ANALISIS SENTIMEN PENGALAMAN PENGGUNA TWITTER TERHADAP CHATGPT MENGGUNAKAN CNN DAN LSTM

Adam Mohammad Nur[1], Bayu Sulistiyanto Ipung Sutejo[2], Gerry Firmansyah[3], Budi Tjahjono[4]
Program Studi Teknik Informatika, Fakultas Ilmu Komputer
Universitas Esa Unggul[1,2,3,4]
Jalan Arjuna Utara No. 9, Kebon Jeruk, Jakarta Barat 11510
adammohammadnur14@student.esaunggul.ac.id[1], bayu.sulistiyanto@esaunggul.ac.id[2],
gerry@esaunggul.ac.id[3], budi.tjahjono@esaunggul.ac.id[4]

Abstrak

Perkembangan *Generative Artificial Intelligence*, khususnya ChatGPT, telah memunculkan berbagai tanggapan pengguna di media sosial Twitter (X), berupa dukungan, kritik, kekhawatiran, maupun informasi netral. Penelitian ini bertujuan menganalisis sentimen pengguna Twitter (X) terhadap ChatGPT serta membandingkan performa model *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) dalam mengklasifikasikan sentimen negatif, netral, dan positif. Penelitian menggunakan pendekatan kuantitatif berbasis *deep learning*. Data dikumpulkan dengan kata kunci "ChatGPT" dan "Chat GPT". Dari 11.210 *tweet* yang diperoleh, 10.586 *tweet* digunakan setelah proses *cleansing* dengan menghapus URL, *mention*, *hashtag*, emoji, simbol, tanda baca, karakter khusus, dan spasi berlebih. Distribusi data terdiri atas 6.281 *tweet* negatif, 3.484 netral, dan 821 positif. Data dibagi dengan rasio 80:20 menjadi 8.468 data latih dan 2.118 data uji. Evaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. CNN memperoleh *accuracy* 69,55%, *macro F1-score* 57,17%, dan *weighted F1-score* 69,36%, sedangkan LSTM memperoleh *accuracy* 72,29%, *macro F1-score* 62,07%, dan *weighted F1-score* 72,78%. LSTM menghasilkan 1.531 prediksi benar, sedangkan CNN 1.473. Berdasarkan hasil tersebut, LSTM menunjukkan performa lebih baik dalam mengklasifikasikan sentimen terhadap ChatGPT.

Kata Kunci: Analisis Sentimen, ChatGPT, Twitter (X), *Convolutional neural network*, *long short-term memory.*

Abstract

The *development of generative artificial intelligence, particularly ChatGPT, has elicited various responses from users on the social media platform Twitter (X), including support, criticism, concerns, and neutral information. This research aims to analyze Twitter (X) users' sentiments toward ChatGPT and compare the performance of the Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models in classifying negative, neutral, and positive sentiments. The research uses a quantitative approach based on deep learning. Data were collected using the keywords "ChatGPT" and "Chat GPT." From the 11,210 tweets obtained, 10,586 were used after the cleansing process, which involved removing URLs, mentions, hashtags, emojis, symbols, punctuation, special characters, and excessive spaces. The data distribution consists of 6,281 negative tweets, 3,484 neutral, and 821 positive. The data was divided with an 80:20 ratio into 8,468 training data and 2,118 test data. Evaluation used accuracy, precision, recall, and F1-score. CNN achieved an accuracy of 69.55%, a macro F1-score of 57.17%, and a weighted F1-score of 69.36%, while LSTM achieved an accuracy of 72.29%, a macro F1-score of 62.07%, and a weighted F1-score of 72.78%. LSTM produced 1,531 correct predictions, while CNN produced 1,473. Based on these results, LSTM shows better performance in classifying sentiment toward ChatGPT.*

Keywords*: Sentiment Analysis, ChatGPT, Twitter (X), Convolutional neural network, long short-term memory.*

PENDAHULUAN

Kemajuan teknologi *Artificial intelligence* (AI) pada era digital memberikan dampak signifikan terhadap berbagai aspek kehidupan, mulai dari pendidikan, industri, komunikasi, hingga aktivitas sosial. Salah satu bentuk perkembangan AI yang paling menonjol adalah *Generative Artificial intelligence*, yaitu teknologi yang mampu menciptakan konten berupa teks, gambar, kode program, maupun respons percakapan berbasis bahasa alami. Miyazaki et al (2024) menjelaskan bahwa *Generative* AI memperoleh perhatian besar karena kemajuannya dalam pembuatan teks, gambar, dan perangkat lunak, sedangkan ChatGPT menjadi salah satu aplikasi paling populer karena kemampuan percakapannya. Appiah & Htait (2024) menyatakan bahwa AI telah menjadi topik yang banyak diperbincangkan di media sosial karena berkaitan dengan persepsi publik, manfaat, risiko, etika, dan dampak sosialnya. Di sisi lain, perkembangan teknologi digital juga menghadirkan berbagai risiko yang perlu diperhatikan. Sutejo et al (2025) menunjukkan bahwa keamanan siber, privasi data, kepatuhan regulasi, integrasi

sistem, dan kualitas data merupakan sejumlah aspek penting dalam pengelolaan risiko transformasi digital. Temuan tersebut menunjukkan bahwa perkembangan teknologi tidak hanya perlu dipandang dari sisi manfaat dan efisiensi, tetapi juga dari sisi keamanan, tata kelola, dan potensi risiko yang ditimbulkannya.

Twitter (X) merupakan platform *microblogging* yang memungkinkan pengguna menyampaikan opini, kritik, dukungan, maupun kekhawatiran secara cepat melalui unggahan singkat. Juwiantho et al (2020) menyatakan bahwa media sosial memungkinkan masyarakat menyampaikan pendapat dan informasi melalui tulisan maupun gambar, sedangkan Isnain et al (2021) menjelaskan bahwa Twitter memungkinkan pengguna mendiskusikan berbagai isu yang sedang ramai diperbincangkan. Mudding & Arifin A Abd Karim (2022) menegaskan bahwa media sosial menghasilkan banyak opini yang dapat dimanfaatkan untuk memahami respons publik terhadap suatu fenomena, termasuk perkembangan ChatGPT. Opini pengguna terhadap ChatGPT tidak selalu seragam; sentimen positif muncul karena ChatGPT dianggap membantu produktivitas, sedangkan sentimen negatif muncul akibat kekhawatiran terhadap akurasi informasi, ketergantungan pada teknologi, etika, dan hak cipta.

Analisis sentimen digunakan sebagai metode untuk mengolah data teks dan mengklasifikasikan opini ke dalam kategori tertentu. Joloudari et al (2023) menjelaskan bahwa analisis sentimen merupakan alat analisis teks yang dapat mendeteksi opini dan emosi dari data tidak terstruktur. Namun, data Twitter memiliki tantangan tersendiri karena bersifat singkat, tidak formal, dan mengandung singkatan, emoji, *hashtag*, *mention*, serta karakter khusus, sehingga memerlukan tahapan *preprocessing* (Juwiantho et al., 2020). Berbagai metode *machine learning* klasik telah diterapkan dalam analisis sentimen, salah satunya adalah *K-Nearest Neighbor* (Isnain et al., 2021), maupun *deep learning*. Ahmad et al (2023) menjelaskan bahwa CNN efektif mengidentifikasi keterkaitan sentimen jangka pendek, sedangkan LSTM memodelkan hubungan sentimen jangka panjang antarkata. Wardianto dan Farikhin (2023) menambahkan bahwa LSTM mampu menyimpan informasi jangka panjang dalam sel memori dan mempelajari konteks untuk tugas klasifikasi.

Meskipun penelitian sebelumnya telah banyak membahas analisis sentimen pada ulasan produk, aplikasi, maupun persepsi AI secara umum, penelitian yang secara khusus menganalisis sentimen pengguna Twitter (X) terhadap ChatGPT dengan membandingkan model CNN dan LSTM masih perlu dikembangkan. Mohbey et al (2022) menggunakan pendekatan CNN-LSTM untuk mengklasifikasikan sentimen *tweet* ke dalam kategori positif, negatif, dan netral, sehingga pendekatan tersebut relevan sebagai dasar penggunaan model *deep learning* pada data Twitter. Berdasarkan uraian tersebut, penelitian ini bertujuan untuk: mengumpulkan dan melakukan *preprocessing* data *tweet* terkait ChatGPT, mengidentifikasi kecenderungan sentimen pengguna Twitter (X) terhadap ChatGPT, menerapkan model CNN dan LSTM untuk klasifikasi sentimen, dan membandingkan performa kedua model berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

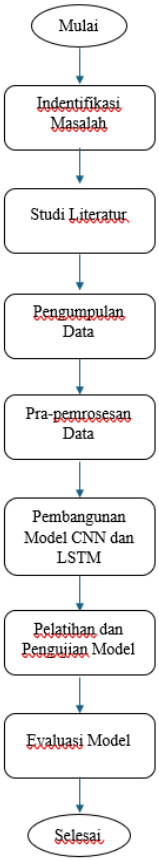
PENELITIAN RELEVAN

Sejumlah penelitian telah menerapkan metode *deep learning* dalam analisis sentimen pada data media sosial. Pipin & Kurniawan (2022) menggunakan LSTM untuk klasifikasi sentimen *multiclass* pada *tweet* berbahasa Indonesia dan memperoleh akurasi sebesar 80,42%. (Alfin et al., 2023) menerapkan CNN pada analisis sentimen layanan Dana dan Sakuku dengan akurasi 81%. Sementara itu, (Prasetyo et al., 2024) menggunakan LSTM pada 1.758 *tweet* terkait isu penundaan Pemilu 2024 dan menghasilkan akurasi sebesar 88%. Penelitian yang membandingkan CNN dan LSTM dilakukan oleh Ahmad et al (2023), dengan hasil LSTM memperoleh akurasi 60,74%, lebih tinggi dibandingkan CNN sebesar 57,99%, meskipun CNN menghasilkan *precision* yang lebih tinggi. Zakharia et al (2025) menerapkan CNN untuk menganalisis sentimen publik di Twitter mengenai pemanfaatan AI dalam seni digital dan memperoleh akurasi 72,28%. Selain itu, (Muammar Khadapi & Pakpahan, 2024) membandingkan LSTM dan BERT pada 4.376 *tweet* terkait Pemilu 2024, dengan LSTM menghasilkan akurasi sekitar 87%, sedangkan BERT sebesar 76,48%. Berbagai penelitian tersebut menunjukkan bahwa CNN dan LSTM dapat diterapkan dalam klasifikasi sentimen pada data Twitter dengan performa yang berbeda sesuai karakteristik data dan objek penelitian. Namun, kajian yang secara khusus membandingkan CNN dan LSTM untuk mengklasifikasikan sentimen pengguna Twitter (X) terhadap ChatGPT ke dalam kategori negatif, netral, dan positif masih perlu dikembangkan. Oleh karena itu, penelitian ini membandingkan kedua model tersebut untuk mengetahui model yang memberikan performa lebih baik dalam analisis sentimen terhadap ChatGPT.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis analisis sentimen dengan metode *deep learning*. Model yang digunakan adalah *Convolutional neural network* (CNN) dan *Long short-term memory* (LSTM). CNN digunakan untuk mengenali pola atau fitur penting pada teks, sedangkan LSTM digunakan untuk memahami urutan dan konteks kata dalam *tweet*. Objek penelitian adalah sentimen pengguna Twitter (X) terhadap tren *Generative Artificial intelligence*, khususnya ChatGPT, yang tercermin dalam kategori sentimen positif, negatif, dan netral.

Sampel penelitian berupa data *tweet* yang dikumpulkan pada periode Januari 2025 hingga Maret 2026 menggunakan kata kunci "ChatGPT" dan "Chat GPT" melalui Twitter API, dengan teknik *purposive sampling*. Data yang berhasil dikumpulkan sebelum *cleansing* berjumlah 11.210 *tweet*. Variabel independen (X) adalah teks *tweet* yang membahas ChatGPT, sedangkan variabel dependen (Y) adalah kategori sentimen hasil klasifikasi model. Tahapan penelitian secara umum diawali dengan identifikasi masalah dan studi literatur, dilanjutkan dengan pengumpulan data *tweet*, pra-pemrosesan data, pembangunan model CNN dan LSTM, pelatihan dan pengujian model, evaluasi performa, hingga penarikan kesimpulan, sebagaimana digambarkan pada Gambar 1.



Gambar 1. *Flowchart* Tahapan Penelitian

Pra-Pemrosesan dan Representasi Data

Tahap pra-pemrosesan dilakukan melalui proses *cleansing* untuk membersihkan data *tweet* dari URL, *mention*, *hashtag*, emoji, simbol, tanda baca, karakter khusus, dan data yang tidak relevan, tanpa mengubah struktur maupun makna kalimat. Data yang telah dibersihkan selanjutnya direpresentasikan ke bentuk numerik melalui tahap *tokenizer* dengan kosakata maksimum 10.000 kata dan token <OOV> untuk kata di luar kosakata. Setiap *tweet* diubah menjadi *sequence* dengan panjang maksimum 200 token melalui proses *post-padding*, sedangkan label sentimen diubah menjadi bentuk *one-hot encoding* dengan nilai 0 untuk negatif, 1 untuk netral, dan 2 untuk positif.

Pembangunan Model CNN

Model CNN dibangun menggunakan *Tensorflow* dengan lapisan *Embedding* (dimensi vektor 128), dua lapisan *Conv1D* (64 dan 32 filter) yang masing-masing diikuti *Batch Normalization* dan

MaxPooling1D, lapisan *GlobalMaxPooling1D* untuk mengambil fitur paling dominan, dua lapisan *Dense* (64 dan 32 unit) dengan *Dropout*, serta lapisan keluaran *Dense*(3, *activation*='*softmax*') untuk tiga kelas sentimen. Model memiliki total 1.335.971 parameter, dengan 1.335.779 parameter dapat dilatih. Model dikompilasi menggunakan *optimizer* Adam dan fungsi *loss categorical_crossentropy*.

Pembangunan Model LSTM

Arsitektur LSTM terdiri atas lapisan *Embedding* berdimensi 128, satu lapisan LSTM dengan 64 unit untuk mempelajari urutan dan konteks kata, lapisan *Dense* (64 unit) dengan regularisasi L2, lapisan *Dropout* untuk mengurangi *overfitting*, serta lapisan *Dense* keluaran dengan aktivasi *softmax* untuk tiga kelas sentimen. Lapisan LSTM memiliki 49.408 parameter, sedangkan lapisan *Dense* pertama memiliki 4.160 parameter.

Pelatihan, Pengujian, dan Evaluasi Model

Dataset dibagi menggunakan rasio 80:20 dengan teknik *stratified random* sampling, sehingga proporsi setiap kelas sentimen tetap terjaga pada data latih dan data uji. Pembagian ini dipilih karena memberikan keseimbangan antara jumlah data yang cukup untuk pelatihan dan pengujian generalisasi model, sekaligus membantu mengantisipasi *overfitting* dan *underfitting* (Ayuningtyas et al., 2024). Kedua model dilatih maksimal 30 *epoch* dengan *batch size* 64 dan *class_weight* untuk menangani ketidakseimbangan kelas. Mekanisme *Earlystopping*, *ReduceLROnPlateau*, dan Model *Checkpoint* digunakan untuk menghentikan pelatihan saat *validation loss* tidak membaik, menurunkan *learning rate*, dan menyimpan bobot model terbaik. Kinerja model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, baik pada tiap kelas maupun secara *macro* dan *weighted average*.

HASIL DAN PEMBAHASAN

Pengumpulan dan *Cleansing* Data

Pengumpulan data *tweet* menggunakan kata kunci "ChatGPT" dan "Chat GPT" menghasilkan 11.210 *tweet*. Setelah melalui proses *cleansing* untuk menghapus URL, *mention*, *hashtag*, emoji, simbol, karakter khusus, duplikasi, dan data tidak relevan, diperoleh 10.586 *tweet* yang digunakan sebagai data penelitian, sebagaimana disajikan pada Tabel 1.

Tabel 1. Hasil *Cleansing* Data *Tweet*

Keterangan	Jumlah Data
Data sebelum *cleansing*	11.210
Data yang tidak digunakan	624
Data setelah *cleansing*	10.586

Data hasil *cleansing* selanjutnya diberi label sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral. Distribusi label sentimen disajikan pada Tabel 2.

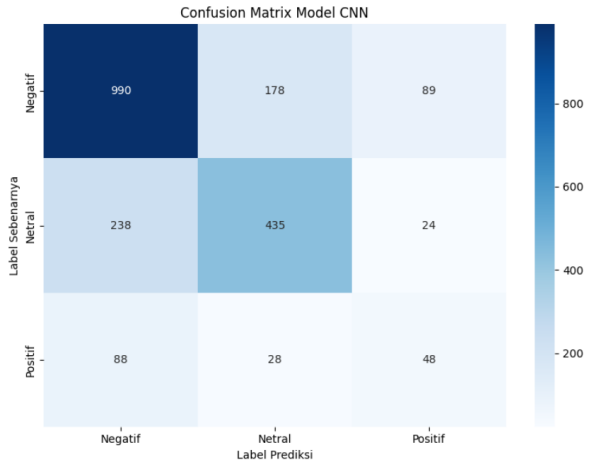
Tabel 2. Distribusi Label Sentimen

Label Sentimen	Jumlah *Tweet*
Positif	821
Negatif	6.281
Netral	3.484
Total	10.586

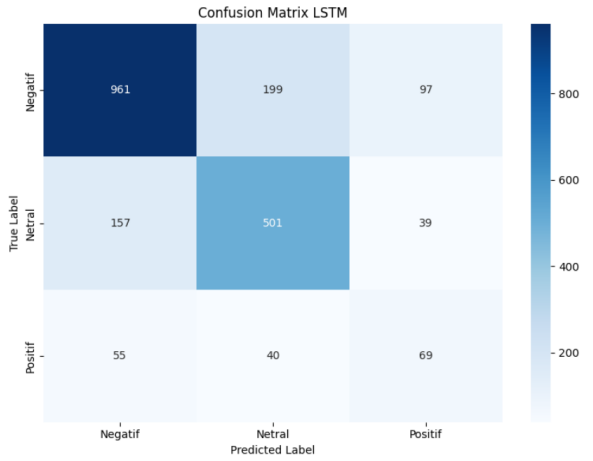
Distribusi data menunjukkan bahwa pembahasan mengenai ChatGPT didominasi oleh sentimen negatif, diikuti netral dan positif. Kondisi ini mengindikasikan ketidakseimbangan kelas yang berpotensi memengaruhi kemampuan model dalam mempelajari pola pada kelas dengan jumlah data lebih sedikit. Data selanjutnya dibagi menggunakan rasio 80:20 menjadi 8.468 data latih dan 2.118 data uji sebelum digunakan pada tahap pelatihan model CNN dan LSTM.

Hasil Pengujian Model CNN dan LSTM

Model CNN dilatih hingga maksimum 30 *epoch*. Pada *epoch* pertama, model memperoleh training *accuracy* sebesar 32,99% dan *validation accuracy* sebesar 59,35%. Nilai training *accuracy* terus meningkat seiring bertambahnya *epoch* hingga mencapai 96,09% pada *epoch* ke-14, sedangkan *validation accuracy* berada pada kisaran 69%–73% pada *epoch-epoch* akhir. *Validation loss* terendah diperoleh pada *epoch* ke-7 sebesar 0,7196, sehingga bobot model pada *epoch* tersebut disimpan sebagai model terbaik melalui mekanisme Model*Checkpoint*, dan pelatihan dihentikan pada *epoch* ke-14 melalui *Earlystopping*. Model kemudian diuji menggunakan 2.118 data uji dan memperoleh test *accuracy* sebesar 69,55%, *macro precision* 57,63%, *macro recall* 56,81%, *macro F1-score* 57,17%, *weighted precision* 69,29%, *weighted recall* 69,55%, dan *weighted F1-score* 69,36%. Performa terbaik CNN diperoleh pada kelas negatif dengan *F1-score* 76,95%, diikuti netral 65,02%, dan positif 29,54%.
Model LSTM juga dilatih hingga maksimum 30 *epoch*. Pada *epoch* pertama, model memperoleh training *accuracy* sebesar 55,77% dan *validation accuracy* sebesar 61,39%, kemudian meningkat hingga training *accuracy* mencapai 96,25% pada *epoch* ke-13. *Validation loss* terendah diperoleh pada *epoch* ke-6 sebesar 0,7633 dengan *validation accuracy* 73,26%, sehingga bobot pada *epoch* tersebut disimpan sebagai model terbaik, dan pelatihan dihentikan pada *epoch* ke-13 melalui *Earlystopping*. Pada tahap pengujian, LSTM memperoleh test *accuracy* sebesar 72,29%, *macro precision* 61,10%, *macro recall* 63,47%, *macro F1-score* 62,07%, *weighted precision* 73,51%, *weighted recall* 72,29%, dan *weighted F1-score* 72,78%. Model LSTM juga menunjukkan performa terbaik pada kelas negatif dengan *F1-score* 79,09%, diikuti netral 69,73%, dan positif 37,40%.
Confusion matrix kedua model disajikan pada Gambar 2 dan Gambar 3. Pada model CNN, dari 1.257 data negatif, 990 diprediksi benar, 178 diprediksi netral, dan 89 diprediksi positif; dari 697 data netral, 435 diprediksi benar; dan dari 164 data positif, hanya 48 diprediksi benar. Pada model LSTM, dari 1.257 data negatif, 961 diprediksi benar; dari 697 data netral, 501 diprediksi benar; dan dari 164 data positif, 69 diprediksi benar. Secara keseluruhan, LSTM menghasilkan 1.531 prediksi benar, sedangkan CNN menghasilkan 1.473 prediksi benar dari total 2.118 data uji, sehingga LSTM unggul sebanyak 58 prediksi benar dibandingkan CNN.



Gambar 2. *Confusion matrix* CNN



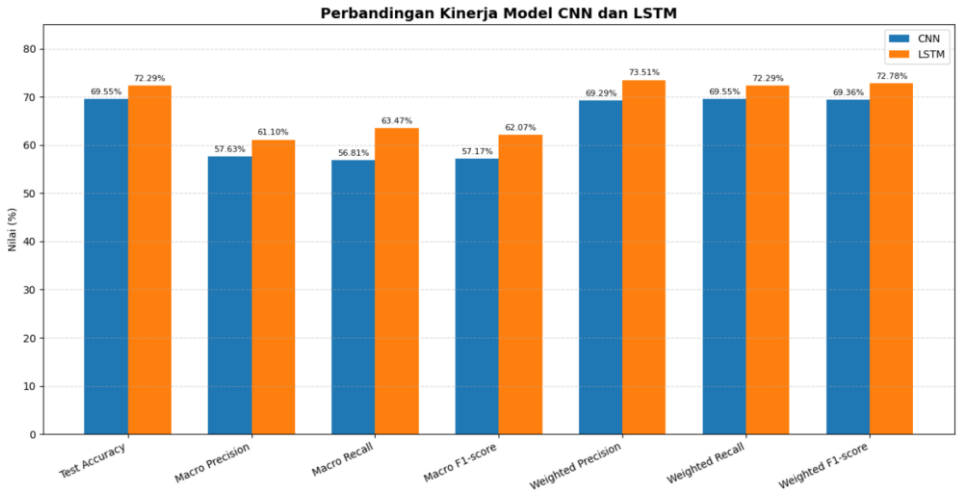
Gambar 3. *Confusion matrix* LSTM

Perbandingan Kinerja Model CNN dan LSTM

Tabel 3 menunjukkan bahwa LSTM unggul pada tujuh dari delapan metrik evaluasi, sedangkan CNN hanya unggul pada nilai test *loss* dengan selisih 0,0339. LSTM memperoleh akurasi 72,29%, lebih tinggi 2,74 poin persentase dibandingkan CNN. Keunggulan terbesar LSTM terlihat pada *macro recall* dengan selisih 6,66 poin persentase, yang menunjukkan LSTM lebih mampu mengenali data pada setiap kelas sentimen secara merata. Perbandingan kinerja kedua model pada seluruh metrik ditampilkan pada Gambar 4.

Tabel 3. Perbandingan Kinerja Model CNN dan LSTM

Metrik	CNN	LSTM
Test *accuracy*	69,55%	72,29%
Test *loss*	0,7196	0,7535
Macro precision	57,63%	61,10%
Macro recall	56,81%	63,47%
Macro F1-score	57,17%	62,07%
Weighted precision	69,29%	73,51%
Weighted recall	69,55%	72,29%
Weighted F1-score	69,36%	72,78%



Gambar 4. Perbandingan Kinerja Model CNN dan LSTM

Selain perbandingan secara keseluruhan, performa kedua model juga dibandingkan pada tiap kelas sentimen berdasarkan nilai *F1-score*, sebagaimana disajikan pada Tabel 4.

Tabel 4. Perbandingan *F1-score* CNN dan LSTM per Kelas Sentimen

Kelas Sentimen	CNN	LSTM	Selisih
Negatif	76,95%	79,09%	2,14 poin
Netral	65,02%	69,73%	4,71 poin
Positif	29,54%	37,40%	7,86 poin

Tabel 4 menunjukkan bahwa LSTM unggul pada seluruh kategori sentimen, dengan selisih terbesar pada kelas positif sebesar 7,86 poin persentase. Kedua model memiliki hasil terbaik dalam mengenali sentimen negatif dan hasil terendah pada sentimen positif, yang mengindikasikan bahwa sentimen positif merupakan kelas paling sulit dikenali oleh kedua arsitektur, kemungkinan besar karena jumlah datanya jauh lebih sedikit dibandingkan kelas negatif dan netral.

Analisis Data

Ketidakseimbangan jumlah data antarkelas turut memengaruhi kemampuan kedua model dalam mengenali setiap kategori sentimen. Jumlah data negatif yang jauh lebih banyak memberikan lebih banyak variasi pola pembelajaran kepada model, sehingga kelas ini dapat dikenali dengan lebih baik, sedangkan jumlah data positif yang jauh lebih sedikit menyebabkan variasi pola bahasa pada kelas tersebut lebih terbatas. Dalam kondisi data yang tidak seimbang, nilai *accuracy* belum sepenuhnya menggambarkan kemampuan model secara merata pada setiap kelas. Oleh karena itu, nilai *macro precision*, *macro recall*, dan *macro F1-score* digunakan untuk melihat kemampuan model mengklasifikasikan ketiga kelas secara seimbang, sedangkan nilai *weighted precision*, *weighted recall*,

dan *weighted F1-score* digunakan untuk mengevaluasi performa model dengan mempertimbangkan proporsi jumlah data pada masing-masing kelas. Pada penelitian ini, nilai *weighted* yang lebih tinggi dibandingkan nilai *macro* pada kedua model turut menunjukkan bahwa performa model lebih banyak dipengaruhi oleh kelas dengan jumlah data yang lebih besar, yaitu kelas negatif.

Pembahasan

Model CNN digunakan untuk mengklasifikasikan sentimen pada data teks dengan memanfaatkan fitur yang telah diekstraksi, sehingga model dapat mempelajari pola dalam data untuk membedakan kategori sentimen (Tukino, 2025). Penerapan CNN pada teks berbahasa Indonesia oleh (Handoko et al., 2024) menunjukkan bahwa arsitektur ini relevan untuk mengekstraksi fitur lokal pada teks media sosial. Namun, hasil penelitian ini menunjukkan bahwa CNN masih memiliki keterbatasan dalam memahami ketergantungan antarkata yang letaknya berjauhan dalam sebuah *tweet*, sehingga performanya pada kelas positif dengan variasi bahasa yang lebih beragam relatif rendah.

Sebaliknya, LSTM dirancang dengan sel memori serta tiga mekanisme gerbang, yaitu *forget gate*, *input gate*, dan *output gate*, yang membantu model mempertahankan informasi penting dalam urutan data yang panjang (Pipin & Kurniawan, 2022). Kemampuan tersebut sejalan dengan hasil penelitian ini, di mana LSTM menunjukkan performa lebih baik pada seluruh kategori sentimen dibandingkan CNN. Temuan ini juga konsisten dengan penelitian (Ahmad et al., 2023) yang menunjukkan LSTM lebih unggul secara keseluruhan dibandingkan CNN pada analisis sentimen ulasan produk, meskipun CNN memiliki keunggulan pada indikator tertentu seperti *precision* atau *loss*. (Pipin & Kurniawan, 2022) juga menunjukkan bahwa LSTM efektif untuk klasifikasi multikelas pada teks media sosial berbahasa Indonesia.

Kedua model menunjukkan performa terendah pada kelas sentimen positif, yang diduga dipengaruhi oleh jumlah data positif yang jauh lebih sedikit dibandingkan kelas negatif dan netral, sehingga model memiliki keterbatasan dalam mempelajari variasi pola bahasa pada kelas tersebut. Dominasi sentimen negatif pada hasil penelitian ini (6.281 *tweet*) sejalan dengan temuan (Hanan et al., 2025) yang juga menemukan dominasi tanggapan negatif dalam analisis sentimen pengguna ChatGPT, namun berbeda dengan temuan (Miyazaki et al., 2024) yang melaporkan sentimen publik terhadap *Generative* AI secara umum cenderung positif. Perbedaan tersebut dapat dipengaruhi oleh ruang lingkup topik, bahasa, periode pengumpulan data, dan konteks penggunaan teknologi yang dianalisis.

Hasil penelitian ini berbeda dengan penelitian (Handoko et al., 2024) mengenai sentimen pengguna Twitter terhadap sistem Sirekap, di mana CNN menunjukkan performa lebih baik dibandingkan pendekatan hibrida CNN-LSTM. Penelitian tersebut menerapkan tahapan *preprocessing* yang lebih lengkap, teknik SMOTE untuk menangani ketidakseimbangan kelas, serta representasi kata yang telah dilatih sebelumnya. Perbedaan ini menunjukkan bahwa kinerja CNN dan LSTM tidak hanya ditentukan oleh karakteristik arsitektur, tetapi juga dipengaruhi oleh topik data, distribusi kelas, teknik representasi teks, tahapan *preprocessing*, dan pengaturan parameter model. Berbeda dengan pendekatan hibrida yang digunakan (Mohbey et al., 2022), penelitian ini membangun CNN dan LSTM secara terpisah agar performa keduanya dapat dibandingkan secara langsung menggunakan data latih, data uji, dan metrik evaluasi yang sama, sehingga penggabungan kedua arsitektur dapat menjadi pertimbangan pada penelitian selanjutnya. Secara keseluruhan, hasil penelitian menunjukkan bahwa karakteristik data, keseimbangan kelas, dan arsitektur model perlu dipertimbangkan dalam menentukan metode klasifikasi sentimen yang paling sesuai.

SIMPULAN

Penelitian ini menganalisis sentimen pengguna Twitter (X) terhadap ChatGPT menggunakan model CNN dan LSTM. Proses pengumpulan data menghasilkan 11.210 *tweet*, yang berkurang menjadi 10.586 *tweet* setelah proses *cleansing*. Hasil pelabelan menunjukkan dominasi sentimen negatif (6.281 *tweet*), diikuti netral (3.484 *tweet*) dan positif (821 *tweet*). Data dibagi dengan rasio 80:20 menjadi 8.468 data latih dan 2.118 data uji untuk melatih dan menguji kedua model.

Model CNN memperoleh akurasi 69,55%, *macro F1-score* 57,17%, dan *weighted F1-score* 69,36%, sedangkan model LSTM memperoleh akurasi 72,29%, *macro F1-score* 62,07%, dan *weighted F1-score* 72,78%, dengan 1.531 prediksi benar dibandingkan 1.473 pada CNN. Kedua model menunjukkan performa terbaik pada kelas negatif dan performa terendah pada kelas positif akibat ketidakseimbangan

jumlah data antarkelas. Dengan demikian, LSTM terbukti lebih sesuai digunakan dibandingkan CNN untuk klasifikasi sentimen pada data *tweet* mengenai ChatGPT karena kemampuannya memahami hubungan dan urutan kata secara lebih baik. Penelitian selanjutnya disarankan menggunakan data yang lebih seimbang, penanganan ketidakseimbangan kelas seperti SMOTE atau *oversampling*, optimasi parameter model, serta eksplorasi arsitektur lain seperti BiLSTM, GRU, atau IndoBERT untuk meningkatkan performa klasifikasi.

DAFTAR PUSTAKA

Ahmad, S., Ridwan, A. M., & Setyawan, G. D. (2023). Analisis Sentimen Product Tools & Home Menggunakan Metode CNN Dan LSTM. *Teknokom*, *6*(2), 133–140. https://doi.org/10.31943/teknokom.v6i2.154

Alfin, A. A., Kurniasari, L., & Yanuartanti, I. (2023). INFORMASI (Jurnal Informatika dan Sistem Informasi) Volume 15 No.1 / Mei / 2023. *INFORMASI (Jurnal Informatika Dan Sistem Informasi)*, *15*(1), 32–45.

Appiah, E., & Htait, A. (2024). AI in the Public Eye: Analysing Social Media Sentiment and Opinion on Artificial Intelligence. *Proceedings of Machine Learning Research*, *295*, 1–15.

Ayuningtyas, P., Khomsah, S., & Sudianto, S. (2024). Pelabelan Sentimen Berbasis Semi-Supervised Learning menggunakan Algoritma LSTM dan GRU. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *9*(3), 217–229. https://doi.org/10.14421/jiska.2024.9.3.217-229

Hanan, D. A., Husodo, A. Y., & Rassy, R. P. (2025). Sentiment Study of ChatGPT on Twitter Data with Hybrid K-Means and LSTM. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *24*(2), 273–284. https://doi.org/10.30812/matrik.v24i2.4791

Handoko, H., Asrofiq, A., Junadhi, J., & Negara, A. S. (2024). Sentiment Analysis of Sirekap Tweets Using CNN Algorithm. *INTENSIF: Jurnal Ilmiah Penelitian Dan Penerapan Teknologi Sistem Informasi*, *8*(2), 312–329. https://doi.org/10.29407/intensif.v8i2.23046

Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(2), 121. https://doi.org/10.22146/ijccs.65176

Joloudari, J. H., Hussain, S., Nematollahi, M. A., Bagheri, R., Fazl, F., Alizadehsani, R., Lashgari, R., & Talukder, A. (2023). *BERT-Deep CNN: State-of-the-Art for Sentiment Analysis of COVID-19 Tweets*. *2*, 306–312.

Juwiantho, H., Setiawan, E. I., Santoso, J., Purnomo, M. H., Informasi, D. T., Tinggi, S., Surabaya, T., Elektro, D. T., Elektro, F. T., Komputer, D. T., Elektro, F. T., Korespondensi, P., Convolutional, D., & Network, N. (2020). *Sentiment Analysis Twitter Bahasa Indonesia Berbasis Word2Vec Menggunakan Deep Convolutional Neural Network*. *7*(1), 181–188. https://doi.org/10.25126/jtiik.202071758

Miyazaki, K., Murayama, T., Uchiba, T., An, J., & Kwak, H. (2024). *Public perception of generative AI on Twitter : an empirical study based on occupation and usage*. https://doi.org/10.1140/epjds/s13688-023-00445-y

Mohbey, K. K., Meena, G., Kumar, S., & Lokesh, K. (2022). *A CNN-LSTM-based hybrid deep learning approach to detect sentiment polarities on Monkeypox tweets*. 1–11. http://arxiv.org/abs/2208.12019

Muammar Khadapi, & Pakpahan, V. M. (2024). Analisis Sentimen Berbasis Jaringan LSTM dan BERT terhadap Diskusi Twitter tentang Pemilu 2024. *JUKI : Jurnal Komputer Dan Informatika*, *6*(2), 130–137. https://doi.org/10.53842/juki.v6i2.681

Mudding, A. A., & Arifin A Abd Karim. (2022). Analisis Sentimen Menggunakan Algoritma Lstm Pada Media Sosial. *Jurnal Publikasi Ilmu Komputer Dan Multimedia*, *1*(3), 181–187. https://doi.org/10.55606/jupikom.v1i3.517

Pipin, S. J., & Kurniawan, H. (2022). *Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat di Twitter Menggunakan LSTM*. *23*(2), 197–208.

Prasetyo, A. P., Sri Kusuma, C., & Rizki, D. (2024). Analisis Sentimen Twitter terhadap Isu Penundaan Pemilu 2024 Menggunakan Metode Long Short-Term Memory (LSTM). *Jurnal Repositor*, *6*(3), 293–304. https://doi.org/10.22219/repositor.v6i3.32643

Sutejo, B. S. I., Aryani, D., & Asianto, A. (2025). Pemodelan Pengaruh Risiko dengan Pendekatan Centrality dan Graph-based Learning Approach pada Transformasi Digital Kesehatan. *JUKI : Jurnal Komputer Dan Informatika*, *7*(2), 249–258. https://doi.org/10.53842/juki.v7i2.1992

Tukino, T. (2025). Penerapan Algoritma Convolutional Neural Network Untuk Klasifikasi Sentimen Pada Layanan e-Commerce. *Jurnal Desain Dan Analisis Teknologi*, *4*(1), 44–53. https://doi.org/10.58520/jddat.v4i1.72

Wardianto, Farikhin, D. M. K. N. (2023). *Analisis Sentimen Berbasis Aspek Ulasan Pelanggan Restoran Menggunakan LSTM Dengan Adam Optimizer*. *7*(1), 67–76.

Zakharia, A., Hasanah, H., & Srirahayu, A. (2025). Analisis Sentimen Publik di Twitter tentang Pemanfaatan AI dalam Seni Digital Menggunakan CNN. *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(3), 536–545.