

# Analisis Sentimen Ulasan Pengguna Aplikasi Media Sosial X Di *Play Store* Menggunakan Algoritma *Long Short-Term Memory* (LSTM) Dan *Gated Recurrent Unit* (GRU)

Wily Arisandi<sup>1</sup>, Sajarwo Anggai<sup>2</sup>, Tukiya<sup>3</sup>

Program Studi Teknik Informatika S-2, Program Pascasarjana, Universitas Pamulang, Kota Tangerang Selatan, Banten

[1arisandiwilywidi@gmail.com](mailto:1arisandiwilywidi@gmail.com), [2sajarwo@gmail.com](mailto:2sajarwo@gmail.com), [3dosen02711@unpam.ac.id](mailto:3dosen02711@unpam.ac.id)

Diterima : 30 Agustus 2025

Disetujui : 20 September 2025

**Abstract**—Meningkatnya penggunaan media sosial X menghasilkan banyak ulasan pengguna di *Google Play Store* yang mencerminkan kepuasan dan pengalaman mereka. Namun, jumlah ulasan yang besar menyulitkan analisis manual, sehingga dibutuhkan metode otomatis seperti *deep learning* untuk mengklasifikasikan sentimen. Penelitian ini bertujuan untuk membandingkan performa dua algoritma *deep learning*, yaitu *Long Short-Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU), dalam melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi media sosial X di *Google Play Store*. Dataset dikumpulkan dalam rentang waktu 12 Februari hingga 14 November 2024, yang digunakan sebanyak 5.100 data ulasan yang telah diberi pelabelan secara manual ke dalam tiga kategori sentimen kelas positif, netral, dan negatif. Data dibagi menjadi tiga bagian, yaitu 70% untuk pelatihan, 15% validasi, dan 15% pengujian, dengan eksperimen 12 skenario *hyperparameter* yang melibatkan variasi nilai *learning rate*, *epoch*, dan *batch size*. Hasil optimalisasi model menunjukkan bahwa model LSTM dengan skenario (0.002-LSTM-100-512) memberikan performa terbaik dengan akurasi 0.842, presisi 0.730, recall 0.719, dan F1-score 0.724. Sementara itu, model GRU terbaik dengan skenario (0.001-GRU-100-256) menghasilkan akurasi 0.837, presisi 0.713, recall 0.690, dan F1-score 0.696. Berdasarkan hasil evaluasi model dengan 12 skenario *hyperparameter*, model LSTM dipilih sebagai model paling optimal untuk tugas analisis sentimen dalam penelitian ini. Model LSTM unggul dalam semua metrik lainnya, terutama F1-score yang menjadi indikator utama dalam penelitian ini karena mencerminkan keseimbangan antara presisi dan recall.

**Keywords** —Analisis Sentimen, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Ulasan Pengguna, Media Sosial X, Google Play Store.

## I. PENDAHULUAN

Perkembangan teknologi digital yang begitu cepat telah mendorong berbagai organisasi untuk terus beradaptasi, agar tetap mampu bersaing dan bertahan dalam lingkungan yang kompetitif. Transformasi digital membawa perubahan signifikan dalam cara manusia berinteraksi, berkomunikasi, dan memanfaatkan teknologi informasi. Implementasi teknologi ini

memengaruhi pola pikir, perilaku, serta cara individu maupun institusi menjalankan aktivitas sehari-hari (Dian Sudiantini et al., 2023).

Salah satu dampak nyata dari perkembangan ini adalah meningkatnya penggunaan media sosial yang mudah diunduh melalui platform seperti *Google Play Store*. Pada platform ini, selain mengunduh aplikasi, pengguna juga dapat memberikan ulasan dan penilaian sebagai bentuk

tanggapan terhadap pengalaman mereka. Ulasan pengguna menjadi faktor penting dalam menentukan reputasi, popularitas, dan tingkat penerimaan sebuah aplikasi.

*Google Play Store* kini juga berfungsi sebagai wadah diskusi antar pengguna, termasuk aplikasi X yang sebelumnya dikenal sebagai Twitter. Setelah akuisisi Twitter oleh Elon Musk pada 27 Oktober 2022, aplikasi ini mengalami proses rebranding menjadi X pada 23 Juli 2023, yang ditandai dengan perubahan logo dan nama aplikasi (Eskiyaturrofikoh & Suryono, 2024). Perubahan tersebut memicu beragam respons dari pengguna, yang tercermin melalui ulasan yang mereka tulis di *Play Store*.

Ulasan yang diberikan pengguna mencerminkan pengalaman dan persepsi mereka terhadap berbagai aspek aplikasi, seperti fitur, performa, tampilan antarmuka, hingga kualitas layanan. Namun, jumlah ulasan yang sangat banyak membuat analisis secara manual menjadi tidak efisien dan memakan waktu. Untuk itu, dibutuhkan pendekatan berbasis analisis sentimen yang dapat mengolah data ulasan secara otomatis. Analisis sentimen bertujuan mengidentifikasi opini atau perasaan yang terkandung dalam teks, baik berupa positif, negatif, maupun netral, sehingga dapat memberikan wawasan yang lebih jelas mengenai tanggapan pengguna (Akbar & Sugiharto, 2023).

Dalam melakukan analisis sentimen, dibutuhkan metode yang efektif untuk memproses teks. Salah satu pendekatan yang populer adalah deep learning, khususnya model yang mampu menangani data sekuensial seperti teks. *Long Short-Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU) merupakan dua jenis algoritma deep learning yang dirancang untuk memahami konteks urutan kata dalam teks. Keunggulan utama kedua algoritma ini terletak pada kemampuannya mengatasi masalah *vanishing gradient*, yang sering terjadi pada model *Recurrent Neural Network* (RNN) tradisional. Dengan demikian, LSTM dan GRU mampu memproses teks yang memiliki konteks kompleks, seperti kalimat yang memuat sentimen

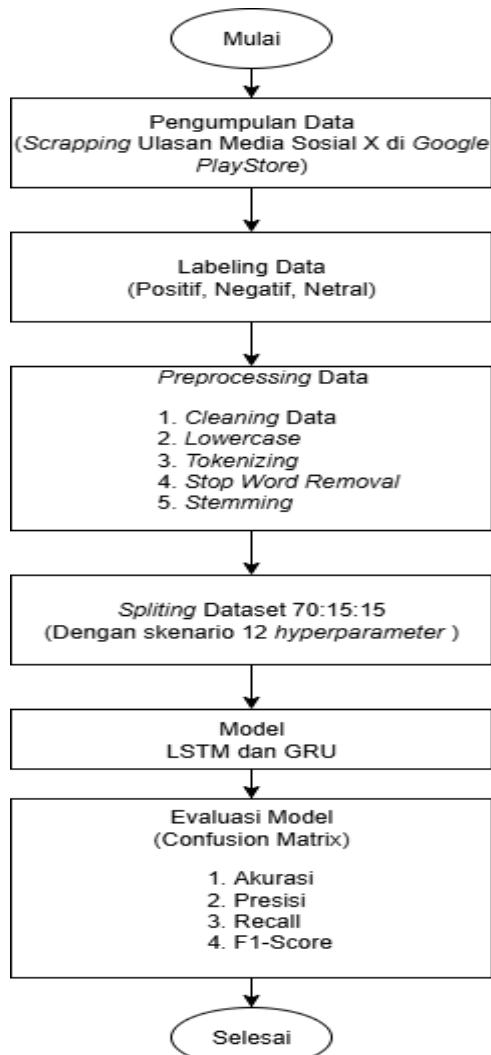
campuran. Misalnya, dalam kalimat "*Fiturnya bagus, tetapi aplikasinya sering crash*", diperlukan pemahaman konteks untuk menangkap sentimen yang sebenarnya.

Hasil penelitian sebelumnya menunjukkan bahwa perbedaan arsitektur LSTM dan GRU dapat memengaruhi kinerja model. Penelitian oleh (Ayuningtyas & Khomsah (2024)) menunjukkan bahwa GRU memiliki akurasi lebih tinggi dibandingkan LSTM dalam klasifikasi sentimen, khususnya ketika menggunakan teknik pembobotan kata seperti *word2vec* dan *Continuous Bag of Word*. Namun, kualitas data dan konfigurasi hyperparameter seperti *learning rate*, jumlah *epoch*, dan *batch size* juga berperan penting dalam menentukan hasil pelatihan dan pengujian model.

Berdasarkan kondisi tersebut, penelitian ini dilakukan untuk membandingkan performa algoritma LSTM dan GRU dalam klasifikasi sentimen ulasan pengguna aplikasi X di *Google Play Store*. Dengan memanfaatkan dataset yang diperoleh melalui teknik *web scraping*, penelitian ini bertujuan untuk mengidentifikasi sentimen dominan, memahami pola persepsi pengguna, serta memberikan rekomendasi berbasis data bagi pengembang aplikasi. Selain itu, hasil penelitian ini diharapkan dapat menjadi referensi bagi komunitas akademik dalam penerapan deep learning untuk analisis teks, sekaligus memberikan pemahaman praktis mengenai pengalaman pengguna setelah terjadinya perubahan besar pada aplikasi X.

## II. METODOLOGI PENELITIAN

Metodologi penelitian ini dirancang untuk menggambarkan tahapan yang dilakukan mulai dari proses pengumpulan data hingga evaluasi model, sehingga hasil yang diperoleh sesuai dengan tujuan penelitian. Secara umum, tahapan penelitian ini meliputi pengumpulan data, prapemrosesan data, pembagian dataset (*splitting*), pelabelan data, pembangunan model, dan evaluasi performa. Alur lengkap proses penelitian dapat dilihat pada Gambar 1.



Gambar 1 Perancangan penelitian

## 2.1 PERSIAPAN DATASET

Data yang digunakan dalam penelitian ini berupa ulasan pengguna aplikasi X yang diperoleh dari *Google Play Store*. Proses pengambilan data dilakukan dengan teknik *web scraping* menggunakan pustaka *google-play-scraper* berbasis Python. Teknik ini dipilih karena mampu mengunduh data ulasan dalam jumlah besar secara otomatis dan terstruktur, tanpa memerlukan akses API resmi.

Ulasan yang dikumpulkan mencakup periode waktu 12 Februari 2024 hingga 14 November 2024. Hasil pengumpulan data awal berjumlah 22.000 baris yang terdiri dari empat atribut utama, yaitu:

1. User Name – identitas pengguna yang memberikan ulasan,

2. Score – penilaian berbentuk bintang (1–5),
3. At – tanggal ulasan diberikan,
4. Content – teks ulasan yang berisi opini pengguna.

Seluruh data yang dikumpulkan kemudian disimpan dalam format **CSV** agar mudah diproses pada tahap prapemrosesan berikutnya.

## 2.2 PREPROCESSING DATA

Tahap prapemrosesan bertujuan untuk meningkatkan kualitas data sebelum digunakan dalam pelatihan model. Data mentah dari web scraping sering kali mengandung **noise**, seperti duplikasi, teks kosong, atau karakter yang tidak relevan. Oleh karena itu, beberapa langkah pembersihan data dilakukan, yaitu:

1. Menghapus duplikasi data, menghilangkan ulasan yang tercatat lebih dari satu kali.
2. Menghapus nilai kosong (null values), menghapus baris yang tidak memiliki teks ulasan.
3. Membersihkan teks dasar, menghapus URL, tag, *mention*, angka, tanda baca, serta spasi ganda.

content	clean_content
tolong bantu saya buat verifikasi identitas ka...	tolong bantu saya buat verifikasi identitas ka...
Banyak masuk iklan tolol;	Banyak masuk iklan tolol
Baru dibuka pagi udah beralih akun! Bukanya ta...	Baru dibuka pagi udah beralih akun Bukanya tam...
Sudah beberapa hari ini akun saya di tangguhka...	Sudah beberapa hari ini akun saya di tangguhka...
Good;	Good

Gambar 2 Hasil *Cleaning* Dataset

Setelah pembersihan awal, dilakukan normalisasi teks untuk menyamakan format data. Adapun tahapan lengkap prapemrosesan adalah sebagai berikut:

### A. LOWERCASE

Seluruh teks ulasan diubah menjadi huruf kecil untuk menyeragamkan data. Hal ini bertujuan agar kata yang sama dengan format berbeda, seperti "Bagus" dan "bagus", diperlakukan sebagai entitas yang sama.

	clean_content	lower_content
0	tolong bantu saya buat verifikasi identitas ka...	tolong bantu saya buat verifikasi identitas ka...
1	Banyak masuk iklan tolol	banyak masuk iklan tolol
2	Baru dibuka pagi udah beralih akun Bukanya tam...	baru dibuka pagi udah beralih akun bukanya tam...
3	Sudah beberapa hari ini akun saya di tangguhka...	sudah beberapa hari ini akun saya di tangguhka...
4	Good	good

Gambar 3 Hasil Lowercase

### B. TOKENIZING

Tokenisasi merupakan proses memecah teks ulasan menjadi unit-unit kata yang lebih kecil yang disebut token. Proses ini penting dalam analisis teks karena memungkinkan sistem untuk memahami struktur dan isi dari teks secara lebih

	lower_content	tokenized_content
0	tolong bantu saya buat verifikasi identitas ka...	[tolong, bantu, saya, buat, verifikasi, identi...
1	banyak masuk iklan tolol	[banyak, masuk, iklan, tolol]
2	baru dibuka pagi udah beralih akun bukanya tam...	[baru, dibuka, pagi, udah, beralih, akun, buka...
3	sudah beberapa hari ini akun saya di tangguhka...	[sudah, beberapa, hari, ini, akun, saya, di, t...
4	good	[good]

terperinci.

Gambar 4 Hasil Tokenizing

### C. STOPWORD REMOVAL

Stopword adalah kata-kata umum yang sering muncul namun tidak memiliki makna signifikan dalam analisis, seperti "dan", "yang", atau "ini". Penghapusan stopwords bertujuan untuk mengurangi dimensi data dan meningkatkan fokus analisis hanya pada kata-kata yang relevan dengan sentimen.

	tokenized_content	no_stopwords
0	[tolong, bantu, saya, buat, verifikasi, identi...	[tolong, bantu, buat, verifikasi, identitas, l...
1	[banyak, masuk, iklan, tolol]	[banyak, masuk, iklan]
2	[baru, dibuka, pagi, udah, beralih, akun, buka...	[baru, dibuka, pagi, udah, beralih, akun, buka...
3	[sudah, beberapa, hari, ini, akun, saya, di, t...	[beberapa, hari, akun, tangguhkan, padahal, ke...
4	[good]	[good]

Gambar 5. Hasil Stopword Removal

### D. STEMMING

Stemming adalah proses mengubah kata berimbuhan menjadi bentuk dasar. Misalnya, kata "dibuka" diubah menjadi "buka", atau "beralih" menjadi "alih". Proses ini memastikan berbagai variasi kata dianalisis sebagai satu entitas yang sama.

no_stopwords	stemmed_content
['tolong', 'bantu', 'buat', 'verifikasi', 'identitas', 'banyak', 'masuk', 'iklan']	['tolong', 'bantu', 'buat', 'verifikasi', 'identitas', 'banyak', 'masuk', 'iklan']
['baru', 'dibuka', 'pagi', 'udah', 'beralih', 'akun', 'beberapa', 'hari', 'akun', 'tangguhkan', 'pada', 'good']	['baru', 'buka', 'pagi', 'udah', 'alih', 'akun', 'buk', 'beberapa', 'hari', 'akun', 'tangguh', 'padahal', 'good']

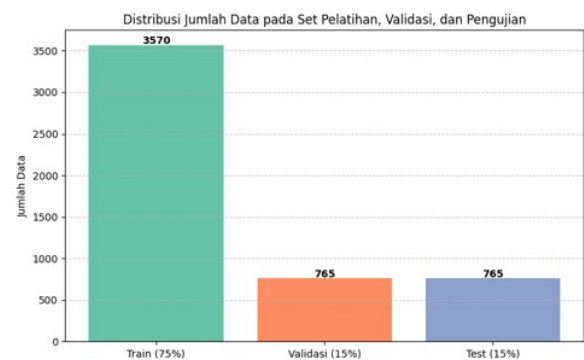
Gambar 5. Hasil Stemming

## 2.3 SPLITTING DATASET

Setelah tahap prapemrosesan selesai, data dibagi menjadi tiga bagian agar proses pelatihan dan evaluasi model lebih terstruktur, yaitu:

- 70% data pelatihan (*training set*)
- 15% data validasi (*validation set*)
- 15% data pengujian (*testing set*)

Dari 12.376 data ulasan yang telah diproses, sebanyak 5.100 data ulasan dilabeli secara manual ke dalam tiga kategori sentimen: positif, netral, dan negatif. Pembagian data dilakukan secara acak (*random shuffle*) untuk memastikan distribusi data yang merata pada setiap subset. Proses pembagian dataset ditunjukkan pada Gambar 6.



Gambar 6. Splitting Dataset

## 2.4 SKENARIO HYPERPARAMETER

Penelitian ini melakukan eksperimen menggunakan 12 skenario hyperparameter untuk masing-masing model LSTM dan GRU. Skenario ini dirancang dengan memvariasikan:

1. Learning rate merupakan kecepatan pembaruan bobot model,
2. Epoch merupakan jumlah iterasi pelatihan,
3. Batch size merupakan jumlah data yang diproses sebelum pembaruan bobot,
4. L2 Regularization merupakan teknik untuk mencegah *overfitting*.

Konfigurasi L2 regularization yang digunakan adalah:

- L2 = 0.001 untuk *learning rate* 0.002
- L2 = 0.01 untuk *learning rate* 0.001

Table 1 Skenario Pemodelan Algoritma LSTM

Learning Rate	L2 Regularization	Epochs	Batch Size	Nama Skenario
0.002	0.001	50	128	0.002-LSTM-50-128
0.002	0.001	50	256	0.002-LSTM-50-256
0.002	0.001	50	512	0.002-LSTM-50-512
0.002	0.001	100	128	0.002-LSTM-100-128
0.002	0.001	100	256	0.002-LSTM-100-256
0.002	0.001	100	512	0.002-LSTM-100-512
0.001	0.01	50	128	0.001-LSTM-50-128
0.001	0.01	50	256	0.001-LSTM-50-256
0.001	0.01	50	512	0.001-LSTM-50-512
0.001	0.01	100	128	0.001-LSTM-100-128
0.001	0.01	100	256	0.001-LSTM-100-256
0.001	0.01	100	512	0.001-LSTM-100-512

Table 2 Skenario Pemodelan Algoritma GRU

Learning Rate	L2 Regularization	Epochs	Batch Size	Nama Skenario
0.002	0.001	50	128	0.002-GRU-50-128
0.002	0.001	50	256	0.002-GRU-50-256
0.002	0.001	50	512	0.002-GRU-50-512
0.002	0.001	100	128	0.002-GRU-100-128
0.002	0.001	100	256	0.002-GRU-100-256
0.002	0.001	100	512	0.002-GRU-100-512
0.001	0.01	50	128	0.001-GRU-50-128
0.001	0.01	50	256	0.001-GRU-50-256
0.001	0.01	50	512	0.001-GRU-50-512
0.001	0.01	100	128	0.001-GRU-100-128
0.001	0.01	100	256	0.001-GRU-100-256
0.001	0.01	100	512	0.001-GRU-100-512

## 2.6 EVALUASI MODEL

Evaluasi performa model dilakukan menggunakan *Confusion Matrix* dan empat metrik utama, yaitu:

1. Akurasi untuk Mengukur persentase prediksi model yang benar dari total data, dengan rumus dibawah ini.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (3.1)$$

2. *Precision* untuk Mengukur proporsi data yang benar-benar positif dari semua prediksi positif, dengan rumus dibawah ini:

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (3.2)$$

Mengukur ketepatan model dalam memprediksi kelas positif, atau berapa banyak prediksi positif yang benar.

3. *Recall* untuk Mengukur seberapa baik model dapat mendeteksi kelas positif, dengan rumus dibawah ini:

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (3.3)$$

Jika semakin tinggi recall, semakin sedikit *false negative* yang dibuat model.

4. *F1-Score* untuk menggabungkan precision dan recall untuk memberikan gambaran menyeluruh performa model, terutama jika ada ketidakseimbangan kelas sentimen dalam data, dengan rumus dibawah ini.

$$F-1\ Score = 2 \frac{Precision \times Recall}{Precision + Recall} \dots\dots\dots (3.4)$$

Jika nilai F1-Score tinggi berarti model memiliki keseimbangan antara precision dan recall.

		Actual Values	
		Positive	Negative
Predicted Values	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

Gambar 7. Ilustrasi *Confusion Matrix*

## III. HASIL DAN PEMBAHASAN

Pada bagian ini dijelaskan hasil eksperimen dan analisis performa model LSTM dan GRU berdasarkan 12 skenario hyperparameter yang telah diuji. Hasil yang dipaparkan mencakup **nilai akurasi, presisi, recall, dan F1-score** untuk setiap skenario.

Selain itu, disajikan juga confusion matrix serta grafik akurasi dan loss untuk memvisualisasikan kinerja model selama proses pelatihan dan validasi. Fokus utama penelitian ini adalah F1-score, karena metrik ini memberikan gambaran keseimbangan antara presisi dan recall pada data yang memiliki distribusi kelas yang tidak seimbang.

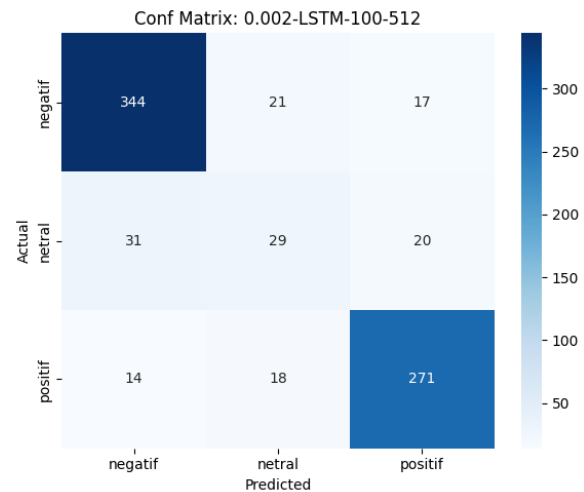
### 3.1 HASIL ANALISIS MODEL LSTM

Eksperimen model LSTM dilakukan pada 12 konfigurasi hyperparameter dengan pembagian data 70% pelatihan, 15% validasi, dan 15% pengujian.

Berdasarkan hasil pengujian, skenario terbaik diperoleh pada **0.002-LSTM-100-512**.

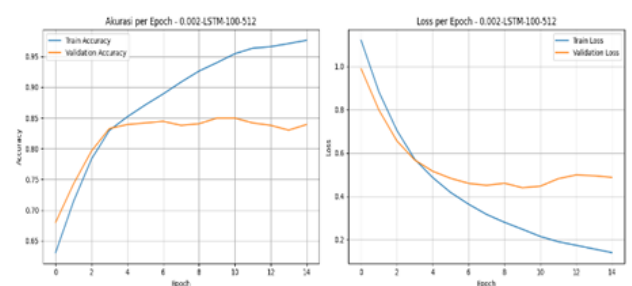
Hasil utama untuk model LSTM ditunjukkan pada poin berikut:

1. Akurasi tertinggi model dengan konfigurasi 0.002-LSTM-100-512 mencapai akurasi 0.842, yang berarti 84,2% dari data uji dapat diprediksi dengan benar.
2. Presisi tertinggi presisi tertinggi juga diperoleh pada skenario yang sama, yaitu 0.730, menunjukkan bahwa model mampu meminimalkan prediksi positif yang salah (*false positive*).
3. *Recall* tertinggi nilai recall tertinggi sebesar 0.723 dicapai oleh skenario 0.002-LSTM-50-512, yang menandakan kemampuan model dalam mengenali data dari kelas yang benar dan meminimalkan *false negative*.
4. *F1-score* tertinggi skenario 0.002-LSTM-100-512 mencatatkan F1-score 0.724, yang menjadi indikator bahwa model ini memiliki keseimbangan terbaik antara presisi dan recall.



Gambar 2 Confusion Matrix 0.002-LSTM-100-512

Pada gambar 8 *Confusion matriks* menunjukkan kinerja model pada tiga kelas yaitu negatif, netral, dan positif. Dapat dilihat bahwa model memiliki performa yang cukup baik dalam memprediksi kelas positif dengan akurasi 271 dari 300. Namun, masih ada kesalahan prediksi yang signifikan, terutama pada kelas negatif dan netral. Misalnya, 344 yang sebenarnya negatif diprediksi dengan benar, tetapi 21 diprediksi salah sebagai netral dan 17 diprediksi salah sebagai positif. Sementara itu, 29 yang sebenarnya netral diprediksi dengan benar, tetapi 31 diprediksi salah sebagai negatif dan 20 diprediksi salah sebagai positif.



Gambar 3 Kurva Akurasi dan Loss 0.002-LSTM-100-512

Pada gambar 9 grafik akurasi per epoch menunjukkan perkembangan akurasi model selama pelatihan dan validasi. Dapat dilihat bahwa akurasi pelatihan dan validasi semakin meningkat selama epochs, menunjukkan bahwa model semakin akurat dalam memprediksi data yang belum pernah dilihat sebelumnya. Akurasi pelatihan mencapai nilai yang cukup tinggi,



sedangkan akurasi validasi fluktuasi sedikit namun umumnya meningkat. Sementara Grafik loss per epoch menunjukkan perkembangan loss model selama pelatihan dan validasi. Dapat dilihat bahwa loss pelatihan dan validasi semakin menurun selama epochs, menunjukkan bahwa model semakin baik dalam mengurangi kesalahan prediksi. Loss pelatihan turun drastis dari epoch pertama hingga epoch ke-14, sedangkan loss validasi fluktuasi sedikit namun umumnya menurun.

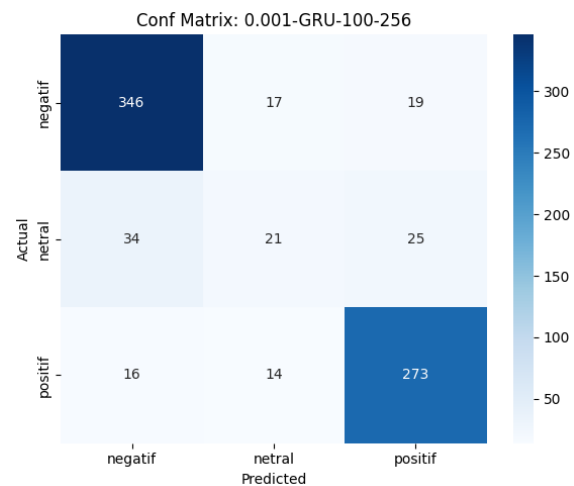
### 3.2 HASIL ANALISIS MODEL GRU

Seperti pada model LSTM, model **GRU** juga diuji menggunakan 12 konfigurasi hyperparameter. Hasil terbaik diperoleh pada skenario **0.001-GRU-100-256**, yang memberikan performa paling optimal dalam klasifikasi sentimen.

Hasil pengujian model GRU dapat dijelaskan sebagai berikut:

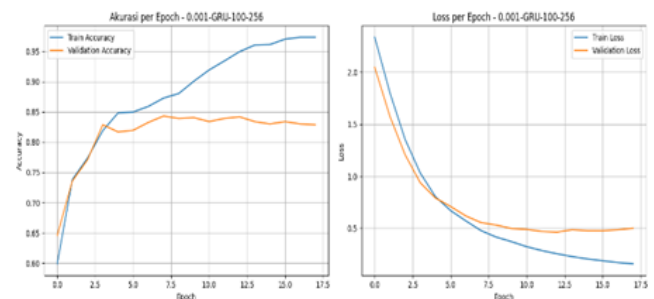
1. Akurasi tertinggi model 0.001-GRU-100-256 mencapai akurasi 0.837, yang berarti 83,7% data uji diprediksi dengan benar.
2. Presisi tertinggi skenario yang sama juga menghasilkan presisi 0.713, menandakan kemampuan model dalam mengurangi *false positive*.
3. *Recall* tertinggi nilai recall tertinggi 0.690 juga dimiliki oleh konfigurasi ini, yang berarti model mampu mengenali data yang benar dari setiap kelas dengan baik.
4. *F1-score* tertinggi terbaik sebesar 0.696, menunjukkan keseimbangan yang baik antara presisi dan *recall*, meskipun nilainya masih di bawah model LSTM terbaik.

Confusion matrix menunjukkan kinerja model pada tiga kelas yaitu negatif, netral, dan positif. Model menunjukkan performa yang cukup baik dalam memprediksi kelas positif dengan akurasi 273 dari 300. Namun, terdapat kesalahan prediksi yang cukup signifikan pada kelas negatif dan netral. Contohnya, 346 yang sebenarnya negatif diprediksi dengan benar, tetapi 17 diprediksi salah sebagai netral dan 19 diprediksi salah



Gambar 4 *Confusion Matrix* (0.001-GRU-100-256)

sebagai positif. Sementara itu, 21 yang sebenarnya netral diprediksi dengan benar, tetapi 34 diprediksi salah sebagai negatif dan 25 diprediksi salah sebagai positif.



Gambar 5 Kurva Akurasi & Loss (0.001-GRU-100-256)

Pada gambar 11 Grafik akurasi per epoch menunjukkan perkembangan akurasi model selama pelatihan dan validasi. Dapat dilihat bahwa akurasi pelatihan dan validasi semakin meningkat selama epochs, menunjukkan bahwa model semakin akurat dalam memprediksi data yang belum pernah dilihat sebelumnya. Akurasi pelatihan mencapai nilai yang cukup tinggi, sedangkan akurasi validasi fluktuasi sedikit namun umumnya meningkat. Sementara Grafik loss per epoch menunjukkan perkembangan loss model selama pelatihan dan validasi. Dapat dilihat bahwa loss pelatihan dan validasi semakin menurun selama epochs, menunjukkan bahwa model semakin baik dalam mengurangi kesalahan prediksi. Loss pelatihan turun drastis dari epoch pertama hingga epoch ke-17, sedangkan loss

validasi fluktuasi sedikit namun umumnya menurun.

### 3.3 Hasil Analisis Model Terbaik LSTM dan GRU

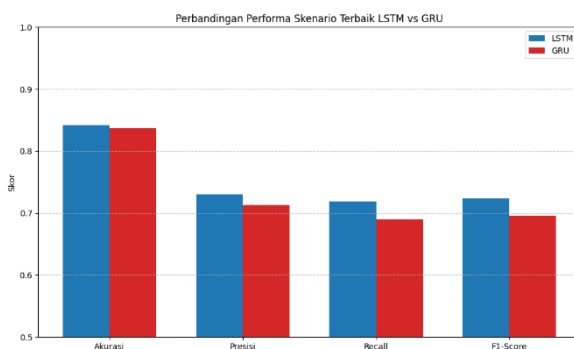
Hasil evaluasi Dari hasil yang sudah dilakukan, diperoleh ringkasan perbandingan performa terbaik pada Tabel 3.

Tabel 3 Performa Skenario Terbaik LSTM vs GRU

Matrix	LSTM (0.002-LSTM-100-512)	GRU (0.001-GRU-100-256)
Akurasi	0.842	0.837
Presisi	0.730	0.713
Recall	0.719	0.690
F1-Score	0.724	0.696

Berdasarkan tabel tersebut:

1. LSTM unggul di semua metrik evaluasi, meskipun selisih akurasi dan presisi relatif kecil.
2. *F1-score* menjadi indikator utama dalam penelitian ini, dan LSTM menunjukkan keseimbangan yang lebih baik antara presisi dan recall.
3. GRU tetap memberikan hasil kompetitif dan dapat menjadi alternatif model jika efisiensi waktu pelatihan menjadi pertimbangan.



Gambar 6 Diagram Batang Performa Skenario Terbaik LSTM vs GRU

Pada Gambar 12 diagram batang menggambarkan perbedaan nilai metrik antar model. Performa LSTM tampak lebih tinggi pada semua metrik dibandingkan GRU, menegaskan bahwa LSTM lebih unggul dalam mengklasifikasikan sentimen ulasan pengguna aplikasi media sosial X.

## IV. KESIMPULAN DAN SARAN

### 4.1 KESIMPULAN

Berdasarkan hasil analisis terhadap 5.100 data ulasan pengguna aplikasi media sosial X yang telah diberi label secara manual, ditemukan bahwa pola sentimen pengguna mencerminkan keberagaman reaksi terhadap perubahan dan pembaruan pada aplikasi tersebut. Distribusi hasil prediksi sentimen oleh model LSTM terbaik dengan skenario (0.002-LSTM-100-512) menunjukkan sentimen negatif sebesar 49,3% mendominasi klasifikasi model dan menunjukkan banyaknya ulasan bernada keluhan atau ketidakpuasan. Sentimen positif sebesar 41,6%, menunjukkan bahwa cukup banyak pengguna memberikan ulasan apresiasi terhadap aplikasi. Sementara sentimen netral sebesar 9,2% menunjukkan bahwa model cenderung lebih sulit mengenali ekspresi netral yang biasanya bersifat ambigu.

Proses klasifikasi sentimen dilakukan dengan menggunakan dua model *deep learning*, yaitu LSTM dan GRU, dengan total 12 skenario pengujian *hyperparameter* untuk masing-masing algoritma. Data dibagi menjadi 70% pelatihan, 15% validasi, dan 15% pengujian, dengan *preprocessing* menyeluruh termasuk pembersihan teks, *lowercase*, tokenisasi, *stopword removal*, dan *stemming*. Ketiga kelas sentimen (positif, negatif, netral) diolah secara seimbang sehingga model dapat mengidentifikasi sentimen.

Hasil penelitian ini menunjukkan bahwa model LSTM unggul pada penelitian ini dibandingkan GRU. Model LSTM terbaik pada skenario (0.002-LSTM-100-512) memperoleh akurasi 84,2%, Presisi: 73,0%, *Recall*: 71,9%, dan *F1-score*: 72,4%. Sementara model GRU terbaik pada skenario (0.001-GRU-100-256) memperoleh akurasi 83,7%, Presisi 71,3%, *Recall* 69,0%, dan *F1-score* 69,6%. Berdasarkan hasil penelitian, maka saran-saran yang dapat diberikan diantaranya sebagai berikut:

1. Disarankan untuk dilakukan pelabelan secara manual untuk meningkatkan akurasi dalam menilai dataset yang diperoleh untuk meminimalisir kesalahan dalam persepsi



kalimat ulasan, diperlukan juga dataset yang lebih banyak untuk meningkatkan performa dan kemampuan generalisasi model terhadap berbagai ekspresi sentimen, semakin banyak dataset semakin baik performas model bekerja.

2. Untuk meningkatkan performa yang lebih baik penelitian mendatang dapat mengkaji penggunaan model seperti *Bidirectional LSTM* (BiLSTM), *Bidirectional GRU* (BiGRU) atau model berbasis transformer seperti BERT, yang lebih unggul dalam berbagai tugas NLP, termasuk analisis sentimen.

#### DAFTAR PUSTAKA

- [1] Abdurrohman, I., & Rahman, A. (N.D.). *Penerapan Natural Language Processing Untuk Analisis Sentimen Terhadap Kebijakan Pemerintah*.
- [2] Akbar, Y., & Sugiharto, T. (2023). Analisis Sentimen Pengguna Twitter Di Indonesia Terhadap Chatgpt Menggunakan Algoritma C4.5 Dan Naïve Bayes (Yuma Akbar 1\*, Tri Sugiharto 2 ) Analisis Sentimen Pengguna Twitter Di Indonesia Terhadap Chatgpt Menggunakan Algoritma C4.5 Dan Naïve Bayes. *Jurnal Sains Dan Teknologi*, 5(1), 115–122. <https://doi.org/10.55338/Saintek.V4i3.1368>
- [3] Aprian Putra Dan, J., & Budi, A. (N.D.). *Penerapan Natural Language Processing Dalam Aplikasi Chatbot Sebagai Media Pencarian Informasi Dengan Menggunakan React (Studi Kasus: Institut Bisnis Dan Informatika Kwik Kian Gie)*.
- [4] Aslam, N., Rustam, F., Lee, E., Washington, P. B., & Ashraf, I. (2022). Sentiment Analysis And Emotion Detection On Cryptocurrency Related Tweets Using Ensemble Lstm-Gru Model. *Ieee Access*, 10, 39313–39324. <https://doi.org/10.1109/Access.2022.3165621>
- [5] Atharsyah, M., Romli, M. A., Utara, J. R., Sleman, K., & Yogyakarta, I. (2024). Implementasi Model Lstm, Gru, Bilstm, Dan Bigru Dalam Prediksi Harga Nikel. In *Jurnal Informatika & Rekayasa Elektronika* (Vol. 7, Issue 2). <http://E-Journal.Stmiklombok.Ac.Id/Index.Php/Jireissn.2620-6900>
- [6] Ayuningtyas, P., & Khomsah, S. (2024). Pelabelan Sentimen Berbasis Semi-Supervised Learning Menggunakan Algoritma Lstm Dan Gru. In *Jurnal Informatika Sunan Kalijaga* (Vol. 9, Issue 3).
- [7] Damaris Y. Koli, Jems Arison Zacharias, & Joni Kanisius Olin. (2023). Pemanfaatan Limbah Dedaunan Kering Sebagai Sumber Pendapatan Pasca Bencana Siklon Seroja Pada Rumah Tangga Di Koordinator Wilayah V Jemaat Gmit Nazaret Oesapa Timur, Kelurahan Oesapa. *Cakrawala: Jurnal Pengabdian Masyarakat Global*, 2(3), 17–22. <https://doi.org/10.30640/Cakrawala.V2i3.1280>
- [8] Dian Sudiantini, Mayang Puspita Ayu, Muhammad Cheirnel All Shawirdra Aswan, Meyliana Alifah Prastuti, & Melani Apriliya. (2023). Transformasi Digital : Dampak, Tantangan, Dan Peluang Untuk Pertumbuhan Ekonomi Digital. *Trending: Jurnal Manajemen Dan Ekonomi*, 1(3), 21–30. <https://doi.org/10.30640/Trending.V1i3.1115>
- [9] Eskiyaturrofikoh, E., & Suryono, R. R. (2024a). Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (Svm). *Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(3), 1408–1419. <https://doi.org/10.29100/Jipi.V9i3.5392>
- [10] Eskiyaturrofikoh, E., & Suryono, R. R. (2024b). Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (Svm). *Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(3), 1408–1419. <https://doi.org/10.29100/Jipi.V9i3.5392>
- [11] Handayani, S. F., Pratiwi, R. W., Dairoh, D., & Af'idah, D. I. (2022a). Analisis Sentimen Pada Data Ulasan Twitter Dengan Long-Short Term Memory. *Jtera (Jurnal Teknologi Rekayasa)*, 7(1), 39. <https://doi.org/10.31544/Jtera.V7.I1.2022.39-46>
- [12] Handayani, S. F., Pratiwi, R. W., Dairoh, D., & Af'idah, D. I. (2022b). Analisis Sentimen Pada Data Ulasan Twitter Dengan Long-Short Term Memory. *Jtera (Jurnal Teknologi Rekayasa)*, 7(1), 39. <https://doi.org/10.31544/Jtera.V7.I1.2022.39-46>
- [13] Hendrawan, A., & Sela, E. I. (2024). Analisis Sentimen Komentar Youtube Tentang Resesi Global 2023 Menggunakan Lstm. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, 5(1), 587–593. <https://doi.org/10.35870/Jimik.V5i1.526>
- [14] Ikhsan, M., Sri, D., & Wahyuni, N. (2023). Analisis Sentimen Terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Long Short-Term Memory. *Ijcsr: The Indonesian Journal Of Computer Science Research*, 1(1). <https://doi.org/10.37905>
- [15] Nurdin, A., Anggo, B., Aji, S., Bustamin, A., & Abidin, Z. (2020). Perbandingan Kinerja Word Embedding Word2vec, Glove, Dan Fasttext Pada Klasifikasi Teks. *Jurnal Teknokompak*, 14(2), 74.
- [16] Nurhakiki, J., Yahfizham, Y., William, J., Ps, I. V, Estate, M., Percut, K., Tuan, S., & Serdang, K. D.

- (2024). Studi Kepustakaan: Pengenalan 4 Algoritma Pada Pembelajaran Deep Learning Beserta Implikasinya. *Jurnal Pendidikan Berkarakter*, 1, 270–281.  
<https://doi.org/10.51903/Pendekar.V2i1.598>
- [17] Putera Khano, M. N. A., Saputro, D. R. S., Sutanto, S., & Wibowo, A. (2023). Sentiment Analysis With Long-Short Term Memory (Lstm) And Gated Recurrent Unit (Gru) Algorithms. *Barekeng: Jurnal Ilmu Matematika Dan Terapan*, 17(4), 2235–2242.  
<https://doi.org/10.30598/Barekengvol17iss4pp2235-2242>
- [18] Raup, A., Ridwan, W., Khoeriyah, Y., Yuliati Zaqiah, Q., & Islam Negeri Sunan Gunung Djati Bandung, U. (N.D.). *Deep Learning Dan Penerapannya Dalam Pembelajaran*.  
<http://jiip.stkipyapisdempu.ac.id>
- [20] Rolangon, A., Weku, A., Sandag, G. A., Mononutu, J. A., Utara, A.-M., Utara, S., Kunci, K., & Twitter, : (N.D.). *Perbandingan Algoritma Lstm Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Rumah Sakit Saat Pandemi Covid-19 The Comparison Of Lstm Algorithms For Twitter User Sentiment Analysis On Hospital Services During The Covid-19 Pandemic*.
- [21] Suarna, N., & Prihartono, W. (2024). Penerapan Nlp (Natural Language Processing) Dalam Analisis Sentimen Pengguna Telegram Di Playstore. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 2).
- [22] Utomo Rahino Putro, D., Maulidina, A., Atthoriq Kuswirasatya, W., & Puspita Sari, A. (2024). *Analisis Sentimen Publik pada “X” Menggunakan Long Short-Term Memory Neural Network* (Vol. 5).  
[www.elektro.itn.ac.id](http://www.elektro.itn.ac.id)
- [23] Wijayanto, I. R., Cholissodin, I., & Sari, Y. A. (2021). *Pengaruh Metode Word Embedding dalam Vector Space Model pada Pemerolehan Informasi Materi IPA Siswa SMP* (Vol. 5, Issue 3). <http://j-ptiik.ub.ac.id>