Jurnal Informatika: Jurnal pengembangan IT

Vol. 10, No. 3, 2025
ISSN : 2477-5126, DOI:10.30591/jpit.v10i3.8607

Perbandingan IndoBERT dan IndoRoBERTa Untuk Analisis Sentimen Pada Film Dokumenter Dirty Vote

Fadhel Muhammad Apriansyah[1], Teguh Ikhlas Ramadhan [2], Cepi Rahmat Hidayat[3], Anggito Karta Wijaya[4]

[123]Program Studi Teknik Informatika, Universitas Perjuangan Tasikmalaya, Tasikmalaya, 46115, Indonesia
[4] Program Studi Sistem Informasi, Universitas Jember, Jember, 68121, Indonesia

Info Artikel

Riwayat Artikel:

Received 2025-03-24
Revised 2025-05-30
Accepted 2025-06-04

Corresponding Author:

Fadhel Muhammad Apriansyah

Email: fmapriansyah3@gmail.com

This is an open access article under the CC BY 4.0 license.

Abstract *–Sentiment analysis is a technique in Natural Language Processing (NLP) used to identify and classify opinions or emotions expressed in text. This study aims to compare the performance of two Transformer-based language models, IndoBERT and IndoRoBERTa, in analyzing sentiment from public comments on the documentary film Dirty Vote. The research process includes data collection, text preprocessing, sentiment labeling using a lexicon of positive and negative words, and model evaluation through the K-Fold Cross-Validation method. The results show that IndoBERT achieved an average accuracy of 99%, while IndoRoBERTa reached 94%. IndoBERT demonstrated more consistent performance in classifying positive, negative, and neutral sentiments. Architectural differences between the models influenced their sensitivity to textual meaning, with IndoBERT training faster and showing stronger alignment with explicitly expressed opinions or criticism. In contrast, IndoRoBERTa tended to classify ambiguous comments as neutral.These findings indicate that IndoBERT is more effective for sentiment analysis of Indonesian-language texts in socio-political contexts. The study contributes to the development of NLP models, particularly in applying Transformer-based approaches to public opinion analysis.*

Keywords: *Sentiment Analysis; IndoBERT; IndoRoBERTa; Dirty Vote; Electoral Fraud; Natural Language Processing.*

Abstrak *–Analisis sentimen merupakan teknik dalam pemrosesan bahasa alami (Natural Language Processing, NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini atau emosi dalam teks. Penelitian ini bertujuan membandingkan kinerja dua model bahasa berbasis Transformer, yakni IndoBERT dan IndoRoBERTa, dalam menganalisis sentimen terhadap komentar publik pada film dokumenter Dirty Vote. Proses penelitian mencakup tahapan pengumpulan data, prapemrosesan teks, pelabelan sentimen berdasarkan daftar kata positif dan negatif (kamus leksikal), serta evaluasi performa model menggunakan metode K-Fold Cross-Validation. Hasil penelitian menunjukkan bahwa IndoBERT memperoleh akurasi rata-rata sebesar 99%, sedangkan IndoRoBERTa mencapai 94%. IndoBERT menunjukkan performa yang lebih stabil dalam mengklasifikasikan sentimen positif, negatif, dan netral. Perbedaan arsitektur antara kedua model memengaruhi sensitivitas dalam mengenali makna teks, di mana IndoBERT lebih cepat dalam pelatihan dan lebih konsisten dalam menangkap ekspresi opini atau kritik secara eksplisit. Sementara itu, IndoRoBERTa cenderung mengategorikan komentar yang ambigu sebagai netral. Temuan ini menunjukkan bahwa IndoBERT lebih efektif dalam tugas analisis sentimen terhadap teks berbahasa Indonesia dalam konteks sosial-politik. Hasil penelitian ini dapat menjadi referensi untuk pengembangan lebih lanjut dalam bidang pemrosesan bahasa alami, khususnya dalam penggunaan model Transformer untuk analisis opini publik.*

Kata Kunci*: Analisis Sentimen, IndoBERT, IndoRoBERTa, Dirty Vote, Kecurangan Pemilu, Pemrosesan Bahasa Alami*

I. PENDAHULUAN

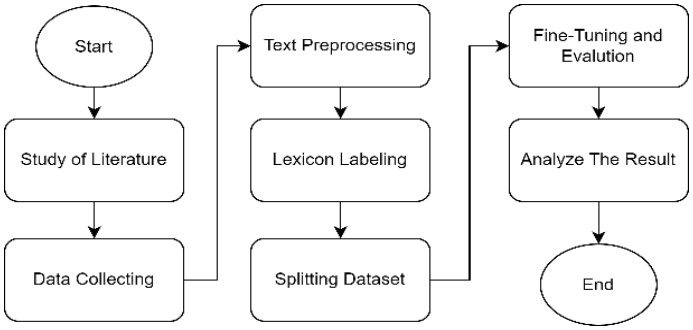
Film dokumenter memiliki peran penting dalam merekam dan menyajikan peristiwa, fakta, serta fenomena nyata yang terjadi di masyarakat [1]. Selain berfungsi sebagai media informasi, film dokumenter juga memberikan wawasan yang lebih mendalam mengenai suatu topik tertentu. Salah satu film dokumenter yang relevan dalam konteks politik di Indonesia adalah *"Dirty Vote"*, yang mengangkat isu kecurangan dalam pemilihan umum dan telah ditonton sebanyak 9,4 juta kali di YouTube [2]. Popularitas film ini menunjukkan adanya ketertarikan publik yang tinggi terhadap isu yang diangkat. Relevansi film ini sebagai studi kasus dalam analisis sentimen terletak pada muatan isu sosial-politik yang sensitif dan mampu memicu beragam respons emosional dari masyarakat. Film ini memunculkan diskusi terbuka, kritik, serta dukungan dari berbagai kalangan melalui media sosial, khususnya di kolom komentar YouTube, sehingga menyediakan data tekstual yang kaya untuk dianalisis. Dengan menganalisis komentar-komentar tersebut, dapat diperoleh pemahaman yang lebih mendalam mengenai persepsi publik terhadap integritas demokrasi dan dinamika politik di Indonesia, menjadikan *Dirty Vote* sebagai objek yang tepat dan kontekstual untuk kajian analisis sentimen. Berdasarkan

data Google Trends pada 23 hingga 29 Februari 2024, tren pencarian film "*Dirty Vote*" di Indonesia mencatat 40 pencarian di web dan 38 pencarian di *YouTube* [3]. Data ini mencerminkan minat masyarakat terhadap topik pemilu dan dugaan praktik kecurangan yang terjadi. Sebagai platform berbagi video daring, *YouTube* memiliki peran strategis dalam penyebaran informasi dan diskusi publik [4]. Dalam era digital, platform seperti *YouTube* memungkinkan masyarakat untuk mengungkapkan opini mereka secara terbuka melalui komentar, ulasan, serta diskusi daring [5]. Seiring dengan perkembangan media sosial dan platform daring, jumlah data tekstual yang dihasilkan meningkat secara signifikan [6]. Data ini mencakup opini, sentimen, serta reaksi emosional yang dapat dianalisis untuk memahami persepsi publik terhadap suatu isu tertentu. Dengan meningkatnya partisipasi masyarakat dalam diskusi daring, analisis sentimen menjadi semakin relevan dalam mengevaluasi opini publik secara sistematis [7]. Namun, tantangan dalam analisis sentimen muncul ketika harus memahami makna tersirat dalam teks, terutama jika mengandung idiom, bahasa gaul, atau konteks budaya tertentu [8]. Oleh karena itu, diperlukan pendekatan yang mampu menangkap kompleksitas bahasa alami dalam analisis sentimen. Model pembelajaran mesin berbasis Transformer, seperti Bidirectional Encoder Representations from Transformers (BERT) dan Robustly Optimized BERT Pretraining Approach (RoBERTa), telah terbukti efektif dalam berbagai tugas pemrosesan bahasa alami [9]. IndoBERT dan IndoRoBERTa yang merupakan varian model yang dikembangkan khusus untuk bahasa Indonesia, telah banyak digunakan dalam penelitian analisis sentimen [10]. Studi sebelumnya menunjukkan bahwa model berbasis Transformer memiliki keunggulan dibandingkan metode tradisional dalam memahami konteks dan makna dalam teks berbahasa Indonesia [11]. IndoBERT dan IndoRoBERTa menawarkan keunggulan dalam analisis sentimen karena kemampuannya dalam memahami konteks kata dalam kalimat. Perbedaan fundamental antara kedua model tersebut terletak pada proses pretraining dan pengoptimalan arsitektur. IndoBERT menggunakan arsitektur BERT asli yang menerapkan masked language modeling dan next sentence prediction sebagai tugas pretraining. Sedangkan IndoRoBERTa mengadopsi pendekatan RoBERTa yang menghilangkan tugas next sentence prediction serta menerapkan dynamic masking yang berbeda pada setiap epoch. Selain itu, IndoRoBERTa dilatih dengan data dan batch size yang lebih besar, sehingga memungkinkan pemahaman konteks yang lebih mendalam dan pelatihan yang lebih efisien [14]. Berkat hal ini, IndoRoBERTa cenderung menunjukkan performa yang lebih unggul serta waktu pelatihan yang lebih singkat dibandingkan IndoBERT. Berdasarkan penelitian yang dilakukan oleh Koto, IndoBERT memiliki akurasi mencapai 91,2% dalam tugas klasifikasi sentimen pada dataset berita dan media sosial berbahasa Indonesia [12]. Sementara itu, IndoRoBERTa, yang merupakan versi lebih optimal dari BERT, mampu meningkatkan akurasi analisis sentimen hingga 92,5% pada tugas serupa sebagaimana dalam penelitian Pratama & Wibowo[13]. Model ini juga lebih efisien dalam pelatihan karena proses pretraining yang dioptimalkan, memungkinkan performa yang lebih baik dengan waktu komputasi yang lebih singkat [14]. Penelitian terdahulu yang dilakukan oleh Putri & Ardiansyah membahas penerapan BERT dan RoBERTa dalam analisis sentimen terhadap kemajuan kecerdasan buatan di Indonesia, meskipun objek penelitian dan dataset yang digunakan berbeda [15]. Selain itu, penelitian oleh Lazuardi & Juarna menunjukkan efektivitas BERT dalam analisis sentimen ulasan aplikasi, yang mengindikasikan potensi penggunaan model ini dalam berbagai domain [16]. Dalam studi lainnya, IndoBERT menunjukkan performa yang lebih unggul dibandingkan model Long Short-Term Memory (LSTM) dan Support Vector Machine (SVM) dalam analisis sentimen dengan selisih akurasi rata-rata sebesar 5–10% [17]. Dengan demikian, pendekatan berbasis Transformer menjadi pilihan utama dalam analisis sentimen di berbagai bidang, termasuk politik dan sosial. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan performa model IndoBERT dan IndoRoBERTa dalam analisis sentimen terhadap film dokumenter "Dirty Vote". Studi ini menggunakan data yang berfokus pada opini publik terhadap isu sosial-politik yang kompleks, sehingga menghadirkan tantangan tersendiri dalam pemahaman konteks bahasa. Selain itu, penelitian ini mengaplikasikan metode evaluasi yang komprehensif untuk memberikan gambaran performa kedua model secara lebih mendalam [18]. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode analisis sentimen yang lebih akurat untuk bahasa Indonesia serta memperkaya kajian komunikasi politik dan pemrosesan bahasa alami [19].

II. METODE

Pada bab ini dijelaskan secara rinci mengenai metode penelitian yang digunakan dalam analisis sentimen terhadap film dokumenter *Dirty Vote* dengan pendekatan model IndoBERT dan IndoRoBERTa. Penelitian ini dirancang dan dilaksanakan melalui beberapa tahapan yang sistematis dan terstruktur guna memastikan keakuratan serta validitas hasil yang diperoleh. Setiap tahap dalam proses penelitian memiliki peranan penting untuk mendukung kualitas analisis sentimen yang dilakukan. Alur tahapan penelitian secara keseluruhan dapat dilihat pada Gambar 1. Gambar 1. menunjukkan diagram alur penelitian yang dimulai dari tahap studi literatur dan pengumpulan data relevan sebagai dasar pengembangan model. Selanjutnya, data yang diperoleh menjalani proses prapemrosesan teks yang bertujuan untuk membersihkan dan menyiapkan data agar sesuai untuk analisis lebih lanjut. Setelah tahap prapemrosesan, dilakukan pemberian label sentimen berbasis leksikon dan pembagian

dataset menjadi subset pelatihan dan pengujian. Pada tahap berikutnya, model IndoBERT dan IndoRoBERTa dilakukan fine-tuning dengan data yang telah dipersiapkan, kemudian dievaluasi performanya menggunakan metrik yang relevan. Terakhir, hasil evaluasi dianalisis secara mendalam untuk menarik kesimpulan dan memberikan rekomendasi sebelum penelitian ini diakhiri.



Gambar 1. Alur Tahapan Penelitian

A. Dataset

Dalam penelitian ini, proses pengumpulan data dilakukan dengan menggunakan teknik web scraping dari platform YouTube, yang dipilih karena merupakan salah satu media sosial paling populer dengan jutaan pengguna aktif yang secara terbuka membagikan opini mereka melalui kolom komentar. YouTube juga menyediakan ruang diskusi yang luas dalam bentuk komentar publik, yang sangat bermanfaat untuk menangkap persepsi, sentimen, dan respons spontan dari masyarakat terhadap isu-isu aktual. Objek yang dianalisis adalah komentar dari video film dokumenter berjudul "Dirty Vote", yang menjadi viral dan menjadi bahan perbincangan luas di berbagai media sosial seperti Twitter (X), Instagram, dan TikTok. Perbincangan tersebut mencakup topik tentang etika pemilu, integritas demokrasi, serta kritik terhadap praktik-praktik politik di Indonesia. Film ini memicu reaksi emosional dan pemikiran kritis dari masyarakat, yang tercermin dalam komentar-komentar publik. Pemilihan komentar dari YouTube sebagai sumber data dilakukan karena komentar di platform ini sering kali bersifat lebih panjang, mendalam, dan kontekstual dibandingkan media sosial lain, sehingga memungkinkan analisis sentimen dan opini publik yang lebih kaya. Selain itu, YouTube menyediakan kemudahan dalam pengambilan data secara otomatis (scraping), sehingga memungkinkan pengumpulan data dalam jumlah besar secara efisien. Melalui proses web scraping, berhasil dikumpulkan sebanyak 72.538 komentar, yang mencerminkan beragam opini, sentimen, serta tanggapan pengguna YouTube terhadap film "Dirty Vote". Pengambilan data dilakukan hingga tanggal 18 Februari 2024, sehingga komentar yang terkumpul mencerminkan berbagai perspektif publik dalam periode waktu yang relevan dengan momentum penayangan dan viralnya film tersebut. Untuk memberikan gambaran lebih jelas mengenai hasil pengumpulan data, Tabel 1 berikut menyajikan sampel data komentar YouTube tentang film *Dirty Vote* yang diperoleh dari proses web scraping.

TABEL 1
SAMPEL DATA KOMENTAR YOUTUBE TENTANG DIRTY VOTE

No	Comment
1	CAPE CAPE BUAT FILM KONYOL MILIHNYA YG KALAH RUGI DONG
2	perwakilan yang tidak puas dengan sistem pemerintahan saat ini, itu aja, tapi yang memandang dari sudut lain gmn?
3	ada yg nonton sampe abis emang?
4	Film yg koar2 turinin presiden 😂😂😂
5	tapi realita nya ga semudah asumsi anda... mau semua gubernur pro ke 1 paslon pun belum tentu rakyat ngikutin para gubernur tersebut.

B. Text Preprocessing

Dalam penelitian ini, proses pengumpulan data dilakukan dengan menggunakan teknik *web scraping* dari platform *YouTube*. Data yang dikumpulkan berupa komentar dari video film dokumenter berjudul "*Dirty Vote*",

yang merupakan salah satu film yang banyak diperbincangkan di media sosial. Teknik *web scraping* digunakan untuk mengekstrak data komentar secara otomatis, sehingga memungkinkan pengambilan data dalam jumlah besar dengan efisiensi tinggi. Hasil dari proses pengumpulan data ini berhasil memperoleh sebanyak 72.538 komentar, yang mencerminkan berbagai opini, sentimen, serta tanggapan pengguna *YouTube* terhadap film tersebut. Pengambilan data dilakukan hingga 18 Februari 2024, sehingga komentar yang terkumpul mencakup berbagai perspektif dari pengguna dalam periode waktu tertentu. Untuk memberikan gambaran lebih jelas mengenai hasil pengumpulan data, Tabel 1 berikut menyajikan sampel data komentar youtube tentang *dirty vote* yang diperoleh dari proses *web scraping*.

1) Casefolding

Tahap pertama dalam prapemrosesan teks adalah *casefolding*, yaitu mengonversi seluruh huruf dalam teks menjadi huruf kecil guna memastikan konsistensi format. Berdasarkan hasil prapemrosesan, jumlah karakter yang berhasil dikonversi melalui proses casefolding mencapai 9.147.357 karakter.

2) Pembersihan Teks (*Text Cleaning*)

Pada tahap ini, dilakukan penghapusan elemen-elemen yang tidak relevan dalam teks, seperti tanda baca, angka, simbol khusus (@, #, $, %, &, dan sebagainya), emoji, kata yang berulang, serta karakter yang tidak memiliki makna dalam analisis. Sebelum proses pembersihan, total jumlah karakter dalam teks adalah 8.684.640, sedangkan setelahnya berkurang menjadi 8.627.992 karakter, menunjukkan bahwa sejumlah elemen yang tidak penting telah dihilangkan.

3) Penghapusan Kata Umum (*Stopword Removal*)

Tahap berikutnya adalah penghapusan kata-kata umum (*stopword removal*), yaitu proses menghilangkan kata yang sering muncul tetapi tidak memberikan makna signifikan dalam analisis sentimen, seperti "saya", "kamu", "kita", dan sebagainya. Sebelum penghapusan, jumlah karakter dalam teks mencapai 1.350.234, sedangkan setelahnya menyusut menjadi 723.754 karakter.

4) Normalisasi Teks (*Normalization*)

Normalisasi teks dilakukan untuk memperbaiki kesalahan ejaan serta menyelaraskan penggunaan kata yang mengandung idiom atau bahasa informal agar sesuai dengan kaidah bahasa baku. Sebelum proses normalisasi, jumlah teks yang tersedia mencapai 8.684.640 karakter, sementara setelahnya berkurang menjadi 8.673.474 karakter, menunjukkan adanya penyempurnaan dan standarisasi dalam teks.

5) *Stemming* atau Lemmatisasi

Stemming dan lemmatisasi bertujuan untuk mengubah kata ke bentuk dasarnya dengan menghilangkan imbuhan (*prefix*, *suffix*, dan *infix*). Lemmatisasi dilakukan dengan merujuk pada Kamus Besar Bahasa Indonesia (*KBBI*) agar setiap kata memiliki bentuk yang sesuai dengan standar bahasa. Jumlah teks sebelum dan setelah proses stemming serta lemmatisasi tetap sebanyak 723.754 karakter.

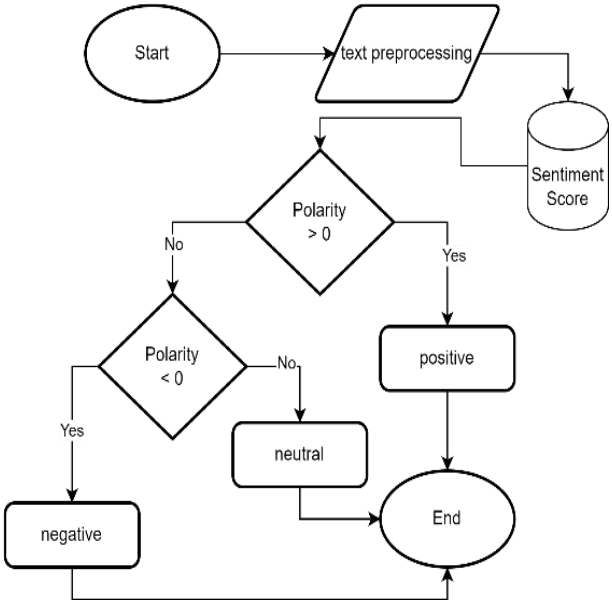
6) Tokenisasi (*Tokenization*)

Tahap terakhir dalam prapemrosesan teks adalah tokenisasi, yaitu proses memecah kalimat menjadi unit kata yang lebih kecil agar lebih mudah dianalisis. Dengan adanya tokenisasi, teks dapat disajikan dalam format yang lebih terstruktur dan siap untuk tahap analisis lebih lanjut.

C. Labeling with Lexicon

Metode *lexicon* digunakan untuk menentukan apakah suatu kalimat, teks, atau komentar tergolong dalam kategori positif, negatif, atau netral (Nahar). Setelah proses prapemrosesan teks selesai, langkah selanjutnya adalah melakukan pelabelan sentimen menggunakan metode *lexicon*. Pada tahap ini, sistem membandingkan setiap kata dalam teks dengan *sentiment lexicon* yang telah disusun sebelumnya berdasarkan kajian dan penelitian terdahulu. Kamus tersebut terdiri dari dua kategori utama, yaitu kamus kata positif yang berisi 2.083 kata dan kamus kata negatif yang berisi 1.510 kata. Kata-kata dalam teks yang sesuai dengan daftar kata positif akan dikategorikan sebagai sentimen positif, sedangkan kata-kata yang ditemukan dalam daftar kata negatif akan diberi label sentimen negatif. Apabila suatu teks tidak mengandung kata-kata yang terdapat dalam kedua kamus tersebut, maka teks tersebut diklasifikasikan sebagai sentimen netral. Dengan pendekatan ini, analisis sentimen dapat dilakukan secara sistematis dan konsisten berdasarkan pola kata dalam teks, sehingga menghasilkan klasifikasi yang lebih objektif, terstruktur, dan dapat diandalkan dalam berbagai konteks analisis. Pada Gambar 2. ditampilkan alur pelabelan *lexicon* yang menggunakan pendekatan *polarity* dalam menentukan kategori sentimen. *Polarity* merupakan konsep yang mencerminkan dua kutub berlawanan dalam suatu entitas, yaitu kutub positif dan negatif. Dalam konteks analisis sentimen, nilai *polarity* digunakan untuk mengukur kecenderungan suatu teks terhadap sentimen tertentu secara kuantitatif. Jika nilai *polarity* lebih dari nol (0), maka

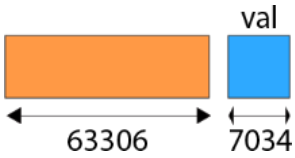
teks tersebut dikategorikan sebagai sentimen positif. Sebaliknya, jika nilai *polarity* kurang dari nol (0), teks tersebut digolongkan sebagai sentimen negatif. Sementara itu, apabila nilai *polarity* sama dengan nol (0), teks dianggap memiliki sentimen netral atau tidak memihak. Pendekatan ini memungkinkan klasifikasi sentimen yang lebih objektif dan berbasis data numerik, sehingga dapat memberikan gambaran yang lebih jelas mengenai orientasi sentimen dalam suatu kumpulan data teks. Dengan demikian, metode *lexicon* berbasis *polarity* ini menjadi salah satu teknik yang efektif dan populer dalam analisis sentimen berbasis teks.



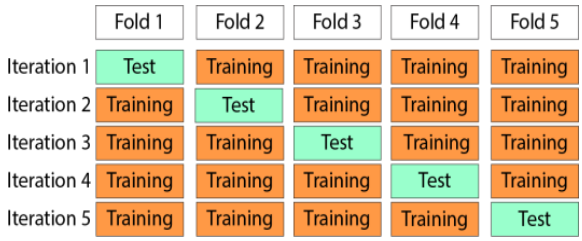
Gambar 2. Flowchart Lexicon Labeling

D. Splitting Dataset

Tahap *splitting dataset* merupakan proses pembagian data untuk melatih dan menguji model agar dapat melakukan prediksi dengan baik. Dalam penelitian ini, data dibagi dengan rasio 80:20, di mana 80% digunakan sebagai data latih (*training set*) untuk membangun model, dan 20% sisanya sebagai data uji (*testing set*) untuk mengevaluasi performa model. Pembagian ini bertujuan agar model mampu mengenali pola secara optimal dan melakukan generalisasi dengan baik pada data baru. Pada Gambar 3, sebelum menerapkan *k-fold cross-validation*, sebagian data utama diambil dan disimpan sebagai subset data validasi. Subset ini digunakan untuk menguji model yang telah dilatih. Dalam penelitian ini, data validasi diambil sebesar 10% dari keseluruhan data, yaitu sebanyak 7.034 komentar. Pada Gambar 4, ditampilkan ilustrasi proses *k-fold cross-validation*. Sebanyak 63.306 komentar diterapkan dalam metode ini, di mana data pelatihan dibagi menjadi *k* bagian atau *folds*. Setiap subset dibagi dengan rasio 80:20, dengan rata-rata 56.272 komentar digunakan sebagai data pelatihan dan 14.068 komentar sebagai data uji dalam setiap iterasi.



Gambar 3. Split Dataset for Validation.



Gambar 4. Ilustrasi Proses K-Fold Cross-Validation

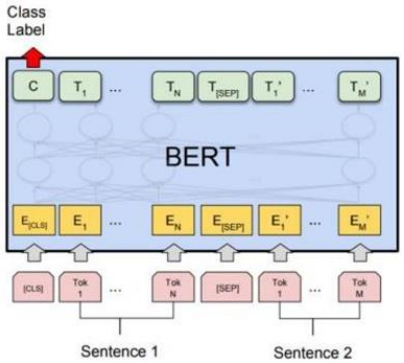
E. Fine-Tuning

Pada tahap ini, dilakukan proses *fine-tuning* guna mengoptimalkan kinerja model bahasa IndoBERT dan IndoRoBERTa agar lebih sesuai dengan tugas analisis sentimen terhadap komentar publik pada film dokumenter *Dirty Vote*. Tujuan dari fine-tuning adalah untuk menyesuaikan parameter model terhadap karakteristik data yang digunakan, sehingga model dapat secara lebih akurat memahami konteks dan mengklasifikasikan sentimen

dalam teks berbahasa Indonesia, khususnya yang berkaitan dengan isu sosial-politik. Proses fine-tuning untuk model IndoBERT dilakukan dengan beberapa parameter utama, antara lain jumlah *epoch* sebanyak 3, *batch size* sebesar 16, dan panjang maksimum token (max sequence length) sebanyak 512 token. Proses pelatihan menggunakan *learning rate* sebesar 2e-5 dan dioptimasi menggunakan algoritma AdamW dengan fungsi kerugian CrossEntropyLoss. Pelatihan dilakukan pada perangkat keras GPU NVIDIA Tesla T4 melalui platform Google Colab, dengan waktu pelatihan rata-rata sekitar 10 menit 41 detik per *epoch*. Sementara itu, untuk model IndoRoBERTa, fine-tuning dilakukan dengan jumlah *epoch* sebanyak 3 dan *batch size* sebesar 8. Panjang maksimum token yang digunakan adalah 60, dengan *learning rate* sebesar 1e-5. Selain itu, model ini juga menggunakan *gradient accumulation steps* sebanyak 2 dan batas maksimum normalisasi gradien (*max gradient norm*) sebesar 1.0. Optimasi model dilakukan menggunakan algoritma AdamW dengan fungsi kerugian CrossEntropyLoss. Proses pelatihan juga dilakukan pada GPU NVIDIA Tesla T4 di Google Colab, dengan waktu pelatihan rata-rata sekitar 21 menit 09 detik per *epoch*. Kedua model diimplementasikan menggunakan pustaka HuggingFace Transformers berbasis framework PyTorch, dan masing-masing model dimuat dari repositori yang telah tersedia, yaitu indobenchmark/indobert-base-p2 untuk IndoBERT dan w11wo/indonesian-roberta-base-sentiment-classifier untuk IndoRoBERTa. Setelah proses fine-tuning selesai, dilakukan evaluasi performa model menggunakan sejumlah metrik klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Metrik-metrik tersebut digunakan untuk menilai sejauh mana model mampu memprediksi label sentimen secara akurat dan konsisten. Proses evaluasi dilakukan dengan pendekatan K-Fold Cross-Validation guna memastikan bahwa model tidak hanya mampu mengenali pola pada data pelatihan, tetapi juga memiliki kemampuan generalisasi yang baik terhadap data baru yang tidak terlihat sebelumnya. Evaluasi ini menjadi dasar dalam membandingkan performa kedua model, serta memberikan gambaran yang lebih menyeluruh mengenai efektivitas masing-masing pendekatan dalam klasifikasi sentimen teks berbahasa Indonesia.

1) IndoBERT

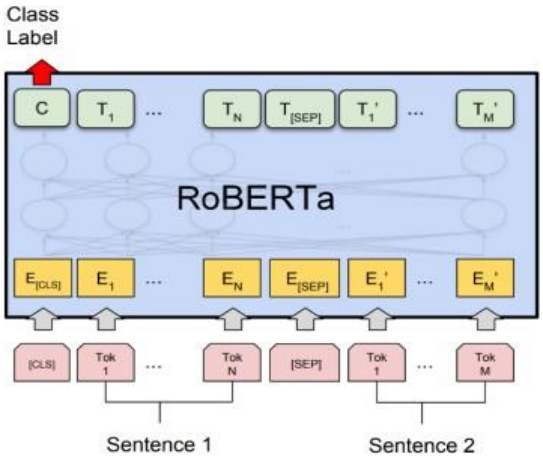
IndoBERT adalah model bahasa alami berbasis transformer yang dikembangkan dari BERT (*Bidirectional Encoder Representations from Transformers*) dan telah dilatih secara khusus untuk bahasa Indonesia. Model ini dirancang untuk memahami konteks kata dalam suatu kalimat secara lebih mendalam, sehingga dapat digunakan dalam berbagai tugas pemrosesan bahasa alami (*Natural Language Processing*), seperti analisis sentimen, klasifikasi teks, pemahaman bahasa, dan ekstraksi informasi. IndoBERT menggunakan pendekatan *masked language modeling* (MLM) dan *next sentence prediction* (NSP) untuk menangkap hubungan antar kata dalam suatu kalimat maupun antar kalimat secara lebih efektif. Dengan strategi ini, model dapat memahami makna kata berdasarkan konteks sekitarnya serta mengidentifikasi hubungan antara dua kalimat. Gambar 5 menampilkan arsitektur IndoBERT yang digunakan dalam penelitian ini. Model yang dipilih adalah indobenchmark/indobert-base-p2, yang merupakan bagian dari proyek IndoNLU. Model ini dikembangkan berdasarkan arsitektur BERT standar tetapi dengan korpus teks yang dikumpulkan secara khusus dari berbagai sumber berbahasa Indonesia, seperti berita, media sosial, dan dokumen resmi. Dengan demikian, IndoBERT lebih optimal dalam menangani teks berbahasa Indonesia dibandingkan dengan BERT standar yang dilatih menggunakan teks dalam bahasa Inggris. Secara teknis, arsitektur IndoBERT terdiri dari embedding layer yang mencakup *token embedding*, *segment embedding*, dan *positional embedding*. Setiap token input akan melewati beberapa lapisan transformer yang menerapkan mekanisme self-attention, memungkinkan model memahami hubungan antar kata dalam teks dengan lebih baik. Pada tahap akhir, token [CLS] digunakan sebagai representasi keseluruhan kalimat dan diteruskan ke lapisan klasifikasi untuk menentukan sentimen atau label yang sesuai berdasarkan teks yang diberikan. Dengan pendekatan ini, IndoBERT menjadi salah satu model terbaik untuk berbagai tugas NLP dalam bahasa Indonesia, menawarkan performa yang lebih akurat dalam memahami, menganalisis, dan mengolah teks dibandingkan model berbasis bahasa lain.



Gambar 5. Arsitektur IndoBERT.

2) IndoRoBERTa

IndoRoBERTa adalah model pemrosesan bahasa alami yang dikembangkan berdasarkan arsitektur RoBERTa (*Robustly Optimized BERT Approach*). Model ini merupakan hasil optimalisasi dari BERT, dengan peningkatan pada teknik pre-training yang lebih efisien, seperti penggunaan jumlah data yang lebih besar, batch size yang lebih tinggi, serta eliminasi tugas *Next Sentence Prediction* (NSP). Dengan optimasi tersebut, IndoRoBERTa mampu memahami konteks bahasa Indonesia dengan lebih baik, sehingga meningkatkan performa dalam tugas-tugas seperti analisis sentimen, klasifikasi teks, dan pemodelan bahasa lainnya. Pada Gambar 6 ditampilkan arsitektur IndoRoBERTa, yang merupakan pengembangan dari model RoBERTa (*Robustly Optimized BERT Approach*). IndoRoBERTa dirancang untuk meningkatkan kinerja pemrosesan bahasa alami (NLP) dalam bahasa Indonesia dengan optimasi yang lebih baik dibandingkan model BERT standar. Dalam penelitian ini, model yang digunakan adalah w11wo/indonesian-robertabase-sentiment-classifier yang telah melalui proses pra-pelatihan secara ekstensif pada korpus bahasa Indonesia untuk memahami struktur dan makna bahasa dengan lebih akurat. IndoRoBERTa bekerja dengan cara memproses token dari kalimat yang diberikan, di mana setiap token direpresentasikan dalam bentuk vektor embedding. Model ini menggunakan token khusus seperti [CLS] sebagai representasi keseluruhan kalimat dan [SEP] untuk memisahkan dua kalimat jika ada lebih dari satu input. Setelah melalui lapisan transformer yang terdiri dari mekanisme *self-attention* dan *feed-forward neural networks*, model menghasilkan keluaran yang digunakan untuk klasifikasi sentimen. Dengan arsitektur yang lebih fleksibel dan teknik pelatihan yang lebih optimal dibandingkan BERT, IndoRoBERTa mampu menangkap hubungan antar kata dalam teks dengan lebih baik, sehingga menghasilkan akurasi yang lebih tinggi dalam tugas klasifikasi sentimen. Model ini menjadi pilihan yang lebih andal untuk analisis sentimen dalam bahasa Indonesia, terutama dalam konteks penelitian ini yang berfokus pada evaluasi sentimen terhadap film dokumenter *Dirty Vote*.



Gambar 6. Arsitektur IndoBERTa

F. Evaluation

Hasil evaluasi model dianalisis untuk membandingkan kinerja antara IndoBERT dan IndoRoBERTa. Perbandingan ini dilakukan dengan mempertimbangkan beberapa faktor yang mempengaruhi performa model, seperti arsitektur model, metode pembagian dataset menggunakan *K-Fold Cross Validation* serta keragaman sentimen dalam data latih. Evaluasi dilakukan dengan mengukur *accuracy*, *precision*, *recall*, dan *F1-score* berdasarkan confusion matrix dari masing-masing model. Analisis hasil evaluasi ini bertujuan untuk mengidentifikasi keunggulan serta kelemahan dari masing-masing model, sehingga dapat disimpulkan model mana yang lebih optimal dalam mengklasifikasikan sentimen berdasarkan metrik evaluasi yang digunakan.

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

Akurasi (*accuracy*) adalah nilai yang dihitung dengan membandingkan jumlah label yang diprediksi dengan benar terhadap total seluruh data.

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

Presisi (*precision*) adalah matrik yang mengukur seberapa besar proporsi prediksi positif yang benar dibandingkan dengan total prediksi positif yang dihasilkan model

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

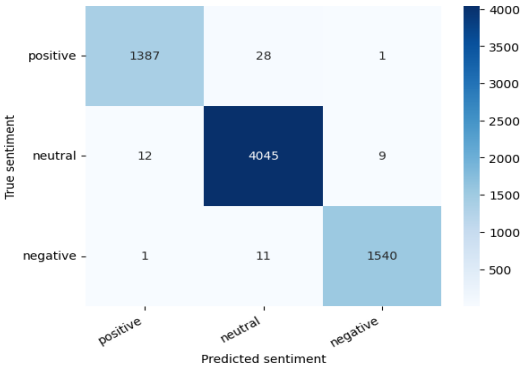
Sensitivitas (*recall*) merupakan metrik yang mengukur proporsi kasus positif yang berhasil diidentifikasi oleh model

$$F1\ Score = 2 \times \frac{Recall \times Precision}{Recall - Precision} \quad (4)$$

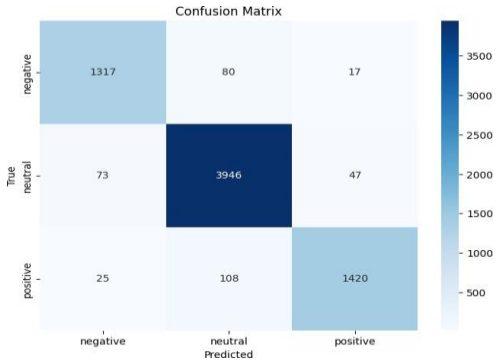
F1 score adalah metrik yang mengukur perbandingan rata-rata harmonis antara recall dan precision

III. HASIL DAN PEMBAHASAN

Dalam tahapan penelitian ini, akan dibahas hasil evaluasi kinerja masing-masing model, yaitu IndoBERT dan IndoRoBERTa, dalam melakukan analisis sentimen terhadap film dokumenter *Dirty Vote*. Analisis ini bertujuan untuk mengukur sejauh mana kedua model mampu memahami dan mengklasifikasikan sentimen dalam berbagai ulasan atau opini publik yang beredar terkait film tersebut. Evaluasi dilakukan berdasarkan sejumlah metrik yang umum digunakan dalam pemodelan klasifikasi, seperti *accuracy*, *precision*, *recall*, dan *F1-score* guna menilai performa masing-masing model secara objektif. Dengan demikian, penelitian ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai efektivitas IndoBERT dan IndoRoBERTa dalam menangani tugas analisis sentimen pada teks berbahasa Indonesia, khususnya dalam konteks opini masyarakat terhadap isu-isu politik yang diangkat dalam film dokumenter ini. Pada Gambar 7, ditampilkan confusion matrix dari model IndoBERT sebagai representasi hasil evaluasi klasifikasi sentimen. Berdasarkan hasil analisis, model ini mampu memprediksi sentimen positif dengan benar sebanyak 1.387 kali, sentimen netral sebanyak 4.045 kali, dan sentimen negatif sebanyak 1.540 kali. Dari data tersebut, dapat diamati bahwa IndoBERT memiliki tingkat akurasi yang lebih tinggi dalam mengklasifikasikan sentimen netral dibandingkan dengan kategori sentimen lainnya. Hal ini mengindikasikan bahwa model memiliki kemampuan yang baik dalam memahami teks dalam bahasa Indonesia, khususnya dalam mengenali opini yang cenderung bersifat netral. Meskipun demikian, terdapat potensi peningkatan dalam klasifikasi sentimen positif dan negatif agar distribusi prediksi lebih seimbang dan akurat. Sementara itu, pada Gambar 8, ditampilkan confusion matrix dari model IndoRoBERTa, yang menunjukkan hasil klasifikasi sentimen ke dalam tiga kategori utama: negatif, netral, dan positif. Dari hasil evaluasi, model ini memiliki performa yang baik dalam mengklasifikasikan sentimen netral, dengan jumlah prediksi benar sebanyak 3.946 kali. Selain itu, IndoRoBERTa juga berhasil mengklasifikasikan sentimen negatif dengan benar sebanyak 1.317 kali dan sentimen positif sebanyak 1.420 kali. Meskipun secara keseluruhan model menunjukkan tingkat akurasi yang tinggi, masih terdapat beberapa kesalahan klasifikasi, seperti 80 kasus di mana sentimen negatif diklasifikasikan sebagai netral dan 108 kasus di mana sentimen positif salah dikategorikan sebagai netral. Kesalahan ini menunjukkan bahwa meskipun model memiliki kinerja yang cukup baik dalam menangkap makna dan emosi dalam teks, masih terdapat potensi perbaikan dalam membedakan sentimen yang lebih bernuansa. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa baik IndoBERT maupun IndoRoBERTa memiliki kemampuan yang cukup baik dalam melakukan analisis sentimen terhadap teks berbahasa Indonesia, khususnya dalam konteks opini publik terhadap film dokumenter *Dirty Vote*. Meskipun terdapat beberapa perbedaan dalam tingkat akurasi antar kategori sentimen, kedua model tetap menunjukkan performa yang kompetitif dalam tugas klasifikasi sentimen. Optimalisasi lebih lanjut, seperti penyesuaian hyperparameter, penggunaan data *augmentation* atau *fine-tuning* dengan dataset yang lebih beragam, dapat menjadi langkah strategis untuk meningkatkan akurasi dan keseimbangan klasifikasi pada model di masa mendatang.



Gambar 7. Confusion Matrix IndoBERT



Gambar 8. Confusion Matrix IndoBERTa.

Pada Tabel 2, hasil evaluasi menunjukkan bahwa model memiliki performa yang sangat baik dalam mengklasifikasikan sentimen. Hal ini ditunjukkan dengan nilai *F1-score, precision*, dan *recall* yang masing-masing mencapai 99%. Nilai tersebut menunjukkan bahwa model mampu mengidentifikasi sentimen dengan tingkat akurasi yang tinggi, baik dalam mendeteksi sentimen positif, netral, maupun negatif. Dengan akurasi keseluruhan sebesar 99%, model ini dapat diandalkan untuk melakukan analisis sentimen dengan tingkat kesalahan yang sangat rendah.

TABEL 2
CONTOH RATA-RATA K-FOLD INDOBERT MODEL

	Precision	*Recall*	*F1-score*	*Support*
Positive	0.99	0.98	0.99	1416
Neutral	0.99	0.99	0.99	4066
Negative	0.99	0.99	0.99	1552
Accuracy			0.99	7034
Macro Avg	0.99	0.99	0.99	7034
Weighted Avg	0.99	9.99	0.99	7034

Pada Tabel 3, hasil evaluasi menggunakan metode *K-Fold Cross Validation* menunjukkan bahwa model IndoRoBERTa memiliki akurasi rata-rata sebesar 94%, yang mengindikasikan bahwa model mampu memprediksi sebagian besar sentimen dengan benar. Nilai *precision* yang tinggi, berkisar antara 90% hingga 95%, menunjukkan bahwa ketika model mengklasifikasikan teks ke dalam suatu kelas, kemungkinan besar teks tersebut memang berasal dari kelas tersebut. Selain itu, nilai recall yang tinggi, yaitu 91% hingga 95%, menandakan bahwa model mampu mengidentifikasi sebagian besar sentimen yang sebenarnya berasal dari kelas tertentu. Skor F*1-score* yang tinggi, berkisar antara 91% hingga 95%, menunjukkan keseimbangan yang baik antara precision dan recall. Hal ini mengindikasikan bahwa model IndoRoBERTa memiliki kemampuan yang sangat baik dalam mengidentifikasi serta membedakan setiap kategori sentimen secara optimal.

TABEL 3
CONTOH TABEL RATA-RATA K-FOLD INDOBERTA MODEL

	Precision	*Recall*	*F1-score*	*Support*
Positive	0.90	0.92	0.91	1416
Neutral	0.95	0.95	0.95	4065
Negative	0.93	0.91	0.92	1552
Accuracy			0.94	7033
Macro Avg	0.93	0.93	0.93	7033
Weighted Avg	0.9	9.94	0.94	7033

Berdasarkan hasil analisis perbandingan antara model IndoBERT dan IndoRoBERTa dalam aspek arsitektur, terdapat perbedaan utama dalam teknik masking yang digunakan. IndoBERT menggunakan static masking, sedangkan IndoRoBERTa menggunakan dynamic masking, meskipun keduanya menggunakan parameter yang sama. Dalam hal performa, IndoBERT menunjukkan keunggulan yang signifikan ketika diterapkan dengan pembagian dataset menggunakan *K-Fold Cross Validation*. Hasil evaluasi menunjukkan bahwa IndoBERT memiliki rata-rata akurasi sebesar 99%, dengan *precision* 99%, *recall* 99%, dan *F1-score* 99%. Sementara itu, IndoRoBERTa memiliki rata-rata akurasi sebesar 94%, dengan *precision* 94%, *recall* 94%, dan *F1-score* 94%. Hasil ini menunjukkan bahwa IndoBERT lebih unggul dalam menangani tugas klasifikasi sentimen dibandingkan dengan IndoRoBERTa, terutama dalam konteks data yang diuji pada penelitian ini. Berdasarkan Tabel 4 yang membandingkan arsitektur model IndoBERT dan IndoRoBERTa, terdapat beberapa perbedaan signifikan yang mempengaruhi kinerja kedua model. IndoBERT memiliki keunggulan dalam pemrosesan teks dengan panjang token maksimum yang lebih besar, yaitu 512 token, sehingga lebih sesuai untuk menangani teks yang lebih panjang dibandingkan IndoRoBERTa yang hanya mendukung panjang token hingga 128. Dari segi cakupan kosakata, IndoRoBERTa memiliki jumlah token yang lebih banyak, yaitu 50.265 token, dibandingkan dengan IndoBERT yang hanya memiliki 50.000 token. Hal ini memungkinkan IndoRoBERTa mengenali lebih banyak variasi kata dalam bahasa Indonesia. Selain itu, IndoRoBERTa memiliki nilai learning rate yang lebih rendah, yaitu 1e-5, yang dapat membantu dalam proses fine-tuning pada data yang lebih kompleks.IndoRoBERTa juga memiliki nilai epsilon yang lebih besar, yaitu 1e-5, yang menunjukkan toleransi lebih tinggi terhadap pembagian dengan nol, sehingga dapat meningkatkan stabilitas model selama proses pelatihan. Namun, dari segi efisiensi komputasi, IndoBERT lebih unggul karena membutuhkan waktu pelatihan yang lebih singkat, yaitu 10 menit 41 detik per epoch, dibandingkan dengan IndoRoBERTa yang memerlukan waktu 21 menit 09 detik per *epoch*. Dengan mempertimbangkan aspek-aspek tersebut, IndoBERT lebih efisien dalam pemrosesan teks panjang dan memiliki waktu pelatihan yang lebih cepat. Sementara itu, IndoRoBERTa unggul dalam cakupan kosakata yang lebih luas serta stabilitas dalam pelatihan.

TABEL 4
PERBANDINGAN ARSITEKTUR MODEL

Feature	*IndoBERT*	*IndoRoBERTa*
Total Layer	*12 (bert.layer)*	*12 (roberta.encoder. layer)*
Maximum Token	*512*	*512*
Dimentional Embedding	*768*	*768*
Token vocabulary	*50.000*	*50.265*
Max Sequence Length	*512*	*128*
Padding Token	*0*	*1*
Token Type Embeddings	*2*	*1*
Learning Rate	*2e-5 (0.00002)*	*1e-5 (0.00001)*
Epsilon	*1e-12 (0.000000000001)*	*1e-5 (0.00001)*
Activation Function (Intermediate	*GELU (Gaussian Error Linear Unit)*	*GELU (Gaussian Error Linear Unit)*
Activation Function (Pooler Layer)	*Tanh*	*Tanh*
Masking Method	*Static Masking*	*Dynamic Masking*
Training Times using T4 GPU	*10 Minutes 41 Second/Epoch*	*21 Minutes 09 Second/Epoch*

Berdasarkan Tabel 5 yang menunjukkan hasil klasifikasi sentimen, terdapat perbedaan dalam cara IndoBERT dan IndoRoBERTa mengategorikan beberapa komentar. IndoBERT cenderung memberikan klasifikasi yang lebih tegas terhadap opini atau kritik tajam, sementara IndoRoBERTa lebih sering menganggap komentar serupa sebagai sentimen netral. Pada komentar *"Itu ga nyudutin paslon tertentu. Semua ada minusnya ada pelanggarannya. Cm ya kalo paslon tertentu banyak minusnya itu salah sendiri, siapa suruh bikin banyak pelanggaran?"*, IndoBERT mengklasifikasikannya sebagai negatif, sedangkan IndoRoBERTa mengategorikannya sebagai netral. Hal serupa juga terjadi pada komentar *"SALAM CERDAS SEBELUM TANGGAL 14. Salam 4 jari"*, yang dikategorikan sebagai positif oleh IndoBERT, tetapi dinilai sebagai negatif

oleh IndoRoBERTa. Pada beberapa komentar lainnya, kedua model memberikan hasil klasifikasi yang sama. Misalnya, komentar *"Terima kasih sudah membuat banyak orang membuka mata dan sadar atas betapa gentingnya keadaan Indonesia saat ini dan Terima kasih sudah membuat dikumenter ini, semoga orang-orang yang masih tutup mata bisa tercerahkan."* dikategorikan sebagai netral oleh kedua model. Konsistensi ini menunjukkan bahwa terdapat beberapa jenis teks yang dikenali dengan cara yang sama oleh kedua model tanpa adanya perbedaan dalam interpretasi sentimen. Perbedaan hasil klasifikasi antara IndoBERT dan IndoRoBERTa mencerminkan bagaimana kedua model memproses informasi sentimen dalam teks. IndoBERT, yang menggunakan pendekatan *static masking*, lebih sensitif terhadap kata-kata yang menunjukkan opini atau kritik, sehingga lebih sering memberikan klasifikasi positif atau negatif. Di sisi lain, IndoRoBERTa yang menerapkan *dynamic masking* lebih fleksibel dalam menangkap konteks, sehingga lebih banyak mengategorikan komentar sebagai netral. Hasil ini menunjukkan bahwa arsitektur model dan metode pretraining yang digunakan dapat memengaruhi hasil klasifikasi sentimen. IndoBERT cenderung menghasilkan prediksi yang lebih tegas terhadap komentar yang mengandung opini atau kritik, sedangkan IndoRoBERTa lebih sering memberikan hasil netral pada komentar yang mengandung ambiguitas.

TABEL 5
HASIL SENTIMEN MODEL

Comments	IndoBERT	IndoRoBERTa
Itu ga nyudutin paslon tertentu. Semua ada minusnya ada pelanggarannya. Cm ya kalo paslon tertentu banyak minusnya itu salah sendiri, siapa suruh bikin banyak pelanggaran?	Negative	Neutral
SALAM CERDAS SEBELUM TANGGAL 14. Salam 4 jari	Positive	Negative
Film politik,gak usah ada film, rakyat pilihanya sapa udah tau, takut kalah ya	Negative	Neutral
Terima kasih sudah membuat banyak ornag membuka mata dan sadar atas betapa gentingya keadaan Indonesia saat ini.	Neutral	Neutral
Terima kasih sudah membuat dikumenter ini, semoga orang-orang yang masih tutup mata bisa tercerahkan!	Neutral	Neutral

Berdasarkan Tabel 6 yang menyajikan hasil perbandingan antara skor leksikon dan klasifikasi sentimen menggunakan model IndoBERT dan IndoRoBERTa terhadap komentar-komentar yang berkaitan dengan film dokumenter *Dirty Vote*, ditemukan adanya perbedaan yang cukup signifikan dalam hal interpretasi sentimen antara kedua model. Untuk memberikan kejelasan dalam analisis ini, penelitian menggunakan skala sentimen yang disusun berdasarkan nilai skor leksikon, di mana skor bernilai antara -2 hingga -1 dikategorikan sebagai sentimen negatif, skor 0 dikategorikan sebagai sentimen netral, dan skor antara +1 hingga +2 dikategorikan sebagai sentimen positif. Skor-skor tersebut ditentukan melalui pendekatan berbasis kamus leksikal sentimen berbahasa Indonesia yang telah disesuaikan dengan konteks linguistik dan sosial masyarakat Indonesia. Sebagai contoh, kata "salah" yang memiliki skor leksikon -2, secara semantik mengandung makna negatif. Model IndoBERT mampu mengklasifikasikan kata ini secara tepat sebagai sentimen negatif, sedangkan IndoRoBERTa justru mengklasifikasikannya sebagai sentimen netral. Hal yang sama terjadi pada kata "cerdas", yang memiliki skor leksikon +2 dan seharusnya dikategorikan sebagai sentimen positif. IndoBERT kembali memberikan hasil klasifikasi yang sesuai, sementara IndoRoBERTa secara keliru mengklasifikasikannya sebagai sentimen negatif. Pada kata "takut" yang juga memiliki skor leksikon -2, IndoBERT mengidentifikasinya sebagai sentimen negatif, sedangkan IndoRoBERTa kembali mengklasifikasikannya sebagai sentimen netral. Perbandingan ini menunjukkan bahwa model IndoBERT memiliki tingkat konsistensi yang lebih tinggi antara skor leksikon dan hasil klasifikasi sentimen. Dengan demikian, IndoBERT dapat dikatakan lebih akurat dalam menginterpretasikan makna emosional dari suatu kata dalam konteks kalimat. Sebaliknya, IndoRoBERTa cenderung menghasilkan klasifikasi netral terhadap kata-kata yang secara leksikal mengandung makna positif maupun negatif yang kuat.

Hal ini mengindikasikan bahwa tingkat akurasi IndoRoBERTa dalam memetakan sentimen pada data validasi dalam penelitian ini relatif lebih rendah dibandingkan IndoBERT.

TABEL 6
HASIL SENTIMEN MODEL

Comments	IndoBERT	IndoRoBERTa
Itu ga nyudutin paslon tertentu. Semua ada minusnya ada pelanggarannya. Cm ya kalo paslon tertentu banyak minusnya itu salah sendiri, siapa suruh bikin banyak pelanggaran?	Negative (-2)	Neutral (-2)
SALAM CERDAS SEBELUM TANGGAL 14. Salam 4 jari	Positive (2)	Negative (2)
Film politik,gak usah ada film, rakyat pilihanya sapa udah tau, takut kalah ya	Negative (-2)	Neutral (-2)
Terima kasih sudah membuat banyak ornag membuka mata dan sadar atas betapa gentingya keadaan Indonesia saat ini.	Neutral (0)	Neutral (0)
Terima kasih sudah membuat dikumenter ini, semoga orang-orang yang masih tutup mata bisa tercerahkan!	Neutral (0)	Neutral (0)

IV. SIMPULAN

Penelitian ini membandingkan kinerja dua model berbasis Transformer, yaitu IndoBERT dan IndoRoBERTa, dalam melakukan analisis sentimen terhadap komentar-komentar publik terkait film dokumenter *Dirty Vote*. Hasil evaluasi menunjukkan bahwa model IndoBERT mencapai rata-rata akurasi sebesar 99%, sementara IndoRoBERTa mencapai akurasi sebesar 94%. Berdasarkan skala pengkategorian akurasi yang digunakan dalam penelitian ini, keduanya termasuk dalam kategori sangat tinggi. Namun demikian, IndoBERT menunjukkan performa yang lebih unggul secara keseluruhan, baik dari segi akurasi, efisiensi pelatihan, maupun konsistensi hasil klasifikasi terhadap pelabelan berbasis leksikon. Dari sisi arsitektur, perbedaan antara pendekatan *static masking* pada IndoBERT dan *dynamic masking* pada IndoRoBERTa turut memengaruhi sensitivitas model terhadap konteks linguistik. IndoBERT menunjukkan keunggulan dalam menangkap opini atau kritik secara eksplisit, sedangkan IndoRoBERTa cenderung mengategorikan komentar ambigu sebagai netral. Hal ini menunjukkan bahwa IndoBERT lebih akurat dan konsisten dalam menangani ekspresi emosional dalam teks berbahasa Indonesia, khususnya pada topik-topik sosial-politik.

Secara ilmiah, penelitian ini memberikan kontribusi penting dalam pengembangan model analisis sentimen berbasis Transformer untuk bahasa Indonesia. Penelitian ini juga memperkuat bukti empiris mengenai efektivitas IndoBERT dalam konteks klasifikasi sentimen terhadap data nyata di media sosial. Adapun manfaat praktis dari temuan ini dapat diterapkan dalam pengembangan sistem pemantauan opini publik, analisis isu politik, dan pengambilan kebijakan berbasis data dalam ranah digital. Sebagai arah penelitian lanjutan (*future work*), disarankan untuk memperluas cakupan data pelatihan dengan melibatkan sumber dan topik yang lebih beragam agar model mampu melakukan generalisasi yang lebih baik. Selain itu, eksplorasi metode hibrida yang menggabungkan pendekatan leksikal dan pembelajaran mendalam dapat menjadi alternatif untuk meningkatkan akurasi pada komentar dengan makna implisit atau bernuansa. Penelitian selanjutnya juga dapat mempertimbangkan penggunaan analisis multimodal, seperti kombinasi teks dan video, untuk memahami sentimen publik secara lebih komprehensif.

DAFTAR PUSTAKA

[1] A. Aditia and A. P. Riyandi, "Pengertian Film Dokumenter: Definisi, Jenis dan Contohnya," Kompas Entertainment, 17 Oktober 2022. [Online]. Available: https://entertainment.kompas.com/read/2022/10/17/173309466/pengertian-film-dokumenter-definisi-jenis-dan-contohnya
[2] YouTube, "Dirty Vote - YouTube View Count," 2024. [Online]. Available: https://www.youtube.com
[3] Google Trends, "Tren Pencarian 'Dirty Vote' di Indonesia (Februari 2024)," 2024. [Online]. Available: https://trends.google.com
[4] K. Munger and J. Phillips, "A Supply and Demand Framework for YouTube Politics," Public Opin. Q., vol. 86, no. S1, pp. 161–188, 2022. [Online]. Available: https://doi.org/10.1093/poq/nfac002
[5] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Found. Trends Inf. Retr., vol. 2, no. 1–2, pp. 1–135, 2008. [Online]. Available: https://doi.org/10.1561/1500000001
[6] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," IEEE Intell. Syst., vol. 28, no. 2, pp. 15–21, 2013. [Online]. Available: https://doi.org/10.1109/MIS.2013.30
[7] B. Liu, Sentiment Analysis and Opinion Mining, vol. 5, no. 1. Synth. Lect. Hum. Lang. Technol., 2012, pp. 1–167. [Online]. Available: https://doi.org/10.2200/S00416ED1V01Y201204HLT016
[8] R. Feldman, "Techniques and Applications for Sentiment Analysis," Commun. ACM, vol. 56, no. 4, pp. 82–89, 2013. [Online]. Available: https://doi.org/10.1145/2436256.2436274
[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1810.04805
[10] B. Wilie et al., "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," arXiv preprint arXiv:2009.05387, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2009.05387
[11] J. U. S. Lazuardi and A. Juarna, "Analisis Sentimen Ulasan Pengguna Aplikasi JOOX pada Android Menggunakan Metode BERT," J. Ilm. Inform. Komput., vol. 28, no. 3, pp. 251–260, 2023. [Online]. Available: https://doi.org/10.35760/ik.2023.v28i3.10090
[12] N. A. R. Putri and Ardiansyah, "Analisis Sentimen terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa," J. Sains Inform., vol. 9, no. 2, pp. 136–145, 2023. [Online]. Available: https://doi.org/10.34128/jsi.v9i2.649
[13] E. P. A. Akhmad, "Analisis Sentimen Ulasan Aplikasi DLU Ferry pada Google Play Store Menggunakan BERT," J. Apl. Pelayaran Kepelabuhanan, vol. 13, no. 2, pp. 104–112, 2023. [Online]. Available: https://doi.org/10.30649/japk.v13i2.94
[14] K. Nahar, A. Jaradat, M. Atoum, and F. Ibrahim, "Sentiment Analysis and Classification of Arab Jordanian Facebook Comments Using Machine Learning," Jordanian J. Comput. Inf. Technol., vol. 0, no. 1, p. 1, 2020. [Online]. Available: https://doi.org/10.5455/jjcit.71-1586289399
[15] K. Koto et al., "IndoBERT: A Pre-trained Indonesian-Specific Language Model," in Proc. 28th Int. Conf. Comput. Linguist. (COLING), 2020.
[16] F. Pratama and H. Wibowo, "Improving Indonesian Sentiment Analysis Using IndoRoBERTa," Indones. J. Comput. Cybern. Syst., vol. 15, no. 3, pp. 198–210, 2021.
[17] A. S. Ramadhan, S. Wibowo, and M. Kurniawan, "Comparing IndoBERT, LSTM, and SVM for Sentiment Analysis on Indonesian News Data," J. Inf. Syst. Eng. Bus. Intell., vol. 9, no. 1, pp. 23–31, 2022.
[18] M. Runimeirati, A. Muis, and F. Muhammad, "Pelatihan Text Mining Menggunakan Bahasa Pemrograman Python," Abdimas Langkanae, vol. 3, no. 1, pp. 36–46, 2023. [Online]. Available: https://doi.org/10.53769/abdimas.3.1.2023.83
[19] A. Wijaya, "Comparative Analysis of Transformer-based Models for Indonesian Sentiment Analysis," Int. J. Artif. Intell. Res., vol. 10, no. 2, pp. 87–95, 2023.