Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

Hybrid CNN-BiLSTM untuk Analisis Sentimen Multi-Platform terhadap Insiden Keamanan Pangan Program Makan Bergizi Gratis

Mohamad Rival Farid Riwaldi*, Aripin

[1] Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
[2] Fakultas Teknik, Program Studi Teknik Biomedis, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1,*]111202214547@mhs.dinus.ac.id, [2]arifin@dsn.dinus.ac.id
Email Penulis Korespondensi: 111202214547@mhs.dinus.ac.id
Submitted: **09/05/2026**; Accepted: **02/06/2026**; Published**: 05/06/2026**

Abstrak−Penurunan prevalensi stunting di Indonesia belum diiringi peningkatan kualitas pelaksanaan program intervensi gizi, salah satunya Program Makan Bergizi Gratis (MBG) yang memicu polemik setelah munculnya insiden keamanan pangan di berbagai daerah. Tingginya arus opini publik di media sosial yang bersifat lintas platform menuntut pendekatan analisis yang mampu menangkap keragaman gaya bahasa dari berbagai sumber secara simultan. Penelitian ini mengusulkan model klasifikasi hybrid Convolutional Neural Network-Bidirectional Long Short-Term Memory (CNN-BiLSTM) untuk menganalisis sentimen komentar publik terkait insiden tersebut, dengan CNN berperan mengekstraksi pola fitur lokal dan BiLSTM memodelkan dependensi urutan kata dari dua arah secara bersamaan. Sebanyak 3.416 komentar dikumpulkan dari lima platform media sosial (X, Instagram, TikTok, YouTube, dan Facebook), kemudian melalui tahapan preprocessing teks dan pelabelan awal berbasis leksikon ke dalam tiga kelas sentimen negatif, netral, dan positif. Untuk memperkuat validitas label, kualitas pelabelan divalidasi melalui anotasi manual oleh dua anotator independen dengan nilai Cohen's Kappa sebesar $\kappa = 0{,}828$. Dataset dibagi 80:20 secara stratified, dengan penerapan class weight untuk mengurangi bias akibat ketidakseimbangan kelas tanpa mengubah jumlah sampel pada setiap kelas. Kinerja model hybrid dibandingkan dengan dua baseline, yaitu CNN dan BiLSTM, menggunakan macro F1-score sebagai metrik utama, sedangkan akurasi digunakan sebagai metrik pendukung. Hasil eksperimen menunjukkan model hybrid CNN-BiLSTM mencapai macro F1-score 90,38% dan akurasi 94,59%, melampaui kedua baseline. Analisis kesalahan klasifikasi mengungkap bahwa kesalahan terbanyak terjadi pada komentar argumentatif, negasi, dan kalimat kontrastif yang mencerminkan keterbatasan pelabelan berbasis leksikon dalam menangkap nuansa bahasa. Secara keseluruhan, pendekatan ini menunjukkan potensi penerapan analisis sentimen berbasis deep learning lintas platform sebagai komponen awal pemantauan opini publik terhadap kebijakan pemerintah berskala nasional. Penelitian ini berkontribusi dalam menyediakan dataset multi-platform berbahasa Indonesia yang tervalidasi secara manual, mengembangkan arsitektur hybrid CNN-BiLSTM dengan skema class weight yang efektif untuk klasifikasi sentimen tiga kelas pada teks informal, serta membuka peluang penerapan deep learning sebagai sarana pemantauan opini publik berbasis data.

Kata Kunci: Analisis Sentimen; Komentar Media Sosial; Hybrid CNN-BiLSTM; Pelabelan Berbasis Leksikon; Program Makan Bergizi Gratis

Abstract−The decline in stunting prevalence in Indonesia has not been accompanied by improvements in the quality of nutritional intervention program implementation, including the Free Nutritious Meal Program (MBG), which sparked public controversy following food safety incidents in several regions. The high volume of cross-platform public opinion on social media requires an analytical approach capable of simultaneously capturing diverse linguistic styles from various sources. This study proposes a hybrid Convolutional Neural Network-Bidirectional Long Short-Term Memory (CNN-BiLSTM) classification model to analyze public sentiment regarding the incidents, with CNN extracting local feature patterns and BiLSTM modeling bidirectional word-sequence dependencies. A total of 3,416 comments were collected from five social media platforms (X, Instagram, TikTok, YouTube, and Facebook), then processed through text preprocessing and initial lexicon-based labeling into three sentiment classes: negative, neutral, and positive. To strengthen label validity, the labeling quality was validated through manual annotation by two independent annotators, yielding a Cohen's Kappa value of $\kappa = 0.828$. The dataset was split using an 80:20 stratified scheme, with class weight applied to reduce bias caused by class imbalance without changing the number of samples in each class. The hybrid model was compared with two baseline models, CNN and BiLSTM, using macro F1-score as the primary metric, while accuracy was used as a supporting metric. The experimental results show that the hybrid CNN–BiLSTM model achieved a macro F1-score of 90.38% and an accuracy of 94.59%, outperforming both baseline models. Misclassification analysis revealed that most errors occurred in argumentative comments, negation, and contrastive sentences, reflecting the limitations of lexicon-based labeling in capturing nuanced language. Overall, this approach demonstrates the potential of cross-platform deep learning-based sentiment analysis as an initial component for monitoring public opinion on national-scale government policies. This study contributes by providing a manually validated multi-platform Indonesian dataset, developing a hybrid CNN-BiLSTM architecture with a class weight scheme effective for three-class sentiment classification on informal text, and opening opportunities for applying deep learning as a means of data-driven public opinion monitoring.

Keywords: Sentiment Analysis; Social Media Comments; Hybrid CNN-BiLSTM; Lexicon-Based Labeling; Free Nutritious Meal Program

1. PENDAHULUAN

Permasalahan gizi kronis masih terjadi di Indonesia, salah satunya adalah stunting yaitu, gangguan pertumbuhan pada anak. Meskipun prevalensi stunting nasional menurun dari 21,6% pada tahun 2022 menjadi 19,8% pada tahun 2024 menurut Survei Status Gizi Indonesia (SSGI), capaian tersebut masih berada di atas ambang batas yang ditargetkan pemerintah yaitu 14%[1][2]. Penurunan ini menunjukkan adanya kemajuan, namun tantangan utama kini bergeser pada kualitas pelaksanaan program intervensi gizi di lapangan[1]. Salah satu bentuk upaya penanganannya adalah

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

Program Makan Bergizi Gratis (MBG) untuk mencukupi gizi pelajar dan ibu hamil. Program ini menjadi salah satu agenda skala prioritas nasional yang diusung oleh Presiden dan Wakil Presiden tahun 2025. Namun implementasi program ini memicu perdebatan publik di berbagai media sosial setelah banyaknya laporan dugaan keracunan Makan Bergizi Gratis di beberapa daerah. Insiden keracunan makanan ini membuat masyarakat khawatir terhadap keamanan makanan dan efektivitas sistem pengawasan distribusi makanan. Kasus ini menunjukkan bahwa pentingnya menerapkan standar kebersihan dan pengawasan yang ketat agar tujuan program penurunan stunting tercapai.

Permasalahan utama yang muncul bukan hanya pada aspek teknis distribusi dan keamanan pangan, tetapi juga pada dampak sosial berupa munculnya ketidakpercayaan publik terhadap efektivitas kebijakan pemerintah. Tingginya arus informasi dan opini yang beredar di media sosial sering kali menimbulkan misinformasi serta memperkuat persepsi negatif terhadap program tersebut. Kurangnya respons komunikasi publik yang cepat dan berbasis data menyebabkan sentimen negatif lebih dominan, sehingga pemerintah sulit memantau persepsi masyarakat secara *real-time*. Dalam konteks ini, opini publik yang terefleksi pada platform media sosial seperti X, Instagram, TikTok, YouTube, dan Facebook menjadi sumber data penting untuk memetakan isu yang perlu respons kebijakan cepat. Analisis sentimen sebagai proses mengekstraksi dan mengklasifikasikan polaritas opini dari teks [3] menjadi pendekatan yang relevan, namun sejumlah keterbatasan teknis mendasar belum tertangani secara memadai. Sebagian besar penelitian menggunakan data dari satu platform tunggal, padahal setiap platform menunjukkan karakteristik linguistik yang berbeda sesuai konteks penggunaannya. Di sisi lain, pendekatan berbasis *machine learning* konvensional terbukti terbatas dalam menangkap dependensi sekuensial jarak jauh pada teks informal, sementara model *deep learning* tunggal seperti CNN atau LSTM masing-masing hanya unggul pada satu aspek dan tidak keduanya secara bersamaan. Keterbatasan ini menjadi kritis pada teks media sosial yang kaya sarkasme, negasi, dan perubahan sentimen di tengah kalimat. Integrasi data multi-platform dengan arsitektur hibrida CNN-BiLSTM pada konteks kebijakan publik berbahasa Indonesia merupakan pendekatan yang belum banyak dieksplorasi, dan kesenjangan inilah yang menjadi justifikasi metodologis penelitian ini.

Penerapan analisis sentimen pada platform X dapat mengungkap distribusi sentimen positif negatif dan netral beserta topik dominan sehingga pengambil kebijakan memperoleh masukan real time untuk penyesuaian strategi komunikasi dan implementasi [4], namun cakupan datanya terbatas pada platform X sehingga opini publik dari platform lain tidak terwakili. Sebuah studi berorientasi kesehatan masyarakat menganalisis postingan yang menyebutkan akun resmi Kementerian Kesehatan di platform X menggunakan pembobotan TF-IDF, dan hasilnya menunjukkan bahwa algoritma SVM mencapai akurasi 79%, sedikit di atas Naive Bayes sebesar 77% [5]. Temuan ini menunjukkan bahwa model berbasis margin seperti SVM dapat lebih baik menangkap pola bahasa publik yang kompleks. Sejalan dengan ini, sebuah studi tahun 2024 mengintegrasikan CNN dengan optimasi swarm partikel untuk analisis sentimen politik, mencapai akurasi 78,2%, meningkat dari baseline dan menyoroti nilai optimasi hyperparameter untuk klasifikasi teks [6]. Sementara itu, sebuah studi yang dilakukan selama pandemi COVID-19 melaporkan bahwa Long Short-Term Memory dengan word embedding mencapai akurasi 81%, mengungguli Naive Bayes 74% dan RNN 71%, menunjukkan kekuatan LSTM dalam menangkap konteks urutan kata dan ketergantungan sekuensial dalam data bahasa alami.[7]

Hasil gabungan dari studi-studi ini menunjukkan pola yang konsisten bahwa model margin seperti SVM unggul dalam fitur sparse dan berbasis vektor, sementara arsitektur berlapis seperti CNN dan LSTM lebih efektif dalam memahami konteks sekuensial. Dalam konteks kebijakan nasional, sebuah studi melaporkan bahwa model SVM yang dikombinasikan dengan teknik SMOTE mencapai akurasi 85,74%, mengungguli Random Forest sebesar 81,53%, menunjukkan bahwa penyeimbangan kelas dapat secara substansial meningkatkan stabilitas model [4]. Temuan ini didukung oleh studi internasional yang mengkonfirmasi bahwa penerapan SMOTE dan variannya dapat meningkatkan kinerja klasifikasi teks, misalnya, peningkatan akurasi SVM dari 71,41% menjadi 83,89% dalam studi tahun 2025 dan evaluasi komprehensif berbagai varian SMOTE dalam Scientific Reports [8], [9].

Selain itu, penelitian global secara konsisten menunjukkan keunggulan arsitektur *deep learning* hibrida dibandingkan model tunggal untuk analisis sentimen. Studi yang mengusulkan ConvBiLSTM dengan menggabungkan CNN satu dimensi dan BiLSTM untuk klasifikasi sentimen Twitter berbahasa Inggris mencapai akurasi 91,13% [10], membuktikan bahwa kombinasi ekstraksi fitur lokal dan pemodelan sekuensial dua arah menghasilkan representasi yang lebih kaya dibandingkan model tunggal, meskipun pengujiannya terbatas pada satu platform dengan klasifikasi biner dua kelas. Studi yang menerapkan CNN dikombinasikan dengan BiGRU pada bahasa Arab mencapai akurasi 90,06% pada klasifikasi biner, namun performanya turun signifikan menjadi 73,17% pada klasifikasi empat kelas [11], mengindikasikan bahwa arsitektur hybrid masih menghadapi tantangan pada skenario multi-kelas. Studi lain yang menggunakan CNN-LSTM pada tweet COVID-19 mencapai akurasi 84,72% [12], namun penggunaan LSTM satu arah membatasi kemampuan model dalam menangkap konteks kata dari dua arah secara bersamaan. Pendekatan BERT-BiLSTM-CNN pada tweet bahasa Turki mencapai akurasi 91,01% [13], namun ketergantungan pada *pre-trained language model* yang spesifik untuk satu bahasa menjadikannya kurang fleksibel untuk diterapkan pada bahasa lain termasuk bahasa Indonesia informal. Studi lain yang meneliti arsitektur hibrida berbasis *attention* untuk klasifikasi sentimen [14], juga menunjukkan bahwa kombinasi mekanisme *attention* dengan ekstraksi fitur lokal dan sekuens dapat meningkatkan stabilitas klasifikasi teks pada data sosial yang sangat beragam secara linguistik. Secara keseluruhan, studi-studi ini menunjukkan potensi arsitektur hybrid, namun sebagian besar diuji pada satu platform, klasifikasi biner, atau bahasa selain Indonesia, sehingga penerapannya pada komentar

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

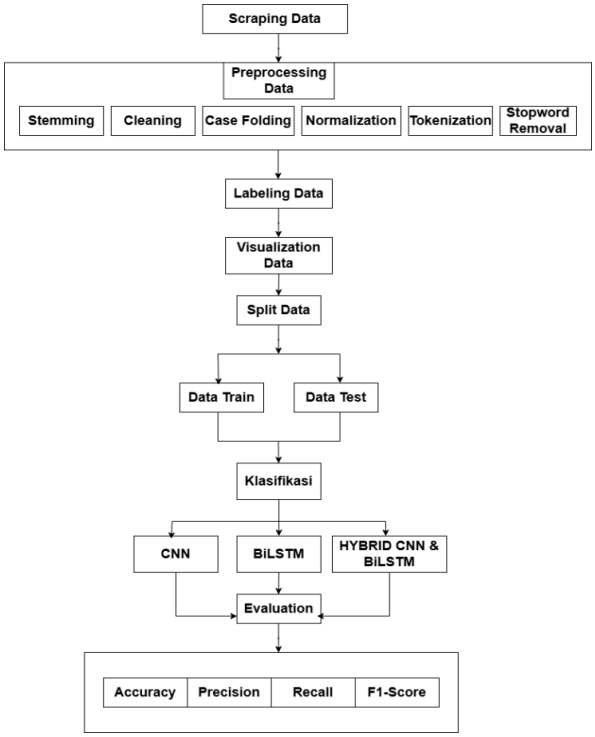
informal lintas platform berbahasa Indonesia dalam konteks kebijakan publik masih memerlukan eksplorasi lebih lanjut.

Kebaruan penelitian ini terletak pada dua aspek utama, yaitu sumber data yang bersifat lintas platform dan rancangan model hibrida CNN-BiLSTM. Data tidak hanya berasal dari satu media sosial, tetapi dikumpulkan dari berbagai platform populer seperti X, Instagram, TikTok, YouTube, dan Facebook yang masing-masing memiliki karakteristik linguistik berbeda yaitu X cenderung menghasilkan teks pendek dan padat, Facebook mengandung kalimat yang lebih panjang dan naratif, sementara TikTok dan Instagram didominasi bahasa informal dengan penggunaan hashtag dan singkatan yang tinggi. Seluruh data tersebut diproses melalui tahap normalisasi dan ketidakseimbangan distribusi sentimen ditangani menggunakan skema class weight, serta dilengkapi proses anotasi manual oleh dua anotator independen terhadap 300 sampel data yang menghasilkan nilai Cohen's Kappa sebesar κ = 0,828. Dari sisi metodologi, penelitian ini menggabungkan kekuatan CNN dalam mengekstraksi pola lokal pada sekuens teks dengan kemampuan BiLSTM dalam memahami dependensi urutan kata secara dua arah, sehingga menghasilkan model hibrida yang lebih adaptif terhadap karakteristik bahasa informal di media sosial berbahasa Indonesia. Pendekatan ini diharapkan mampu memberikan hasil analisis yang lebih akurat dan komprehensif dalam memetakan opini publik terkait Program Makan Bergizi Gratis, serta dapat dimanfaatkan sebagai dasar awal pemantauan sentimen dan penyusunan kebijakan yang lebih responsif. Secara keseluruhan, penelitian ini berkontribusi pada tiga aspek. Pertama, tersedianya dataset multi-platform berbahasa Indonesia yang dikumpulkan dari lima platform media sosial dan divalidasi melalui anotasi manual. Kedua, dikembangkannya arsitektur hybrid CNN-BiLSTM yang dilengkapi skema *class weight* untuk menangani ketidakseimbangan kelas pada klasifikasi sentimen tiga kelas. Ketiga, terbukanya peluang pemanfaatan pendekatan *deep learning* sebagai sarana awal pemantauan opini publik berbasis data dalam konteks kebijakan nasional.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menerapkan tahapan metodologi yang terstruktur dalam melakukan analisis sentimen terkait keracunan program MBG yang mencakup scraping data, preprocessing data, labeling data, visualisasi data, split data, klasifikasi model, dan evaluasi model. Visualisasi alur tahapan tersebut terdapat pada Gambar 1.



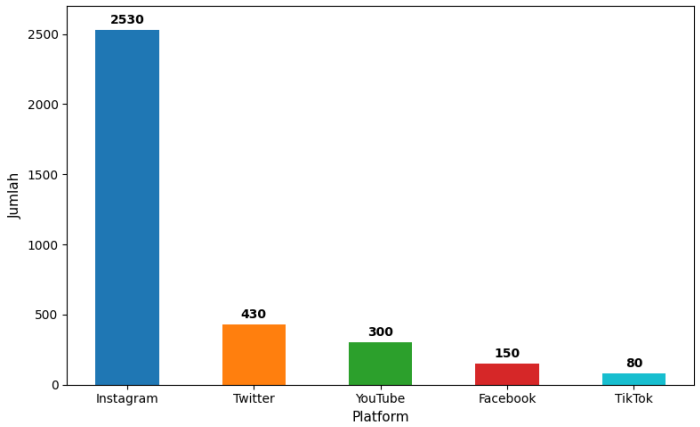
Gambar 1. Flowchart tahapan penelitian

Gambar 1 menjelaskan langkah pertama adalah mengumpulkan data komentar publik, selanjutnya data dilakukan preprocessing untuk memvalidasi kualitas data sebelum dimasukkan ke dalam pemodelan pembelajaran mesin. Setelah melalui tahap preprocessing, data diberi label sesuai kategori sentiment. Selanjutnya dilakukan visualisasi data untuk menampilkan word cloud per kelas sentimen, diagram batang top 10 frekuensi kata, serta grafik n-gram (n = 3) untuk memetakan kata dan frasa mana yang mendominasi dalam korpus. Pembagian data dibagi menjadi dua bagian yaitu data uji dan data latih, dilanjutkan dengan klasifikasi model dan evaluasi.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

2.2 Scraping Data

Pengumpulan data pada penelitian ini diperoleh melalui lima platform media sosial meliputi X, Instagram, Tiktok, Youtube, dan Facebook. Proses pengumpulan data dilakukan menggunakan pendekatan scraping data, yaitu proses pengambilan data secara otomatis untuk mengumpulkan data yang relevan[15]. Metode disesuaikan pada setiap platform untuk memperoleh komentar yang relevan dengan isu keracunan MBG. Pada platform X, data dikumpulkan menggunakan tweet-harvest berbasis python dengan kata kunci utama yaitu "keracunan MBG" untuk menjaring opini publik yang berhubungan langsung dengan isu tersebut. Pada Instagram, Tiktok, Youtube, dan Facebook data diperoleh melalui apify, komentar yang diambil dari unggahan atau video yang memiliki tingkat respon publik tinggi. Setelah proses penggabungan komentar serta penghapusan komentar yang bersifat duplikasi, diperoleh 3.416 komentar yang digunakan sebagai dataset utama. Komentar yang dikumpulkan merupakan komentar yang secara eksplisit maupun implisit menyinggung kasus keracunan MBG, baik dalam bentuk opini, keluhan, dukungan, maupun diskusi terhadap kebijakan pemerintah. Distribusi jumlah komentar berdasarkan platform ditunjukan pada Gambar 2.



Gambar 2. Distribusi komentar berdasarkan platform media sosial

Gambar 2 menunjukkan bahwa komentar terbanyak dari platform Instagram dibandingkan media sosial lainnya. Dominasi ini tidak terlepas dari karakteristik sumber data, di mana salah satu unggahan yang dijadikan objek scraping merupakan postingan dengan tingkat interaksi sangat tinggi yaitu sekitar 2500 komentar. Kondisi tersebut membuat jumlah komentar yang terkumpul dari Instagram jauh lebih besar dibandingkan X, Youtube, Facebook dan Tiktok.

2.3 Preprocessing Data

Tahap *preprocessing* bertujuan menghasilkan teks yang bersih dan terstandarisasi sehingga siap digunakan pada proses pelabelan dan pemodelan. Tahapan yang diterapkan meliputi *cleaning* untuk menghapus karakter tidak relevan seperti tautan, *mention*, *hashtag*, angka, dan emotikon, kemudian *case folding* untuk mengubah teks menjadi huruf kecil, dilanjutkan normalisasi untuk memperbaiki kata tidak baku, tokenisasi untuk memecah teks menjadi deretan kata, *stopword removal* untuk menghapus kata umum yang kurang informatif, dan *stemming* untuk mengembalikan kata ke bentuk dasarnya. Penerapan mekanisme *preprocessing* yang tepat sebelum pelatihan dapat meningkatkan akurasi model secara signifikan [16]. Hasil akhir preprocessing disajikan pada Tabel 1 yang memperlihatkan teks menjadi lebih ringkas, terstruktur, dan siap digunakan dalam proses pelabelan.

Tabel 1. Hasil Pre-Processing Teks Komentar

No	Tahapan	Teks
1	Data awal	SELAMAT DAN SUKSES ATAS PROGRAM MBG BAPAK . HEBAT..... LUAR BIASA.....
2	Cleaning	SELAMAT DAN SUKSES ATAS PROGRAM MBG BAPAK HEBAT LUAR BIASA
3	Case Folding	selamat dan sukses atas program mbg bapak hebat luar biasa
4	Normalisasi	selamat dan sukses atas program mbg bapak hebat luar biasa
5	Tokenisasi	['selamat', 'dan', 'sukses', 'atas', 'program', 'mbg', 'bapak', 'hebat', 'luar', 'biasa']
6	Stopword removal	selamat sukses program mbg hebat
7	Stemming	selamat sukses program mbg hebat

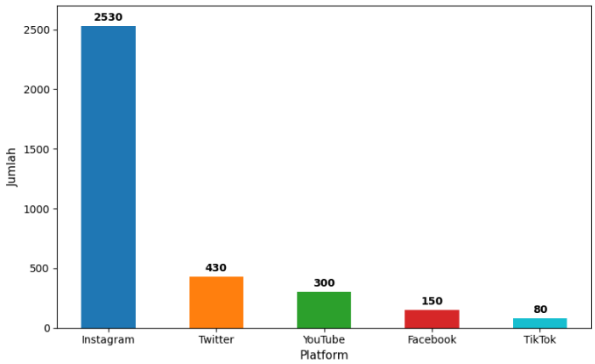
Berdasarkan Tabel 1, setiap tahapan memberikan kontribusi dalam menyederhanakan teks secara bertahap. *Cleaning* menghilangkan karakter dan spasi berlebih, *case folding* menyeragamkan huruf, normalisasi memperbaiki bentuk kata, tokenisasi memecah kalimat menjadi token, *stopword removal* menghapus kata umum seperti "dan"

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

dan "atas", serta *stemming* mengembalikan kata ke bentuk dasarnya. Hasil akhir menunjukkan bahwa teks yang semula panjang dan tidak terstruktur berhasil diringkas menjadi representasi yang lebih bersih dan siap digunakan dalam proses pelabelan dan pemodelan.

2.4 Labelling Data

Pelabelan data merupakan proses menentukan label sentimen terhadap data agar dapat digunakan dalam pelatihan dan pengujian model [17]. Pada penelitian ini, pelabelan dilakukan secara otomatis menggunakan pendekatan *lexicon-based*, yaitu metode yang menilai polaritas sentimen berdasarkan kamus kata berbobot[18]. Setiap komentar diberi skor numerik pada rentang -5 hingga 5, di mana skor -5 sampai -1 menyatakan sentimen Negatif, skor 0 menyatakan sentimen Netral, dan skor 1 sampai 5 menyatakan sentimen Positif. Skor tersebut kemudian dipetakan ke dalam tiga kelas label yang digunakan pada dataset pelatihan dan pengujian. Setelah seluruh komentar melalui proses pelabelan, diperoleh total 3.416 komentar dengan komposisi 2.240 komentar Negatif (65,6%), 967 komentar Netral (28,3%), dan 209 komentar Positif (6,1%), sebagaimana ditunjukkan pada Gambar 3.



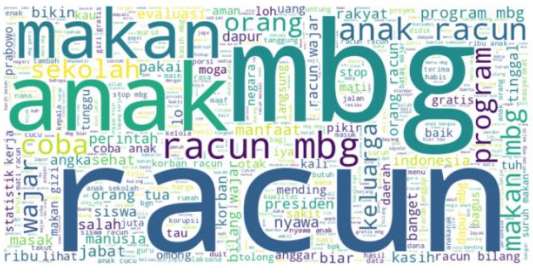
Gambar 3. Distribusi Komentar Berdasarkan Platform Media Sosial

Untuk memvalidasi kualitas pelabelan otomatis tersebut, dilakukan proses anotasi manual oleh dua anotator independen terhadap 300 sampel data yang secara purposif untuk mencakup ketiga kelas sentimen. Kedua anotator menentukan label sentimen secara terpisah tanpa saling berkomunikasi, kemudian tingkat kesepakatan antar keduanya diukur menggunakan Cohen's Kappa. Sebanyak 29 data yang tidak mencapai kesepakatan antar anotator tetap menggunakan label leksikon sebagai penentu akhir. Hasil pengukuran menunjukkan nilai $\kappa = 0{,}828$, mengindikasikan bahwa pelabelan berbasis leksikon memiliki konsistensi yang tinggi dengan penilaian manusia dan layak digunakan sebagai ground truth dalam pelatihan model.

Dominasi kelas negatif menunjukkan adanya ketidakseimbangan kelas yang cukup signifikan antara kelas mayoritas dan kelas minoritas. Kondisi ini berpotensi menyebabkan model lebih cenderung memprediksi kelas negatif dibandingkan kelas netral maupun positif. Untuk mengurangi bias tersebut, penelitian ini menerapkan skema *class weight* pada tahap pelatihan. Perlu ditegaskan bahwa *class weight* tidak mengubah jumlah sampel pada setiap kelas, melainkan memberikan bobot penalti yang lebih besar terhadap kesalahan prediksi pada kelas dengan jumlah sampel lebih sedikit. Dengan demikian, distribusi data aktual tetap sama seperti dataset awal, yaitu 2.240 data negatif, 967 data netral, dan 209 data positif. Namun, kontribusi kesalahan dari kelas netral dan positif menjadi lebih diperhitungkan dalam proses optimasi model.

2.5 Visualisasi Data

Visualisasi data digunakan untuk menyajikan data hasil preprocessing ke dalam bentuk grafis untuk mempermudah interpretasi dan pemahaman hasil analisis[19]. Penelitian ini menggunakan tiga teknik visualisasi, yaitu *wordcloud*, diagram frekuensi kata, dan *n-gram*. *Wordcloud* menampilkan distribusi frekuensi kata secara visual, di mana kata yang lebih sering muncul ditampilkan dengan ukuran lebih besar sehingga tema utama dan kata kunci yang dominan dapat diidentifikasi tanpa harus membaca seluruh komentar secara manual. Hasil *wordcloud* komentar terkait kasus keracunan MBG disajikan pada Gambar 4.



Gambar 4. Wordcloud dari komentar MBG

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

Gambar 4 menunjukkan wordcloud dari keseluruhan korpus komentar setelah melalui tahapan preprocessing. Ukuran kata yang lebih besar menandakan frekuensi kemunculan yang lebih tinggi. Terlihat bahwa kata-kata seperti "racun". "anak", "makan" dan "mbg" muncul paling dominan, yang paling menunjukkan bahwa isu keracunan makanan serta dampak nya terhadap anak menjadi fokus utama percakapan publik. Pola ini konsisten dengan distribusi sentimen yang menunjukkan tingginya proporsi komentar bernada negatif.

2.6 Split Data

Setelah Setelah melalui tahap preprocessing dan pelabelan, dataset dibagi menjadi data latih dan data uji dengan beberapa skenario rasio, yaitu 60:40, 70:30, dan 80:20. Rasio 80:20 dipilih sebagai konfigurasi utama karena memberikan performa terbaik dengan jumlah data uji yang masih representatif, yaitu 2.732 komentar sebagai data latih dan 684 komentar sebagai data uji. Pembagian data dilakukan secara *stratified* sehingga proporsi kelas Negatif, Netral, dan Positif pada data latih dan data uji tetap mengikuti distribusi awal. Mengingat distribusi kelas pada dataset tidak seimbang, tahap pelatihan menerapkan skema *class weight* pada fungsi loss. Skema ini tidak mengubah jumlah sampel pada setiap kelas, tetapi memberikan bobot penalti yang lebih besar terhadap kesalahan prediksi pada kelas dengan jumlah sampel lebih sedikit. Dengan demikian, model diharapkan tidak terlalu bias terhadap kelas mayoritas tanpa perlu melakukan penambahan data sintetis atau pengurangan data mayoritas. Sebagian data latih juga dipisahkan sebagai data validasi untuk memantau dinamika pelatihan, mengevaluasi perubahan nilai loss dan accuracy pada setiap epoch, serta mengidentifikasi adanya indikasi overfitting.

2.7 Klasifikasi

Penelitian ini menggunakan tiga arsitektur untuk mengklasifikasikan sentimen, yaitu CNN, BiLSTM dan hybrid CNN-BiLSTM. Ketiga model ini dipilih karena telah banyak digunakan dan terbukti efektif pada penelitian analisis sentimen berbasis teks, CNN kuat pada mengekstraksi pola lokal, BiLSTM efektif untuk ketergantungan sekuensial, dan kombinasi hybrid CNN-BiLSTM memanfaatkan kedua kelebihan tersebut karena itu arsitektur ini dipilih sebagai model utama[20]. Pada tahap klasifikasi, setiap komentar hasil labeling dikategorikan ke dalam tiga kelas sentimen yaitu negatif, netral, dan positif. Label sentimen yang semula berbentuk teks dikodekan menjadi label numerik menggunakan LabelEncoder dari scikit-learn dan direpresentasikan dalam bentuk one-hot vector selama pelatihan.

Penerapan model hybrid CNN-BiLSTM pada penelitian ini mencakup tahapan pemetaan teks ke bentuk numerik, pembentukan sekuens dengan panjang tetap, perancangan arsitektur model, serta proses pelatihan dan evaluasi model. Pada tahap pemetaan teks kolom stemming data digunakan sebagai masukan model, sedangkan kolom sentimen berperan sebagai label target. Seluruh teks komentar kemudian melalui proses tokenisasi dengan membangun kamus kata dari data dan membatasi kosakata pada kata kata yang sering muncul. Setiap teks diubah menjadi deretan indeks kata dan diseragamkan menjadi sekuens dengan panjang tetap sehingga seluruh sampel memiliki dimensi input yang konsisten untuk di proses oleh model. Untuk memperjelas peran masing-masing arsitektur, pembahasan berikut disusun berurutan CNN, BiLSTM lalu hybrid CNN-BiLSTM.

2.7.1 Model CNN

Convolutional Neural Network (CNN) merupakan model deep learning yang secara otomatis mempelajari hierarki fitur spasial melalui backpropagation dengan menyusun blok convolution, pooling dan fully connected[21]. Pada penelitian ini menerapkan 1D-CNN untuk teks dengan terlebih dahulu mengubah kalimat menjadi urutan embedding. Filter convolution mengekstraksi pola lokal, pooling melakukan down sampling dan fitur akhir dipetakan ke probabilitas kelas melalui lapisan fully connected dan softmax[22]. Persamaan convolution yang digunakan adalah:

$$y[i] = \sum_{\{m=0\}}^{\{k-1\}} w[m]\, x[i+m] \quad (1)$$

Persamaan (1) menghitung keluaran konvolusi pada posisi i dengan menjumlahkan hasil kali antara potongan urutan input x dan koefisien kernel w sepanjang k elemen. Indeks m merepresentasikan pergeseran jendela kernel di sekitar posisi i setiap suku $w[m]$" "$x[i+m]$ menangkap kecocokan pola lokal pada urutan. Bobot kernel yang sama dipakai di seluruh posisi (*weight sharing*), sehingga deteksi pola menjadi konsisten dan jumlah parameter tetap efisien. Jika padding digunakan, elemen dekat batas tetap terdefinisi sedangkan stride menentukan jarak perpindahan jendela antarperhitungan. Nilai $y[i]$ untuk semua i membentuk peta fitur yang selanjutnya dapat diringkas dengan *pooling* sebelum tahap klasifikasi.

2.7.2 Model BiLSTM

Long Short-Term Memory (LSTM) merupakan varian *recurrent neural network* (RNN) yang dirancang untuk memodelkan hubungan urutan antarkata pada data sekuensial [23]. Dalam arsitektur BiLSTM, sekuens teks dibaca dari dua arah maju dan mundur. Penelitian ini menggunakan arsitektur dua lapis Bidirectional LSTM untuk memodelkan pola urutan kata pada teks media sosial. Mekanisme ini dirancang melalui tiga gerbang utama yaitu input gate, forget gate, dan output gate yang secara dinamis mengontrol penambahan, penyimpanan, dan penghapusan informasi di dalam memori [24]. Pada penelitian ini, BiLSTM diimplementasikan sebagai salah satu arsitektur pembanding baseline terhadap model Hybrid CNN-BilSTM untuk klasifikasi sentimen tiga kelas. Model dibangun dalam bentuk jaringan sekuensial yang terdiri atas lapisan embedding berdimensi 128, dua lapis Bidirectional LSTM

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

dengan 64 unit pada lapisan pertama dan 32 unit pada lapisan kedua, diikuti lapisan dropout dengan rasio 0.5, lapisan dense berukuran 64 unit dengan fungsi aktivasi ReLU, serta lapisan keluaran softmax dengan tiga neuron yang memetakan representasi fitur ke kelas sentimen negatif, netral, dan positif.

Proses pembagian data mengikuti skema pembagian data latih dan data uji sebesar 80% dan 20% secara stratified serta penerapan class weight pada loss, dengan pengaturan 10 epoch, ukuran 32 batch, dan metrik evaluasi. Dengan adanya model BiLSTM sebagai baseline, kinerja model hybrid CNN-BiLSTM pada bagian hasil dapat dibandingkan secara lebih objektif. Karakteristik teks pada media sosial yang cenderung pendek, padat, dan banyak mengandung bahasa gaul, singkatan, serta emotikon menjadikan pemodelan konteks sebagai aspek yang penting dalam klasifikasi sentimen. BiLSTM mampu menangkap konteks dari kata sebelumnya dan sesudahnya secara bersamaan sehingga representasi kalimat yang dihasilkan menjadi lebih informatif dibandingkan LSTM satu arah. Kemampuan ini diharapkan membantu model membedakan nuansa sentimen yang halus, misalnya pada ungkapan sarkasme atau opini yang disampaikan secara tidak langsung, sehingga kapasitas pembeda baseline terhadap tiga kelas sentimen menjadi lebih baik.

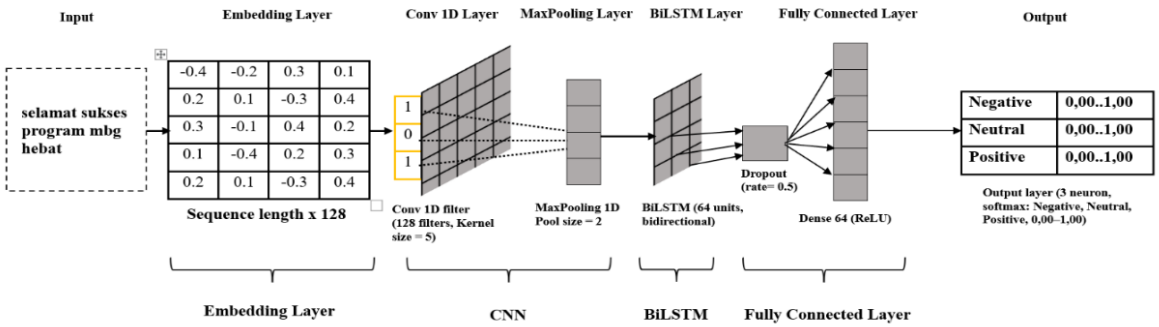
Selain itu, penggunaan BiLSTM sebagai baseline memberikan landasan yang jelas untuk menilai kontribusi tambahan dari konvolusi pada model hybrid. Apabila performa model hybrid menunjukkan peningkatan yang signifikan dibandingkan BiLSTM murni, maka peningkatan tersebut dapat diinterpretasikan sebagai efek positif kombinasi ekstraksi fitur lokal oleh CNN dan pemodelan ketergantungan jangka panjang oleh BiLSTM. Sebaliknya, jika selisih kinerja relatif kecil, hasil tersebut mengindikasikan bahwa informasi sekuensial yang ditangkap oleh BiLSTM sudah cukup representatif untuk menyelesaikan tugas klasifikasi sentimen pada dataset ini. Dengan demikian, analisis perbandingan antara baseline dan model hybrid pada bagian hasil dan pembahasan dapat memberikan gambaran yang lebih komprehensif mengenai kelebihan dan keterbatasan masing-masing arsitektur.

2.7.3 Model Hybrid CNN BiLSTM

Penelitian ini menerapkan arsitektur hybrid CNN-BiLSTM untuk analisis sentimen. CNN berperan sebagai ekstrak fitur lokal sedangkan BiLSTM memodelkan dependensi jarak jauh yang tidak tercakup oleh convolution dan pooling lokal [25]. Dalam implementasi ini, teks ditokenisasi menggunakan tokenizer dengan batas 5.000 kosa kata dan token OOV, kemudian dipetakan embedding berdimensi 128 yang dilatih end-to-end bersama lapisan lain. Ekstraksi dilakukan oleh Conv1D dengan 128 filter dan kernel 5, diikuti MaxPooling1D berukuran 2 untuk mereduksi panjang sekuens secara terkontrol dan menjaga efisiensi. Representasi hasil CNN diproses oleh Bidirectional LSTM 64 unit untuk menangkap konteks dua arah, diikuti dengan Dropout dengan rate 0.5, lapisan Dense 64 dengan aktivasi Relu. Lapisan terakhir berupa softmax tiga kelas. Skema data menggunakan 80% data latih dan 20% data uji. Output dari lapisan BiLSTM hingga lapisan softmax dioptimasi menggunakan fungsi *loss categorical cross-entropy* untuk menangani ketidakseimbangan distribusi kelas. Bobot untuk setiap kelas dihitung berdasarkan proporsi sampel pada data latih sebagaimana didefinisikan pada persamaan (2).

$$w_c = \frac{N}{K * n_c} \tag{2}$$

Dimana N menyatakan jumlah seluruh sampel latih, K jumlah kelas dan n_c jumlah sampel pada kelas ke- c. Distribusi kelas sentimen pada data latih tidak seimbang, di mana kelas negatif jauh lebih dominan dibandingkan kelas netral dan positif. Untuk memitigasi ketidakseimbangan kelas pada model, seluruh model dilatih dengan menerapkan skema class weight pada fungsi loss yang dihitung menggunakan fungsi compute class weight dari library scikit learn berdasarkan frekuensi masing masing kelas. Dengan cara ini, kelas netral dan positif yang jumlah sampelnya lebih sedikit memperoleh bobot kesalahan yang lebih besar, sehingga kontribusi kesalahan pada kelas minoritas menjadi lebih signifikan tanpa mengubah jumlah data pada tiap kelas. Konfigurasi tersebut membentuk alur pemrosesan pada model hybrid CNN-BiLSTM yang digunakan pada penelitian ini . Secara ringkas, alur dimulai dari input teks, embedding, dilanjutkan dengan ekstraksi fitur lokal oleh CNN, pemodelan konteks dua arah oleh BiLSTM, penggunaan Dropout untuk mencegah overfitting, dan klasifikasi tiga kelas pada lapisan softmax, sebagaimana ditunjukan pada Gambar 5.



Gambar 5. Arsitektur Model Hybrid CNN-BiLSTM

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

Arsitektur pada Gambar 5. digunakan sebagai konfigurasi utama dalam seluruh percobaan pada penelitian ini. Pada model hybrid CNN-BiLSTM, setiap komentar yang telah melalui tahap preprocessing terlebih dahulu direpresentasikan sebagai vektor indeks dengan panjang tetap, kemudian dijadikan input vector bagi embedding layer. Lapisan embedding memetakan setiap indeks kata ke vektor berdimensi 128, sehingga satu komentar diubah menjadi embedding matrix berukuran sequence length x 128 yang menyimpan representasi kata dalam bentuk vektor real yang dapat dipelajari selama proses pelatihan. Embedding layer ini selanjutnya diproses oleh Convolutional 1D layer yang menggunakan 128 filter dengan ukuran kernel 5. Setiap filter digeser di sepanjang dimensi urutan kata untuk menangkap pola lokal berbasis 3-gram, misalnya frasa bernada negatif ataupun ekspresi dukungan pada komentar media sosial. Proses convolutional 1D ini menghasilkan serangkaian 1D feature maps yang merepresentasikan keberadaan dan kekuatan pola tersebut di sepanjang sekuens.

Seluruh feature maps yang dihasilkan CNN kemudian dialirkan ke MaxPooling layer dengan ukuran pool 2. Pada lapisan ini, nilai aktivasi pada setiap dua posisi berurutan dalam sekuens digabungkan dengan cara mengambil nilai yang paling besar, sehingga panjang sekuens berkurang setengahnya namun hanya merespon yang paling kuat dari tiap filter yang dipertahankan. Keluaran dari lapisan MaxPooling ini selanjutnya digunakan sebagai masukan bagi lapisan BiLSTM dengan 64 unit. Di dalam lapisan BiLSTM, urutan fitur diproses secara bersamaan dari dua arah yaitu, maju membaca urutan kata dari kiri dan kanan, sedangkan mundur membaca urutan kata dari kanan dan kiri. Dengan cara ini, representasi pada setiap posisi kata tidak hanya melihat konteks sebelumnya, tetapi juga mempertimbangkan kata-kata yang muncul setelahnya, sehingga informasi yang digunakan model menjadi lebih lengkap. Mekanisme dua arah ini penting untuk menangani komentar media sosial yang sering memuat negasi, sarkasme, atau perubahan sentimen di tengah kalimat. Dengan membaca dari kiri ke kanan dan dari kanan ke kiri, BiLSTM dapat memahami hubungan antarkata yang posisinya saling berjauhan dalam satu kalimat, sehingga makna sentimen yang tersembunyi di balik pola bahasa tersebut dapat ditangkap dengan lebih akurat. Output BiLSTM kemudian diregularisasi menggunakan lapisan Dropout dengan lapisan rasio 0.5 yang secara acak menonaktifkan sebagian neuron selama proses pelatihan agar model tidak terlalu hafal data latih dan bisa lebih general ke data uji. Pada penelitian ini, dropout diterapkan setelah lapisan BiLSTM sehingga setiap iterasi pelatihan hanya sebagian unit yang aktif. Dengan cara tersebut, model terdorong untuk menyebarkan informasi ke lebih banyak neuron dan tidak bergantung pada pola spesifik dari data latih yang mungkin bersifat noise atau hanya muncul sebagian kecil komentar. Representasi yang sudah melalui proses dropout ini kemudian diteruskan ke lapisan fully connected layer dengan 64 neuron dan fungsi aktivasi ReLU untuk memetakan fitur hasil ekstraksi CNN-BiLSTM ke ruang yang lebih ringkas dan diskriminatif. Tahap terakhir adalah output layer dengan tiga neuron dan fungsi aktivasi softmax yang menghasilkan probabilitas untuk setiap kelas sentimen, yaitu negatif, netral, dan positif, dengan rentang nilai 0–1. Kelas sentimen akhir untuk suatu komentar ditentukan berdasarkan neuron dengan nilai probabilitas tertinggi pada lapisan softmax tersebut.

2.8 Evaluasi

Evaluasi model bertujuan untuk menilai kemampuan model dalam mengklasifikasikan data uji yang tidak digunakan pada proses pelatihan. Pada penelitian ini, kinerja model dievaluasi menggunakan accuracy, precision, recall, dan F1-score. Mengingat distribusi kelas pada dataset tidak seimbang, evaluasi tidak hanya berfokus pada accuracy, tetapi juga menempatkan macro F1-score sebagai metrik utama agar performa setiap kelas, termasuk kelas minoritas, tetap diperhitungkan secara adil.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precison = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1-Score = 2x\frac{Presisi\ x\ Recall}{Presisi+Recall} \quad (6)$$

True True Positive (TP) menunjukkan jumlah data pada suatu kelas yang berhasil diprediksi dengan benar sebagai kelas tersebut. False Positive (FP) menunjukkan jumlah data dari kelas lain yang keliru diprediksi sebagai kelas tersebut. False Negative (FN) menunjukkan jumlah data pada suatu kelas yang keliru diprediksi sebagai kelas lain. Sementara itu, True Negative (TN) menunjukkan jumlah data dari kelas lain yang tidak diprediksi sebagai kelas tersebut. Pada klasifikasi multi-kelas, perhitungan TP, FP, FN, dan TN dilakukan untuk masing-masing kelas dengan pendekatan one-vs-rest. Accuracy mengukur proporsi prediksi benar dari seluruh data uji. Namun, pada dataset yang tidak seimbang seperti penelitian ini, accuracy yang tinggi dapat menyesatkan karena model berpotensi lebih banyak benar pada kelas mayoritas. Oleh karena itu, precision digunakan untuk mengukur ketepatan prediksi model pada suatu kelas, sedangkan recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang benar-benar termasuk kelas tersebut. F1-score digunakan sebagai rata-rata harmonis antara precision dan recall sehingga mampu memberikan gambaran performa yang lebih seimbang. Pada penelitian ini digunakan macro F1-score, yaitu rata-rata F1-score dari seluruh kelas dengan bobot yang sama tanpa mempertimbangkan jumlah sampel pada masing-masing kelas. Dengan demikian, performa pada kelas minoritas seperti kelas Positif tetap diperhitungkan

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

secara proporsional dalam evaluasi keseluruhan model.

3. HASIL DAN PEMBAHASAN

Eksperimen dilakukan menggunakan tiga arsitektur model deep learning, yaitu CNN, BiLSTM, dan Hybrid CNN-BiLSTM, yang dievaluasi pada tiga skenario rasio pembagian data, yaitu 60:40, 70:30, dan 80:20. Dataset terdiri dari 3.416 komentar dari lima platform media sosial dengan distribusi kelas yang tidak seimbang, yaitu Negatif 65,6%, Netral 28,3%, dan Positif 6,1%. Untuk menangani ketidakseimbangan kelas tersebut, seluruh model dilatih menggunakan skema *class weight*, yaitu pembobotan penalti pada fungsi loss tanpa mengubah jumlah sampel pada setiap kelas. Pelatihan dilakukan menggunakan konfigurasi TensorFlow/Keras dengan optimizer Adam, batch size 32, dan 10 epoch. Kinerja model dievaluasi dengan menempatkan macro F1-score sebagai metrik utama, sedangkan akurasi, precision, recall, F1-score per kelas, dan confusion matrix digunakan sebagai metrik pendukung untuk menganalisis performa model secara lebih komprehensif. Skenario 80:20 ditetapkan sebagai konfigurasi utama karena menghasilkan performa terbaik dengan jumlah data uji yang masih representatif untuk evaluasi.

3.1 Pengujian Berbagai Rasio Pembagian Data Latih dan Data Uji

Untuk mengkaji pengaruh skenario pembagian data terhadap kinerja model, tiga rasio diuji pada CNN, BiLSTM, dan Hybrid CNN-BiLSTM secara *stratified* untuk memastikan proporsi setiap kelas sentimen tetap konsisten dengan distribusi awal, sebagaimana disajikan pada Tabel 2. Pengujian dilakukan secara terpisah pada setiap arsitektur model untuk memperoleh gambaran yang objektif mengenai sensitivitas masing-masing model terhadap variasi proporsi data latih dan data uji.

Tabel 2. Hasil Akurasi Seluruh Model

Model	60:40	70:30	80:20
CNN	92,47%	93,17%	93,27%
BiLSTM	88,15%	89,27%	89,04%
Hybrid CNN-BiLSTM	91,08%	92,00%	94,59%

Berdasarkan Tabel 2, ketiga model menunjukkan peningkatan akurasi seiring bertambahnya proporsi data latih. Model Hybrid CNN-BiLSTM secara konsisten memperoleh akurasi tertinggi pada setiap rasio, dengan peningkatan sebesar 3,51% dari rasio 60:40 ke rasio 80:20. Peningkatan ini jauh lebih besar dibandingkan CNN yang hanya meningkat 0,80% dan BiLSTM sebesar 0,89%, mengindikasikan bahwa arsitektur Hybrid yang lebih kompleks lebih responsif terhadap penambahan data latih. Model CNN menunjukkan peningkatan yang stabil, sementara BiLSTM mengalami sedikit fluktuasi dengan akurasi menurun dari 89,27% pada rasio 70:30 menjadi 89,04% pada rasio 80:20, yang mengindikasikan sensitivitas arsitektur rekuren terhadap variasi distribusi data latih. Berdasarkan hasil tersebut, rasio 80:20 dipilih sebagai konfigurasi utama karena menghasilkan akurasi tertinggi untuk seluruh model, khususnya Hybrid CNN-BiLSTM yang mencapai 94,59%. Rasio ini juga memberikan jumlah data uji yang memadai 684 sampel untuk evaluasi yang representatif, terutama pada kelas Positif yang hanya mencakup 6,1% dari total dataset. Penambahan proporsi data latih di atas 80% diperkirakan akan mengurangi jumlah data uji secara signifikan sehingga berpotensi menghasilkan estimasi kinerja yang kurang dapat diandalkan.

3.2 Hasil Evaluasi

Evaluasi kinerja model merupakan tahap krusial dalam penelitian ini untuk mengukur sejauh mana model yang dikembangkan mampu melakukan klasifikasi sentimen secara akurat dan andal. Evaluasi dilakukan pada konfigurasi rasio 80:20 yang telah ditetapkan sebagai konfigurasi utama berdasarkan hasil perbandingan antar variasi rasio pembagian data sebelumnya. Pada konfigurasi ini, data latih berjumlah 2.732 komentar dan data uji berjumlah 684 komentar dengan pembagian yang dilakukan secara stratified untuk memastikan proporsi setiap kelas sentimen terwakili secara proporsional pada kedua subset data. Pendekatan stratified sampling dipilih dengan pertimbangan bahwa distribusi kelas pada dataset bersifat tidak seimbang, sehingga diperlukan mekanisme yang menjamin keterwakilan kelas minoritas pada data uji agar hasil evaluasi mencerminkan kondisi nyata secara lebih representatif. Untuk memahami kinerja model secara lebih rinci, dilakukan analisis metrik per kelas pada model Hybrid CNN-BiLSTM sebagaimana disajikan pada Tabel 3.

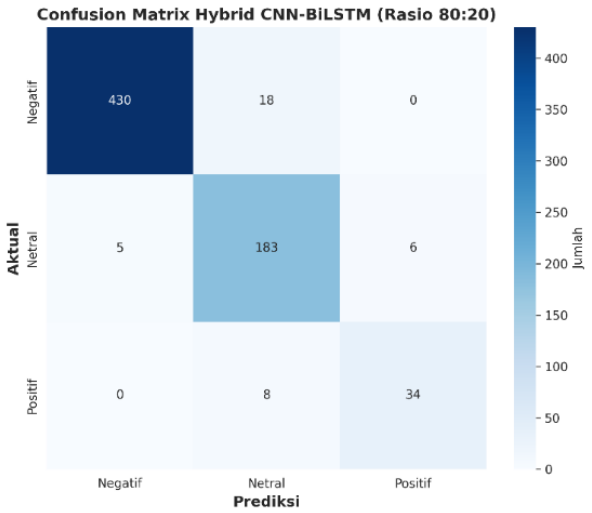
Tabel 3. Precision, Recall, dan F1-Score Per Kelas Model Hybrid CNN-BiLSTM

Kelas	Precision	Recall	F1-score
Negatif	98,85%	95,98%	97,40%
Netral	87,56%	94,33%	90,82%
Positif	85,00%	80,95%	82,93%

Tabel 3 menunjukkan kelas Negatif memperoleh nilai *precision*, *recall*, dan *F1-score* tertinggi di antara ketiga kelas, yaitu masing-masing sebesar 98,85%, 95,98%, dan 97,40%. Nilai tersebut menunjukkan bahwa model mampu mengenali komentar bernada negatif dengan sangat baik sekaligus jarang salah mengklasifikasikannya ke kelas lain.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

Kondisi ini konsisten dengan karakteristik dataset, di mana kelas Negatif mendominasi distribusi karena isu keracunan secara alami memicu lebih banyak komentar berisi kritik, keluhan, dan kekhawatiran dibandingkan komentar Netral maupun Positif. Meskipun skema *class weight* telah diterapkan untuk menyeimbangkan kontribusi setiap kelas selama pelatihan, dominasi jumlah dan keragaman ekspresi pada komentar Negatif tetap berkontribusi pada performa tertinggi di kelas ini. Pada kelas Netral, model memperoleh *precision* 87,56%, *recall* 94,33%, dan *F1-score* 90,82%. Nilai *recall* yang tinggi mengindikasikan bahwa sebagian besar komentar yang bersifat informatif atau deskriptif tanpa muatan emosi dapat terdeteksi dengan baik sebagai Netral. Nilai *precision* yang sedikit lebih rendah dibandingkan kelas Negatif menunjukkan masih terdapat sejumlah komentar dari kelas lain yang terklasifikasikan sebagai Netral, misalnya komentar yang mengandung nada kecewa ringan namun secara konteks lebih mendekati Netral. Di antara ketiga kelas, kelas Positif merupakan yang paling menantang bagi model. Model memperoleh *precision* 85,00%, *recall* 80,95%, dan *F1-score* 82,93%. Nilai *precision* yang cukup tinggi menunjukkan bahwa ketika model memprediksi suatu komentar sebagai Positif, prediksi tersebut umumnya tepat. Adapun nilai *recall* yang lebih rendah dibandingkan kelas Negatif dan Netral mengindikasikan masih terdapat komentar Positif yang salah diklasifikasikan ke kelas Netral. Hal ini disebabkan oleh keterbatasan jumlah sampel kelas Positif pada dataset, sehingga model tidak memperoleh cukup contoh untuk mempelajari variasi ekspresi positif secara menyeluruh. Class weight digunakan untuk menangani ketidakseimbangan kelas tanpa mengubah jumlah data fisik pada masing-masing kelas. Pendekatan ini memberikan bobot penalti lebih besar terhadap kesalahan prediksi pada kelas minoritas, sehingga model tidak hanya mengoptimalkan prediksi pada kelas mayoritas. Namun, mengingat distribusi data penelitian ini masih sangat tidak seimbang, khususnya kelas Positif yang hanya berjumlah 209 komentar, class weight tidak dimaknai sebagai metode yang sepenuhnya menghilangkan bias kelas**.** Oleh karena itu, untuk memastikan bahwa akurasi 94,59% tidak ditafsirkan sebagai satu-satunya indikator keberhasilan model, evaluasi utama penelitian ini menggunakan macro F1-score sebesar 90,38%, yang menghitung rata-rata F1-score setiap kelas dengan bobot yang sama tanpa mempertimbangkan jumlah sampel per kelas. Berbeda dengan akurasi yang dapat menyesatkan pada data tidak seimbang, macro F1-score memastikan kinerja pada kelas Positif yang hanya mencakup 6,1% dataset tetap diperhitungkan secara proporsional. Nilai macro F1-score 90,38% mengindikasikan bahwa model tidak hanya bekerja pada kelas mayoritas, tetapi juga memiliki performa yang cukup baik pada kelas minoritas, meskipun sensitivitas terhadap kelas Positif masih perlu ditingkatkan. Untuk memperoleh gambaran yang lebih komprehensif mengenai distribusi kesalahan prediksi model, dilakukan analisis confusion matrix yang memvisualisasikan jumlah prediksi benar dan salah untuk setiap kelas sentimen. Confusion matrix memberikan informasi yang lebih rinci dibandingkan metrik agregat seperti akurasi, karena memungkinkan identifikasi pola kesalahan spesifik antar kelas. Hasil confusion matrix model Hybrid CNN-BiLSTM pada rasio 80:20 disajikan pada Gambar 6.

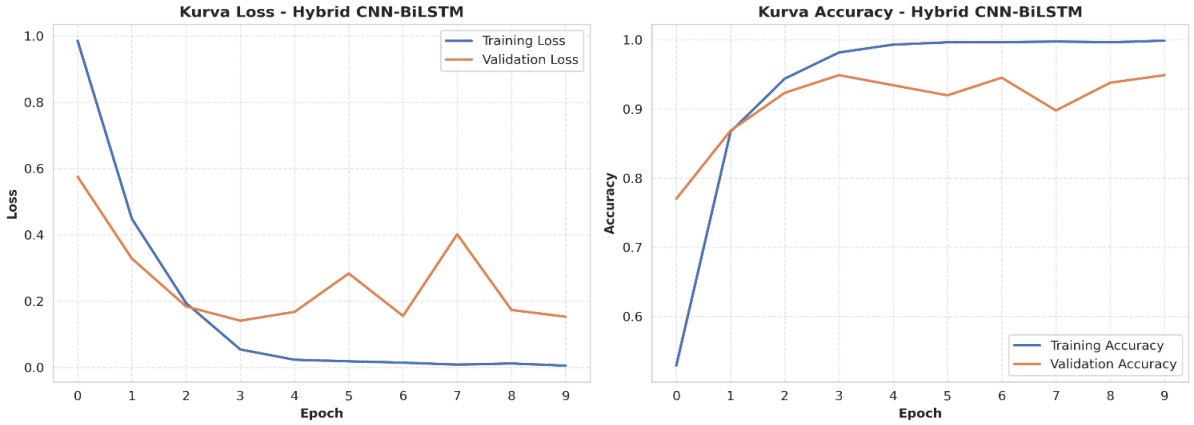


Gambar 6. Confusion Matrix Model Hybrid CNN–BiLSTM

Secara visual, pola evaluasi per kelas tersebut tercermin pada *confusion matrix* Gambar 6. Pada kelas Negatif, sebanyak 430 komentar berhasil diprediksi dengan benar, sedangkan 18 komentar bergeser ke kelas Netral dan tidak ada yang salah terbaca sebagai Positif. Pada kelas Netral, 183 komentar terklasifikasikan dengan benar, sementara 5 komentar salah diklasifikasikan sebagai Negatif dan 6 komentar lainnya sebagai Positif. Pada kelas Positif, tidak terdapat komentar yang salah diklasifikasikan sebagai Negatif, 34 komentar diprediksi benar sebagai Positif, sedangkan 8 komentar bergeser ke kelas Netral. Pola ini menguatkan temuan pada Tabel 3 bahwa model sangat andal pada kelas Negatif dan cukup stabil pada kelas Netral, namun masih cenderung mengklasifikasikan komentar Positif yang ekspresinya halus ke dalam kelas Netral. Secara keseluruhan, model Hybrid CNN-BiLSTM telah menunjukkan kinerja yang baik pada ketiga kelas sentimen, dengan ruang peningkatan yang tersisa terutama pada sensitivitas terhadap kelas Positif. Selain evaluasi berbasis metrik klasifikasi, proses pelatihan model juga dianalisis melalui kurva *loss* dan *accuracy* per *epoch* untuk menilai stabilitas dan konvergensi model selama pelatihan. Kurva pelatihan model

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

Hybrid CNN-BiLSTM disajikan pada Gambar 7.



Gambar 7. Kurva Pelatihan Model Hybrid CNN-BiLSTM

Berdasarkan Gambar 7, training loss menurun konsisten dari sekitar 1,0 pada epoch pertama hingga mendekati 0 pada epoch kesepuluh, sementara validation loss menunjukkan tren penurunan dari 0,58 menjadi 0,16 dengan fluktuasi pada beberapa epoch, khususnya epoch keempat dan ketujuh. Fluktuasi tersebut dapat dipengaruhi oleh distribusi kelas yang tidak seimbang, terutama karena jumlah sampel kelas minoritas relatif terbatas sehingga perubahan kecil pada data validasi dapat memengaruhi nilai loss. Di sisi lain, training accuracy yang mendekati 100% sementara validation accuracy berada pada kisaran 90-95% menunjukkan adanya indikasi overfitting. Kondisi ini mengindikasikan bahwa model berpotensi mempelajari pola spesifik pada data latih, terutama mengingat ukuran dataset yang relatif terbatas dan distribusi kelas yang tidak seimbang. Meskipun validation accuracy tidak menunjukkan penurunan drastis, hasil pelatihan tetap perlu ditafsirkan secara hati-hati. Oleh karena itu, overfitting dinyatakan sebagai salah satu keterbatasan penelitian. Penelitian lanjutan dapat menerapkan strategi seperti early stopping, regularisasi tambahan, atau augmentasi teks untuk meningkatkan kemampuan generalisasi model.

3.3 Analisis Kesalahan Klasifikasi

Untuk memahami keterbatasan model secara lebih mendalam, dilakukan analisis terhadap 37 komentar yang salah diklasifikasikan dari total 684 data uji (5,41%). Distribusi kesalahan menunjukkan bahwa 18 komentar Negatif salah diprediksi sebagai Netral, 11 komentar Netral salah diprediksi ke kelas lain, dan 8 komentar Positif salah diprediksi sebagai Netral. Kualitas label referensi berbasis leksikon telah divalidasi melalui anotasi manual terhadap 300 sampel data oleh dua anotator independen dengan nilai Cohen's Kappa sebesar κ = 0,828. Namun, analisis kualitatif menunjukkan bahwa sebagian kesalahan klasifikasi tetap terjadi pada komentar dengan nuansa bahasa yang kompleks, seperti kalimat argumentatif, negasi, sarkasme, dan polaritas campuran, sebagaimana disajikan pada Tabel 4.

Tabel 4. Contoh Kesalahan Klasifikasi Model Hybrid CNN-BiLSTM

No	Teks	Label Referensi	Prediksi	Penyebab Kesalahan
1.	"Malah yang gak wajar cara berpikirnya, keracunan adalah masalah, ada 1 keracunan itu masalah. Kaya kasus pembunuhan ada terbunuh 1 harus diungkap karena terkait nyawa manusia"	Negatif	Netral	Kalimat bersifat argumentatif dengan analogi perbandingan, sehingga meski mengandung kritik kuat terhadap MBG, model menangkapnya sebagai Netral karena tidak ada kata emosi negatif yang eksplisit dalam leksikon
2.	"Ya kalo secara data emang kecil sih, cuma bukan berarti program ini sukses. Bisa dikatakan sukses kalo sampe gaada kasus keracunan lagi"	Positif	Netral	Kata 'sukses' muncul berulang sehingga leksikon memberi label Positif, namun model tidak dapat menangkap bahwa 'sukses' digunakan dalam konteks negasi yaitu program belum bisa dikatakan sukses selama masih ada keracunan
3.	"Demi untung dan menguntungkan segelintir orang saja nyawa anak anak dalam bahaya MBG makanan busuk gratis"	Negatif	Netral	Kata 'untung' dan 'menguntungkan' di awal kalimat ditangkap model sebagai sinyal positif sehingga mengaburkan muatan negatif kuat dari frasa 'nyawa anak anak dalam bahaya' dan 'makanan busuk gratis'

Berdasarkan Tabel 4, terdapat beberapa pola kesalahan utama. Pertama, komentar negatif yang disampaikan

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

dalam bentuk argumentatif atau informatif cenderung diprediksi sebagai Netral karena model tidak selalu menangkap intensitas kritik secara eksplisit. Kedua, komentar yang mengandung negasi atau struktur kontrastif, seperti penggunaan kata "sukses" dalam konteks "bukan berarti program ini sukses", menimbulkan ambiguitas polaritas sehingga model menggeser prediksi ke kelas Netral. Ketiga, bahasa informal, kalimat tidak baku, dan ekspresi sarkastis pada media sosial masih menjadi tantangan karena makna sentimen tidak selalu muncul melalui kata-kata emosional yang eksplisit. Temuan ini menunjukkan bahwa kesalahan klasifikasi tidak hanya dipengaruhi oleh arsitektur model, tetapi juga oleh kompleksitas bahasa informal dan keterbatasan label referensi berbasis leksikon dalam menangkap konteks pragmatik tertentu. Oleh karena itu, hasil analisis kesalahan ini memperjelas bahwa komentar dengan ekspresi ambigu, negasi, sarkasme, dan polaritas campuran masih menjadi area yang paling menantang bagi model.

3.4 Pembahasan

Untuk mengkaji posisi kontribusi penelitian ini dalam perkembangan literatur analisis sentimen, hasil yang diperoleh dibandingkan dengan beberapa penelitian terdahulu yang relevan sebagaimana disajikan pada Tabel 5. Perbandingan ini mencakup aspek dataset, model yang digunakan, serta hasil performa yang diperoleh. Perbandingan ini bertujuan untuk menempatkan kontribusi penelitian dalam konteks perkembangan metode analisis sentimen berbasis *deep learning,* dengan mempertimbangkan bahwa setiap penelitian menggunakan dataset dan konteks yang berbeda.

Tabel 5. Perbandingan Berdasarkan Metode Terdahulu

Peneliti	Dataset	Model	Akurasi	F1-Score
Samonte *et al.* (2023)	Tweet COVID-19	Hybrid CNN-LSTM, CNN, dan LSTM	Hybrid CNN-LSTM: 84,72%; CNN: 84,29%; LSTM: 83,81%	85,00%
Alawi & Bozkurt (2024)	Tweet X Berbahasa Turki	BERT-BiLSTM-CNN	BERT- BiLSTM CNN:91,01 %;	88,01%
Triningsih *et al.* (2025)	Tweet X Program Makan Bergizi Gratis	SVM dan Random Fo rest	SVM: 85,74%; Random Forest: 81,53%	-
Penelitian ini (2026)	Komentar 5 platform	Hybrid CNN-BiLSTM	Hybrid CNN-BiLSTM: 94,59%	90,38%

Berdasarkan Tabel 5, penelitian Triningsih et al. yang menggunakan dataset dengan topik sama namun pendekatan *machine learning* konvensional memperoleh akurasi tertinggi 85,74% menggunakan SVM. Meskipun SVM terbukti efektif, pendekatan konvensional memiliki keterbatasan dalam menangkap representasi semantik yang lebih dalam pada teks tidak terstruktur seperti komentar media sosial. Samonte et al. yang menerapkan Hybrid CNN-LSTM pada dataset tweet COVID-19 memperoleh akurasi 84,72%, mengonfirmasi bahwa penggabungan dua arsitektur *deep learning* menghasilkan performa lebih baik dibandingkan model tunggal. Penelitian ini mengembangkan pendekatan serupa dengan mengganti LSTM satu arah menjadi BiLSTM yang memodelkan konteks dari dua arah secara simultan, dan pada konteks dataset penelitian ini menghasilkan akurasi 94,59%. Adapun Alawi dan Bozkurt yang menggunakan BERT-BiLSTM-CNN memperoleh akurasi 91,01% pada dataset tweet berbahasa Turki, sementara pada konteks dataset penelitian ini model Hybrid CNN-BiLSTM mencapai akurasi 94,59% tanpa memerlukan *pre-trained language model*. Secara keseluruhan, model Hybrid CNN-BiLSTM pada penelitian ini mencapai *macro* F1-score 90,38% dan akurasi 94,59%, menunjukkan performa yang kompetitif dibandingkan penelitian pembanding. Keunggulan ini menunjukkan bahwa kombinasi ekstraksi fitur lokal oleh CNN dan pemodelan konteks dua arah oleh BiLSTM yang diperkuat skema *class weight* merupakan pendekatan yang efektif untuk analisis sentimen komentar media sosial berbahasa Indonesia pada isu kebijakan publik berskala nasional.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa model Hybrid CNN-BiLSTM efektif untuk analisis sentimen komentar publik multi-platform terkait insiden keamanan pangan Program Makan Bergizi Gratis, dengan macro F1-score sebesar 90,38% sebagai metrik utama dan akurasi sebesar 94,59% sebagai metrik pendukung. Secara ilmiah, kombinasi CNN untuk mengekstraksi pola lokal dan BiLSTM untuk memodelkan dependensi urutan kata dua arah terbukti relevan dalam menangani variasi bahasa informal pada komentar media sosial dari berbagai platform. Pendekatan ini penting karena komentar dari X, Instagram, TikTok, YouTube, dan Facebook menunjukkan keragaman karakteristik linguistik yang berbeda sesuai konteks penggunaan masing-masing platform. Secara praktis, model ini dapat dimanfaatkan sebagai komponen awal dalam monitoring opini publik terhadap kebijakan pemerintah, misalnya untuk mengidentifikasi kecenderungan sentimen negatif, mengenali isu dominan yang memicu keresahan publik, serta membantu penyusunan respons komunikasi yang lebih cepat dan berbasis data. Namun, hasil penelitian ini tidak dimaknai sebagai model yang sepenuhnya bebas dari bias kelas, karena performa pada kelas Positif masih

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

perlu ditingkatkan akibat jumlah datanya yang relatif terbatas. Selain itu, penggunaan label referensi berbasis leksikon masih memiliki keterbatasan dalam menangkap sarkasme, negasi, ironi, dan kalimat kontrastif. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas anotasi manual dan menambah representasi data pada kelas minoritas, serta menguji model pada isu kebijakan publik lain agar kemampuan generalisasinya dapat dievaluasi secara lebih kuat. Dari sisi kontribusi, penelitian ini berhasil menghasilkan dataset multi-platform berbahasa Indonesia yang tervalidasi secara manual sebagai sumber data yang dapat dimanfaatkan pada penelitian lanjutan. Selain itu, arsitektur hybrid CNN-BiLSTM dengan skema *class weight* yang dikembangkan terbukti efektif untuk klasifikasi sentimen tiga kelas pada teks informal, sekaligus membuka peluang penerapan *deep learning* sebagai komponen pemantauan opini publik yang responsif dan berbasis data.

REFERENCES

[1] Badan Kebijakan Pembangunan Kesehatan, "Survei Status Gizi Indonesia 2024 I Tim Penyusun SSGI 2024 Dalam Angka," 2025. [Online]. Available: https://www.badankebijakan.kemkes.go.id/survei-status-gizi-indonesia-ssgi-2024/

[2] Badan Kebijakan Pembangunan Kesehatan, "Buku Saku Hasil Survei Status Gizi Indonesia (SSGI) 2022." [Online]. Available: https://www.badankebijakan.kemkes.go.id/laporan-hasil-survei/

[3] R. N. Ikhsani and F. F. Abdulloh, "Optimasi SVM dan Decision Tree Menggunakan SMOTE Untuk Mengklasifikasi Sentimen Masyarakat Mengenai Pinjaman Online," *Jurnal Media Informatika Budidarma*, vol. 7, no. 4, p. 1667, Oct. 2023, doi: 10.30865/mib.v7i4.6809.

[4] E. Triningsih, M. Afdal, I. Permana, and N. Evrilyan Rozanda, "Analisis Sentimen Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning Pada Sosial Media X," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 4, pp. 2240–2250, 2025, doi: 10.47065/bits.v6i4.6534.

[5] F. A. Ryandi, D. Pratiwi, and S. Sari, "Analisis Sentimen Masyarakat Di Media Sosial X Terhadap Kemenkes Dengan Naive Bayes dan SVM," *Jurnal Sains dan Teknologi*, vol. 7, no. 1, pp. 1–6, 2025, doi: 10.55338/saintek.v7i1.4615.

[6] R. A. Rudiyanto and E. B. Setiawan, "Sentiment Analysis Using Convolutional Neural Network (CNN) and Particle Swarm Optimization on Twitter," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 9, no. 2, pp. 188–195, Feb. 2024, doi: 10.33480/jitk.v9i2.5201.

[7] M. Z. Rahman, Y. A. Sari, and N. Yudistira, "Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 11, pp. 5120–5127, 2021, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10188

[8] D. Andriyani, F. Ahmad, and E. P. Sandy, "The Effect of SMOTE Application on Support Vector Machine Performance in Sentiment Classification on Imbalanced Datasets," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 4, no. 2, pp. 752–757, Feb. 2025, doi: 10.59934/jaiea.v4i2.742.

[9] S. F. Taskiran, B. Turkoglu, E. Kaya, and T. Asuroglu, "A comprehensive evaluation of oversampling techniques for enhancing text classification performance," *Sci. Rep.*, vol. 15, no. 1, p. 21631, Jul. 2025, doi: 10.1038/s41598-025-05791-7.

[10] S. Tam, R. Ben Said, and O. O. Tanriover, "A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification," *IEEE Access*, vol. 9, pp. 41283–41293, 2021, doi: 10.1109/ACCESS.2021.3064830.

[11] M. Mhamed, R. Sutcliffe, X. Sun, J. Feng, E. Almekhlafi, and E. A. Retta, "Improving Arabic Sentiment Analysis Using CNN-Based Architectures and Text Preprocessing," *Comput. Intell. Neurosci.*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/5538791.

[12] M. J. C. Samonte, A. T. G. Dela Rosa, L. J. C. Rivera, and J. S. E. Silo, "Using Hybrid CNN-LSTM Model for Sentiment Analysis of COVID-19 Tweets," in *2023 13th International Conference on Software Technology and Engineering (ICSTE)*, IEEE, Oct. 2023, pp. 133–142. doi: 10.1109/ICSTE61649.2023.00029.

[13] A. B. Alawi and F. Bozkurt, "A hybrid machine learning model for sentiment analysis and satisfaction assessment with Turkish universities using Twitter data," *Decision Analytics Journal*, vol. 11, p. 100473, Jun. 2024, doi: 10.1016/j.dajour.2024.100473.

[14] W. Meng, Y. Wei, P. Liu, Z. Zhu, and H. Yin, "Aspect Based Sentiment Analysis With Feature Enhanced Attention CNN-BiLSTM," *IEEE Access*, vol. 7, pp. 167240–167249, 2019, doi: 10.1109/ACCESS.2019.2952888.

[15] L. Hidayati, L. P. Kusuma, D. Agustini, and V. Y. P. Ardhana, "Implementasi Web Scraping Untuk Pengumpulan Data Media Sosial Lingkup Pemerintah Provinsi NTB," *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 7, no. 1, pp. 63–72, Mar. 2024, doi: 10.47080/simika.v7i1.3200.

[16] K. Kusrini, M. R. Arif Yudianto, and H. Al Fatta, "The effect of Gaussian filter and data preprocessing on the classification of Punakawan puppet images with the convolutional neural network algorithm," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, p. 3752, Aug. 2022, doi: 10.11591/ijece.v12i4.pp3752-3761.

[17] M. Herrmann, M. Obaidi, L. Chazette, and J. Klünder, "On the subjectivity of emotions in software projects: How reliable are pre-labeled data sets for sentiment analysis?," *Journal of Systems and Software*, vol. 193, p. 111448, Nov. 2022, doi: 10.1016/j.jss.2022.111448.

[18] U. Ependi, S. Aliya, and A. Wibowo, "Sentiment Analysis of Covid-19 Handling in Indonesia Based on Lexicon Weighting," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 1, pp. 76–82, Mar. 2023, doi: 10.32736/sisfokom.v12i1.1615.

[19] H. M. Shakeel, S. Iram, H. Al-Aqrabi, T. Alsboui, and R. Hill, "A Comprehensive State-of-the-Art Survey on Data Visualization Tools: Research Developments, Challenges and Future Domain Specific Visualization Framework," *IEEE Access*, vol. 10, pp. 96581–96601, 2022, doi: 10.1109/ACCESS.2022.3205115.

[20] V. R. Prasetyo, M. F. Naufal, and K. Wijaya, "Sentiment Analysis of ChatGPT on Indonesian Text using Hybrid CNN and Bi-LSTM," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 9, no. 2, pp. 327–333, Apr. 2025, doi: 10.29207/resti.v9i2.6334.

[21] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018, doi: 10.1007/s13244-018-0639-9.

[22] E. Y. Hidayat and D. Handayani, "Penerapan 1D-CNN untuk Analisis Sentimen Ulasan Produk Kosmetik Berdasar Female

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 184−197
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9896

Daily Review," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 8, no. 3, pp. 153–163, Jan. 2023, doi: 10.25077/TEKNOSI.v8i3.2022.153-163.

[23] M. E. Alzahrani, T. H. H. Aldhyani, S. N. Alsubari, M. M. Althobaiti, and A. Fahad, "Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–10, May 2022, doi: 10.1155/2022/3840071.

[24] A. R. Gunawan and R. F. Alfa Aziza, "Sentiment Analysis Using LSTM Algorithm Regarding Grab Application Services in Indonesia," *Journal of Applied Informatics and Computing*, vol. 9, no. 2, pp. 322–332, Mar. 2025, doi: 10.30871/jaic.v9i2.8696.

[25] L. Khan, A. Amjad, K. M. Afaq, and H.-T. Chang, "Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media," *Applied Sciences*, vol. 12, no. 5, p. 2694, Mar. 2022, doi: 10.3390/app12052694.