

**ANALISIS KINERJA MODEL LONG SHORT-TERM
MEMORY (LSTM) UNTUK KLASIFIKASI EMOSI PADA
TEKS BERBAHASA INDONESIA**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

PUTRI YAHNDA ALHA ZHALIANTI

22.11.5353

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS KINERJA MODEL LONG SHORT-TERM
MEMORY (LSTM) UNTUK KLASIFIKASI EMOSI PADA
TEKS BERBAHASA INDONESIA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

PUTRI YAHNDA ALHA ZHALIANTI

22.11.5353

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS KINERJA MODEL LONG SHORT-TERM
MEMORY (LSTM) UNTUK KLASIFIKASI EMOSI PADA
TEKS BERBAHASA INDONESIA**

yang disusun dan diajukan oleh

Putri Yahnda Alha Zhalianti

22.11.5353

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Januari 2026

Dosen Pembimbing,


Mulia Sulistivono, S. Kom., M. Kom.
NIK. 190302248

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS KINERJA MODEL LONG SHORT-TERM
MEMORY (LSTM) UNTUK KLASIFIKASI EMOSI PADA
TEKS BERBAHASA INDONESIA

yang disusun dan diajukan oleh

Putri Yahnda Alha Zhalianti

22.11.5353

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ali Mustopa, S.Kom., M.Kom.
NIK. 190302192

Ria Andriani, S.Kom., M.Kom.
NIK. 190302458

Mulia Sulistiyono, S.Kom., M.Kom.
NIK. 190302248



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini,

Nama mahasiswa : Putri Yahnda Alha Zhalianti
NIM : 22.11.5353

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Kinerja Model Long Short-Term Memory (LSTM) untuk Klasifikasi Emosi Pada Teks Berbahasa Indonesia

Dosen Pembimbing : Mulia Sulistiyono, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 Januari 2026

Yang Menyatakan,



The image shows a handwritten signature in black ink over a rectangular stamp. The stamp contains the text '10000' at the top, a Garuda emblem in the center, and 'METERAI TERPAJ' below it. At the bottom of the stamp, the alphanumeric code 'VER12ANX2554159 90' is printed.

Putri Yahnda Alha Zhalianti

HALAMAN PERSEMBAHAN

Dengan ini, saya persembahkan skripsi ini dengan penuh syukur kepada orang-orang serta pihak-pihak yang telah menjadi bagian penting dalam perjalanan saya untuk menyelesaikan skripsi ini. Persembahan ini merupakan ungkapan rasa terima kasih penulis atas segala dukungan dan doa yang dipanjatkan untuk penulis. Oleh karena itu, penelitian ini saya persembahkan kepada:

1. Allah SWT, terima kasih untuk kelancaran dan kemudahan-Nya, sehingga saya mampu menyelesaikan penulisan skripsi ini.
2. Kedua orang tua saya, Bapak Piter Setiawan dan Ibu Maryani Muput, terima kasih karena selalu memberikan doa, kasih sayang, serta dukungan tanpa henti.
3. Dosen Pembimbing saya, Bapak Mulia Sulistiyono, Skom., M.Kom., yang telah dengan sabar meluangkan waktu untuk memberikan arahan, bimbingan, serta motivasi selama penyusunan skripsi ini.
4. Saudara kandung saya, Revana Zahra Mahfuzah, Revina Zahra Mahfuzah, Azham Mikhael Azzikhra, Varisha Shareen Setiawan, serta tante saya, Rizki Amalia, dan keluarga besar yang tidak dapat disebutkan satu per satu. Terima kasih telah memberikan dukungan dan motivasi sepanjang perjalanan ini.
5. Nana Syarifuddin, Aisyah, Amalia Fitria Nurazizah, Puspitasari Anisa Putri, Yasmin Kartika Iswari, Heni Sapta Mawar, dan Vita Suci Hidayah, yang telah menjadi *support system* terbaik, serta teman untuk berdiskusi dan menjadi pendengar yang baik selama proses ini.
6. Teman-teman kelas IF13 dan seluruh rekan satu angkatan, atas kebersamaan dan saling mendukung selama masa kuliah.
7. Terakhir, untuk diri saya sendiri. Terima kasih karena telah berhasil melalui ini semua. Terima kasih sudah berusaha dan bertahan hingga akhir dan selalu menyelesaikan apa yang sudah dimulai.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul **"Analisis Kinerja Model Long Short-Term Memory (LSTM) untuk Klasifikasi Emosi Pada Teks Berbahasa Indonesia"** dengan baik. Penulisan skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Saya menyadari bahwa dalam proses penyusunan skripsi ini, penulis banyak mendapatkan bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi Informatika.
4. Bapak Mulia Sulistiyono, S.Kom., M.Kom., selaku Dosen Pembimbing yang telah memberikan bimbingan, arahan serta masukan bagi penulis hingga skripsi ini dapat terselesaikan dengan baik.
5. Seluruh Dosen dan Staf Universitas AMIKOM Yogyakarta yang telah memberikan bekal ilmu pengetahuan dan pelayanan selama penulis menempuh pendidikan.
6. Kedua orang tua atas segala dukungan, doa, motivasi dan kasih sayang yang diberikan tanpa putus, baik secara langsung maupun tidak langsung.

Saya menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun sangat saya harapkan demi perbaikan di masa mendatang. Semoga karya ini dapat memberikan manfaat bagi pembaca dan perkembangan ilmu pengetahuan.

Yogyakarta, 20 Januari 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	10

2.2.1	Emosi	10
2.2.2	<i>Anger</i> (Marah).....	10
2.2.3	<i>Sad</i> (Sedih).....	10
2.2.4	<i>Fear</i> (Takut).....	10
2.2.5	<i>Joy</i> (Senang).....	11
2.2.6	<i>Love</i> (Cinta)	11
2.2.7	Klasifikasi	11
2.2.8	<i>Natural Language Processing (NLP)</i>	11
2.2.9	<i>Deep Learning</i>	12
2.2.10	<i>Long Short-Term Memory (LSTM)</i>	12
2.2.11	<i>Confusion Matrix</i>	14
BAB III METODE PENELITIAN		16
3.1	Objek Penelitian.....	16
3.2	Alur Penelitian	17
3.2.1	Pengumpulan Data	18
3.2.2	Data Preprocessing.....	19
3.2.3	Representasi Teks	19
3.2.4	Label Encoding	20
3.2.5	Split Data	20
3.2.6	Pembuatan Model	21
3.2.7	Evaluasi Model	21
3.3	Alat dan Bahan.....	22
3.3.1	Data Penelitian	22
3.3.2	Alat/instrumen.....	22
BAB IV HASIL DAN PEMBAHASAN		23

4.1	Pengumpulan Data	23
4.2	<i>Preprocessing</i> Data	24
4.2.1	Menghapus Data Duplikat	24
4.2.2	<i>Cleaning</i>	25
4.2.3	Normalisasi	26
4.2.4	<i>Tokenizing</i>	27
4.2.5	<i>Stopword Removal</i>	27
4.3	Visualisasi Data Teks	28
4.3.1	Visualisasi <i>WordCloud</i> Keseluruhan	28
4.3.2	Visualisasi <i>WordCloud</i> Per Kelas	29
4.4	Representasi Teks	30
4.4.1	<i>Tokenization</i>	31
4.4.2	<i>Padding</i>	31
4.5	<i>Label Encoding</i>	31
4.6	Pembagian Data	32
4.7	LSTM Model	33
4.8	Evaluasi Model	34
BAB V PENUTUP		37
5.1	Kesimpulan	37
5.2	Saran	38
REFERENSI		39

DAFTAR TABEL

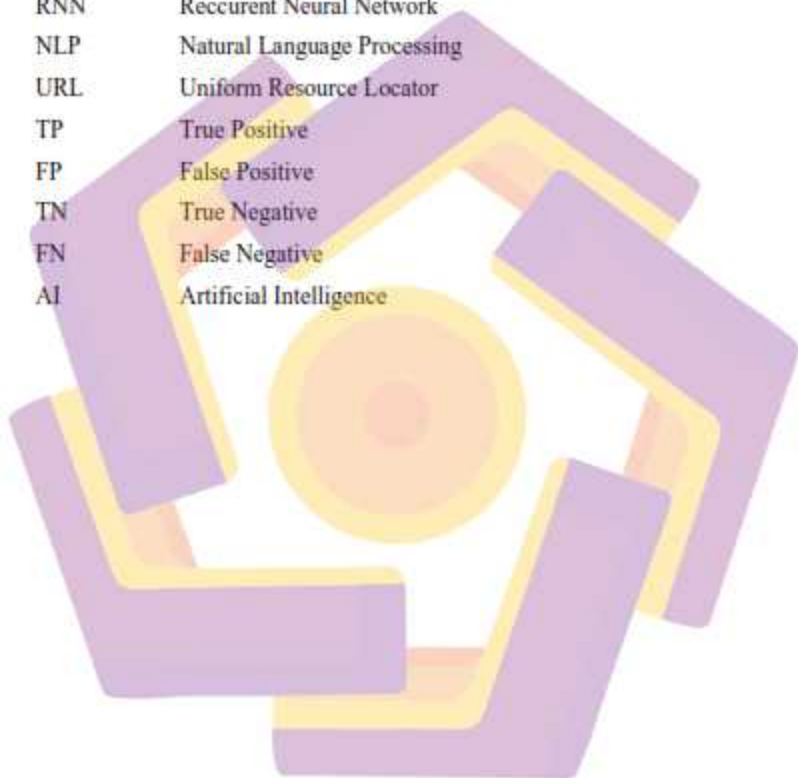
Tabel 2. 1 Keaslian Penelitian.....	7
Tabel 2. 2 Confusion Matrix.....	14
Tabel 4. 1 Rincian Sumber Data.....	23
Tabel 4. 2 Sampel Data Mentah.....	23
Tabel 4. 3 Hasil Cek Data Duplikat.....	24
Tabel 4. 4 Hasil Cleaning.....	26
Tabel 4. 5 Hasil Normalisasi.....	26
Tabel 4. 6 Hasil Tokenizing.....	27
Tabel 4. 7 Hasil Stopword Removal.....	28
Tabel 4. 8 Hasil Label Encoding.....	32
Tabel 4. 9 Classification Report.....	34

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur LSTM.....	13
Gambar 3. 1 Alur Penelitian	17
Gambar 4. 1 Source Code Cek Data Duplikat	24
Gambar 4. 2 Source Code Mengatasi Data Duplikat.....	25
Gambar 4. 3 Hasil Mengatasi Data Duplikat.....	25
Gambar 4. 4 Source Code Normalisasi.....	26
Gambar 4. 5 Source Code Cleaning.....	25
Gambar 4. 6 Source Code Tokenizing	27
Gambar 4. 7 Source Code Stopword Removal.....	28
Gambar 4. 8 WordCloud Keseluruhan.....	29
Gambar 4. 9 WordCloud Per Kelas	29
Gambar 4. 10 Source Code dan Output Tokenization dan Padding	30
Gambar 4. 11 Source Code Label Encoding.....	31
Gambar 4. 12 Source Code Pembagian Data.....	32
Gambar 4. 13 Source Code LSTM	34
Gambar 4. 14 Confusion Matrix.....	35

DAFTAR LAMBANG DAN SINGKATAN

LSTM	Long Short-Term Memory
BERT	Bidirectional Encoder Representations from Transformers
CNN	Convolutional Neural Network
RNN	Reccurent Neural Network
NLP	Natural Language Processing
URL	Uniform Resource Locator
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
AI	Artificial Intelligence



DAFTAR ISTILAH

<i>Accuracy</i>	Persentase jumlah prediksi benar dari seluruh tweet yang diuji oleh model
<i>Classification Report</i>	Laporan evaluasi yang menyajikan metrik utama dalam klasifikasi untuk menilai kinerja model
<i>Confusion Matrix</i>	<i>Matrix</i> yang menunjukkan performa model klasifikasi yang membandingkan nilai prediksi model dengan nilai sebenarnya (<i>ground truth</i>) dari data uji.
<i>Cleaning</i>	Tahap <i>preprocessing</i> untuk mendeteksi dan memperbaiki atau menghapus data yang rusak, tidak akurat, atau tidak relevan dari sebuah dataset, seperti tanda baca, angka, username, URL, dan simbol.
<i>Deep Learning</i>	Sub-bidang <i>Machine Learning</i> yang menggunakan jaringan saraf tiruan berlapis untuk memodelkan pola kompleks dalam data.
<i>Dropout</i>	Teknik regularisasi dalam jaringan saraf tiruan untuk mencegah <i>overfitting</i> dengan cara menonaktifkan sejumlah neuron secara acak selama proses pelatihan.
<i>Embedding Layer</i>	Lapisan dalam model jaringan saraf yang berfungsi mengubah kata-kata menjadi vektor numerik kontinu
<i>F1-Score</i>	Rata-rata harmonis antara <i>Precision</i> dan <i>Recall</i> , digunakan sebagai metrik evaluasi ketika dataset tidak seimbang.
<i>False Negative</i>	Kesalahan prediksi di mana model memprediksi data sebagai kelas negatif, padahal sebenarnya data tersebut adalah positif.
<i>False Positive</i>	Kesalahan prediksi di mana model memprediksi data sebagai kelas positif, padahal sebenarnya data tersebut adalah negatif.
<i>Label Encoding</i>	Teknik mengubah data kategori atau label menjadi format angka numerik agar dapat diproses oleh algoritma komputer.

<i>Normalisasi</i>	Proses penyeragaman teks, seperti mengubah bahasa gaul (slang) menjadi bahasa baku, serta memperbaiki singkatan.
<i>Overfitting</i>	Kondisi di mana model terlalu mempelajari detail dan <i>noise</i> pada data latih hingga berdampak negatif pada performanya saat menguji data baru.
<i>Padding</i>	Proses penambahan nilai pada data sekuens agar semua input memiliki panjang yang sama sebelum diproses oleh model.
<i>Precision</i>	Tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem (rasio prediksi benar positif terhadap total prediksi positif).
<i>Preprocessing</i>	Tahapan awal pengolahan data mentah untuk mengubahnya menjadi format yang lebih bersih dan terstruktur sebelum masuk ke tahap pemodelan.
<i>Recall</i>	Tingkat keberhasilan sistem dalam menemukan kembali informasi (rasio prediksi benar positif terhadap total data yang sebenarnya positif).
<i>Stopword Removal</i>	Proses penghapusan kata-kata umum yang sering muncul namun dianggap tidak memiliki makna penting dalam analisis sentimen seperti "yang", "dan", "di".
<i>Tokenizing</i>	Proses pemecahan kalimat atau dokumen teks menjadi potongan-potongan kecil yang disebut token.
<i>True Negative</i>	Keadaan ketika model memprediksi kelas negatif dengan benar sesuai dengan kenyataan.
<i>True Positive</i>	Keadaan ketika model memprediksi kelas positif dengan benar sesuai dengan kenyataan.
<i>Web Scrapping</i>	Teknik pengambilan data secara otomatis dari situs web (seperti media sosial) untuk dikumpulkan dan dianalisis.
<i>WordCloud</i>	Metode visualisasi data teks di mana ukuran kata merepresentasikan frekuensi kemunculannya dalam dokumen; semakin sering muncul, semakin besar ukurannya.

INTISARI

Penelitian ini bertujuan untuk menganalisis kinerja model Long Short-Term Memory (LSTM) dalam mengklasifikasikan emosi pada teks berbahasa Indonesia ke dalam lima kategori, yaitu Anger, Sad, Fear, Joy, dan Love. Dataset yang digunakan merupakan gabungan dari sumber GitHub, Kaggle, Twitter/X, dan YouTube dengan total 10.667 data teks yang mencakup variasi bahasa formal dan informal. Proses penelitian meliputi tahap preprocessing seperti pembersihan data, normalisasi bahasa gaul, dan stopword removal sebelum data diolah oleh model LSTM. Hasil pengujian menunjukkan bahwa model LSTM berhasil mencapai tingkat akurasi sebesar 81%, dengan performa terbaik pada klasifikasi emosi Love (F1-Score 0,83) dan performa terendah pada emosi Sad (F1-Score 0,77). Temuan ini menunjukkan bahwa model LSTM memiliki stabilitas yang baik dalam mengenali pola emosi pada teks sekuensial berbahasa Indonesia, meskipun masih terdapat ambiguitas prediksi antara kategori emosi negatif.

Kata kunci: Klasifikasi Emosi, *LSTM*, *Text Mining*, *Natural Language Processing*, Bahasa Indonesia.

ABSTRACT

This research aims to analyze the performance of the Long Short-Term Memory (LSTM) model for emotion classification in Indonesian text, categorized into five classes: Anger, Sad, Fear, Joy, and Love. The study utilizes a combined dataset of 10,667 text entries sourced from GitHub, Kaggle, Twitter/X, and YouTube, reflecting a diverse range of formal and informal linguistic contexts. The methodology involves comprehensive preprocessing steps, including data cleaning, slang normalization, and stopword removal, followed by model training using the LSTM architecture. The experimental results demonstrate that the LSTM model achieves an overall accuracy of 81%, with the highest performance observed in the Love category (F1-Score 0.83) and the lowest in the Sad category (F1-Score 0.77). These findings indicate that while the LSTM model is stable and effective for Indonesian emotion classification, challenges remain regarding predictive ambiguity between negative emotion categories.

Keyword: Emotion Classification, Indonesian Language, LSTM, Natural Language Processing, Text Mining