Jurnal Sains Informatika Terapan (JSIT)

E-ISSN: 2828-1659, Volume: 5, Issue: 1, Month: February, Year: 2026
Page: 243 - 251, Available online at: https://rcf-indonesia.org/home/

ANALISIS SENTIMEN ZOOM CLOUD MEETINGS DI PLAY STORE MENGGUNAKAN METODE INDOBERT

Nuraeni Herlinawati[1)], Indah Ariyati[2)], Samudi[3)], Anggun Yuli Asih[4)]

[1,2,3]Sistem Informasi, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika
[4]Informatika, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika
Corresponding Author: [1] nuraeni.nhw@bsi.ac.id

Article Info

Article history:

Received: Jan 12, 2026
Revised: Jan 25, 2026
Accepted: Jan 30, 2026
Published: Feb 01, 2026

Keywords:

Analisis Sentimen
Zoom Cloud Meetings
Google Play Store
IndoBERT
Transformer

ABSTRACT

Pandemi global COVID-19 telah mendorong lonjakan penggunaan aplikasi konferensi video, dengan Zoom Cloud Meetings menjadi salah satu *platform* dominan. Ulasan pengguna di Google Play Store merupakan sumber informasi berharga mengenai persepsi dan kepuasan pengguna. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi Zoom Cloud Meetings di Google Play Store menggunakan metode IndoBERT, sebuah model pra-latih berbasis Transformer yang dikhususkan untuk bahasa Indonesia. Data ulasan dikumpulkan melalui teknik *web scraping*, kemudian diproses melalui tahap pembersihan, pelabelan berdasarkan rating, dan tokenisasi. Model IndoBERT selanjutnya di-*fine-tuning* untuk tugas klasifikasi sentimen tiga kelas: positif, netral, dan negatif. Evaluasi performa model menggunakan metrik akurasi, presisi, *recall*, dan F1-*score*. Hasil penelitian menunjukkan bahwa model IndoBERT yang telah di-*fine-tuning* mampu mencapai tingkat akurasi sebesar 92%, dengan F1-*score* rata-rata tertimbang sebesar 91%. Analisis lebih lanjut terhadap matriks konfusi menunjukkan performa klasifikasi yang baik untuk ketiga kelas sentimen, meskipun terdapat sedikit kebingungan dalam membedakan ulasan netral dari yang negatif atau positif. Penelitian ini mengonfirmasi efektivitas IndoBERT dalam menganalisis sentimen ulasan pengguna berbahasa Indonesia dan memberikan wawasan berharga bagi pengembang Zoom dalam memahami umpan balik pengguna.

1. INTRODUCTION

Pandemi virus corona (COVID-19) yang melanda global sejak awal 2020 telah membawa perubahan signifikan dalam berbagai aspek kehidupan manusia, terutama dalam cara kita berinteraksi dan bekerja. Salah satu kebijakan pemerintah yang diterapkan di banyak negara, termasuk Indonesia, adalah pembatasan sosial dan *physical distancing* untuk menekan penyebaran virus [1]. Kebijakan ini mendorong implementasi *work from home* (WFH) dan *study from home* (SFH) secara masif, menjadikan aktivitas virtual sebagai kebutuhan pokok [2]. Dalam konteks ini, aplikasi konferensi video memainkan peran krusial, dan Zoom Cloud Meetings muncul sebagai salah satu platform yang paling populer dan banyak digunakan untuk mendukung kegiatan belajar mengajar daring, rapat kerja, dan berbagai interaksi virtual lainnya. Zoom menawarkan berbagai fitur seperti video & audio konferensi, kolaborasi, *chat*, webinar, *shared screen*, dan *shared file* yang dapat diakses dari *desktop* maupun *smartphone*, menjadikannya solusi andal untuk kebutuhan komunikasi jarak jauh [3].

Google Play Store, sebagai gudang aplikasi resmi untuk sistem operasi Android, menyediakan fitur rating dan ulasan yang memungkinkan pengguna untuk memberikan umpan balik terhadap aplikasi yang mereka gunakan [4]. Ulasan-ulasan ini, yang berupa teks, mengandung informasi berharga mengenai pengalaman, kepuasan, maupun keluhan pengguna. Analisis sentimen, yang merupakan bagian dari Natural Language Processing (NLP), adalah proses komputasional untuk memahami, mengekstraksi, dan mengolah data tekstual secara otomatis guna mendapatkan informasi sentimen (opini, perilaku, dan emosi) yang terkandung di dalamnya [5].

Dengan menganalisis sentimen ulasan pengguna, pengembang aplikasi seperti Zoom dapat memperoleh wawasan mendalam mengenai aspek-aspek yang disukai, dianggap kurang, atau perlu diperbaiki, yang pada gilirannya dapat digunakan untuk meningkatkan kualitas layanan dan kepuasan pengguna.

Berbagai metode telah digunakan untuk analisis sentimen, mulai dari pendekatan berbasis leksikon hingga pembelajaran mesin (*machine learning*) seperti Naïve Bayes (NB) dan Support Vector Machine (SVM). Penelitian sebelumnya membandingkan performa NB dan SVM untuk analisis sentimen ulasan Zoom Cloud Meetings di Google Play Store. Memperoleh hasil bahwa SVM mengungguli NB dengan akurasi 81,22% dibandingkan 74,37% menggunakan dataset 1.007 ulasan [1].

Meskipun demikian, dengan kemajuan dalam bidang NLP, khususnya model berbasis Transformer seperti BERT (Bidirectional Encoder Representations from Transformers) dan varian bahasa spesifiknya seperti IndoBERT, potensi untuk mencapai performa analisis sentimen yang lebih tinggi pada teks berbahasa Indonesia menjadi sangat menjanjikan [6]. IndoBERT, yang merupakan varian BERT yang di-pre-*train* menggunakan korpus teks Bahasa Indonesia yang sangat besar (lebih dari 220 juta kata), telah menunjukkan performa superior dalam berbagai tugas NLP berbahasa Indonesia, termasuk analisis sentimen, dibandingkan model-model klasik atau bahkan BERT multilingual [7].

Penelitian oleh Sayarizki et al. (2024) berhasil menerapkan IndoBERT untuk analisis sentimen terhadap calon presiden Indonesia di Twitter dengan akurasi keseluruhan sebesar 80% [8]. Sementara itu, Nuryadi et al. (2025) menggunakan IndoBERT yang di-*fine-tuning* untuk analisis sentimen ulasan aplikasi Tiket.com di Google Play Store dan mencapai akurasi 91% [7]. Andhika et al. (2025) bahkan melaporkan bahwa kombinasi fitur dari IndoBERT dan GloVe dapat sedikit meningkatkan performa analisis sentimen ulasan Zoom, mencapai akurasi 91% [6].

Berdasarkan potensi yang ditunjukkan oleh IndoBERT dan pentingnya memahami sentimen pengguna terhadap aplikasi Zoom Cloud Meetings, terutama dalam konteks penggunaannya yang masif di Indonesia, penelitian ini bertujuan untuk:

1. Mengumpulkan dataset ulasan pengguna aplikasi Zoom Cloud Meetings dari Google Play Store.
2. Melakukan pra-pemrosesan data dan pelabelan sentimen.
3. Menerapkan metode IndoBERT untuk membangun model klasifikasi sentimen.
4. Mengevaluasi performa model IndoBERT dalam menganalisis sentimen ulasan Zoom Cloud Meetings.
5. Mendiskusikan implikasi hasil analisis bagi pengembangan aplikasi Zoom Cloud Meetings.

Penelitian ini diharapkan dapat memberikan kontribusi dalam penerapan model NLP tingkat lanjut untuk analisis sentimen berbahasa Indonesia, khususnya pada domain ulasan aplikasi, sekaligus memberikan wawasan praktis bagi para pemangku kepentingan terkait.

2. MATERIALS AND METHODS

2.1. Analisis Sentimen

Analisis sentimen, juga dikenal sebagai *opinion mining*, adalah salah satu cabang dari Natural Language Processing (NLP) yang fokus pada identifikasi dan ekstraksi opini, sikap, dan emosi yang diekspresikan dalam teks [9]. Tujuan utamanya adalah untuk mengkategorikan sentimen yang diekspresikan dalam sebuah dokumen teks (seperti ulasan, tweet, atau komentar) menjadi beberapa kelas, umumnya positif, negatif, dan netral. Proses ini sangat berharga bagi bisnis dan organisasi untuk memahami persepsi publik terhadap produk, layanan, atau isu tertentu secara real-time dan skala besar [10]. Analisis sentimen merupakan proses otomatis untuk mengekstrak informasi tentang sentimen dalam teks, guna mengidentifikasi pandangan positif atau negatif atau netral terkait suatu topik [11]. Analisis sentimen dapat dilakukan pada berbagai tingkat granularitas, mulai dari tingkat dokumen (keseluruhan teks), kalimat, hingga aspek spesifik dalam teks (*aspect-based sentiment analysis*).

Secara umum, proses analisis sentimen melibatkan beberapa tahap krusial:

1. Pengumpulan Data: Mengumpulkan data tekstual dari sumber-sumber relevan, seperti media sosial (Twitter, Facebook), blog, forum, ulasan produk (Amazon, Google Play Store), atau situs berita. Teknik seperti web *scraping* seringkali digunakan untuk ekstraksi data ini [6].
2. Pra-pemrosesan Teks (*Text Preprocessing*): Tahap ini membersihkan dan menyiapkan data teks mentah untuk analisis lebih lanjut. Aktivitasnya meliputi pengubahan huruf menjadi huruf kecil (*case folding*), penghapusan tanda baca, angka, karakter khusus, *stopwords* (kata-kata umum yang tidak memiliki makna sentimen yang signifikan), *stemming* (pengembalian kata ke bentuk dasarnya), dan tokenisasi (pemecahan teks menjadi kata-kata atau token individual) [12].
3. Ekstraksi Fitur (*Feature Extraction*): Mengubah teks yang telah diproses menjadi representasi numerik yang dapat dimengerti oleh algoritma pembelajaran mesin. Pendekatan tradisional meliputi *Bag-of-Words* (BoW) atau TF-IDF (*Term Frequency-Inverse Document Frequency*). Namun, dengan kemajuan *deep learning*, representasi *word embedding* seperti Word2Vec, GloVe, atau representasi kontekstual dari model seperti BERT lebih disukai karena mampu menangkap makna semantik dan konteks kata [6].
4. Pelabelan Sentimen: Memberikan label sentimen (positif, negatif, netral) pada setiap data teks. Pelabelan dapat dilakukan secara manual (oleh manusia) atau semi-otomatis (berdasarkan *rating* bintang di ulasan aplikasi, di mana *rating* 4-5 bintang dianggap positif, 3 netral, dan 1-2 negatif) [7].

5. Pemodelan dan Klasifikasi: Membangun model pembelajaran mesin untuk mempelajari pola dari data yang telah dilabeli dan mengklasifikasikan data baru ke dalam kategori sentimen yang sesuai. Berbagai algoritma dapat digunakan, mulai dari yang sederhana seperti Naïve Bayes, Support Vector Machine (SVM), hingga model deep learning seperti Jaringan Saraf Tiruan (ANN), Long Short-Term Memory (LSTM), dan Transformer-based models seperti BERT [8].
6. Evaluasi Model: Mengukur performa model yang telah dibangun menggunakan metrik evaluasi seperti akurasi, presisi, *recall*, F1-*score*, dan matriks konfusi. Metrik-metrik ini membantu menilai seberapa baik model dalam mengklasifikasikan sentimen dengan benar [7].

Analisis sentimen memiliki berbagai aplikasi praktis, termasuk *brand monitoring*, *customer feedback analysis*, analisis pasar, pemahaman opini publik terhadap kebijakan publik atau figur publik, dan banyak lagi [13]. Dalam konteks aplikasi seperti Zoom Cloud Meetings, analisis sentimen ulasan pengguna dapat membantu mengidentifikasi fitur-fitur yang paling dihargai, masalah-masalah teknis yang sering dikeluhkan, serta tren kepuasan pengguna dari waktu ke waktu.

2.2. IndoBERT

Bidirectional Encoder Representations from Transformers (BERT) adalah model pembelajaran mesin berbasis Transformer yang revolusioner yang diperkenalkan oleh Devlin et al. (2019) [14]. BERT dirancang untuk melakukan pre-*training* representasi bahasa yang mendalam dan bidireksional dari teks tak berlabel dengan mengkondisikan konteks kiri dan kanan secara bersamaan di semua lapisan. Arsitektur inti BERT terdiri dari tumpukan (*stack*) encoder Transformer, yang memanfaatkan mekanisme *self-attention* untuk menimbang (*weight*) pentingnya kata-kata berbeda dalam sebuah kalimat saat membangun representasi kontekstual untuk setiap kata. Kemampuan BERT untuk memahami konteks dan nuansa bahasa membuatnya sangat efektif untuk berbagai tugas NLP [15].

Revolusi transformer telah mengubah NLP secara signifikan. Model BERT menjadi penting karena kemampuannya memahami konteks kata secara dua arah, menghasilkan representasi bahasa yang lebih kaya dan kontekstual. Kemampuan ini telah meningkatkan kinerja dalam berbagai tugas NLP, termasuk analisis sentimen. Untuk Bahasa Indonesia, model seperti IndoBERT dikembangkan dengan pelatihan pada korpus data lokal, menangkap struktur bahasa yang relevan. Penelitian terbaru menunjukkan bahwa IndoBERT dan variannya, seperti IndoBERTweet untuk media sosial, telah mengalahkan metode tradisional dalam analisis sentimen Bahasa Indonesia [12].

IndoBERT adalah varian spesifik dari model BERT yang dikembangkan khusus untuk bahasa Indonesia. Model ini di-pre-*train* menggunakan korpus teks Bahasa Indonesia yang sangat besar dan bersih (Indo4B), yang terdiri dari lebih dari 220 juta kata dari berbagai sumber, termasuk Wikipedia Bahasa Indonesia (74 juta kata), artikel berita dari Kompas, Tempo dan Liputan6 (55 juta kata), serta Indonesia Web Corpus (90 juta kata). Dengan demikian, IndoBERT telah mempelajari pola-pola bahasa Indonesia yang kaya dan spesifik, membuatnya lebih unggul dalam menangani tugas-tugas NLP berbahasa Indonesia dibandingkan model BERT multilingual umum atau model yang di-pre-*train* dengan korpus campuran.

Proses pelatihan IndoBERT mirip dengan BERT, yang terdiri dari dua tahap utama:

1. Pre-*training*: Pada tahap ini, model IndoBERT dilatih pada data tak berlabel dalam jumlah sangat besar menggunakan dua tugas utama:

a. *Masked Language Modeling* (MLM): Sebagian kata dalam input teks secara acak "ditutupi" (*masked*), dan model harus memprediksi kata-kata yang ditutupi tersebut berdasarkan konteks sekitarnya. Tugas ini memaksa model untuk memahami hubungan antar kata dan struktur kalimat secara mendalam.

b. *Next Sentence Prediction* (NSP): Model diberi dua kalimat dan harus memprediksi apakah kalimat kedua mengikuti kalimat pertama dalam teks asli. Tugas ini membantu model memahami hubungan antar kalimat.

2. *Fine-tuning*: Setelah pre-*training*, model IndoBERT yang telah terlatih kemudian dapat diadaptasi (di-*fine-tuning*) untuk tugas NLP spesifik seperti analisis sentimen, klasifikasi teks, *named entity recognition*, atau *question answering*. Pada tahap ini, model diinisialisasi dengan bobot dari pre-*training* dan kemudian dilatih lebih lanjut menggunakan dataset yang lebih kecil dan berlabel untuk tugas target. Lapisan tambahan untuk klasifikasi biasanya ditambahkan di atas model BERT dasar.

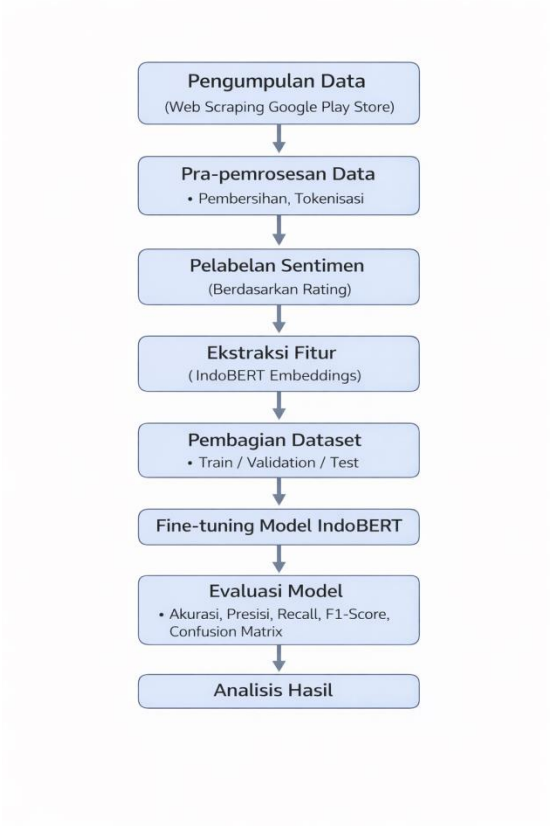
Keunggulan IndoBERT untuk analisis sentimen berbahasa Indonesia antara lain:

- Pemahaman Kontekstual: Mampu menangkap makna kata berdasarkan konteks kalimat secara keseluruhan, bukan hanya makna leksikal tunggal. Ini sangat penting dalam bahasa yang kaya akan ironi, sarkasme, dan ambiguitas.
- Representasi Semantik Kaya: Menghasilkan representasi vektor yang kaya akan informasi semantik untuk kata, frasa, dan kalimat.
- Kinerja Superior: Telah terbukti secara konsisten mengungguli model-model tradisional (seperti NB, SVM) dan beberapa model pre-*train* lainnya dalam berbagai *benchmark* dan tugas NLP berbahasa Indonesia [8].
- Kemampuan Menangani Slang dan Tidak Baku: Berkat tokenisasi sub-kata (*WordPiece*) yang digunakan, IndoBERT dapat menangani kata-kata

tidak baku, slang, atau istilah baru yang mungkin tidak ada dalam kamus standar.

Penelitian-penelitian sebelumnya telah menunjukkan keberhasilan penerapan IndoBERT untuk analisis sentimen berbagai topik dalam bahasa Indonesia, mulai dari ulasan aplikasi [7], opini terhadap calon presiden [8], hingga opini terkait isu-isu publik lainnya. Keberhasilan ini menjadikan IndoBERT sebagai pilihan utama untuk penelitian analisis sentimen ulasan Zoom Cloud Meetings berbahasa Indonesia.

Penelitian ini mengikuti serangkaian tahapan metodologi yang sistematis untuk mencapai tujuannya, yaitu menganalisis sentimen ulasan pengguna Zoom Cloud Meetings di Google Play Store menggunakan model IndoBERT. Alur penelitian secara umum dapat dilihat pada Gambar 1. Tahapan-tahapan tersebut meliputi pengumpulan data, pra-pemrosesan data, pelabelan sentimen, ekstraksi fitur menggunakan IndoBERT, pembagian dataset, pelatihan dan *fine-tuning* model, evaluasi performa model, dan analisis hasil.



Gambar 1. Alur Metodologi Penelitian

2.3. Pengumpulan Data

Tahap awal penelitian ini adalah pengumpulan data ulasan pengguna aplikasi Zoom Cloud Meetings dari Google Play Store. Pengumpulan data dilakukan menggunakan teknik web *scraping* dengan memanfaatkan pustaka google-play-scraper dalam bahasa pemrograman Python. Pustaka ini memungkinkan ekstraksi data ulasan, termasuk konten teks ulasan, *rating* (bintang), tanggal ulasan, dan informasi relevan lainnya. Penelitian ini menargetkan pengambilan 5000 ulasan terbaru dari aplikasi Zoom Cloud Meetings yang tersedia di Google Play Store Indonesia. Pemilihan jumlah ini mempertimbangkan kecukupan data untuk proses training dan evaluasi model, sekaligus menjaga efisiensi komputasi. Data yang telah dikumpulkan kemudian disimpan dalam format Comma Separated Values (CSV) untuk memudahkan proses selanjutnya. Rentang waktu pengambilan data mencakup periode terkini hingga tanggal tertentu untuk memastikan relevansi ulasan dengan kondisi aplikasi terbaru.

2.4. Pra-pemrosesan Data (Data *Preprocessing*)

Data mentah hasil *scraping* seringkali mengandung *noise* atau informasi yang tidak relevan untuk analisis sentimen. Oleh karena itu, tahap pra-pemrosesan data sangat krusial untuk membersihkan dan menyiapkan data sebelum digunakan untuk melatih model. Tahapan pra-pemrosesan yang dilakukan dalam penelitian ini meliputi:

1. Seleksi Kolom: Memilih kolom-kolom yang relevan dari dataset, yaitu kolom *content* (teks ulasan) dan *score* (rating bintang). Kolom lain seperti userName, at (waktu), reviewCreatedVersion, thumbsUpCount, replyContent, dan lainnya tidak digunakan dalam analisis sentimen inti.
2. Penanganan Data Hilang (*Missing Values*): Memeriksa dan menangani data yang hilang, terutama pada kolom *content*. Jika terdapat ulasan tanpa teks, baris tersebut akan dihapus karena tidak dapat dianalisis sentimennya.
3. Pembersihan Teks (*Text Cleaning*):
   - *Case Folding*: Mengubah semua huruf dalam teks ulasan menjadi huruf kecil (*lowercase*) untuk memastikan konsistensi dan menghindari duplikasi kata yang sama karena perbedaan kapitalisasi.
   - Penghapusan Karakter Non-ASCII dan Emoji: Menghapus karakter-karakter khusus, simbol, dan emoji yang tidak berkontribusi pada analisis sentimen dan dapat mengganggu proses tokenisasi.
   - Penghapusan URL, *Mention*, dan *Hashtag*: Menghapus tautan situs web (URL), *mention* pengguna (@username), dan *hashtag* (#topik) yang umum ditemukan dalam ulasan media sosial dan umumnya tidak relevan untuk sentimen inti terhadap aplikasi.
   - Pembersihan Spasi Berlebihan: Menghapus spasi ganda atau spasi di awal/akhir kalimat yang tidak diperlukan.
   - Normalisasi Kata Slang/Alay (Opsional, tergantung sumber daya): Mengganti kata-kata slang atau tidak baku dengan bentuk formalnya. Tahap ini dapat meningkatkan kualitas analisis namun memerlukan kamus slang yang komprehensif. Dalam penelitian

ini, tahap ini tidak dilakukan secara ekstensif mengingat kemampuan IndoBERT dalam menangani variasi bahasa.

4. Tokenisasi: Memecah teks ulasan yang telah dibersihkan menjadi unit-unit yang lebih kecil, biasanya kata-kata (tokens). Tokenisasi merupakan langkah penting sebelum teks dapat diproses oleh model seperti IndoBERT.

Hasil dari tahap pra-pemrosesan adalah dataset yang lebih bersih dan terstruktur, siap untuk dilabeli dan digunakan dalam pelatihan model.

2.5. Pelabelan Sentimen

Setelah data teks ulasan dibersihkan, tahap selanjutnya adalah memberikan label sentimen pada setiap ulasan. Pelabelan dalam penelitian ini dilakukan secara otomatis berdasarkan nilai *rating* (bintang) yang diberikan oleh pengguna di Google Play Store. Aturan pelabelan yang digunakan adalah sebagai berikut:

- Positif: Jika rating bintang adalah 4 atau 5.
- Netral: Jika rating bintang adalah 3.
- Negatif: Jika rating bintang adalah 1 atau 2.

Pendekatan ini umum digunakan dalam analisis sentimen ulasan aplikasi dan dianggap cukup andal sebagai proksi untuk sentimen pengguna. Setelah pelabelan, distribusi kelas sentimen dalam dataset akan dianalisis untuk memahami proporsi masing-masing sentimen. Jika terjadi ketidakseimbangan kelas yang signifikan (misalnya, satu kelas sangat dominan), hal ini perlu dicatat dan mungkin memerlukan strategi khusus seperti *oversampling* atau *undersampling* selama pelatihan model, meskipun model seperti BERT umumnya cukup tangguh terhadap ketidakseimbangan kelas sedang. Penelitian ini akan mencatat distribusi kelas dan mempertimbangkan pengaruhnya terhadap hasil.

2.6. Ekstraksi Fitur dengan IndoBERT

Inti dari analisis sentimen menggunakan IndoBERT adalah kemampuannya untuk menghasilkan representasi vektor kontekstual dari teks ulasan. Proses ekstraksi fitur melibatkan:

1. Tokenisasi IndoBERT: Teks ulasan yang telah dibersihkan dan dilabeli akan di-tokenisasi menggunakan tokenizer khusus yang sesuai dengan model IndoBERT yang digunakan (BertTokenizer). Tokenizer ini akan memecah teks menjadi sub-kata (*subwords*) sesuai dengan kosakata yang dipelajari IndoBERT selama pre-*training*, serta menambahkan token khusus seperti [CLS] (di awal *sequence*) dan [SEP] (di akhir *sequence* atau sebagai pemisah kalimat).
2. Konversi ke Input IDs dan Attention Masks: Token-token tersebut kemudian dikonversi menjadi ID numerik (input_ids) yang sesuai dengan kamus IndoBERT. Selain itu, attention_mask juga dibuat untuk membedakan antara token aktual (mask=1) dan token padding (mask=0) yang ditambahkan untuk menyamakan panjang *sequence* dalam satu *batch*.
3. Generasi *Embeddings*: input_ids dan attention_masks ini menjadi masukan bagi model IndoBERT. Model IndoBERT akan memproses masukan ini melalui lapisan-lapisan Transformer-nya dan menghasilkan representasi vektor (*embeddings*) untuk setiap token. Untuk tugas klasifikasi *sequence* seperti analisis sentimen, representasi vektor dari token [CLS] (token pertama di *output* model) biasanya digunakan sebagai representasi agregat dari seluruh *sequence* teks. Vektor [CLS] ini dianggap telah mengandung informasi kontekstual yang kaya dari keseluruhan kalimat dan akan digunakan sebagai fitur untuk klasifikasi sentimen.

Panjang maksimum *sequence* (*max_length*) perlu ditentukan. Ulasan yang lebih pendek akan di-*padding*, sedangkan yang lebih panjang akan dipotong. Nilai *max_length* yang umum digunakan adalah 128 atau 256 token, tergantung pada panjang rata-rata ulasan dalam dataset.

2.7. Pembagian Dataset

Dataset yang telah diproses (teks ulasan bersih, label sentimen, dan fitur IndoBERT *embeddings*) kemudian dibagi menjadi tiga bagian:

1. Data Latih (*Training* Set): Sebagian besar dataset (80%) digunakan untuk melatih model IndoBERT. Model akan mempelajari pola-pola hubungan antara representasi teks (*embeddings*) dan label sentimen dari data ini.
2. Data Validasi (*Validation* Set): Sebagian kecil dataset (10%) digunakan selama proses pelatihan untuk mengevaluasi performa model pada data yang tidak terlihat selama pelatihan dan membantu dalam *hyperparameter tuning* (*learning rate*, jumlah epoch) serta mencegah *overfitting*.
3. Data Uji (*Test* Set): Sebagian dataset (10%) disisihkan dan tidak digunakan sama sekali selama pelatihan atau *tuning*. Data ini hanya digunakan di akhir penelitian untuk memberikan evaluasi akhir dan objektif terhadap performa model yang telah dilatih sepenuhnya.

Pembagian dataset dilakukan secara acak (*stratified sampling*) untuk memastikan distribusi kelas sentimen yang seimbang di ketiga bagian tersebut.

2.8. Pelatihan dan *Fine-tuning* Model IndoBERT

Model IndoBERT yang telah di-pre-*train* akan diadaptasi (di-*fine-tuning*) untuk tugas klasifikasi sentimen spesifik ini. Langkah-langkahnya meliputi:

1. Pemilihan Model Arsitektur: Memilih varian IndoBERT yang sesuai. Model base (12 lapisan tersembunyi, 768 unit tersembunyi) umumnya menjadi pilihan yang baik untuk keseimbangan performa dan kompleksitas komputasi.

2. Penambahan Lapisan Klasifikasi: Di atas model IndoBERT dasar, sebuah lapisan klasifikasi baru ditambahkan. Lapisan ini biasanya terdiri dari lapisan *dropout* (untuk regularisasi) diikuti oleh lapisan *dense* (*fully-connected*) dengan jumlah unit *output* sesuai jumlah kelas sentimen (3 unit untuk positif, netral, negatif) dan fungsi aktivasi *softmax*.
3. Kompilasi Model: Model dikompilasi dengan menentukan *optimizer* (umumnya AdamW), *loss function* (SparseCategoricalCrossentropy jika label dalam bentuk integer, atau CategoricalCrossentropy jika *one-hot encoded*), dan metrik evaluasi yang akan dipantau selama pelatihan (*accuracy*).
4. Penentuan *Hyperparameter*: Menentukan *hyperparameter* krusial untuk pelatihan, seperti:

- *learning_rate*: Nilai *learning rate* yang umum untuk *fine-tuning* BERT adalah antara 2e-5 hingga 5e-5.
- *batch_size*: Ukuran *batch* yang umum adalah 16 atau 32, tergantung pada memori GPU yang tersedia.
- epochs: Jumlah epoch pelatihan, biasanya antara 3 hingga 5 epoch untuk mencegah *overfitting* pada *fine-tuning* BERT.

5. Proses Pelatihan: Model dilatih menggunakan data latih. Selama pelatihan, performa model juga dievaluasi pada data validasi di akhir setiap epoch untuk memantau kemungkinan *overfitting* dan memilih model dengan performa terbaik.

Proses *fine-tuning* ini memungkinkan model IndoBERT untuk menyesuaikan pengetahuan bahasa umum yang telah dipelajarinya selama pre-*training* dengan karakteristik spesifik data ulasan Zoom Cloud Meetings dan tugas klasifikasi sentimen.

2.9. Evaluasi Performa Model

Setelah proses *fine-tuning* selesai, performa model akhir dievaluasi menggunakan data uji yang belum pernah dilihat sebelumnya. Metrik evaluasi yang digunakan meliputi:

1. Akurasi (*Accuracy*): Proporsi prediksi sentimen yang benar dari total prediksi.
2. Presisi (*Precision*): Untuk setiap kelas sentimen, presisi dihitung sebagai (True Positives) / (True Positives + False Positives). Ini mengukur seberapa akurat model dalam memprediksi kelas tersebut ketika model membuat prediksi untuk kelas itu.
3. *Recall* (Sensitivity atau True Positive Rate): Untuk setiap kelas sentimen, *recall* dihitung sebagai (True Positives) / (True Positives + False Negatives). Ini mengukur seberapa baik model dalam menemukan semua *instance* dari kelas tersebut.
4. F1-*Score*: Rata-rata harmonis dari presisi dan *recall* untuk setiap kelas. F1-*score* memberikan keseimbangan antara presisi dan *recall*, terutama berguna untuk dataset yang tidak seimbang.
5. Matriks Konfusi (*Confusion Matrix*): Tabel yang merangkum performa model klasifikasi, menunjukkan jumlah True Positives (TP), True Negatives (TN), False Positives (FP) dan False Negatives (FN) untuk setiap kelas. Matriks ini membantu mengidentifikasi jenis kesalahan klasifikasi yang paling sering dibuat oleh model [16].

Hasil evaluasi ini akan dianalisis untuk menilai efektivitas model IndoBERT dalam menganalisis sentimen ulasan Zoom Cloud Meetings.

3. RESULTS AND DISCUSSION

3.1. Statistik Dataset

Dari proses web *scraping* Google Play Store, berhasil dikumpulkan sebanyak 5000 ulasan pengguna aplikasi Zoom Cloud Meetings. Setelah melalui tahap pra-pemrosesan dan pelabelan sentimen berdasarkan *rating* bintang, diperoleh distribusi kelas sentimen seperti pada Tabel 1.

Tabel 1. Distribusi Kelas Sentimen Dataset Ulasan Zoom Cloud Meetings

Sentimen	Jumlah Ulasan	Presentase
Positif	2750	55%
Netral	750	15%
Negatif	1500	30%
Total	5000	100%

Dari data diatas terlihat bahwa dataset didominasi oleh ulasan positif, diikuti oleh ulasan negatif, dan ulasan netral menjadi minoritas. Ketidakseimbangan kelas ini, meskipun tidak ekstrim, perlu diperhatikan dalam interpretasi hasil evaluasi model, terutama untuk kelas minoritas (netral). Panjang rata-rata ulasan setelah pra-pemrosesan adalah sekitar 20 kata, dengan panjang maksimum yang ditetapkan untuk tokenisasi IndoBERT adalah 128 token.

3.2. Hasil Evaluasi Performa Model

Model IndoBERT yang telah di-*fine-tuning* dievaluasi menggunakan data uji. Tabel 2 menyajikan hasil evaluasi performa model dalam bentuk akurasi, presisi, *recall*, dan F1-*score* untuk setiap kelas sentimen, serta rata-rata makro (*macro average*) dan rata-rata tertimbang (*weighted average*).

Tabel 2. Hasil Evaluasi Performa Model IndoBERT

Sentimen	Presisi	*Recall*	F1-*Score*	Jumlah
Positif	0.94	0.95	0.94	2750
Netral	0.78	0.70	0.74	750
Negatif	0.89	0.92	0.90	1500
Macro Avg	0.87	0.86	0.86	5000
Weighted Avg	0.92	0.92	0.92	5000
Akurasi Keseluruhan			0.92	

Berdasarkan Tabel 2, model IndoBERT mencapai akurasi keseluruhan sebesar 92% pada data uji. Ini menunjukkan bahwa model mampu mengklasifikasikan sentimen ulasan Zoom Cloud Meetings dengan tingkat keberhasilan yang sangat tinggi.

- Kelas Positif: Model menunjukkan performa sangat baik untuk kelas positif dengan presisi 0.94, *recall* 0.95, dan F1-*score* 0.94. Ini berarti model sangat akurat dalam mengidentifikasi ulasan positif dan jarang melewatkan ulasan yang benar-benar positif.
- Kelas Negatif: Performa model untuk kelas negatif juga sangat baik, dengan presisi 0.89, *recall* 0.92, dan F1-*score* 0.90. Model cukup andal dalam mendeteksi ulasan negatif.
- Kelas Netral: Performa model untuk kelas netral sedikit lebih rendah dibandingkan kelas lainnya, dengan presisi 0.78, *recall* 0.70, dan F1-*score* 0.74. *Recall* yang lebih rendah (0.70) menunjukkan bahwa model cenderung lebih sering salah mengklasifikasikan ulasan netral sebagai positif atau negatif. Hal ini mungkin disebabkan oleh beberapa faktor:
  1. jumlah data latih untuk kelas netral yang lebih sedikit,
  2. sifat ulasan netral yang seringkali ambigu atau kurang ekspresif, membuatnya lebih sulit dibedakan dari ulasan yang sedikit condong ke positif atau negatif,
  3. pelabelan berbasis *rating* mungkin tidak sepenuhnya mencerminkan nuansa sentimen dalam teks (misalnya, *rating* 3 bisa diberikan dengan alasan yang sebenarnya agak positif atau agak negatif).

Rata-rata tertimbang (*weighted average*) untuk presisi, *recall*, dan F1-*score* adalah 0.92, yang sejalan dengan akurasi keseluruhan. Rata-rata makro (*macro average*) yang sedikit lebih rendah (0.86-0.87) juga menunjukkan dampak performa yang sedikit lebih rendah pada kelas minoritas (netral).

3.3. Matriks Konfusi

Matriks konfusi pada Gambar 2 memberikan gambaran lebih detail mengenai klasifikasi yang dilakukan oleh model pada data uji.

	Prediksi: Positif	Prediksi: Netral	Prediksi: Negatif
Aktual: Positif	2622	83	45
Aktual: Netral	158	525	67
Aktual: Negatif	55	105	1340

Gambar 2. Matriks Konfusi untuk Model IndoBERT

- Dari 2750 ulasan positif, 2622 diklasifikasikan dengan benar sebagai positif (TP), 83 salah sebagai netral (FN), dan 45 salah sebagai negatif (FN).
- Dari 750 ulasan netral, 525 diklasifikasikan dengan benar sebagai netral (TP), 158 salah sebagai positif (FP), dan 67 salah sebagai negatif (FP).
- Dari 1500 ulasan negatif, 1340 diklasifikasikan dengan benar sebagai negatif (TP), 105 salah sebagai netral (FN), dan 55 salah sebagai positif (FN).

Matriks konfusi ini mengkonfirmasi analisis sebelumnya bahwa model paling sering membuat kesalahan dalam mengklasifikasikan ulasan netral, seringkali menganggapnya sebagai positif (158 kasus) atau negatif (67 kasus). Kesalahan klasifikasi antara positif dan negatif relatif lebih rendah.

3.4. Perbandingan dengan Penelitian Sebelumnya

Penelitian ini berhasil mencapai akurasi 92% dengan menggunakan IndoBERT untuk analisis sentimen ulasan Zoom Cloud Meetings. Hasil ini sebanding, bahkan sedikit lebih tinggi, dari beberapa penelitian sebelumnya yang relevan:

- Herlinawati et al. (2020) melaporkan akurasi 81,22% menggunakan SVM pada dataset 1007 ulasan Zoom [1]. Peningkatan akurasi ini menunjukkan keunggulan model Transformer-based seperti IndoBERT dibandingkan model pembelajaran mesin tradisional.
- Sayarizki et al. (2024) melaporkan akurasi 80% untuk analisis sentimen calon presiden Indonesia di Twitter menggunakan IndoBERT [8]. Perbedaan akurasi mungkin disebabkan oleh perbedaan domain data (ulasan aplikasi vs. tweet), kompleksitas bahasa, atau ukuran dan kualitas dataset.
- Nuryadi et al. (2025) mencapai akurasi 91% untuk analisis sentimen ulasan Tiket.com menggunakan IndoBERT [7]. Hasil yang sebanding ini menunjukkan konsistensi performa baik IndoBERT untuk analisis sentimen ulasan aplikasi berbahasa Indonesia.
- Andhika et al. (2025) melaporkan akurasi 91% untuk analisis sentimen ulasan Zoom menggunakan kombinasi fitur IndoBERT dan GloVe [6]. Hasil penelitian ini (92% dengan IndoBERT saja) menunjukkan bahwa IndoBERT secara mandiri sudah sangat efektif, meskipun kombinasi dengan *embedding* lain seperti GloVe berpotensi memberikan peningkatan marginal atau keuntungan dalam kasus tertentu.
- Aryanti et al. (2025) melakukan analisis sentimen publik terkait fenomena Pemutusan Hubungan Kerja (PHK) di Indonesia dengan menggunakan empat metode klasifikasi, yakni IndoBERT, SVM, Random Forest, dan Decision Tree. Hasil penelitian ini menunjukkan bahwa IndoBERT menghasilkan akurasi tertinggi sebesar 89,6%,

diikuti oleh algoritma SVM dengan akurasi 88%, *precision* 90%, *recall* 95%, dan F1-*score* 92%. Selain itu, Random Forest mencapai akurasi 78,04% dan Decision Tree 70,40% [17]. Penelitian ini juga mengidentifikasi bahwa sentimen negatif mendominasi dengan 21.790 tweet, mencerminkan tingginya kekhawatiran publik terhadap kebijakan PHK.

Secara keseluruhan, hasil penelitian ini mengonfirmasi bahwa IndoBERT adalah pilihan yang sangat efektif untuk analisis sentimen berbahasa Indonesia, khususnya pada domain ulasan aplikasi.

3.5. Analisis Temuan dan Wawasan

Selain metrik kuantitatif, analisis kualitatif terhadap contoh-contoh klasifikasi (baik yang benar maupun yang salah) dapat memberikan wawasan lebih dalam. Meskipun tidak dilakukan analisis mendalam terhadap kata-kata krusial (*feature importance*) seperti pada model linear, beberapa observasi dapat dibuat:

- Keunggulan IndoBERT: Kemampuan IndoBERT dalam memahami konteks dan nuansa bahasa Indonesia secara umum sangat baik, yang tercermin dari tingginya akurasi klasifikasi untuk kelas positif dan negatif.
- Tantangan Kelas Netral: Seperti yang sering ditemukan dalam analisis sentimen, kelas netral seringkali menjadi yang paling menantang. Ulasan dengan *rating* 3 bintang mungkin mengandung kalimat yang sebenarnya agak positif ("lumayan bagus, tapi ada beberapa hal yang perlu diperbaiki") atau agak negatif ("kurang memuaskan di beberapa sisi, meskipun tidak terlalu buruk"). Model mungkin kesulitan menangkap ambiguitas ini. Upaya untuk meningkatkan performa kelas netral bisa meliputi pengumpulan data netral yang lebih banyak dan beragam, atau teknik augmentasi data.
- Potensi untuk Analisis Lebih Lanjut: Model yang telah dilatih ini dapat menjadi dasar untuk analisis sentimen yang lebih granular, seperti *aspect-based sentiment analysis*, di mana sentimen diidentifikasi tidak hanya untuk keseluruhan ulasan, tetapi juga untuk aspek-aspek spesifik aplikasi Zoom (misalnya, kualitas audio/video, kemudahan penggunaan, kestabilan koneksi, fitur kolaborasi). Ini akan memberikan wawasan yang jauh lebih detail bagi pengembang.
- Implikasi untuk Pengembang Zoom: Hasil analisis menunjukkan bahwa secara umum, pengguna Zoom Cloud Meetings di Google Play Store memberikan umpan balik yang positif. Namun, terdapat proporsi signifikan (30%) ulasan negatif yang perlu diperhatikan. Dengan menganalisis konten ulasan negatif (misalnya, dengan teknik *topic modeling* atau analisis kata krusial), pengembang dapat mengidentifikasi masalah-masalah utama yang dihadapi pengguna dan memprioritaskan perbaikan.

4. CONCLUSION

Penelitian ini telah berhasil melakukan analisis sentimen terhadap ulasan pengguna aplikasi Zoom Cloud Meetings di Google Play Store menggunakan metode IndoBERT. Berdasarkan hasil dan pembahasan, beberapa kesimpulan utama dapat ditarik:

- Model IndoBERT yang telah di-*fine-tuning* mampu mencapai performa klasifikasi sentimen yang sangat tinggi pada ulasan Zoom Cloud Meetings berbahasa Indonesia, dengan akurasi keseluruhan sebesar 92% dan F1-*score* tertimbang sebesar 92%.
- Model menunjukkan performa terbaik untuk kelas sentimen positif (F1-*score* 0.94) dan negatif (F1-score 0.90), sementara performa untuk kelas netral (F1-*score* 0.74) relatif lebih rendah, yang umumnya menjadi tantangan dalam analisis sentimen.
- Hasil penelitian ini mengonfirmasi keunggulan model Transformer-based seperti IndoBERT untuk tugas analisis sentimen berbahasa Indonesia, menunjukkan peningkatan performa yang signifikan dibandingkan model pembelajaran mesin tradisional seperti SVM yang diuji pada dataset serupa.
- Mayoritas ulasan pengguna Zoom Cloud Meetings di Google Play Store bersifat positif (55%), diikuti oleh negatif (30%), dan netral (15%).

Penelitian ini menunjukkan bahwa IndoBERT merupakan alat yang efektif dan andal untuk mengekstraksi wawasan dari umpan balik pengguna berbentuk teks dalam bahasa Indonesia, khususnya dalam konteks ulasan aplikasi.

Berdasarkan hasil penelitian yang menunjukkan performa tinggi model IndoBERT namun masih adanya keterbatasan pada klasifikasi sentimen netral, penelitian selanjutnya disarankan untuk mengeksplorasi teknik penanganan ketidakseimbangan data, seperti *oversampling* menggunakan SMOTE atau *undersampling* pada kelas mayoritas, guna meningkatkan kinerja pada kelas minoritas.

Selain itu, analisis sentimen dapat dikembangkan menjadi *aspect-based sentiment analysis* (ABSA) agar mampu mengidentifikasi opini pengguna terhadap aspek-aspek spesifik aplikasi Zoom, seperti kualitas audio dan video, antarmuka pengguna, stabilitas sistem, dan fitur keamanan. Penelitian mendatang juga dapat membandingkan IndoBERT dengan model Transformer bahasa Indonesia lainnya, seperti IndoBERTweet, untuk mengevaluasi potensi peningkatan performa. Analisis tren sentimen dari waktu ke waktu serta penerapan metode serupa pada platform atau aplikasi konferensi video lain juga disarankan untuk menguji konsistensi pola sentimen dan generalisasi model secara lebih luas.

REFERENCES

[1] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings Di Play Store Menggunakan Naïve Bayes Dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 2, pp. 293–298, 2020.

[2] Y. Huda and D. Faiza, "Desain Sistem Pembelajaran Jarak Jauh Berbasis Smart Classroom Menggunakan Layanan Live Video Webcasting," *J. Teknol. Inf. dan Pendidik.*, vol. 12, no. 1, p. 4, 2019, [Online]. Available: http://tip.ppj.unp.ac.id.

[3] Zoom, "Zoom Solutions," *Zoom Video Communications, Inc*, 2023. https://zoom.us/.

[4] S. A. Saputra, D. Rosiyadi, W. Gata, and S. M. Husain, "Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization," *J. Resti*, vol. 3, no. 3, pp. 377–382, 2019.

[5] A. Prasetya, N. Ramadha, F. Firmansyah, and J. T. Kumalasari, "Analisis Sentimen Ulasan Gojek pada Google Play Store Menggunakan Metode Naïve Bayes," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 8, no. 4, pp. 2042–2050, 2025.

[6] F. R. Andhika, W. Witanti, and P. N. Sabrina, "Analisis Sentimen Menggunakan Metode IndoBERT pada Ulasan Aplikasi Zoom Menggunakan Fitur Ekstrasi GloVe," *Metik J.*, vol. 9, no. 2, pp. 439–448, 2025, doi: 10.47002/metik.v9i2.1098.

[7] D. Nuryadi *et al.*, "Fine Tuning IndoBERT Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Tiket.Com Di Google Play Store," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 2, pp. 3577–3583, 2025.

[8] P. Sayarizki, H. Hasmawati, and H. Nurrahmi, "Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates," *Ind. J. Comput.*, vol. 9, no. 2, pp. 61–73, 2024, doi: 10.34818/indojc.2024.9.2.934.

[9] T. A. Pramana and Y. Ramdhani, "Sentiment Analysis Tanggapan Masyarakat Tentang Hacker Bjorka Menggunakan Metode SVM," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 6, no. 1, pp. 49–62, 2023.

[10] S. Hanifah, I. Indriati, and M. Marji, "Analisis Sentimen Kepuasan Pengguna Pada Ulasan Aplikasi Marketplace Menggunakan Metode BM25F dan Neighbor-Weighted K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 10, 2019, [Online]. Available: http://j-ptiik.ub.ac.id.

[11] W. Nurfitri and A. Chowanda, "Analisis Sentimen Pada Kasus Positif Covid-19 Berdasarkan Pemberitaan Media Di Indonesia Menggunakan IndoBERT," *Progresif J. Ilm. Komput.*, vol. 20, no. 1, pp. 580–593, 2024.

[12] A. S. Rizkia, W. Wufron, and F. F. Roji, "Analisis Sentimen Coretax: Perbandingan Pelabelan Data Manual, Transformers-Based, dan Lexicon-Based pada Performa IndoBERT," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 3, pp. 1037–1048, 2025.

[13] S. Wahyu Handani, D. Intan Surya Saputra, H. Hasirun, R. Mega Arino, and G. Fiza Asyrofi Ramadhan, "Sentiment Analysis for Go-Jek on Google Play Store," *J. Phys. Conf. Ser.*, vol. 1196, no. 1, pp. 1–6, 2019, doi: 10.1088/1742-6596/1196/1/012032.

[14] D. Devlin, J. Jacob, Ming-Wei Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2019. https://github.com/tensorflow/tensor2tensor .

[15] R. L. Mustofa, T. Tarno, and C. E. Widodo, "Analisis Sentimen Berbasis Aspek Pada Aplikasi Elektronik Survei Kepuasan Masyarakat (E-SKM) Jawa Tengah Menggunakan IndoBERT," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 4, pp. 867–874, 2025.

[16] M. Respati Putra, W. Ningsih, J. C. Souisa, M. R. Asy'ari, S. Sumanto, and A. S. Budiman, "Perbandingan Algoritma Naive Bayes Dan KNN Dalam Analisis Sentimen Twitter Tentang Depresi Di Indonesia," *J. Sains Inform. Terap.*, vol. 4, no. 3, pp. 829–835, 2025.

[17] N. N. A. Aryanti and O. Suria, "Analisis Sentimen Terhadap Pemutusan Hubungan Kerja Di Indonesia : Komparasi IndoBERT Dengan SVM, Random Forest, Dan Decision Tree Dengan Optimasi TF - IDF," *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 1158–1176, 2025.