Analisis Sentimen Multimodal terhadap Opini Publik Mengenai Kesehatan Mental di Media Sosial X dengan Metode CNN BiLSTM dan Ekspansi fitur FastText

Tugas Akhir

diajukan untuk memenuhi salah satu syarat

memperoleh gelar sarjana

pada Program Studi S1 Informatika

Fakultas Informatika Universitas Telkom

1301210496

Nadim Rafli Hamzah

Program Studi Sarjana S1 Informatika

Fakultas Informatika

Universitas Telkom

Bandung 2025

LEMBAR PENGESAHAN

Analisis Sentimen Multimodal terhadap Opini Publik Mengenai Kesehatan Mental di Media Sosial X dengan Metode CNN BiLSTM dan Ekspansi fitur FastText

Multimodal Sentiment Analysis of Public Opinion on Mental Health on Social Media X with CNN BiLSTM Method with FastText Feature Expansion

NIM : 1301210496

Nadim Rafli Hamzah

Tugas akhir ini telah diterima dan disahkan untuk memenuhi sebagian syarat memperoleh gelar

pada Program Studi Sarjana S1 Informatika

Fakultas Informatika

Universitas Telkom

Bandung, 15 Juli 2025

Menyetujui

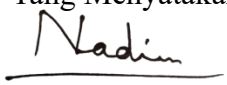
Pembimbing	Ketua Program Studi Sarjana Informatika
Dr. Erwin Budi Setiawan, S.Si., M.T.	Mahmud Dwi Sulistiyo, S.T., M.T., Ph.D.
NIP: 00760045	NIP: 13880017

LEMBAR PERNYATAAN

Dengan ini saya, Nadim Rafli Hamzah, menyatakan sesungguhnya bahwa Tugas Akhir saya dengan judul Analisis Sentimen Multimodal terhadap Opini Publik Mengenai Kesehatan Mental di Media Sosial X dengan Metode CNN BiLSTM dan Ekspansi fitur FastText beserta dengan seluruh isinya adalah merupakan hasil karya sendiri, dan saya tidak melakukan penjiplakan yang tidak sesuai dengan etika keilmuan yang belaku dalam masyarakat keilmuan, serta produk dari tugas akhir bukan merupakan produk dari *Generative AI* . Saya siap menanggung resiko/sanksi yang diberikan jika di kemudian hari ditemukan pelanggaran terhadap etika keilmuan dalam Laporan TA atau jika ada klaim dari pihak lain terhadap keaslian karya,

Bandung, 15 Juli 2025

Yang Menyatakan



Nadim Rafli Hamzah
1301210496

Nadim Rafli Hamzah[1], Erwin Budi Setiawan[2]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]nadimdim@students.telkomuniversity.ac.id,
[2]erwinbudisetiawan@telkomuniversity.ac.id,

Abstrak

Dalam hal kesehatan masyarakat global, kesehatan mental menjadi perhatian penting. Salah satu situs media sosial yang paling populer, X telah berkembang menjadi forum bagi orang-orang untuk mendiskusikan masalah kesehatan mental dan berbagi cerita pribadi. Menganalisis sentimen dalam percakapan online ini penting untuk memahami persepsi publik dan memandu intervensi kesehatan mental. Penelitian ini mengusulkan model analisis sentimen menggunakan multimodal yang memanfaatkan data tekstual dan visual, dengan fitur teks yang diekstraksi melalui CNN-BiLSTM, TF-IDF, dan FastText, dan fitur gambar menggunakan VGG-16. Klasifikasi sentimen dilakukan dengan menggunakan model Hybrid CNN-BiLSTM dengan mekanisme perhatian. Model ini menggunakan fusi tingkat menengah untuk mengintegrasikan fitur teks dan gambar, diikuti dengan tingkat keputusan untuk menggabungkan output dari model teks saja, gambar saja, dan multimodal. 24.742 pasangan tweetgambar dikumpulkan dari platform X dan dianotasi melalui sistem pemungutan suara mayoritas. Untuk membangun korpus kemiripan FastText, 63.512 data dari portal berita digital CNN (*Cable News Network*) dan X digabungkan. Dengan akurasi 87,92%, model multimodal mengungguli model teks saja sebesar 0,09% dan model gambar saja sebesar 25,10%. Hasil ini menunjukkan keefektifan data modalitas, ekstraksi fitur yang komprehensif, dan multimodal.

Kata kunci: ***Analisis Sentimen, FastText, Hybrid CNN-BiLSTM, TF-IDF, VGG-16***

Abstract

In terms of global public health, mental health is a significant concern. One of the most popular social media sites, X has developed into a forum for people to discuss mental health issues and share personal stories. Analyzing sentiment in these online conversations is important to understand public perception and guide mental health interventions. This research proposes a sentiment analysis model using multimodal that utilizes both textual and visual data, with text features extracted via CNN-BiLSTM, TF-IDF, and FastText, and image features using VGG-16. Sentiment classification is performed using a Hybrid CNN-BiLSTM model with an attention mechanism. This model uses mid-level fusion to integrate text and image features, followed by decision-level fusion to combine the output of text-only, image-only, and multimodal models. 24,742 tweet-image pairs were collected from the X platform and annotated through a majority voting system. To build the FastText similarity corpus, 63,512 data from digital news portals CNN (Cable News Network) and X were combined. With an accuracy of 87.92%, the multimodal model outperformed the text-only model by 0.09% and the image-only model by 25.10%. These results demonstrate the effectiveness of modality data, comprehensive feature extraction, and multimodal.

Keywords: ***FastText, Hybrid CNN-BiLSTM, Sentiment Analysis, TF-IDF, VGG-16***

1. Pendahuluan

Latar Belakang

Produktivitas masyarakat dan kualitas hidup individu dipengaruhi secara signifikan oleh kesehatan mental. Menurut WHO, depresi memengaruhi lebih dari 264 juta orang di seluruh dunia, angka yang terus meningkat setiap tahunnya. Sayangnya, stigma sosial masih menjadi penghalang utama untuk mencari bantuan profesional. Oleh karena itu, penelitian di bidang kesehatan mental sangat penting untuk merumuskan solusi preventif dan responsif terhadap gangguan psikologis [1].

Dengan lebih dari 450 juta pengguna aktif bulanan, X merupakan salah satu jaringan media sosial terbesar yang menyediakan data real-time dalam bentuk teks pendek dan media visual. Banyak pengguna yang secara terbuka membagikan perasaan, pengalaman pribadi, dan bahkan keinginan untuk bunuh diri melalui cuitan mereka, menjadikan X sebagai sumber data yang potensial untuk deteksi dini gangguan mental. Setelah pandemi COVID-19, tren penggunaan X untuk mengekspresikan masalah kesehatan mental telah meningkat, seperti yang terlihat dari tagar populer seperti *#MentalHealthAwareness* dan *#Depression* [2], [3].

Cuitan di X tidak hanya berisi informasi verbal, tetapi juga visual melalui gambar yang menyertainya. Analisis multimodal yang menggabungkan teks dan gambar telah terbukti lebih unggul daripada pendekatan unimodal karena memberikan konteks yang lebih lengkap. Sebagai contoh, teks yang netral dapat menunjukkan tanda-tanda depresi ketika digabungkan dengan gambar yang menyedihkan. Model seperti CNN-BiLSTM dan penyematan FastText telah terbukti meningkatkan akurasi prediksi kondisi mental dalam pendekatan ini [4].

Pengembangan kerangka kerja multimodal yang menggabungkan *Hybrid* CNN-BiLSTM dan TF-IDF adalah kontribusi utama dari penelitian ini. Kerangka kerja ini disempurnakan melalui FastText untuk perluasan fitur dalam ekstraksi fitur tekstual dan untuk prediksi sentimen pada subjek kesehatan mental, VGG-16 untuk ekstraksi fitur visual. Sementara itu, strategi fusi menengah menggabungkan karakteristik visual dan tekstual, dan output dari model teks saja dan gambar saja digabungkan pada fusi tingkat keputusan. Para penulis mengklaim bahwa belum ada penelitian sebelumnya yang meneliti analisis sentimen multimodal dengan ekstraksi TF-IDF dalam model prediksi sentimen. Oleh karena itu, tujuan dari penelitian ini adalah untuk membuat kerangka kerja seperti itu karena, menurut hasil penelitian sebelumnya, hal tersebut dapat meningkatkan akurasi.

Topik dan Batasannya

Kesehatan mental merupakan masalah besar yang sangat memengaruhi kualitas hidup individu dan produktivitas masyarakat. Tetapi stigma sosial sering menghalangi orang untuk mencari bantuan. Sebaliknya, media sosial, terutama X, telah berkembang menjadi platform utama bagi pengguna untuk mengungkapkan opini dan emosi yang berkaitan dengan kesehatan mental. X menyediakan banyak data yang dapat digunakan untuk mengidentifikasi lebih baik psikologi pengguna dengan fitur teks pendek dan fitur multimodal seperti gambar.

Perumusan Masalah

1. Bagaimana cara mengimplementasikan analisis sentimen multimodal menggunakan metode CNN-BiLSTM dengan ekspansi fitur FastText untuk mengingkatkan akurasi klasifikasi sentimen?
2. Bagaimana performansi sistem analisis sentimen multimodal yang dibangun menggunakan model CNN-BiLSTM dan ekspansi fitur FastText dalam mengklasifikasi sentimen?

Tujuan

Tujuan utama penelitian ini adalah untuk membuat sistem multimodal yang dapat menganalisis opini publik tentang kesehatan mental di X. Untuk mendapatkan pola semantik teks yang lebih baik, teknik yang digunakan termasuk model *Convolutional Neural Network* (CNN) dan *Bidirectional Long Short-Term Memory* (BiLSTM), yang didukung oleh embedding FastText. Penelitian ini membuat pendekatan multimodal yang memberikan pemahaman data yang lebih luas dengan menggabungkan analisis teks dan gambar.

Berikut beberapa tujuannya:

1. Mengimplementasikan analisis sentimen multimodal dengan menggunakan metode CNN-BiLSTM dan ekspansi fitur FastText untuk meningkatkan akurasi klasifikasi sentimen.
2. Mengukur performansi sistem analisis sentimen multimodal dengan menggunakan metode CNN-BiLSTM dan ekspansi fitur FastText.

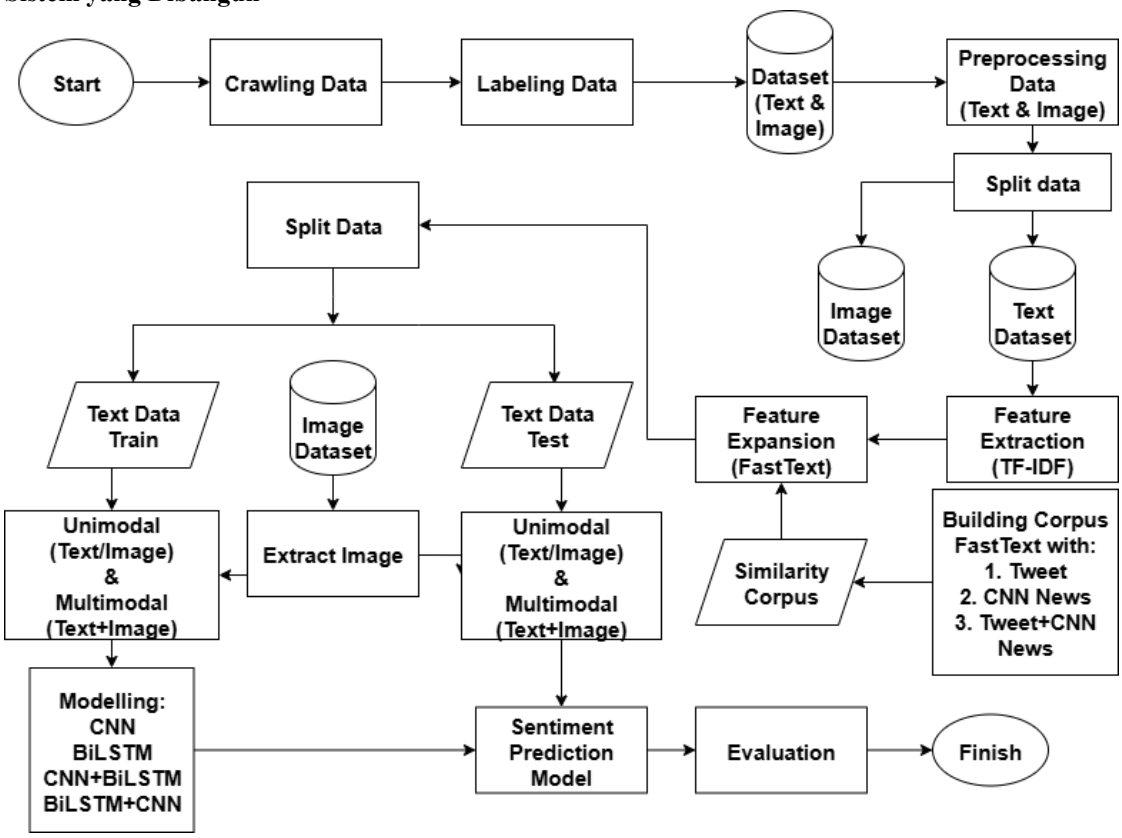
2. Studi Terkait

Beberapa penelitian telah meneliti penerapan teknologi pada analisis kesehatan mental di media sosial, khususnya X, yang menyediakan data yang melimpah untuk memahami opini publik tentang kesehatan mental. Pada penelitian pertama mengembangkan model CNN-BiLSTM yang didukung oleh FastText untuk mendeteksi gangguan mental dan keinginan untuk bunuh diri dari data X. Penelitian ini menunjukkan bahwa pendekatan multimodal dapat meningkatkan akurasi prediksi karena FastText memiliki kemampuan untuk menangkap representasi semantik teks dengan lebih baik.[5]

Penelitian kedua menampilkan peningkatan hasil analisis sentimen berbasis X dengan menggunakan fusi fitur pada tingkat representasi. Penelitian ini berhasil menunjukkan bahwa fusi fitur tingkat tinggi dapat memberikan hasil yang lebih akurat pada analisis data berbasis teks pendek seperti X, karena model yang digunakan menggabungkan CNN-BiLSTM dengan embedding FastText yang berfokus pada klasifikasi emosi melalui perhatian pada karakteristik data.[6]

Untuk mendeteksi depresi di media sosial, penelitian yang dilakukan oleh Ghosh dan Al Banna (2023) memperkenalkan pendekatan *hybrid* berbasis BiLSTM-CNN dengan arsitektur perhatian. Mereka menggunakan penyematan FastText untuk mengekstrak fitur teks yang lebih rinci, dan untuk meningkatkan interpretabilitas prediksi, mereka menerapkan anotasi pada model. Metode ini dapat digunakan untuk memahami masalah kesehatan mental dari sudut pandang pengguna media sosial.[7]

3. Sistem yang Dibangun



Gambar 1. Alur kerja sistem menggunakan arsitektur hibrida CNN-BiLSTM

Pipeline prediksi sentimen yang menggunakan model hibrida CNN-BiLSTM dengan ekspansi fitur FastText digambarkan dalam flowchart pada Gambar 1. C*rawling data*, *Pre-processing* data, ekstraksi fitur TF-IDF, perluasan fitur FastText, dan model CNN, BiLSTM, *hybrid* CNN-BiLSTM, dan *hybrid* BiLSTM-CNN adalah beberapa langkah yang membentuk sistem.

3.1. Crawling Data

Crawling data mengumpulkan sejumlah besar tweet tekstual dan visual untuk analisis sentimen percakapan X seputar kesehatan mental [8]. Memanfaatkan kata kunci, periode tanggal, atau akun pengguna, alat seperti Tweet-harvest mengekstrak tweet yang relevan, menyimpan data dalam format CSV bersama dengan stempel waktu, URL gambar, teks tweet, dan metadata untuk analisis multimodal. Sebaran data hasil dari *crawling* dapat dilihat pada tabel 1.

Tabel 1. Jumlah dataset berdasarkan kata kunci

Keyword	*Quantity*
broken	2,078
burnout	3,000
frustration	2,500
unloved	3,000
anxiety	2,800
selflove	3,570
mentalhealthmatters	2,500
wellbeing	2,794
toxic	2,500
Total	24,742

3.2. Labeling Data

Dalam analisis sentimen multimodal, pelabelan data sangat penting, terutama untuk cuitan X yang berkaitan dengan kesehatan mental [8]. Setelah pengumpulan data, setiap tweet yang mencakup teks dan gambar diklasifikasikan sebagai negatif atau positif dengan melibatkan 5 anotator dengan prinsip *majority vote* yang ditampilkan pada tabel 2. Kedua modalitas tersebut diberi label, dan nada emosionalnya bisa jadi sama atau berbeda. Para anotator bekerja sama dan menyelesaikan perbedaan pendapat melalui dialog untuk menjaga konsistensi. Prosedur metode ini menghasilkan kumpulan data yang dapat dipercaya yang diperlukan untuk evaluasi dan pelatihan model hibrida CNN-BiLSTM, meningkatkan kapasitas model untuk menguraikan berbagai ekspresi emosional.

Tabel 2. Distribusi label sentimen dari dataset

Category	Label	Quantity
Text	*Positive*	12,429
	Negative	12,313
Image	*Positive*	12,429
	Negative	12,313
Total		24,742

3.3. Pre-Processing Data

Untuk meningkatkan kualitas data *input* dan kinerja klasifikasi, *pre-processing* sangat penting untuk analisis sentimen multimodal. *Pre-processing* data untuk ekstraksi TF-IDF melibatkan beberapa langkah *Pre-processing* menggunakan pustaka NLTK. Setelah tokenisasi, lemmatization, dan penghilangan stopword, data teks diubah menjadi vektor kata dan menjalani analisis CNN untuk mengekstrak fitur lokal dan analisis BiLSTM untuk memahami konteks sekuensial. Sementara itu, data gambar mengalami augmentasi (rotasi, flipping), konversi warna, dan normalisasi ukuran [9].

1) *Pre-Processing* Teks menjamin data yang jelas dan konsisten.
   a) *Cleaning*: Gunakan ekspresi reguler untuk menghilangkan referensi, tagar, URL, angka, dan simbol.
   b) *Case Folding*: Untuk memastikan konsistensi, ubah teks menjadi huruf kecil.
   c) *Normalization*: Membakukan bahasa gaul dan menggunakan ejaan yang tepat.
   d) *Stopword Removal*: Menghilangkan istilah umum seperti “dan” dan “yang.”
   e) *Stemming*: Memecah kata menjadi bagian-bagian yang paling dasar.
   f) *Tokenization*: Gunakan NLTK untuk penyematan untuk membagi teks menjadi token.

2) *Pre-processing* gambar dilakukan dengan menggunakan *Python Imaging Library* untuk memastikan semua gambar kompatibel dengan model VGG 16 sebelum ekstraksi fitur. Langkah-langkah *pre-processing* meliputi:
   a) *Image Resizing*: Mengubah ukuran semua gambar ke ukuran tetap 224x224 piksel.
   b) *Format Conversion*: Gambar dalam format non-RGB dikonversi ke RGB untuk memastikan

konsistensi.

c) *Channel Reordering*: Fungsi input *Pre-processing* modul VGG-16 digunakan untuk mengubah gambar dari RGB ke BGR.

3.4. Ekstraksi Fitur TF-IDF

Ekstraksi fitur dilakukan setelah tahap persiapan *Pre-processing* data. Tahap pertama dari klasifikasi teks adalah ekstraksi fitur. Tujuannya adalah untuk mengubah teks menjadi representasi vektor di mana setiap kata diberi bobot [10]. Langkah awal dalam representasi X adalah pemodelan fitur N-gram, yang meliputi unigram, bigram, dan trigram. Ekstraksi fitur yang digunakan dalam penelitian kali ini *Term Frequency-Inverse Document Frequency* (TF-IDF). Pendekatan ini banyak digunakan oleh banyak peneliti karena dianggap efisien, mudah digunakan, dan memberikan hasil yang dapat dipercaya [11]. TF-IDF merupakan kombinasi dari *Term Frequency* (TF), yang menghitung frekuensi term dalam sebuah dokumen, dan *Inverse Document Frequency* (IDF), yang membobotkan term dalam sebuah dokumen. Akibatnya, kata-kata yang sering digunakan dalam berbagai dokumen akan memiliki nilai yang lebih rendah daripada kata-kata yang lebih jarang digunakan. Rumus TF-IDF diberikan oleh persamaan 1 dan 2.

$$W_{\mathrm{dt}} = TF_{\mathrm{dt}} \; x \; IDF_{\mathrm{dt}} \qquad (1)$$

$$IDF_{\mathrm{dt}} = (\log (\frac{N}{dt})) \qquad (2)$$

W adalah bobot kata d relatif terhadap kata t dalam Persamaan (1) dan (2), df adalah jumlah dokumen yang menyertakan kata yang terdeteksi, dan N adalah jumlah total dokumen dalam set data.

3.5. Ekspansi Fitur FastText

Paradigma representasi kata yang disebut FastText diciptakan untuk mengatasi kelemahan penyematan konvensional seperti Word2Vec. Tidak seperti Word2Vec, yang hanya menggunakan vektor kata penuh, FastText membuat representasi kata berdasarkan subkata atau n-gram, yang memungkinkan FastText untuk mendapatkan lebih banyak informasi dari kata-kata yang tidak ada di dalam kosakata. Untuk bahasa dengan morfologi yang kompleks, seperti bahasa Indonesia, pendekatan berbasis subkata menjadi sangat penting[12]. Hal ini dikarenakan awalan, akhiran, dan imbuhan sering kali mengubah arti kata secara signifikan[13].

Tabel 4. Korpus *similarity* Fasttext

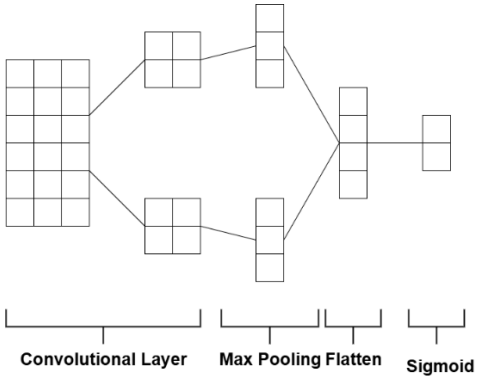
Corpus	*CNN News*	*Tweets*	*Tweets + CNN News*
Total	**63,512**	**24,742**	**88,254**

Untuk membuat *similarity corpus*, setiap kata yang ditemukan dalam data X dan atau CNN News digabungkan. Selanjutnya, FastText digunakan untuk menghitung kesamaan antar kata menggunakan tiga jenis data: data tweet, CNN News dan kombinasi keduanya. *Similarity corp*us akhir dibuat dengan menguji setiap dataset, termasuk Tweet dataset, CNN News dataset dan *combined* dataset seperti yang ditampilkan pada tabel 4. Dengan menggunakan FastText untuk menghasilkan *similarity corpus*.

3.6. Algoritma Klasifikasi

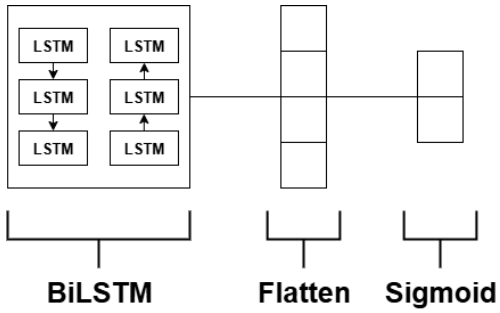
Seperti yang diilustrasikan pada Gambar 1, vektor yang dihasilkan akan digunakan sebagai input untuk pengembangan model *Hybrid Deep Learning* yang menggunakan *Convolutional Neural Networks* (CNN) dan *Bidirectional Long Short-Term Memory* (BiLSTM) untuk mendeteksi dengan melakukan kategorisasi setelah prosedur perluasan fitur selesai.

Salah satu jenis jaringan saraf yang menggunakan struktur konvolusi untuk mengekstrak vektor fitur lokal disebut *Convolutional Neural Network* (CNN) [14]. Input dari lapisan berikutnya dari jaringan *multilayer* CNN adalah output dari lapisan sebelumnya. Hasilnya, jaringan ini terdiri dari beberapa lapisan tersembunyi, *input,* dan *output* [15]. Arsitektur CNN[16] yang digunakan pada penelitian ini ditampilkan pada Gambar 2.



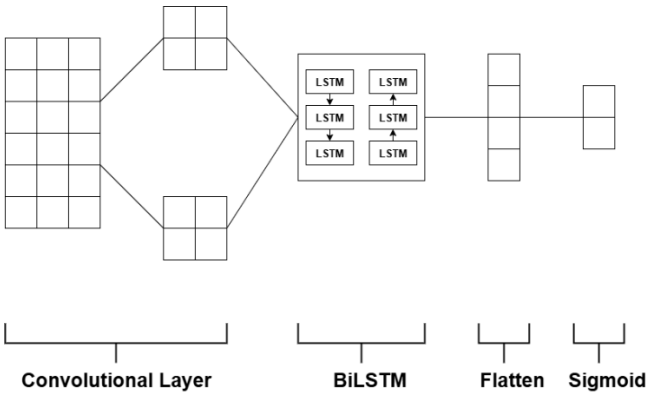
Gambar 2. Arsitektur CNN[16]

Bi-directional Long Short-Term Memory (BiLSTM) adalah kombinasi dari LSTM maju dan mundur. Untuk mengekspresikan sebuah kalimat, model LSTM belajar melalui pelatihan tentang apa yang harus diingat dan apa yang harus dilupakan. Namun demikian, informasi tidak dapat dikodekan dari belakang ke depan menggunakan model LSTM. Oleh karena itu, ketergantungan semantik dua arah harus ditangkap oleh BiLSTM [14]. Arsitektur BiLSTM[16] yang digunakan pada penelitian ini ditampilkan pada Gambar 3.

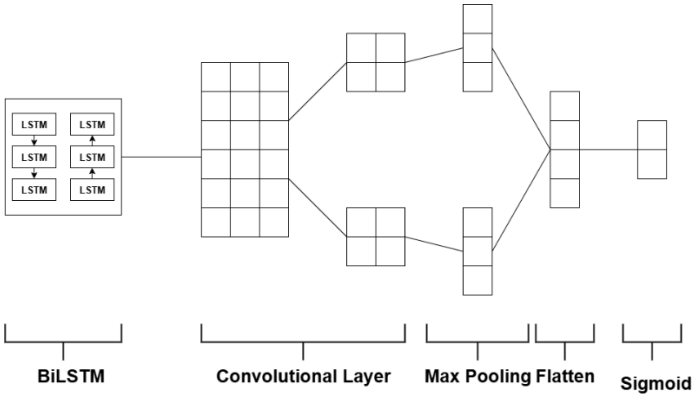


Gambar 3. Arsitektur BiLSTM

Oleh karena itu, dengan menggunakan keunggulan dari kedua model tersebut, proyek ini akan bereksperimen dengan menggabungkan CNN dan BiLSTM untuk menghasilkan model *Hybrid* CNN-BiLSTM dan *Hybrid* BiLSTM-CNN. BiLSTM dapat menangkap ketergantungan jarak jauh dan konteks masa lalu dan masa depan, sedangkan CNN dapat menangkap pola lokal [16]. Gambaran rinci tentang arsitektur *Hybrid* CNN-BiLSTM[16] dan *Hybrid* BiLSTM-CNN[16] dapat dilihat pada Gambar 4 dan 5.



Gambar 4. Arsitektur CNN-BiLSTM



Gambar 5. Arsitektur BiLSTM-CNN

4. Evaluasi

CNN, BiLSTM, CNN-BiLSTM, dan BiLSTM-CNN adalah empat model algoritma klasifikasi yang akan menjadi subjek dari beberapa pengujian dalam penelitian ini. Di sini, temuan-temuan tersebut akan dijelaskan dan diuraikan ke dalam beberapa skenario berikut.

4.1. Skenario 1

Rasio pembagian data pada skenario pertama 90:10, 80:20, dan 70:30. Diuji untuk menemukan rasio data *training* optimal yang akan diterapkan pada skenario berikutnya. Selain mengevaluasi rasio data, skenario ini juga menguji fitur maksimum optimal yang akan digunakan pada skenario berikutnya. Skala fitur maksimum berkisar antara 5.000 hingga 15.000 fitur.

Nilai akurasi dari kedua model, dengan menggunakan fitur maksimum 10.000 untuk semua proporsi data yang dibagi, ditampilkan pada Tabel 5. *Baseline* yang digunakan untuk skenario berikut ini adalah TF-IDF dengan split 80:20. Karena, jika dibandingkan dengan split yang lain, temuan pada Tabel 5 data split 80:20 memperoleh nilai akurasi terbaik.

Tabel 5. Nilai akurasi dengan split rasio

Split Ratio	***Accuracy (%)***	
	CNN	***BiLSTM***
90:10	85.97	87.38
80:20	**86.21**	**87.79**
70:30	85.72	87.55

Tabel 5. Nilai akurasi dengan fitur maksimal

Data Ratio	***Max Feature***	***Accuracy (%)***	
		CNN	***BiLSTM***
80:20	5000	85.11	83.32
	10000	**86.21**	**87.79**
	15000	86.24	85.61

4.2. Skenario 2

Skenario kedua akan membandingkan TF-IDF n-gram seperti Unigram, Bigram, Trigram, Unigram+Bigram, atau Allgram. Karena split data 80:20 memiliki tingkat akurasi tertinggi di antara split data pada skenario sebelumnya, maka kedua model yang akan diuji pada skenario ini menggunakan split data 80:20 dan fitur maksimum 10.000.

Nilai akurasi kedua model dari perbandingan n-gram dengan split 80:20 dan fitur maksimum 10.000 ditampilkan pada Tabel 6. Model CNN dengan n-gram Unigram + Bigram dan model BiLSTM dengan n-gram Unigram adalah *baseline* yang digunakan untuk skenario berikutnya.

Tabel 6. Nilai akurasi dengan perbandingan n-gram

N-Gram	***Accuracy (%)***	
	CNN	***BiLSTM***
Unigram (*Baseline*)	86.21	**87.79**
Bigram	75.80	77.82
Trigam	63.46	63.67
Unigram + Bigram	**86.57**	87.30
Allgram	86.36	87.20

4.3. Skenario 3

Pada skenario berikutnya, model CNN dan BiLSTM diuji dengan menggunakan fitur perluasan yang diperoleh dari korpus kemiripan FastText. Dalam proses ini, fitur perluasan membantu meningkatkan representasi teks dengan menangkap lebih banyak hubungan semantik antar kata. Kami juga memeriksa beberapa peringkat "N" teratas dari istilah yang paling mirip dalam korpus untuk menentukan metrik kemiripan optimal yang akan menghasilkan presisi maksimum. Top 1, Top 5, Top 10, Top 15, Top 20, dan Top 25 adalah beberapa peringkat teratas yang diperiksa dalam investigasi ini.

Model CNN memiliki akurasi yang meningkat dengan kurasi hingga 0,23%, menurut hasil pengujian model tunggal pada Tabel 7. Representasi semantik berhasil ditingkatkan oleh FastText. Namun, akurasi model BiLSTM tidak meningkat.

Tabel 7. Nilai akurasi model CNN & BiLSTM dengan fasttext

Model	***Rank***	***Accuracy (%)***			
		Baseline	***Tweet***	***News***	***Tweet+News***
CNN	Top 1	86.57	85.74	85.81	86.29
	Top 5		86.72	86.22	**86.77**
	Top 10		85.88	86.84	86.29
	Top 15		86.24	84.85	86.24
	Top 20		-	-	-
	Top 25		-	-	-
BiLSTM	Top 1	87.79	85.55	86.24	85.91
	Top 5		85.95	86.03	86.22
	Top 10		85.81	86.43	86.58
	Top 15		86.48	85.68	**86.89**
	Top 20		-	-	-
	Top 25		-	-	-

4.4. Skenario 4

Pada skenario berikutnya, fitur-fitur yang diperluas yang diambil dari korpus kemiripan FastText juga digunakan untuk menguji model BiLSTM-CNN dan *Hybrid* CNN-BiLSTM. Dalam proses ini, fitur-fitur yang diperluas membantu meningkatkan representasi teks dengan menangkap lebih banyak hubungan semantik antar kata. Kami juga memeriksa beberapa peringkat "N" teratas dari istilah yang paling mirip dalam korpus untuk menentukan metrik kemiripan optimal yang akan menghasilkan presisi maksimum. Top 1, Top 5, Top 10, Top 15, Top 20, dan Top 25 adalah beberapa peringkat teratas yang diperiksa dalam investigasi ini.

Berdasarkan temuan pengujian pada Tabel 8, akurasi model hibrida untuk CNN-BiLSTM dan BiLSTM-CNN masing-masing adalah 87,66% dan 87,84%, yang mengindikasikan bahwa kedua model tersebut telah meningkat jika dibandingkan dengan model tunggal.

Tabel 8. Nilai akurasi model CNN-BILSTM & BILSTM-CNN dengan fasttext

Model	***Rank***	***Accuracy (%)***		
		Tweet	***News***	***Tweet+News***
CNN-BiLSTM	Top 1	86.99	85.16	86.74
	Top 5	86.99	84.72	86.53
	Top 10	87.24	85.19	86.32
	Top 15	**87.66**	85.59	86.67
	Top 20	87.57	-	-
	Top 25	-	-	-
BiLSTM-CNN	Top 1	87.38	87.41	87.95
	Top 5	87.36	87.06	87.65
	Top 10	87.44	86.81	87.32
	Top 15	87.75	86.74	87.21
	Top 20	**87.84**	-	-
	Top 25	87.75	-	-

4.5. Skenario 5

Pada skenario terakhir, efektivitas pemodelan multimodal dievaluasi dengan membandingkan kinerja model teks dan gambar independent (unimodal) yang menggabungkan prediksi dari teks dan gambar.

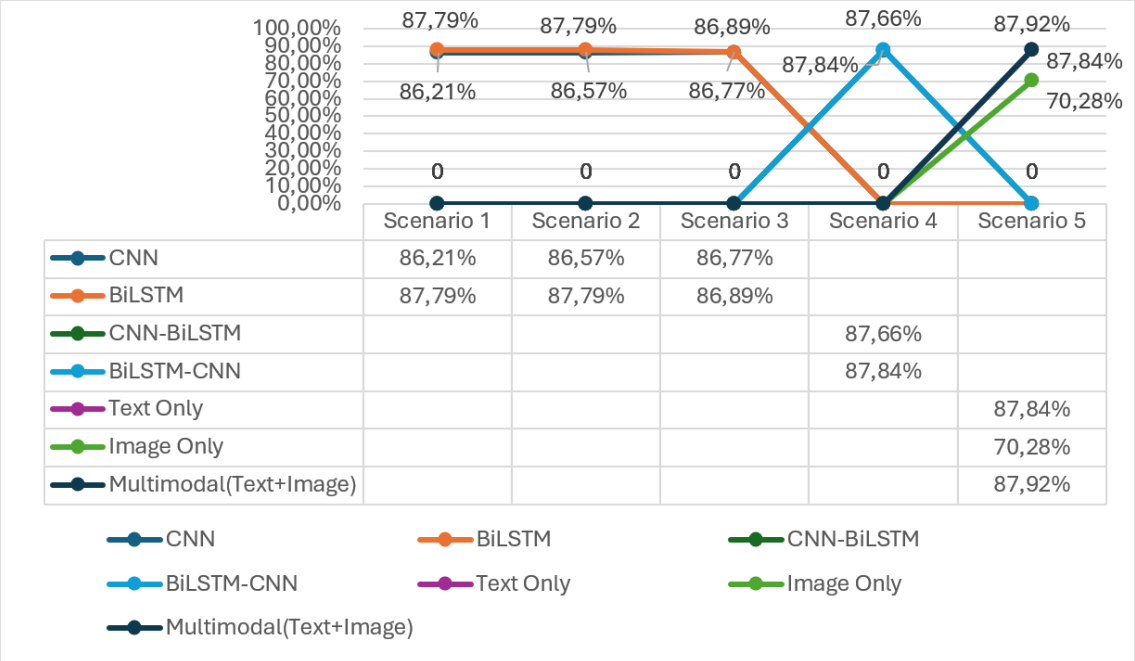
Hasil Tabel 9 menunjukkan bahwa, dengan akurasi 87,92%, pendekatan multimodal-yang memadukan teks dan gambar-memberikan hasil yang paling baik secara keseluruhan. Hasil ini menunjukkan peningkatan sebesar +0.09% dibandingkan model berbasis teks dan +25.10% dibandingkan model berbasis gambar, yang menyoroti keefektifan penggabungan tingkat keputusan dalam meningkatkan kinerja klasifikasi sentimen.

Tabel 9. Evaluasi unimodal dan multimodal

Feature	**Accuracy (%)**
Text-based	87,84
Image-based	70,28
Multimodal (*Text+Image*)	**87,92**

5. Diskusi

Untuk menemukan model terbaik, sejumlah skenario pengujian dilakukan dalam investigasi ini. Akurasi model berbasis teks, berbasis gambar, dan multimodal di masing-masing dari lima skenario yang diteliti ditunjukkan pada Gambar 6. Skenario pertama berfungsi sebagai *baseline*, dengan akurasi awal masingmasing model sebesar 86,21% untuk CNN dan 87,79% untuk BiLSTM. Pada skenario kedua, terdapat sedikit peningkatan yang signifikan untuk model CNN yang meningkat menjadi 86,57% dan BiLSTM tetap pada 87,79%. Tingkat akurasi model CNN kemudian meningkat menjadi 0,23% pada kasus ketiga, sedangkan model BiLSTM tidak. Pada 87,66% untuk CNN-BiLSTM dan 87,84% untuk BiLSTM-CNN, model hibrida CNN-BiLSTM dan BiLSTM-CNN mengungguli model tunggal pada kasus keempat. Pada skenario kelima, kinerja model teks dan gambar diuji secara terpisah untuk membandingkan keampuhan model multimodal. Hasilnya, tingkat akurasi untuk gambar adalah 70,28%, sedangkan tingkat akurasi untuk multimodal adalah 87,92%.



Gambar 6. Performa Akurasi Varian Model pada seluruh skenario

6. Kesimpulan

Pada penelitian ini dikembangkan analisis sentimen multimodal dengan model hybrid CNN-BiLSTM dan ekspansi fitur FastText. Penelitian ini menggunakan 24.742 pasangan tweet-gambar berbahasa Inggris dari media sosial X setelah preprocessing dan hanya menggunakan label positif dan negatif. Untuk membangun korpus kemiripan FastText, 63.512 entri dari portal berita digital CNN (*Cable News Network*) dan kumpulan data X digabungkan. Semua skenario dievaluasi menggunakan model BiLSTM-CNN yang diintegrasikan dengan mekanisme perhatian dan dinilai menggunakan rasio pembagian 80:20. Konfigurasi ekstraksi fitur teks yang optimal adalah konfigurasi yang dicapai dengan menggabungkan TF-IDF menggunakan Allgram dengan maksimum 10.000 fitur, perluasan menggunakan 15 kata yang mirip dari korpus FastText. Untuk ekstraksi fitur visual menggunakan VGG-16. Hasil terbaik secara keseluruhan adalah hasil yang dicapai dengan menggunakan pendekatan multimodal, yang menggabungkan teks dan gambar, dengan akurasi 87,92%. Hasil ini menunjukkan peningkatan sebesar 0,09% dibandingkan dengan model berbasis *text unimodal* dan 25,10% dibandingkan dengan model berbasis *image unimodal*. Kekuatan dari penelitian ini terletak pada pendekatan komprehensif untuk ekstraksi fitur, memanfaatkan keunggulan yang saling melengkapi dari setiap modalitas dan tingkat fusi untuk memahami dan mengklasifikasikan sentimen dalam wacana kesehatan mental secara lebih efektif.

Daftar Pustaka

[1] A. Abdurrahim and D. H. F. Fudholi, "Mental Health Prediction Model on Social Media Data Using CNN-BiLSTM," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 4, no. 1, 2024, doi: 10.22219/kinetik.v9i1.1849.

[2] Y. Cao, J. Dai, Z. Wang, Y. Zhang, X. Shen, and Y. Liu, "Systematic Review : Text Processing Algorithms in Machine Learning and Deep Learning for Mental Health Detection on Social Media," pp. 1–48.

[3] E. Lee, H. Kim, Y. Esener, and T. McCall, "Online Social Connections of Black American College Students Pre- and Peri-COVID-19 Pandemic: Network Science Approach (Preprint)," *J Med Internet Res*, vol. 26, 2023, doi: 10.2196/55531.

[4] S. A. Sazan, M. H. Miraz, and A. B. M. Muntasir Rahman, "Enhancing Depressive Post Detection in Bangla: A Comparative Study of TF-IDF, BERT and FastText Embeddings," *Annals of Emerging Technologies in Computing*, vol. 8, no. 3, pp. 34–49, 2024, doi: 10.33166/AETiC.2024.03.003.

[5] Ö. Ezerceli and R. Dehkharghani, *Mental disorder and suicidal ideation detection from social media using deep neural networks*, vol. 7, no. 3. Springer Nature Singapore, 2024. doi: 10.1007/s42001-024-00307-1.

[6] H. Elfaik and E. H. Nfaoui, "Leveraging feature-level fusion representations and attentional bidirectional RNN-CNN deep models for Arabic affect analysis on Twitter," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 1, pp. 462–482, 2023, doi: 10.1016/j.jksuci.2022.12.015.

[7] T. Ghosh, M. H. Al Banna, M. J. Al Nahian, M. N. Uddin, M. S. Kaiser, and M. Mahmud, "An attention-based hybrid architecture with explainability for depressive social media text detection in Bangla," *Expert Syst Appl*, vol. 213, no. PC, p. 119007, 2023, doi: 10.1016/j.eswa.2022.119007.

[8] N. H. Di Cara, V. Maggio, O. S. P. Davis, and C. M. A. Haworth, "Methodologies for Monitoring Mental Health on Twitter: Systematic Review," 2023, JMIR Publications Inc. doi: 10.2196/42734.

[9] S. W. Jannah, "Analisis Sentimen Multimodal Berbasis Aspek untuk Ulasan Wisatawan Menggunakan Metode Deep Learning," Skripsi Institut Teknologi 2025.[Online].Available:https://repository.its.ac.id/117461/

[10] Febiana Anistya and Erwin Budi Setiawan, "Hate Speech Detection on Twitter in Indonesia with Feature Expansion Using GloVe," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 5, no. 6, pp. 1044 1051, Dec. 2021, doi: 10.29207/resti.v5i6.3521.

[11] A. Zeyer, P. Doetsch, P. Voigtlaender, R. Schluter, and H. Ney, "A comprehensive study of deep bidirectional LSTM RNNS for acoustic modeling in speech recognition," ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp. 2462–2466, 2017, doi: 10.1109/ICASSP.2017.7952599.

[12] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," Trans Assoc Comput Linguist, vol. 5, pp. 135–146, 2017, doi: 10.1162/tacl_a_00051.

[13] Yue, W., Li, L. (2020). Sentiment analysis using Word2vec-CNN-BiLSTM classification. In 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), Paris, France, pp. 1-5. https://doi.org/10.1109/SNAMS52053.2020.9336549

[14] Rhanoui, M., Mikram, M., Yousfi, S., Barzali, S. (2019). A CNN-BiLSTM model for document-level sentiment analysis. Machine Learning and Knowledge Extraction, 1(3): 832-847. https://doi.org/10.3390/make1030048

[15] Toktarova, A., Syrlybay, D., Myrzakhmetova, B., Anuarbekova, G., Rakhimbayeva, G., Zhylanbaeva, B., Suieuova, N., Kerimbekov, M. (2023). Hate speech detection in social networks using machine learning and deep learning methods. International Journal of Advanced Computer Science and Applications, 14(5): 396-406. https://doi.org/10.14569/IJACSA.2023.0140542

[16] M. A. S. Nasution and E. B. Setiawan, "Enhancing Cyberbullying Detection on Indonesian Twitter: Leveraging FastText for Feature Expansion and Hybrid Approach Applying CNN and BiLSTM," *Revue d'Intelligence Artificielle*, vol. 37, no. 4, pp. 929–936, 2023, doi: 10.18280/ria.370413.