

Analisis Sentimen Pengguna X dan Youtube Terhadap Carmen Hearts2Hearts Menggunakan Metode IndoBERT

Fellycia Caroline^{a1}, Syalsabilla Valentisyesa^{a2}, Muhammad Rizky Pribadi^{a3}

^aProgram Studi Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang, Sumatera Selatan

Jl. Rajawali No.14, 9 Ilir, Kec. Ilir Tim. II, Kota Palembang, Sumatera Selatan, Indonesia

¹fellyciacaroline_2327250010@mhs.mdp.ac.id

²syalsabillavalentisyesa_2327250054@mhs.mdp.ac.id

³rizky@mdp.ac.id

Abstract

The rapid growth of social media has increased the amount of public opinion expressed online, particularly on platforms such as X and YouTube, where users actively share their views regarding public figures and entertainment topics. This study aims to analyze public sentiment toward Carmen, a member of the K-pop group Hearts2Hearts, using the IndoBERT model for sentiment classification. Data were collected from X and YouTube comments through web scraping techniques and combined into a single dataset to obtain more diverse opinions. The research process involved several stages, including text preprocessing, manual sentiment labeling, dataset splitting, model training, and evaluation. The preprocessing stage consisted of duplicate data removal, case folding, noise removal, tokenization, stopword removal, and stemming to improve data quality before classification. The dataset was categorized into three sentiment classes: positive, neutral, and negative, then divided into training and testing data using an 80:20 ratio. The IndoBERT model was trained using transformer-based deep learning to understand the context of Indonesian-language text more effectively. Evaluation results showed that the model achieved an accuracy of 72.41%, precision of 75.82%, recall of 72.41%, and F1-score of 71.15%, indicating that IndoBERT performs effectively in classifying sentiment on Indonesian social media data despite challenges such as informal language and ambiguous expressions.

Keywords: Sentiment Analysis, Carmen, IndoBERT, Natural Language Processing

1. Pendahuluan

Perkembangan media sosial telah meningkatkan jumlah opini publik yang disampaikan secara daring, khususnya pada platform X yang memungkinkan pengguna memberikan respons secara cepat terhadap isu tertentu [1]. Penggunaan analisis sentimen di Indonesia juga mengalami peningkatan seiring banyaknya data opini yang berasal dari media sosial [2]. Media sosial dan forum daring dapat digunakan sebagai sumber data untuk menganalisis diskursus dan sentimen publik terhadap suatu isu [3]. Dalam industri hiburan, khususnya K-pop, interaksi penggemar di media sosial berlangsung sangat aktif sehingga idol maupun grup musik sering menjadi topik pembicaraan publik di berbagai platform digital. Salah satu figur yang menjadi perhatian publik adalah Carmen, anggota grup Hearts2Hearts. Opini terkait Carmen tidak hanya ditemukan pada platform X, tetapi juga pada kolom komentar YouTube yang digunakan penggemar untuk memberikan respons terhadap konten video, penampilan, maupun aktivitas grup. Banyaknya opini yang muncul pada kedua platform tersebut menghasilkan data teks dalam jumlah besar yang dapat dimanfaatkan untuk mengetahui persepsi publik melalui analisis sentimen.

Analisis sentimen merupakan teknik dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi polaritas opini dari suatu teks, seperti positif, negatif, dan

netral. Analisis sentimen pada media sosial banyak digunakan untuk mengetahui respons dan persepsi publik terhadap suatu isu berdasarkan data teks yang dihasilkan pengguna [4]. Namun, analisis sentimen pada data media sosial memiliki tantangan tersendiri karena teks pada platform X maupun komentar YouTube umumnya menggunakan bahasa informal, singkatan, campuran bahasa, serta istilah slang yang sulit dipahami oleh metode klasifikasi tradisional [5]. Selain itu, penggunaan satu sumber data saja terkadang menghasilkan jumlah data yang terbatas sehingga diperlukan penggabungan data dari beberapa platform media sosial untuk memperoleh variasi opini yang lebih beragam.

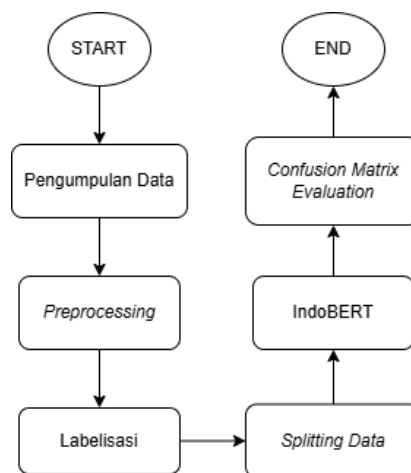
Seiring berkembangnya teknologi *deep learning*, model berbasis *transformer* seperti BERT mampu memahami konteks kalimat secara lebih baik dibandingkan metode *machine learning* konvensional [6]. Untuk pemrosesan bahasa Indonesia, dikembangkan IndoBERT yang telah dilatih khusus menggunakan korpus berbahasa Indonesia sehingga lebih efektif dalam memahami konteks teks bahasa Indonesia [7]. Penggunaan IndoBERT dalam analisis sentimen juga menunjukkan performa yang baik pada berbagai penelitian terkini karena mampu menangani bahasa informal dan konteks kalimat yang kompleks pada media sosial [8].

Penelitian yang dilakukan oleh Putri dkk. menggunakan IndoBERT untuk menganalisis sentimen Twitter mengenai Pemilihan Presiden Indonesia 2024 dan memperoleh nilai akurasi sebesar 60% pada proses klasifikasi sentiment [9]. Penelitian lain oleh Alfatah menerapkan model *transformer* IndoBERT pada data Twitter Indonesia dan menunjukkan bahwa IndoBERT memperoleh performa yang lebih baik dibandingkan beberapa metode klasifikasi tradisional dalam analisis sentiment [10]. Selain itu, penelitian oleh Pota dkk. menunjukkan bahwa model berbasis BERT efektif digunakan dalam analisis sentimen Twitter karena mampu memahami konteks teks secara lebih baik pada data media sosial [11].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk melakukan analisis sentimen terhadap opini pengguna X dan komentar YouTube mengenai Carmen Hearts2Hearts menggunakan metode IndoBERT. Penelitian ini diharapkan dapat memberikan gambaran mengenai persepsi publik terhadap Carmen serta menjadi referensi dalam penerapan model *transformer* untuk analisis sentimen media sosial berbahasa Indonesia.

2. Metode Penelitian

Penelitian ini menggunakan metode analisis sentimen berbasis *Natural Language Processing* (NLP) dengan model IndoBERT untuk mengklasifikasikan opini publik terhadap Carmen pada platform X dan YouTube. Tahapan penelitian dilakukan secara bertahap mulai dari pengumpulan data hingga evaluasi model untuk memperoleh hasil analisis sentimen yang sesuai dengan tujuan penelitian. Alur tahapan penelitian dapat dilihat pada Gambar 1.



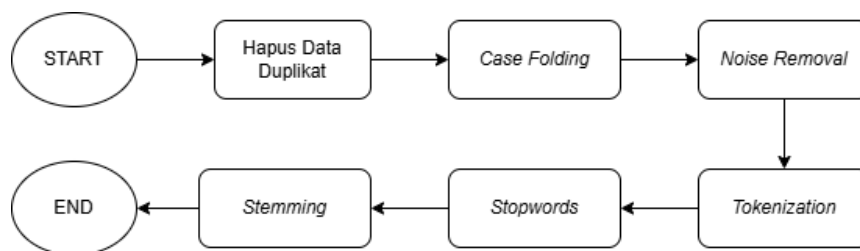
Gambar 1. Alur Tahapan Penelitian

2.1. Pengumpulan Data

Pada tahap pengumpulan data, data diperoleh dari platform X dan kolom komentar YouTube yang berkaitan dengan Carmen Hearts2Hearts. Pengambilan data dilakukan dengan menggunakan *scraping*. *Web Data Scraping* (WDS) merupakan proses pengambilan serta pengumpulan data dari *website* secara otomatis dan terstruktur untuk memperoleh informasi yang dibutuhkan [12]. Pada platform X dilakukan *scraping* berdasarkan kata kunci seperti “Carmen”, “Carmen Hearts2Hearts”, “Carmen H2H”, “Carmen Vocal”, “Carmen Suara”, dan “Carmen Fancam”. Selain itu, data komentar YouTube diperoleh dari video yang membahas Carmen. Penggabungan data dari kedua platform dilakukan karena jumlah data pada platform X dinilai masih terbatas sehingga diperlukan sumber data tambahan untuk memperoleh variasi opini yang lebih beragam.

2.2. Text Preprocessing

Text Preprocessing merupakan tahapan awal dalam proses klasifikasi teks yang bertujuan untuk menyiapkan data sebelum memasuki tahap pengolahan berikutnya. Pada tahap ini, teks akan dibersihkan dan diubah ke dalam bentuk yang lebih terstruktur agar informasi yang dihasilkan menjadi lebih baik dan dapat digunakan secara optimal pada proses selanjutnya [13]. *Preprocessing* terdiri dari beberapa tahapan yang dilakukan untuk membersihkan serta menyiapkan teks ulasan sebelum diproses lebih lanjut. Alur dari tahapan tersebut dapat dilihat pada Gambar 2.



Gambar 2. Tahapan Preprocessing

a. Hapus Data Duplikat

Proses penghapusan data duplikat dilakukan untuk menjaga agar data tetap akurat dan konsisten serta mengurangi adanya data yang berulang. Tahapan ini penting dalam analisis data karena dapat meminimalkan bias, mempercepat proses pengolahan data, dan meningkatkan kualitas hasil analisis yang diperoleh [14].

b. Case Folding

Case Folding merupakan tahapan yang dilakukan dengan mengubah seluruh huruf dalam dokumen menjadi huruf kecil agar format teks menjadi lebih seragam. Sementara itu, karakter selain huruf akan dianggap sebagai delimiter atau tanda pemisah antar kata [15].

c. Noise Removal

Noise removal merupakan tahapan yang dilakukan untuk menghapus unsur-unsur yang tidak diperlukan dalam proses analisis sentimen, seperti karakter khusus, tanda baca, URL, serta kata-kata penghubung yang tidak memiliki pengaruh terhadap hasil analisis [16].

d. Tokenization

Tokenization merupakan proses memecah teks ulasan menjadi kumpulan kata-kata. Pada tahap ini, tanda baca maupun simbol yang bukan termasuk huruf akan dihilangkan, seperti “ , . , / , ; ,) dan karakter lainnya yang tidak diperlukan dalam proses analisis [17].

e. Stopwords

Stopword merupakan salah satu tahapan preprocessing yang dilakukan dengan menghapus kata-kata umum yang dianggap tidak memiliki makna penting dalam proses analisis teks. Kata-kata tersebut biasanya sering muncul dalam kalimat, tetapi tidak memberikan pengaruh besar terhadap informasi utama yang ingin dianalisis. Pada tahap ini, kata-kata yang terdapat dalam daftar stopwords akan dihapus dari dataset agar proses pengolahan data menjadi lebih efektif. Penghapusan stopwords juga dapat mengurangi dimensi data, meringankan beban komputasi, serta membantu meningkatkan kinerja algoritma machine learning dalam proses analisis [18].

f. Stemming

Stemming merupakan tahapan preprocessing yang dilakukan dengan mengubah kata berimbuhan menjadi kata dasar melalui penghapusan prefiks, sufiks, maupun infiks [19]. Contohnya seperti kata “berjalan” menjadi “jalan” dan “makanan” menjadi “makan”.

2.3. Labelisasi

Tahap labelisasi data dilakukan setelah proses *preprocessing* selesai dilakukan. Pada tahap ini, setiap data teks dari platform X dan komentar YouTube diberikan label sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral. Penjelasan dari masing-masing kategori sentimen adalah sebagai berikut [19]:

a. Sentimen Positif

Komentar yang menunjukkan perasaan positif, seperti pujian, dukungan, atau rasa senang.

b. Sentiment Negatif

Komentar yang mengandung perasaan negatif, seperti kritik, kekecewaan, atau kemarahan.

c. Sentiment Netral

Komentar yang bersifat netral, tidak berkaitan dengan topik pembahasan, atau tidak menunjukkan emosi yang jelas.

Proses pelabelan dilakukan secara manual dengan membaca dan memahami konteks setiap komentar agar hasil labelisasi lebih sesuai dengan makna opini yang disampaikan pengguna. Tahap ini bertujuan untuk menghasilkan dataset yang dapat digunakan pada proses pelatihan model IndoBERT sehingga model mampu mempelajari pola sentimen dari data teks berbahasa Indonesia.

2.4. Splitting Data

Tahap *splitting data* dilakukan setelah proses labelisasi selesai dilakukan. Pada tahap ini, dataset yang telah diberi label dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Pembagian dataset dilakukan dengan rasio 80:20, di mana 80% data digunakan sebagai data latih dan 20% sisanya digunakan sebagai data uji.

Data latih digunakan untuk proses pelatihan model IndoBERT agar model dapat mempelajari pola dan karakteristik sentimen dari data teks yang telah diberikan label. Sementara itu, data uji digunakan untuk mengevaluasi kemampuan model dalam melakukan klasifikasi sentimen pada data yang belum pernah dipelajari sebelumnya. Pembagian data dilakukan secara acak (*random splitting*) untuk mengurangi kemungkinan bias pada dataset dan membantu menghasilkan evaluasi model yang lebih objektif.

2.5. IndoBERT

IndoBERT merupakan varian model *pre-trained* BERT yang telah dilatih menggunakan data berbahasa Indonesia dan digunakan untuk melakukan analisis sentimen pada dataset teks bahasa Indonesia [20]. Pada tahap ini, model IndoBERT dilatih menggunakan data pelatihan yang telah melalui proses *preprocessing*. Proses pelatihan dilakukan dengan beberapa pengaturan hyperparameter, seperti `num_train_epochs = 5` untuk menentukan jumlah epoch pelatihan, `learning_rate = 2e-5` untuk mengatur laju pembelajaran, serta `per_device_train_batch_size = 8` dan `per_device_eval_batch_size = 8` untuk menentukan jumlah data yang diproses pada setiap iterasi training dan evaluasi. Selain itu, digunakan `weight_decay = 0.01` untuk membantu mengurangi overfitting dan `max_length = 128` pada tahap tokenisasi untuk membatasi jumlah token pada setiap komentar.

2.6. Confusion Matrix Evaluation

Confusion matrix digunakan untuk mengevaluasi hasil klasifikasi model dengan melihat jumlah prediksi yang benar maupun salah. Pada *confusion matrix* terdapat empat kondisi utama, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Dari kondisi tersebut, dihitung metrik evaluasi utama sebagai berikut [21]:

a. Akurasi

Akurasi digunakan untuk mengukur tingkat kesesuaian antara hasil prediksi model dengan data sebenarnya. Nilai akurasi dihitung berdasarkan jumlah prediksi yang benar dibandingkan dengan keseluruhan data menggunakan persamaan (1).

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

b. Presisi

Precision digunakan untuk mengukur tingkat ketepatan hasil prediksi model terhadap data yang diprediksi positif. Nilai precision dihitung menggunakan persamaan (2).

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

c. Recall

Recall digunakan untuk mengukur kemampuan model dalam menemukan data yang sesuai berdasarkan label yang digunakan. Nilai recall dihitung menggunakan persamaan (3).

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

d. F1-Score

F1-Score digunakan untuk mengukur keseimbangan antara nilai precision dan recall melalui rata-rata harmonis dari kedua metrik tersebut. Nilai F1-Score dapat dihitung menggunakan persamaan (4).

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

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan komentar berbahasa Indonesia yang diperoleh dari dua *platform* media sosial, yaitu X (Twitter) dan YouTube. Pengumpulan data dilakukan melalui proses *scraping* untuk memperoleh komentar pengguna yang relevan dengan topik penelitian. Data yang diperoleh kemudian disimpan dalam format CSV (*Comma Separated Values*) untuk memudahkan proses pengolahan lebih lanjut.

Tahapan pengumpulan data dilakukan dalam beberapa langkah, yaitu pengambilan data dari *platform* X, pengambilan data dari *platform* YouTube, penggabungan kedua dataset, serta proses pelabelan sentimen secara manual.

a. Dataset Hasil Scraping dari Platform X

Tahap pertama dilakukan pengambilan komentar dari *platform* X (Twitter) menggunakan teknik *web scraping*. Dataset yang diperoleh berisi komentar asli pengguna beserta informasi pendukung lainnya yang relevan.

Contoh hasil *scraping* dari *platform* X ditunjukkan pada Tabel 1.

Tabel 1. Contoh Hasil Scraping dari Platform X

full_text
Malam itu terasa berbeda, sunyi tapi seperti ada yang mengawasi. Seorang wanita tiba-tiba terbangun karena mendengar suara aneh dari kamar mandinya. Awalnya ia mengira itu hanya hal biasa, tapi suara itu semakin jelas, semakin dekat. Halu SELAMAT DATANG DI RUMAH CARMEN.
<i>I'm proud of them both, but to be honest suara Carmen kurang di kupingku, ringingnya agak nyerang.</i>
Sebagai orang yang punya <i>light voice</i> , tiap kali ada yang bilang suara Carmen cempreng rasanya pengen aku toyor satu-satu. Mereka tidak bisa membedakan mana suara cempreng dan mana tipe suara yang ringan. Orang yang suaranya cempreng akan sulit memakai teknik vokal apa pun, apalagi teknik <i>belting</i> .
<i>Of course my favorite person is CARMEN, alasan: (1) dari Indonesia, (2) suara bagus banget, (3) lucu banget, (4) cantik, (5) bangga sama dia sudah bisa sejauh ini.</i>

b. Dataset Hasil Scraping dari Platform YouTube

Tahap berikutnya dilakukan pengambilan data komentar dari *platform* YouTube menggunakan teknik *scraping* komentar video. Dataset ini berisi komentar pengguna yang berkaitan dengan objek penelitian

Contoh hasil *scraping* dari *platform* YouTube ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil Scraping dari Platform YouTube

author	comment	type	voteCount	replyCount	publishedTimeText
@shela_mouttt	Jyujur aku nonton ini	comment	0	0	23 minutes ago

author	comment	type	voteCount	replyCount	publishedTimeText
	gara-gara pengen liat Carmen ngomong bahasa Indonesia				
@SuriAzzahra-g8q	Ada beberapa part Carmen ngomong B. Indo 🤔👍	comment	0	0	1 hour ago
@jeongleean.s2u	Carmen kenapa kamu gak konser di Surabaya?	comment	0	0	3 hours ago
@jeongleean.s2u	I love Carmen	comment	0	0	3 hours ago

c. Dataset Gabungan Platform X dan YouTube

Setelah data dari kedua *platform* berhasil dikumpulkan, seluruh komentar digabungkan menjadi satu dataset terpadu. Proses penggabungan ini bertujuan untuk memperkaya variasi data serta meningkatkan kemampuan model dalam mengenali berbagai bentuk ekspresi sentimen.

Contoh dataset hasil penggabungan ditunjukkan pada Tabel 3.

Tabel 3. Contoh Hasil Gabungan dari Platform X dan YouTube

comment
Jyujur aku nonton ini gara-gara pengen liat Carmen ngomong bahasa Indonesia
Ada beberapa part Carmen ngomong B. Indo
Carmen kenapa kamu gak konser di Surabaya?
Guys buat 10 juta view yokkkk biar mereka segera dibawa ke Bali semua
Kakak sekali kali bahasa indonesia

3.2. Preprocessing

Dataset yang telah dikumpulkan dari platform X (Twitter) dan YouTube selanjutnya melalui tahap *text preprocessing* untuk meningkatkan kualitas data sebelum digunakan dalam proses klasifikasi sentimen. Tahap *preprocessing* dilakukan untuk membersihkan teks dari karakter yang tidak diperlukan serta mengubah komentar menjadi format yang lebih terstruktur agar dapat diproses oleh model IndoBERT.

Beberapa tahapan *preprocessing* yang dilakukan meliputi *data cleaning*, *case folding*, *tokenization*, serta penghilangan kata-kata yang kurang bermakna (*stopword removal*) dan proses *stemming* untuk mengubah kata ke bentuk dasarnya.

Pada tahap *data cleaning*, dilakukan penghapusan karakter khusus, tautan (*URL*), emoji yang rusak (*encoding error*), serta simbol-simbol yang tidak relevan. Selanjutnya, tahap *case folding*

dilakukan dengan mengubah seluruh huruf menjadi huruf kecil (*lowercase*) agar konsisten dalam proses analisis.

Setelah itu, komentar dipecah menjadi unit-unit kata melalui proses *tokenization*. Pada beberapa kasus, kata-kata umum yang tidak memberikan makna signifikan dihapus menggunakan *stopword removal*, kemudian kata yang memiliki imbuhan dikembalikan ke bentuk dasarnya melalui proses *stemming*.

Contoh hasil *preprocessing* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Preprocessing

No	Proses Preprocessing	Hasil
1	<i>Comment</i> (Data Awal)	Jyujur aku nonton ini gara gara pengen liat Carmen ngomong bahasa Indonesia
2	<i>Data Cleaning</i>	Jyujur aku nonton ini gara gara pengen liat Carmen ngomong bahasa Indonesia
3	<i>Case Folding</i>	jyujur aku nonton ini gara gara pengen liat carmen ngomong bahasa indonesia
4	<i>Tokenization</i>	['jyujur', 'aku', 'nonton', 'ini', 'gara', 'gara', 'pengen', 'liat', 'carmen', 'ngomong', 'bahasa', 'indonesia']
5	<i>Stopword Removal</i>	['jyujur', 'nonton', 'gara', 'gara', 'pengen', 'liat', 'carmen', 'ngomong', 'bahasa', 'indonesia']
6	<i>Stemming</i>	['jyujur', 'nonton', 'gara', 'gara', 'ken', 'liat', 'carmen', 'ngomong', 'bahasa', 'indonesia']

3.3. Labelisasi

Setelah melalui tahap *text preprocessing*, dataset selanjutnya dilakukan proses pelabelan data secara manual (*manual labeling*). Pada tahap ini, setiap komentar dianalisis berdasarkan makna dan konteks kalimat, kemudian dikategorikan ke dalam tiga kelas sentimen, yaitu *Negative*, *Neutral*, dan *Positive*.

Hasil pelabelan kemudian disimpan dalam *file* hasil *labeling_manual.csv* dengan atribut utama berupa *comment*, *tokens*, dan *label_manual_rule*. Contoh hasil pelabelan ditunjukkan pada Tabel 5.

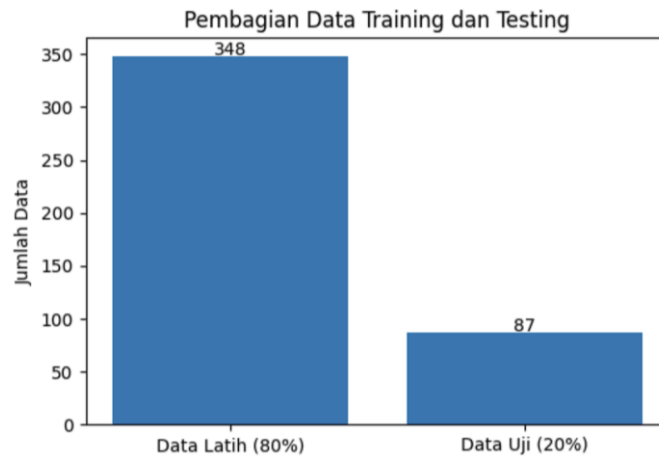
Tabel 5. Hasil Pelabelan Manual

comment	tokens	label_manual_rule
Jyujur aku nonton ini gara-gara pengen liat Carmen ngomong bahasa Indonesia	['jyujur', 'nonton', 'gara', 'gara', 'ken', 'liat', 'carmen', 'ngomong', 'bahasa', 'indonesia']	<i>Neutral</i>
Ada beberapa part Carmen ngomong B. Indo	['part', 'carmen', 'ngomong', 'bindo']	<i>Neutral</i>
Carmen kenapa kamu gak konser di Surabaya?	['carmen', 'gakkonser', 'surabaya']	<i>Neutral</i>
Guys buat 10 juta view yokkkk biar mereka segera dibawa ke Bali semua	['guys', 'juta', 'view', 'yokkkk', 'biar', 'bawa', 'bal']	<i>Neutral</i>

3.4. *Splitting Data*

Setelah proses pelabelan selesai, dataset dibagi menjadi data latih (*training set*) dan data uji (*testing set*) menggunakan *fungsi train_test_split()* dari *library Scikit-learn*.

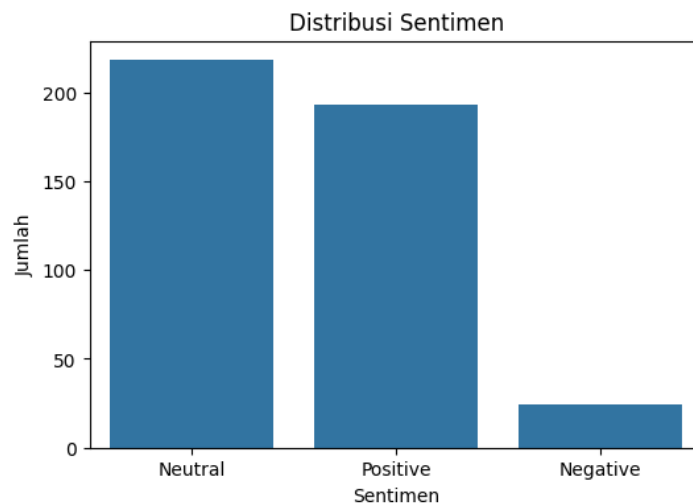
Pembagian data dilakukan dengan proporsi, 80% data latih dan 20% data uji. Seperti yang dapat dilihat pada Gambar 3.



Gambar 3. Pembagian Data Training dan Testing

3.5. *IndoBERT*

Model klasifikasi sentimen dibangun menggunakan *AutoModelForSequenceClassification* dari *library Hugging Face Transformers* dengan jumlah kelas sebanyak 3 label sentimen (*Positive, Neutral, Negative*) yang hasilnya dapat dilihat pada Gambar 4.



Gambar 4. Distribusi Sentimen Hasil Prediksi

3.6. *Evaluasi Model*

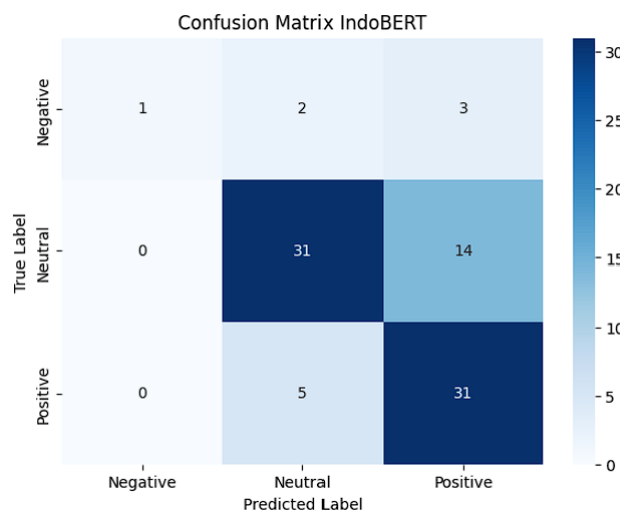
Evaluasi model dalam penelitian ini akan menggunakan confusion matriks, yang hasilnya dapat dilihat pada Tabel 6.

Tabel 6. Hasil Metrik Evaluasi

Metrik	Nilai
Accuracy	72,41%
Precision	75,82%
Recall	72,41%
F1-Score	71,15%

Berdasarkan hasil evaluasi, model IndoBERT mampu melakukan klasifikasi sentimen dengan performa yang baik dan seimbang. Nilai accuracy sebesar 72,41% menunjukkan bahwa model dapat mengidentifikasi polaritas sentimen secara konsisten pada sebagian besar data uji. Performa ini didukung oleh nilai precision sebesar 75,82% dan recall sebesar 72,41%, yang menandakan bahwa model memiliki ketepatan yang tinggi dalam memprediksi kelas sentimen sekaligus mampu menjangkau mayoritas data target dengan baik. Keseimbangan performa ini dipertegas oleh nilai F1-score yang stabil di angka 71,15%.

Selain evaluasi numerik tersebut, dilakukan analisis menggunakan confusion matrix untuk melihat distribusi kesalahan klasifikasi antar kelas. Gambar 5 menunjukkan visualisasi confusion matriks hasil prediksi model.



Gambar 5. Confusion Matrix Model IndoBERT

Visualisasi confusion matrix menunjukkan bahwa sebagian besar kesalahan klasifikasi terjadi antara kategori Neutral dan Positive, yang mengindikasikan bahwa beberapa komentar memiliki konteks yang ambigu atau mengandung ekspresi yang sulit dibedakan secara tegas oleh model.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, penerapan model IndoBERT untuk analisis sentimen pada komentar berbahasa Indonesia menunjukkan performa yang cukup baik dalam mengklasifikasikan sentimen ke dalam tiga kategori, yaitu negative, neutral, dan positive. Berdasarkan hasil evaluasi, model memperoleh tingkat akurasi sekitar 70%, yang menunjukkan bahwa model telah mampu mengenali pola sentimen pada sebagian besar data, meskipun masih terdapat beberapa kesalahan klasifikasi.

Kemampuan model dalam memahami konteks bahasa Indonesia melalui pendekatan transformer-based deep learning memberikan hasil yang cukup efektif, terutama dalam

menangkap makna kalimat yang kompleks. Namun demikian, performa model belum sepenuhnya optimal, yang kemungkinan dipengaruhi oleh beberapa faktor seperti keterbatasan jumlah data latih, ketidakseimbangan distribusi kelas, serta adanya penggunaan bahasa informal atau ambigu dalam komentar yang menyulitkan proses klasifikasi.

Analisis tambahan melalui confusion matrix, distribusi sentimen, dan word cloud menunjukkan bahwa model mampu memberikan gambaran umum mengenai kecenderungan opini pengguna serta mengidentifikasi kata-kata dominan pada masing-masing kategori sentimen. Hal ini membuktikan bahwa meskipun tingkat akurasi belum tinggi, model tetap dapat dimanfaatkan sebagai alat bantu untuk analisis opini secara otomatis.

Untuk penelitian selanjutnya, performa model dapat ditingkatkan melalui penambahan jumlah dataset, proses preprocessing yang lebih mendalam seperti normalisasi bahasa informal, serta perbandingan dengan model lain atau optimasi hyperparameter agar diperoleh hasil klasifikasi yang lebih akurat dan stabil.

Daftar Pustaka

- [1] V. E. Sidauruk dan W. Herowati, "Indobert-Based Sentiment Analysis of Political Discourse on Platform X: The Case Of Prabowo-Gibran Administration," *J. Appl. Informatics Comput.*, vol. 10, no. 1, hal. 673–683, 2026, doi: 10.30871/jaic.v10i1.11586.
- [2] B. Setiawan, "A Review of Sentiment Analysis Applications in Indonesia Between 2023-2024," *JIEET (Journal Inf. Eng. Educ. Technol.)*, vol. 8, no. 2, hal. 71–83, 2024, doi: 10.26740/jieet.v8n2.p71-83.
- [3] N. Ambarsari, K. Abitsha, R. Nazra, dan O. Rukmana, "Netnografi : Analisis Sentimen Publik Terhadap Kenaikan Tarif Ppn 12%," *J. Ekon. Manajemen, Akunt. dan Keuang.*, vol. 7, no. 2, hal. 1–12, 2026, doi: 10.53697/emak.v7i2.3739.
- [4] M. R. Manoppo, I. C. Kolang, D. N. Fiat, R. Michelly, dan C. Mawara, "Analisis Sentimen Publik di Media Sosial Terhadap Kenaikan PPN 12% di Indonesia Menggunakan IndoBERT," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 4, no. 2, hal. 152–163, 2025, doi: 10.69916/jkbt.v4i2.322.
- [5] A. S. Rizkia, Wufron, dan 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, hal. 1037–1048, 2025, doi: 10.57152/malcom.v5i3.2151.
- [6] M. A. Mayzaroh, D. F. Ningsih, N. Destriani, dan M. C. T. Manullang, "Benchmarking PyCaret AutoML Against IndoBERT Fine-Tuning For Sentiment Analysis On Indonesian IKN Twitter Data," *arXiv Prepr. arXiv2604.25392*, no. 1, hal. 1–7, 2026.
- [7] F. Koto, A. Rahimi, J. H. Lau, dan T. Baldwin, "IndoLEM and IndoBERT : A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proceedings of the 28th International Conference on Computational Linguistics, 2020*, hal. 757–770. doi: 10.18653/v1/2020.coling-main.66.
- [8] M. A. A. Saputra, A. Alamsyah, D. P. Ramadhani, T. S. Siadari, dan H. Fakhurroja, "IndoBERT-Sentiment: Context-Conditioned Sentiment Classification for Indonesian Text," *arXiv Prepr. arXiv2604.07057*, hal. 1–8, 2026.
- [9] D. I. Putri, A. N. Alfian, M. Y. Putra, dan P. D. Mulyo, "IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election," *J. Appl. Informatics Comput.*, vol. 8, no. 1, hal. 7–12, 2024, doi: 10.30871/jaic.v8i1.7440.
- [10] D. Alfatah, "Penerapan Model Transformer Untuk Deteksi Sentimen Pada Data Twitter Berbahasa Indonesia," *J. Komput.*, vol. 2, no. 2, hal. 67–70, 2024, doi: 10.70963/jk.v2i2.99.
- [11] M. Pota, M. Ventura, R. Catelli, dan M. Esposito, "An Effective BERT-Based Pipeline for Twitter Sentiment Analysis : A Case Study in Italian," *Sensors*, vol. 21, no. 1, hal. 1–21, 2020, doi: 10.3390/s21010133.
- [12] D. S. A. Mumtaz dan N. W. A. Majid, "Web Scraping Dan Fine-Tuning IndoBERT Untuk Analisis Sentimen Berbasis Aspek (ABSA) Pada Data Twitter/X: Studi Kasus Topik Kurikulum Merdeka," *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 11, no. 1, hal. 949–963, 2026, doi: 10.36341/rabit.v11i1.7155.

- [13] S. Khairunnisa, Adiwijaya, dan S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. April, hal. 406–414, 2021, doi: 10.30865/mib.v5i2.2835.
- [14] K. Khotimah, Martanto, A. R. Dikananda, dan A. Rifa, "Analisis Sentimen Ulasan Aplikasi Pintu Di Google Play Store Menggunakan Algoritma Naïve Bayes," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 13, no. 1, 2025, doi: 10.23960/jitet.v13i1.5789 ANALISIS.
- [15] Styawati, N. Hendrastuty, A. R. Isnain, dan A. Y. Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, hal. 150–155, 2021, doi: 10.30591/jpit.v6i3.2870.
- [16] Y. Saputri, B. N. Sari, dan S. Defiyanti, "Analisis Sentimen Terhadap Program MSIB Dalam X 2024 Menggunakan Lexicon-Based Dan Naïve Bayes Classifier," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 5, hal. 9043–9048, 2025, doi: 10.36040/jati.v9i5.15258.
- [17] O. Irnawati dan K. Solecha, "Analisis Sentimen Ulasan Aplikasi Flip Menggunakan Naïve Bayes dengan Seleksi Fitur PSO," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 2, hal. 189–199, 2022, doi: 10.46772/intech.v4i02.868.
- [18] S. P. Tanzil dan M. R. Pribadi, "Analisis Sentimen Pengguna Instagram terhadap Timnas Indonesia U-23 pada Piala AFC menggunakan Algoritma K-Nearest Neighbor (K-NN) dengan Synthetic Minority Over-sampling Technique (SMOTE)," *TELEKONTRAN J. Ilm. Telekomun. Kendali dan Elektron. Terap.*, vol. 12, no. 1, hal. 68–80, 2024, doi: 10.34010/telekontran.v12i1.12869.
- [19] B. Franko, N. Wilyanto, dan H. Irsyad, "Analisis Sentimen Terhadap Naturalisasi Pemain pada Youtube Menggunakan Decision Tree dan Naive Bayes," *J. Sess. (Software Dev. Digit. Bus. Intell. Comput. Eng.)*, vol. 3, no. 1, hal. 8–16, 2024, doi: 10.57203/session.v3i1.2024.8-16.
- [20] A. A. P. Simarmata dan T. B. Sasongko, "Sentiment Analysis on BRImo Application Reviews Using IndoBERT," *J. Appl. Informatics Comput.*, vol. 9, no. 3, hal. 851–862, 2025, doi: 10.30871/jaic.v9i3.8162.
- [21] W. Tanuwijaya, C. E. Setiawan, H. Irsyad, dan A. Rahman, "Implementasi TF-IDF dan Cosine Similarity untuk Penyaringan Dokumen Berita Program Makan Siang Gratis Pemerintah Indonesia," *Device J. Inf. Syst. Comput. Sci. Inf. Technol.*, vol. 6, no. 2, hal. 322–334, 2025, doi: 10.46576/device.v6i2.6724.