



# Analisis Sentimen Ulasan Aplikasi CapCut Berbasis IndoBERT dengan Validasi Silang

Tasya Nurdin<sup>1</sup>, Dodo Zaenal Abidin<sup>2</sup>, Kurniabudi<sup>3</sup>

<sup>1-3</sup>Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa, Indonesia

Email: [tasyanurdin504@gmail.com](mailto:tasyanurdin504@gmail.com)<sup>1</sup>, [dodozaenalabidin@gmail.com](mailto:dodozaenalabidin@gmail.com)<sup>2</sup>, [kurniabudi@unama.ac.id](mailto:kurniabudi@unama.ac.id)<sup>3</sup>

Alamat: Jl. Jenderal Sudirman, Thehok, Jambi

\*Penulis Korespondensi: [tasyanurdin504@gmail.com](mailto:tasyanurdin504@gmail.com)

**Abstract.** This study conducts sentiment analysis of Indonesian user reviews of the CapCut application using IndoBERT and compares two evaluation schemes: a single 80/20 train-test split and stratified 5-fold cross-validation ( $k=5$ ). A total of 1,048,575 reviews were collected from the Google Play Store through web scraping and labeled into three sentiment classes based on rating: negative (1–2), neutral (3), and positive (4–5). After preprocessing—cleaning, case folding, banned-word removal, normalization—and duplicate removal, 517,962 reviews were retained. IndoBERT Base P1 was fine-tuned using fixed hyperparameters (batch size 32, learning rate  $2e-5$ , up to 4 epochs, early stopping patience 2), while undersampling was applied to the training set to address class imbalance. Performance was assessed using accuracy, precision, recall, F1-score, and ROC-AUC, supported by confusion matrix and ROC-curve visualizations. The single split achieved an accuracy of 0.756, whereas cross-validation produced a mean accuracy of 0.740. Across both schemes, the positive class achieved the best performance (F1-score 0.850; ROC-AUC 0.918–0.919), while the neutral class remained the most challenging (precision 0.198–0.206; F1-score 0.280–0.283). Overall, cross-validation is recommended for reporting because it reduces dependence on a single partition and provides a more representative estimate across multiple splits.

**Keywords:** CapCut; Cross-validation; IndoBERT; Sentiment analysis; Transformer.

**Abstrak.** Penelitian ini melakukan analisis sentimen ulasan berbahasa Indonesia pada aplikasi CapCut menggunakan IndoBERT serta membandingkan dua skema evaluasi, yaitu single train-test split (80/20) dan stratified 5-fold cross-validation ( $k=5$ ). Sebanyak 1.048.575 ulasan dikumpulkan dari Google Play Store melalui web scraping dan dilabeli menjadi tiga kelas berdasarkan rating: negatif (1–2), netral (3), dan positif (4–5). Setelah pra-pemrosesan teks (cleaning, case folding, penghapusan kata terlarang, normalisasi) dan penghapusan duplikasi, diperoleh 517.962 ulasan untuk eksperimen. Model IndoBERT Base P1 di-fine-tune dengan parameter tetap (batch size 32, learning rate  $2e-5$ , maksimal 4 epoch, dan early stopping patience 2), serta diterapkan undersampling pada data latih untuk mengurangi ketidakseimbangan kelas. Evaluasi menggunakan accuracy, precision, recall, F1-score, dan ROC-AUC, serta didukung confusion matrix dan kurva ROC. Hasil menunjukkan accuracy pada skema single split sebesar 0,756, sedangkan cross-validation menghasilkan accuracy rata-rata 0,740. Performa terbaik konsisten pada kelas positif (F1-score 0,850; ROC-AUC 0,918–0,919), sementara kelas netral paling menantang (precision 0,198–0,206; F1-score 0,280–0,283). Secara keseluruhan, cross-validation direkomendasikan untuk pelaporan karena mengurangi ketergantungan pada satu kali pembagian data dan memberikan evaluasi yang lebih representatif melalui beberapa pembagian data.

**Kata kunci:** Analisis sentimen; CapCut; IndoBERT; Validasi silang; Transformer.

## 1. LATAR BELAKANG

Perkembangan pesat Natural Language Processing (NLP) mendorong pemanfaatannya di berbagai industri, khususnya untuk klasifikasi teks seperti analisis sentimen guna memahami umpan balik pelanggan (Kumar & Renuka, 2023). Analisis sentimen (penambahan opini) berfokus pada identifikasi opini, penilaian, sikap, dan emosi terhadap suatu entitas yang diekspresikan dalam teks, umumnya dikategorikan menjadi positif, netral, atau negatif (Liu, 2020). Teknik ini semakin penting karena masifnya data daring yang dimanfaatkan dalam

konteks bisnis, politik, maupun media sosial (Srivastava & Lin, 2025). Dalam ekosistem aplikasi digital, ulasan pengguna menjadi sumber data utama yang mencerminkan kepuasan dan keluhan, sehingga analisisnya krusial untuk mendukung keputusan strategis (Dąbrowski et al., 2022). Salah satu objek yang relevan adalah aplikasi CapCut, yang telah diunduh lebih dari satu miliar kali di Google Play Store (CapCut Video Editor on Google Play, n.d.), sehingga menghasilkan ulasan yang sangat besar, beragam, informal, dan tidak terstruktur.

Keterbatasan muncul ketika input dibatasi pada tingkat kalimat sehingga pemahaman konteks menjadi kurang optimal (Tao & Fang, 2020). Hal ini mendorong pergeseran menuju arsitektur Transformer yang terbukti lebih unggul pada berbagai benchmark, misalnya BERT dan variannya yang melampaui model klasik seperti SVM, LSTM, dan Naïve Bayes pada dataset IMDb (Sudhir & Suresh, 2021).

IndoBERT adalah sebuah model bahasa deep learning yang merupakan versi Indonesia dari BERT, yang cara kerjanya didasarkan pada arsitektur BERT dan dilatih dengan teknik transfer learning pada sebuah korpus masif berisi lebih dari 220 juta kata dari sumber-sumber berbahasa Indonesia formal, sehingga mampu mempelajari pola-pola bahasa pada tingkat yang sangat tinggi dan telah terbukti memberikan hasil yang lebih unggul daripada MalayBERT pada pengujian dataset IndoLEM (Asri et al., 2025).

Meskipun IndoBERT telah terbukti unggul, banyak penelitian masih mengandalkan evaluasi berbasis pembagian data latih–uji tunggal yang rentan bias dan kurang merepresentasikan kemampuan generalisasi model (Hua et al., 2021). Hal ini menegaskan kebutuhan metode evaluasi yang lebih robust, seperti validasi silang (*cross-validation*), agar estimasi kinerja lebih andal dan tidak bias, terutama karena penerapan NLP di dunia nyata dapat berdampak pada keputusan sensitif di berbagai domain (Abidin et al., 2025).

Karena itu, penelitian ini membandingkan evaluasi IndoBERT pada ulasan CapCut menggunakan single split dan stratified 5-fold *cross-validation*, agar hasil kinerja lebih representatif dan tidak bergantung pada satu pembagian data.

## 2. KAJIAN TEORITIS

NLP adalah bidang Kecerdasan Buatan (AI) yang berfokus pada interaksi antara komputer dan bahasa manusia. Bidang ini melibatkan pengembangan algoritma dan model statistik yang memungkinkan komputer memahami, menafsirkan, menghasilkan, dan memanipulasi bahasa alami dengan cara yang bermakna. *NLP* mencakup berbagai tugas, termasuk penerjemahan mesin, analisis sentimen, dan pengenalan suara, yang bertujuan menjembatani kesenjangan antara komunikasi manusia dan pemahaman komputer (Ferrari &

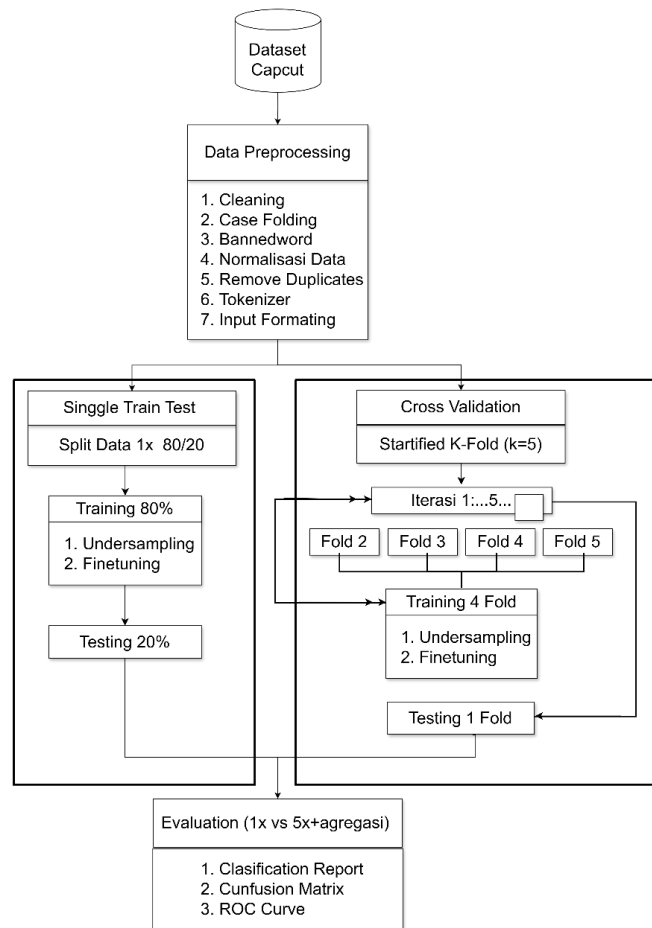
Ginde, 2025). Analisis sentimen, atau penambangan opini, merupakan bidang studi yang berfokus pada analisis opini, sentimen, penilaian, sikap, dan emosi terhadap entitas seperti produk, layanan, organisasi, individu, peristiwa, isu, maupun topik yang diekspresikan dalam teks tertulis (Liu, 2020). Melalui pendekatan ini, polaritas sentimen dapat dikenali dalam ulasan produk, komentar di media sosial, maupun umpan balik pengguna (Ferdous et al., 2024). Selain itu, analisis sentimen membantu menilai opini subjektif dalam skala besar, yang sulit dilakukan secara manual, terutama ketika data mencapai ribuan atau bahkan jutaan entri (Suhadi et al., 2025).

IndoBERT (*Indonesian Encoder Representations from Transformers*) merupakan model pre-training berbasis arsitektur BERT (Bidirectional Encoder Representations from Transformers) yang dilatih secara khusus untuk bahasa Indonesia menggunakan korpus besar Indo4B. Model ini memanfaatkan struktur transformer yang terdiri dari lapisan-lapisan encoder untuk mengubah token input menjadi representasi vektor yang kaya konteks (Apriyadi & Styawati, 2025). IndoBERT dikembangkan melalui pendekatan transfer learning dan dilatih pada korpus masif berbahasa Indonesia yang mencakup lebih dari 220 juta kata dari sumber-sumber berbahasa Indonesia formal, sehingga mampu mempelajari pola bahasa pada tingkat yang tinggi dan menunjukkan kinerja yang unggul (Asri et al., 2025). Studi oleh Baihaqi dan Munandar (Baihaqi & Munandar, 2023) melaporkan bahwa *IndoBERT* jauh menunjukkan performa yang baik dalam analisis sentimen.

Walaupun IndoBERT terbukti unggul, banyak penelitian masih bergantung pada evaluasi single train–test split yang rentan bias dan kurang merepresentasikan kemampuan generalisasi model (Hua et al., 2021). Sehingga diperlukan metode evaluasi yang lebih robust seperti validasi silang (*cross-validation*) untuk memberikan estimasi kinerja yang andal dan tidak bias, untuk benar-benar membuktikan keandalannya (Abidin et al., 2025).

### **3. METODE PENELITIAN**

Penelitian ini menggunakan tahapan kerja yang terstruktur untuk memastikan seluruh proses, mulai dari akuisisi data hingga evaluasi model, berjalan secara konsisten. Alur eksperimen penelitian ditunjukkan pada Gambar 1.



**Gambar 1.** Alur Eksperimen

Alur penelitian pada gambar 1 dijelaskan secara detail sebagai berikut :

### Pengumpulan dan Pelabelan Data

Data dikumpulkan dari ulasan pengguna aplikasi CapCut di Google Play Store menggunakan teknik web scraping, sehingga diperoleh sebanyak 1.048.575 ulasan berbahasa Indonesia. Pelabelan sentimen dilakukan berdasarkan skor rating, yaitu rating 1–2 dikategorikan sebagai sentimen negatif, rating 3 sebagai sentimen netral, dan rating 4–5 sebagai sentimen positif, sesuai dengan pendekatan yang telah divalidasi pada penelitian sebelumnya (Mola et al., 2025).

### Pra-pemrosesan Data

Data ulasan dipra-pemroses untuk meningkatkan kualitas dan konsistensi teks melalui cleaning (menghapus karakter tidak informatif), case folding, penghapusan kata terlarang (*bannedword*), serta normalisasi untuk mengurangi noise dan menyeragamkan variasi kata. Setelah itu dilakukan penghapusan data duplikat agar hanya ulasan unik dan valid yang digunakan sehingga menghindari bias. Selanjutnya, teks ditokenisasi dan diformat sebagai input model. Setelah pra-pemrosesan dan penghapusan duplikasi, jumlah data akhir yang digunakan dalam eksperimen adalah 517.962 ulasan.

## **Pembagian Data dan Skenario Eksperimen**

Eksperimen dilakukan menggunakan dua skema evaluasi, yaitu single train–test split (80/20) sebagai baseline dan stratified 5-fold cross-validation ( $k=5$ ). Pada skema train–test split, data dibagi satu kali menjadi 80% data latih dan 20% data uji. Pada skema cross-validation, data dibagi menjadi lima fold secara stratified; pada setiap iterasi, empat fold digunakan sebagai data latih dan satu fold sebagai data uji secara bergantian. Pengacakan dikontrol menggunakan random state untuk menjaga reproduibilitas dan keseimbangan distribusi kelas. Proses penyeimbangan kelas melalui undersampling diterapkan hanya pada data latih (baik pada baseline maupun pada setiap fold), agar evaluasi pada data uji tetap merepresentasikan kondisi sebenarnya.

## **Fine-Tuning dan Klasifikasi Menggunakan IndoBERT**

Model yang digunakan adalah IndoBERT Base P1 (indobenchmark/indobert-base-p1). Pada setiap skenario eksperimen, model di-*fine-tune* pada data latih yang telah di-undersampling dengan parameter tetap (*batch size 32, learning rate 2e-5*, maksimal 4 *epoch*, dan *early stopping patience 2*), kemudian digunakan untuk mengklasifikasikan sentimen pada data uji baik pada skema 80:20 maupun 5-fold cross-validation.

## **Evaluasi dan Visualisasi Hasil**

Performa IndoBERT dievaluasi dengan dua skema: train–test split 80/20 dan stratified 5-fold cross-validation. Pada split, metrik dihitung sekali dari 20% data uji; sedangkan pada cross-validation, pelatihan–pengujian dilakukan di tiap fold lalu metrik dirata-ratakan agar lebih representatif. Evaluasi memakai accuracy, precision, recall, F1-score, dan ROC-AUC, serta divisualisasikan dengan confusion matrix dan kurva ROC. Hasilnya dirangkum dalam tabel dan grafik untuk membandingkan baseline IndoBERT (split) dengan IndoBERT (cross-validation)

## **4. HASIL DAN PEMBAHASAN**

Pada bagian ini dijelaskan hasil kinerja model berdasarkan hasil eksperimen. Kinerja model menampilkan perbandingan nilai *accuracy, precision, recall, F1-score*, dan ROC-AUC, serta visualisasi confusion matrix dan kurva ROC.

### **Pra-pemrosesan (*Preprocessing*)**

Tahap pra-pemrosesan meliputi *Cleaning, Case Folding, Bannedword*, dan Normalisasi. Setelah itu dilakukan penghapusan duplikasi sehingga hanya ulasan unik yang digunakan. Selanjutnya, teks ditokenisasi dan diformat sebagai input model. Jumlah data

akhir yang dipakai dalam eksperimen adalah 517.962 ulasan. Ringkasan jumlah data serta contoh perubahan teks pada setiap tahap ditampilkan pada Tabel 1 dan Tabel 2.

**Tabel 1.** Ringkasan Dataset

Tahap	Jumlah Ulasan
Data hasil scraping	1.048.575
Data akhir setelah pra-pemrosesan dan hapus duplikat	517.962

**Tabel 2.** Hasil Pra-pemrosesan Data

No.	Sebelum	Cleaning	Case Folding	Bannedword	Normalisasi
1	aplikasi buruk masuk keluar terus ngedit satu ajah gak beres beres ðŸŽ	aplikasi buruk masuk keluar terus ngedit satu ajah gak beres beres	aplikasi buruk masuk keluar terus ngedit satu ajah gak beres beres	aplikasi buruk masuk keluar terus ngedit satu ajah gak beres beres	aplikasi buruk masuk keluar terus mengedit satu saja tidak beres beres
2	ayo lah asu	ayo lah asu	ayo lah asu	ayo lah asu	ayo lah anjing
3	aplikasi ini sangat cocok untuk mengedit	aplikasi ini sangat cocok untuk mengedit	aplikasi ini sangat cocok untuk mengedit	aplikasi ini sangat cocok untuk mengedit	aplikasi ini sangat cocok untuk mengedit

### Perbandingan Kinerja Antar Skenario

Perbandingan kinerja model IndoBERT pada dua pendekatan evaluasi yang berbeda untuk melihat perbedaan performa model terhadap pembagian data yang digunakan, sebagaimana ditunjukkan pada Tabel 3.

**Tabel 3.** Hasil Evaluasi IndoBERT pada Skema Single Train-Test Split (80/20) dan Stratified 5-Fold Cross-Validation

Model		Accuracy	Precision	Recall	F1-Score	ROC-AUC
IndoBERT ( <i>base</i> )	Negatif	0,756	0,771	0,742	0,756	0,910
	Netral	0,756	0,206	0,435	0,280	0,773
	Positif	0,756	0,905	0,801	0,850	0,918
IndoBERT ( <i>cross-validation</i> )	Negatif	0,740	0,776	0,687	0,729	0,904
	Netral	0,740	0,198	0,496	0,283	0,773
	Positif	0,740	0,908	0,798	0,850	0,919

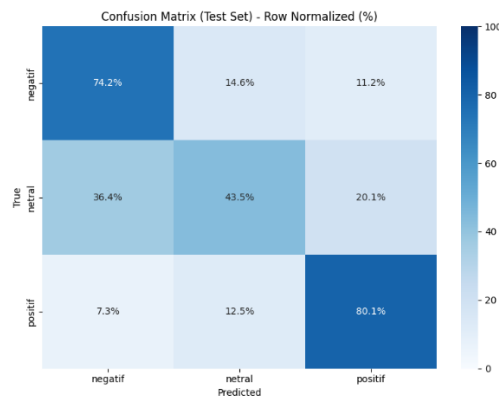
Berdasarkan Tabel 3, IndoBERT (*base*) memperoleh accuracy 0,756, sedangkan IndoBERT (*cross-validation*) menghasilkan accuracy rata-rata 0,740. Secara per kelas, performa terbaik konsisten pada kelas positif (F1-score 0,850; ROC-AUC 0,918–0,919) yang menunjukkan kemampuan model mengenali sentimen positif cukup kuat. Kelas negatif juga menunjukkan performa baik, meskipun terjadi penurunan pada skema cross-validation. Sebaliknya, kelas netral menjadi yang paling menantang pada kedua skema (precision sekitar 0,20; F1-score sekitar 0,28), yang mengindikasikan adanya tumpang tindih karakteristik bahasa antara netral dengan dua kelas lainnya. Walaupun single split menghasilkan accuracy lebih

tinggi, *cross-validation* lebih direkomendasikan untuk pelaporan karena memberikan evaluasi yang lebih representatif melalui beberapa pembagian data.

### **Analisis Confusion Matrix**

*Confusion matrix* digunakan untuk menganalisis pola klasifikasi dan sumber kesalahan model pada tiap kategori sentimen. Pada skenario base, confusion matrix dihitung dari 20% subset data uji. Sementara pada skenario cross-validation, confusion matrix dibentuk dari penggabungan prediksi data uji di semua fold, sehingga evaluasinya lebih menyeluruh.

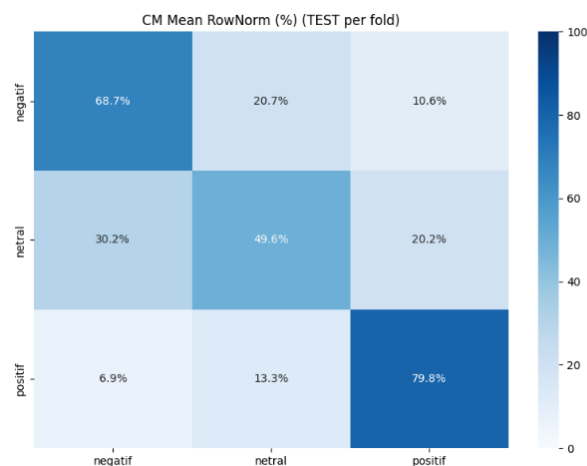
#### **IndoBERT(base)**



**Gambar 2.** Confusion Matrix IndoBERT (*base*)

Gambar 2 menunjukkan confusion matrix IndoBERT (*base*) tanpa cross validation pada data uji, di mana model cukup baik mengenali kelas negatif (74,2%) dan terutama positif (80,1%), tetapi masih lemah pada kelas netral (43,5%) yang sering tertukar menjadi negatif (36,4%) atau positif (20,1%). Secara umum, model lebih mudah membedakan sentimen positif–negatif dibanding netral.

#### **IndoBERT (cross validation)**



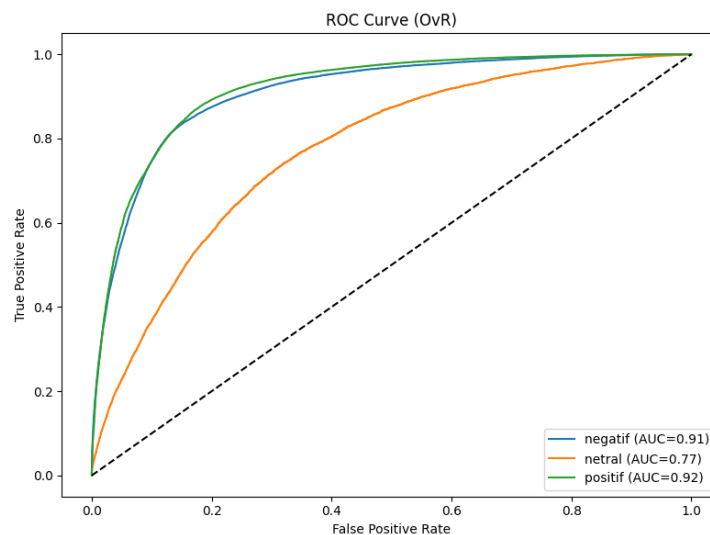
**Gambar 3.** Confusion Matrix IndoBERT (*cross-validation*)

Gambar 3, menunjukkan confusion matrix model IndoBERT dengan cross validation. menunjukkan model paling baik mengenali kelas positif (79,8%) dan cukup baik pada kelas negatif (68,7%), sedangkan kelas netral masih paling sulit (49,6%) karena sering tertukar menjadi negatif (30,2%) atau positif (20,2%). Secara umum, model lebih mudah membedakan sentimen positif–negatif dibanding netral.

## Analisis ROC Curve

### *IndoBERT (base)*

Gambar 4, menampilkan kurva Receiver Operating Characteristic (ROC) dari model IndoBERT tanpa *cross validation* yang menggambarkan hubungan antara True Positive Rate dan False Positive Rate pada berbagai nilai ambang klasifikasi.

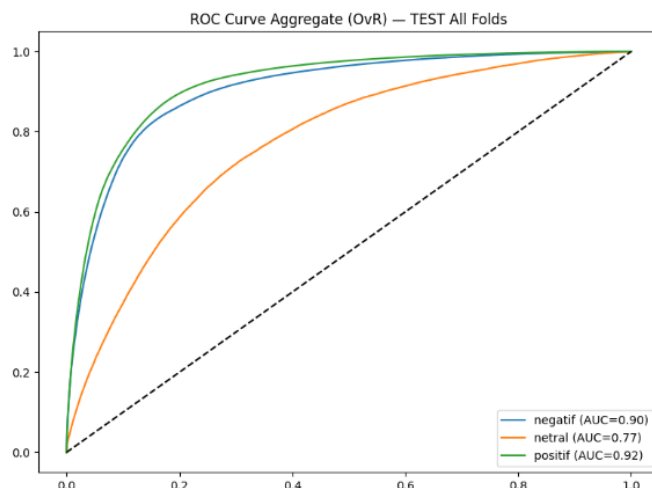


**Gambar 4.** ROC Curve IndoBERT(*base*)

Berdasarkan Gambar 4 menunjukkan bahwa IndoBERT memiliki kemampuan diskriminatif yang baik dalam memisahkan kelas sentimen pada data pengujian, sehingga evaluasi tidak hanya bergantung pada akurasi tetapi juga pada kualitas pemisahan kelas di berbagai ambang keputusan.

### *IndoBERT (cross validation)*

Gambar 5, menampilkan kurva Receiver Operating Characteristic (ROC) dari model IndoBERT dengan *cross validation* yang menggambarkan hubungan antara True Positive Rate dan False Positive Rate pada berbagai nilai ambang klasifikasi.



**Gambar 5.** ROC Curve IndoBERT(*cross-validation*)

Pada Gambar 5 menunjukkan bahwa model memiliki performa yang lebih kuat dalam membedakan sentimen positif dan negatif, sedangkan kemampuan pemisahan untuk kelas netral cenderung lebih rendah dibanding dua kelas lainnya.

## 5. KESIMPULAN DAN SARAN

Penelitian ini membandingkan kinerja IndoBERT Base P1 pada dua skema evaluasi, yaitu single train–test split (80/20) dan stratified 5-fold cross-validation (k=5), dengan alur kerja mencakup pengumpulan ulasan CapCut, pelabelan berbasis rating, pra-pemrosesan, tokenisasi, hingga fine-tuning. Hasil menunjukkan skema single split menghasilkan accuracy total 0,756, sedangkan cross-validation menghasilkan accuracy rata-rata 0,740. Walaupun pengujian tunggal memberikan nilai lebih tinggi, cross-validation dinilai lebih representatif karena menghasilkan estimasi performa yang lebih representatif dan tidak bergantung pada satu kali pembagian data. Ditinjau per kelas, performa terbaik konsisten pada kelas positif (F1-score 0,850; ROC-AUC 0,918–0,919) dan cukup baik pada kelas negatif, sementara kelas netral masih paling menantang, tercermin dari precision sekitar 0,198–0,206 dan F1-score sekitar 0,280–0,283 sehingga sering tumpang tindih dengan dua kelas lainnya. Keterbatasan penelitian ini meliputi pelabelan berbasis rating yang berpotensi menimbulkan noise label, penggunaan undersampling yang dapat mengurangi variasi informasi pada data latih, serta cakupan data yang terbatas pada ulasan CapCut di Google Play Store sehingga generalisasi lintas platform maupun lintas periode belum teruji. Ke depan, disarankan meningkatkan kualitas label melalui validasi manual atau semiautomatis, mengeksplor penanganan ketidakseimbangan kelas selain undersampling (misalnya *class-weight* atau augmentasi data),

memfokuskan perbaikan pada kelas netral, serta membandingkan dengan model Transformer lain dan memperluas sumber data agar temuan lebih general dan robust.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing atas arahan, masukan, dan bimbingan yang diberikan selama pelaksanaan penelitian hingga penyusunan artikel ini. Penulis juga menyampaikan apresiasi kepada sivitas akademika yang telah menyediakan dukungan keilmuan serta fasilitas penelitian serta kepada pihak-pihak yang membantu dalam proses pengumpulan data serta penelaahan naskah.

## DAFTAR REFERENSI

- Abidin, D. Z., Afuan, L., & Toscany, A. N. (2025). *A Comprehensive Benchmarking Pipeline for Transformer-Based Sentiment Analysis using Cross-Validated Metrics*. 6(4).
- Apriyadi, C., & Styawati, S. (2025). Sentiment Analysis of Cyber Attacks in Bank Syariah Indonesia Using SVM and Indobert Method. *Jurnal Teknik Informatika (Jutif)*, 6(2), 819–838. <https://doi.org/10.52436/1.jutif.2025.6.2.2636>
- Asri, Y., Kuswardani, D., Ramadhana, S. A., TS, J. F. P., Marbun, D. U. N., Fatimah, F. N., & Qoriza, Z. (2025). *OPTIMALISASI ANALISIS SENTIMEN DENGAN SPELLING CORRECTOR*. Uwais Inspirasi Indonesia. <https://books.google.co.id/books?id=WltnEQAAQBAJ>
- Baihaqi, W. M., & Munandar, A. (2023). Sentiment Analysis of Student Comment on the College Performance Evaluation Questionnaire Using Naïve Bayes and IndoBERT. *JUITA : Jurnal Informatika*, 11(2), 213. <https://doi.org/10.30595/juita.v11i2.17336>
- CapCut Video Editor on Google Play*. (n.d.). <https://play.google.com/store/apps/details?id=com.lemon.lvoverseas&hl=id>
- Dąbrowski, J., Letier, E., Perini, A., & Susi, A. (2022). Analysing app reviews for software engineering: A systematic literature review. *Empirical Software Engineering*, 27(2). <https://doi.org/10.1007/s10664-021-10065-7>
- Ferdous, S. M., Newaz, S. N. E., Mugdha, S. B. S., & Uddin, M. (2024). Sentiment Analysis in the Transformative Era of Machine Learning: A Comprehensive Review. *Statistics, Optimization & Information Computing*, 13(1), 331–346. <https://doi.org/10.19139/soic-2310-5070-2113>
- Ferrari, A., & Ginde, G. (2025). *Handbook on Natural Language Processing for Requirements Engineering*. Springer Nature Switzerland. <https://books.google.co.id/books?id=vaVMEQAAQBAJ>
- Hua, H., Li, X., Dou, D., Xu, C., & Luo, J. (2021). Noise Stability Regularization for Improving BERT Fine-tuning. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3229–3241. <https://doi.org/10.18653/v1/2021.naacl-main.258>
- Kumar, L. A., & Renuka, D. K. (2023). *Deep Learning Approach for Natural Language Processing, Speech, and Computer Vision: Techniques and Use Cases*. CRC Press. <https://books.google.co.id/books?id=HmqwEAAAQBAJ>

- Liu, B. (2020). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press. <https://books.google.co.id/books?id=PdX7DwAAQBAJ>
- Mola, S. A. S., Djawa, S. N. R., & Mauko, A. Y. (2025). *Text Mining: Analisis Sentimen dengan Naïve Bayes*. Kaizen Media Publishing. <https://books.google.co.id/books?id=qrxNEQAAQBAJ>
- Srivastava, G., & Lin, C. W. (2025). Guest Editorial Special Section on Fuzzy-Deep Neural Network Learning in Sentiment Analysis. In *IEEE Transactions on Fuzzy Systems* (Vol. 33, Issue 1, pp. 1–2). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/TFUZZ.2024.3520662>
- Sudhir, P., & Suresh, V. D. (2021). Comparative study of various approaches, applications and classifiers for sentiment analysis. *Global Transitions Proceedings*, 2(2), 205–211. <https://doi.org/10.1016/j.gltip.2021.08.004>
- Suhadi, E., Kunci, K., Ekspor, B., & Analisis, S. (2025). ANALISIS SENTIMEN APLIKASI BISA EKSPOR PADA ULASAN PENGGUNA DI GOOGLE PLAY DENGAN NAÏVE BAYES. In *JIKA* (No. 1; Vol. 9, pp. 2722–2713).
- Tang, T., Tang, X., & Yuan, T. (2020). Fine-Tuning BERT for Multi-Label Sentiment Analysis in Unbalanced Code-Switching Text. *IEEE Access*, 8, 193248–193256. <https://doi.org/10.1109/ACCESS.2020.3030468>
- Tao, J., & Fang, X. (2020). Toward multi-label sentiment analysis: A transfer learning based approach. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-019-0278-0>