

**PREDIKSI KONDISI MENTAL PENGGUNA SOSIAL MEDIA
SOSIAL X PENGGUNA BERDASARKAN DATA TEKS
MENGUNAKAN LONG SHORT-TERM MEMORY**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

FAUZIAH GUSRI YASINTA

22.11.4932

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**PREDIKSI KONDISI MENTAL PENGGUNA SOSIAL MEDIA
SOSIAL X PENGGUNA BERDASARKAN DATA TEKS
MENGUNAKAN LONG SHORT-TERM MEMORY**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1 Informatika



disusun oleh

FAUZIAH GUSRI YASINTA

22.11.4932

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

HALAMAN PERSETUJUAN

SKRIPSI

**PREDIKSI KONDISI MENTAL PENGGUNA SOSIAL MEDIA
SOSIAL X PENGGUNA BERDASARKAN DATA TEKS
MENGUNAKAN LONG SHORT-TERM MEMORY**

yang disusun dan diajukan oleh

Fauziah Gusri Yasinta

22.11.4932

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 April 2026

Dosen Pembimbing,



I Made Artha Agastya. S. T., M.Eng., Ph.D.
NIK. 190302352

HALAMAN PENGESAHAN

SKRIPSI

**PREDIKSI KONDISI MENTAL PENGGUNA SOSIAL MEDIA
SOSIAL X PENGGUNA BERDASARKAN DATA TEKS
MENGUNAKAN LONG SHORT-TERM MEMORY**

yang disusun dan diajukan oleh

Fauziah Gusri Yasinta

22.11.4932

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 April 2026

Susunan Dewan Penguji

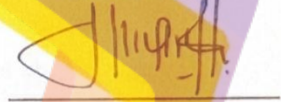
Nama Penguji

Dr. Hartatik, S.T., M.Cs.
NIK. 190302232

Yudha Riwanto, M.Kom.
NIK. 190302702

I Made Artha Agastya, S.T., M.Eng., Ph.D.
NIK. 190302352

Tanda Tangan



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

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fauziah Gusri Yasinta
NIM : 22.11.4932

Menyatakan bahwa Skripsi dengan judul berikut:

**PREDIKSI KONDISI MENTAL PENGGUNA SOSIAL MEDIA SOSIAL X
PENGGUNA BERDASARKAN DATA TEKS MENGGUNAKAN LONG SHORT-
TERM MEMORY**

Dosen Pembimbing : I Made Artha Agastya, S.T., M.Eng., Ph.D.

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 ketidakbenaran 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 April 2026

Yang Menyatakan,



Fauziah Gusri Yasinta

HALAMAN PERSEMBAHAN

Karya tulis ini saya persembahkan kepada:

1. Allah Yang Maha Penagshih dan Maha Penyayang.
2. Kedua orang tua saya selalu meridhoi dan mendoakan anaknya sehingga dapat menyelesaikan studinya.
3. Para dosen, terutama untuk Bapak I Made Artha Agastya, S.T., M.Eng., Ph.D., yang sudah membimbing dengan sabar dan tulus selama proses penyusunan skripsi.
4. Kakak saya, Galang Gusri Fernanda S, Kom., yang selalu menyempatkan waktu untuk mengingatkan saya dalam mengerjakan karya tulis ini.
5. Para sahabat yang selalu ada untuk memberi semangat dan menyertai dalam setiap tahap proses.

Terima kasih atas doa dan dukungan dari kalian semua.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana pada S1-Informatika di Universitas AMIKOM.

Dalam penyusunan skripsi ini, penulis menyadari bahwa banyak pihak yang telah memberikan bantuan, bimbingan, serta dukungan. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak I Made Artha Agastya, S.T., M.Eng., Ph.D., selaku dosen pembimbing yang telah memberikan arahan, bimbingan, dan motivasi selama proses penyusunan skripsi ini.
2. Para dosen penguji yang telah memberikan saran dan masukan yang membangun.
3. Seluruh dosen dan staf di Amikom yang telah memberikan ilmu dan bantuan selama masa perkuliahan.
4. Kedua orang tua tercinta yang selalu memberikan doa, dukungan, dan semangat kepada penulis.
5. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan pihak-pihak yang membutuhkan.

Yogyakarta, 20 April 2026

Penulis

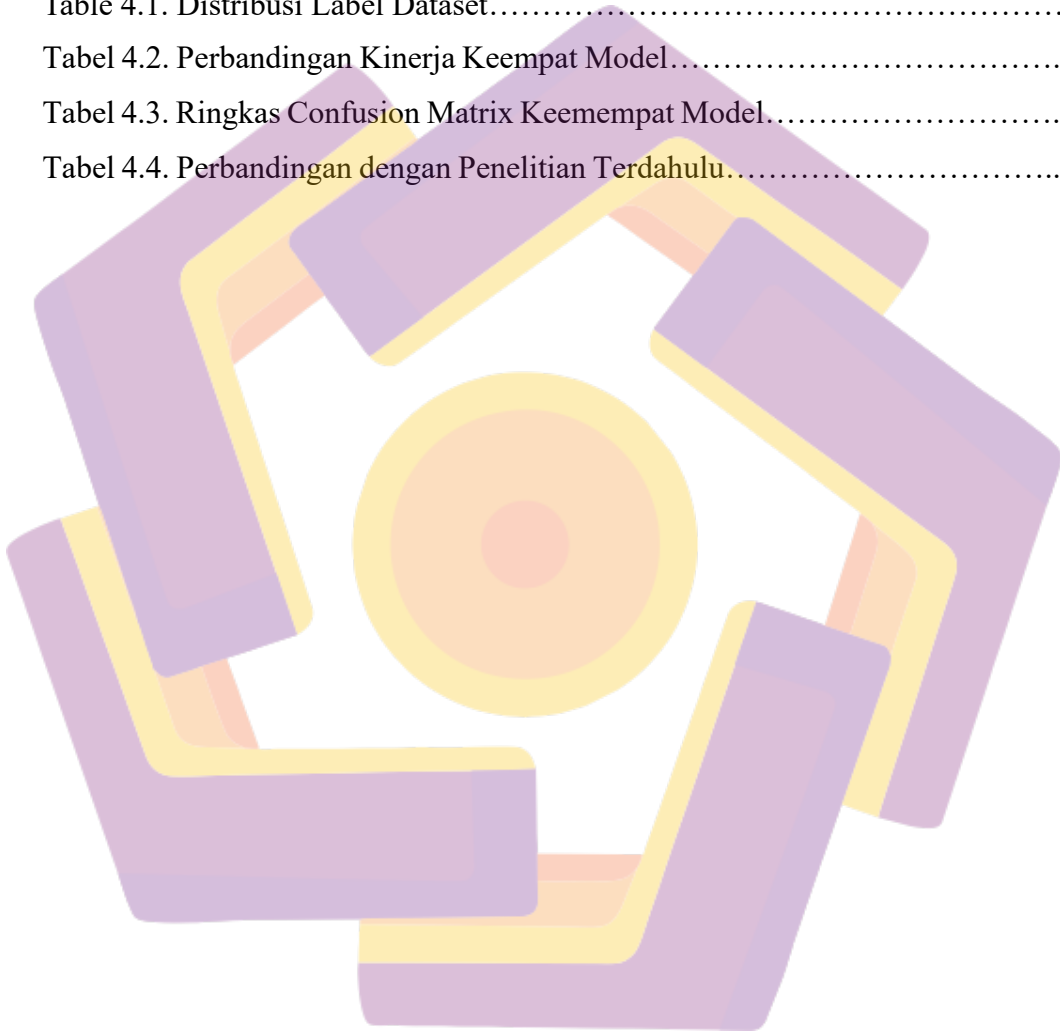
DAFTAR ISI

SKRIPSI.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xv
INTISARI	xviii
ABSTRACT.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
1.6 Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA	9
2.1 Studi Literatur	9
2.2 Dasar Teori	14
2.2.1 Kesehatan Mental dan Deteksi Dini.....	14
2.2.2. Media Sosial sebagai Sumber Data	15
2.2.3. Natural Language Processing (NLP) dan Pra-pemrosesan Teks.....	15
2.2.4. Word Embedding	17
2.2.5. Long Short-Term Memory (LSTM)	18
2.2.6. Bidirectional LSTM (BiLSTM)	19
2.2.7. Attention Mechanism	20
2.2.8. Multi-Head Attention.....	21
2.2.9. Metrik Evaluasi.....	22

BAB III METODE PENELITIAN	25
3.1 Objek Penelitian.....	25
3.2 Alur Penelitian	25
3.3 Alat dan Bahan Penelitian	43
3.3.1 Data Penelitian.....	43
3.3.2 Pre-trained Word Embedding Dolma 300D	44
3.3.3 Bahan Penelitian	45
BAB IV HASIL DAN PEMBAHASAN	49
4.1 Hasil Eksplorasi Data	49
4.1.1. Distribusi Label	49
4.1.2. Visualisasi WordCloud.....	50
4.1.3. Statistik Panjang Teks	51
4.1.4. Cakupan Pre-trained Embedding	52
4.2. Hasil Pelatihan Model	52
4.2.1. Model LSTM	53
4.2.2. Model BiLSTM	54
4.2.3. Model LSTM+Attention.....	55
4.2.4. Model LSTM+MultiHead	56
4.2.5. Perbandingan Antar Model	56
4.3. Evaluasi Kinerja Model.....	58
4.3.1. Analisis Metrik Evaluasi	58
4.3.2. Analisis Confusion Matrix	59
4.3.3. Ringkasan Temuan.....	62
4.4. Pembahasan	63
4.4.1. Analisis Keunggulan Model BiLSTM	63
4.4.2. Analisis Model LSTM Standar.....	64
4.4.3. Analisis Model LSTM+Attention.....	64
4.4.4. Analisis Model LSTM+MultiHead	65
4.4.5. Perbandingan dengan Penelitian Terdahulu	66
4.4.6. Keterbatasan Penelitian.....	67
BAB V PENUTUP	69
5.1 Kesimpulan	69
5.2 Saran	70
REFERENSI	73

DAFTAR TABEL

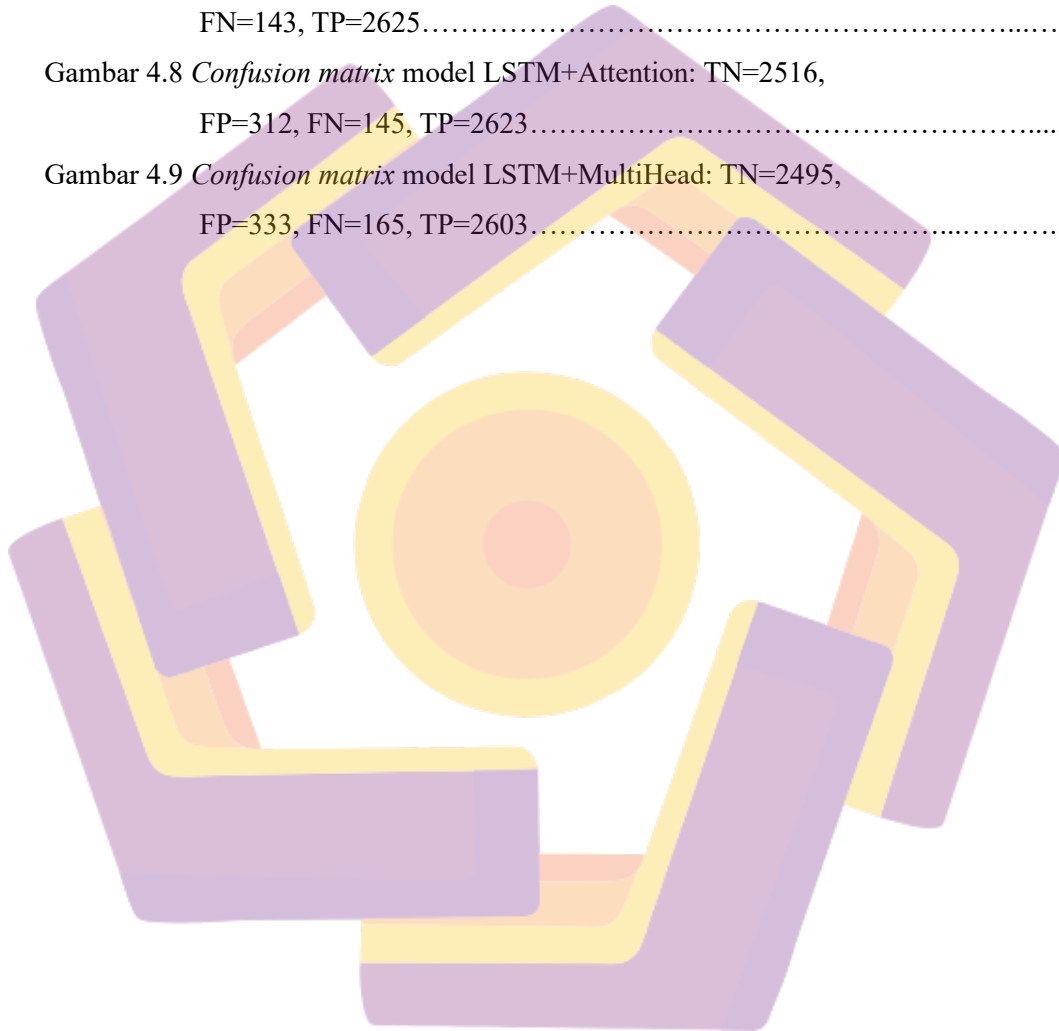
Tabel 2.1 Keaslian Penelitian.....	13
Table 3.1 Spesifikasi Perangkat Lokal	51
Table 3.2 Spesifikasi Lingkungan Komputasi Kaggle.....	51
Tabel 3.3 Parameter Penelitian.....	51
Table 4.1. Distribusi Label Dataset.....	53
Tabel 4.2. Perbandingan Kinerja Keempat Model.....	61
Tabel 4.3. Ringkas Confusion Matrix Keempat Model.....	62
Tabel 4.4. Perbandingan dengan Penelitian Terdahulu.....	70



DAFTAR GAMBAR

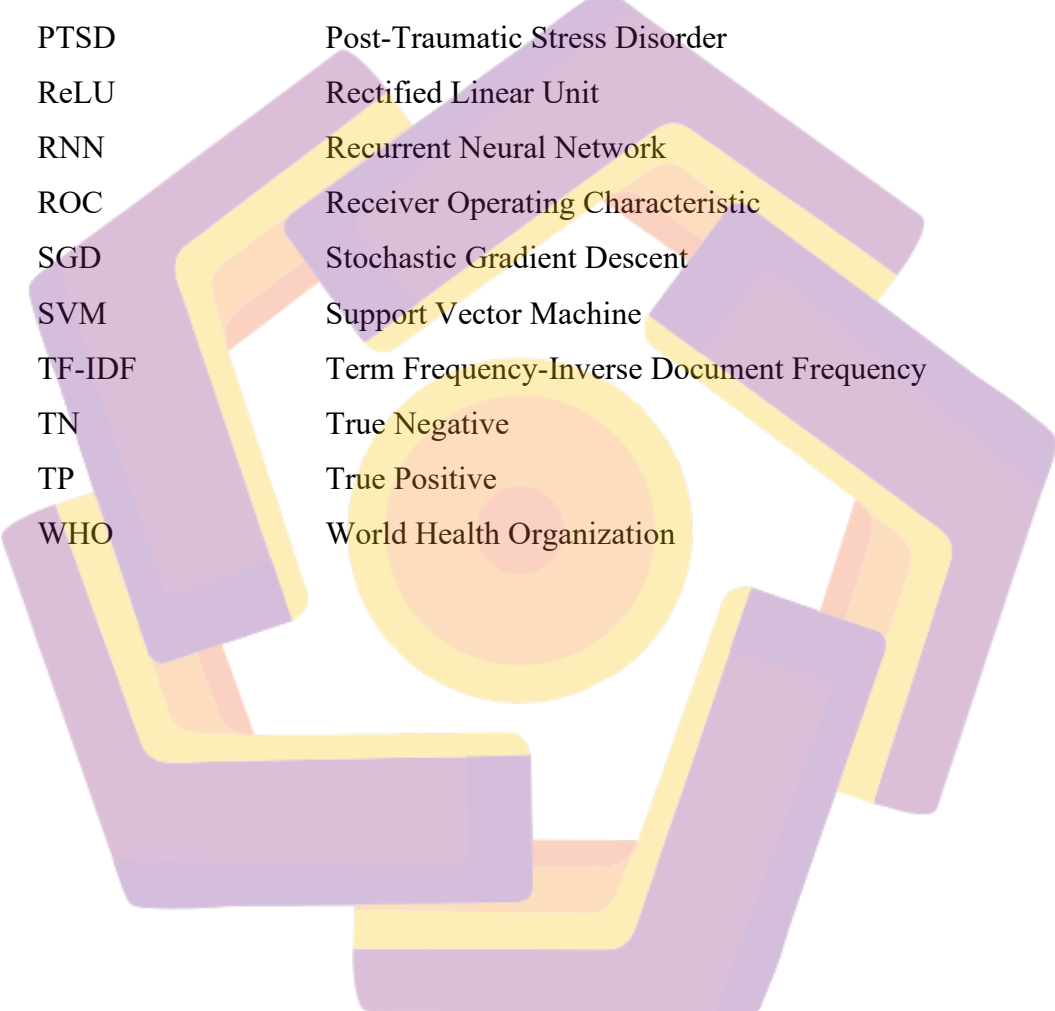
Gambar 2.1 Spektrum Kesehatan Mental: dari Fungsi Optimal Hingga Distres Berat.....	14
Gambar 2.2 Alur Analisis Sentimen pada Data Twitter (Media Sosial).....	15
Gambar 2.3 Tahapan Pra-pemrosesan Teks: dari Teks Mentah hingga Representasi Numerik.....	16
Gambar 2.4 Visualisasi Proyeksi Dua Dimensi dari Word Embedding dalam Ruang Vektor.....	17
Gambar 2.5 Arsitektur Sel Long Short-Term Memory (LSTM) dengan Tiga Gerbang Utama.....	18
Gambar 2.6 Arsitektur Bidirectional LSTM (BiLSTM) dengan Pemrosesan Dua Arah (Forward dan Backward).....	19
Gambar 2.7 Mekanisme Attention dengan Query, Keys, Values, dan Perhitungan Bobot.....	20
Gambar 2.8 Arsitektur Multi-Head Attention dengan Empat Kepala (Heads) yang Bekerja secara Paralel.....	22
Gambar 2.9 Confusion Matrix untuk Klasifikasi Biner (kiri) dan Kurva ROC dengan AUC (kanan).....	23
Gambar 3.1 Diagram Alir Penelitian dari Pengumpulan Data hingga Analisis dan Perbandingan Keempat Model	26
Gambar 3.2 Arsitektur Model LSTM Standar yang Digunakan dalam Penelitian	32
Gambar 3.3 Arsitektur Model Bidirectional LSTM (BiLSTM) yang Digunakan dalam Penelitian	33
Gambar 3.4 Arsitektur Model LSTM dengan Mekanisme Attention (LSTM+Attention) yang Digunakan dalam Penelitian	35
Gambar 3.5 Alur Penelitian dari Pengumpulan Data hingga Analisis dan Perbandingan Keempat Model Arsitektur LSTM.....	38
Gambar 4.1 Perbandingan kata dominan pada teks Kesehatan mental (label 1) dan teks umum (label 0).....	50
Gambar 4.2 Grafik loss dan akurasi model LSTM selama 50 epoch	53
Gambar 4.3 Grafik loss dan akurasi model BiLSTM selama 50 epoch	54

Gambar 4.4 Grafik loss dan akurasi model LSTM+Attention selama 50 epoch.....	55
Gambar 4.5 Grafik loss dan akurasi model LSTM+MultiHead selama 50 epoch.....	56
Gambar 4.6 <i>Confusion matrix</i> model LSTM: TN=2515, FP=313, FN=131, TP=2637.....	60
Gambar 4.7 <i>Confusion matrix</i> model BiLSTM: TN=2578, FP=250, FN=143, TP=2625.....	60
Gambar 4.8 <i>Confusion matrix</i> model LSTM+Attention: TN=2516, FP=312, FN=145, TP=2623.....	61
Gambar 4.9 <i>Confusion matrix</i> model LSTM+MultiHead: TN=2495, FP=333, FN=165, TP=2603.....	62



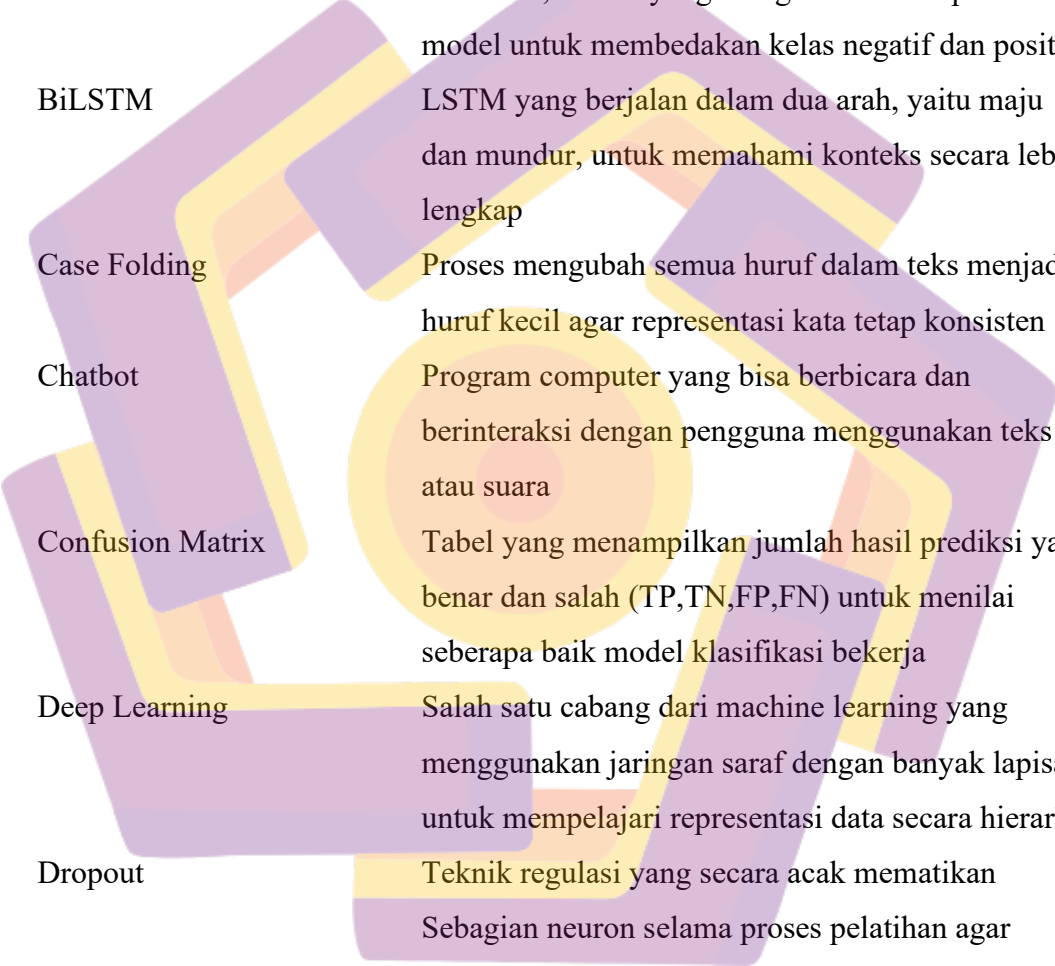
DAFTAR LAMBANG DAN SINGKATAN

a_t	Output kandidat memori pada LSTM pada waktu ke- t
b	Bias pada model <i>deep learning</i>
C	Confusion matrix
c_t	Sel memori pada LSTM pada waktu ke- t
d	Dimensi vektor <i>Embedding</i>
E	Metriks <i>Embedding</i>
f_t	<i>Forget gate</i> pada LSTM pada waktu ke- t
h_t	<i>Hidden state</i> pada LSTM pada waktu ke- t
i_t	<i>Input gate</i> pada LSTM pada waktu ke- t
k	<i>Key</i> pada mekanisme <i>attention</i>
L_{CE}	<i>Cross-entropy loss</i>
MAX_WORDS	Jumlah kata terbanyak yang digunakan dalam tokenizer
o_t	<i>Output gate</i> pada LSTM pada waktu ke- t
P	Presisi (<i>precision</i>)
q	<i>Query</i> pada mekanisme <i>attention</i>
R	<i>Recall</i>
t	Indeks waktu pada data sekuensial
TPR	<i>True Positive Rate</i>
v	<i>Value</i> pada mekanisme <i>attention</i>
V	Ukuran kosakata (<i>vocabulary</i>)
α	Bobot <i>attention</i>
β	Parameter pembobot pada F1-score dan regularisasi
η	<i>Learning rate</i> pada <i>gradient descent</i>
λ	Parameter regularisasi L_2
μ	Rata-rata (<i>mean</i>)
σ	Fungsi aktivasi sigmoid
AUC	Area Under the Curve
BiLSTM	Bidirectional Long Short-Term Memory
EDA	Exploratory Data Analysis

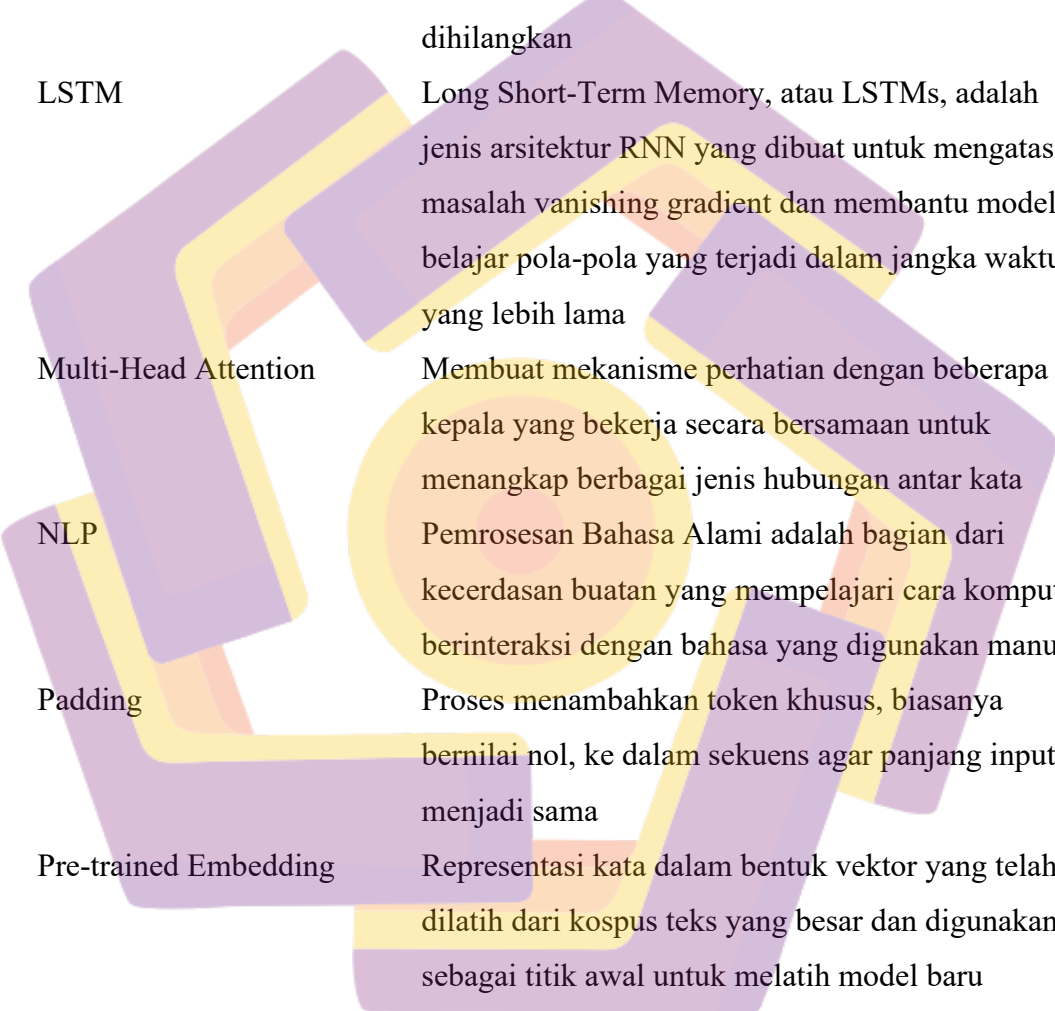


FN	False Negative
FP	False Positive
I-NAMHS	Indonesia-National Adolescent Mental Health Survey
LSTM	Long Short-Term Memory
MLP	Multi Layer Perceptron
NLP	Natural Language Processing
PTSD	Post-Traumatic Stress Disorder
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
TN	True Negative
TP	True Positive
WHO	World Health Organization

DAFTAR ISTILAH



Attention	Mekanisme ini memungkinkan model memprioritaskan input yang paling penting saat memproses data sekuensial.
AUC-ROC	Area Di Bawah Kurva—Karakteristik Operasi Penerima, metrik yang mengukur kemampuan model untuk membedakan kelas negatif dan positif
BiLSTM	LSTM yang berjalan dalam dua arah, yaitu maju dan mundur, untuk memahami konteks secara lebih lengkap
Case Folding	Proses mengubah semua huruf dalam teks menjadi huruf kecil agar representasi kata tetap konsisten
Chatbot	Program computer yang bisa berbicara dan berinteraksi dengan pengguna menggunakan teks atau suara
Confusion Matrix	Tabel yang menampilkan jumlah hasil prediksi yang benar dan salah (TP,TN,FP,FN) untuk menilai seberapa baik model klasifikasi bekerja
Deep Learning	Salah satu cabang dari machine learning yang menggunakan jaringan saraf dengan banyak lapisan untuk mempelajari representasi data secara hierarkis
Dropout	Teknik regulasi yang secara acak mematikan Sebagian neuron selama proses pelatihan agar menghindari terlalu banyaknya penyesuaian pada data pelatihan
Epoch	Satu siklus pelathan lengkap di mana model memproses seluruh data latih sekali saja
F1-Score	Rata-rata harmonic dari presisi dan recall digunakan untuk menilai keseimbangan dalam performa klasifikasi



False Negative (FN)	Kasus di mana model mengatakan negative tetapi sebenarnya hasilnya positif
False Positive (FP)	Kasus dimana model mengatakan positif, tetapi dalam kenyataannya adalah negative
Forget Gate	Gerbang pada LSTM yang menentukan informasi mana dari sel memori sebelumnya yang akan dihilangkan
LSTM	Long Short-Term Memory, atau LSTMs, adalah jenis arsitektur RNN yang dibuat untuk mengatasi masalah vanishing gradient dan membantu model belajar pola-pola yang terjadi dalam jangka waktu yang lebih lama
Multi-Head Attention	Membuat mekanisme perhatian dengan beberapa kepala yang bekerja secara bersamaan untuk menangkap berbagai jenis hubungan antar kata
NLP	Pemrosesan Bahasa Alami adalah bagian dari kecerdasan buatan yang mempelajari cara komputer berinteraksi dengan bahasa yang digunakan manusia
Padding	Proses menambahkan token khusus, biasanya bernilai nol, ke dalam sekuens agar panjang input menjadi sama
Pre-trained Embedding	Representasi kata dalam bentuk vektor yang telah dilatih dari kospus teks yang besar dan digunakan sebagai titik awal untuk melatih model baru
Presisi	Proporsi dari prediksi positif yang benar dibandingkan dengan total jumlah prediksi positif
Recall	Proporsi data yang benar-benar positif dan berhasil diprediksi dengan akurat
RNN	Jaringan saraf yang dirancang untuk memproses data berurutan dengan cara berulang

Tokenisasi	Proses memecah teks menjadi bagian-bagian kecil seperti kata atau bagian dari kata
True Positive (TP)	True Positive adalah kasus di mana model menghasilkan prediksi positif dan hasilnya benar-benar positif
True Negative (TN)	True Negative, yaitu kasus dimana model menghasilkan prediksi negative dan sebenarnya kondisinya juga negatif
Word Embedding	Representasi kata dalam bentuk angka yang menangkap makna dan struktur kalimat



INTISARI

Kesehatan mental adalah bagian penting dalam hidup seseorang yang sering tidak diperhatikan, dan kasus gangguan mental seperti depresi serta kecemasan semakin banyak terjadi. Deteksi dini sulit dilakukan karena adanya stigma sosial dan kurangnya akses ke layanan profesional. Media sosial seperti Twitter/X menjadi tempat bagi orang-orang untuk menyampaikan perasaan dan pengalaman psikologis mereka, sehingga teks dari platform ini bisa digunakan untuk memprediksi kondisi mental seseorang dengan menggunakan metode Natural Language Processing (NLP) dan deep learning. Penelitian ini ingin membandingkan bagaimana sempat jenis arsitektur LSTM bekerja, yaitu LSTM biasa, LSTM berarah dua arah (BiLSTM), LSTM yang menggunakan mekanisme perhatian (LSTM+Attention), dan LSTM yang menggunakan multi-head attention (LSTM+MultiHead), dalam memprediksi kondisi mental pengguna dari data teks. Dataset yang digunakan adalah Mental Health Corpus, yang terdiri dari 27.977 sampel teks dan memiliki dua kategori label, yaitu normal dan indikasi gangguan mental. Model dilatih menggunakan embedding yang sudah dilatih sebelumnya yaitu Dolma 300D, dan penilaian dilakukan dengan mengukur beberapa metrik seperti akurasi, presisi, recall, skor F1, serta AUC-ROC. Hasil penelitian menunjukkan bahwa model BiLSTM memiliki kinerja terbaik dengan akurasi mencapai 92,98%, presisi 91,30%, recall 94,83%, F1-score 93,04%, serta AUC-ROC 98,15%. Model berikutnya adalah LSTM+Attention dengan akurasi dengan akurasi 91,83% dan F1-score 91,99%, kemudian LSTM dengan akurasi 92,07% dan F1-score 92,24%, serta LSTM+MultiHead dengan akurasi 91,10% dan F1-score 91,27%. Penelitian ini membantu memilih arsitektur yang sesuai untuk system deteksi dini Kesehatan mental yang menggunakan teks, serta memberikan pemahaman tentang bagaimana efektivitas dari mekanisme bidireksional dan perhatian.

Kata Kunci: Prediksi Kesehatan Mental, LSTM, BiLSTM, Attention, Multi-Head Attention, Klasifikasi Teks, Media Sosial.

ABSTRACT

Mental health is an important part of a person's life that is often overlooked, and cases of mental disorders such as depression and anxiety are increasingly common. Early detection is difficult due to social stigma and lack of access to professional services. Social media such as Twitter/X has become a place for people to share their feelings and psychological experiences, so text from these platforms can be used to predict a person's mental state using Natural Language Processing (NLP) and deep learning methods. This study aims to compare how several types of LSTM architectures work, namely regular LSTM, bidirectional LSTM (BiLSTM), LSTM using attention mechanisms (LSTM+Attention), and LSTM using multi-head attention (LSTM+MultiHead), in predicting a user's mental state from text data. The dataset used is the Mental Health Corpus, which consists of 27,977 text samples and has two label categories: normal and indicative of mental disorders. The model was trained using a pre-trained embedding, namely Dolma 300D, and assessment was carried out by measuring several metrics such as accuracy, precision, recall, F1 score, and AUC-ROC. The results showed that the BiLSTM model performed best, achieving 92.98% accuracy, 91.30% precision, 94.83% recall, 93.04% F1-score, and 98.15% AUC-ROC. The next best model was LSTM+Attention, with 91.83% accuracy and 91.99% F1-score, followed by LSTM with 92.07% accuracy and 92.24% F1-score, and LSTM+MultiHead with 91.10% accuracy and 91.27% F1-score. This research helps select an appropriate architecture for a text-based mental health early detection system and provides insight into the effectiveness of bidirectional and attention mechanisms.

Keywords: Mental Health Prediction, LSTM, BiLSTM, Attention, Multi-Head Attention, Text Classification, Social Media