

Optimasi Analisis Sentimen Twitter Tentang Isu Kesehatan Mental dengan Bi-LSTM pada Dataset Tidak Berimbang

Indah Rani Fatmawati*, Muhammad Pajar Kharisma Putra

Fakultas Teknik dan Ilmu Komputer, Program Studi Informatika, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹ indah_rani_fatmawati@teknokrat.ac.id, ² pajarkharisma@teknokrat.ac.id

Email Penulis Korespondensi: indah_rani_fatmawati@teknokrat.ac.id

Submitted: 31/01/2025; Accepted: 04/03/2025; Published: 07/03/2025

Abstrak—Penelitian ini bertujuan untuk menganalisis sentimen pengguna Twitter terkait isu kesehatan mental menggunakan model Bidirectional Long Short-Term Memory (BiLSTM). Dataset terdiri dari 52,681 entri yang mencakup tujuh kategori kesehatan mental: Anxiety, Bipolar, Depression, Normal, Personality Disorder, Stress, dan Suicidal. Metode yang digunakan meliputi pra-pemrosesan data, pembagian data, serta pelatihan model dengan teknik penyesuaian bobot kelas untuk menangani ketidakseimbangan data. Hasil pelatihan menunjukkan peningkatan akurasi dari 16,02% pada epoch pertama menjadi 88,48% pada epoch ke-10, dengan akurasi evaluasi sebesar 74,21%. Model menunjukkan kinerja terbaik pada kelas Anxiety dengan nilai F1-score sebesar 0,90. Namun, model masih mengalami keterbatasan dalam mengklasifikasikan kelas minoritas seperti Bipolar dan Personality Disorder akibat jumlah data yang sedikit serta kompleksitas ekspresi bahasa dalam kategori tersebut. Oleh karena itu, diperlukan peningkatan jumlah data dan teknik pemrosesan bahasa yang lebih adaptif untuk meningkatkan kinerja model pada kategori dengan jumlah data terbatas.

Kata Kunci: Sentimen; Kesehatan Mental; Twitter; BiLSTM; Klasifikasi

Abstract—This study aims to analyze Twitter user sentiments related to mental health issues using the Bidirectional Long Short-Term Memory (BiLSTM) model. The dataset consists of 52,681 entries covering seven mental health categories: Anxiety, Bipolar, Depression, Normal, Personality Disorder, Stress, and Suicidal. The methods used include data pre-processing, data splitting, and model training with class weight adjustment techniques to handle data imbalance. The training results show an increase in accuracy from 16.02% in the first epoch to 88.48% in the 10th epoch, with an evaluation accuracy of 74.21%. The model shows the best performance in the Anxiety class with an F1-score of 0.90. However, the model still experiences limitations in classifying minority classes such as Bipolar and Personality Disorder due to the small amount of data and the complexity of language expressions in these categories. Therefore, an increase in the amount of data and more adaptive language processing techniques are needed to improve model performance in categories with limited data.

Keywords: Sentiment; Mental Health; Twitter; BiLSTM; Classification

1. PENDAHULUAN

Kesehatan mental merupakan salah satu aspek penting dalam kehidupan manusia yang sering kali terabaikan dalam diskusi publik. Menurut Organisasi Kesehatan Dunia (WHO), kesehatan mental mencakup kemampuan individu untuk mengenali potensi dirinya, mengatasi tekanan hidup sehari-hari, bekerja secara produktif, serta memberikan kontribusi kepada komunitasnya [1]. Namun, stigma dan kurangnya pemahaman mengenai kesehatan mental masih menjadi hambatan signifikan bagi individu untuk mendapatkan bantuan yang mereka butuhkan [2]. Di berbagai budaya, isu kesehatan mental masih dianggap tabu untuk dibahas secara terbuka, sehingga menciptakan rasa isolasi bagi banyak individu [3]. Dalam konteks ini, penting untuk memahami persepsi masyarakat terhadap kesehatan mental agar dapat mengatasi stigma dan meningkatkan kesadaran publik.

Media sosial, khususnya *Twitter*, telah menjadi platform utama untuk diskusi publik mengenai berbagai isu, termasuk kesehatan mental [4]. Platform ini memungkinkan pengguna untuk berbagi pengalaman pribadi, memberikan dukungan, serta menyuarakan pendapat mereka tentang topik yang relevan. Melalui analisis sentimen pada *tweet*, kita dapat memahami bagaimana masyarakat memandang kesehatan mental. Informasi ini dapat memberikan wawasan penting untuk memahami pola pikir masyarakat terhadap isu-isu kesehatan mental dan mengidentifikasi stigma yang masih ada.

Penelitian terkait analisis sentimen di *Twitter* telah berkembang pesat dalam beberapa tahun terakhir. Alisyafira et al. (2024) menggunakan metode *Naive Bayes* dan *Support Vector Machine (SVM)* untuk menganalisis data *Twitter* terkait kesehatan mental pascapandemi COVID-19. Hasil penelitian menunjukkan bahwa 38,34% dari *tweet* bersifat netral, 32,56% positif, dan 29,09% negatif. Penelitian ini menyoroti pentingnya media sosial dalam berbagi informasi dan dukungan terkait kesehatan mental [5]. Penelitian ini menjadi bukti bahwa berbagai metode tradisional telah memberikan dasar yang kuat untuk memahami persepsi masyarakat, meskipun perkembangan metode yang lebih mutakhir tetap diperlukan.

Selanjutnya, Astari (2021) dalam penelitiannya, "Analisis Sentimen Multi-Kelas pada Media Sosial Menggunakan Metode *Long Short-Term Memory (LSTM)*," menunjukkan bahwa model *LSTM* yang dikembangkan berhasil mencapai akurasi 91,9% dalam mengklasifikasikan emosi dalam teks berbahasa Indonesia. Penelitian ini menekankan pentingnya kemajuan teknologi dalam analisis sentimen untuk memahami isu kesehatan mental di media sosial. Dengan demikian, *LSTM* telah menjadi salah satu metode yang sangat andal dalam menangkap pola sentimen yang kompleks [6].

Karina et al. (2023) juga mengaplikasikan algoritma *Naive Bayes* untuk menganalisis *tweet* terkait isu kesehatan mental. Hasil penelitian menunjukkan bahwa sentimen positif mendominasi diskusi mengenai kesehatan mental di *Twitter*. Meskipun metode yang digunakan terbatas pada klasifikasi dasar, penelitian ini memberikan kontribusi signifikan dalam mengidentifikasi pola umum dalam diskusi kesehatan mental [7]. Metode ini menjadi dasar untuk eksplorasi lebih lanjut menggunakan pendekatan yang lebih canggih.

Windjatika dan Maharani (2022) dalam penelitian "*Depression Detection on Social Media Twitter Using Long Short-Term Memory*" menyoroti pentingnya mendeteksi gangguan depresi melalui analisis data dari 159 pengguna *Twitter*. Penelitian ini menggunakan metode *Word2Vec* dan *LSTM*, dengan hasil akurasi sebesar 77,95% dan *F1-score* sebesar 57,14%. Hasil ini menunjukkan bahwa model deep learning memiliki potensi besar untuk meningkatkan analisis sentimen, meskipun masih terdapat tantangan yang perlu diatasi, terutama dalam hal peningkatan presisi dan pemahaman konteks [8].

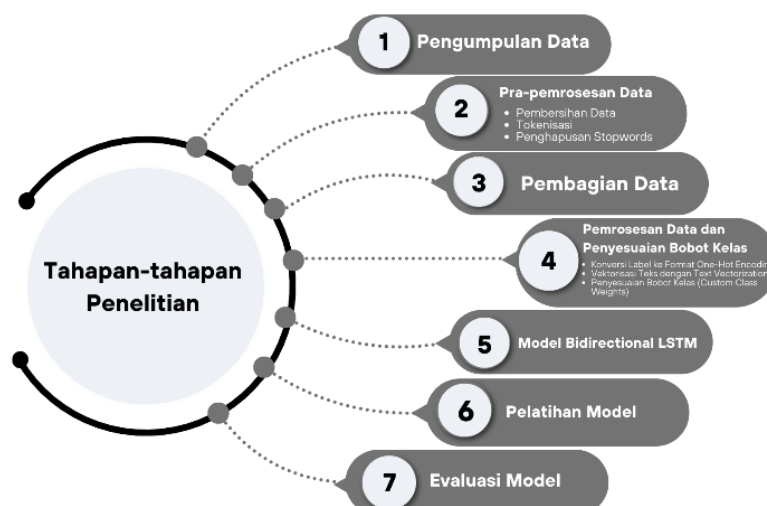
Penelitian terbaru oleh Wibowo et al. (2024) dalam penelitian "Implementasi *Long Short-Term Memory* Dalam Analisis Sentimen Pengguna Aplikasi *Twitter* Yang Mengandung Ujaran Kebencian" menunjukkan efektivitas algoritma *LSTM* dengan akurasi mencapai 83% dalam analisis sentimen ujaran kebencian. Meskipun penelitian ini tidak secara langsung berfokus pada isu kesehatan mental, pendekatannya relevan untuk memahami potensi *LSTM* dalam berbagai konteks analisis sentimen, termasuk topik yang sensitif [9].

Dalam penelitian ini, metode *Bidirectional Long Short-Term Memory (BiLSTM)* diusulkan untuk menganalisis sentimen pengguna *Twitter* terhadap isu kesehatan mental. Model *BiLSTM* dipilih karena kemampuannya menangkap konteks dari dua arah, baik dari kata-kata sebelumnya maupun setelahnya, sehingga memungkinkan pemahaman yang lebih baik terhadap makna kalimat secara keseluruhan [10]. Dibandingkan dengan metode sebelumnya seperti *Naive Bayes* dan *Support Vector Machine (SVM)* yang mengandalkan fitur statis, atau *Recurrent Neural Network (RNN)* yang memiliki keterbatasan dalam menangani *long-term dependencies*, *BiLSTM* mampu mempertahankan informasi lebih lama dan memahami hubungan antar kata secara lebih akurat. Penggunaan *BiLSTM* memberikan peluang baru untuk menangkap hubungan konteks linguistik secara lebih mendalam dibandingkan beberapa pendekatan sebelumnya. Selain itu, kelebihan model ini terletak pada kemampuannya untuk mengolah data yang bersifat sekuensial secara efisien, sehingga menghasilkan analisis yang lebih akurat [11].

Penelitian ini bertujuan untuk memberikan kontribusi pada upaya meningkatkan kesadaran dan pemahaman masyarakat terhadap kesehatan mental. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi potensi stigma yang masih ada di masyarakat, sehingga dapat menjadi dasar untuk strategi edukasi dan intervensi yang lebih efektif. Dengan hasil yang dicapai, penelitian ini tidak hanya memberikan kontribusi dalam ranah akademik melalui penerapan model *BiLSTM*, tetapi juga memiliki dampak praktis yang signifikan dalam upaya meningkatkan kesadaran kesehatan mental di masyarakat. Hasil penelitian ini diharapkan dapat mendorong diskusi yang lebih terbuka dan inklusif mengenai kesehatan mental, sekaligus menjadi rujukan bagi penelitian lebih lanjut di bidang analisis sentimen dan kesehatan mental.

2. METODOLOGI PENELITIAN

Dalam penelitian ini, fokus utama ditujukan pada metode klasifikasi menggunakan *Bidirectional LSTM (BiLSTM)*, meskipun terdapat berbagai metode klasifikasi lain yang tersedia. Proses untuk mendapatkan hasil yang optimal melibatkan beberapa tahapan, mulai dari pengumpulan data, hingga pra-pemrosesan data. Setelah data diproses, langkah selanjutnya adalah menerapkan algoritma *BiLSTM* untuk melakukan klasifikasi sentimen berdasarkan data yang tersedia. Gambar 1 menunjukkan tahapan-tahapan penelitian secara keseluruhan.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini dikumpulkan dari berbagai sumber terpercaya, termasuk platform media sosial seperti *Twitter* dan *Reddit*. Data yang digunakan berjumlah 52,681 entri yang diberi label sesuai kategori kesehatan mental, yaitu *Normal*, *Depression*, *Suicidal*, *Anxiety*, *Bipolar*, *Stress*, dan *Personality Disorder*. Data ini telah melalui proses validasi dan pembersihan untuk memastikan kualitasnya.

2.2 Pra-pemrosesan Data

Pra-pemrosesan adalah tahap persiapan data yang bertujuan untuk memudahkan pengolahan dan analisis data [12]. Langkah-langkah yang dilakukan antara lain:

- a. Pembersihan Data
Menghapus *URL*, karakter tidak relevan, dan tanda baca, serta menurunkan teks menjadi huruf kecil.
- b. Tokenisasi dan Penghapusan *Stopwords*
Mengubah setiap kalimat menjadi daftar kata menggunakan pustaka seperti `nlTK.word_tokenize`. Tokenisasi ini dilakukan untuk memudahkan dalam tahapan prapemrosesan selanjutnya [13]. Serta proses penghapusan *stopword*, yaitu kata-kata yang sering muncul namun tidak penting dan relevan, seperti kata hubung (konjungsi), kata kepemilikan, dan kata ganti orang dilakukan penghapusan dan penghilangan [14].

2.3 Pembagian Data

Setelah pra-pemrosesan data, pembagian dataset dilakukan menggunakan fungsi `train_test_split` dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian. Teknik *random sampling* diterapkan dalam proses ini, dengan parameter `random_state=42` untuk memastikan hasil yang konsisten dalam setiap eksperimen. Pendekatan ini memungkinkan model untuk belajar dari berbagai sampel data secara acak, sehingga dapat menangkap pola yang beragam dalam analisis sentimen kesehatan mental. Jika diperlukan, teknik *stratified sampling* dapat diterapkan dengan menambahkan parameter `stratify=y` pada fungsi `train_test_split`, sehingga proporsi setiap kategori tetap terjaga dalam kedua subset. Hal ini dapat membantu model dalam menangani dataset yang memiliki distribusi kelas tidak seimbang.

2.4 Pemrosesan Data dan Penyesuaian Bobot Kelas

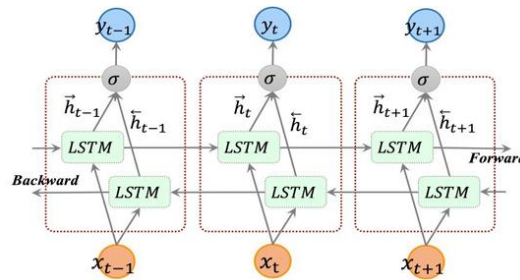
Pada penelitian ini, dilakukan langkah-langkah pemrosesan data untuk memastikan data siap digunakan oleh model. Langkah-langkah tersebut meliputi:

- a. Konversi Label ke Format *One-Hot Encoding*
Dalam metode ini setiap kategori atau label diubah menjadi vektor biner dengan panjang yang setara dengan jumlah total kategori yang berbeda dalam variabel tersebut [15]. label pada data latih dan data uji dikonversi ke format *one-hot encoding* menggunakan fungsi `pd.get_dummies`. Langkah ini bertujuan untuk mengubah label kategori menjadi representasi numerik yang mudah diproses oleh model pembelajaran mesin, sekaligus memastikan bahwa setiap kelas memiliki representasi unik.
- b. Vektorisasi Teks dengan *Text Vectorization*
Selanjutnya, data teks diubah menjadi bentuk numerik menggunakan lapisan *TextVectorization* dari *TensorFlow*. Lapisan ini berfungsi untuk membuat representasi numerik dari data teks dengan membatasi jumlah kata unik yang digunakan (hingga 20.000 kata) dan memastikan setiap teks memiliki panjang tetap (200 kata). Proses ini mempermudah model dalam mempelajari pola dari data teks sekaligus menjaga konsistensi panjang data, sehingga pelatihan model dapat berlangsung lebih efisien.
- c. Penyesuaian Bobot Kelas (*Custom Class Weights*)
Selain itu, penyesuaian bobot kelas (*custom class weights*) dilakukan untuk mengatasi ketidakseimbangan jumlah data antar kategori dalam data latih. Proses ini memberikan bobot yang lebih besar pada kelas dengan jumlah data lebih sedikit dan bobot yang lebih kecil pada kelas dengan jumlah data lebih banyak. Penyesuaian bobot kelas bertujuan untuk meningkatkan performa model pada kelas minoritas, sehingga model dapat memberikan prediksi yang lebih adil dan akurat untuk semua kategori, terutama pada data yang tidak seimbang. Bobot kelas yang telah dihitung kemudian digunakan selama pelatihan untuk mengarahkan model agar mempertimbangkan semua kelas secara proporsional.

2.5 Bidirectional LSTM

Model klasifikasi dibangun menggunakan arsitektur *Bidirectional LSTM* dengan *framework TensorFlow* dengan *layer embedding* untuk menangkap representasi semantik kata-kata, lapisan *LSTM* untuk menangkap konteks temporal data teks, dan *dropout* untuk mencegah *overfitting*. *Bidirectional Long Short Term Memory* atau *Bi-LSTM* adalah model pemrosesan berurutan yang terdiri dari dua *LSTM* : satu mengambil *input* dalam arah maju (*forward direction*), dan yang lainnya di arah belakang (*backward direction*) [16]. *LSTM* adalah salah satu tipe dari *RNN* yang dapat membuat jaringan mampu mempertahankan ketergantungannya dalam jangka panjang antara data pada waktu tertentu dengan beberapa data pada waktu sebelum-sebelumnya [17]. Dengan adanya dua *layer* yang berlawanan arah tersebut, model

dapat mempelajari informasi masa lalu dan informasi masa mendatang untuk setiap *sequence input* [18]. Desain model juga mencakup penggunaan dropout untuk mengurangi risiko *overfitting* selama pelatihan. Arsitektur model *Bi-LSTM* yang digunakan dapat dilihat pada Gambar 2.



Gambar 2. Arsitektur Model *Bi-LSTM*

Gambar 2 menunjukkan bahwa dalam arsitektur BiLSTM, proses pada *forward LSTM* dapat ditulis sebagai $\vec{h}_t = LSTM(X_t, h_{t-1})$ sedangkan untuk proses *backward LSTM* ditulis sebagai $\overleftarrow{h}_t = LSTM(X_t, h_{t+1})$ Output akhir dari *BiLSTM* adalah penggabungan dari *forward* dan *backward* yang ditulis sebagai $h_t = (\vec{h}_t, \overleftarrow{h}_t)$ [19].

2.6 Pelatihan Model

Pelatihan dilakukan menggunakan fungsi loss *categorical_crossentropy* dan optimizer Adam. Untuk mengatasi ketidakseimbangan kelas, penyesuaian bobot kelas diterapkan. Pelatihan berlangsung selama 20 epoch dengan *batch size* 64, serta menggunakan teknik *early stopping* untuk menghentikan pelatihan jika tidak ada peningkatan signifikan pada data validasi.

2.7 Evaluasi Model

Evaluasi model dilakukan menggunakan berbagai metrik performa, seperti akurasi, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Proses evaluasi juga meliputi pengujian dengan berbagai parameter, seperti ukuran *batch*, jumlah *epoch*, dan jumlah *neuron* pada setiap *layer*. Evaluasi ini bertujuan untuk menilai kualitas model dalam mengklasifikasikan data, serta untuk memastikan model tidak *overfitting* atau *underfitting*.

3. HASIL DAN PEMBAHASAN

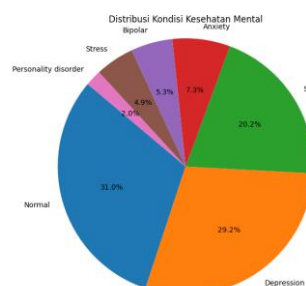
3.1 Distribusi Kelas

Dataset yang digunakan dalam penelitian ini terdiri dari 52,681 entri yang terbagi ke dalam beberapa kategori kondisi kesehatan mental. Distribusi jumlah data pada masing-masing kategori dapat dilihat pada Tabel 1 berikut:

Tabel 1. Distribusi Kelas

Status	Jumlah data
Anxiety	3,841
Bipolar	2,777
Depression	15,404
Normal	16,343
Personality Disorder	1,077
Stress	2,585
Suicidal	10,652

Distribusi ini divisualisasikan menggunakan diagram *pie* untuk memberikan gambaran proporsi setiap kategori secara keseluruhan. Diagram *pie* dapat dilihat pada Gambar 3.



Gambar 3. Distribusi Kondisi Kesehatan Mental

Berdasarkan Tabel 1 dan Gambar 3, terlihat bahwa dataset memiliki distribusi kelas yang tidak seimbang, di mana kategori Normal, *Depression*, dan *Suicidal* memiliki jumlah data yang jauh lebih banyak dibandingkan kelas lainnya, seperti *Bipolar* dan *Personality Disorder*. Ketidakseimbangan ini dapat berdampak pada performa model, karena algoritma cenderung lebih akurat dalam memprediksi kelas dengan jumlah data yang dominan, sementara kelas minoritas mungkin mengalami tingkat kesalahan prediksi yang lebih tinggi. Untuk mengatasi tantangan ini, beberapa strategi dapat diterapkan, seperti *oversampling* pada kelas minoritas untuk meningkatkan jumlah sampel, *undersampling* pada kelas mayoritas untuk menyeimbangkan distribusi, atau augmentasi data guna menciptakan variasi data yang lebih representatif. Selain itu, teknik penyesuaian bobot kelas (*class weighting*) juga dapat digunakan untuk memberikan perhatian lebih pada kelas yang lebih sedikit, sehingga model dapat belajar secara lebih optimal dan menghasilkan prediksi yang lebih akurat.

3.2 Pra Pemrosesan Data

Tahap pra-pemrosesan data sangat penting untuk memastikan bahwa data yang digunakan siap dan berkualitas baik untuk analisis selanjutnya. Pra-pemrosesan data teks juga mendukung persiapan data untuk digunakan dalam model-model *NLP* yang kompleks [20]. Pada tahap ini, beberapa langkah dilakukan, mulai dari membersihkan data, memecahnya menjadi token-token, hingga menghapus kata-kata yang tidak terlalu berarti bagi analisis. Berikut penjelasan lebih lanjut mengenai proses yang dilakukan:

a. Pembersihan Data

Langkah pertama yang dilakukan adalah membersihkan data dari elemen-elemen yang tidak relevan, seperti *URL*, karakter khusus, dan tanda baca. Hal ini dilakukan untuk memastikan bahwa teks yang digunakan tidak mengandung informasi yang tidak diperlukan. Selain itu, semua teks diubah menjadi huruf kecil untuk menghindari perbedaan penulisan yang dapat mempengaruhi analisis. Hasil dari pembersihan data ini dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pembersihan Data

Status	Statement	Cleaned_statement
Anxiety	I've been so restless since this morning	ive been so restless since this morning
Bipolar	How do you cope with mixed hypomania? Help please	how do you cope with mixed hypomania help please
Depression	someone please message me oh my god help	someone please message me oh god help
Normal	last night wearing patch even more tired	last night wearing patch even more tired
Personality Disorder	I slut shame. Edit: You've made your point.	i slut shame edit youve made your point
Stress	Insomnia How do you guys fight insomnia?	insomnia how do you guys fight insomnia
Suicidal	Count down to the end, No longer coping.	count down to the end no longer coping

b. Tokenisasi

Setelah data dibersihkan, tahap berikutnya adalah tokenisasi. Tokenisasi bertujuan untuk memecah teks panjang menjadi unit-unit lebih kecil, yang disebut token [21]. Setiap kata dalam teks dipisahkan untuk memudahkan model dalam memproses informasi. Hasil dari tokenisasi dapat dilihat pada Tabel 3.

Tabel 3. Hasil Tokenisasi

Status	Cleaned_statement	tokens
Anxiety	ive been so restless since this morning	['ive', 'been', 'so', 'restless', 'since', 'this', 'morning']
Bipolar	how do you cope with mixed hypomania help please	['how', 'do', 'you', 'cope', 'with', 'mixed', 'hypomania', 'help', 'please']
Depression	someone please message me oh god help	['someone', 'please', 'message', 'me', 'oh', 'my', 'god', 'help']
Normal	last night wearing patch even more tired	['last', 'night', 'wearing', 'patch', 'even', 'more', 'tired']
Personality Disorder	i slut shame edit youve made your point	['i', 'slut', 'shame', 'edit', 'youve', 'made', 'your', 'point']
Stress	insomnia how do you guys fight insomnia	['insomnia', 'how', 'do', 'you', 'guys', 'fight', 'insomnia']
Suicidal	count down to the end no longer coping	['count', 'down', 'to', 'the', 'end', 'no', 'longer', 'coping']

c. Penghapusan *Stopwords*

Langkah selanjutnya adalah menghapus *stopwords*. *Stopwords* adalah kata-kata umum yang sering muncul dalam kalimat namun tidak memberikan informasi yang signifikan untuk analisis, seperti kata "*the*", "*is*", dan "*and*" [22]. Penghapusan *Stopwords* diharapkan dapat menghasilkan dataset dengan kualitas yang lebih baik sehingga

performa nilai *f1-score* model yang dibangun menjadi lebih optimal. Hasil dari penghapusan *stopwords* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Penghapusan *Stopwords*

status	Cleaned_statement	tokens
Anxiety	ive been so restless since this morning	ive restless since morning
Bipolar	how do you cope with mixed hypomania help please	cope mixed hypomania help please
Depression	someone please message me oh god help	someone please message oh god help
Normal	last night wearing patch even more tired	last night wearing patch even tired
Personality	i slut shame edit youve made your point	slut shame edit youve made point
Disorder		
Stress	insomnia how do you guys fight insomnia	insomnia guys fight insomnia
Suicidal	count down to the end no longer coping	count end longer coping

3.3 Pembagian Data

Setelah preprocessing, dataset dibagi menjadi 80% data pelatihan dan 20% data pengujian menggunakan fungsi `train_test_split` dari *Scikit-learn*. Pembagian ini memastikan model belajar dari data pelatihan (X_{train} , y_{train}) dan diuji pada data yang belum terlihat (X_{test} , y_{test}). Total data dalam penelitian ini adalah 52.681 entri. sehingga pembagian data dapat dihitung sebagai berikut:

$$\text{Jumlah data latih} = 80\% \times \text{total data} = 0.8 \times 52,681 = 42,144$$

$$\text{Jumlah data uji} = 20\% \times \text{total data} = 0.2 \times 52,681 = 10,537$$

Hasil perhitungan ini konsisten dengan keluaran dari implementasi kode, yaitu 42,144 data latih dan 10,537 data uji. Untuk memastikan pembagian yang konsisten setiap kali kode dijalankan, digunakan parameter `random_state=42`.

3.4 Pemrosesan Data dan Penyesuaian Bobot Kelas

a. Konversi Label ke Format *One-Hot Encoding*

Label pada data latih (y_{train}) dan data uji (y_{test}) dikonversi ke format *one-hot encoding* menggunakan fungsi `pd.get_dummies`. Dengan 7 kelas, yaitu *Normal*, *Depression*, *Suicidal*, *Anxiety*, *Bipolar*, *Stress*, dan *Personality Disorder*, setiap label diubah menjadi vektor biner, misalnya:

1. "Normal" $\rightarrow [1, 0, 0, 0, 0, 0, 0]$

2. "Depression" $\rightarrow [0, 1, 0, 0, 0, 0, 0]$, dan seterusnya.

Hasilnya disimpan dalam variabel $y_{train_encoded}$ untuk data latih dan $y_{test_encoded}$ untuk data uji, yang akan digunakan sebagai input untuk pelatihan model.

b. Vektorisasi Teks dengan *Text Vectorization*

Data teks pada X_{train} dan X_{test} diubah menjadi representasi numerik menggunakan lapisan *TextVectorization* dari TensorFlow. Proses ini mencakup:

1. Membatasi jumlah token unik hingga 20.000 kata (`max_tokens=20000`), yang disesuaikan berdasarkan data latih.

2. Mengonversi setiap kalimat menjadi urutan angka dengan panjang tetap 200 kata (`output_sequence_length=200`). Kata-kata yang melebihi batas akan dipotong, dan yang lebih pendek akan diisi dengan padding.

Hasil vektorisasi disimpan dalam $X_{train_vectorized}$ dan $X_{test_vectorized}$, yang digunakan sebagai input numerik untuk model.

c. Penyesuaian Bobot Kelas (*Custom Class Weights*)

Penyesuaian bobot kelas dilakukan untuk mengatasi ketidakseimbangan kelas pada dataset menggunakan pendekatan *custom class weights*. Dengan metode ini, bobot dihitung secara otomatis berdasarkan distribusi kelas menggunakan fungsi `compute_class_weight` dari *Scikit-learn*. Pendekatan ini memastikan bahwa kelas dengan jumlah data kecil (*minority classes*) mendapatkan bobot lebih tinggi, sehingga model lebih memperhatikan kelas-kelas tersebut selama pelatihan. Metode ini menghitung bobot berdasarkan rumus berikut:

$$W = \frac{N}{k \cdot n_i} \quad (1)$$

Bobot kelas ke- i (W) dihitung berdasarkan jumlah total data (N) yang terdiri dari 52.681 sampel dan jumlah total kelas (k) yang mencakup tujuh kategori. Setiap kelas memiliki jumlah data masing-masing yang dinyatakan sebagai n_i , yaitu jumlah data dalam kelas- i . Bobot ini digunakan untuk menyeimbangkan distribusi data antar kelas dalam analisis yang dilakukan.

Contoh Perhitungan untuk kelas *Personality disorder*:

Total data (N) = 52.681.

Total kelas (k) = 7.

Jumlah data pada kelas *Personality disorder* (n_i) = 1.077.

Bobot dihitung sebagai berikut: Bobot dihitung sebagai berikut:

$$W = \frac{52.681}{7 \cdot 1.077} = 7.167$$

Hasil perhitungan menghasilkan Tabel 5 bobot kelas berikut:

Tabel 5. Hasil Perhitungan Bobot Kelas

Class	Weight
Anxiety	1.950903
Bipolar	2.675810
Depression	0.486000
Normal	0.461877
Personality Disorder	7.167347
Stress	2.935432
Suicidal	0.708803

Tabel 5 menunjukkan bahwa kelas *Personality Disorder* memiliki bobot tertinggi karena jumlah datanya paling sedikit, sedangkan Normal dan *Depression* memiliki bobot lebih rendah. Bobot ini digunakan dalam parameter `class_weight` untuk memastikan model lebih fokus pada kelas minoritas, sehingga meningkatkan keseimbangan performa pada semua kelas.

3.5 Hasil Pelatihan Model

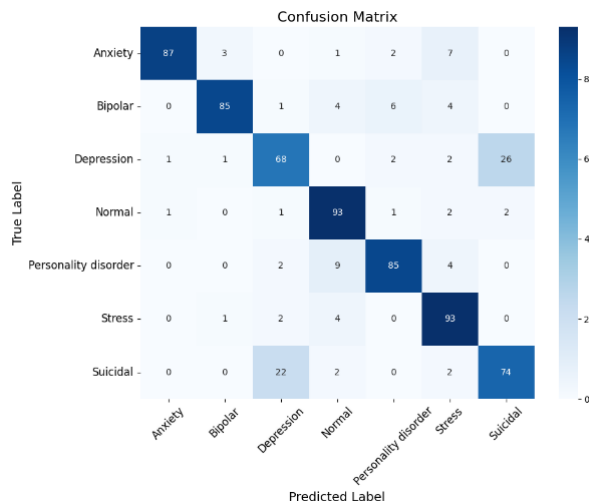
Selama proses pelatihan, performa model menunjukkan peningkatan signifikan pada data latih, dengan akurasi meningkat dari 16,02% di *epoch* pertama menjadi 88,48% di *epoch* ke-10, sementara nilai loss menurun secara konsisten. Namun, akurasi validasi mencapai batas maksimal 73,83% pada *epoch* ke-10 dan mengalami stagnasi sejak *epoch* ke-7. Fenomena ini menunjukkan adanya *overfitting* ringan, meskipun telah diterapkan teknik seperti *dropout* dan *early stopping* untuk membatasi hal tersebut. Tabel 6 menunjukkan hasil pelatihan model pada setiap *epoch*, termasuk akurasi dan loss untuk data latih, serta akurasi dan *loss* untuk data validasi.

Tabel 6. Hasil Pelatihan Model Setiap Epoch

Epoch	Train Accuracy	Train Loss	Validation Accuracy	Validation Loss
1	0.1602	1.9139	0.2814	1.6106
2	0.2322	1.8328	0.3204	1.5453
3	0.3739	1.6563	0.5470	1.1394
4	0.5627	1.2842	0.6355	0.9555
5	0.6750	0.9850	0.6925	0.8564
6	0.7542	0.7173	0.7257	0.7513
7	0.8051	0.4816	0.7421	0.7205
8	0.8413	0.3672	0.7269	0.7712
9	0.8681	0.2934	0.7343	0.8625
10	0.8848	0.2508	0.7383	0.8462

Hasil proses pelatihan model pada Tabel 6 menunjukkan model mampu mempelajari pola pada data latih dan menggeneralisasikan pada data uji meski ada sedikit fluktuasi akurasi dan loss. *Early stopping* dengan *patience* 3 berhasil mencegah *overfitting*, menghentikan pelatihan di *epoch* ke-10 karena *validation loss* sudah stabil, meski maksimum *epoch* ditetapkan 20.

Model diuji menggunakan 100 sampel data baru yang tidak terlibat dalam proses pelatihan, dengan tujuan untuk mengevaluasi kemampuan model dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya. Data ini diambil secara merata dari setiap kelas dan hasilnya diintegrasikan ke dalam *confusion matrix*. Berikut adalah *confusion matrix* yang dihasilkan dapat dilihat pada Gambar 4.



Gambar 4. Confusion Matrix model uji

Gambar 4 menunjukkan bahwa model cukup akurat dalam mengklasifikasikan kelas *Anxiety*, *Bipolar*, *Normal*, dan *Stress*, dengan prediksi benar di atas 85%. Namun, terdapat kesalahan signifikan pada kelas *Depression* dan *Suicidal*, di mana banyak data *Depression* salah diklasifikasikan sebagai *Suicidal* dan sebaliknya. Kelas *Personality Disorder* juga mengalami kesalahan dengan beberapa data dikategorikan sebagai *Normal* atau *Depression*. Kesalahan ini menunjukkan bahwa model masih kesulitan membedakan kelas dengan kemiripan semantik tinggi, sehingga diperlukan perbaikan seperti augmentasi data, pembobotan kelas, atau penggunaan model berbasis transformer. Analisis kesalahan klasifikasi dilakukan untuk memahami kekeliruan model berdasarkan data dengan contoh kasus *false positive* dan *false negative*.

- Pada kasus *false negative* di kelas *Suicidal*, model salah memprediksi sebagai *Depression* meskipun data asli mengindikasikan perilaku membahayakan diri sendiri yang menjadi ciri khas kelas *Suicidal*. Sebagai contoh, pada pernyataan "*Only one person is keeping me alive. Hit my head off my desk till it bled,*" frasa ini mencerminkan tingkat keputusasaan dan tindakan ekstrem, tetapi model kesulitan membedakan intensitas emosional tersebut dari kelas *Depression*. Hal ini menunjukkan bahwa model belum sepenuhnya mampu mengenali tingkat keparahan emosi yang terungkap.
- Sementara itu, pada kasus *false positive* di kelas *Normal*, seperti pada pernyataan "*Laying here with my wrists cut open. Struggling to feel anything at all. I do this too many nights...kill me I am really not okay. I am dying,*" model memprediksi sebagai *Normal*, padahal pernyataan ini jelas menunjukkan karakteristik dari kelas *Depression*. Penggunaan kata-kata seperti "*kill me*" dan "*not okay*" seharusnya menjadi indikator eksplisit untuk kelas tersebut, namun model tampaknya tidak mampu menangkap intensitas emosional secara akurat. Kekeliruan ini mengindikasikan bahwa model memiliki keterbatasan dalam memahami konteks emosional yang mendalam, terutama ketika data memiliki fitur yang ambigu atau tidak cukup eksplisit. Oleh karena itu, diperlukan peningkatan dalam proses pelatihan model, seperti penyesuaian pada teknik pemrosesan teks atau pelatihan dengan data yang lebih representatif dan beragam untuk mengatasi kelemahan tersebut.

Penyebab utama kesalahan model berasal dari distribusi data yang tidak seimbang, di mana kelas mayoritas mendominasi dataset sementara kelas minoritas kekurangan data. Hal ini membuat model cenderung bias terhadap kelas mayoritas. Selain itu, kemiripan semantik antar kelas, seperti antara *Stress* dan *Anxiety* atau *Depression* dan *Suicidal*, membuat model kesulitan membedakan keduanya. Indikasi *overfitting* juga terlihat dari akurasi data latih yang mencapai 88,48%, sementara akurasi validasi stagnan di sekitar 73%-74% sejak epoch ke-7.

3.6 Hasil Evaluasi Model

Setelah pelatihan, model dievaluasi menggunakan data uji dan prediksi kelas dikonversi ke dalam label. Metrik yang dihitung meliputi akurasi, *F1-score* berbobot, dan *confusion matrix*. Akurasi mengukur persentase prediksi benar, *F1-score* menilai keseimbangan presisi dan *recall*, sementara *confusion matrix* membandingkan prediksi dengan label aktual. Hasil evaluasi disajikan pada Tabel 7.

Tabel 7. Hasil Evaluasi Model

Kelas	Precision	Recall	F1-score	Jumlah Data (Support)
Normal	0.76	0.81	0.78	755
Depression	0.86	0.78	0.82	527
Suicidal	0.73	0.63	0.68	3016
Anxiety	0.94	0.87	0.90	3308
Bipolar	0.42	0.78	0.55	237
Stress	0.44	0.64	0.52	536

Personality Disorder	0.63	0.69	0.66	2158
Akurasi Rata-rata Makro	0.68	0.74	0.70	10537
Rata-rata Berbobot	0.76	0.74	0.75	10537

Tabel 7 diatas menunjukkan bahwa model mencapai akurasi sebesar 74% dengan *F1-score* rata-rata berbobot 0,75, menunjukkan keseimbangan antara *precision* dan *recall* secara keseluruhan. Kelas mayoritas seperti *Anxiety* memiliki performa terbaik dengan *F1-score* 0,90, diikuti oleh *Depression* dengan *F1-score* 0,82. Sementara itu, kelas minoritas seperti *Bipolar* dan *Stress* menunjukkan performa rendah dengan *F1-score* masing-masing 0,55 dan 0,52 akibat distribusi data yang terbatas. Kelas *Personality Disorder* mencapai *F1-score* 0,66, meskipun *recall*-nya cukup tinggi (69%), menunjukkan adanya potensi perbaikan untuk meningkatkan *precision*.

Kesalahan umum ditemukan pada prediksi kelas *Suicidal* menjadi *Depression* akibat pola emosi yang mirip, serta prediksi *Depression* menjadi *Normal* yang disebabkan oleh sensitivitas model yang rendah terhadap intensitas emosional tertentu. Distribusi data yang tidak seimbang menjadi penyebab utama kesalahan ini, di mana kelas mayoritas lebih diutamakan oleh model dibandingkan kelas minoritas. Kemiripan semantik antar kelas, terutama antara *Suicidal* dan *Depression*, juga menjadi tantangan dalam membedakan pola teks yang kompleks.

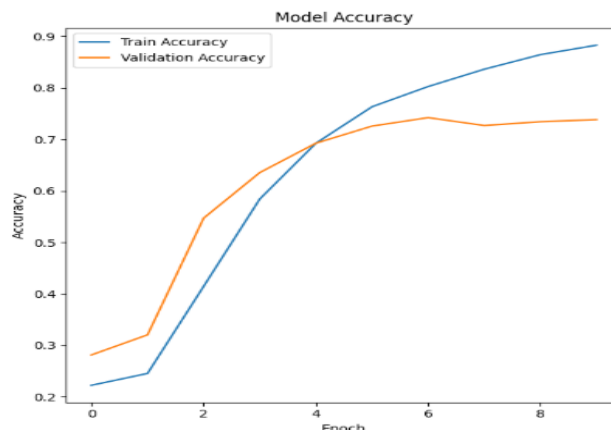
Untuk meningkatkan performa model, beberapa langkah perbaikan dapat diterapkan. Pertama, *oversampling* atau augmentasi data dapat membantu menyeimbangkan distribusi kelas minoritas. Kedua, penerapan pembobotan kelas diperlukan untuk memberikan perhatian lebih pada kelas dengan jumlah data yang kecil selama pelatihan. Ketiga, penyempurnaan pada tahap *preprocessing*, seperti penambahan fitur *sentiment scores* atau *keyword matching*, dapat meningkatkan diferensiasi antar kelas. Terakhir, penggunaan model yang lebih kompleks seperti *BERT* atau *RoBERTa* dapat meningkatkan kemampuan model dalam menangkap konteks emosional secara mendalam dan memperbaiki performa klasifikasi.

3.7 Visualisasi Hasil Evaluasi

Hasil evaluasi juga divisualisasikan dalam beberapa grafik untuk mempermudah interpretasi, antara lain:

a. Grafik Model Akurasi

Grafik model akurasi dapat dilihat pada Gambar 5 berikut.

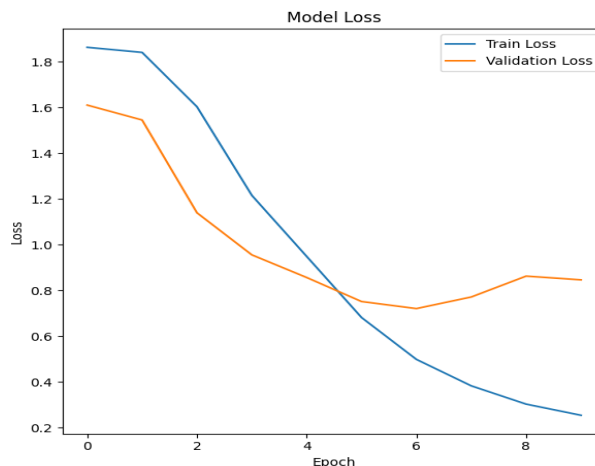


Gambar 5. Grafik Model Akurasi

Pada Gambar 5 dapat dilihat bahwa grafik akurasi menunjukkan peningkatan signifikan pada data latih dan validasi di awal pelatihan. Namun, setelah *epoch* ke-8, akurasi data latih terus meningkat sementara akurasi validasi stagnan, mengindikasikan *overfitting*. Meski demikian, model mencapai akurasi validasi 74% pada *epoch* ke-7, menunjukkan kemampuannya dalam menangkap pola data. Penggunaan *dropout* dan *early stopping* membantu mengurangi *overfitting*, meskipun tidak sepenuhnya menghilangkannya. Hasil ini menunjukkan model efektif untuk klasifikasi multi-kelas kesehatan mental dan berpotensi untuk pengembangan lebih lanjut.

b. Grafik Model Loss

Grafik ini menunjukkan perkembangan *loss* pada data latih dan data uji selama pelatihan. Grafik ini berguna untuk melihat bagaimana model meminimalkan kesalahan prediksi seiring waktu. Grafik model *loss* dapat dilihat pada Gambar 6.

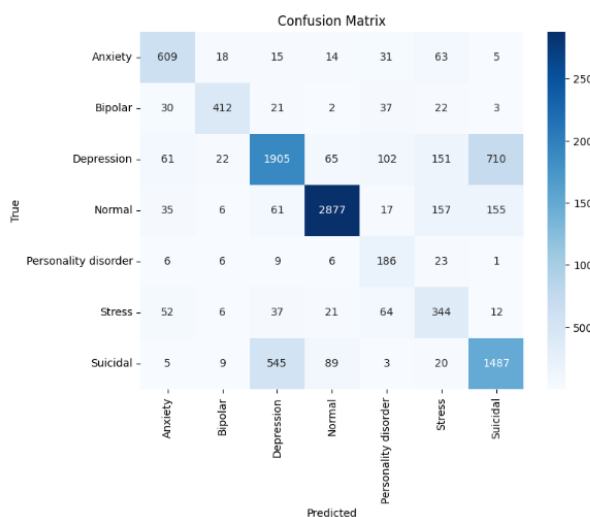


Gambar 6. Grafik Model Loss

Pada Gambar 6 dapat dilihat bahwa grafik model loss menunjukkan perubahan nilai *Train Loss* dan *Validation Loss* selama proses pelatihan. Pada awal pelatihan, kedua nilai loss cukup tinggi, menandakan bahwa model masih dalam tahap penyesuaian terhadap pola dalam data. Seiring bertambahnya *epoch*, *Train Loss* mengalami penurunan yang konsisten, menunjukkan bahwa model semakin baik dalam mengenali data latih. *Validation Loss* juga mengalami penurunan pada beberapa *epoch* pertama, tetapi mulai mengalami peningkatan setelah titik tertentu. Kondisi ini mengindikasikan *overfitting*, di mana model terlalu menyesuaikan diri dengan data latih sehingga kinerjanya terhadap data validasi menurun. Untuk mengatasi hal tersebut, dapat diterapkan teknik seperti *early stopping* untuk menghentikan pelatihan sebelum *overfitting* terjadi, regularisasi untuk mengurangi kompleksitas model, atau data *augmentation* guna meningkatkan variasi data latih sehingga model mampu melakukan generalisasi dengan lebih baik.

c. Confusion Matrix

Confusion Matrik divisualisasikan menggunakan *heatmap* untuk memudahkan interpretasi kesalahan klasifikasi model. Ini menggambarkan seberapa baik model dalam memprediksi setiap kelas dibandingkan dengan label aktual. Hasil *Confusion Matrix* dapat dilihat pada Gambar 7.

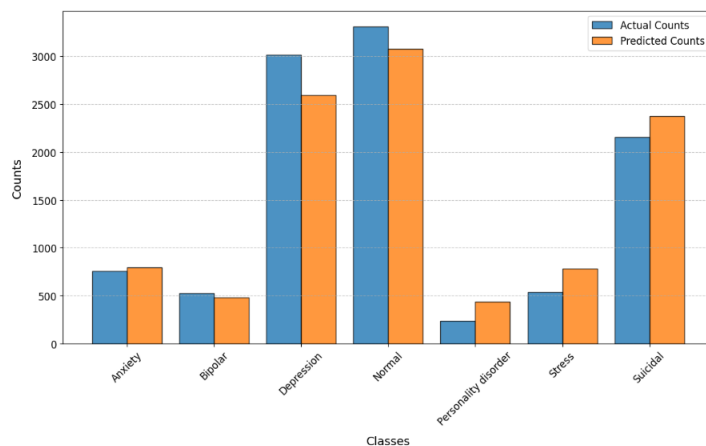


Gambar 7. Confusion Matrix

Berdasarkan *confusion matrix* pada Gambar 7, model menunjukkan kinerja baik pada kelas mayoritas, seperti Normal (2.877 prediksi benar) dan Depression (1.905 prediksi benar). Namun, terdapat kekeliruan pada kelas *Suicidal* yang sering diklasifikasikan sebagai *Depression* (545 kesalahan). Kelas dengan distribusi data kecil, seperti *Personality Disorder* dan *Bipolar*, memiliki prediksi benar yang lebih sedikit, menunjukkan kesulitan model pada kelas minoritas. Hasil ini menunjukkan potensi model pada klasifikasi multi-kelas, dengan ruang perbaikan pada kelas minoritas melalui peningkatan bobot kelas atau augmentasi data.

d. Perbandingan Jumlah Kelas Aktual dan Prediksi

Grafik batang dibawah ini membandingkan jumlah aktual dan prediksi untuk setiap kelas, memberikan gambaran apakah model cenderung mengklasifikasikan data salah pada beberapa kelas. Hasil grafik batang perbandingan jumlah kelas aktual dan prediksi dapat dilihat pada Gambar 8.



Gambar 8. Perbandingan Jumlah Kelas Aktual dan Prediksi per Kelas

Pada Gambar 8 terlihat bahwa grafik perbandingan data aktual dan prediksi menunjukkan performa model dalam klasifikasi tujuh kelas. Model akurat pada kelas mayoritas seperti Normal dan *Depression*, namun kesulitan pada kelas minoritas seperti *Personality Disorder* dan *Bipolar*. Kelas *Suicidal* menunjukkan hasil prediksi seimbang. Secara keseluruhan, model lebih akurat pada kelas mayoritas, sementara kelas minoritas membutuhkan perbaikan melalui strategi seperti penyeimbangan data atau pembobotan.

e. *Word Cloud* per Kelas

Visualisasi *word cloud* untuk setiap kelas ditampilkan pada Gambar 9 gambar tersebut menggambarkan kata-kata yang paling sering muncul dalam setiap kategori kesehatan mental. Ukuran kata menunjukkan frekuensinya, di mana kata yang lebih sering muncul ditampilkan dalam ukuran yang lebih besar.



Gambar 9. *Word Cloud* Setiap Kelas

Pada Gambar 9 diatas *word cloud* menunjukkan kata-kata yang paling sering muncul pada data teks yang diklasifikasikan ke dalam tujuh kelas (*Anxiety*, *Bipolar*, *Depression*, Normal, *Personality Disorder*, *Stress*, dan *Suicidal*). Kata-kata seperti "feel," "know," "life," dan "want" terlihat dominan di hampir semua kelas, menunjukkan bahwa topik emosional dan eksistensial menjadi tema utama pada data. Kelas mayoritas seperti Normal dan *Depression* memiliki variasi kata yang lebih banyak dibandingkan kelas minoritas seperti *Bipolar* dan *Personality Disorder*, yang mencerminkan distribusi data yang tidak merata. Secara keseluruhan, hasil evaluasi menunjukkan bahwa model dapat mengklasifikasikan data dengan akurat, namun masih ada beberapa kesalahan pada beberapa kelas tertentu. Hasil ini memberikan wawasan yang berharga tentang kinerja model dan area yang perlu diperbaiki dalam iterasi berikutnya.

4. KESIMPULAN

Model *BiLSTM* yang diterapkan dalam penelitian ini menunjukkan hasil yang cukup baik dalam mengklasifikasikan sentimen terkait kondisi kesehatan mental pada *Twitter*. Akurasi yang diperoleh mencapai 74,21%, dengan performa terbaik pada kelas *Anxiety* dan *Depression*. Penggunaan penyesuaian bobot kelas efektif dalam meningkatkan



perhatian model terhadap kelas minoritas, meskipun beberapa kelas seperti *Bipolar* dan *Personality Disorder* masih membutuhkan perbaikan lebih lanjut Untuk meningkatkan akurasi pada kelas minoritas, teknik seperti *SMOTE* untuk *oversampling*, *back-translation* untuk augmentasi data, atau pendekatan berbasis *transformer* seperti *fine-tuning BERT* dapat diterapkan. Dengan optimasi lebih lanjut, model ini berpotensi meningkatkan analisis sentimen kesehatan mental di media sosial.

REFERENCES

- [1] A. Nur Haryanti, M. Bintang Syah Putra, N. Larasati, V. Nureel Khairunnisa, and L. A. Dyah Dewi, “Analisis Kondisi Kesehatan Mental di Indonesia Dan Strategi Penanganannya,” *Student Res. J.*, vol. 2, no. 3, pp. 28–40, 2024, doi: 10.55606/srjyappi.v2i3.1219.
- [2] A. Asfahani, E. Yuniarti, L. Husnita, P. Pahmi, and N. S. Jamin, “Peningkatan Kesadaran Masyarakat Tentang Pentingnya Kesehatan Mental Melalui Edukasi Pendidikan Sosial,” *Communnity Dev. J.*, vol. 5, no. 2, pp. 3633–3639, 2024.
- [3] N. P. Sari and E. D. Prahastiwi, “Konsep Pemeliharaan Kesehatan Mental Pada Remaja Muslim Melalui Pendidikan Keimanan (Telaah Pemikiran Prof. Dr. Zakiah Daradjat),” *J. Pemikir. Keislam. dan Kemanus.*, vol. 8, no. 2, pp. 45–58, 2024.
- [4] Z. A. Rakhman, I. D. Florina, and S. Edy, “Peran Media Sosial Dalam Mendorong Diskusi Terbuka Tentang Kesehatan Mental,” *J. Ilmu Komun.*, vol. 1, no. 1, pp. 34–40, 2024.
- [5] A. S. Nandila, “Analisis Sentimen Pada Media Sosial Twitter Terhadap Isu Kesehatan Mental Pasca Pandemi Covid-19 Dengan Menggunakan Metode Naïve Bayes & Support Vector Machine (SVM),” Universitas Nasional, 2024.
- [6] Y. Astari, Afiyati, and S. W. Rozaqi, “Analisis Sentimen Multi-Class pada Sosial media Menggunakan Metode Long Short-Term Memory (LSTM),” *J. Linguist. Komputasional*, vol. 3, no. 2, p. 6, 2021.
- [7] K. Aulia and L. Amelia, “Analisis Sentimen Twitter Pada Isu Mental Health Dengan Algoritma Klasifikasi Naive Bayes,” *Siliwangi J. (Seri Sains Teknol.)*, vol. 6, no. 2, pp. 60–65, 2020.
- [8] H. H. Windjatika and W. Maharani, “Depression Detection on Social Media Twitter Using Long Short-Term Memory,” *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 1835, 2022, doi: 10.30865/mib.v6i4.4457.
- [9] L. Yosia Wibowo, N. Annisa, P. Ananda Khairunnisa, V. Handrianus Pranatawijaya, and R. Priskila, “Implementasi Long Short-Term Memory Dalam Analisis Sentimen Pengguna Aplikasi Twitter Yang Mengandung Ujaran Kebencian,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 3170–3174, 2024, doi: 10.36040/jati.v8i3.9654.
- [10] A. Karami, M. Lundy, F. Webb, and Y. K. Dwivedi, “Twitter and Research: A Systematic Literature Review through Text Mining,” *IEEE Access*, vol. 8, pp. 67698–67717, 2020, doi: 10.1109/ACCESS.2020.2983656.
- [11] R. W. Pratiwi, Y. Sari, and Y. Suyanto, “Attention-Based BiLSTM for Negation Handling in Sentimen Analysis,” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 14, no. 4, p. 397, 2020, doi: 10.22146/ijccs.60733.
- [12] M. F. B. A. A. Syam, G.H.M., A. Salim, D. F. Surianto, “Analisis Teknik Preprocessing pada Sentimen Masyarakat Terkait Konflik Israel-Palestina Menggunakan Support Vector Machine,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 3, pp. 1464–1472, 2024.
- [13] M. A. Nur and N. Wardhani, “Optimasi Normalisasi Kata Pada Data Twitter Untuk Meningkatkan Akurasi Analisis Sentimen (Studi Kasus Respon Masyarakat Terhadap Layanan Teman Bus),” *J. Fokus Elektroda (Energi List. Telekomun. Komputer, Elektron. dan Kendali)*, vol. 7, no. 4, pp. 237–243, 2022.
- [14] S. J. Angelina, A. Bijaksana, P. Negara, and H. Muhardi, “Analisis Pengaruh Penerapan Stopword Removal Pada Performa Klasifikasi Sentimen Tweet Bahasa Indonesia,” *JUARA (J. Apl dan Ris. Inform.)*, vol. 02, no. 1, pp. 165–173, 2023, doi: 10.26418/juara.v2i1.69680.
- [15] C. Herdian, A. Kamila, and I. G. Agung Musa Budidarma, “Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi,” *Technol. J. Ilm.*, vol. 15, no. 1, p. 93, 2024, doi: 10.31602/tji.v15i1.13457.
- [16] R. Onsu, D. febrina Sengkey, and F. D. Kambey, “Implementasi Bi-LSTM Dengan Ekstraksi Fitur Word2vec Untuk Pengembangan Analisis Sentimen Aplikasi Identitas Kependudukan Digital,” *J. Teknol. Terpadu*, vol. 10, no. 1, pp. 46–55, 2024.
- [17] A. Saputra, R. C. Sigitta Hariyono, and N. M. Saraswati, “Analisis Sentimen Pengguna Aplikasi MyPertamina Menggunakan Algoritma Bidirectional Long Short Term Memory,” *J. Eksplora Inform.*, vol. 13, no. 2, pp. 156–163, 2024, doi: 10.30864/eksplora.v13i2.973.
- [18] Y. Karyadi, “Prediksi Kualitas Udara Dengan Metoda LSTM, Bidirectional LSTM, dan GRU,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 671–684, 2022, doi: 10.35957/jatisi.v9i1.1588.
- [19] Z. Hameed and B. Garcia-Zapirain, “Sentiment Classification Using a Single-Layered BiLSTM Model,” *IEEE Access*, vol. 8, pp. 73992–74001, 2020, doi: 10.1109/ACCESS.2020.2988550.
- [20] Teknokrat, “Optimalkan Analisis Data dengan Pentingnya Pra-pemrosesan Data Teks,” FTIK Teknokrat. Accessed: Jan. 20, 2025. [Online]. Available: <https://ftik.teknokrat.ac.id/optimalikan-analisis-data-dengan-pentingnya-pra-pemrosesan-data-teks/>
- [21] DqLab, “Menggali Sentimen dengan Sistem Operasi SQL,” DQLab. Accessed: Jan. 20, 2025. [Online]. Available: <https://dqlab.id/menggali-sentimen-dengan-sistem-operasi-sql>
- [22] A. Kartika Sari, I. Akhmad, N. A. Dinda, Islamiyah, and G. Stephanie Elfriede, “Analisis Sentimen Twitter Menggunakan Machine Learning untuk Identifikasi Konten Negatif,” *Adopsi Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 64–73, 2024, doi: 10.30872/atasi.v3i1.1373.