JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9697
Hal 666-678
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Implementasi BERT dan IndoRoBERTa untuk Klasifikasi Sentimen Opini Publik tentang Kecerdasan Buatan dalam Pendidikan di YouTube

Muhammad Mutawakkil Alallah[1], Mokhammad Amin Hariyadi[1*], Triyo Supriyatno[1] , Eri Riana[2]

[1]Fakultas Saintek, Magister Informatika, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia
[2] Fakultas Teknologi Informasi (FTI), Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia
Email: [1]240605220005@student.uin-malang.ac.id, [2*]adyt2002@uin-malang.ac.id, [3]triyo@pai.uin-malang.ac.id, [4]eri.eea@bsi.ac.id
Email Penulis Korespondensi: adyt2002@uin-malang.ac.id
Submitted **13-04-2026;** Accepted **27-04-2026;** Published **30-04-2026**

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen komentar YouTube berbahasa Indonesia terkait kecerdasan buatan (Artificial Intelligence) dalam bidang pendidikan menggunakan pendekatan deep learning berbasis Transformer, yaitu Bidirectional Encoder Representations from Transformers (BERT) dan IndoRoBERTa. Data penelitian diperoleh melalui YouTube Data API dengan total 10.834 komentar yang mencerminkan opini publik terhadap penerapan AI dalam pendidikan. Dataset dilakukan proses pelabelan manual ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif, kemudian melalui tahapan pre-processing data yang meliputi case folding, text cleaning, normalisasi teks, tokenisasi, serta filtering topik. Hasil eksperimen menunjukkan bahwa pada skenario baseline tanpa fine-tuning, kedua model menghasilkan performa rendah dengan akurasi di bawah 41%. Namun setelah dilakukan fine-tuning, terjadi peningkatan signifikan, di mana IndoRoBERTa mencapai akurasi 91,54% dengan F1-score 0,9134, sedangkan BERT mencapai akurasi 84,63% dengan F1-score 0,8413. Hasil ini menunjukkan bahwa model berbasis Transformer yang telah disesuaikan dengan data spesifik mampu meningkatkan kemampuan dalam memahami konteks linguistik bahasa Indonesia yang bersifat informal dan tidak terstruktur. Selain itu, IndoRoBERTa menunjukkan kinerja yang lebih stabil dalam menangani ketidakseimbangan kelas dibandingkan BERT. Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan berbasis Transformer efektif dalam analisis sentimen media sosial dan dapat digunakan untuk memahami persepsi publik terhadap implementasi kecerdasan buatan dalam pendidikan secara lebih akurat dan komprehensif.

Kata Kunci: Klasifikasi Sentimen; Kecerdasan Buatan; YouTube; BERT; IndoRoBERTa; Transformer

Abstract

This study aims to analyze the sentiment of Indonesian-language YouTube comments related to artificial intelligence (AI) in the field of education using a Transformer-based deep learning approach, namely Bidirectional Encoder Representations from Transformers (BERT) and IndoRoBERTa. The research data were obtained through the YouTube Data API, consisting of 10,834 comments reflecting public opinion on the implementation of AI in education. The dataset was manually labeled into three sentiment categories: positive, neutral, and negative, followed by a preprocessing stage including case folding, text cleaning, text normalization, tokenization, and topic filtering. The experimental results show that in the baseline scenario without fine-tuning, both models achieved low performance with accuracy below 41%. However, after fine-tuning, a significant improvement was observed, where IndoRoBERTa achieved an accuracy of 91.54% with an F1-score of 0.9134, while BERT reached an accuracy of 84.63% with an F1-score of 0.8413. These results indicate that Transformer-based models adapted to specific datasets are capable of better capturing the contextual and linguistic characteristics of informal and unstructured Indonesian text. In addition, IndoRoBERTa demonstrates more stable performance in handling class imbalance compared to BERT. Overall, this study demonstrates that Transformer-based approaches are effective for sentiment analysis in social media and can be used to more accurately and comprehensively understand public perceptions of the implementation of artificial intelligence in education.

Keywords: Sentiment Classification; Artificial Intelligence; YouTube; BERT; IndoRoBERTa; Transformer

1. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*) pada periode 2021 hingga 2025 menunjukkan akselerasi yang sangat signifikan dan berdampak luas pada berbagai sektor strategis, termasuk industri, kesehatan, pendidikan, dan hiburan [1], [2]. Dalam konteks pendidikan, AI tidak hanya berfungsi sebagai alat bantu pembelajaran, tetapi juga berperan dalam mentransformasi sistem pendidikan melalui otomatisasi proses administratif, analisis data pembelajaran dalam skala besar, serta penerapan sistem prediktif untuk mendukung pengambilan keputusan berbasis data [3]. Meskipun demikian, implementasi AI dalam pendidikan tidak terlepas dari berbagai permasalahan yang kompleks, terutama terkait aspek sosial, etika, dan psikologis[4], [5]. Beberapa isu yang sering muncul meliputi kekhawatiran terhadap penggantian peran tenaga pendidik, potensi penyalahgunaan data pribadi, bias algoritma yang dapat memengaruhi keadilan sistem, serta meningkatnya ketergantungan terhadap teknologi otomatis [6]. Permasalahan tersebut menunjukkan bahwa penerapan AI tidak hanya memerlukan kesiapan teknologi, tetapi juga pemahaman yang komprehensif terhadap persepsi publik sebagai pengguna utama sistem tersebut.

Dalam era digital saat ini, persepsi publik terhadap suatu teknologi banyak diekspresikan melalui platform media sosial [7], [8]. Salah satu platform yang memiliki tingkat interaksi tinggi adalah YouTube, yang memungkinkan pengguna untuk memberikan tanggapan secara langsung melalui kolom komentar [9]. Komentar-komentar tersebut mencerminkan opini publik yang bersifat spontan, tidak terstruktur, dan sering kali mengandung emosi yang kuat terhadap suatu isu, termasuk penerapan AI dalam pendidikan [10]. Namun, karakteristik komentar YouTube yang cenderung informal, mengandung bahasa slang, singkatan, serta variasi linguistik yang tinggi menjadi tantangan tersendiri dalam proses analisis [11], [12], [13]. Kondisi ini menimbulkan permasalahan utama dalam penelitian, yaitu

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 666
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9697
Hal 666-678
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

bagaimana mengidentifikasi dan mengklasifikasikan sentimen publik secara akurat dari data teks yang tidak terstruktur dan kompleks [14]. Oleh karena itu, diperlukan pendekatan berbasis *Natural Language Processing* (NLP), khususnya analisis sentimen, untuk mengolah dan mengekstraksi informasi dari data tersebut secara sistematis dan terukur [15].

Sebagai solusi terhadap permasalahan tersebut, model berbasis Transformer seperti BERT (*Bidirectional Encoder Representations from Transformers*) dan IndoRoBERTa menjadi pendekatan yang relevan untuk digunakan [16], [17]. Model Transformer dikenal memiliki kemampuan dalam memahami konteks semantik secara dua arah, sehingga lebih efektif dalam menangkap makna kata dalam suatu kalimat dibandingkan metode tradisional [6]. Model pra-latih umum telah banyak digunakan dalam berbagai tugas NLP, namun memiliki keterbatasan ketika diterapkan pada bahasa selain bahasa Inggris [18]. Untuk mengatasi hal tersebut, model spesifik bahasa dikembangkan menggunakan korpus bahasa Indonesia, sehingga mampu menangkap karakteristik linguistik lokal, termasuk penggunaan bahasa informal yang umum ditemukan pada media sosial. Dengan demikian, kombinasi dan perbandingan antara kedua model ini diharapkan dapat memberikan solusi optimal dalam klasifikasi sentimen komentar YouTube berbahasa Indonesia.

Sejumlah penelitian terdahulu telah mengkaji analisis sentimen menggunakan berbagai pendekatan dan sumber data. Metode *Naïve Bayes* dan *Support Vector Machine* (SVM) dengan representasi TF-IDF telah digunakan untuk mengklasifikasikan komentar YouTube, namun menunjukkan keterbatasan dalam menangkap konteks semantik [3]. Pendekatan berbasis *Convolutional Neural Network* (CNN) dan *Bi-directional Long Short-Term Memory* (Bi-LSTM) menunjukkan peningkatan kinerja dibanding metode konvensional dalam analisis emosi publik [19]. Selain itu, pemanfaatan data Reddit digunakan untuk mengkaji persepsi masyarakat terhadap teknologi *deepfake* [20], sementara analisis opini publik terhadap institusi pendidikan dilakukan menggunakan data Twitter [21]. Penggunaan model berbasis Transformer juga telah menunjukkan peningkatan performa klasifikasi sentimen secara signifikan dibandingkan pendekatan sebelumnya [6].

Meskipun penelitian-penelitian tersebut menunjukkan perkembangan signifikan dalam bidang analisis sentimen, terdapat beberapa keterbatasan yang dapat diidentifikasi. Pertama, sebagian besar penelitian masih berfokus pada platform media sosial seperti Twitter dan Reddit, sementara pemanfaatan data dari YouTube, khususnya komentar berbahasa Indonesia, masih relatif terbatas. Kedua, topik yang dikaji belum secara spesifik membahas persepsi publik terhadap penerapan AI dalam konteks pendidikan. Ketiga, studi komparatif antara model Transformer umum dan model spesifik bahasa dalam konteks bahasa Indonesia masih jarang dilakukan secara sistematis. Keterbatasan-keterbatasan ini menunjukkan adanya kesenjangan penelitian (*research gap*) yang perlu diatasi untuk memperoleh pemahaman yang lebih komprehensif mengenai efektivitas model NLP dalam menganalisis sentimen berbasis konteks lokal dan domain spesifik.

Berdasarkan identifikasi permasalahan dan kesenjangan penelitian tersebut, penelitian ini bertujuan untuk: (1) mengevaluasi kinerja model berbasis Transformer dalam mengklasifikasikan sentimen komentar YouTube berbahasa Indonesia terkait penerapan AI dalam pendidikan; serta (2) membandingkan efektivitas model pra-latih umum dan model spesifik bahasa menggunakan metrik evaluasi standar seperti akurasi, presisi, *recall*, dan *F1-score*. Penelitian ini diharapkan dapat memberikan kontribusi akademik dalam pengembangan metode analisis sentimen berbasis *deep learning*, khususnya dalam konteks bahasa Indonesia. Selain itu, hasil penelitian ini juga diharapkan dapat memberikan manfaat praktis berupa informasi berbasis data mengenai persepsi publik terhadap implementasi AI dalam pendidikan, sehingga dapat menjadi dasar pertimbangan bagi pemangku kepentingan dalam merumuskan kebijakan dan strategi pengembangan teknologi yang lebih adaptif, inklusif, dan berkelanjutan.

Kontribusi utama penelitian ini adalah melakukan komparasi empiris antara model BERT dan IndoRoBERTa pada data komentar YouTube berbahasa Indonesia dalam konteks pendidikan, serta mengevaluasi dampak fine-tuning terhadap peningkatan performa model dalam kondisi data tidak seimbang.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

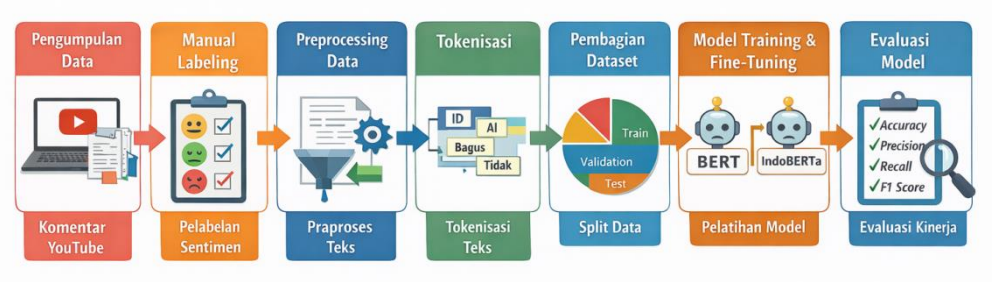
Penelitian ini menggunakan pendekatan eksperimen kuantitatif untuk mengevaluasi kinerja model *deep learning* berbasis Transformer dalam klasifikasi sentimen opini publik terhadap kecerdasan buatan dalam pendidikan. Model yang digunakan adalah BERT dan IndoRoBERTa, yang dibandingkan untuk mengetahui performa terbaik dalam menganalisis komentar berbahasa Indonesia.

Tahapan penelitian dilakukan secara sistematis mulai dari pengumpulan data hingga evaluasi model. Secara umum, alur penelitian terdiri dari: (1) pengumpulan data komentar YouTube, (2) pelabelan sentimen secara manual, (3) *pre-processing* data, (4) tokenisasi, (5) pembagian dataset, (6) pelatihan dan *fine-tuning* model, serta (7) evaluasi kinerja model. Setiap tahapan dirancang untuk memastikan bahwa model mampu memahami pola linguistik serta polaritas sentimen dari data yang dianalisis.

Alur tahapan penelitian secara keseluruhan ditunjukkan pada Gambar 1.

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 667
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9697
Hal 666-678
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 1. Tahapan Penelitian

Gambar 1 menggambarkan proses penelitian secara sistematis mulai dari pengumpulan data hingga evaluasi model. Setiap tahapan saling terintegrasi untuk memastikan bahwa data yang diolah memiliki kualitas yang baik dan model yang dihasilkan memiliki performa optimal.

2.2 Pengumpulan Data

Dataset yang digunakan berupa komentar dari video yang membahas kecerdasan buatan dalam pendidikan pada platform YouTube. Pemilihan YouTube sebagai sumber data didasarkan pada tingginya volume interaksi pengguna yang merepresentasikan opini publik secara terbuka.

Pengumpulan data dilakukan secara otomatis menggunakan *YouTube Data API* dengan memanfaatkan serangkaian kata kunci yang relevan, antara lain "AI pendidikan", "kecerdasan buatan pendidikan", "AI dalam pendidikan", "AI pembelajaran", "AI di sekolah", "AI di kampus", "ChatGPT pendidikan", "ChatGPT di sekolah", "AI tutor", "chatbot pendidikan", "AI generatif pendidikan", "manfaat AI pendidikan", "dampak AI pada pendidikan", "tantangan AI pendidikan", "etika AI dalam pendidikan", "kecurangan akademik AI", "personalisasi pembelajaran AI", serta "transformasi digital pendidikan". Berdasarkan kata kunci tersebut, video yang sesuai dengan topik penelitian diidentifikasi, kemudian komentar pengguna diekstraksi dan dikumpulkan sebagai dataset yang digunakan dalam penelitian ini.

Hasil *crawling* awal memperoleh 11.667 komentar dari 168 video dan 146 kanal YouTube. Dataset mencakup atribut teks komentar, jumlah *likes*, waktu publikasi, ID video, kata kunci, judul video, dan nama kanal. Setelah proses validasi berupa penghapusan data duplikat, data kosong, serta penyaringan rentang waktu, diperoleh 10.834 komentar valid dari 167 video dan 145 kanal.

Komentar yang digunakan berada pada rentang waktu 31 Maret 2022 hingga 31 Desember 2025, yang mencerminkan kondisi terkini diskusi publik terkait AI dalam pendidikan.

Untuk memberikan gambaran umum mengenai data yang digunakan dalam penelitian ini, ringkasan dataset disajikan pada Tabel 1.

Tabel 1. Ringkasan Dataset

Deskripsi	Nilai
Total komentar terkumpul	11.667
Komentar valid	10.834
Total video	167
Total kanal	145
Rentang waktu	2022–2025

Berdasarkan Tabel 1, terlihat bahwa jumlah komentar valid yang digunakan dalam penelitian ini mencapai 10.834 data yang berasal dari 167 video dan 145 kanal YouTube. Rentang waktu data yang digunakan, yaitu tahun 2022 hingga 2025, menunjukkan bahwa dataset mencerminkan kondisi terkini terkait diskusi publik mengenai kecerdasan buatan dalam pendidikan. Jumlah data yang relatif besar ini memberikan dasar yang kuat untuk proses pelatihan model deep learning agar mampu menangkap pola sentimen secara lebih representatif.

Untuk memberikan gambaran nyata mengenai data yang digunakan dalam penelitian ini, sebagian contoh dataset hasil crawling ditampilkan pada Tabel 2. Data yang ditampilkan merupakan komentar asli pengguna YouTube yang belum melalui tahap *pre-processing*.

Tabel 2. Contoh Dataset Komentar YouTube (Data Mentah)

Comment	Published At	Video ID	Title	Channel
Tidak ada yang bisa kalah kita sebagai manusia,, itu hanya orang2 tidak ingat ilmu agama, karna manusia Allah SWT yng menciptakan nya, manusia mahluk yng sempurna	2025-12-24	Glvgjr-gPUQ	[FULL] KICK ANDY - PROF STELLA: OTAK VS AI	METRO TV
Otak lebih pintar dari teknologi	2025-12-24	Glvgjr-gPUQ	[FULL] KICK ANDY - PROF STELLA:	METRO TV

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 668
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Ya tetap otak lah, tidak mungkin ada AI, tanpa ada otak yg ber ide	2025-11-21	Glvgjr-gPUQ	OTAK VS AI [FULL] KICK ANDY - PROF STELLA: OTAK VS AI	METRO TV
saya kira cupu ternyata suhu pintar menguraikan, apa sebenarnya AI	2025-09-10	Glvgjr-gPUQ	[FULL] KICK ANDY - PROF STELLA: OTAK VS AI	METRO TV
Posisi penting, gaji kecil. Posisi ga penting, perusahaan milik Negara merugi, gajinya gede bangetttt	2025-08-04	Glvgjr-gPUQ	[FULL] KICK ANDY - PROF STELLA: OTAK VS AI	METRO TV

Berdasarkan Tabel 2, dapat diamati bahwa dataset memiliki karakteristik bahasa yang tidak terstruktur, termasuk penggunaan huruf kapital yang tidak konsisten, singkatan (seperti yg, yng), pengulangan huruf, serta tanda baca yang berlebihan. Selain itu, beberapa komentar mengandung opini subjektif yang berkaitan dengan persepsi terhadap kecerdasan buatan, baik secara eksplisit maupun implisit.

Variasi linguistik tersebut mencerminkan sifat alami data media sosial yang bersifat informal dan kontekstual. Hal ini menjadi tantangan utama dalam proses analisis sentimen, sehingga diperlukan tahapan *pre-processing* data yang komprehensif untuk meningkatkan kualitas data sebelum digunakan dalam pelatihan model.

Selain itu, atribut seperti published_at, video_id, title, dan channel memberikan konteks tambahan yang dapat digunakan untuk analisis lanjutan, seperti analisis temporal atau analisis sumber konten. Namun, dalam penelitian ini, fokus utama analisis difokuskan pada teks komentar sebagai representasi opini publik.

2.3 Pelabelan Sentimen

Komentar yang telah dikumpulkan kemudian diberi label sentimen secara manual ke dalam tiga kategori, yaitu negatif, netral, dan positif. Pelabelan ini bertujuan untuk menghasilkan dataset *ground truth* yang akurat sebagai dasar dalam proses pelatihan model [22], [23].

Proses pelabelan dilakukan oleh dua orang validator secara independen untuk meningkatkan objektivitas dan konsistensi penilaian [24]. Penentuan label didasarkan pada makna semantik dan konteks kalimat. Komentar yang mengandung dukungan atau apresiasi terhadap topik dikategorikan sebagai positif, komentar yang mengandung kritik atau penolakan dikategorikan sebagai negatif, sedangkan komentar yang bersifat informatif tanpa menunjukkan opini yang jelas dikategorikan sebagai netral.

Untuk memastikan konsistensi antar validator, dilakukan pengukuran tingkat kesepakatan menggunakan metode *inter-annotator agreement* [25], [26]. Apabila terdapat perbedaan label, maka dilakukan diskusi hingga mencapai kesepakatan akhir [27].

2.4 *Preprocessing* Data

Komentar yang telah dikumpulkan kemudian diberi label sentimen secara manual ke dalam tiga kategori, yaitu negatif, netral, dan positif. Pelabelan dilakukan untuk menghasilkan dataset *ground truth* yang akurat sebagai dasar pelatihan model [22], [28].

Tahap Preprocessing dilakukan untuk meningkatkan kualitas data dan mengurangi noise. Proses ini meliputi:

1. Case Folding: mengubah seluruh teks menjadi huruf kecil [23], [29].
2. Text Cleaning: menghapus URL, emoji, dan karakter non-alfanumerik [30], [31]
3. Stopword Removal: menghapus kata umum menggunakan Sastrawi [32], [33].
4. Normalisasi Teks: mengubah kata tidak baku menjadi baku [34], [35].
5. Penghapusan Duplikat: menghindari bias data [36], [37].
6. Filtering Topik: memastikan relevansi komentar dengan topik AI pendidikan [38], [39].

Tahapan ini bertujuan agar model hanya memproses informasi yang relevan secara semantik.

2.5 Tokenisasi

Data hasil praproses kemudian diubah menjadi token menggunakan tokenizer masing-masing model [40]. Tokenisasi mengonversi teks menjadi representasi numerik agar dapat diproses oleh model [41], [42]. BERT menggunakan metode *WordPiece*, sedangkan IndoRoBERTa menggunakan *Byte-Pair Encoding* (BPE). Panjang maksimum token ditetapkan sebesar 128 untuk menjaga efisiensi komputasi.

2.6 Pembagian Dataset

Dataset akhir yang digunakan dalam penelitian ini berjumlah 4.726 data setelah melalui tahapan filtering topik dan penghapusan data tidak relevan. Proses filtering dilakukan secara ketat untuk memastikan bahwa hanya komentar yang benar-benar berkaitan dengan topik kecerdasan buatan dalam pendidikan yang digunakan dalam pelatihan model. Hal ini menyebabkan berkurangnya jumlah data secara signifikan dari 10.834 menjadi 4.726, namun meningkatkan kualitas dan relevansi dataset.

Dataset kemudian dibagi menjadi data latih, validasi, dan uji dengan rasio 70:15:15 menggunakan metode stratified sampling untuk menjaga proporsi distribusi kelas pada setiap subset [43], [44].

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 669
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9697
Hal 666-678
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Pembagian dataset ke dalam data latih, validasi, dan pengujian dilakukan menggunakan metode stratified sampling. Distribusi data pada masing-masing subset ditunjukkan pada Tabel 3.

Tabel 3. Pembagian Dataset

Dataset	Negatif	Netral	Positif	Total
Training	270	2745	293	3308
Validation	57	587	65	709
Test	57	587	65	709

Berdasarkan Tabel 3, distribusi kelas pada setiap subset relatif seimbang secara proporsional, sehingga dapat meminimalkan bias model terhadap kelas tertentu. Hal ini penting untuk memastikan bahwa model memiliki kemampuan generalisasi yang baik saat diuji pada data baru.

2.7 Pelatihan dan *Fine-Tuning* Model

Penelitian ini menggunakan dua model Transformer, yaitu BERT dan IndoRoBERTa. Implementasi dilakukan menggunakan pustaka *Hugging Face* Transformers.

Eksperimen dilakukan dalam dua skenario, yaitu:

1. *Baseline*, yaitu kondisi di mana model digunakan secara langsung tanpa dilakukan pelatihan ulang (*fine-tuning*), sehingga performa yang dihasilkan merefleksikan kemampuan awal dari model pralatih dalam memahami data tanpa penyesuaian terhadap domain spesifik penelitian [41], [45].
2. *Fine-Tuning*, yaitu proses pelatihan ulang model pralatih menggunakan dataset penelitian, dengan tujuan menyesuaikan parameter model agar lebih optimal dalam memahami karakteristik data dan meningkatkan kinerja klasifikasi pada domain yang diteliti [46].

Parameter pelatihan yang digunakan dalam proses fine-tuning model disajikan pada Tabel 4

Tabel 4. Parameter Pelatihan

Parameter	Nilai
Learning Rate	2e-5
Batch Size	16
Epoch	3
Optimizer	Adam
Max Length	128

Berdasarkan Tabel 4, parameter pelatihan yang digunakan mengikuti konfigurasi umum pada model Transformer [10]., seperti learning rate sebesar 2e-5 dan jumlah epoch sebanyak 3. Pengaturan parameter ini bertujuan untuk menjaga keseimbangan antara performa model dan efisiensi komputasi selama proses pelatihan. Pelatihan dilakukan menggunakan GPU NVIDIA T4 untuk meningkatkan efisiensi komputasi.

2.8 Evaluasi Model

Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur kinerja klasifikasi.

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

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

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

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

Keterangan:
TP (True Positive), TN (True Negative), FP (False Positive), dan FN (False Negative).
Selain itu, digunakan confusion matrix untuk menganalisis distribusi kesalahan klasifikasi pada masing-masing kelas sentimen dengan membandingkan hasil prediksi dan label sebenarnya. Analisis ini memberikan gambaran rinci terkait jumlah true positive, true negative, false positive, dan false negative pada setiap kelas, sehingga dapat menunjukkan kemampuan model dalam membedakan kelas sentimen [47], [48]. Selain itu, confusion matrix juga membantu mengidentifikasi pola kesalahan, seperti kecenderungan model pada kelas tertentu, serta memberikan evaluasi yang lebih mendalam terutama pada kondisi data tidak seimbang [49], [50].

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 670
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9697
Hal 666-678
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

3. HASIL DAN PEMBAHASAN

3.1 Dataset dan Statistik Data

Dataset dalam penelitian ini berjumlah 10.834 komentar YouTube yang berkaitan dengan pembahasan kecerdasan buatan di bidang pendidikan. Data tersebut diperoleh melalui *YouTube Data API* dan kemudian dilakukan pelabelan secara manual berdasarkan polaritas sentimen masing-masing komentar. Setiap komentar diklasifikasikan ke dalam tiga kategori sentimen, yaitu negatif, netral, dan positif.

Distribusi dari dataset yang telah dilabeli ditampilkan pada Tabel 5.

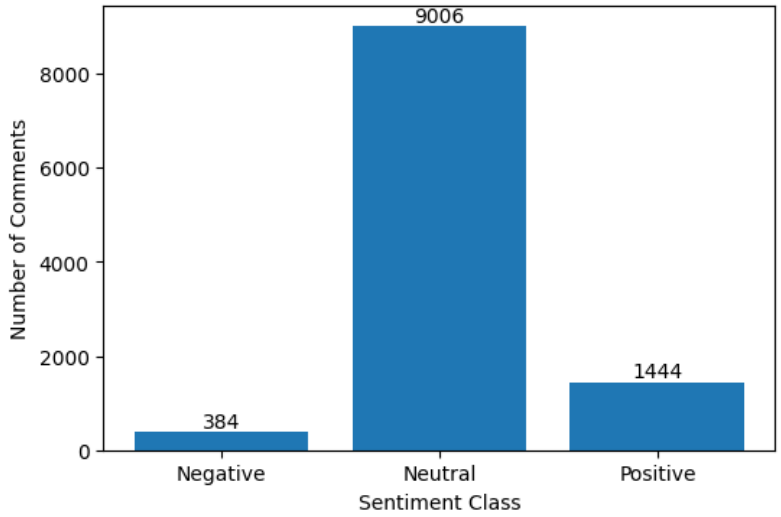
Tabel 5. Distribusi Sentimen Dataset

Sentimen	Label	Jumlah
Negatif	0	384
Netral	1	9006
Positif	2	1444
Total	-	10.834

Berdasarkan Tabel 5, terlihat bahwa kelas sentimen netral mendominasi dataset dengan persentase lebih dari 80%. Kondisi ini menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) yang berpotensi memengaruhi performa model, terutama dalam mengklasifikasikan kelas minoritas seperti sentimen negatif.

Distribusi ini juga ditunjukkan pada Gambar 2, yang memperlihatkan dominasi kelas netral dibandingkan kelas lainnya. Ketidakseimbangan ini menjadi tantangan dalam proses klasifikasi karena model cenderung bias terhadap kelas mayoritas.

Visualisasi distribusi kelas sentimen disajikan pada Gambar 2.



Gambar 2. Distribusi kelas sentimen dalam dataset

Gambar 2 menunjukkan bahwa kelas netral mendominasi distribusi data secara signifikan. Hal ini memperkuat temuan pada Tabel 5 mengenai adanya ketidakseimbangan kelas yang perlu diperhatikan dalam proses pelatihan model.

3.1.1 Hasil *Pre-processing* Data

Tahap *pre-processing* dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam pelatihan model. Hasil dari setiap tahapan dijelaskan sebagai berikut:

a. *Case Folding*

Proses ini mengubah seluruh teks menjadi huruf kecil untuk menjaga konsistensi format. Hasil dari proses *case folding* disajikan pada Tabel 6.

Tabel 6. Hasil *Case Folding*

Comment	Case Folding
Tidak ada yang bisa kalah kita sebagai manusia,, itu hanya orang2 tidak ingat ilmu agama, karna manusia Allah SWT yng menciptakan nya, manusia mahluk yng sempurna	tidak ada yang bisa kalah kita sebagai manusia,, itu hanya orang2 tidak ingat ilmu agama, karna manusia allah swt yng menciptakan nya, manusia mahluk yng sempurna
Otak lebih pintar dari teknologi	otak lebih pintar dari teknologi
Ya tetap otak lah, tidak mungkin ada Ai, tanpa ada otak yg ber ide	ya tetap otak lah, tidak mungkin ada ai, tanpa ada otak yg ber ide

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 671
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9697
Hal 666-678
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

saya kira cupu ternyata suhu pintar menguraikan , apa sebenarnya Ai .	saya kira cupu ternyata suhu pintar menguraikan , apa sebenarnya ai .
Posisi penting, gaji kecil. Posisi ga penting, perusahaan milik Negara merugi, gajinya gede bangetttt.....	posisi penting, gaji kecil. posisi ga penting, perusahaan milik negara merugi, gajinya gede bangetttt.....

Berdasarkan Tabel 6, proses case folding berhasil menyeragamkan seluruh teks menjadi huruf kecil, sehingga mengurangi variasi penulisan yang tidak diperlukan dan membantu model dalam mengenali pola kata secara lebih konsisten.

b. *Text Cleaning*

Tahap ini menghapus karakter yang tidak relevan seperti tanda baca dan simbol. Hasil dari proses *text cleaning* disajikan pada Tabel 7.

Tabel 7. Hasil *Text Cleaning*

Case Folding	Text Cleaning
tidak ada yang bisa kalah kita sebagai manusia,, itu hanya orang2 tidak ingat ilmu agama, karna manusia allah swt yng menciptakan nya, manusia mahluk yng sempurna	tidak ada yang bisa kalah kita sebagai manusia itu hanya orang tidak ingat ilmu agama karna manusia allah swt yng menciptakan nya manusia mahluk yng sempurna
otak lebih pintar dari teknologi	otak lebih pintar dari teknologi
ya tetap otak lah, tidak mungkin ada ai, tanpa ada otak yg ber ide	ya tetap otak lah tidak mungkin ada ai tanpa ada otak yg ber ide
saya kira cupu ternyata suhu pintar menguraikan , apa sebenarnya ai .	saya kira cupu ternyata suhu pintar menguraikan apa sebenarnya ai
posisi penting, gaji kecil. posisi ga penting, perusahaan milik negara merugi, gajinya gede bangetttt.....	posisi penting gaji kecil posisi ga penting perusahaan milik negara merugi gajinya gede bangetttt

Berdasarkan Tabel 7, proses text cleaning berhasil menghapus karakter yang tidak relevan seperti tanda baca dan simbol, sehingga menghasilkan teks yang lebih bersih dan siap untuk diproses lebih lanjut.

c. *Stopword Removal*

Menghapus kata umum yang tidak memiliki makna signifikan. Hasil dari proses *stopword removal* disajikan pada Tabel 8.

Tabel 8. Hasil *Stopword Removal*

Cleaning	Stopword Removal
tidak ada yang bisa kalah kita sebagai manusia itu hanya orang tidak ingat ilmu agama karna manusia allah swt yng menciptakan nya manusia mahluk yng sempurna	kalah manusia orang ingat ilmu agama karna manusia allah swt yng menciptakan nya manusia mahluk yng sempurna
otak lebih pintar dari teknologi	otak lebih pintar teknologi
ya tetap otak lah tidak mungkin ada ai tanpa ada otak yg ber ide	tetap otak lah mungkin ai otak yg ber ide
saya kira cupu ternyata suhu pintar menguraikan apa sebenarnya ai	kira cupu ternyata suhu pintar menguraikan apa sebenarnya ai
posisi penting gaji kecil posisi ga penting perusahaan milik negara merugi gajinya gede bangetttt	posisi penting gaji kecil posisi ga penting perusahaan milik negara merugi gajinya gede bangetttt

Berdasarkan Tabel 8, proses stopword removal mampu menghilangkan kata-kata umum yang tidak memiliki kontribusi signifikan terhadap makna kalimat, sehingga meningkatkan fokus model pada kata-kata penting.

d. *Text Normalization*

Mengubah kata tidak baku menjadi bentuk baku. Hasil dari proses *text normalization* disajikan pada Tabel 9.

Tabel 9. *Hasil Text Normalisasi*

Stopword Removal	Text Normalization
kalah manusia orang ingat ilmu agama karna manusia allah swt yng menciptakan nya manusia mahluk yng sempurna	kalah manusia orang ingat ilmu agama karna manusia allah swt yang menciptakan nya manusia mahluk yang sempurna
otak lebih pintar teknologi	otak lebih pintar teknologi
tetap otak lah mungkin ai otak yg ber ide	tetap otak lah mungkin ai otak yang ber ide
kira cupu ternyata suhu pintar menguraikan apa sebenarnya ai	kira cupu ternyata suhu pintar menguraikan apa sebenarnya ai
posisi penting gaji kecil posisi ga penting perusahaan milik negara merugi gajinya gede bangetttt	posisi penting gaji kecil posisi tidak penting perusahaan milik negara merugi gajinya gede banget

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 672
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9697
Hal 666-678
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Berdasarkan Tabel 9, proses normalisasi berhasil mengubah kata tidak baku menjadi bentuk baku, sehingga meningkatkan konsistensi linguistik dalam dataset.

e. *Duplicate Removal*

Menghapus data duplikat untuk meningkatkan kualitas dataset. Hasil dari proses *duplicate removal* disajikan pada Tabel 10.

Tabel 10. Hasil Penghapusan Duplikat

Deskripsi	Jumlah Data
Jumlah data sebelum penghapusan	10.834
Jumlah data setelah penghapusan	10.398
Jumlah duplikasi yang dihapus	436

Berdasarkan Tabel 10, proses penghapusan duplikasi berhasil mengurangi jumlah data yang redundan, sehingga meningkatkan kualitas dataset dan menghindari bias dalam pelatihan model.

f. Filtering Topik

Tahap filtering topik dilakukan untuk memastikan bahwa data yang digunakan dalam penelitian benar-benar relevan dengan pembahasan mengenai kecerdasan buatan dalam pendidikan. Proses ini dilakukan dengan pendekatan berbasis keyword matching menggunakan daftar kata kunci yang mencakup istilah terkait artificial intelligence, teknologi pembelajaran, serta konteks pendidikan seperti sekolah, kampus, guru, siswa, pembelajaran, dan istilah etika serta dampak penggunaan AI.

Berdasarkan hasil implementasi filtering, jumlah data sebelum proses penyaringan adalah sebanyak 10.398 komentar, kemudian setelah dilakukan filtering berdasarkan kesesuaian kata kunci tematik diperoleh 4.726 komentar yang relevan. Dengan demikian, sebanyak 5.672 data dinyatakan tidak relevan dan dihapus dari dataset karena tidak mengandung konteks pembahasan yang sesuai dengan fokus penelitian.

Hasil ini menunjukkan bahwa pendekatan filtering berbasis keyword mampu meningkatkan relevansi dataset secara signifikan dengan memfokuskan data pada topik penelitian, yaitu kecerdasan buatan dalam konteks pendidikan, sehingga data yang digunakan dalam tahap analisis selanjutnya lebih bersih, terarah, dan sesuai dengan tujuan penelitian.

Hasil dari proses *filtering* topik disajikan pada Tabel 11.

Tabel 11. Hasil *Filtering*

Deskripsi	Jumlah
Sebelum	10.398
Sesudah	4.726
Terhapus	5.672

Berdasarkan Tabel 11, terjadi pengurangan jumlah data yang signifikan setelah proses filtering, yang menunjukkan bahwa banyak data awal tidak relevan dengan topik penelitian. Proses ini penting untuk memastikan bahwa model hanya dilatih menggunakan data yang sesuai dengan konteks penelitian.Tahapan *pre-processing* ini terbukti mampu mengurangi *noise* serta meningkatkan relevansi data terhadap topik penelitian.

3.1.2 Hasil Tokenisasi

Tokenisasi dilakukan untuk mengubah teks menjadi unit kata atau sub-kata. Hasil dari proses tokenisasi disajikan pada Tabel 12.

Tabel 12. Hasil Tokenisasi

Text Normalization	Tokenization
kalah manusia orang ingat ilmu agama karna manusia allah swt yang menciptakan nya manusia mahluk yang sempurna	[kalah, manusia, orang, ingat, ilmu, agama, karna, manusia, allah, swt, yang, menciptakan, nya, manusia, mahluk, yang, sempurna]
otak lebih pintar teknologi	[otak, lebih, pintar, teknologi]
tetap otak lah mungkin ai otak yang ber ide	[tetap, otak, lah, mungkin, ai, otak, yang, ber, ide]
kira cupu ternyata suhu pintar menguraikan apa sebenarnya ai	[kira, cupu, ternyata, suhu, pintar, menguraikan, apa, sebenarnya, ai]
posisi penting gaji kecil posisi tidak penting perusahaan milik negara merugi gajinya gede banget	[posisi, penting, gaji, kecil, posisi, tidak, penting, perusahaan, milik, negara, merugi, gajinya, gede, banget]

Berdasarkan Tabel 12, proses tokenisasi berhasil mengubah teks menjadi unit token yang dapat diproses oleh model. Representasi ini memungkinkan model untuk memahami struktur dan hubungan antar kata dalam kalimat.

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 673
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9697
Hal 666-678
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

3.1.3 Implementasi dan Pengujian Model

Pengujian dilakukan dalam dua skenario, yaitu *baseline* dan *fine-tuning*.

a. *Baseline* Model

Eksperimen pertama mengevaluasi kinerja awal model tanpa penerapan *fine-tuning* tambahan. Dalam skenario ini, model pra-latih BERT dan IndoRoBERTa langsung diterapkan pada dataset.

Hasil evaluasi model pada skenario baseline ditampilkan pada Tabel 13.

Tabel 13. Hasil *Baseline*

Model	Accuracy	Precision	Recall	F1-score
BERT	0.3977	0.6286	0.3977	0.4631
IndoRoBERTa	0.4020	0.6128	0.4020	0.4635

Berdasarkan Tabel 13, kedua model menunjukkan performa yang relatif rendah, yang mengindikasikan bahwa model pra-latih belum mampu menangkap konteks spesifik domain tanpa proses penyesuaian lebih lanjut. IndoRoBERTa sedikit mengungguli BERT dengan akurasi 40,20%, sedangkan BERT mencapai 39,77%. Temuan ini menunjukkan bahwa model pra-latih saja belum cukup untuk menangkap karakteristik kontekstual spesifik dari diskusi terkait kecerdasan buatan dalam pendidikan.

b. *Fine-Tuning* Model

Pada eksperimen kedua, kedua model dilakukan *fine-tuning* menggunakan dataset berlabel. Proses ini memungkinkan model menyesuaikan parameter sesuai dengan tugas klasifikasi sentimen.

Hasil evaluasi model setelah proses fine-tuning ditampilkan pada Tabel 14.

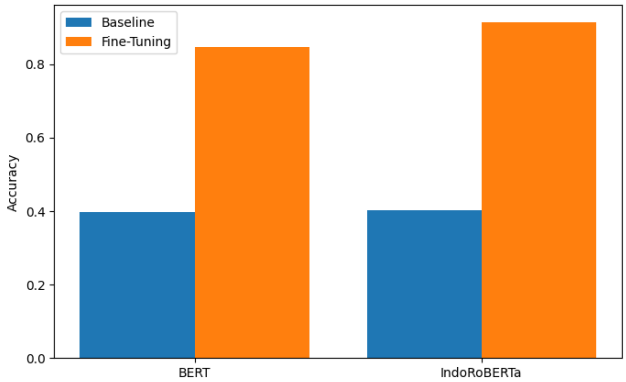
Tabel 14. Hasil *Fine-Tuning*

Model	Accuracy	Precision	Recall	F1-score
BERT	0.8463	0.8379	0.8463	0.8413
IndoRoBERTa	0.9154	0.9132	0.9154	0.9134

Berdasarkan Tabel 14, terlihat adanya peningkatan performa yang signifikan setelah dilakukan fine-tuning IndoRoBERTa memperoleh tingkat akurasi tertinggi sebesar 91,54%, lebih unggul dibandingkan BERT yang mencapai akurasi sebesar 84,63%. Temuan ini mengindikasikan bahwa proses *fine-tuning* membantu model dalam memahami pola linguistik yang bersifat kontekstual pada komentar berbahasa Indonesia secara lebih baik.

Untuk menggambarkan peningkatan tersebut, dilakukan perbandingan antara model *baseline* dan model hasil *fine-tuning*.

Perbandingan performa model sebelum dan sesudah fine-tuning ditampilkan pada Gambar 4.



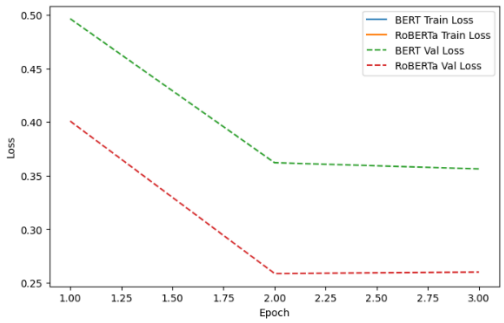
Gambar 4. Perbandingan akurasi antara model *baseline* dan model *fine-tuning*

Gambar 4 menunjukkan bahwa kedua model mengalami peningkatan performa yang signifikan setelah proses fine-tuning. IndoRoBERTa mencatat peningkatan paling tinggi, dengan akurasi yang naik dari 40,20% pada tahap *baseline* menjadi 91,54% setelah *fine-tuning*. Sementara itu, BERT juga menunjukkan peningkatan yang cukup besar, dari 39,77% menjadi 84,63%. Temuan ini menunjukkan bahwa pelatihan yang disesuaikan dengan tugas (*task-specific training*) membantu model Transformer dalam memahami pola linguistik kontekstual pada komentar berbahasa Indonesia dengan lebih baik.

Untuk memperoleh gambaran lebih mendalam mengenai proses pelatihan selama *fine-tuning*, kurva pembelajaran (*learning curves*) yang menggambarkan *training loss* dan *validation loss* dari kedua model ditampilkan pada Gambar 5.

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 674
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9697
Hal 666-678
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



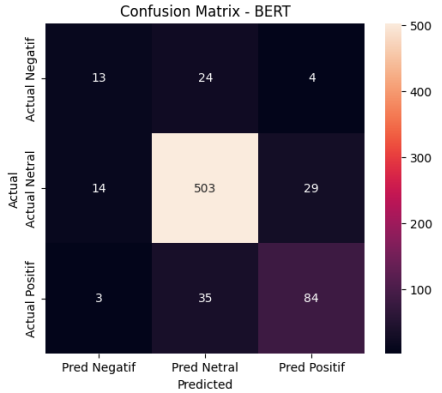
Gambar 5. Kurva *Training* dan *Validation Loss* per Epoch

Gambar 5 menampilkan kurva *training loss* dan *validation loss* dari model BERT dan IndoRoBERTa selama proses *fine-tuning*. Hasil yang diperoleh menunjukkan bahwa kedua model mengalami penurunan nilai *training loss* dan *validation loss* secara bertahap dari epoch 1 hingga epoch 2, yang menandakan proses pembelajaran dan optimasi berjalan dengan baik. Setelah memasuki epoch 2, nilai *loss* cenderung berada pada kondisi stabil, yang mengindikasikan bahwa model telah mencapai titik konvergensi.

IndoRoBERTa memiliki nilai *validation loss* yang lebih rendah dibandingkan BERT pada setiap epoch, yang menunjukkan kemampuan generalisasi yang lebih baik terhadap data yang belum pernah digunakan sebelumnya. Selain itu, perbedaan antara *training loss* dan *validation loss* pada kedua model relatif kecil, sehingga dapat disimpulkan bahwa tingkat overfitting yang terjadi sangat rendah. Hasil ini menegaskan bahwa pemilihan hiperparameter serta konfigurasi pelatihan yang diterapkan sudah sesuai, dan proses *fine-tuning* mampu meningkatkan performa model secara efektif.

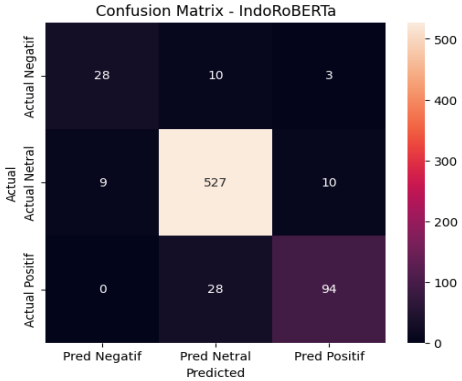
c. Analisis Confusion Matrix

Untuk melakukan analisis lebih mendalam terhadap kinerja klasifikasi, digunakan metode *confusion matrix*. *Confusion matrix* menggambarkan keterkaitan antara label hasil prediksi dengan label sebenarnya dalam dataset. Kinerja model IndoRoBERTa dapat dilihat pada Gambar 6.



Gambar 6. Confusion Matrix BERT

Berdasarkan gambar 6, model BERT menunjukkan performa yang baik dalam mengklasifikasikan sentimen netral yang merupakan kelas dominan dalam dataset. Namun demikian, kemampuan model dalam mengidentifikasi sentimen negatif masih relatif lebih rendah, mengingat kelas tersebut memiliki jumlah data yang paling sedikit. Kinerja model IndoRoBERTa dapat dilihat pada Gambar 7.



Gambar 7. Confusion Matrix IndoRoBERTa

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 675
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9697
Hal 666-678
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Berdasarkan gambar 7, IndoRoBERTa menunjukkan performa klasifikasi yang lebih seimbang antar kelas sentimen, serta kemampuan yang lebih baik dalam mengenali sentimen positif dan negatif.

3.2 Pembahasan

Hasil penelitian pada Tabel 5 menunjukkan bahwa distribusi dataset tidak seimbang dengan dominasi kelas netral yang mencapai lebih dari 80% dari total data, sehingga kondisi ini berpengaruh langsung terhadap performa model karena kecenderungan model dalam mempelajari pola dari kelas mayoritas dapat menurunkan kemampuan generalisasi pada kelas minoritas [3], [14]. Hal tersebut tercermin pada hasil baseline (Tabel 13) di mana BERT dan IndoRoBERTa hanya mencapai akurasi di bawah 41%, yang mengindikasikan bahwa model pralatih tanpa penyesuaian belum mampu menangkap karakteristik linguistik yang kompleks dan kontekstual dari komentar YouTube berbahasa Indonesia [6], [18].

Setelah dilakukan fine-tuning (Tabel 14), terjadi peningkatan performa yang signifikan pada kedua model, di mana IndoRoBERTa mencapai akurasi 91,54% dan BERT 84,63%, yang menunjukkan bahwa proses penyesuaian parameter terhadap data spesifik mampu meningkatkan kemampuan representasi semantik model secara efektif [16], [17]. Temuan ini diperkuat oleh Gambar 4 yang memperlihatkan perbedaan yang jelas antara performa baseline dan fine-tuned.

Selanjutnya, analisis kurva pembelajaran pada Gambar 5 menunjukkan bahwa training loss dan validation loss mengalami penurunan secara stabil hingga epoch ke-2 dan kemudian cenderung konvergen, yang menandakan proses pelatihan berjalan optimal. Selain itu, IndoRoBERTa secara konsisten menunjukkan nilai validation loss yang lebih rendah dibandingkan BERT, yang mencerminkan kemampuan generalisasi yang lebih baik serta stabilitas model selama proses pelatihan [6].

Lebih lanjut, analisis confusion matrix pada Gambar 6 dan Gambar 7 menunjukkan bahwa BERT cenderung lebih bias terhadap kelas netral, sedangkan IndoRoBERTa menghasilkan distribusi prediksi yang lebih seimbang pada seluruh kelas, termasuk kelas positif dan negatif, sehingga menunjukkan kemampuan klasifikasi yang lebih robust terhadap ketidakseimbangan data [14].

Jika dikaitkan dengan konteks penelitian ini, hasil tersebut menguatkan fenomena bahwa persepsi publik terhadap penerapan kecerdasan buatan dalam pendidikan yang terekam melalui komentar YouTube bersifat beragam, tidak terstruktur, dan cenderung didominasi opini netral sebagaimana dijelaskan dalam latar belakang penelitian [7], [9]. Hal ini menegaskan bahwa karakteristik data media sosial tidak hanya mencerminkan aspek teknis, tetapi juga dinamika sosial, etika, dan psikologis masyarakat dalam merespons teknologi AI [4], [5]. Oleh karena itu, kemampuan model dalam memahami konteks linguistik informal menjadi sangat penting untuk merepresentasikan opini publik secara lebih akurat [11], [12], [15].

Jika dibandingkan dengan pendekatan sebelumnya seperti Naïve Bayes, SVM, CNN, dan Bi-LSTM, model Transformer memberikan keunggulan karena mampu menangkap konteks dua arah serta hubungan semantik yang lebih mendalam pada teks tidak terstruktur, sehingga lebih efektif digunakan dalam analisis sentimen pada data media sosial [19], [6]. Secara keseluruhan, IndoRoBERTa menunjukkan kinerja terbaik dalam seluruh skenario pengujian, baik dari aspek akurasi, stabilitas pelatihan, maupun kemampuan dalam memahami konteks linguistik yang kompleks, sehingga relevan untuk digunakan dalam analisis persepsi publik terhadap implementasi kecerdasan buatan di bidang pendidikan [1], [2].

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa analisis sentimen komentar YouTube berbahasa Indonesia terkait kecerdasan buatan dalam pendidikan menggunakan model Transformer BERT dan IndoRoBERTa menunjukkan bahwa dataset sebanyak 10.834 komentar yang diperoleh melalui YouTube Data API memiliki ketidakseimbangan kelas yang signifikan dengan dominasi kelas netral lebih dari 80% yang berdampak pada rendahnya performa baseline di bawah 41%. Penerapan fine-tuning pada kedua model mampu meningkatkan kinerja secara signifikan, di mana IndoRoBERTa memperoleh akurasi sebesar 91,54% dan BERT sebesar 84,63%, yang menunjukkan efektivitas adaptasi model terhadap konteks bahasa Indonesia informal. Hasil evaluasi juga memperlihatkan bahwa IndoRoBERTa memiliki kemampuan klasifikasi yang lebih seimbang pada seluruh kelas sentimen dibandingkan BERT, terutama pada kelas minoritas negatif. Selain itu, hasil confusion matrix menunjukkan performa prediksi yang lebih stabil, sedangkan kurva loss mengindikasikan proses konvergensi yang baik dengan risiko overfitting yang rendah. Secara keseluruhan, IndoRoBERTa terbukti lebih unggul dalam menangkap konteks linguistik dan menangani ketidakseimbangan data, namun penelitian ini masih memiliki keterbatasan pada sumber data yang hanya berasal dari satu platform serta belum diterapkannya teknik penyeimbangan data seperti oversampling atau data augmentation. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas cakupan data lintas platform serta meningkatkan teknik preprocessing agar menghasilkan model yang lebih general dan adaptif pada domain analisis sentimen pendidikan digital.

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 676
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9697
Hal 666-678
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada seluruh pihak yang telah memberikan dukungan dalam pelaksanaan penelitian ini, baik dalam bentuk bimbingan, arahan, maupun bantuan teknis maupun nonteknis sehingga penelitian dapat berjalan dengan baik dan selesai sesuai dengan yang direncanakan, khususnya kepada dosen pembimbing yang telah memberikan masukan konstruktif selama proses penyusunan, serta kepada pihak institusi yang telah memfasilitasi kebutuhan akademik dan akses sumber daya yang diperlukan, tidak lupa kepada keluarga dan rekan-rekan yang senantiasa memberikan doa, motivasi, serta dukungan moral selama proses penelitian berlangsung, sehingga kontribusi dan dukungan dari berbagai pihak tersebut sangat berarti dalam penyelesaian penelitian ini.

REFERENCES

[1] H. S. J. Chew dan P. Achananuparp, "Perceptions and Needs of Artificial Intelligence in Health Care to Increase Adoption," vol. 24, hal. 1–19, 2022, doi: 10.2196/32939.
[2] R. Sahar dan M. Munawaroh, *Artificial intelligence in higher education with bibliometric and content analysis for future research agenda*. Springer International Publishing, 2025. doi: 10.1007/s43621-025-01086-z.
[3] M. Jamil, H. Hadiyanto, dan R. Sanjaya, "Sentiment Analysis: Classifying Public Comments on YouTube in Disaster Management Simulation in Indonesia Using Naïve Bayes and Support Vector Machine," *Ingénierie des systèmes d Inf.*, vol. 29, no. 2, hal. 437–446, Apr 2024, doi: 10.18280/isi.290205.
[4] J. Kamali, M. F. Alpat, dan A. Bozkurt, "AI ethics as a complex and multifaceted challenge : decoding educators ' AI ethics alignment through the lens of activity theory," *Int. J. Educ. Technol. High. Educ.*, 2024, doi: 10.1186/s41239-024-00496-9.
[5] H. Zhu, Y. Sun, dan J. Yang, "Towards responsible artificial intelligence in education: a systematic review on identifying and mitigating ethical risks," *Humanit. Soc. Sci. Commun.*, 2025, doi: https://doi.org/10.1057/s41599-025-05252-6.
[6] A. Bello, S. C. Ng, dan M. F. Leung, "A BERT Framework to Sentiment Analysis of Tweets," *Sensors*, vol. 23, no. 1, 2023, doi: 10.3390/s23010506.
[7] G. Yumitro, R. Febriani, A. Roziqin, dan A. Indraningtyas, "Bibliometric analysis of international publication trends on social media and terrorism by using the Scopus database," *Front. inCommunication*, 2021, doi: https://doi.org/10.3389/fcomm.2023.1140461.
[8] H. Shaheen, "Social Media Marketing: A Bibliometric Analysis from Scopus," *Futur. Bus. J.*, 2024, doi: 10.1186/s43093-025-00465-2.
[9] D. Sun dan Y. Li, "behavioral sciences Influence of Strategic Crisis Communication on Public Perceptions during Public Health Crises : Insights from YouTube Chinese Media," *Behav. Sci. (Basel).*, 2024, doi: https://doi.org/10.3390/bs14020091.
[10] A. Raza *et al.*, "An improved deep convolutional neural network-based YouTube video classification using textual features," *Heliyon*, vol. 10, no. 16, hal. e35812, 2024, doi: 10.1016/j.heliyon.2024.e35812.
[11] D. A. Musleh *et al.*, "Arabic Sentiment Analysis of YouTube Comments: NLP-Based Machine Learning Approaches for Content Evaluation," *Big Data Cogn. Comput.*, vol. 7, no. 3, 2023, doi: 10.3390/bdcc7030127.
[12] S. M. Islam *et al.*, *"Challenges and future in deep learning for sentiment analysis : a comprehensive review and a proposed novel hybrid approach,"* vol. 57, no. 3. Springer Netherlands, 2024. doi: 10.1007/s10462-023-10651-9.
[13] E. A. El-dahshan, M. M. Bassiouni, A. Hagag, R. K. Chakrabortty, H. Loh, dan U. R. Acharya, "RESCOVIDTCNnet : A residual neural network-based framework for COVID-19 detection using TCN and EWT with chest X-ray images," *Expert Syst. Appl.*, vol. 204, no. April, hal. 117410, 2022, doi: 10.1016/j.eswa.2022.117410.
[14] Y. Mao, Q. Liu, dan Y. Zhang, "Sentiment analysis methods , applications , and challenges : A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, hal. 102048, 2024, doi: 10.1016/j.jksuci.2024.102048.
[15] K. L. Tan, C. P. Lee, dan K. M. Lim, "A Survey of Sentiment Analysis : Approaches, Datasets, and Future Research," *Appl. Sci.*, 2023, doi: https://doi.org/10.3390/app13074550.
[16] X. Gong, W. Ying, S. Zhong, dan S. Gong, "Text Sentiment Analysis Based on Transformer and Augmentation," *Orig. Res.*, vol. 13, no. May, 2022, doi: 10.3389/fpsyg.2022.906061.
[17] S. Tzimiris, S. Nikiforos, M. N. Nikiforos, K. L. Kermanidis, dan D. Mouratidis, "A Comparative Evaluation of Transformer-Based Language Models for Topic-Based Sentiment Analysis," *Electron.*, 2025, [Daring]. Tersedia pada: https://www.mdpi.com/2079-9292/14/15/2957
[18] M. Jojoa, P. Eftekhar, B. Nowrouzi, K. Begonya, dan G. Zapirain, "Natural language processing analysis applied to COVID - 19 open - text opinions using a distilBERT model for sentiment categorization," *AI Soc.*, vol. 39, no. 3, hal. 883–890, 2024, doi: 10.1007/s00146-022-01594-w.
[19] A. Nahid, D. Pramesti, A. D. Fathurahman, dan H. Fakhrurroja, "Exploring Sentiment Analysis for the Indonesian Presidential Election Through Online Reviews Using Multi-Label Classification with a Deep Learning Algorithm," *Information*, hal. 1–33, 2024, doi: https://doi.org/10.3390/info15110705.
[20] Z. Xu, X. Wen, G. Zhong, dan Q. Fang, "Public perception towards deepfake through topic modelling and sentiment analysis of social media data," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, 2025, doi: 10.1007/s13278-025-01445-8.
[21] A. B. Alawi dan F. Bozkurt, "A hybrid machine learning model for sentiment analysis and satisfaction assessment with Turkish universities using Twitter data," *Decis. Anal. J.*, vol. 11, no. April, hal. 100473, 2024, doi: 10.1016/j.dajour.2024.100473.
[22] W. Van Atteveldt, M. A. C. G. Van Der Velden, dan M. Boukes, "The Validity of Sentiment Analysis : Comparing Manual Annotation , Crowd-Coding , Dictionary Approaches , and Machine Learning Algorithms The Validity of Sentiment Analysis : Comparing Manual Annotation ," *Commun. Methods Meas.*, vol. 15, no. 2, hal. 121–140, 2021, doi: 10.1080/19312458.2020.1869198.
[23] A. Alaei, Y. Wang, V. Bui, dan B. Stantic, "Target-Oriented Data Annotation for Emotion and Sentiment Analysis in Tourism Related Social Media Data," *Futur. Internet*, 2023.
[24] S. Frenda, G. Abercrombie, V. Basile, A. Pedrani, dan D. Bernardi, "Perspectivist approaches to natural language processing : a survey," *Lang. Resour. Eval.*, vol. 59, no. 2, hal. 1719–1746, 2025, doi: 10.1007/s10579-024-09766-4.

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 677
This Journal is licensed under a Creative Commons Attribution 4.0 International License

[25] C. A. Martínez-miwa dan M. Castelán, "On reliability of annotations in contextual emotion imagery," *Sci. Data*, hal. 1–12, 2023, doi: 10.1038/s41597-023-02435-1.
[26] K. Lindén, T. Jauhiainen, dan S. Hardwick, "FinnSentiment : a Finnish social media corpus for sentiment polarity annotation," *Lang. Resour. Eval.*, vol. 57, no. 2, hal. 581–609, 2023, doi: 10.1007/s10579-023-09644-5.
[27] N. Stefanovitch, "Holistic Inter-Annotator Agreement and Corpus Coherence Estimation in a Large-scale Multilingual Annotation Campaign," *Antol. ACL*, hal. 71–86, 2023, doi: 10.18653/v1/2023.emnlp-main.6.
[28] I. Your, T. Data, R. Ground, T. A. Quality, dan C. Delineation, "Is Your Training Data Really Ground Truth ? A Quality Assessment of Manual Annotation for Individual Tree Crown Delineation," *Remote Sens.*, 2024, doi: https://doi.org/10.3390/app13074550.
[29] M. Siino, I. Tinnirello, dan M. La Cascia, "Is text preprocessing still worth the time ? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers," *Inf. Syst.*, vol. 121, no. December 2023, hal. 102342, 2024, doi: 10.1016/j.is.2023.102342.
[30] Y. Fissha, H. Ikeda, H. Toriya, dan T. Adachi, "Application of Bayesian Neural Network (BNN) for the Prediction of Blast-Induced Ground Vibration," *Appl. Sci.*, 2023, doi: https://doi.org/10.3390/app13053128.
[31] J. R. Jim, M. A. R. Talukder, P. Malakar, dan M. M. Kabir, "Recent advancements and challenges of NLP-based sentiment analysis : A state-of-the-art review," *Nat. Lang. Process. J.*, vol. 6, no. January, hal. 100059, 2024, doi: 10.1016/j.nlp.2024.100059.
[32] A. Sharma, A. Z. Ansari, R. Kakulavarapu, M. H. Stensen, M. A. Riegler, dan H. L. Hammer, "Predicting Cell Cleavage Timings from Time-Lapse Videos of Human Embryos," *Big Data Cogn. Comput.*, 2023, doi: https://doi.org/10.3390/bdcc7020091.
[33] H. T. Y. Achsan, H. Suhartanto, W. C. Wibowo, D. A. Dewi, dan K. Ismed, "Automatic Extraction of Indonesian Stopwords," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 2, hal. 166–171, 2023, [Daring]. Tersedia pada: https://scholar.ui.ac.id/en/publications/automatic-extraction-of-indonesian-stopwords/
[34] S. S. Louvros, M. Paraskevas, dan T. Chrysikos, "QoS-Aware Resource Management in 5G and 6G Cloud-Based Architectures with Priorities," *Information*, 2023, doi: https://doi.org/10.3390/info14030175.
[35] J. Khan dan S. Lee, "applied sciences Enhancement of Text Analysis Using Context-Aware Normalization of Social Media Informal Text," *Appl. Sci.*, 2021, doi: https://doi.org/10.3390/app11178172.
[36] R. E. Petruse, "applied sciences Enhancing Metal Forging Tools and Moulds : Advanced Repairs and Optimisation Using Directed Energy Deposition Hybrid Manufacturing," *Appl. Sci.*, 2024, doi: https://doi.org/10.3390/app14020567.
[37] Y. Mu, M. Jin, X. Song, dan N. Aletras, "Enhancing Data Quality through Simple De-duplication : Navigating Responsible Computational Social Science Research," *Proc. 2024 Conf. Empir. Methods Nat. Lang. Process.*, hal. 12477–12492, 2024, [Daring]. Tersedia pada: https://aclanthology.org/2024.emnlp-main.694/
[38] M. Hu, X. Li, M. Li, R. Zhu, dan B. Si, "A Framework for Analyzing Fraud Risk Warning and Interference Effects by Fusing Multivariate Heterogeneous Data : A Bayesian Belief Network," *Entropy*, 2023, doi: https://doi.org/10.3390/e25060892.
[39] R. V. Santin dan S. O. Rezende, "Systematic review on aspect-based sentiment analysis in cross-domain," *Artif. Intell. Rev.*, 2025, [Daring]. Tersedia pada: https://link.springer.com/article/10.1007/s10462-025-11437-x
[40] F. Alrashidi dan H. I. Mathkour, "An Empirical Study of Transformer - Based Neural Machine Translation for English to Arabic," *Information*, 2026, doi: https://doi.org/10.3390/info17020198.
[41] E. Kotei dan R. Thirunavukarasu, "A Systematic Review of Transformer-Based Pre-Trained Language Models through Self-Supervised Learning," *Information*, 2023, doi: https://doi.org/10.3390/info14030187.
[42] A. Baiocchi, I. Spinelli, A. Nicolosi, dan S. Scardapane, "Adaptive token selection for scalable point cloud transformers," *Neural Networks*, vol. 188, no. April, hal. 107477, 2025, doi: 10.1016/j.neunet.2025.107477.
[43] T. Huo, D. H. Glueck, E. A. Shenkman, dan K. E. Muller, "Stratified split sampling of electronic health records," *BMC Med. Res. Methodol.*, vol. 0, hal. 1–7, 2023, doi: https://doi.org/10.1186/s12874-023-01938-0.
[44] V. R. Joseph, A. Vakayil, V. R. Joseph, dan A. Vakayil, "SPlit : An Optimal Method for Data Splitting SPlit : An Optimal Method for Data Splitting ABSTRACT," *Technometrics*, vol. 0, no. 0, hal. 1–23, 2022, doi: 10.1080/00401706.2021.1921037.
[45] K. Deturck *et al.*, "Ertim at SemEval-2023 Task 2 : Fine-tuning of Transformer Language Models and External Knowledge Leveraging for NER in Farsi , English , French and Chinese," *Antol. ACL*, hal. 2211–2215, 2023, doi: 10.18653/v1/2023.semeval-1.306.
[46] N. Ding *et al.*, "Parameter-efficient fine-tuning of large-scale pre-trained language models," *Nat. Mach. iIntelligence*, vol. 5, no. March, 2023, doi: 10.1038/s42256-023-00626-4.
[47] L. Ashbaugh dan Y. Zhang, "A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning," *Computers*, 2024, doi: https://doi.org/10.3390/computers13120340.
[48] G. Kumar, W. Pankaj, K. Varshney, A. Gupta, dan S. Kumar, "Sentiment Analysis and Comprehensive Evaluation of Supervised Machine Learning Models Using Twitter Data on Russia – Ukraine War," *SN Comput. Sci.*, 2023, doi: 10.1007/s42979-023-01790-5.
[49] S. Sathyanarayanan dan B. R. Tantri, "Confusion Matrix-Based Performance Evaluation Metrics," *African J. Biomed. Res.*, vol. 27, no. 4, 2024, doi: https://doi.org/10.53555/AJBR.v27i4S.4345.
[50] J. S. A. Ruiz dan M. Michalak, "Classification performance assessment for imbalanced multiclass data," *Sci. Rep.*, hal. 1–10, 2024, doi: 10.1038/s41598-024-61365-z.

Copyright © 2026 The **Muhammad Mutawakkil Alallah**, Page 678
This Journal is licensed under a Creative Commons Attribution 4.0 International License