Jurnal Ilmu Komputer dan Sistem Informasi (JIRSI) **e-ISSN 2830-3954**
Volume: 5, Nomor: 3, September 2026: 793-804 **p-ISSN 2830-6031**
https://jurnal.unity-academy.sch.id/index.php/jirsi/index

Analisis Sentimen Debat Capres dan Cawapres Indonesia 2024 Menggunakan Metode Bert

Sentiment Analysis of the 2024 Indonesian Presidential and Vice-Presidential Debates Using the BERT Method

Mhd. Furqan[1], Aidil Halim Lubis[2], Zaki Musyaffa[3]
Universitas Islam Negeri Sumatera Utara, Fakultas Sains dan Teknologi
Jurusan Ilmu Komputer
E-mail: [1]mhdfurqan@uinsu.ac.id, [2]aidilhalimlubis@uinsu.ac.id, [3]zamusya22@gmail.com

Abstrak

Debat Calon Presiden dan Calon Wakil Presiden (Capres-Cawapres) Indonesia 2024 di kanal YouTube KPU Republik Indonesia memicu ribuan komentar publik yang mencerminkan sentimen masyarakat terhadap para kandidat, namun opini tersebut belum banyak dikaji secara sistematis. Penelitian ini bertujuan menganalisis sentimen komentar YouTube terhadap debat tersebut menggunakan metode Bidirectional Encoder Representations from Transformers (BERT). Sebanyak 2.500 komentar dikumpulkan melalui YouTube Data API dari lima sesi debat, kemudian dibersihkan melalui tahap pre-processing dan diberi label otomatis menggunakan kamus InSetLexicon ke dalam tiga kategori: positif, negatif, dan netral. Setelah penghapusan data duplikat, diperoleh 2.362 data bersih yang dibagi menjadi data latih dan data uji dengan rasio 80:20. Model IndoBERT kemudian di-fine-tuning untuk mengklasifikasikan sentimen komentar. Hasil penelitian menunjukkan distribusi sentimen didominasi oleh komentar positif (1.107), diikuti negatif (660) dan netral (595). Evaluasi model menggunakan confusion matrix menghasilkan akurasi sebesar 76,5%, dengan precision, recall, dan F1-score tertinggi pada kelas positif masing-masing sebesar 87,78%, 85,39%, dan 86,57%. Hasil ini menunjukkan bahwa metode BERT cukup efektif dalam menganalisis sentimen komentar berbahasa Indonesia pada konteks debat politik di platform YouTube, sekaligus memberikan kontribusi berupa pemetaan opini publik terhadap kandidat dan isu-isu debat yang dapat menjadi rujukan bagi penelitian analisis sentimen politik berbasis bahasa Indonesia selanjutnya.

Kata Kunci: *Analisis Sentimen, BERT, Debat Capres Cawapres 2024, InSetLexicon, YouTube*

Abstract

The 2024 Indonesian Presidential and Vice Presidential Candidate Debate, broadcast on the YouTube channel of the Indonesian General Elections Commission (KPU), generated thousands of public comments reflecting societal sentiment toward the candidates, yet these opinions have not been systematically examined. This study aims to analyze the sentiment of YouTube comments on the debate using the Bidirectional Encoder Representations from Transformers (BERT) method. A total of 2,500 comments were collected via the YouTube Data API from five debate sessions, then cleaned through a pre-processing stage and automatically labeled using the InSetLexicon dictionary into three categories: positive, negative, and neutral. After removing duplicate data, 2,362 clean records were obtained and divided into training and testing data with an 80:20 ratio. The IndoBERT model was then fine-tuned to classify comment sentiment. The results show that the sentiment distribution was dominated by positive comments (1,107), followed by negative (660) and neutral (595) comments. Model evaluation using a confusion matrix yielded an accuracy of 76.5%, with the highest precision, recall, and F1-score values in the positive class at 87.78%, 85.39%, and 86.57%, respectively. These findings indicate that the BERT method is fairly effective in analyzing the sentiment of

Indonesian-language comments in the context of political debates on YouTube, while also contributing a mapping of public opinion toward candidates and debate issues that can serve as a reference for future Indonesian-language political sentiment analysis research.

Keywords: *Sentiment Analysis, BERT, 2024 Presidential Debate, InSetLexicon, YouTube*

1. PENDAHULUAN

Bahasa pada hakikatnya berorientasi pada enam fungsi utama berdasarkan perspektif *Roman Jakobson* yakni referensial, emotif, puitis, fatis, konatif, dan metalingual yang bertindak sebagai sarana pembentuk makna terstruktur dalam konteks komunikasi interaktif. Dalam konteks Debat Capres-Cawapres Indonesia 2024, manifestasi keenam fungsi bahasa tersebut tefleksikan secara pragmatis melalui strategi retorika para kandidat: fungsi referensial digunakan untuk memaparkan data konkret, visi-misi, serta program kerja; fungsi emotif mengekspresikan ketegasan moral dan kepedulian emosional; fungsi konatif berperan memersuasi dan mengajak konstituen menentukan pilihan politik; fungsi fatis menjaga alur interaksi dan sapaan antar kandidat; fungsi metalingual menglarifikasi definisi teknis atau istilah kebijakan; serta fungsi puitis dimanfaatkan melalui penggunaan slogan, jargon, dan pantun estetis guna memperkuat daya ingat pemilih[1].

Debat calon presiden dan calon wakil presiden Pemilu 2024 dilaksanakan dalam lima tahapan rangkaian debat publik resmi yang diawali dengan debat pertama calon presiden pada Selasa, 12 Desember 2023. Rangkaian debat publik ini berlangsung sepanjang periode kampanye pemilu dari bulan Desember 2023 hingga awal Februari 2024 sebagai forum penentuan elektabilitas kandidat[2].

Tahapan debat Capres 2024 meliputi penyampaian visi-misi, adu argumentasi atas isu krusial, hingga evaluasi pasca debat melalui analisis survei dan sentimen media sosial. Hasil analisis menunjukkan bahwa debat berdampak signifikan pada dinamika pergeseran elektabilitas: pasangan Prabowo Subianto–Gibran Rakabuming Raka mengalami peningkatan elektabilitas yang dominan (mencapai 45,2% hingga 51,8% pascadebat), sementara pasangan Anies Baswedan–Muhaimin Iskandar dan Ganjar Pranowo–Mahfud MD cenderung mengalami penurunan akibat strategi serangan politik dalam debat yang justru menciptakan bumerang dan meningkatkan simpati publik kepada pasangan lawan[3].

Natural Language Processing (NLP) dijelaskan sebagai teknik dan algoritma pemrosesan bahasa alami yang digunakan untuk menganalisis, memproses, serta mengklasifikasikan data teks berskala besar dari media sosial secara otomatis dan efisien ke dalam kategori sentimen positif, negatif, atau netral. Teknik NLP ini mencakup tahapan pemrosesan teks (*preprocessing*) seperti tokenisasi (memecah teks menjadi unit kata), *stemming* (mengubah kata ke bentuk dasar), serta *lemmatization*, yang dilanjutkan dengan penerapan algoritma *machine learning* atau metode guna mengevaluasi persepsi dan opini publik secara *real-time*[4].

Opinion Mining atau Analisis sentimen dapat diartikan sebagai teknik analisis terhadap data teks dari media sosial (seperti *Youtube*) untuk mengidentifikasi dan mengevaluasi opini, sikap, serta persepsi publik terhadap kasus penelitian dengan mengelompokkannya ke dalam kategori sentimen positif, negatif, atau netral[5].

Analisis ini berperan penting dalam mengevaluasi opini, persepsi, dan kecenderungan respons publik terhadap performa para calon presiden.

Penelitian ini menggunakan metode BERT yang dimana BERT (*Bidirectional Encoder Representations from Transformers*) merupakan model representasi bahasa berarsitektur *Transformer* yang dilatih secara dua arah (*bidirectional*). Metode ini menghasilkan *contextual embeddings* yang dinamis dengan memahami makna kata berdasarkan konteks sebelum dan sesudahnya secara bersamaan. Melalui pendekatan *pre-training* pada korpus teks berskala besar serta proses *fine-tuning*, BERT mampu meningkatkan kinerja berbagai tugas Pemrosesan Bahasa Alami (*Natural Language Processing*) secara signifikan[6].

Peneliti memilih *platform* media sosial *Youtube* sebagai media/instrumen penelitian, Dimana *YouTube* merupakan *platform* media sosial berbasis video populer dan kini telah berkembang menjadi media edukasi yang efektif serta interatif bagi berbagai generasi. *Platform* ini menawarkan berbagai manfaat pembelajaran yang informatif, praktis, hemat biaya, interaktif, serta mudah dibagikan, sehingga mampu memfasilitasi peneliti dalam hal sarana/alat untuk mengumpulkan data[7].

Penelitian analisis sentimen berbahasa Indonesia telah banyak dilakukan pada berbagai domain. Azzahra dan Mailoa menganalisis sentimen masyarakat terhadap RSUD Salatiga menggunakan *Support Vector Machine* (SVM) dengan kernel linear dan *TF-IDF* pada 414 ulasan terproses, menghasilkan akurasi 84,00%, presisi 84,00%, *recall* 83,25%, dan *F1-score* 83,53%, dengan dominasi respons positif sebesar 55,8%[8]. Keterbatasan penelitian tersebut terletak pada pelabelan yang hanya mencakup dua kelas (positif dan negatif) tanpa kelas netral, sebagaimana diakui penulisnya bahwa penambahan kelas netral diharapkan dapat memberikan wawasan yang lebih konkret. Selain itu, ekstraksi fitur TF-IDF bekerja berdasarkan pembobotan frekuensi kata sehingga tidak menangkap makna kontekstual dan urutan kata dalam kalimat.

Pada penggunaan model *deep learning*, Nuryadi dkk. *melakukan fine-tuning IndoBERT* untuk menganalisis sentimen ulasan aplikasi *Tiket.com* di *Google Play Store* dan memperoleh akurasi keseluruhan 91% dengan *F1-score* 93% untuk kelas positif dan 94% untuk kelas negatif, tetapi *recall* kelas netral masih rendah yaitu 65%[9]. Wirayudha dkk. menerapkan metode *IndoBERT* pada 6.000 ulasan aplikasi *Access by KAI* dengan akurasi 76%, di mana nilai pada kelas netral menjadi yang terendah dibandingkan kelas lainnya yaitu 47%[10].

Persoalan yang sama terlihat pada kedua penelitian berbasis *IndoBERT* tersebut, yaitu kelas netral selalu menjadi kelas dengan performa terendah. Ketidakseimbangan kelas (*class imbalance*) pada data membuat kinerja klasifikasi condong atau bias ke arah kelas mayoritas[11], dan metode *deep learning* pun terbukti tetap terpengaruh oleh persoalan ketidakseimbangan kelas[12]. Di sisi lain, koreksi terhadap ketidakseimbangan kelas tidak selalu diperlukan dan bahkan dapat berdampak buruk terhadap keandalan model prediksi, sehingga penanganannya menuntut kehati-hatian[13].

Berdasarkan kajian penelitian terdahulu di atas, terdapat kesenjangan penelitian (*research gap*) yang melandasi penelitian ini. Pertama, dari sisi objek, studi analisis sentimen berbahasa Indonesia sebelumnya didominasi domain komersial dan layanan publik seperti ulasan rumah sakit[8], ulasan aplikasi

transportasi dan perjalanan[9][10], serta tren mata uang kripto di *Twitter*[5], sedangkan kajian terhadap debat Capres dan Cawapres 2024 selama ini berfokus pada aspek kesantunan berbahasa[2] dan elektabilitas kandidat berbasis survei[3], bukan analisis sentimen komputasional terhadap opini publik di media sosial. Kedua, dari sisi metode, penelitian yang masih mengandalkan *machine learning* konvensional seperti *SVM* dan *TF-IDF*[5][8] tidak mampu menangkap makna kontekstual kata, padahal *BERT* terbukti menghasilkan *contextual embeddings* yang meningkatkan kinerja berbagai tugas *Natural Language Processing* secara signifikan[6]. Ketiga, dari sisi permasalahan, penelitian berbasis *IndoBERT* sebelumnya[9][10] sama-sama belum berhasil menangani kelas netral dengan baik akibat ketidakseimbangan kelas[11][12], sementara koreksi ketidakseimbangan yang keliru justru dapat menurunkan keandalan model[13].

Penelitian ini mengisi kesenjangan tersebut melalui tiga kontribusi yang berbeda dari penelitian sebelumnya. Pertama, objek yang dianalisis adalah komentar publik pada lima sesi debat Capres dan Cawapres Indonesia 2024 di kanal *YouTube* resmi KPU Republik Indonesia, sehingga data yang diolah memiliki karakteristik bahasa politik informal yang sarat singkatan dan kata gaul, berbeda dengan ulasan aplikasi maupun layanan publik. Kedua, penelitian ini melakukan finetuning model *IndoBERT* untuk klasifikasi tiga kelas sentimen (positif, negatif, dan netral) dengan pelabelan otomatis berbasis kamus *InSetLexicon*, berbeda dengan penelitian terdahulu yang hanya menggunakan dua kelas dan pelabelan manual[8]. Ketiga, sebagai pendekatan solusi tingkat data, pengambilan komentar dilakukan dalam porsi yang sama pada setiap sesi debat sehingga variasi konteks tiap sesi terwakili secara seimbang dalam dataset. Hasil penelitian ini dapat menjadi sumber informasi penting bagi para pengambil keputusan politik, peneliti, dan praktisi yang terlibat dalam proses pemilihan umum, karena dengan memahami sentimen publik yang muncul selama debat capres dan cawapres, mereka dapat membuat keputusan yang lebih informasional dan responsif terhadap kebutuhan dan keinginan masyarakat.

2. METODOLOGI PENELITIAN

2.1. Waktu dan Jadwal Kerja Penelitian

Penelitian ini dilaksanakan pada semester genap tahun ajaran 2023/2024, dengan waktu dan jadwal yang tercantum dalam tabel berikut:

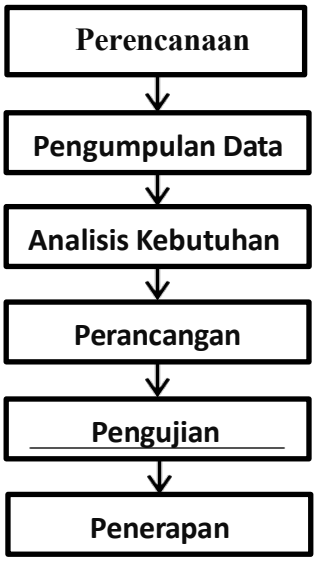
Tabel 1. Waktu dan Jadwal Penelitian

No	Jadwal	Waktu Penelitian			
		Juli	Agustus	September	Oktober
1.	Perencanaan				
2.	Pengumpulan Data				
3.	Analisis Dan Perancangan				
4.	Penerapan				
5.	Pengujian				

2.2. Alur kerja Penelitian

2.2.1 Perencanaan

Perencanaan merupakan gambaran dari tahapan awal penelitian hingga akhir penelitian yang digambarkan kedalam satu diagram yang saling berurutan. Perencanaan penelitian dapat dilihat pada gambar berikut.



Gambar 1. Alur Kerja Penelitian

2.2.2 Pegumpulan Data

A. Studi Pusaka

Dalam penelitian ini peneliti menghimpun informasi dari beragam sumber, termasuk jurnal, skripsi, buku ajar, dan internet sebagai dukungan untuk penelitian ini

B. Observasi & *Scraping* Data

Data yang digunakan dalam penelitian ini didapatkan melalui kolom komentar video debat capres & cawapres indonesia 2024 pada *platfrom YouTube* menggunakan *Google collab,* berikut ringkasan alur kerja di *Google collab:*

- *Install Library*: Menginstal alat ekstraksi (*youtube-comment-downloader*).
- *Execute Script*: Mengambil komentar berdasarkan *URL* video dan membatasi jumlah data jika diperlukan.
- *Data Structuring*: Mengubah format data *JSON/Dictionary* menjadi tabel *Pandas DataFrame*.
- *Export & Download*: Menyimpan ke *.csv* dan mengunduhnya via pustaka *google.colab.files.*

2.2.3 Analisis Kebutuhan

Analisis kebutuhan adalah proses menganalisis spesifikasi, kebutuhan dan kondisi dari suatu sistem yang akan diteliti. Dalam penelitian ini sistem yang dimaksud adalah pemodelan sistem analisis sentimen menggunakan bahasa

pemrogaman *python*. Analisis kebutuhan pada penelitian ini terbagi menjadi 2, yaitu:

A. Kebutuhan Fungsional

Kebutuhan fungsional mencerminkan langkah-langkah yang harus dijalankan pada suatu sistem untuk memastikan bahwa sistem dapat berjalan dengan baik dan memenuhi kebutuhan yang telah ditetapkan. Berikut adalah daftar kebutuhan fungsional yang akan dimplementasikan dalam sistem:

- Sistem dapat menangani proses input *dataset training* dan *dataset testing.*
- Sistem dapat menangani proses *preprocessing data*, diantaranya yaitu *case folding, cleaning, tokenizing, normalization, stopword removal,dan stemming.*
- Sistem dapat menangani proses klasifikasi dan pelabelan sentimen menjadi tiga kelas, yaitu positif, negatif,dan netral.

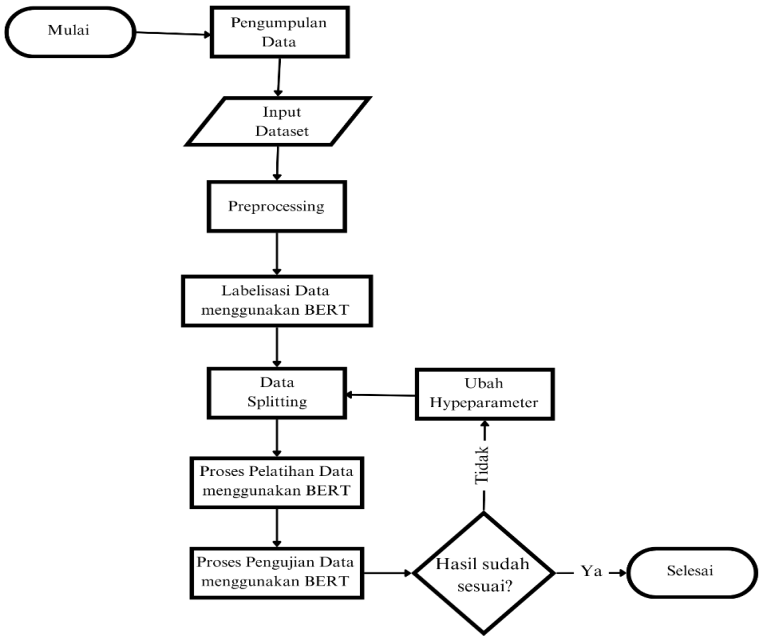
B. Kebutuhan Non-fungsional

Kebutuhan non-fungsional merupakan kebutuhan tambahan untuk mendukung terwujudnya suatu sistem. Berikut kebutuhan non-fungsional yang akan dibangun pada sistem, yaitu:

- Pemodelan sistem nantinya akan dijalankan menggunakan *Google Collab.*
- Pemodelan sistem nantinya akan dibangun dengan menggunakan bahasa pemrograman *python*.

2.2.4 Perancangan

Penelitian ini dirancang dengan membangun *flowchart* sistemnya. Berikut ini merupakan *flowchart* perancangan sistem penelitian:



Gambar 2. *Flowchart* Sistem

2.2.5 Pengujian

A. Pengumpulan Data

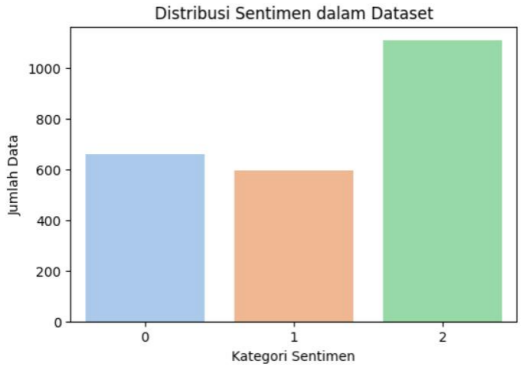
Data Metode ini menggunakan pustaka *Python youtube-comment-downloader* karena tidak memerlukan *API Key*, tidak terbatas kuota harian, dan sangat mudah dijalankan langsung di lingkungan *Colab*. Data diambil dari video Youtube debat dengan menginstal pustaka downloader dan pandas lalu mengganti url video dengan link video *YouTube* debat di akun resmi *@KPURepublikIndonesia.* Setiap sesi diambil sebanyak 500 komentar, sehingga total data yang terkumpul berjumlah 2.500.

B. *Input Dataset* dan *Text Preprocessing*

Data mentah diproses melalui beberapa tahap: (1) *text cleaning* untuk membersihkan karakter tidak relevan seperti tanda baca, angka, dan simbol; (2) *case folding* untuk menyeragamkan huruf menjadi huruf kecil; (3) tokenisasi untuk memecah kalimat menjadi unit kata; (4) normalisasi kata tidak baku atau slang menggunakan kamus gaul yang diperoleh dari *repositori GitHub*; (5) penghapusan *stopword* menggunakan daftar kata dari NLTK; dan (6) *stemming* ke bentuk kata dasar menggunakan algoritma Sastrawi. Setelah seluruh data melalui proses tersebut dan komentar duplikat dihapus, jumlah data bersih yang diperoleh adalah 2.362.

C. Pelabelan Sentimen

Pelabelan sentimen dilakukan secara otomatis menggunakan kamus InSetLexicon, yang berisi daftar kata beserta bobot polaritas positif dan negatif dalam bahasa Indonesia. Skor sentimen suatu komentar dihitung dari akumulasi bobot kata-kata penyusunnya; skor total positif dikategorikan sebagai sentimen positif, skor negatif sebagai sentimen negatif, dan skor nol sebagai sentimen netral. Label kategori kemudian dipetakan ke bentuk numerik: negatif = 0, netral = 1, dan positif = 2 untuk keperluan pelatihan model. Distribusi hasil pelabelan pada 2.362 data ditunjukkan pada Gambar 3.



Gambar 3. Distribusi Label Sentimen Hasil *InSetLexicon*

Gambar menunjukkan bahwa dari 2.362 komentar, sentimen positif mendominasi dengan 1.107 komentar (46,9%), diikuti sentimen negatif sebanyak 660 komentar (27,9%), dan sentimen netral sebanyak 595 komentar (25,2%). Distribusi ini mengindikasikan bahwa mayoritas penonton memberikan tanggapan positif terhadap jalannya debat Capres dan Cawapres 2024.

kemampuan sistem dalam menganalisis sentimen adalah tujuan utama dari pengujian sistem tersebut. Proses ini melibatkan pengambilan sentimen dari data yang diperoleh melalui teknik *scraping*, kemudian melibatkan langkah-langkah seperti *pre-processing*, labelisasi, pembobotan kata dengan menggunakan *library bert- multilingual-uncased,* Sistem yang telah dibangun akan dievaluasi menggunakan metode matriks konfusi sebagai indikator performa.

Pemilihan matriks konfusi sebagai metode pengujian dalam penelitian ini didasarkan pada kegunaannya pada umumnya untuk mengukur tingkat akurasi dalam konteks *data mining.* Matriks konfusi berisikan informasi ketepatan prediksi sentimen dan membuat parameter akurasi, *recall, precision*, dan *f1-score.*

2.2.6 Penerapan

langkah terakhir adalah implementasi dan penerapan model untuk analisis sentimen secara menyeluruh terhadap kolom komentar *YouTube* pada debat capres/cawapres Indonesia 2024.

Implementasi model melibatkan penggunaan model yang telah dilatih untuk melakukan analisis sentimen pada data komentar baru yang diperoleh dari *YouTube*.

A. Fine-Tuning Model *IndoBERT*

Data yang telah berlabel dibagi dengan rasio 80:20 menjadi 1.888 data latih dan 473 data uji. Setiap komentar ditokenisasi menggunakan *tokenizer* bawaan *IndoBERT (indobenchmark/indobert-base-p2),* dengan penambahan token khusus *[CLS]* dan *[SEP],* serta *attention mask* dan *token type ids.* Panjang *sequence* maksimum ditetapkan sebesar 80 token berdasarkan distribusi panjang komentar dalam dataset. Model kemudian *di-fine-tune* menggunakan konfigurasi *hyperparameter* pada Tabel 2.

Tabel 2. Konfigurasi *Hyperparameter Fine-Tuning IndoBERT*

Parameter	Nilai
Pre-trained model	*IndoBERT (indobert-base-p2)*
Jumlah *epoch*	50
Batch size	32
Learning rate	5e-5
Optimizer	Adam
Loss function	*Sparse Categorical Crossentropy*
Max sequence length	80 token

B. Evaluasi Model

Performa model dievaluasi menggunakan *confusion matrix* dengan menghitung akurasi, *precision, recall, dan F1-score* sebagaimana persamaan (1) sampai (4).

$$Akurasi = \frac{(TP+TN)}{Total\ Data} (1)$$

$$Precision = \frac{TP}{(TP+FP)} (2)$$

$$Recall = \frac{TP}{(TP+FN)} (3)$$

$$FI - Score = 2 \times \frac{(Precision \times Recall)}{(Precision+Recall)} (4)$$

Selain itu, hasil analisis sentimen dapat diintegrasikan ke dalam sistem atau *platform* yang lebih besar untuk memberikan wawasan yang lebih dalam tentang dinamika politik dan opini publik.

Dengan implementasi dan penerapan model analisis sentimen yang efektif, kita dapat meningkatkan pemahaman tentang respons publik terhadap debat politik dan memberikan kontribusi yang berarti dalam meningkatkan kualitas diskusi politik di ruang media sosial, terkhusunya media sosial *Youtube*.

3. HASIL DAN PEMBAHASAN

Proses *fine-tuning* dilakukan selama 50 *epoch* menggunakan data latih. Gambar 2 menunjukkan performa model selama pelatihan berdasarkan metrik akurasi dan *loss.*

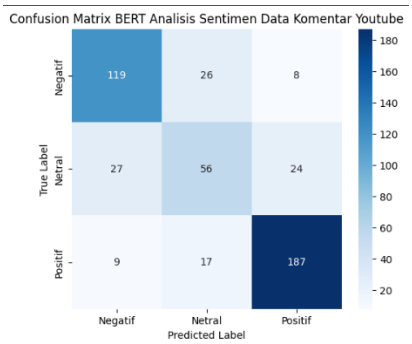
Gambar 4. Grafik Akurasi dan *Loss Model IndoBERT* Selama Pelatihan

Grafik pada Gambar 4 memperlihatkan bahwa akurasi pelatihan meningkat tajam pada *epoch-epoch* awal dan terus naik hingga mendekati 100%, sementara *loss* menurun mendekati nol meskipun terdapat sedikit fluktuasi. Pola ini menunjukkan model belajar dengan baik terhadap data latih, meskipun akurasi yang jauh lebih tinggi pada data latih dibandingkan data uji (76,5%) juga mengindikasikan adanya potensi *overfitting* yang perlu diperhatikan pada penelitian selanjutnya, misalnya melalui regularisasi atau e*arly stopping.*

Evaluasi pada 473 data uji menghasilkan *confusion matrix* pada Tabel 2 dan visualisasinya pada Tabel 3.

Tabel 3. *Confusion Matrix* Hasil Prediksi pada Data Uji

Aktual \ Prediksi	Negatif	Netral	Positif
Negatif	119	26	8
Netral	27	56	24
Positif	9	17	187



Gambar 5. Visualisasi *Confusion Matrix* Model IndoBERT

Berdasarkan Tabel 2, jumlah prediksi benar (diagonal utama) adalah 119 + 56 + 187 = 362 dari total 473 data uji, sehingga akurasi keseluruhan model sebesar 362/473 = 0,765 atau 76,5%. Perhitungan *precision*, *recall*, dan *F1-score* untuk masing-masing kelas ditunjukkan pada Tabel 3.

Tabel 4. Hasil Evaluasi *Precision, Recall,* dan *F1-Score*

Kelas	*Precision*	*Recall*	*F1-Score*
Negatif	77,80%	76,80%	77,28%
Netral	52,34%	56,57%	54,36%
Positif	87,78%	85,39%	86,57%

Tabel 4 menunjukkan bahwa model *IndoBERT* memberikan performa terbaik pada kelas positif dengan *F1-score* sebesar 86,57%, diikuti kelas negatif sebesar 77,28%, dan kelas netral sebagai yang terendah sebesar 54,36%. Rendahnya performa pada kelas netral kemungkinan disebabkan oleh ambiguitas makna serta tumpang tindih kosakata antara komentar netral dengan komentar positif atau negatif, yang merupakan tantangan umum pada pelabelan otomatis berbasis *lexicon*. Dibandingkan penelitian terdahulu yang mencapai akurasi 91,8% [4] dan 95% [6], akurasi 76,5% pada penelitian ini relatif lebih rendah, yang diduga dipengaruhi oleh

karakteristik bahasa informal dan slang pada komentar *YouTube* debat politik, ketidakseimbangan jumlah data antar kelas, serta potensi *label noise* dari pendekatan pelabelan berbasis *lexicon* yang bersifat berbasis aturan *(rule-based)* dan tidak sepenuhnya mempertimbangkan konteks kalimat.

Adapun distribusi prediksi sentimen model pada seluruh data uji menunjukkan jumlah prediksi positif sebanyak 219 komentar, negatif 155 komentar, dan netral 99 komentar, yang sejalan dengan proporsi label aktual pada dataset dan mengonfirmasi bahwa sentimen positif tetap menjadi kecenderungan dominan dalam tanggapan publik terhadap debat Capres dan Cawapres 2024.

KESIMPULAN

Penelitian ini menerapkan model *IndoBERT* untuk menganalisis sentimen 2.362 komentar *YouTube* pada lima sesi debat Capres dan Cawapres Indonesia 2024 di kanal resmi KPU Republik Indonesia, dengan label sentimen yang dihasilkan secara otomatis menggunakan *InSetLexicon*. Hasil pengujian pada 473 data uji menunjukkan akurasi keseluruhan sebesar 76,5%, dengan performa terbaik pada kelas positif *(precision 87,78%, recall 85,39%, F1-score 86,57%)* dan performa terendah pada kelas netral (*F1-score 54,36%).* Distribusi sentimen menunjukkan dominasi opini positif (1.107 komentar) dibandingkan negatif (660 komentar) dan netral (595 komentar). Hasil ini menunjukkan bahwa BERT, melalui varian *IndoBERT*, cukup efektif dalam menangkap konteks polaritas pada komentar politik berbahasa Indonesia, meskipun tantangan berupa bahasa informal, ketidakseimbangan kelas, dan noise dari pelabelan berbasis *lexicon* masih perlu diatasi.

Untuk penelitian selanjutnya, disarankan memperluas cakupan data dengan mengambil komentar dari berbagai platform media sosial lain seperti *Twitter*, *Instagram*, atau *Facebook* guna memperoleh perspektif yang lebih beragam. Penggunaan model lain seperti *RoBERTa* atau varian *transformer* berbahasa Indonesia lainnya juga dapat dieksplorasi untuk dibandingkan performanya dengan *IndoBERT*, disertai penanganan ketidakseimbangan kelas dan evaluasi lebih lanjut terhadap kualitas pelabelan otomatis. Pengembangan *dashboard* visualisasi interaktif juga akan mempermudah penyajian tren sentimen dari waktu ke waktu maupun perbandingan sentimen antar kandidat secara lebih komprehensif.

DAFTAR PUSTAKA

[1] M. Luthfialana, M. Hasyim, and A. M. A. Anshory, "Fungsi Bahasa dalam Cerpen Berjuta Rasanya Karya Tere Liye: Perspektif Roman Jakobson," *Bahasa: Jurnal Keilmuan Pendidikan Bahasa dan Sastra Indonesia*, vol. 6, no. 1, pp. 1–11, 2024, doi: 10.26499/bahasa.v6i1.744.

[2] U. Khasanah *et al.*, "Pelanggaran Kesantunan Berbahasa dalam Debat Capres dan Cawapres 2024 Relevansinya sebagai Sumber Bahan Ajar," *Hortatori: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, vol. 8, no. 2, pp. 256–265, 2024.

[3] Citra Widayanti and Yulita Nilam Fridiyanti, “Analisis Pengaruh Debat Calon Presiden 2024 Pertama Terhadap Elektabilitas Calon Presiden Perspektif

Pandangan Masyarakat," *Journal of Social and Economics Research*, vol. 5, no. 2, pp. 1720–1731, 2024, doi: 10.54783/jser.v5i2.259.

[4] G. Prima Ertansyah, R. Tri Cahya Kusuma, and A. Arum Sari, "Analisis Sentimen Pada Media Sosial Menggunakan Teknik Natural Language Processing," *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis*, pp. 183–189, 2025, doi: 10.47701/qgcey104.

[5] A. A. Corrs, A. Syam, and V. Aris, "Analisis Sentimen Tren Cryptocurrency Menggunakan Machine Learning," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 3, no. 4, pp. 59–66, 2025, doi: 10.31004/riggs.v3i4.486.

[6] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, and B. Alshemaimri, "BERT applications in natural language processing: a review," *Artificial Intelligence Review*, vol. 58, no. 6, 2025, doi: 10.1007/s10462-025-11162-5.

[7] Tri Ayu Mareta, Desty Endrawati Subroto, Lailaturrohmah Aulia, Siti Nuryanah, and Ratu Najwa Fadilah, "Peran Media Sosial Youtube sebagai Media Edukasi dalam Pendidikan Generasi Z," *Guruku: Jurnal Pendidikan dan Sosial Humaniora*, vol. 3, no. 1, pp. 98–106, 2025, doi: 10.59061/guruku.v3i1.894.

[8] W. L. Azzahra and E. Mailoa, "Analisis Sentimen terhadap RSUD Salatiga Menggunakan SVM dan TF-IDF," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 6, no. 1, pp. 478–489, 2025, doi: 10.35870/jimik.v6i1.1208.

[9] D. Nuryadi *et al.*, "Fine Tuning Indobert Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Tiket.Com Di Google Play Store," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 3577–3583, 2025, doi: 10.36040/jati.v9i2.13204.

[10] A. Wirayudha, M. Murniyati, and R. Rosdiana, "Analisis Sentimen Terhadap Ulasan Access By KAI Pada Google Play Store Menggunakan Metode Indobert," *Portal Riset dan Inovasi Sistem Perangkat Lunak*, vol. 3, no. 1, pp. 9–20, 2025, doi: 10.59696/prinsip.v3i1.69.

[11] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in big data," *Journal of Big Data*, vol. 5, no. 1, 2018, doi: 10.1186/s40537-018-0151-6.

[12] K. Ghosh, C. Bellinger, R. Corizzo, P. Branco, B. Krawczyk, and N. Japkowicz, *The class imbalance problem in deep learning*, vol. 113, no. 7. Springer US, 2024. doi: 10.1007/s10994-022-06268-8.

[13] A. Carriero, K. Luijken, A. de Hond, K. G. M. Moons, B. van Calster, and M. van Smeden, "The Harms of Class Imbalance Corrections for Machine Learning Based Prediction Models: A Simulation Study," *Statistics in Medicine*, vol. 44, no. 3–4, 2025, doi: 10.1002/sim.10320.