

KOMPARASI MODEL DEEP LEARNING PADA ULASAN APLIKASI DRAMABOX DI GOOGLE PLAY STORE

Adam Idhofi Rakasiwi

Universitas Pembangunan Nasional “Veteran” Jawa Timur
Correspondence author email: 22082010026@student.upnjatim.ac.id

Arsa Cahaya Pradipta

Universitas Pembangunan Nasional “Veteran” Jawa Timur
22082010015@student.upnjatim.ac.id

Al-Faiz Azzam Aryaputra

Universitas Pembangunan Nasional “Veteran” Jawa Timur
22082010022@student.upnjatim.ac.id

Abstract

The rapid growth of streaming platforms and the increasing popularity of micro-drama have led to the emergence of applications such as DramaBox, which generate a large number of user reviews as a source of public opinion. However, sentiment analysis on streaming application reviews is still limited, particularly in comparing the performance of deep learning models. This study aims to compare the performance of Bi-LSTM and IndoBERT models in conducting sentiment analysis on user reviews of the DramaBox application. The data were collected through web scraping from the Google Play Store, resulting in a total of 13,701 user reviews. The data then underwent preprocessing before being divided into data train and data test. Model evaluation was carried out using accuracy, precision, recall, F1-score, confusion matrix, and 5-fold cross-validation. The results show that the IndoBERT model achieved better performance with an accuracy of 0.90 compared to 0.87 for Bi-LSTM, and also demonstrated higher stability based on cross-validation results. Therefore, IndoBERT is considered more effective in understanding the context of the Indonesian language and producing more accurate sentiment analysis on the DramaBox user review dataset.

Keywords: Sentiment Analysis, Deep Learning, Bi-LSTM, IndoBERT, DramaBox.

Abstrak

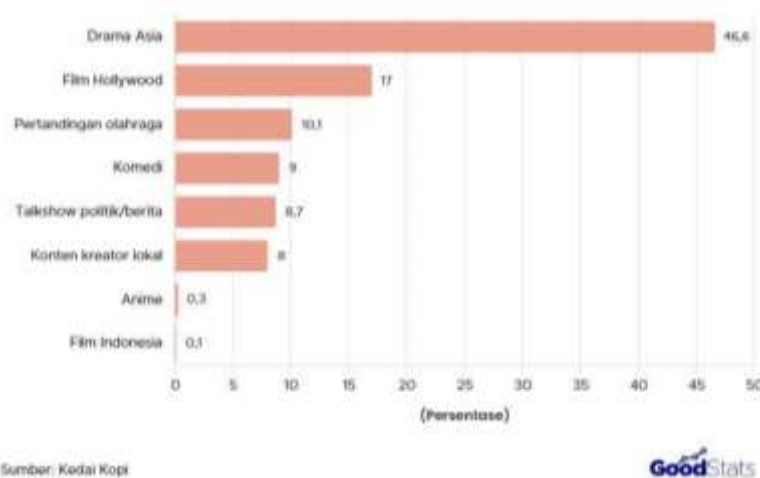
Perkembangan platform streaming dan meningkatnya popularitas micro-drama mendorong munculnya berbagai aplikasi seperti DramaBox yang menghasilkan banyak ulasan pengguna sebagai sumber opini publik. Namun, analisis sentimen terhadap ulasan aplikasi streaming masih terbatas, khususnya dalam membandingkan performa model deep learning. Penelitian ini bertujuan untuk membandingkan kinerja model Bi-LSTM dan IndoBERT dalam melakukan analisis sentimen terhadap ulasan pengguna aplikasi DramaBox. Data diperoleh melalui web scraping dari Google Play Store sebanyak 13.701 ulasan pengguna. Data kemudian melalui tahap pra pemrosesan sebelum dibagi menjadi data latih dan

data uji. Evaluasi model dilakukan menggunakan accuracy, precision, recall, F1-score, confusion matrix, dan 5-fold cross-validation. Hasil penelitian menunjukkan bahwa model IndoBERT memperoleh performa lebih baik dengan accuracy sebesar 0.90 dibandingkan Bi-LSTM sebesar 0.87, serta memiliki stabilitas yang lebih tinggi berdasarkan hasil cross-validation. Dengan demikian, IndoBERT dinilai lebih efektif dalam memahami konteks bahasa Indonesia dan menghasilkan analisis sentimen yang lebih akurat pada dataset ulasan aplikasi DramaBox.

Kata Kunci : Analisis Sentimen, Deep Learning, Bi-LSTM, IndoBERT, DramaBox

PENDAHULUAN

Perkembangan teknologi digital dan *platform streaming* telah mengubah pola konsumsi hiburan masyarakat secara signifikan, terutama melalui layanan Video on Demand (VoD) yang memungkinkan akses konten kapan saja dan di mana saja. Penelitian menunjukkan bahwa algoritma rekomendasi dan *platform streaming* berperan besar dalam membentuk pola konsumsi media global (Lee et al., 2025). Studi oleh Lee et al. (2025) yang menganalisis data Netflix di 71 negara menemukan bahwa preferensi konten dipengaruhi faktor geografis dan budaya, namun tetap didorong oleh distribusi digital global. Fenomena ini mendorong meningkatnya popularitas drama Asia secara internasional serta memicu persaingan antar negara dalam menghasilkan konten yang inovatif.



Gambar 1. Statistik Jenis Tayangan yang Paling Sering Ditonton Masyarakat Indonesia Tahun 2025

Popularitas drama Asia di Indonesia dapat dilihat dari berbagai laporan survei mengenai preferensi hiburan masyarakat. Berdasarkan data GoodStats, 46.6% masyarakat Indonesia menjadikan drama Asia sebagai jenis tayangan yang paling sering ditonton, diikuti film Hollywood sebesar 17% dan pertandingan olahraga sebesar 10.1% (Shahibah, 2025). Selain itu laporan GoodStats lainnya menunjukkan bahwa film dan serial dari Korea Selatan menjadi konten paling favorit bagi masyarakat Indonesia dengan persentase sekitar 64%, diikuti oleh konten dari negara lain seperti China dan

Jepang (Wafa, 2026). Temuan tersebut menunjukkan bahwa drama Asia memiliki pengaruh besar terhadap pola konsumsi hiburan masyarakat Indonesia.

Selain drama Korea, drama China juga menunjukkan peningkatan popularitas yang signifikan seiring berkembangnya distribusi konten digital melalui platform *streaming* dan media sosial. Studi menunjukkan bahwa sekitar 50.4% pengguna aplikasi video pendek di China telah menonton *micro-drama* berdurasi kurang dari tiga menit, yang mencerminkan tingginya konsumsi format tersebut (Li, 2024). Selain itu, sebanyak 74.19% responden telah mengonsumsi *micro-drama* selama lebih dari dua tahun melalui platform seperti Tencent Video, iQIYI, dan Douyin (Li, 2024). Karakteristik *micro-drama* yang berdurasi singkat dengan alur cepat dan konflik dramatis menjadikannya efektif dalam menarik perhatian pengguna *mobile* serta memperluas jangkauan audiens global (Hanney, 2025).



Gambar 2. Aplikasi DramaBox

Pertumbuhan industri *micro-drama* juga mendorong munculnya berbagai aplikasi *streaming*, salah satunya DramaBox yang menyediakan ribuan episode drama pendek dari berbagai *genre* dalam format video vertikal untuk konsumsi *mobile*. Aplikasi ini diluncurkan pada tahun 2023 dan dalam waktu singkat telah tersedia di lebih dari 200 negara dengan lebih dari 2000 judul drama pendek (GetNews, 2025). Selain itu, laporan Sensor Tower yang dikutip Business Insider, menyebutkan bahwa DramaBox menghasilkan sekitar \$120 juta pendapatan global pada kuartal pertama 2025, yang menunjukkan tingginya minat pengguna terhadap format hiburan ini (Moses, 2026). Dengan meningkatnya jumlah pengguna, ulasan pada platform seperti Google Play Store juga semakin banyak menjadi sumber data opini publik yang penting untuk dianalisis.

Banyaknya ulasan pengguna pada aplikasi *streaming* seperti DramaBox menghasilkan teks yang sangat besar dan tidak terstruktur sehingga sulit dianalisis secara manual. Oleh karena itu, pendekatan NLP digunakan untuk mengekstraksi informasi dari teks ulasan pengguna, salah satunya melalui teknik analisis sentimen yang

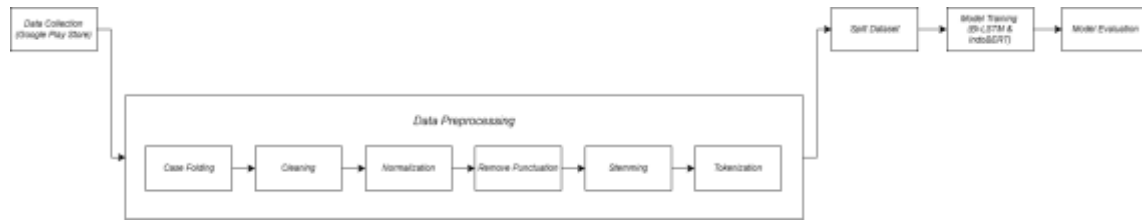
bertujuan mengidentifikasi opini pengguna dalam bentuk sentimen positif, negatif, atau netral (Setiawan, 2025). Salah satu algoritma *deep learning* yang sering digunakan dalam analisis sentimen adalah Bidirectional Long Short-Term Memory (Bi-LSTM) karena mampu mempelajari hubungan konteks antar kata dalam urutan teks secara dua arah sehingga dapat memahami makna kalimat secara lebih komprehensif (Nasution et al., 2025). Dengan kemampuan tersebut, Bi-LSTM menjadi salah satu metode *deep learning* yang banyak digunakan dalam penelitian analisis sentimen berbasis teks.

Selain Bi-LSTM, model berbasis *transformer* seperti BERT juga banyak digunakan dalam analisis sentimen karena mampu memahami konteks bahasa secara lebih kompleks melalui mekanisme *self-attention*. IndoBERT sebagai varian BERT yang dilatih khusus menggunakan korpus bahasa Indonesia terbukti mampu memahami struktur linguistik secara lebih baik dibandingkan model multibahasa (Nugroho et al., 2021). Penelitian sebelumnya menunjukkan bahwa IndoBERT dan Bi-LSTM sama-sama mampu menghasilkan performa klasifikasi yang tinggi pada berbagai dataset ulasan pengguna (Farizi & Sibaroni, 2025). Namun, penelitian yang secara khusus membandingkan kedua model pada ulasan aplikasi *streaming* masih terbatas, sehingga penelitian ini dilakukan untuk mengetahui model yang paling efektif dalam menganalisis sentimen pengguna aplikasi DramaBox.

METODE PENELITIAN

Penelitian ini menggunakan jenis penelitian kuantitatif dengan pendekatan eksperimental komputasional, yang bertujuan membandingkan kinerja model *deep learning* IndoBERT dan Bi-LSTM dalam analisis sentimen ulasan aplikasi DramaBox di Google Play Store. Pendekatan ini dipilih karena memungkinkan pengukuran objektif performa model melalui metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Penelitian sebelumnya menunjukkan bahwa model *deep learning* seperti Bi-LSTM dan IndoBERT mampu memahami konteks bahasa dalam teks dengan lebih baik dibandingkan metode *machine learning* karena mampu menangkap hubungan semantik antar kata dalam suatu kalimat (Pramana et al., 2025).

Pengumpulan data pada penelitian ini dilakukan menggunakan teknik *web scraping* untuk mengambil ulasan pengguna aplikasi DramaBox dari Google Play Store. Proses *scraping* dilakukan menggunakan bahasa pemrograman Python dengan memanfaatkan *library* Google-Play-Scraper. Data yang diambil meliputi teks ulasan pengguna, rating yang diberikan, tanggal ulasan, serta informasi lain yang tersedia pada halaman ulasan aplikasi. Data yang dikumpulkan kemudian digunakan dalam proses analisis sentimen menggunakan model *deep learning*.



Gambar 3. Alur Penelitian

Gambar 3 merupakan diagram alur penelitian yang dimulai dari tahap pengumpulan data ulasan aplikasi DramaBox. Data yang diperoleh kemudian melalui tahap *preprocessing* yang meliputi *case folding*, *cleaning*, *normalization*, *remove punctuation*, *stemming*, dan *tokenization* sebelum dilakukan pembagian dataset menjadi data *train* dan data *test*. Proses ini bertujuan untuk mengurangi *noise* dalam teks serta menyeragamkan bentuk kata agar lebih mudah diproses oleh algoritma *machine learning*. Selanjutnya dataset digunakan dalam proses pelatihan model *deep learning* menggunakan algoritma Bi-LSTM dan IndoBERT, kemudian performa model dievaluasi menggunakan *confusion matrix* dengan *accuracy*, *precision*, *recall*, dan *F1-score* untuk menentukan model yang paling efektif.

Tahap analisis data dalam penelitian ini dilakukan dengan melatih model Bi-LSTM dan IndoBERT menggunakan dataset ulasan pengguna yang telah melalui proses *preprocessing*. Model Bi-LSTM merupakan pengembangan dari Recurrent Neural Network (RNN) yang mampu memproses teks secara dua arah sehingga dapat memahami hubungan konteks antar kata dalam suatu kalimat. Sementara IndoBERT merupakan model berbasis *transformer* yang telah dilatih menggunakan korpus bahasa Indonesia dalam jumlah besar sehingga mampu memahami konteks bahasa Indonesia secara lebih baik.

Evaluasi performa model dalam penelitian ini menggunakan metode *confusion matrix*. *Confusion Matrix* terdiri dari empat komponen utama yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Berdasarkan nilai tersebut, beberapa metrik evaluasi dapat dihitung untuk menilai kinerja model seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Rumus perhitungannya adalah sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Keterangan:

- TP = Kasus positif berhasil diprediksi benar oleh model

- TN = Kasus negatif berhasil diprediksi benar oleh model
- FP = Kasus negatif salah diprediksi sebagai positif
- FN = Kasus positif salah diprediksi sebagai negatif

Nilai metrik tersebut kemudian digunakan untuk membandingkan performa model Bi-LSTM dan IndoBERT guna menentukan model yang paling efektif dalam menganalisis sentimen ulasan pengguna aplikasi DramaBox.

HASIL DAN PEMBAHASAN

1. Data Collection

Pada tahap *data collection*, data ulasan pengguna aplikasi DramaBox berhasil dikumpulkan dari Google Play Store menggunakan teknik *web scraping* berbasis Python. Proses *scraping* dilakukan dengan memanfaatkan *library* google-play-Scraper pada platform Google Colab. Melalui proses tersebut, sistem berhasil memperoleh informasi yang tersedia pada halaman ulasan aplikasi, seperti isi ulasan, rating yang diberikan, tanggal ulasan, serta *link* ulasan. Contoh data ulasan yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1 berikut.

Tabel 1. Data Ulasan Aplikasi DramaBox

review	rating	date	review_link
“aplikasi najis..baru banget download aplikasi ini, taunya buka episode aja harus nonton iklan 20detik..sangat buang waktu..”	1	18/02/2026 21:27:22	https://play.google.com/store/apps/details?id=com.storymatrix.drama&reviewId=e7d11aco-78do-484d-89d3-23eb8073b05c
“bgus bnget aku kasih lima bintang”	5	06/11/2025 9:18:01	https://play.google.com/store/apps/details?id=com.storymatrix.drama&reviewId=c6ef44d3-4864-469b-adfo-of3b183c7ac4

Berdasarkan proses pengumpulan data yang dilakukan pada periode Oktober 2025 hingga Maret 2026, penelitian ini berhasil memperoleh sebanyak 13.701 ulasan pengguna. Seluruh data tersebut kemudian dikumpulkan dan disimpan dalam bentuk dataset dengan format CSV agar memudahkan proses pengolahan data pada tahap *data preprocessing*.

2. Data Preprocessing

Tahap *data preprocessing* dilakukan untuk menyiapkan data ulasan agar dapat diproses secara optimal oleh model *deep learning*. Data ulasan dari Google Play Store

umumnya masih berupa teks tidak terstruktur yang mengandung berbagai *noise* seperti variasi penulisan huruf, simbol dan kata tidak baku. Oleh karena itu, diperlukan proses pembersihan dan standarisasi agar teks menjadi lebih terstruktur dan mudah dipahami oleh model. Proses ini bertujuan mengubah teks mentah menjadi format yang lebih konsisten sehingga dapat diproses secara efektif oleh algoritma *deep learning*. Berikut adalah beberapa tahap yang dapat dilihat pada tabel 2, yaitu:

Tabel 2. Tahapan Data Preprocessing

Langkah	Penjelasan	Teks Sebelum	Teks Sesudah
Preprocessing			
Case Folding	Mengubah seluruh huruf dalam teks menjadi huruf kecil.	kebanyakan iklan itupun dikunci semua kan harus nonton iklan lamaa sampe bosen di apk sebelah minim iklan, BANYAK IKLAN LAMA LAGI 🤔🤔🤔🤔🤔	kebanyakan iklan itupun dikunci semua kan harus nonton iklan lamaa sampe bosen di apk sebelah minim iklan, banyak iklan lama lagi 🤔🤔🤔🤔🤔
Cleaning	Menghapus elemen yang tidak diperlukan seperti URL, angka, atau karakter khusus dalam teks.	kebanyakan iklan itupun dikunci semua kan harus nonton iklan lamaa sampe bosen di apk sebelah minim iklan, banyak iklan lama lagi 🤔🤔🤔🤔🤔	kebanyakan iklan itupun dikunci semua kan harus nonton iklan lamaa sampe bosen di apk sebelah minim iklan, banyak iklan lama lagi
Normalization	Mengubah kata tidak baku atau slang menjadi bentuk kata baku.	kebanyakan iklan itupun dikunci semua kan harus nonton iklan lamaa sampe bosen di apk sebelah minim iklan, banyak iklan lama lagi	kebanyakan iklan itupun dikunci semua kan harus nonton iklan lama sampe bosen di aplikasi sebelah minim iklan, banyak iklan lama lagi
Remove Punctuation	Menghapus tanda baca seperti titik, koma, tanda seru, dan simbol lainnya yang tidak diperlukan.	kebanyakan iklan itupun dikunci semua kan harus nonton iklan lama sampe bosen di aplikasi sebelah minim iklan, banyak iklan lama lagi	kebanyakan iklan itupun dikunci semua kan harus nonton iklan lama sampe bosen di aplikasi sebelah minim iklan banyak iklan lama lagi

Stemming	Mengubah kata berimbuhan menjadi bentuk kata dasar.	kebanyakan iklan banyak iklan itu kunci itupun dikunci semua kan harus nonton iklan lama sampe bosan di aplikasi sebelah minim iklan banyak iklan lagi	banyak iklan itu kunci semua kan harus nonton iklan lama sampe bosan di aplikasi sebelah minim iklan banyak iklan lama lagi
Tokenization	Memecah kalimat menjadi unit kata atau token.	banyak iklan itu kunci semua kan harus nonton iklan lama sampe bosan di aplikasi sebelah minim iklan banyak iklan lama lagi	['banyak', 'iklan', 'itu', 'kunci', 'semua', 'kan', 'harus', 'nonton', 'iklan', 'lama', 'sampe', 'bosan', 'di', 'aplikasi', 'sebelah', 'minim', 'iklan', 'banyak', 'iklan', 'lama', 'lagi']

Tabel 2 tersebut memberikan gambaran bagaimana setiap tahapan *preprocessing* berkontribusi dalam membersihkan dan menstandarisasi data teks sehingga siap digunakan dalam proses analisis. Selanjutnya, dataset yang telah diproses akan memasuki tahap data splitting untuk membagi data menjadi data latih dan data uji sebelum proses pembuatan model dilakukan.

3. Split Dataset

```

from sklearn.model_selection import train_test_split

X = df['stemmed']
y = df['label']

X_train, X_test, y_train, y_test = train_test_split(
    X, y,
    test_size=0.2,
    random_state=42,
    stratify=y
)

```

Gambar 4. Kode Split Dataset

Tahapan *split dataset* dapat dilihat pada Gambar 4 dilakukan untuk membagi data menjadi data latih dan data uji. Pada tahap ini menggunakan fungsi *train_test_split* dari *library scikit_learn*, dimana variabel X berisi data teks hasil *preprocessing* pada kolom *stemmed* dan variabel y berisi label sentimen. Dataset kemudian dibagi menjadi 80% data latih dan 20% data uji menggunakan parameter *test_size=0.2*. Parameter *random_state=42* digunakan agar proses pembagian data tetap konsisten setiap kali

kode dijalankan, sedangkan *stratify=y* bertujuan menjaga proporsi setiap label tetap seimbang pada data latih dan data uji.

4. Model Training (Bi-LSTM & IndoBERT)

4.1. Bi-LSTM

Pada tahap ini dilakukan proses pelatihan model menggunakan arsitektur Bidirectional Long Short-Term Memory (Bi-LSTM) untuk analisis sentimen pada ulasan pengguna aplikasi DramaBox. Rincian struktur arsitektur model Bi-LSTM yang digunakan dalam penelitian ini dapat dilihat pada Tabel 3 berikut.

Tabel 3. Model Bi-LSTM

Layer	Output Shape	Parameter Count
Embedding	(None, 100, 128)	2,560,000
Bidirectional LSTM	(None, 128)	98,816
Dropout	(None, 128)	0
Dense	(None, 32)	4,128
Dense (Output)	(None, 3)	99
Total		2,663,043

Berdasarkan Tabel 3, model Bi-LSTM yang digunakan terdiri dari beberapa lapisan utama yaitu *embedding layer*, *bidirectional LSTM*, *dropout*, serta *dense layer* sebagai lapisan klasifikasi. *Embedding layer* mengubah teks menjadi representasi vektor numerik berdimensi 128 sehingga dapat diproses oleh model *deep learning*. Selanjutnya lapisan *bidirectional LSTM* mengekstraksi informasi kontekstual dari urutan teks sebelum diteruskan ke *dense layer*. Lapisan output menggunakan fungsi aktivasi *softmax* untuk menghasilkan probabilitas klasifikasi terhadap tiga kelas sentimen yaitu positif, negatif, dan netral.

4.2. IndoBERT

Tahapan ini dimulai dengan proses pelatihan model IndoBERT untuk melakukan analisis sentimen pada ulasan pengguna aplikasi DramaBox. Rincian arsitektur model IndoBERT yang digunakan dalam penelitian ini dapat dilihat pada Tabel 4 berikut.

Tabel 4. Model IndoBERT

Layer	Output Shape	Parameter Count	Connected To
Input IDs	(None, 128)	0	-
Attention Mask	(None, 128)	0	-
Token Embedding	(None, 128)	~23M	Input IDs
Positional Embedding	(None, 128, 768)	~0.4M	Token Embedding

Transformer Encoder (12 layers)	(None, 128, 768)	~85M	Embedding Layer
Pooler (CLS representation)	(None, 128, 768)	~590K	Transformer Encoder
Dropout	(None, 768)	0	Pooler
Classification Dense	(None, 3)	2,307	Dropout
Total		124,000,000	

Berdasarkan Tabel 4, model IndoBERT menerima input berupa *input IDs* dan *attention mask* yang kemudian diproses melalui lapisan *embedding* untuk merepresentasikan token ke dalam bentuk vektor numerik. Selanjutnya representasi tersebut diproses oleh 12 lapisan *transformer encoder* untuk mengekstraksi informasi kontekstual dari seluruh token dalam kalimat. Hasil representasi dari token [CLS] kemudian diteruskan ke lapisan *pooler* dan *dropout* sebelum masuk ke lapisan klasifikasi. Lapisan *classification dense* dengan fungsi aktivasi *softmax* menghasilkan probabilitas klasifikasi terhadap tiga kelas sentimen yaitu positif, negatif, dan netral.

5. Model Evaluation

Evaluasi model dilakukan untuk mengukur kemampuan tiap model. Proses evaluasi dilakukan menggunakan beberapa metrik yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, serta analisis *confusion matrix* untuk melihat distribusi kesalahan prediksi pada setiap kelas. Penggunaan metrik evaluasi ini penting karena setiap metrik menggambarkan aspek performa model yang berbeda.

Tabel 5. Perbandingan Kinerja Model

Model	Label	Precision	Recall	F1-score	Accuracy
Bi-LSTM	negatif	0.90	0.94	0.92	0.87
	netral	0.75	0.60	0.67	
	positif	0.81	0.87	0.84	
IndoBERT	negatif	0.95	0.93	0.94	0.90
	netral	0.73	0.81	0.77	
	positif	0.96	0.88	0.92	

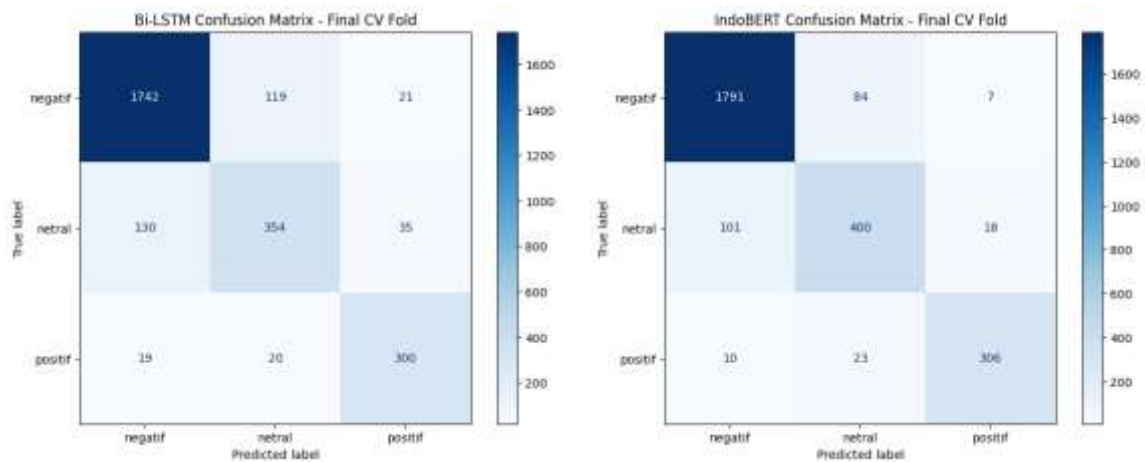
Berdasarkan Tabel 5, model Bi-LSTM memperoleh nilai *accuracy* sebesar 0.87, sedangkan model IndoBERT memperoleh *accuracy* sebesar 0.90. Pada model Bi-LSTM, performa terbaik terdapat pada kelas negatif dengan nilai *F1-score* 0.92, sementara kelas netral memiliki performa paling rendah dengan *F1-score* 0.67. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam membedakan ulasan netral dengan

sentimen lainnya. Sementara itu, model IndoBERT menunjukkan performa yang lebih baik pada hampir seluruh kelas dengan nilai *F1-score* 0.94 pada kelas negatif, 0.77 pada kelas netral, dan 0.92 pada kelas positif. Secara keseluruhan, hasil tersebut menunjukkan bahwa model IndoBERT mampu memberikan performa yang lebih baik dibandingkan model Bi-LSTM pada dataset ulasan pengguna aplikasi DramaBox.

Tabel 6. Cross Validation

Model	Fold Accuracy					Avg ± SD
	1	2	3	4	5	Acc.
Bi-LSTM	87%	87%	88%	87%	87%	87% ± 0.55%
IndoBERT	91%	90%	91%	90%	91%	90% ± 0.22%

Tabel 6 di atas merupakan hasil evaluasi menggunakan *5-fold cross-validation* untuk menguji kestabilan performa model pada beberapa pembagian dataset. Metode ini membagi dataset menjadi beberapa bagian (*fold*) dimana setiap *fold* secara bergantian digunakan sebagai data uji sementara bagian lainnya digunakan sebagai data latih, kemudian hasil performa dari setiap *fold* dirata-ratakan untuk memperoleh estimasi kinerja model yang lebih stabil. Hasil pengujian menunjukkan bahwa model Bi-LSTM memperoleh rata-rata *accuracy* sebesar 87% dengan standar deviasi 0.55%, sedangkan model IndoBERT memperoleh rata-rata *accuracy* sebesar 90% dengan standar deviasi 0.22%. Nilai standar deviasi yang kecil menunjukkan bahwa performa kedua model relatif konsisten pada setiap *fold*, namun IndoBERT memiliki tingkat akurasi yang lebih tinggi dan variasi performa yang lebih kecil dibandingkan Bi-LSTM, sehingga dapat dikatakan memiliki kestabilan dan kemampuan generalisasi yang lebih baik pada dataset ulasan pengguna aplikasi DramaBox.



Gambar 5. Confusion Matrix

Analisis selanjutnya dilakukan menggunakan *confusion matrix* seperti yang tertera pada Gambar 5. *Confusion matrix* ini berfungsi untuk melihat distribusi prediksi benar dan kesalahan kinerja model pada setiap kelas sentimen. Berdasarkan hasil *confusion matrix*, model Bi-LSTM mampu memprediksi sebagian besar data terutama pada kelas negatif, namun masih terdapat beberapa kesalahan pada kelas netral yang diprediksi sebagai negatif maupun positif. Sementara itu, IndoBERT menunjukkan distribusi yang lebih dominan pada diagonal matriks dengan jumlah kesalahan yang lebih sedikit, sehingga menunjukkan performa model yang lebih baik dibandingkan model Bi-LSTM.

Analisis/Diskusi

Hasil pengolahan data menunjukkan bahwa ulasan pengguna DramaBox memiliki keragaman bahasa yang tinggi, ditandai dengan penggunaan kata tidak baku, singkatan, *typo*, dan ekspresi emosional. Hal ini umum pada data digital yang bersifat informal, sehingga tahap pra-pemrosesan seperti *case folding*, *cleaning*, *normalization*, *remove punctuation*, *stemming*, dan *tokenization* menjadi sangat penting untuk mengurangi *noise* dan meningkatkan kualitas representasi teks. Tanpa proses ini, model akan kesulitan mengenali pola bahasa yang konsisten dan performa klasifikasi dapat menurun

Perbedaan performa model dipengaruhi oleh arsitektur yang digunakan. Bi-LSTM mampu menangkap konteks lokal melalui pemrosesan sekuensial dua arah, namun memiliki keterbatasan dalam memahami hubungan kata jarak jauh. Sebaliknya, IndoBERT dengan mekanisme *self-attention* dapat memahami konteks global secara lebih efektif, terutama pada teks tidak baku dan kompleks. Selain itu, karakteristik data juga berpengaruh, di mana sentimen negatif lebih mudah dikenali dibandingkan sentimen netral yang cenderung ambigu. Evaluasi menggunakan *precision*, *recall*, dan *F1-score* memberikan gambaran performa yang lebih lengkap, sementara *cross-validation* menunjukkan konsistensi dan kemampuan generalisasi model. Analisis *confusion matrix* juga mengungkap bahwa kesalahan masih sering terjadi pada kelas netral.

Hasil ini menunjukkan bahwa pemilihan model, kualitas data, dan proses *preprocessing* merupakan faktor penting dalam analisis sentimen. Meskipun hasil penelitian cukup baik, terdapat keterbatasan seperti penggunaan dataset dari satu platform dan potensi bias pelabelan.

KESIMPULAN

Berdasarkan hasil yang telah dilakukan, kedua model *deep learning* yaitu Bi-LSTM dan IndoBERT mampu mengkategorikan sentimen ulasan pengguna aplikasi DramaBox dengan performa yang cukup baik. Model Bi-LSTM memperoleh nilai *accuracy* sebesar 0.87, sedangkan model IndoBERT memperoleh *accuracy* yang lebih tinggi yaitu 0.90.

Hasil evaluasi menggunakan *precision*, *recall*, dan *F1-score* juga menunjukkan bahwa IndoBERT memiliki performa yang lebih baik hampir pada seluruh kelas sentimen dibandingkan Bi-LSTM. Hal ini menunjukkan bahwa model berbasis *transformer* seperti BERT memiliki kemampuan untuk dapat memahami konteks bahasa yang lebih baik dibandingkan model berbasis RNN seperti Bi-LSTM

Hasil evaluasi menggunakan *cross-validation* dan *confusion matrix* juga menunjukkan bahwa kedua model memiliki performa yang cukup stabil. Model IndoBERT memiliki rata-rata *accuracy cross-validation* sebesar 90% dengan standar deviasi 0.22%, sedangkan Bi-LSTM memperoleh rata-rata *accuracy* sebesar 87% dengan standar deviasi 0.55%. Berdasarkan seluruh hasil evaluasi tersebut dapat disimpulkan bahwa model IndoBERT merupakan model yang paling optimal untuk melakukan analisis sentimen ulasan pengguna aplikasi DramaBox dalam penelitian ini.

DAFTAR PUSTAKA

- Farizi, A. F. A., & Sibaroni, Y. (2025). Implementation of BILSTM and IndoBERT for sentiment analysis of TikTok reviews. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 10(1), 96–106. <https://doi.org/10.29100/jipi.v10i1.5815>
- GetNews. (2025). DramaBox short dramas lead a new wave in video entertainment. *Barchart.com*. <https://www.barchart.com/story/news/30499772/dramabox-short-dramas-lead-a-new-wave-in-video-entertainment>
- Lee, N., Lim, J., Choi, M., & Jeong, H. (2025). Global streams, local currents: A data analysis on global VOD content consumption. *Journal of the Korean Physical Society*, 88(5), 606–616. <https://doi.org/10.1007/s40042-025-01520-5>
- Li, X. (2024). Understanding audience reception of Chinese internet “Cool dramas.” *SHS Web of Conferences*, 207, 02010. <https://doi.org/10.1051/shsconf/202420702010>
- Moses, L. (2026). *Disney-backed DramaBox is seeking new funding as it tries to win the micro drama race in the US*. Business Insider. Retrieved March 12, 2026, from <https://www.businessinsider.com/dramabox-seeks-new-funding-micro-drama-apps-gain-global-momentum-2026-1>
- Nasution, K., Saddami, K., Roslidar, R., Akhyar, A., Fathurrahman, F., & Aulia, N. (2025). Comparative study of BILSTM and GRU for sentiment analysis on Indonesian E-Commerce product reviews using deep sequential modeling. *Jurnal Teknik Informatika (Jutif)*, 6(4), 1881–1896. <https://doi.org/10.52436/1.jutif.2025.6.4.4878>
- Nugroho, K. S., Sukmadewa, A. Y., Dw, H. W., Bachtiar, F. A., & Yudistira, N. (2021). BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews. *Association for Computing Machinery*, 258–264. <https://doi.org/10.1145/3479645.3479679>
- Pramana, M. W. A., Putri, D. P. S., & Purnawan, I. K. A. (2025). Comparison of IndoBERT and Bi-LSTM models for Indonesian law violation text classification. *Jurnal Informatika Jurnal Pengembangan IT*, 10(4), 1033–1043. <https://doi.org/10.30591/jpit.v10i4.8795>

- Setiawan, B. (2025). A review of sentiment analysis applications in Indonesia between 2023-2024. *Journal of Information Engineering and Educational Technology*, 8(2), 71–83. <https://doi.org/10.26740/jieet.v8n2.p71-83>
- Shahibah, A. (2025). *Drama Asia Jadi Tontonan Favorit Publik RI 2025, Termasuk Drakor dan Dracin*. GoodStats Data. Retrieved March 12, 2026, from <https://data.goodstats.id/statistic/drama-asia-jadi-tontonan-favorit-publik-ri-2025-termasuk-drakor-dan-dracin-Zy2ke>
- Wafa, I. (2026). *Film & Series Asal Negara Mana yang Jadi Favorit Publik Indonesia?* GoodStats Data. Retrieved March 12, 2026, from <https://data.goodstats.id/statistic/film-series-asal-negara-mana-yang-jadi-favorit-publik-indonesia-VjjaX>