

## Perbandingan Performa CNN dan LSTM untuk Klasifikasi Sentimen Multibahasa pada *Review* Aplikasi Facebook

<sup>1</sup>Rajulul Anshar, <sup>2</sup>Zulfa Razi, <sup>3</sup>Ilal Mahdi  
<sup>1,2,3</sup>Teknik Informatika, Universitas Jabal Ghafur, Sigli

E-mail: <sup>1</sup>rajululanshar@gmail.com, <sup>2</sup>zulfarazihb@gmail.com,  
<sup>3</sup>ilalmahdi90@gmail.com

### ABSTRAK

Pesatnya pertumbuhan pengguna media sosial, khususnya Facebook, menghasilkan volume ulasan aplikasi yang sangat besar dan multibahasa, sehingga menimbulkan tantangan tersendiri dalam analisis sentimen, terutama akibat fenomena *code-mixing* dan keragaman linguistik. Sebagian besar pendekatan analisis sentimen yang ada masih bersifat monolingual, sehingga kurang efektif dalam menangani data lintas bahasa. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa algoritma *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) dalam melakukan klasifikasi sentimen multibahasa pada ulasan aplikasi Facebook. Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan total 9.600 ulasan aplikasi yang dikumpulkan dari Google Play Store, mencakup bahasa Indonesia, Inggris, dan Melayu, serta diklasifikasikan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral. Representasi teks dilakukan menggunakan *Language-Agnostic BERT Sentence Embedding* (LaBSE) untuk menangani variasi bahasa tanpa proses penerjemahan. Model CNN dirancang untuk mengekstraksi fitur lokal melalui filter konvolusi, sedangkan model LSTM memanfaatkan mekanisme sekuensial untuk memahami konteks kalimat. Evaluasi performa dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa model CNN memberikan performa terbaik dengan akurasi pengujian sebesar 91,18% dan nilai F1-score sebesar 0,91, sementara model LSTM mencapai akurasi sebesar 79,51% dengan F1-score sebesar 0,79. Meskipun kedua model mampu menangani data multibahasa secara umum, performa klasifikasi pada ulasan berbahasa Inggris dan sentimen netral masih menjadi tantangan utama akibat kompleksitas linguistik dan penggunaan bahasa informal. Secara keseluruhan, penelitian ini menunjukkan bahwa kombinasi embedding multibahasa dan arsitektur CNN lebih efektif untuk klasifikasi sentimen multibahasa pada domain ulasan media sosial.

**Kata kunci** : *Analisis Sentimen, Multibahasa, CNN, LSTM, LaBSE, Deep Learning.*

### ABSTRACT

The rapid growth of social media users, particularly on Facebook, has generated a massive volume of multilingual application reviews, posing significant challenges

for sentiment analysis due to code-mixing and linguistic diversity. Most existing sentiment analysis approaches remain monolingual, making them less effective in handling cross-lingual data. This study aims to evaluate and compare the performance of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) algorithms in classifying multilingual sentiment within Facebook app reviews. Using a quantitative experimental approach, 9,600 reviews comprising Indonesian, English, and Malay were collected from the Google Play Store and categorized into positive, negative, and neutral sentiments. Text representation was handled using Language-Agnostic BERT Sentence Embedding (LaBSE) to manage language variations without requiring translation. The CNN model was designed to extract local features via convolutional filters, while the LSTM model utilized sequential mechanisms to capture sentence context. Performance was evaluated using accuracy, precision, recall, and F1-score. Experimental results show that the CNN model outperformed LSTM, achieving a test accuracy of 91.18% and an F1-score of 0.91, compared to LSTM's accuracy of 79.51% and F1-score of 0.79. While both models generally handled multilingual data well, classifying English reviews and neutral sentiments remained a challenge due to linguistic complexity and informal language. Overall, this research demonstrates that the combination of multilingual embeddings and CNN architecture is highly effective for multilingual sentiment classification in the social media domain.

**Keyword** : *Sentiment Analysis, Multilingual, CNN, LSTM, LaBSE, Deep Learning.*

## 1. PENDAHULUAN

Pesatnya perkembangan teknologi informasi telah mengubah pola interaksi antara pengguna dan pengembang aplikasi. Media sosial, khususnya Facebook, telah menjadi platform utama bagi pengguna untuk mengekspresikan opini mereka secara terbuka. Ulasan yang diberikan oleh pengguna pada aplikasi ini mengandung informasi krusial mengenai tingkat kepuasan dan kualitas layanan. Namun, volume data ulasan yang sangat besar serta karakteristik teks yang bersifat heterogen dan multibahasa menjadi kendala utama dalam pengolahan data secara manual. Analisis sentimen hadir sebagai solusi untuk mengklasifikasikan opini tersebut secara otomatis ke dalam kategori

positif atau negatif guna memahami persepsi publik (Buntoro, 2017).

Tantangan dalam analisis sentimen pada media sosial di Indonesia menjadi lebih kompleks karena adanya penggunaan bahasa yang beragam dan tidak terstruktur. Menurut Riccosan dan Saputra (2025), klasifikasi sentimen pada ulasan aplikasi mobile di Indonesia memerlukan pendekatan yang mampu menangkap konteks semantik dari teks yang sering kali bersifat multibahasa. Selain itu, penggunaan bahasa campuran (*code-mixed*) sering ditemukan dalam interaksi digital, yang menuntut model klasifikasi untuk memiliki fleksibilitas tinggi (Mathur & Shrivastava, 2024). Faktor lain yang sangat menentukan akurasi

dalam proses ini adalah tahap pra-pemrosesan (*preprocessing*), karena pembersihan data yang tepat terbukti secara signifikan dapat meningkatkan kinerja model klasifikasi (Hidayatullah et al., 2019).

Dalam beberapa tahun terakhir, pendekatan *Deep Learning* telah mendominasi penelitian di bidang Pemrosesan Bahasa Alami (*Natural Language Processing*). Arsitektur seperti *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) telah banyak diuji efektivitasnya dalam menangani data teks. CNN memiliki keunggulan dalam mengekstraksi fitur lokal dan pola n-gram yang penting untuk klasifikasi kalimat (Kim, 2014), sementara LSTM dikenal karena kemampuannya dalam menangani ketergantungan sekuensial jangka panjang, yang sangat berguna dalam memahami konteks kalimat yang panjang dan kompleks (Al-Smadi et al., 2018).

Penelitian sebelumnya telah mencoba membandingkan kedua arsitektur ini dalam berbagai domain, seperti deteksi *cyberbullying* dan emosi di media sosial (Sari et al., 2023; Kustiwa et al., 2024). Namun, perbandingan performa antara CNN dan LSTM secara khusus pada ulasan aplikasi Facebook dengan konteks multibahasa masih memerlukan eksplorasi lebih lanjut. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi efektivitas kedua model tersebut guna menentukan arsitektur mana yang memberikan performa terbaik dalam mengklasifikasikan sentimen pengguna secara akurat pada domain ulasan aplikasi mobile.

## 2. LANDASAN TEORI

### A. Analisis Sentimen

Analisis sentimen merupakan bagian dari bidang *data mining* dan *machine learning* yang berfokus pada identifikasi serta ekstraksi opini, sentimen, dan emosi yang terkandung dalam teks (Sammud & Webb, 2017). Tujuan utamanya adalah untuk menentukan polaritas dari sebuah pernyataan, apakah cenderung positif, negatif, atau netral. Dalam pengembangannya, analisis ini sangat bergantung pada kualitas data input dan teknik representasi teks yang digunakan untuk membedakan makna di balik kata-kata (Buntoro, 2017).

### B. Analisis Sentimen Multibahasa

Analisis sentimen multibahasa (*Multilingual Sentiment Analysis*) merupakan studi tentang pengenalan sentimen pada teks yang melibatkan lebih dari satu bahasa. Kompleksitas dalam domain ini muncul akibat perbedaan struktur bahasa dan adanya teks yang tercampur (*code-mixed*). Menurut Elhadidi (2025), penggunaan model berbasis *deep learning* dan *transformer* memberikan kemampuan yang lebih baik dalam menyelaraskan representasi makna antar bahasa dibandingkan metode tradisional. Hal ini senada dengan pandangan Mathur dan Shrivastava (2024) yang menyatakan bahwa penanganan teks multibahasa memerlukan model yang mampu mengenali konteks linguistik yang luas agar tidak terjadi kehilangan informasi selama proses klasifikasi.

### C. Convolutional Neural Network (CNN)

CNN adalah jenis jaringan saraf tiruan yang awalnya populer dalam pengolahan citra, namun kini terbukti sangat efektif untuk klasifikasi teks. Arsitektur CNN bekerja dengan menerapkan lapisan konvolusi untuk mengekstraksi fitur-fitur penting atau pola kata kunci dari kalimat (Kim, 2014). Kemampuannya dalam menangkap fitur lokal melalui filter (kernel) membuat CNN sangat efisien dalam mengenali frasa-frasa yang membawa muatan sentimen tertentu, bahkan pada data teks yang mengandung sarkasme atau emosi yang kompleks (Poria et al., 2017).

### D. Long Short-Term Memory (LSTM)

LSTM adalah pengembangan dari *Recurrent Neural Network* (RNN) yang dirancang khusus untuk mengatasi masalah *vanishing gradient*. Arsitektur ini memiliki mekanisme internal berupa *gates* (pintu) yang memungkinkan model untuk menyimpan informasi penting dalam jangka waktu lama dan membuang informasi yang dianggap tidak relevan (Al-Smadi et al., 2018). Karakteristik ini membuat LSTM sangat unggul dalam memproses data urutan (*sequence data*) seperti kalimat, di mana urutan kata sangat menentukan makna keseluruhan teks (Sari et al., 2023).

### E. Studi Komparatif CNN dan LSTM

Studi mengenai perbandingan kedua model ini telah dilakukan dalam berbagai konteks. Misalnya, dalam penelitian deteksi emosi dan sentimen

masyarakat di Indonesia, ditemukan bahwa kedua model memiliki kekuatan yang berbeda tergantung pada karakteristik dataset yang digunakan (Rasyidah & Trisna, 2022). Sementara CNN sering kali lebih cepat dalam proses pelatihan karena strukturnya yang paralel, LSTM cenderung lebih unggul dalam menangkap nuansa kontekstual pada kalimat yang lebih panjang (Kustiwa et al., 2024). Pemahaman mendalam mengenai generalisasi pada *deep learning* sangat penting untuk memastikan bahwa model yang dibangun tidak hanya akurat pada data latih, tetapi juga handal pada data ulasan yang belum pernah dilihat sebelumnya (Zhang et al., 2017).

## 3. METODOLOGI

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental untuk membandingkan efektivitas dua arsitektur *deep learning* dalam tugas klasifikasi sentimen multibahasa. Alur penelitian dirancang secara sistematis mulai dari akuisisi data hingga evaluasi performa model.

### A. Karakteristik Dataset

Data penelitian diperoleh melalui proses *web scraping* pada platform Google Play Store, dengan fokus pada ulasan aplikasi Facebook. Sebanyak 9.600 data ulasan dikumpulkan untuk mencakup variasi bahasa yang luas, yaitu bahasa Indonesia, Inggris, dan Melayu. Dataset ini dipilih karena memiliki kompleksitas tinggi akibat penggunaan bahasa campuran (*code-mixing*) dan istilah slang media sosial. Seluruh data tersebut kemudian

dikategorikan ke dalam tiga kelas sentimen: positif, negatif, dan netral, yang menjadi dasar bagi model untuk mempelajari pola opini pengguna.

### B. Pra-Pemrosesan Data

Sebelum proses pemodelan, teks mentah melewati tahap pra-pemrosesan untuk meningkatkan kualitas data. Langkah ini meliputi pembersihan teks dari elemen non-informatif seperti URL dan simbol, penyeragaman huruf (*case folding*), serta pemecahan kalimat menjadi unit token. Untuk menjembatani perbedaan bahasa, penelitian ini mengimplementasikan *Language-Agnostic BERT Sentence Embedding* (LaBSE). Teknologi ini mentransformasikan teks ke dalam bentuk vektor numerik berdimensi 768. Keunggulan LaBSE terletak pada kemampuannya memetakan berbagai bahasa ke dalam ruang vektor yang sama, sehingga model dapat memahami makna semantik lintas bahasa tanpa memerlukan tahap penerjemahan manual yang berisiko menghilangkan konteks.

### C. Pengembangan Model dan Parameterisasi

Eksperimen ini membandingkan arsitektur *Convolutional Neural Network* (CNN) yang berfokus pada ekstraksi fitur lokal, dengan *Long Short-Term Memory* (LSTM) yang dirancang untuk menangkap konteks sekuensial jangka panjang. Data didistribusikan dengan rasio 70% untuk pelatihan, 15% untuk validasi, dan 15% untuk pengujian akhir. Untuk memastikan hasil yang objektif, kedua model dilatih menggunakan

konfigurasi *hyperparameter* yang optimal sebagai berikut:

Tabel 1. Konfigurasi hyperparameter model CNN dan LSTM

Parameter	Konfigurasi / Nilai
Model Embedding	LaBSE
Embedding Dimension	768
Dataset Split	70:15:15
Skema Validasi	5-Fold Cross Validation
Optimizer	Adam
Learning Rate	0.001
Loss Function	Categorical Cross-Entropy
Batch Size	32
Epochs	50
CNN Architecture	Conv1D + GlobalMaxPooling
LSTM Architecture	Single Layer LSTM (128 units)
Activation Output	Softmax

Proses pelatihan juga mengintegrasikan mekanisme *5-fold cross-validation* untuk menjamin stabilitas performa dan meminimalkan risiko *overfitting* pada salah satu subset data.

### D. Evaluasi Performa

Tahap akhir penelitian adalah melakukan pengujian pada data uji (*test set*) yang belum pernah dikenali oleh model sebelumnya. Efektivitas dari CNN dan LSTM diukur secara komprehensif menggunakan metrik akurasi untuk melihat performa global, serta presisi, *recall*, dan *F1-score* untuk mengevaluasi ketajaman model

dalam mendeteksi masing-masing kelas sentimen secara seimbang. Hasil dari metrik-metrik ini menjadi dasar utama dalam menentukan algoritma yang paling unggul untuk klasifikasi teks multibahasa.

#### 4. HASIL DAN PEMBAHASAN

##### A. Implementasi dan Pelatihan Model

Proses pelatihan dilakukan menggunakan skema *5-Fold Cross-Validation* untuk menjaga objektivitas evaluasi. Pengujian dilakukan pada dua arsitektur utama, yaitu *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM), dengan memanfaatkan *Language-Agnostic BERT Sentence Embedding* (LaBSE) sebagai representasi fitur.

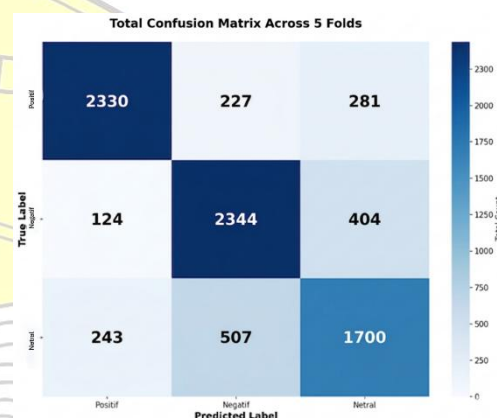
##### 1. Performa Pelatihan CNN

Model CNN dilatih dengan fokus pada ekstraksi fitur lokal melalui filter *n-gram* (trigram, *four-gram*, dan *five-gram*). Berdasarkan hasil eksperimen, model CNN menunjukkan stabilitas yang sangat tinggi. Rekapitulasi hasil pelatihan dapat dilihat pada Tabel 2.

Tabel 2. rekapitulasi hasil rata-rata *5-fold cross-validation* CNN

Metrik Evaluasi	Nilai Rata-rata (Mean)	Standar Deviasi (Std Dev)
Akurasi	0.7811 (78,11%)	± 0.0053
Presisi (Macro)	0.7796	-
Recall (Macro)	0.7770	-
F1-Score (Macro)	0.7776	-

Rendahnya nilai standar deviasi (0,0053) mengindikasikan bahwa model CNN memiliki tingkat ketahanan (*robustness*) yang tinggi terhadap variasi *pembagian* data. Penggunaan *class weights* terbukti membantu model dalam mempertahankan nilai *recall* pada kelas minoritas (netral dan negatif) di tengah karakteristik bahasa informal media sosial.



Gambar 1. Akumulasi prediksi CNN (total *5-fold*)

##### 2. Performa Pelatihan LSTM

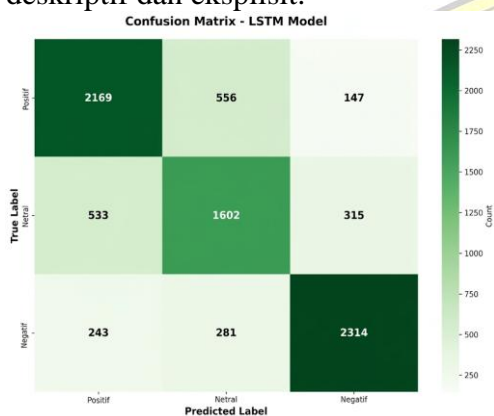
Pelatihan LSTM difokuskan pada penangkapan ketergantungan sekuensial jangka panjang. Meskipun menunjukkan pola pembelajaran yang stabil, performa rata-rata LSTM berada di bawah CNN sebagaimana dirinci pada Tabel 3.

Tabel 3. Rekapitulasi hasil rata-rata *5-fold cross-validation* lstm

Metrik Evaluasi	Nilai Rata-rata (Mean)	Standar Deviasi (Std Dev)
Akurasi	0.7457 (74,57%)	± 0.0056
Presisi (Macro)	0.7432	-

Recall (Macro)	0.7415	-
F1-Score (Macro)	0.7416	-

LSTM tercatat paling stabil dalam mengenali sentimen negatif dengan *F1-score* mendekati 0,82, menunjukkan efektivitas arsitektur rekuren dalam menangkap pola kalimat keluhan yang cenderung deskriptif dan eksplisit.



Gambar 2. Akumulasi prediksi LSTM (total 5-fold)

### B. Pembahasan dan Evaluasi Akhir (Testing)

Tahap pengujian akhir dilakukan menggunakan *test set* independen sebanyak 1.440 data untuk memvalidasi performa model pada data yang belum pernah dilihat sebelumnya.

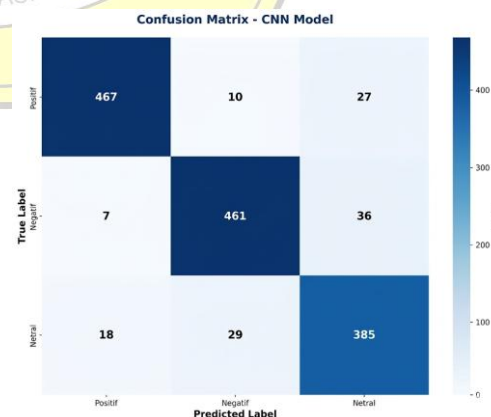
#### 1. Analisis Performa CNN

Model CNN mengalami peningkatan performa signifikan pada tahap *testing* dengan akurasi mencapai 91,18%. Analisis per kelas menunjukkan sentimen positif meraih skor tertinggi, sementara kelas netral menjadi tantangan terbesar karena ambiguitas linguistik.

Tabel 4. Metrik evaluasi per kelas pada tahap *testing* CNN

Kelas	Presisi	Recall	F1-Score
Positif	0.9492	0.9266	0.9378
Negatif	0.9220	0.9147	0.9183
Netral	0.8594	0.8912	0.8750
MACRO AVG	0.9102	0.9108	0.9104

Meskipun model CNN mencapai akurasi tinggi secara keseluruhan, analisis mendalam pada tabel di atas menunjukkan bahwa kelas netral memiliki *F1-score* terendah (0.8750) dibandingkan kelas lainnya. Hal ini disebabkan oleh karakteristik ulasan aplikasi yang sering kali berisi pernyataan faktual tanpa kata sifat yang menunjukkan polaritas kuat, sehingga filter *n-gram* pada CNN kesulitan mengekstraksi fitur diskriminatif. Sebagai contoh, ulasan seperti 'Aplikasi sudah diperbarui' atau 'Sedang mencoba fitur baru' sering kali terklasifikasi sebagai positif karena kedekatan semantik dalam ruang vektor LaBSE dengan konteks pembaruan sistem yang umumnya bernada baik.



Gambar 3. Heatmap confusion matrix akhir CNN

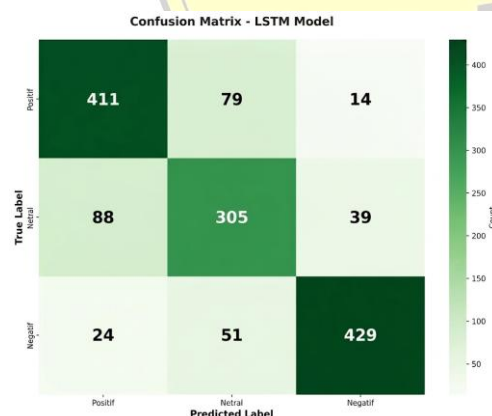
Rendahnya tingkat "kesalahan fatal" (tertukarnya sentimen positif menjadi negatif atau sebaliknya) yang hanya sebesar 1,4% membuktikan bahwa model memiliki pemahaman polaritas yang sangat kuat.

## 2. Analisis Performa LSTM

Model LSTM menghasilkan akurasi final sebesar 79,51%. Berbeda dengan CNN, LSTM memiliki margin kesalahan yang lebih lebar dalam membedakan polaritas yang berlawanan, terutama pada ulasan singkat di mana ketergantungan jangka panjang sulit terbentuk.

Tabel 5. Metrik evaluasi per kelas pada tahap *testing* LSTM

Kelas	Presisi	Recall	F1-Score
Positif	0.7859	0.8155	0.8004
Netral	0.7011	0.7060	0.7036
Negatif	0.8900	0.8512	0.8702
MACRO AVG	0.7923	0.7909	0.7914



Gambar 4. Heatmap confusion matrix akhir LSTM

## C. Perbandingan Kompartatif dan Analisis Bahasa

Perbandingan akhir menunjukkan dominasi model CNN atas LSTM dengan selisih akurasi mencapai 11,67%.

Tabel 6. Perbandingan metrik evaluasi final CNN dan LSTM

Model	CNN	LSTM
Akurasi	91,18%	79,51%
Presisi (Macro)	0.9102	0.7923
Recall (Macro)	0.9108	0.7909
F1-Score (Macro)	0.9104	0.7914

Keunggulan CNN dipengaruhi oleh kemampuannya menangkap fitur *n-gram* pada ulasan media sosial yang singkat dan padat. Selain itu, dilakukan analisis berdasarkan distribusi bahasa untuk melihat pengaruh karakteristik linguistik (Tabel 7).

Tabel 7. Perbandingan akurasi berdasarkan bahasa

Model	Indonesia (id)	Melayu (ms)	Inggris (en)
CNN	98,75%	98,54%	76,25%
LSTM	85,83%	79,79%	72,92%

Penurunan performa yang signifikan pada kategori Bahasa Inggris (76,25% pada CNN) menjadi titik kritis dalam penelitian ini. Investigasi lebih lanjut menunjukkan bahwa fenomena *code-mixing* dan penggunaan *slang* regional (seperti penggunaan singkatan informal atau dialek non-standar) menyebabkan representasi *embedding* LaBSE menjadi lebih tersebar (*scattered*). Berbeda dengan Bahasa Indonesia dan Melayu yang cenderung memiliki struktur lebih konsisten dalam *dataset*

ini, ulasan berbahasa Inggris dari pengguna global memiliki variasi linguistik yang lebih ekstrem. Hal ini mengindikasikan bahwa meskipun LaBSE bersifat *language-agnostic*, keterbatasan dalam memahami nuansa budaya dan bahasa prokem (*slang*) pada domain media sosial tetap menjadi tantangan teknis yang memerlukan tahap normalisasi teks yang lebih intensif di masa mendatang.

Hasil menunjukkan performa sangat tinggi pada Bahasa Indonesia dan Melayu (di atas 98% untuk CNN). Penurunan signifikan terjadi pada Bahasa Inggris (76,25% pada CNN dan 72,92% pada LSTM). Hal ini diidentifikasi bukan karena kelemahan model, melainkan tingginya fenomena *code-mixing*, *code-switching*, dan penggunaan slang pada ulasan regional Amerika Serikat, yang menyebabkan representasi embedding LaBSE menjadi lebih tersebar dan sulit dipisahkan secara semantik dibandingkan bahasa regional Nusantara.

Keunggulan CNN atas LSTM dengan selisih akurasi 11,67% menegaskan bahwa untuk teks media sosial yang bersifat singkat (*short-text*), ekstraksi fitur lokal melalui mekanisme konvolusi lebih efektif daripada penangkapan ketergantungan sekuensial jangka panjang pada LSTM. Pada ulasan Facebook yang padat, informasi sentimen biasanya terkonsentrasi pada frasa tertentu (*keyword-heavy*), sehingga kemampuan CNN dalam mendeteksi pola *n-gram* memberikan hasil yang lebih robust dibandingkan LSTM yang cenderung membutuhkan struktur kalimat yang lebih lengkap untuk

membentuk memori sekuensial yang akurat.

## 5. KESIMPULAN

Berdasarkan hasil penelitian dan analisis komparatif yang telah dilakukan, dapat disimpulkan bahwa arsitektur *Convolutional Neural Network* (CNN) menunjukkan performa yang lebih unggul dibandingkan *Long Short-Term Memory* (LSTM) dalam tugas klasifikasi sentimen multibahasa pada ulasan aplikasi Facebook. Integrasi *Language-Agnostic BERT Sentence Embedding* (LaBSE) terbukti efektif dalam menyediakan representasi fitur yang stabil untuk ulasan dalam bahasa Indonesia, Melayu, dan Inggris. Superioritas model CNN terlihat dari pencapaian akurasi pengujian akhir sebesar 91,18%, yang secara signifikan mengungguli model LSTM dengan akurasi 79,51%. Hal ini mengindikasikan bahwa kemampuan CNN dalam mengekstraksi fitur lokal melalui mekanisme *n-gram* lebih efektif untuk menangani teks media sosial yang cenderung singkat dan padat dibandingkan dengan pendekatan sekuensial pada LSTM.

Dalam aspek kinerja multibahasa, model ini menunjukkan performa yang sangat tinggi pada ulasan berbahasa Indonesia dan Melayu, masing-masing mencapai akurasi 98,75% dan 98,54%. Namun, tantangan signifikan muncul pada kategori bahasa Inggris yang hanya mencapai akurasi 76,25%. Rendahnya akurasi pada bahasa Inggris tersebut dipicu oleh tingginya variasi penggunaan bahasa informal, *slang*, serta fenomena *code-mixing* pada ulasan regional yang mempersulit proses klasifikasi. Meskipun

demikian, penggunaan skema *5-fold cross-validation* menunjukkan nilai standar deviasi yang sangat rendah, yaitu  $\pm 0,0053$ , yang membuktikan bahwa model CNN memiliki tingkat ketahanan (*robustness*) serta konsistensi yang tinggi terhadap variasi distribusi data.

Sebagai langkah pengembangan di masa depan, penelitian ini menyarankan perlunya penerapan teknik normalisasi linguistik yang lebih mendalam untuk menangani ketidakteraturan bahasa. Selain itu, penggunaan metode *fine-tuning* pada model berbasis *Transformer* yang lebih spesifik sangat direkomendasikan untuk menangani teks *slang* dan *code-mixing*. Langkah-langkah tersebut diharapkan dapat meningkatkan akurasi klasifikasi, khususnya pada kategori bahasa Inggris dan ulasan dengan karakteristik bahasa informal yang kompleks.

## 6. DAFTAR PUSTAKA

- Al-Smadi, M., Qawasmeh, O., Talafha, B., & Quwaider, M. (2018). Using LSTM for detecting emotions in Arabic tweets. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 9(8), 456–463.
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, 2(1), 32–41.
- Elhadidi, E. A. (2025). Multilingual Sentiment Analysis: A Review of Deep Learning Transformer and Ensemble Models. *International Journal of Computers and Information*, 1(7), 91–103.
- Hidayatullah, A., Prabowo, H., & Rizki, A. (2019). Pengaruh preprocessing terhadap akurasi analisis sentimen di media sosial. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 6(2), 200–208.
- Kim, Y. (2014). Convolutional neural networks for sentence classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751.
- Kustiwa, R., Akbar, R., & Pinandito, H. (2024). Studi Perbandingan CNN-LSTM dan LSTM untuk Deteksi Emosi di Twitter. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 12(1), 10–18.
- Mathur, A., & Shrivastava, M. (2024). Multilingual Sentiment Analysis and Code-Mixed Text Processing: A Review. *Journal of Artificial Intelligence Research*, 75, 1–30.
- Poria, S., Cambria, E., Hazarika, D., & Vij, P. (2017). A deeper look into sarcastic tweets using deep convolutional neural networks. *Proceedings of the 2017 Conference on Computational Natural Language Learning (CoNLL)*, 160–170.
- Rasyidah, N., & Trisna, N. (2022). Perbandingan CNN dan Bi-LSTM pada Analisis Sentimen dan Emosi Masyarakat Indonesia di Media Sosial Twitter selama Pandemi COVID-19. *Jurnal Sistem Informasi*, 18(1), 25–35.
- Riccosan, R., & Saputra, K. E. (2025). Multilabel classification sentiment analysis on Indonesian mobile app reviews. *IAES International Journal of Artificial*

*Intelligence (IJ-AI)*, 14(5), 4226–4234.

Sammut, C., & Webb, G. I. (Eds.). (2017). *Encyclopedia of Machine Learning and Data Mining*. Springer.

Sari, N. P., Wahyuni, D. R., & Nasution, M. K. M. (2023). Analisa Perbandingan CNN dan LSTM untuk Klasifikasi Pesan Cyberbullying pada Twitter. *Jurnal Teknologi dan Sistem Komputer*, 11(2), 215–223.

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017). Understanding deep learning requires rethinking generalization. *International Conference on Learning Representations (ICLR)*.

