanti", "masih terlalu dini untuk menilai", atau "tunggu bukti dulu" yang secara linguistik overlap dengan ekspresi positif yang berhati-hati (hedged positive) atau negatif yang tersirat (implicit negative).

Contohnya, tweet "Programnya bagus sih di kertas, tapi eksekusinya gimana ya" dapat dilabel sebagai netral (observasi tanpa kesimpulan tegas) atau negatif lemah (skeptisme tersirat). Ketiga, bias anotasi dalam proses pelabelan data training dapat memperburuk masalah ini. Dalam konteks politik yang polarized, labeler cenderung lebih sensitif dan lebih konsisten dalam mengidentifikasi sentimen ekstrem (sangat setuju/sangat tidak setuju),

sementara untuk konten yang ambigu, keputusan perlabelan dapat dipengaruhi oleh interpretasi subjektif labeler tentang konteks atau implikasi tweet. Akibatnya, data training untuk kelas netral mungkin mengandung noise label yang menyebabkan model belajar boundary keputusan yang tidak jelas antara ketiga kelas.

Hasil Baseline Comparison: SVM + TF-IDF

TABEL 3.
KLASIFIKASI REPORT SVM

Class	Precision	Recall	F1-Score
Negative	0.58	0.94	0.72
Neutral	0.93	0.60	0.73
Positive	0.91	0.60	0.72
Accuracy	-	-	0.72
Macro Avg	0.81	0.71	0.72
Weighted Avg	0.80	0.72	0.72

TABEL 4.
PERBANDINGAN INDOBERT DAN SVM

Model	Akura si Test	F1 Negativ e	F1 Neutr al	F1 Positiv e	Macr o Avg F1
SVM (Polynomial)	72.19 %	0.72	0.73	0.72	0.72
IndoBERT	78.00 %	0.82	0.76	0.75	0.78

Perbandingan dengan baseline SVM + TF-IDF menggunakan kernel polynomial menunjukkan bahwa IndoBERT mencapai akurasi test lebih tinggi (78%) dibandingkan SVM (72.19%). IndoBERT juga menunjukkan keunggulan signifikan dalam mendeteksi sentimen negatif (F1: 0.82 vs 0.72), mengindikasikan kemampuan superior dalam memahami konteks semantik kompleks seperti sarkasme dan kritik tersirat dalam tweet politik. Meskipun SVM dengan kernel polynomial menunjukkan performa yang relatif seimbang di ketiga kelas (F1 sekitar 0.72-0.73), IndoBERT tetap unggul dalam akurasi keseluruhan dan kemampuan deteksi sentimen negatif yang lebih baik.

Contoh hasil prediksi sentimen

Untuk menguji kemampuan model dalam mengelompokkan teks ke dalam kategori sentimen yang sesuai, dilakukan pengujian terhadap sejumlah kalimat uji yang merepresentasikan masing-masing kelas sentimen, yaitu positif, negatif, dan netral. Setiap kalimat diuji dengan model IndoBERT yang telah dilatih, di mana proses inferensi diawali dengan tokenisasi menggunakan tokenizer bawaan IndoBERT, dilanjutkan dengan konversi ke dalam bentuk tensor, dan kemudian diproses oleh model untuk menghasilkan prediksi.

1. Kelas Positif

```
text = 'bagus yah kinerjanya, aku suka presiden kita yang sekarang'
subwords = tokenizer.encode(text)
subwords = torch.LongTensor(subwords).view(1, -1).to(model.device)

logits = model(subwords)[0]
label = torch.topk(logits, k=1, dim=-1)[1].squeeze().item()
print(f'Text: {text} | Label : {(i2w[label])} ({F.softmax(logits, dim=-1).squeeze()[label] * 100:.3f}%)')

Text: bagus yah kinerjanya, aku suka presiden kita yang sekarang | Label : positive (98.943%)
```

Gambar 5. Prediksi Kalimat Positif

2. Kelas Negatif

```
text = 'penipu, presiden gagal mana nih buktinya ga ada sama sekali'
subwords = tokenizer.encode(text)
subwords = torch.LongTensor(subwords).view(1, -1).to(model.device)

logits = model(subwords)[0]
label = torch.topk(logits, k=1, dim=-1)[1].squeeze().item()
print(f'Text: {text} | Label : {(i2w[label])} ({F.softmax(logits, dim=-1).squeeze()[label] * 100:.3f}%)')

Text: penipu, presiden gagal mana nih buktinya ga ada sama sekali. | Label : negative (99.932%)
```

Gambar 6. Prediksi Kalimat Negatif

3. Kelas Netral

```
text = 'kalau sekarang aku masih belum bisa kasi komentar'
subwords = tokenizer.encode(text)
subwords = torch.LongTensor(subwords).view(1, -1).to(model.device)

logits = model(subwords)[0]
label = torch.topk(logits, k=1, dim=-1)[1].squeeze().item()
print(f'Text: {text} | Label : {(i2w[label])} ({F.softmax(logits, dim=-1).squeeze()[label] * 100:.3f}%)')

Text: kalau sekarang aku masih belum bisa kasi komentar | Label : neutral (99.845%)
```

Gambar 7. Prediksi Kalimat Netral

Pada contoh pertama, kalimat “bagus yah kinerjanya, aku suka presiden kita yang sekarang” berhasil diklasifikasikan sebagai sentimen positif dengan probabilitas sebesar 98.94%. Hal ini menunjukkan bahwa model mampu menangkap ekspresi apresiatif dan dukungan terhadap pihak tertentu. Contoh kedua menampilkan kalimat “penipu, presiden gagal mana nih buktinya ga ada sama sekali”, yang diklasifikasikan dengan tepat sebagai negatif dengan tingkat kepercayaan 99.52%. Kalimat ini secara eksplisit memuat unsur ketidakpuasan dan kekecewaan, yang dikenali dengan baik oleh model. Sementara itu, pada contoh ketiga, kalimat “kalau sekarang aku masih belum bisa kasi komentar” diprediksi sebagai netral dengan probabilitas 99.843%. Kalimat ini bersifat informatif dan tidak mengandung muatan emosional yang kuat, sehingga model secara akurat mengenalinya sebagai opini netral. Secara keseluruhan, hasil prediksi terhadap kalimat-kalimat uji ini memperkuat temuan bahwa model IndoBERT mampu mengidentifikasi konteks sentimen dengan tingkat akurasi yang tinggi, baik untuk ekspresi yang eksplisit maupun yang bersifat implisit.

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan dan mengevaluasi model analisis sentimen berbasis IndoBERT untuk menganalisis persepsi publik terhadap kinerja pemerintahan Prabowo-Gibran di platform X hingga 30 Juni 2025. Dengan menggunakan dataset sebanyak 2.612 tweet yang telah melalui preprocessing komprehensif

(pembersihan noise, koreksi slang, normalisasi, dan lemmatisasi), model IndoBERT yang telah di-*fine-tune* mencapai akurasi test sebesar 78%, dengan performa terbaik pada deteksi sentimen negatif (F1-score: 0.82), diikuti oleh sentimen neutral (0.76) dan positif (0.75).

Studi ini juga membandingkan IndoBERT dengan metode baseline SVM + TF-IDF menggunakan kernel polynomial. Hasil menunjukkan bahwa IndoBERT mengungguli SVM secara signifikan dengan akurasi test 78% berbanding 72.19%. Keunggulan IndoBERT terutama terlihat pada deteksi sentimen negatif (F1-score 0.82 vs 0.72), yang menunjukkan kemampuan superior model transformer dalam memahami konteks semantik yang lebih kompleks, termasuk sarkasme dan kritik tersirat pada teks politik berbahasa Indonesia di media sosial. Sementara SVM menunjukkan precision tinggi untuk kelas neutral (0.93) dan positif (0.91), model ini mengalami overfitting yang signifikan (akurasi training 99.24% vs test 72.19%) dan F1-score yang lebih rendah di semua kelas dibandingkan IndoBERT.

Analisis confusion matrix mengungkapkan bahwa model cenderung lebih sering mengklasifikasikan tweet neutral sebagai positif, mengindikasikan bahwa sentimen neutral tetap menjadi tantangan klasifikasi yang paling sulit dalam konteks teks media sosial yang politis dan noisy. Hal ini mencerminkan kompleksitas inheren dalam komunikasi politik digital, di mana batas antara sentimen neutral dan positif sering kali ambigu dan bergantung pada nuansa bahasa yang halus.

Penelitian ini juga mengakui beberapa keterbatasan penting. Pertama, ukuran dataset yang relatif terbatas (2.612 tweet) mungkin belum sepenuhnya mewakili keragaman opini publik yang lebih luas. Kedua, representativitas pengguna platform X tidak mencerminkan seluruh spektrum demografi Indonesia, karena pengguna media sosial cenderung lebih muda, urban, dan berpendidikan tinggi. Ketiga, distribusi kelas sentimen yang tidak seimbang dapat mempengaruhi performa model pada kelas minoritas.

Untuk penelitian mendatang, beberapa arah pengembangan disarankan. Pertama, perluasan dataset dengan jumlah sampel yang lebih besar dan periode waktu yang lebih panjang untuk meningkatkan generalisasi model. Kedua, eksplorasi teknik penanganan ketidakseimbangan kelas yang lebih canggih, seperti *oversampling*, *undersampling*, atau *class weighting*. Ketiga, perbandingan dengan arsitektur transformer lain seperti XLM-RoBERTa atau model bahasa besar (Large Language Models) untuk memperkuat benchmarking. Keempat, pengembangan model analisis sentimen berbasis aspek (*aspect-based sentiment analysis*) untuk memberikan wawasan yang lebih granular tentang sentimen publik terhadap program atau kebijakan spesifik pemerintahan. Kelima, integrasi analisis multimodal yang menggabungkan teks, gambar, dan video untuk pemahaman yang lebih holistik tentang diskursus politik digital.

DAFTAR PUSTAKA

- [1] S. K. R. Indonesia, "Prabowo Subianto dan Gibran Rakabuming Raka Resmi Dilantik Sebagai Presiden dan Wakil Presiden RI," 2024.
- [2] L. F. L. Siregar, "Assessing Prabowo-Gibran's Victory: An Exit-Poll Aftermath Analysis of the 2024 Presidential Election," Fulcrum, 2024, [Online]. Available: <https://fulcrum.sg>
- [3] C. T. C. W. Lim, "Indonesian president Prabowo's first 100 days marked by u-turns, missteps ... and sky-high popularity," Guard., 2025, [Online]. Available: <https://www.theguardian.com>
- [4] T. C. T. B. Hu, "Prabowo's tightrope walk to 2029," East Asia Forum, 2025, [Online]. Available: <https://eastasiaforum.org>
- [5] P. Sayarizki, Hasmawati, and H. Nurrahmi, "Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates," Indones. J. Comput., vol. 9, no. 2, pp. 61–72, 2024, doi: 10.34818/INDOJC.2024.9.2.934.
- [6] A. I. Kamil, O. N. Pratiwi, and D. Witarsyah, "Analisis Sentimen dan Pemodelan Topik terhadap Aplikasi Pembelajaran Online pada Platform Google Play," JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Informatika), vol. 10, no. 2, pp. 836–849, 2025, doi: 10.29100/jipi.v10i2.6023.
- [7] C. Jayadianti and others, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," Ilk. J. Ilm., vol. 15, no. 1, pp. 1–10, 2023, [Online]. Available: <https://jurnal.fikom.umi.ac.id/index.php/ILKOM/article/view/1505>
- [8] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoBERT: Pre-training for Indonesian language understanding," in Proceedings of the 28th International Conference on Computational Linguistics, 2020, pp. 757–770.
- [9] H. Imaduddin and others, "Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 8, pp. 13–22, 2023, doi: 10.14569/IJACSA.2023.0140813.
- [10] A. Romadhony and others, "Sentiment Analysis on a Large Indonesian Product Review Dataset," J. Inf. Syst. Eng. Bus. Intell., vol. 10, no. 1, pp. 167–178, 2024, doi: 10.20473/jisebi.10.1.167-178.
- [11] Habibi and others, "Analysis of Indonesia Politics Polarization before 2019 President Election Using Sentiment Analysis and Social Network Analysis," in 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), 2019, pp. 40–45. doi: 10.5815/ijmecs.2019.11.04.
- [12] M. N. Fattanti, "Twitter dan Masa Depan Politik Indonesia: Analisis Perkembangan Komunikasi Politik Lokal melalui Internet," J. Ilmu Pengetah. dan Teknol. Komun., vol. 16, no. 1, pp. 17–30, 2014.
- [13] A. S. Cahyono, "Pengaruh Media Sosial terhadap Perubahan Sosial Masyarakat di Indonesia," Publiciana, vol. 9, no. 1, pp. 140–157, 2016, doi: 10.36563/publiciana.v9i1.79.
- [14] T. Baharuddin, Z. Qodir, H. Jubba, and A. Nurmandi, "Prediction of Indonesian presidential candidates in 2024 using sentiment analysis and text search on Twitter," Int. J. Commun. Soc., vol. 4, no. 2, pp. 204–213, 2022, doi: 10.31763/ijcs.v4i2.512.
- [15] K. Diantoro, A. Soderi, A. Rohman, and A. T. Sitorus, "Sentiment Analysis of Public Opinion on the 2024 Presidential Election in Indonesia Using Twitter Data with the K-NN Method," Digit. J. Comput. Sci. Appl., vol. 1, no. 1, pp. 1–10, 2023, doi: 10.61978/digitus.v1i1.27.
- [16] D. I. Putri, A. N. Alfian, M. Y. Putra, and P. D. Mulyo, "IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election," J. Appl. Informatics Comput., 2024, doi: 10.30871/jaic.v8i1.7440.
- [17] H. Ahmadian and others, "Hybrid Models for Emotion Classification and Sentiment Analysis in Indonesian Language," Appl. Comput. Intell. Soft Comput., 2024, doi: 10.1155/2024/2826773.
- [18] F. Nurrizky and S. Dwiasnati, "Comparison of Naive Bayes and Support Vector Machine (SVM) Algorithms Regarding The Popularity of Presidential Candidates In The Upcoming 2024 Presidential Election," Comput. Eng. Appl. J., vol. 13, no. 1, pp. 17–28, 2024, doi: 10.18495/comengapp.v13i1.459.

- [19] D. S. Nugroho and others, "Analisis Sentimen Dugaan Pelanggaran Pemilu 2024 Berdasarkan Tweet Menggunakan Algoritma Naïve Bayes Classifier," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 3, pp. 1169–1176, 2024, doi: 10.57152/malcom.v4i3.1496.
- [20] F. A. D. Aryanti, A. Luthfiarta, and D. A. I. Soerozo, "Aspect-Based Sentiment Analysis with LDA and IndoBERT Algorithm on Mental Health App: Riliv," J. Appl. Informatics Comput., 2025, doi: 10.30871/jaic.v9i2.8958.
- [21] P. Subarkah and others, "Indonesian Police in the Twitterverse: A Sentiment Analysis Perspectives," in 2023 IEEE 7th International Conference on Information Technology, Information Systems and Electrical Engineering, 2023.
- [22] L. Geni, E. Yulianti, and D. I. Sensuse, "Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models," J. Ilm. Tek. Elektro Komput. dan Inform., vol. 10, no. 2, 2024.
- [23] S. Hayuningtyas, "Penggunaan Twitter sebagai Media Pencitraan Politik (Studi Netnografi pada Akun Twitter @Gerindra)," 2024.
- [24] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 dari Media Sosial Twitter Menggunakan Python," JURIKOM (Jurnal Ris. Komputer), 2024.
- [25] M. Ahmad, S. Aftab, I. Bashir, and M. S. Hameed, "Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy," J. Big Data, vol. 10, no. 95, 2023, doi: 10.1186/s40537-023-00782-9.
- [26] K. S. Nugroho, "Dasar Text Preprocessing dengan Python," 2021.
- [27] DQLab, "Tahapan Text Preprocessing dalam Teknik Pengolahan Data," 2021.
- [28] M. Yunus, "Basic Text Preprocessing menggunakan NLTK," 2021.
- [29] W. Muliady and H. Widiputra, "Preprocessing of Slang Words for Sentiment Analysis on Public Perceptions in Twitter," in Uncertainty Reasoning and Knowledge Engineering, IntechOpen, 2012. doi: 10.1109/urke.2012.6319524.
- [30] B. Gunawan and others, "Text Preprocessing for Optimal Accuracy in Indonesian Sentiment Analysis Using a Deep Learning Model with Word Embedding," 2021.
- [31] Learn Himpasikom, "Teknik NLP: Mengenal Text Pre-Processing," 2022.
- [32] NLTK Project, "Natural Language Toolkit Documentation."
- [33] K. S. Nugroho, "Text Preprocessing Techniques for Indonesian Language," in Python Data Science Handbook, 2021.
- [34] M. Adriani, J. Asian, B. Nazief, S. M. M. Tahaghoghi, and H. E. Williams, "Stemming Indonesian: A confix-stripping approach," ACM Trans. Asian Lang. Inf. Process., vol. 6, no. 4, pp. 1–33, 2007.
- [35] F. Z. Tala, "A Study of Stemming Effect on Information Retrieval in Bahasa Indonesia," University of Amsterdam, 1999.
- [36] H. Utama, E. Daniati, and A. Masruro, "Weak supervision dengan pendekatan labeling function untuk analisis sentimen pada Twitter," Indones. J. Comput. Sci. Res., vol. 3, no. 1, pp. 49–57, 2024, doi: 10.59095/ijcsr.v3i1.93.
- [37] X. Liu, Y. Zhang, H. Wang, L. Chen, M. Li, and Q. Zhou, "A structured sentiment analysis dataset based on public comments from various domains," Sci. Data, vol. 11, p. Article 234, 2024, doi: 10.1038/s41597-024-03067-8.
- [38] CloudFactory, "Dataset Split in Machine Learning," 2024.
- [39] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, and I. Nogues, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1285–1298, 2016.
- [40] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," J. King Saud Univ. - Comput. Inf. Sci., vol. 36, no. 4, p. 102048, 2024.
- [41] A. Gholam, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation," 2018.
- [42] T. D. Purnomo and J. Sutopo, "Comparison of Pre-trained BERT-based Transformer Models for Regional Language Text Sentiment Analysis in Indonesia," Int. J. Sci. Technol., vol. 3, no. 3, pp. 11–21, 2024, doi: 10.56127/ijst.v3i3.1739.
- [43] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv Prepr. arXiv1810.04805, 2018, doi: 10.48550/arXiv.1810.04805.
- [44] H. M. Ramdhan, M. D. Purbolaksono, and Bunyamin, "Sentiment Analysis of Beauty Product Reviews Using the IndoBERT Method and Naive Bayes Classification," in Proceedings of the 2024 12th International Conference on Information and Communication Technology (ICoICT), 2024, pp. 397–404. doi: 10.1109/ICoICT61617.2024.10698198.
- [45] S. Collins, "How to Fine Tune BERT for Real Time Sentiment Analysis," 2023.
- [46] B. Wilie and others, "IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding," in Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, 2020, pp. 843–857.
- [47] M. N. Farizi, "Implementation of BiLSTM and IndoBERT for Sentiment Analysis of TikTok Reviews," JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform., vol. 10, no. 4, pp. 2332–2344, 2024, doi: 10.29100/jipi.v10i4.5815.
- [48] S. Azhari, N. Raharningsih, R. D. Dana, and Mulyawan, "Peningkatan akurasi analisis sentimen pada aplikasi Loklok dengan metode Naïve Bayes," J. Inform. dan Tek. Elektro Terap., vol. 13, no. 1, pp. 1132–1146, 2025, doi: 10.23960/jitet.v13i1.5848.
- [49] L. Nursingah, Ruuhwan, and T. Mufizar, "Analisis sentimen pengguna aplikasi X terhadap program makan siang gratis dengan metode Naïve Bayes classifier," J. Inform. dan Tek. Elektro Terap., vol. 12, no. 3, pp. 1615–1622, 2024, doi: 10.23960/jitet.v12i3.4336.
- [50] A. F. Sadeli and I. I. Lawanda, "Recall, Precision, and F-Measure for Evaluating Information Retrieval System in Electronic Document Management Systems (EDMS)," Khizanah al-Hikmah J. Ilmu Perpustakaan, Informasi, dan Kearsipan, vol. 11, no. 2, 2023, doi: 10.24252/kah.v11i2a8.
- [51] M. D. Al Fahreza, A. Luthfiarta, M. Rafid, and M. Indrawan, "Analisis sentimen: Pengaruh jam kerja terhadap kesehatan mental generasi Z," J. Appl. Comput. Sci. Technol., vol. 5, no. 1, pp. 16–25, 2024, doi: 10.52158/jacost.v5i1.715.
- [52] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171–4186.
- [53] Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 Task 4: Sentiment analysis in Twitter. Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 502–518.