ANALISIS SENTIMEN TERHADAP ULASAN ONLINE UNTUK KENDARAAN LISTRIK MENGGUNAKAN DEEP LEARNING DI PLATFORM X

Fatih Al Jabar H.M., Bagus Hardiansyah
Teknik Informatika, Universitas 17 Agustus 1945 Surabaya
Jalan Semolowaru No. 45 Surabaya, Indonesia
fatihaljabar@gmail.com

ABSTRAK

Meningkatnya perhatian global terhadap kendaraan listrik telah memicu diskursus publik yang masif di platform X, sehingga diperlukan metode analisis sentimen yang akurat untuk memahami persepsi masyarakat. Penelitian ini bertujuan mengevaluasi dan membandingkan kinerja empat model *Deep Learning*, yakni *Convolutional Neural Network* (CNN), *Long Short-Term Memory* (LSTM), *Bidirectional* LSTM (BiLSTM), dan model hibrida CNN-BiLSTM dengan mekanisme *Attention*. Metodologi penelitian melibatkan pengumpulan dari tweet berbahasa Indonesia periode Januari 2023 hingga Agustus 2025 yang dikumpulkan melalui *web-scraping* dan dilabeli menggunakan skor leksikon berbasis kuantil. Tahapan *preprocessing* meliputi *case folding, noise removal*, serta *stemming* menggunakan pustaka Sastrawi untuk menjamin kualitas data input. Hasil eksperimen menunjukkan bahwa model hibrida CNN-BiLSTM-Attention mencapai kinerja paling unggul dengan akurasi 75% dan *macro F1-score* 0,76. Model BiLSTM dan CNN masing-masing mencatat akurasi 73% dan 72%, sementara model LSTM standar mengalami kegagalan konvergensi (*model collapse*) karena ketidakmampuannya menangani kompleksitas data. Sinergi antara ekstraksi fitur lokal CNN dan pemahaman konteks sekuensial dua arah dari BiLSTM terbukti secara signifikan meningkatkan kemampuan klasifikasi, meskipun tantangan utama masih terletak pada ambiguitas antara kelas positif dan netral.

Kata kunci : *Analisis Sentimen, Deep Learning, CNN, BiLSTM, Kendaraan Listrik, Platform X*

1. PENDAHULUAN

Eskalasi perhatian global terhadap teknologi kendaraan listrik (electric vehicles/EV) telah memicu dinamika diskusi publik yang signifikan di berbagai platform digital, sebuah fenomena yang berjalan seiring dengan meningkatnya kesadaran kolektif mengenai isu lingkungan dan urgensi transisi menuju energi terbarukan [1]. Eskalasi perhatian global terhadap teknologi kendaraan listrik (electric vehicles/EV) telah memicu dinamika diskusi publik yang signifikan di berbagai platform digital, sebuah fenomena yang berjalan seiring dengan meningkatnya kesadaran kolektif mengenai isu lingkungan dan urgensi transisi menuju energi terbarukan [2]. Kondisi ini memunculkan permasalahan mendasar mengenai bagaimana mengolah data tekstual yang berlimpah namun mentah tersebut menjadi wawasan strategis yang dapat ditindaklanjuti oleh para pemangku kepentingan.

Guna menjawab tantangan pengolahan data tersebut, analisis sentimen hadir sebagai instrumen krusial dalam ranah Natural Language Processing (NLP) yang berfungsi untuk mengidentifikasi dan mengekstraksi orientasi pandangan masyarakat terhadap suatu entitas[2]. Meskipun pendekatan Machine Learning tradisional pernah menjadi standar, perkembangannya kini mulai bergeser ke arah penerapan metode Deep Learning karena keterbatasan model tradisional dalam menangani dimensi data yang besar. Metode Deep Learning menawarkan kapabilitas yang jauh lebih unggul dalam menangkap pola non-linear yang kompleks serta mengekstraksi fitur kontekstual secara otomatis dari korpus teks yang masif tanpa memerlukan rekayasa fitur manual yang rumit. Salah satu arsitektur yang terbukti efektif adalah Convolutional Neural Network (CNN), yang memiliki keunggulan dalam mengenali pola lokal seperti frasa kunci atau n-gram yang berkontribusi signifikan terhadap makna sentimen [3]. Di sisi lain, untuk menangani aspek sekuensial bahasa, model Long Short-Term Memory (LSTM) beserta variannya seperti Bidirectional LSTM (BiLSTM) menawarkan kemampuan untuk memahami dependensi jangka panjang dan konteks kalimat secara mendalam dengan memperhatikan urutan kata dari dua arah [1].

Meskipun penerapan model tunggal seperti CNN atau LSTM telah banyak dilakukan, kompleksitas bahasa alami di media sosial sering kali memerlukan pendekatan yang lebih canggih untuk mencapai akurasi optimal. Pengintegrasian berbagai arsitektur menjadi sebuah model hybrid menawarkan potensi solusi untuk menutupi kelemahan masing-masing model tunggal. Secara spesifik, penggabungan kemampuan ekstraksi fitur lokal dari CNN dengan pemahaman konteks sekuensial dari BiLSTM, yang diperkuat dengan mekanisme Attention, memungkinkan model untuk tidak hanya memahami konteks kalimat tetapi juga secara dinamis memfokuskan bobot komputasi pada bagian teks yang paling relevan terhadap keputusan klasifikasi [4]. Penelitian terdahulu memang telah mengeksplorasi efektivitas model-model ini secara terpisah, namun studi komparatif yang secara spesifik membandingkan performa antara arsitektur CNN, LSTM, dan model hybrid CNN-BiLSTM-Attention pada domain spesifik kendaraan listrik di Indonesia masih sangat terbatas

[5]. Hal ini mengindikasikan adanya celah penelitian yang perlu diisi untuk menentukan arsitektur mana yang paling tangguh dalam menangani karakteristik unik bahasa ulasan di media sosial Indonesia yang kerap kali bersifat informal dan ambigu..

Berdasarkan uraian permasalahan dan peluang tersebut, penelitian ini bertujuan untuk mengevaluasi dan membandingkan secara komprehensif kinerja tiga arsitektur Deep Learning, yaitu CNN, LSTM, dan model hybrid CNN-BiLSTM-Attention dalam mengklasifikasikan sentimen publik ke dalam kategori positif, negatif, dan netral. Penelitian ini dirancang untuk menguji hipotesis mengenai efektivitas penggabungan fitur spasial dan temporal dalam analisis teks melalui serangkaian evaluasi menggunakan metrik standar seperti akurasi, presisi, recall, dan F1-score. Hasil dari penelitian ini diharapkan tidak hanya memberikan kontribusi teoritis bagi pengembangan metodologi analisis opini berbasis Deep Learning, tetapi juga menghasilkan implikasi praktis berupa masukan strategis bagi industri otomotif nasional untuk meningkatkan kualitas produk dan layanan berdasarkan pemahaman yang akurat terhadap persepsi nyata konsumen di media sosial.

2. TINJAUAN PUSTAKA

Analisis sentimen dalam ranah Natural Language Processing (NLP) didefinisikan sebagai proses identifikasi otomatis terhadap dataset berbasis teks guna mengekstraksi pandangan atau opini terhadap suatu aspek tertentu [2]. Implementasi teknologi ini telah meluas secara signifikan di berbagai sektor strategis, mulai dari e-commerce hingga transportasi, untuk mendukung pengambilan keputusan berbasis data melalui pemahaman mendalam terhadap kebutuhan dan preferensi pengguna [6].

Efektivitas model pembelajaran mendalam sangat bergantung pada kualitas data masukan, sehingga tahap preprocessing text menjadi krusial untuk mentransformasi data mentah platform X yang tidak terstruktur dan sarat akan gangguan menjadi format yang konsisten. Proses ini diawali dengan case folding untuk penyeragaman huruf, pembersihan teks menggunakan ekspresi reguler guna mereduksi noise seperti URL, mention, dan simbol non-alfanumerik, serta diakhiri dengan stemming menggunakan library Sastrawi untuk mereduksi kata berimbuhan menjadi bentuk dasarnya [2].

Dalam melakukan ekstraksi fitur dan klasifikasi, arsitektur Convolutional Neural Network (CNN) diterapkan karena kemampuannya dalam mengenali pola lokal atau frasa kunci (n-gram) melalui operasi konvolusi dan pooling [3]. CNN mampu memindai matriks representasi kata secara hierarkis untuk menangkap fitur-fitur dominan yang berkontribusi langsung terhadap makna sentimen kalimat [7]. Namun, untuk menangkap dependensi jangka panjang yang sering ditemukan dalam teks sekuensial, arsitektur Long Short-Term Memory (LSTM) digunakan sebagai solusi atas permasalahan vanishing gradient pada jaringan saraf berulang standar [1]. LSTM memanfaatkan mekanisme gerbang (gates) untuk mengatur aliran informasi, sehingga model dapat mempertahankan konteks informasi dari langkah waktu sebelumnya secara lebih efektif.

Integrasi model yang lebih lanjut diwujudkan melalui penggunaan Bidirectional LSTM (BiLSTM) yang dilengkapi dengan Attention Mechanism guna menangkap nuansa emosional dari dua arah, yakni konteks sebelum dan sesudah suatu kata [4]. Penambahan lapisan Attention memungkinkan model untuk memberikan bobot atau skor perhatian yang berbeda pada setiap kata, sehingga fokus klasifikasi tertuju pada bagian teks yang paling informatif terhadap keputusan sentimen akhir [4].

Proses pelabelan data dalam penelitian ini didukung oleh metode lexicon yang menentukan orientasi sentimen berdasarkan perhitungan skor dari kamus kata positif dan negatif [8]. Untuk mengevaluasi kinerja arsitektur yang dikembangkan, digunakan instrumen confusion matrix yang membandingkan hasil prediksi sistem dengan data aktual secara objektif [9]. Evaluasi dilakukan melalui metrik akurasi, presisi, recall, dan F1-score guna memastikan model memiliki kemampuan generalisasi yang stabil dalam membedakan kelas sentimen positif, negatif, dan netral [2], [9].

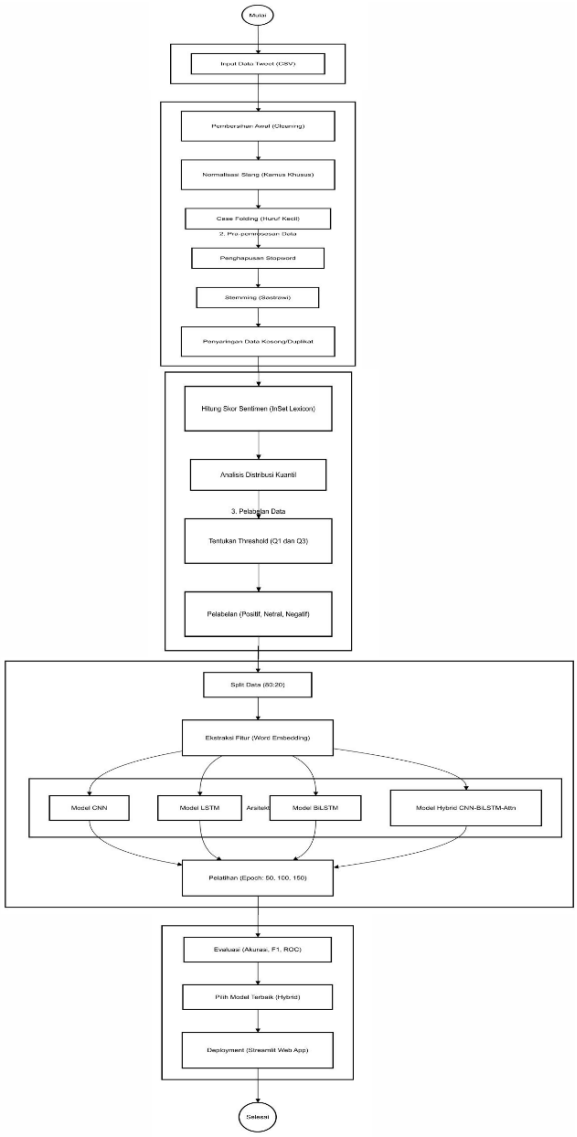
3. METODE PENELITIAN

3.1. Data dan Perangkat Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan data tekstual berupa opini publik mengenai kendaraan listrik yang bersumber dari platform media sosial X (sebelumnya Twitter). Data dikumpulkan menggunakan teknik *scraping* dengan kata kunci "mobil listrik" dalam rentang waktu 1 Januari 2023 hingga 30 Agustus 2025. Pemilihan rentang waktu ini bertujuan untuk menangkap dinamika opini terkini seiring dengan peningkatan adopsi kendaraan listrik di Indonesia. Data yang diambil bersifat publik dan diolah tanpa melibatkan interaksi langsung dengan pengguna untuk menjaga privasi subjek data.

Proses komputasi dan analisis data dilakukan menggunakan perangkat lunak berbasis bahasa pemrograman Python. Lingkungan pengembangan utama memanfaatkan Jupyter Notebook dengan dukungan pustaka analisis data seperti Pandas dan NumPy, serta pustaka *Natural Language Processing* (NLP) seperti NLTK dan Sastrawi untuk pemrosesan teks Bahasa Indonesia. Implementasi model *Deep Learning* dibangun menggunakan kerangka kerja Keras/TensorFlow, sementara visualisasi data menggunakan Matplotlib dan Seaborn. Spesifikasi perangkat keras yang digunakan mencakup prosesor Intel Core i5 Gen 11 dan RAM 8 GB untuk memastikan efisiensi selama proses pelatihan model.

3.2. Prosedur Penelitian



Gambar 1. Alur Tahapan Penelitian

Alur penelitian dirancang secara sistematis mulai dari akuisisi data hingga evaluasi model. Tahapan utama meliputi pengumpulan data, pra-pemrosesan (*preprocessing*), pelabelan data otomatis, pembagian data (*data splitting*), pembangunan arsitektur model, pelatihan, dan evaluasi kinerja. Setiap tahapan dirancang untuk memastikan kualitas data input yang optimal bagi model *Deep Learning* yang dikembangkan.

3.3. Pra-pemrosesan Data (*Data Preprocessing*)

Data mentah dari media sosial memiliki karakteristik tidak terstruktur dan mengandung banyak *noise*. Oleh karena itu, serangkaian tahapan pra-pemrosesan diterapkan untuk meningkatkan kualitas data sebelum masuk ke tahap pemodelan. Proses ini dimulai dengan *cleaning*, yaitu penghapusan elemen non-tekstual seperti URL, *mention*, tagar, emoji, dan karakter non-alfanumerik. Selanjutnya, dilakukan normalisasi kata slang menggunakan kamus khusus (*update_combined_slang_words.txt*) untuk mengubah istilah tidak baku menjadi bentuk baku yang sesuai dengan kaidah bahasa Indonesia.

Tahap berikutnya adalah *case folding* untuk menyeragamkan seluruh teks menjadi huruf kecil, diikuti oleh proses *stemming* menggunakan pustaka Sastrawi [6]. *Stemming* bertujuan mereduksi variasi morfologis kata dengan mengubah kata berimbuhan menjadi kata dasarnya, sehingga mengurangi dimensi fitur yang harus dipelajari model. Langkah terakhir adalah penyaringan data (*filtering*) untuk menghapus duplikasi dan data kosong yang mungkin terbentuk setelah proses pembersihan. Contoh transformasi data selama proses ini dapat dilihat pada Tabel 1.

Tabel 1. Contoh Transformasi Data pada Tahap Pra-pemrosesan

Tahapan	**Teks Input (*Raw* / *Intermediate*)**	**Hasil Pemrosesan (*Output*)**
Case Folding	@pepekoko3 @nguyen_mt Dunia akan berakselarasi dalam kecepatan penuh ke multipolaritas bila kekaisaran DS MIC berubah jadi relasi win win deal ala Dasasila Bandung. Yang kini menjadi realitas geopol via B&RI AIIB Beijing. Jakarta salah 1 bukti empirisnya. Monopoli mobil Jepang tergerus Wuling BYD.	dunia akan berakselarasi dalam kecepatan penuh ke multipolaritas bila kekaisaran ds mic berubah jadi relasi win win deal ala dasasila bandung yang kini menjadi realitas geopol via b amp ri aiib beijing jakarta salah bukti empirisnya monopoli mobil jepang tergerus wuling byd
Stemming	*Output dari Case Folding di atas*	dunia berakselarasi cepat penuh multipolaritas kaisar ds mic ubah relasi menang menang deal ala dasasila bandung realitas geopol via bb amp ri aiib beijing jakarta salah bukti empiris monopoli mobil jepang gerus wuling byd
Case Folding	manifesting bisa beli mobil byd sendiri	manifesting bisa beli mobil byd sendiri
Stemming	*Output dari Case Folding di atas*	manifesting beli mobil byd sendiri
Case Folding	Orang orang lahiran dikasih hadiah sama suaminya ada yang dikasih mobil dikasih hp.. bismillah BYD Seal atau Madza CX-3 (gak mungut mungut yaAllah)	orang orang lahiran dikasih hadiah sama suaminya ada yang dikasih mobil dikasih hp bismillah byd seal atau madza cx gak mungut mungut yaallah

Tahapan	Teks Input (*Raw* / *Intermediate*)	Hasil Pemrosesan (*Output*)
Stemming	*Output dari Case Folding di atas*	orang orang lahir kasih hadiah suami kasih mobil kasih hp bismillah byd seal madza cx mungut mungut yaallah
Case Folding	BOCORAN PERUSAK HARGA Wuling Starlight bakal masuk Indo Batere 50-70 kWh harga sepertiga-setengah BYD Seal (~250-350) Cakep gak? === LEAKED Wuling Starlight is about to enter the SE Asian market Battery: 50-70 kWh Priced at 1/3 of BYD Seal How does it look? https://t.co/2vSgQDGC3t	bocoran perusak harga wuling starlight bakal masuk indo batere kwh harga sepertiga setengah byd seal cakep gak leaked wuling starlight is about to enter the se asian market battery kwh priced at of byd seal how does it look
Stemming	*Output dari Case Folding di atas*	bocor usak harga wuling starlight masuk indonesia batere kwh harga tiga byd seal cakep leaked wuling starlight is about enter the asi market battery kwh priced at of byd seal how does it look
Case Folding	PERANG BINTANG BERLANJUT BYD bikin perang harga di China. Saham² produsen mobil di China pun rontok. Ini menunjukkan pertumpahan darah di akhir tahun ini yg bisa mjd (efek) domino kata Tu Le Direktur Pelaksana Sino Auto Insights. Sekarang Evergrande di industri otomotif https://t.co/vmkj6DVnGW	perang bintang berlanjut byd bikin perang harga di china saham² produsen mobil di china pun rontok ini menunjukkan pertumpahan darah di akhir tahun ini yg bisa mjd efek domino kata tu le direktur pelaksana sino auto insights sekarang evergrande di industri otomotif
Stemming	*Output dari Case Folding di atas*	perang bintang lanjut byd bikin perang harga china saham produsen mobil china rontok tumpah darah efek domino le direktur laksana sino auto insights evergrande industri otomotif

3.4. Pelabelan Data Berbasis Leksikon (*Lexicon-Based Labeling*)

Mengingat volume data yang besar, proses pelabelan dilakukan secara otomatis menggunakan metode *Lexicon-Based Sentiment Analysis*. Penelitian ini menggunakan kamus leksikon InSet yang berisi daftar kata dengan bobot sentimen positif dan negatif. Berbeda dengan pendekatan leksikon konvensional yang kaku, penelitian ini menerapkan metode *Quantile-based Labeling* untuk menentukan kelas sentimen (Positif, Negatif, Netral).

Dalam metode ini, skor sentimen setiap *tweet* dihitung berdasarkan akumulasi bobot kata yang terdeteksi dalam kamus. Distribusi skor seluruh dataset kemudian dianalisis untuk menentukan ambang batas (*threshold*) yang dinamis menggunakan nilai kuantil (Q1 dan Q3). Tweet dengan skor di bawah Q1 dikategorikan sebagai Negatif, skor di atas Q3 sebagai Positif, dan skor di antara kedua nilai tersebut sebagai Netral. Pendekatan adaptif ini bertujuan untuk menghasilkan pembagian kelas yang lebih seimbang dan representatif terhadap karakteristik distribusi data aktual [8].

3.5. Skenario Pembagian Data dan Representasi Fitur

Dataset yang telah dilabeli dibagi menjadi tiga subset proporsional: 70% sebagai data latih (*training set*), 15% sebagai data validasi (*validation set*), dan 15% sebagai data uji (*testing set*). Pembagian ini dilakukan secara acak namun terkontrol (*stratified*) untuk menjaga keseimbangan distribusi kelas sentimen di setiap subset.

Untuk representasi fitur, teks dikonversi menjadi vektor numerik menggunakan teknik *Word Embeddings*. Setiap kalimat diproses dengan panjang sekuens yang seragam, yaitu 128 token, menggunakan teknik *padding*. Lapisan *embedding* ini memungkinkan model mempelajari representasi semantik kata dalam ruang vektor berdimensi tinggi, di mana kata-kata dengan makna serupa akan memiliki representasi vektor yang berdekatan.

3.6. Arsitektur Model Deep Learning

Penelitian ini mengembangkan dan membandingkan tiga arsitektur *Deep Learning* untuk klasifikasi sentimen, yaitu *Convolutional Neural Network* (CNN), *Bidirectional Long Short-Term Memory* (BiLSTM), dan model hibrida CNN-BiLSTM-Attention.

a. Convolutional Neural Network (CNN)
   Model CNN dirancang untuk mengekstraksi fitur lokal (n-gram) yang signifikan dari teks. Arsitektur ini menggunakan lapisan *Conv1D* yang disusun secara hierarkis untuk menangkap pola kata atau frasa pendek. Fitur hasil konvolusi kemudian direduksi dimensinya menggunakan lapisan *MaxPooling1D* dan *GlobalMaxPooling1D* untuk mengambil fitur paling dominan sebelum diklasifikasikan oleh lapisan *Dense* [3].
b. Bidirectional LSTM (BiLSTM)
   Model BiLSTM digunakan untuk menangkap dependensi jangka panjang dan konteks kalimat secara menyeluruh. Berbeda dengan LSTM standar, BiLSTM memproses urutan input dari dua arah: maju (*forward*) dan mundur (*backward*). Mekanisme ini memungkinkan model memahami konteks kata berdasarkan informasi yang mendahului dan mengikutinya, sehingga

menghasilkan representasi semantik yang lebih kaya [4].

c. Hybrid CNN-BiLSTM-Attention
Model hibrida ini menggabungkan keunggulan ekstraksi fitur lokal dari CNN dan pemahaman konteks sekuensial dari BiLSTM, diperkuat dengan mekanisme *Attention*. Arsitektur model hibrida, seperti dirinci pada Tabel 2, dimulai dengan lapisan CNN untuk mengekstraksi fitur spasial, diikuti oleh lapisan BiLSTM untuk memodelkan urutan waktu. Output dari BiLSTM kemudian diproses oleh lapisan *Attention* yang memberikan bobot lebih besar pada kata-kata yang paling relevan terhadap sentimen, sehingga meningkatkan fokus model pada informasi krusial [10].

Tabel 2. Konfigurasi Arsitektur Model Hybrid CNN-BiLSTM-Attention

Layer Type	Input Shape	Output Shape	Jumlah Parameter	Fungsi Utama
InputLayer	(None, 128)	(None, 128)	0	Menerima input sekuens token dengan panjang tetap (128 kata).
Embedding	(None, 128)	(None, 128, 128)	3,840,000	Mengubah token kata menjadi representasi vektor numerik berdimensi 128.
Conv1D (Layer 1)	(None, 128, 128)	(None, 128, 64)	24,640	Ekstraksi fitur n-gram lokal/pola awal dari vektor kata.
Conv1D (Layer 2)	(None, 128, 64)	(None, 128, 64)	12,352	Memperdalam ekstraksi fitur dari hasil konvolusi sebelumnya.
MaxPooling1D (1)	(None, 128, 64)	(None, 64, 64)	0	Mereduksi dimensi fitur (downsampling) untuk mengurangi beban komputasi.
Conv1D (Layer 3)	(None, 64, 64)	(None, 64, 64)	12,352	Ekstraksi fitur tingkat lanjut pada representasi yang lebih padat.
Conv1D (Layer 4)	(None, 64, 64)	(None, 64, 64)	12,352	Ekstraksi fitur tingkat lanjut tahap kedua.
MaxPooling1D (2)	(None, 64, 64)	(None, 32, 64)	0	Reduksi dimensi lanjutan sebelum masuk ke lapisan RNN.
Bidirectional (LSTM)	(None, 32, 64)	(None, 32, 256)	197,632	Mempelajari dependensi konteks kalimat dari dua arah (maju & mundur).
Attention Layer	(None, 32, 256)	(None, 256)	256	Memberikan bobot pada fitur/kata yang paling relevan terhadap sentimen.
Dense (ReLU)	(None, 256)	(None, 64)	16,448	Transformasi fitur non-linier tingkat tinggi.
Dropout	(None, 64)	(None, 64)	0	Menonaktifkan sebagian neuron secara acak untuk mencegah *overfitting*.
Dense (Output)	(None, 64)	(None, 3)	195	Menghasilkan probabilitas klasifikasi (Positif, Negatif, Netral) menggunakan Softmax.
Total Parameter	-	-	4,116,227	Total bobot yang dapat dilatih (*trainable parameters*).

3.7. Pelatihan dan Evaluasi Model

Proses pelatihan dilakukan menggunakan *optimizer* Adam dan fungsi *loss Categorical Crossentropy*. Untuk mencegah *overfitting*, diterapkan mekanisme *Early Stopping* yang memantau kinerja model pada data validasi dan menghentikan pelatihan jika tidak terjadi peningkatan performa dalam sejumlah *epoch* tertentu. Variasi jumlah *epoch* (50, 100, 150) juga diuji untuk menemukan titik konvergensi optimal.

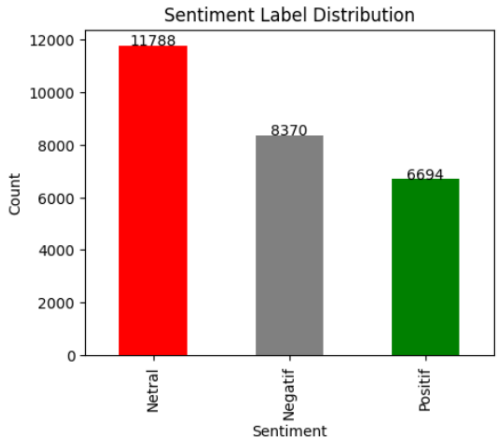
Evaluasi kinerja model dilakukan menggunakan *Confusion Matrix* untuk mengukur metrik Akurasi, Presisi, *Recall*, dan *F1-Score*. Metrik-metrik ini memberikan gambaran komprehensif mengenai kemampuan model dalam mengklasifikasikan setiap kelas sentimen secara tepat. Selain itu, kurva ROC (*Receiver Operating Characteristic*) digunakan untuk menganalisis kemampuan diskriminatif model antar kelas [9]. Model dengan performa terbaik selanjutnya diimplementasikan ke dalam purwarupa aplikasi berbasis web untuk validasi fungsional.

4. HASIL DAN PEMBAHASAN

4.1. Karakteristik Data dan Analisis Eksplorasi

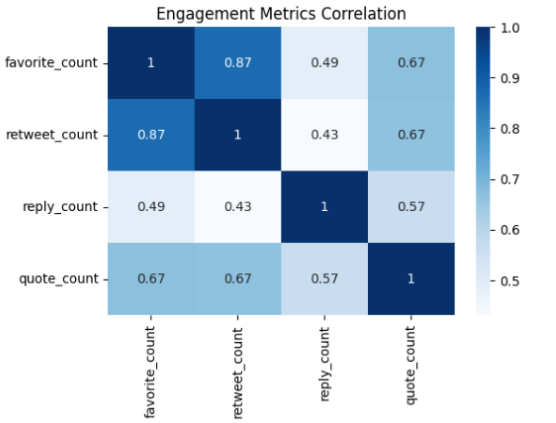
Penelitian ini menggunakan dataset yang dihimpun melalui teknik *web-scraping* pada platform media sosial X (sebelumnya Twitter) dengan kata kunci "mobil listrik". Periode pengambilan data mencakup rentang waktu yang ekstensif, mulai dari 1 Januari 2023 hingga 30 Agustus 2025, menghasilkan korpus data mentah sebanyak hampir 27.000 *tweet*. Data tersebut kemudian melalui tahapan pra-pemrosesan dan pelabelan sentimen menggunakan InSet Lexicon, sebuah kamus leksikon yang dikhususkan untuk sentimen Bahasa Indonesia. Berdasarkan analisis distribusi sentimen, data menunjukkan sebaran yang cukup merata meskipun terdapat dominasi pada kelas Netral. Dominasi ini merupakan fenomena wajar dalam diskursus publik mengenai topik teknologi otomotif, di mana opini subjektif sering bercampur dengan informasi deskriptif umum. Keseimbangan relatif ini diharapkan mampu memfasilitasi model *Deep Learning* untuk

mempelajari karakteristik tiap kelas secara objektif tanpa kecenderungan *overfitting* pada satu polaritas tertentu.



Gambar 2. Bar Chart distribusi banyaknya tiap sentimen pada data

Analisis tren kronologis memperlihatkan adanya lonjakan volume diskusi yang signifikan sejak Mei 2024, mengindikasikan peningkatan atensi publik terhadap ekosistem kendaraan listrik di Indonesia. Meskipun volume diskusi meningkat, tidak ditemukan polarisasi ekstrem; proporsi sentimen tetap seimbang seiring berjalannya waktu. Selain analisis sentimen, eksplorasi terhadap struktur teks menunjukkan bahwa mayoritas *tweet* memiliki panjang 5 hingga 15 kata. Analisis korelasi metrik keterlibatan (*engagement*) melalui *heatmap* menunjukkan hubungan linier yang sangat kuat (87%) antara jumlah *retweet* dan *favorite*. Hal ini mengimplikasikan pola perilaku pengguna di mana interaksi penyebaran informasi (*retweet*) sering kali disertai dengan persetujuan atau penyimpanan konten (*favorite*).



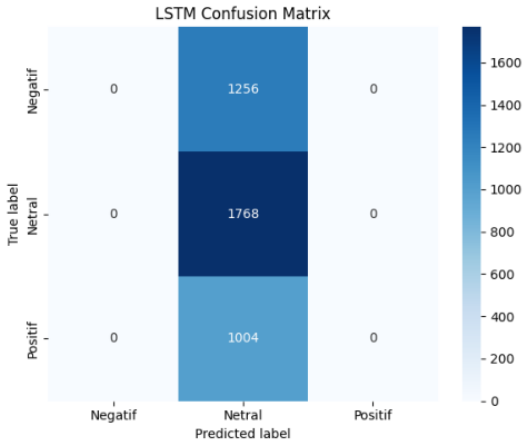
Gambar 3. Korelasi antar Metrik Engagement

4.2. Evaluasi Kinerja Arsitektur CNN dan BiLSTM

Evaluasi kinerja dimulai dengan pengujian model *Convolutional Neural Network* (CNN). Model ini menunjukkan performa yang cukup baik dengan akurasi 72% dan mampu mempelajari pola data seiring penurunan *loss* pelatihan, meskipun terindikasi adanya *overfitting* ringan. Berdasarkan *Classification Report*, CNN menunjukkan keunggulan signifikan dalam mengidentifikasi sentimen Negatif dengan *F1-Score* tertinggi (0.75) dan *Recall* 0.83. Temuan ini sejalan dengan karakteristik CNN yang efektif dalam mengekstraksi fitur lokal atau *n-grams* [3], sehingga frasa negatif yang eksplisit dapat terdeteksi dengan baik. Namun, model ini menghadapi kendala dalam membedakan sentimen Positif dari Netral, di mana terjadi misklasifikasi pada 268 sampel Positif yang diprediksi sebagai Netral, serta presisi terendah pada kelas Netral (0.71).

Sebagai pembanding, model *Bidirectional Long Short-Term Memory* (BiLSTM) diterapkan untuk menangkap konteks sekuensial dua arah. Model ini mencatatkan kinerja yang sedikit lebih unggul dan lebih seimbang dibandingkan CNN, dengan akurasi 73% dan *Macro Avg F1-Score* 0.73. BiLSTM menunjukkan stabilitas yang lebih baik antar kelas, terutama pada kelas Negatif (F1-Score 0.77) dan peningkatan *Recall* pada kelas Positif (0.74). Kemampuan BiLSTM memproses informasi dari arah maju dan mundur memungkinkannya memahami konteks kalimat yang lebih kompleks dibandingkan CNN. Analisis kurva ROC mengonfirmasi bahwa BiLSTM memiliki kemampuan diskriminatif yang sangat baik pada kelas Negatif (AUC 0.922) dan Positif (AUC 0.907), meskipun kelas Netral tetap menjadi tantangan tersulit untuk dipisahkan.

4.3. Analisis Kegagalan Model LSTM Standar



Gambar 4. Confusion Matrix Model LSTM (menunjukkan prediksi menumpuk di Netral)

Berbeda dengan varian Bidirectional, arsitektur LSTM standar (satu arah) mengalami kegagalan total dalam proses pembelajaran, sebuah fenomena yang dikenal sebagai *model collapse*. Grafik pelatihan menunjukkan stagnasi pada akurasi dan *loss* yang tidak menurun, mengindikasikan ketidakmampuan model memperbarui bobot secara efektif. Evaluasi pada data uji menghasilkan akurasi hanya sebesar

43%, angka yang identik dengan proporsi kelas mayoritas (Netral) dalam dataset. Lebih lanjut, model ini menghasilkan *F1-Score* 0.00 untuk kelas Positif dan Negatif, serta nilai AUC 0.500 untuk seluruh kelas, yang setara dengan tebakan acak. Kegagalan ini menegaskan bahwa untuk data tekstual *tweet* Bahasa Indonesia yang kompleks dan tidak terstruktur, arsitektur LSTM satu arah tidak memadai tanpa mekanisme tambahan untuk menangkap dependensi jangka panjang secara efektif [1].

4.4. Efektivitas dan Analisis Komponen Model Hybrid CNN-BiLSTM-Attention

Model *Hybrid* yang mengintegrasikan CNN, BiLSTM, dan mekanisme *Attention* terbukti menjadi arsitektur paling superior dalam penelitian ini, mencapai akurasi tertinggi 75% dan *Macro F1-Score* 0.76. Keunggulan model ini bukan sekadar hasil agregasi, melainkan akibat sinergi hierarkis antar komponennya. Analisis mendalam terhadap mekanisme model menunjukkan peran spesifik tiap lapisan: komponen CNN (Conv1D) bekerja efektif sebagai penapis awal yang mendeteksi frasa lokal bermuatan sentimen (seperti "sangat hemat" atau "baterai boros") sekaligus mereduksi *noise*. Fitur-fitur tersebut kemudian disusun ulang konteksnya oleh lapisan BiLSTM, yang mampu menghubungkan kata-kata berjauhan, misalnya mengaitkan negasi "tidak" di awal kalimat dengan kata sifat di akhir kalimat, sehingga mencegah misklasifikasi.

Peningkatan kinerja yang paling signifikan dikontribusikan oleh mekanisme *Attention* [4], yang memberikan bobot lebih pada kata-kata determinan. Hal ini tercermin dari kemampuan model membedakan nuansa kalimat panjang yang diakhiri dengan opini kuat, menghasilkan *Recall* tinggi pada kelas Netral dan Presisi tertinggi pada kelas Positif (0.89). Secara keseluruhan, model *Hybrid* memiliki kemampuan diskriminatif terbaik dengan AUC 0.935 untuk kelas Negatif dan 0.923 untuk kelas Positif. Meskipun demikian, tantangan untuk memisahkan kelas Positif dan Netral secara sempurna masih tersisa, terlihat dari beberapa sampel Positif yang masih terklasifikasi sebagai Netral. Tabel di bawah ini merangkum perbandingan kinerja kuantitatif dari seluruh model yang diuji.

Tabel 3. Tabel Perbandingan Kinerja Model

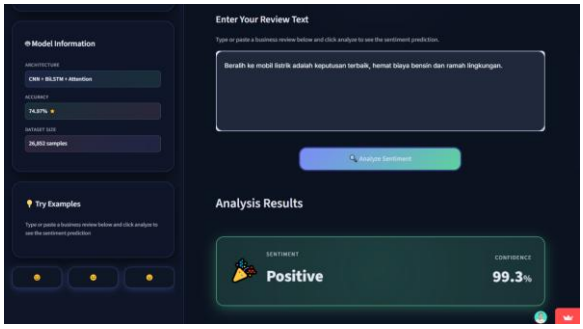
Model	Accuracy	F1 (Macro Avg)	F1 (Weighted Avg)	ROC-AUC (Weighted Avg)
CNN + BiLSTM	**0.75**	**0.76**	**0.75**	**~0.90**
BiLSTM	0.73	0.73	0.73	~0.88
CNN	0.72	0.72	0.72	~0.88
LSTM	0.43	0.20	0.26	0.50

4.5. Analisis Stabilitas Epoch dan Pemilihan Model Final

Selain arsitektur, durasi pelatihan (*epoch*) dievaluasi untuk menentukan titik konvergensi optimal. Eksperimen dilakukan pada 50, 100, dan 150 *epoch*. Hasil menunjukkan pola yang kontras antara model tunggal dan *hybrid*. Model CNN mengalami degradasi performa (*overfitting*) ketika dilatih terlalu lama, dengan penurunan F1-Score dari 0.7444 (epoch 50) menjadi 0.7207 (epoch 150). Sebaliknya, model *Hybrid* CNN-BiLSTM-Attention menunjukkan karakteristik kurva pembelajaran yang positif dan stabil, di mana kinerjanya terus meningkat hingga mencapai puncak F1-Score 0.7599 pada *epoch* 150. Hal ini membuktikan bahwa arsitektur *Hybrid* memiliki kapasitas belajar (*learning capacity*) yang lebih dalam dan lebih tahan terhadap *overfitting* dibandingkan model tunggal, menjadikannya kandidat terbaik untuk tahap implementasi.

4.6. Implementasi Sistem (Deployment)

Sebagai validasi akhir, model *Hybrid* terbaik diimplementasikan ke dalam aplikasi web berbasis *Streamlit* untuk menguji fungsionalitasnya secara *real-time*. Sistem dibangun dengan mengintegrasikan model terlatih (*.h5*), *tokenizer*, dan *label encoder*. Pengujian kualitatif dilakukan dengan memasukkan input teks baru yang merepresentasikan sentimen Positif, Negatif, dan Netral. Hasil pengujian menunjukkan bahwa sistem mampu melakukan pra-pemrosesan (seperti *stemming* dan *cleaning*) secara otomatis dan memberikan prediksi yang akurat dengan tingkat keyakinan (*confidence score*) yang tinggi. Misalnya, input kalimat negatif terdeteksi dengan tepat sebagai "Negatif", sementara kalimat informatif diklasifikasikan sebagai "Netral". Keberhasilan *deployment* ini memvalidasi bahwa model tidak hanya unggul secara metrik statistik, tetapi juga *robust* dan fungsional untuk digunakan dalam analisis opini publik di dunia nyata.



*G*ambar 5. Uji Coba Prediksi Kalimat Positif pada Aplikasi Streamlit

5. KESIMPULAN DAN SARAN

Berdasarkan hasil pemodelan dan evaluasi komprehensif yang telah dilakukan, dapat disimpulkan bahwa arsitektur hibrida CNN-BiLSTM dengan mekanisme *Attention* merupakan model paling efektif dalam klasifikasi sentimen kendaraan listrik dengan

capaian akurasi 75% dan *macro F1-score* 0,75, mengungguli BiLSTM (74%), CNN (71%), serta LSTM standar yang mengalami kegagalan total (*model collapse*). Keberhasilan tersebut didorong oleh sinergi CNN dalam mengekstraksi fitur lokal dan BiLSTM dalam memahami konteks sekuensial dua arah, meskipun tantangan sistemik berupa ambiguitas antara kelas positif dan netral masih menyebabkan rendahnya nilai *recall* positif sebesar 0,63 akibat gaya bahasa informatif yang halus. Guna meningkatkan performa di masa depan, disarankan untuk melakukan *tuning* hiperparameter secara lebih ekstensif, mengadopsi model Transformer seperti IndoBERT untuk menangkap nuansa bahasa yang lebih mendalam, serta mengintegrasikan fitur berbasis leksikon atau *feature engineering* tambahan seperti analisis emoji untuk memperkuat daya diskriminasi model terhadap sentimen yang kompleks.

DAFTAR PUSTAKA

[1] Y. Mao and Q. Li, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 101890, 2024, doi: 10.1016/j.jksuci.2024.101890.

[2] A. K. Halimi, "Analisis Sentimen Masyarakat Indonesia Terhadap Pembelajaran Online Dari Media Sosial Twitter Menggunakan Lexicon Dan K-Nearest Neighbor," *COREAI J. Kecerdasan Buatan, Komputasi dan Teknol. Inf.*, vol. 2, no. 1, pp. 1–10, 2021, doi: 10.33650/coreai.v2i1.2283.

[3] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar: Association for Computational Linguistics, 2014, pp. 1746–1751. doi: 10.3115/v1/D14-1181.

[4] Q. Zhou and Z. Wang, "NLP at IEST 2018: BiLSTM-Attention and LSTM-Attention via Soft Voting in Emotion Classification," in *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2018, pp. 1–6. doi: 10.18653/v1/W18-6226.

[5] A. S. Widagdo and A. Kurniawan, "Analisis Sentimen Mobil Listrik di Indonesia Menggunakan Long-Short Term Memory (LSTM)," *J. Fasilkom*, vol. 13, no. 3, pp. 416–423, 2023, doi: 10.37859/jf.v13i3.5678.

[6] R. S. Merdiansah, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 20–29, 2024, doi: 10.55338/jikomsi.v7i1.2895.

[7] B. Liu, "Text sentiment analysis based on CBOW model and deep learning in big data environment," *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 6, pp. 1–10, 2020, doi: 10.1007/s12652-018-1095-6.

[8] S. Mujilahwati, "Analisis Sentimen Pengguna Aplikasi ChatGPT Berdasarkan Rating Menggunakan Metode Lexicon," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 9, no. 1, pp. 1–8, 2023, doi: 10.36341/rabit.v9i1.3845.

[9] I. Julianto, "Analisis Sentimen Terhadap Sistem Informasi Akademik Institut Teknologi Garut," *J. Algoritm.*, vol. 19, no. 1, pp. 45–52, 2022, doi: 10.33364/algoritma/v.19-1.1112.

[10] W. Meng and Y. Wei, "Aspect Based Sentiment Analysis With Feature Enhanced Attention CNN-BiLSTM," *IEEE Access*, vol. 7, pp. 167240–167249, 2019, doi: 10.1109/ACCESS.2019.295240.