



Perbandingan Long Short-Term Memory dan Bidirectional Long Short-Term Memory pada Analisis Sentimen Ulasan Wisata

Adam Malik*, Nur Ariesanto Ramdhan, Otong Saeful Bachri

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Muhadi Setiabudi, Brebes, Indonesia

Email: ^{1,*}adamfans61@gmail.com, ²ariesantoramdhan@gmail.com, ³otongsaulbahriumus@gmail.com

Email Penulis Korespondensi: adamfans61@gmail.com

Abstrak—Ulasan wisata pada *Google Maps* dapat digunakan untuk memahami persepsi pengunjung terhadap kualitas destinasi, termasuk aspek yang sering menjadi sumber kepuasan maupun keluhan. Penelitian ini menganalisis ulasan wisata Pantai Alam Indah karena destinasi tersebut memiliki jumlah ulasan yang cukup banyak, bentuk teks yang beragam, serta distribusi sentimen yang tidak seimbang. Tujuan penelitian ini adalah membandingkan performa *Long Short-Term Memory* dan *Bidirectional Long Short-Term Memory* dalam analisis sentimen ulasan wisata. Data dikumpulkan melalui *scraping Google Maps*, kemudian diproses melalui tahapan *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*. *Dataset* akhir berjumlah 2.702 ulasan, terdiri dari 2.024 ulasan positif dan 678 ulasan negatif. Untuk mengurangi pengaruh ketidakseimbangan kelas, proses pelatihan menerapkan *class weighting* dengan bobot lebih besar pada kelas negatif. Kedua model menggunakan *embedding* 128 dimensi, *dense layer* 64 neuron, fungsi aktivasi *sigmoid* pada *output layer*, *optimizer Adam*, *batch size* 32, *epoch* 10, dan *validation split* 0,2. Selain klasifikasi sentimen, ulasan juga dikelompokkan ke dalam aspek kebersihan, fasilitas, harga, dan umum menggunakan pendekatan *rule-based* berbasis kata kunci. Hasil evaluasi menunjukkan bahwa *Bidirectional Long Short-Term Memory* memperoleh *accuracy* 75,79%, *precision* 76,33%, *recall* 75,79%, dan *F1-score* 76,04%, sedangkan *Long Short-Term Memory* memperoleh *accuracy* 75,05%, *precision* 75,48%, *recall* 75,05%, dan *F1-score* 75,25%. Selisih performa kedua model relatif kecil, sehingga hasil perlu ditafsirkan secara hati-hati. Analisis aspek menunjukkan bahwa aspek umum mendominasi sentimen positif, sedangkan aspek kebersihan menjadi aspek dengan sentimen negatif terbanyak.

Kata Kunci: *Long Short-Term Memory*; *Bidirectional Long Short-Term Memory*; Analisis Sentimen; Ulasan Wisata; *Class Weighting*

Abstract—Tourism reviews on *Google Maps* can be used to understand visitors' perceptions of destination quality, including aspects that often become sources of satisfaction or complaints. This study analyzes reviews of Pantai Alam Indah because the destination has a large number of reviews, diverse text forms, and an imbalanced sentiment distribution. The purpose of this study is to compare the performance of *Long Short-Term Memory* and *Bidirectional Long Short-Term Memory* in tourism review sentiment analysis. The data were collected through *Google Maps scraping* and processed through *case folding*, *cleaning*, *tokenizing*, *stopword removal*, and *stemming*. The final *dataset* consisted of 2,702 reviews, including 2,024 positive reviews and 678 negative reviews. To reduce the effect of class imbalance, the training process applied *class weighting* by assigning a higher weight to the negative class. Both models used a 128-dimensional *embedding layer*, a 64-neuron *dense layer*, a *sigmoid* activation function in the *output layer*, *Adam optimizer*, *batch size* of 32, 10 *epochs*, and a *validation split* of 0.2. In addition to sentiment classification, the reviews were grouped into cleanliness, facilities, price, and general aspects using a keyword-based *rule-based* approach. The evaluation results show that *Bidirectional Long Short-Term Memory* achieved an *accuracy* of 75.79%, *precision* of 76.33%, *recall* of 75.79%, and *F1-score* of 76.04%, while *Long Short-Term Memory* achieved an *accuracy* of 75.05%, *precision* of 75.48%, *recall* of 75.05%, and *F1-score* of 75.25%. The performance difference between the two models was relatively small, so the results should be interpreted carefully. The aspect analysis shows that the general aspect dominated positive sentiment, while cleanliness had the highest number of negative sentiments.

Keywords: Long Short-Term Memory; Bidirectional Long Short-Term Memory; Sentiment Analysis; Tourism Reviews; Class Weighting

1. PENDAHULUAN

Perkembangan teknologi digital membuat masyarakat semakin mudah memperoleh informasi, memberikan penilaian, dan membagikan pengalaman terhadap suatu destinasi wisata. Salah satu platform yang banyak digunakan untuk aktivitas tersebut adalah *Google Maps*. Melalui fitur ulasan dan rating, pengunjung dapat menyampaikan pendapat mengenai tempat wisata yang pernah dikunjungi. Ulasan tersebut dapat berisi pengalaman positif, kritik, saran, maupun keluhan terhadap kondisi tertentu pada destinasi wisata. Oleh karena itu, ulasan *Google Maps* dapat dimanfaatkan sebagai sumber data untuk menganalisis opini pengunjung karena memuat tanggapan langsung pengguna terhadap suatu lokasi wisata (Ipmawati et al., 2024).

Pantai Alam Indah merupakan destinasi wisata pantai di Kota Tegal yang banyak memperoleh tanggapan dari pengunjung melalui *Google Maps*. Ulasan yang diberikan pengunjung dapat menggambarkan berbagai persepsi terhadap pengalaman berwisata, seperti kondisi lingkungan, kebersihan, fasilitas, harga, dan kenyamanan tempat. Informasi tersebut penting untuk mengetahui bagaimana pengunjung menilai kualitas destinasi secara lebih luas. Akan tetapi, jumlah ulasan yang banyak dan bentuk teks yang tidak terstruktur membuat analisis manual menjadi kurang efektif. Proses manual membutuhkan waktu lebih lama, berisiko subjektif, dan sulit diterapkan pada data ulasan dalam jumlah besar. Oleh sebab itu, diperlukan pendekatan komputasional yang mampu mengolah ulasan secara lebih sistematis.

Analisis sentimen merupakan metode pengolahan teks yang dapat digunakan untuk mengetahui kecenderungan opini dalam suatu ulasan. Melalui pendekatan ini, ulasan pengunjung dapat diklasifikasikan ke dalam sentimen positif



atau negatif. Analisis sentimen banyak diterapkan pada data ulasan digital karena mampu membantu memahami persepsi pengguna terhadap layanan, produk, maupun destinasi wisata secara lebih efisien. Dalam konteks pariwisata, hasil analisis sentimen dapat memberikan gambaran mengenai tingkat kepuasan atau ketidakpuasan pengunjung berdasarkan pengalaman yang mereka tuliskan. Penelitian sebelumnya menunjukkan bahwa ulasan *Google Maps* pada objek wisata dapat dimanfaatkan untuk menganalisis sentimen wisatawan dan menghasilkan informasi yang berguna sebagai bahan evaluasi pengelolaan destinasi (Sudiatmika et al., 2025).

Meskipun analisis sentimen dapat menunjukkan kecenderungan opini secara umum, informasi yang dihasilkan akan lebih bernilai apabila ulasan juga dipetakan berdasarkan aspek yang dibahas. Hal ini karena satu ulasan tidak selalu hanya memuat satu pendapat. Seorang pengunjung, misalnya, dapat memberikan komentar positif terhadap keindahan pantai, tetapi pada saat yang sama menyampaikan keluhan mengenai kebersihan atau fasilitas. Penelitian terdahulu menjelaskan bahwa analisis sentimen pada tingkat dokumen memiliki keterbatasan karena ulasan yang mengandung beberapa opini tetap diklasifikasikan ke dalam satu polaritas utama. Kondisi tersebut dapat membuat informasi mengenai sumber kepuasan atau keluhan menjadi kurang spesifik. Oleh karena itu, pengelompokan aspek diperlukan agar hasil analisis ulasan wisata dapat memberikan gambaran yang lebih jelas untuk mendukung evaluasi layanan (Jayanto et al., 2022).

Pada penelitian ini, ulasan Pantai Alam Indah dianalisis tidak hanya berdasarkan sentimen positif dan negatif, tetapi juga dikelompokkan ke dalam aspek kebersihan, fasilitas, harga, dan umum. Pengelompokan tersebut dilakukan untuk mengetahui aspek yang paling banyak memperoleh tanggapan positif maupun negatif dari pengunjung. Pendekatan ini sesuai dengan penelitian sebelumnya yang menyatakan bahwa analisis sentimen berbasis aspek dapat membantu mengidentifikasi topik yang muncul dalam ulasan sekaligus menentukan polaritas sentimen terhadap aspek tersebut. Dengan demikian, hasil penelitian tidak hanya menampilkan kecenderungan opini secara keseluruhan, tetapi juga menyajikan informasi yang lebih terarah bagi pengelola wisata (Af'idah et al., 2023).

Beragam pendekatan telah digunakan dalam penelitian analisis sentimen, mulai dari metode *machine learning* hingga *deep learning*. Pada data berbentuk teks, metode *deep learning* banyak digunakan karena mampu mempelajari pola kata serta hubungan antar kata dalam suatu kalimat. *Long Short-Term Memory (LSTM)* merupakan salah satu model *deep learning* yang dirancang untuk mengolah data sekuensial. Model ini memiliki kemampuan menyimpan informasi penting melalui mekanisme gerbang, sehingga dapat digunakan untuk mempelajari pola pada data teks. Dalam analisis sentimen, *LSTM* sering digunakan karena mampu memahami hubungan kata dalam urutan kalimat dan mengurangi kelemahan *RNN* dalam menangani ketergantungan jangka panjang (Jayanto et al., 2022).

Selain *LSTM*, terdapat model pengembangan yang dikenal sebagai *Bidirectional Long Short-Term Memory (BiLSTM)*. Perbedaan utama antara *LSTM* dan *BiLSTM* terletak pada arah pemrosesan data teks. *LSTM* mempelajari urutan kata dari satu arah, sedangkan *BiLSTM* memproses teks dari dua arah, yaitu dari awal ke akhir dan dari akhir ke awal. Mekanisme dua arah tersebut memungkinkan *BiLSTM* memperoleh informasi konteks yang lebih lengkap karena model tidak hanya mempertimbangkan kata sebelumnya, tetapi juga kata sesudahnya. Kemampuan ini penting dalam analisis sentimen karena makna suatu kata dalam ulasan sering dipengaruhi oleh konteks kalimat secara keseluruhan. Oleh sebab itu, *LSTM* dan *BiLSTM* relevan digunakan untuk mengolah ulasan berbahasa Indonesia. Pada domain wisata, *BiLSTM* juga telah menunjukkan performa yang baik dalam analisis sentimen ulasan destinasi wisata Indonesia (Af'idah et al., 2023; Cahyaningtyas et al., 2021).

Sejumlah penelitian sebelumnya telah membahas analisis sentimen pada ulasan wisata digital. Penelitian mengenai ulasan *Google Maps* Taman Nasional Bromo Tengger Semeru menunjukkan bahwa data ulasan pengguna dapat dimanfaatkan untuk memahami persepsi wisatawan terhadap aspek tertentu pada destinasi wisata (Bahri & Suadaa, 2023). Penelitian lain pada ulasan objek wisata Baturaden menjelaskan bahwa variasi bahasa, karakteristik teks ulasan, serta ketidakseimbangan data menjadi tantangan dalam analisis sentimen pada bidang pariwisata (Febrianto et al., 2024). Selain itu, penelitian terhadap ulasan tempat wisata di *Google Maps* menggunakan algoritma *Support Vector Machine* menunjukkan bahwa ulasan digital dapat digunakan untuk mengetahui kecenderungan opini pengunjung terhadap suatu destinasi (Ipmawati et al., 2024).

Berdasarkan penelitian-penelitian tersebut, analisis sentimen pada ulasan wisata telah banyak dilakukan dengan berbagai metode, baik *machine learning* maupun *deep learning*. Namun, masih terdapat beberapa celah penelitian yang perlu diperhatikan. Pertama, sebagian penelitian masih berfokus pada klasifikasi sentimen secara umum, sehingga informasi mengenai aspek yang menjadi sumber kepuasan atau keluhan pengunjung belum dijelaskan secara rinci. Kedua, perbandingan *Long Short-Term Memory* dan *Bidirectional Long Short-Term Memory* pada ulasan wisata berbahasa Indonesia masih perlu dikaji dengan memperhatikan karakteristik data yang digunakan. Ketiga, ulasan *Google Maps* memiliki bentuk teks yang singkat, tidak baku, beragam, serta cenderung memiliki distribusi kelas yang tidak seimbang. Kondisi tersebut dapat memengaruhi kemampuan model dalam mengenali pola sentimen, terutama pada kelas dengan jumlah data lebih sedikit.

Ulasan Pantai Alam Indah memiliki karakteristik yang sesuai dengan permasalahan tersebut karena data yang digunakan didominasi oleh sentimen positif, sedangkan sentimen negatif memiliki jumlah yang lebih sedikit. Selain itu, isi ulasan tidak hanya memuat penilaian umum, tetapi juga membahas aspek tertentu seperti kebersihan, fasilitas, harga, dan pengalaman berkunjung. Oleh karena itu, penelitian ini tidak hanya membandingkan performa *Long Short-Term Memory* dan *Bidirectional Long Short-Term Memory*, tetapi juga memetakan ulasan berdasarkan aspek untuk mengetahui bagian yang paling banyak memperoleh tanggapan positif maupun negatif.

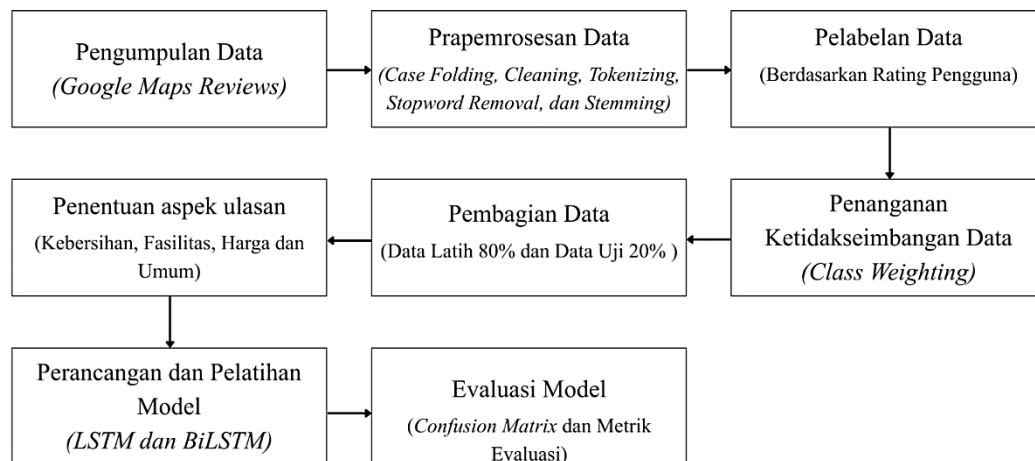
Penelitian ini bertujuan untuk membandingkan performa *Long Short-Term Memory* dan *Bidirectional Long Short-Term Memory* dalam analisis sentimen ulasan wisata. Pengelompokan aspek kebersihan, fasilitas, harga, dan umum digunakan sebagai analisis tambahan untuk mengetahui aspek yang paling banyak memperoleh sentimen positif dan negatif. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi berupa perbandingan performa dua model *deep learning* pada data ulasan wisata yang tidak seimbang serta informasi aspek yang dapat digunakan sebagai bahan evaluasi pengelolaan destinasi wisata.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membandingkan performa *LSTM* dan *BiLSTM* dalam analisis sentimen ulasan Pantai Alam Indah. Data yang digunakan berupa ulasan pengunjung yang diperoleh dari *Google Maps*. Secara umum, tahapan penelitian meliputi pengumpulan data, prapemrosesan teks, penentuan aspek, pelabelan sentimen, pembagian data, pembangunan model *LSTM* dan *BiLSTM*, serta evaluasi performa model. Metode eksperimen digunakan karena penelitian ini menguji dua model pada *dataset*, parameter pelatihan, dan skema evaluasi yang sama, sehingga hasil perbandingan dapat dianalisis secara lebih objektif.

2.1 Tahapan Penelitian

Penelitian ini diawali dengan proses pengumpulan data ulasan Pantai Alam Indah yang bersumber dari *Google Maps* menggunakan teknik *scraping*. Data yang telah diperoleh kemudian diseleksi untuk menghapus ulasan kosong, data yang tidak sesuai, serta elemen teks yang tidak diperlukan. Setelah itu, data diproses melalui tahapan prapemrosesan agar teks ulasan menjadi lebih bersih dan memiliki format yang seragam. Tahap prapemrosesan yang dilakukan meliputi *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*. Tahapan tersebut digunakan untuk menyiapkan teks agar lebih layak digunakan dalam proses analisis sentimen dan pemodelan (Aufar et al., 2023). Setelah data selesai diproses, ulasan dikelompokkan ke dalam beberapa aspek menggunakan pendekatan *rule-based* berbasis kata kunci. Aspek yang digunakan dalam penelitian ini meliputi kebersihan, fasilitas, harga, dan umum. Tahap berikutnya adalah pemberian label sentimen berdasarkan rating pengguna, kemudian data dibagi menjadi data latih dan data uji. Data latih digunakan untuk membangun model *LSTM* dan *BiLSTM*, sedangkan data uji digunakan untuk mengukur kemampuan model dalam mengklasifikasikan sentimen. Evaluasi dilakukan menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Alur penelitian secara keseluruhan disajikan pada Gambar 1.



Gambar 1. Bagan Alir Tahapan Penelitian

Berdasarkan Gambar 1, penelitian dimulai dari pengumpulan data ulasan *Google Maps*, kemudian dilanjutkan dengan prapemrosesan data untuk membersihkan dan menyeragamkan teks ulasan. Setelah itu, data diberi label sentimen berdasarkan rating pengguna dan dikelompokkan ke dalam aspek kebersihan, fasilitas, harga, dan umum. Data yang telah siap kemudian dibagi menjadi data latih dan data uji. Karena distribusi sentimen pada *dataset* tidak seimbang, tahap pelatihan model menerapkan *class weighting* untuk mengurangi kecenderungan model terhadap kelas mayoritas. Selanjutnya, model *LSTM* dan *BiLSTM* dilatih menggunakan data latih dan dievaluasi menggunakan data uji melalui *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari ulasan pengunjung Pantai Alam Indah pada *Google Maps*. Pengambilan data dilakukan dengan bantuan *Apify* untuk memperoleh teks ulasan dan rating yang diberikan oleh pengguna. Dari proses pengumpulan data tersebut, diperoleh *dataset* awal sebanyak 5.609 ulasan. Namun, data awal tersebut belum sepenuhnya siap digunakan karena masih memuat ulasan kosong, karakter khusus, simbol, emoji, serta bagian teks yang tidak relevan dengan kebutuhan analisis.



Oleh karena itu, data terlebih dahulu melalui proses seleksi dan prapemrosesan sebelum digunakan pada tahap pemodelan. Setelah proses tersebut dilakukan, jumlah data akhir yang digunakan dalam penelitian ini menjadi 2.702 ulasan. *Dataset* akhir ini selanjutnya digunakan untuk proses pengelompokan aspek, pelabelan sentimen, pembagian data, pelatihan model *LSTM* dan *BiLSTM*, serta evaluasi performa model.

Tabel 1. Jumlah Data Penelitian

Keterangan	Jumlah Data
Data awal sebelum prapemrosesan	5.609
Data akhir setelah prapemrosesan	2.702

2.3 Prapemrosesan Data

Prapemrosesan data dilakukan untuk meningkatkan kualitas teks ulasan sebelum digunakan dalam proses klasifikasi sentimen. Ulasan yang berasal dari platform digital umumnya masih memiliki bentuk yang tidak teratur, seperti penggunaan huruf kapital yang tidak konsisten, tanda baca, angka, simbol, tautan, emoji, serta kata-kata yang tidak memberikan kontribusi penting terhadap analisis. Oleh karena itu, tahap prapemrosesan diperlukan agar data teks menjadi lebih bersih, konsisten, dan siap diolah oleh model. Proses ini menjadi tahap awal yang penting dalam klasifikasi teks karena berfungsi menyiapkan data mentah menjadi bentuk yang lebih sesuai untuk analisis dan pemodelan (Aufar et al., 2023).

Tahapan prapemrosesan yang digunakan dalam penelitian ini meliputi *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*. *Case folding* dilakukan untuk menyeragamkan huruf dengan mengubah seluruh teks menjadi huruf kecil. *Cleaning* digunakan untuk menghapus elemen yang tidak dibutuhkan, seperti tanda baca, angka, simbol, tautan, emoji, dan karakter lain yang tidak relevan. Selanjutnya, *tokenizing* memecah kalimat menjadi unit kata agar teks lebih mudah diproses. Tahap *stopword removal* digunakan untuk menghilangkan kata umum yang tidak terlalu berpengaruh terhadap sentimen. Setelah itu, *stemming* diterapkan untuk mengubah kata berimbuhan menjadi bentuk dasar. Hasil akhir dari proses ini disimpan pada kolom stemming dan digunakan sebagai masukan dalam model *LSTM* dan *BiLSTM*.

2.4 Penentuan Aspek

Penentuan aspek dilakukan untuk mengidentifikasi topik atau bagian tertentu yang dibahas dalam ulasan pengunjung. Pada penelitian ini, aspek ditentukan menggunakan pendekatan *rule-based* berbasis kata kunci. Proses ini dilakukan dengan mencocokkan kata-kata pada ulasan hasil prapemrosesan dengan daftar kata kunci yang telah disusun sebelumnya. Aspek yang digunakan terdiri dari kebersihan, fasilitas, harga, dan umum.

Aspek kebersihan mencakup ulasan yang membahas kondisi lingkungan wisata, seperti kebersihan area pantai, keberadaan sampah, atau keadaan tempat yang kurang terawat. Aspek fasilitas digunakan untuk ulasan yang berkaitan dengan sarana pendukung, seperti toilet, tempat parkir, mushola, akses jalan, tempat duduk, dan fasilitas lain yang tersedia di kawasan wisata. Aspek harga digunakan untuk ulasan yang menyinggung tiket masuk, biaya parkir, tarif, atau harga layanan. Sementara itu, aspek umum digunakan untuk ulasan yang berisi penilaian keseluruhan mengenai suasana, kenyamanan, keindahan, dan pengalaman berkunjung. Pengelompokan aspek ini bertujuan agar hasil analisis tidak hanya menunjukkan sentimen positif dan negatif secara umum, tetapi juga memperlihatkan bagian yang paling banyak mendapatkan perhatian dari pengunjung.

Tabel 2. Kata Kunci Aspek

Aspek	Kata Kunci
Kebersihan	bersih, kotor, sampah, berserak, buang, kumuh, jorok
Fasilitas	parkir, toilet, wc, kamar mandi, mandi, bilas, warung, mushola, fasilitas, area, duduk, jalan, akses
Harga	harga, murah, mahal, tiket, bayar, biaya, tarif, masuk, rb, jangkau, gratis, htm
Umum	bagus, indah, nyaman, cocok, sejuk, keren, mantap, enak, ramai, sunset, pantai, laut, wisata, keluarga, libur, rekreasi, santai.

Pendekatan *rule-based* dipilih karena penelitian ini berfokus pada perbandingan performa model sentimen, sedangkan aspek digunakan sebagai pemetaan tambahan untuk mengetahui topik kepuasan dan keluhan pengunjung. Meskipun demikian, pendekatan ini memiliki keterbatasan karena bergantung pada daftar kata kunci yang telah ditentukan sebelumnya. Variasi bahasa, sinonim, dan konteks kalimat tertentu dapat menyebabkan sebagian ulasan tidak teridentifikasi secara optimal. Oleh karena itu, hasil pengelompokan aspek pada penelitian ini ditafsirkan sebagai analisis deskriptif, bukan sebagai ekstraksi aspek otomatis berbasis model.

2.5 Pelabelan Sentimen dan Pembagian Data

Pelabelan sentimen pada penelitian ini dilakukan dengan memanfaatkan rating yang diberikan oleh pengunjung di *Google Maps*. Ulasan dengan rating 4 dan 5 dimasukkan ke dalam kelas sentimen positif, sedangkan ulasan dengan rating 1, 2, dan 3 dimasukkan ke dalam kelas sentimen negatif. Penggunaan rating sebagai dasar penentuan label sentimen juga diterapkan pada penelitian sebelumnya, yaitu dengan mengelompokkan rating 4 dan 5 sebagai sentimen positif serta rating 1, 2, dan 3 sebagai sentimen negatif pada data ulasan aplikasi (Maulana et al., 2023).

Tabel 3. Aturan Pelabelan Sentimen

Rating	Label Sentimen	Encoding
1	Negatif	0
2	Negatif	0
3	Negatif	0
4	Positif	1
5	Positif	1

Setelah proses pelabelan selesai, setiap label sentimen diubah ke dalam bentuk numerik agar dapat diproses oleh model. Label negatif direpresentasikan dengan nilai 0, sedangkan label positif direpresentasikan dengan nilai 1. *Dataset* akhir yang digunakan berjumlah 2.702 ulasan, terdiri dari 2.024 ulasan positif dan 678 ulasan negatif. Selanjutnya, data dibagi menggunakan skema 80:20, yaitu 2.161 data sebagai data latih dan 541 data sebagai data uji. Data latih digunakan untuk proses pelatihan model *LSTM* dan *BiLSTM*, sedangkan data uji digunakan untuk menilai kemampuan model dalam memprediksi sentimen pada data yang belum digunakan selama proses pelatihan.

2.6 Penanganan Ketidakseimbangan Data

Dataset akhir pada penelitian ini memiliki distribusi kelas yang tidak seimbang, yaitu 2.024 ulasan positif dan 678 ulasan negatif. Ketidakseimbangan data tersebut dapat membuat model lebih cenderung mengenali kelas mayoritas, yaitu sentimen positif, dibandingkan kelas minoritas. Untuk mengurangi pengaruh tersebut, penelitian ini menerapkan *class weighting* pada tahap pelatihan model *Long Short-Term Memory* dan *Bidirectional Long Short-Term Memory*.

Bobot kelas dihitung menggunakan pendekatan *balanced class weight* berdasarkan jumlah data pada setiap kelas. Hasil perhitungan menghasilkan bobot sebesar 1,9935 untuk kelas negatif dan 0,6673 untuk kelas positif. Dengan pemberian bobot tersebut, kesalahan prediksi pada kelas negatif memperoleh perhatian lebih besar selama proses pelatihan. Penerapan *class weighting* digunakan agar model tidak hanya mempelajari pola kelas positif yang jumlahnya lebih dominan, tetapi juga tetap mempertimbangkan pola pada kelas negatif.

2.7 Perancangan Model LSTM dan BiLSTM

Penelitian ini menggunakan dua model *deep learning*, yaitu *LSTM* dan *BiLSTM*, untuk membandingkan kemampuan keduanya dalam mengklasifikasikan sentimen ulasan. Kedua model tersebut dipilih karena sesuai untuk mengolah data teks yang memiliki bentuk urutan. *LSTM* memproses urutan kata secara searah, sedangkan *BiLSTM* mempelajari teks dari dua arah, yaitu dari awal ke akhir dan dari akhir ke awal. Dengan mekanisme tersebut, *BiLSTM* dapat menangkap informasi konteks yang lebih luas dalam suatu kalimat. *BiLSTM* mampu memahami konteks teks dari dua arah, sehingga dapat memperoleh performa lebih baik dibandingkan *LSTM* dan *GRU* pada tugas analisis sentimen (Suhaeni et al., 2024).

Sebelum data dimasukkan ke dalam model, teks pada kolom stemming terlebih dahulu dikonversi menjadi bentuk numerik menggunakan *Tokenizer*. Setelah itu, proses *padding* diterapkan agar seluruh ulasan memiliki panjang urutan yang sama. Model *LSTM* dan *BiLSTM* dirancang dengan konfigurasi yang sebanding agar perbandingan performa dapat dilakukan secara adil. Kedua model menggunakan *embedding layer*, lapisan *LSTM* atau *BiLSTM*, *dropout layer*, *dense layer*, dan *output layer*. Proses kompilasi model menggunakan fungsi kerugian *binary_crossentropy*, *optimizer Adam*, dan metrik akurasi. Perbandingan antara *LSTM* dan *BiLSTM* relevan digunakan untuk mengetahui perbedaan kinerja kedua model dalam analisis sentimen ulasan pengguna (Gunawan & Aziza, 2025).

Tabel 4. Konfigurasi Model *LSTM* dan *BiLSTM*

Komponen	LSTM	BiLSTM
<i>Input data</i>	Kolom <i>stemming</i>	Kolom <i>stemming</i>
<i>Target</i>	Label sentimen	Label sentimen
<i>Embedding layer</i>	128 dimensi	128 dimensi
<i>Recurrent layer</i>	<i>LSTM</i>	<i>Bidirectional LSTM</i>
Jumlah unit	128	64
<i>Dropout</i>	0,3	0,5
<i>Dense Layer</i>	64 <i>neuron</i> , <i>ReLU</i>	64 <i>neuron</i> , <i>ReLU</i>
<i>Output layer</i>	1 <i>neuron</i> , <i>sigmoid</i>	1 <i>neuron</i> , <i>sigmoid</i>
<i>Loss function</i>	<i>binary_crossentropy</i>	<i>binary_crossentropy</i>
<i>Optimizer</i>	<i>Adam</i>	<i>Adam</i>
<i>Epoch</i>	10	10
<i>Batch size</i>	32	32
<i>Validation split</i>	0,2	0,2
<i>Class weight</i>	Negatif 1,9935; Positif 0,6673	Negatif 1,9935; Positif 0,6673

Berdasarkan Tabel 4, kedua model menggunakan data masukan, target, *embedding layer*, *optimizer*, *loss function*, *batch size*, *epoch*, *validation split*, dan *class weight* yang sama. Perbedaan utama terdapat pada lapisan



recurrent yang digunakan. *Long Short-Term Memory* memproses urutan teks secara searah, sedangkan *Bidirectional Long Short-Term Memory* memproses teks dari dua arah. Penggunaan konfigurasi yang sebanding bertujuan agar hasil perbandingan lebih menggambarkan perbedaan arsitektur model, bukan perbedaan skema pelatihan.

2.8 Evaluasi Model

Evaluasi model dilakukan untuk mengukur kemampuan *LSTM* dan *BiLSTM* dalam mengklasifikasikan sentimen ulasan Pantai Alam Indah. Pengujian dilakukan menggunakan data uji sebanyak 541 ulasan yang sebelumnya tidak digunakan pada proses pelatihan. Metrik yang digunakan dalam evaluasi meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. Keempat metrik tersebut digunakan agar hasil pengujian tidak hanya dilihat dari tingkat ketepatan prediksi secara umum, tetapi juga dari kemampuan model dalam mengenali setiap kelas sentimen (Cherradi & El Haddadi, 2024).

Selain menggunakan metrik evaluasi, penelitian ini juga memanfaatkan *confusion matrix* untuk melihat rincian hasil prediksi model. *Confusion matrix* menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar maupun yang mengalami kesalahan prediksi pada kelas positif dan negatif. Melalui matriks tersebut, performa *LSTM* dan *BiLSTM* dapat dibandingkan secara lebih jelas berdasarkan jumlah prediksi benar dan salah pada masing-masing kelas.

Tabel 5. *Confusion Matrix*

Prediksi / Aktual	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Berdasarkan komponen pada *confusion matrix*, *True Positive (TP)* menunjukkan data positif yang diprediksi benar sebagai positif, sedangkan *True Negative (TN)* menunjukkan data negatif yang diprediksi benar sebagai negatif. *False Positive (FP)* menunjukkan data negatif yang keliru diprediksi sebagai positif, sedangkan *False Negative (FN)* menunjukkan data positif yang keliru diprediksi sebagai negatif. Nilai *TP*, *TN*, *FP*, dan *FN* tersebut kemudian digunakan untuk menghitung *accuracy*, *precision*, *recall*, dan *F1-score*. Rumus evaluasi yang digunakan adalah sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Keterangan: *TP*, *TN*, *FP*, dan *FN* merupakan komponen *confusion matrix* yang digunakan untuk menghitung metrik evaluasi. *Accuracy* menunjukkan proporsi prediksi benar dari seluruh data uji. *Precision* menunjukkan ketepatan model dalam menghasilkan prediksi positif. *Recall* menunjukkan kemampuan model dalam mengenali data positif yang benar. *F1-score* menunjukkan keseimbangan antara *precision* dan *recall*.

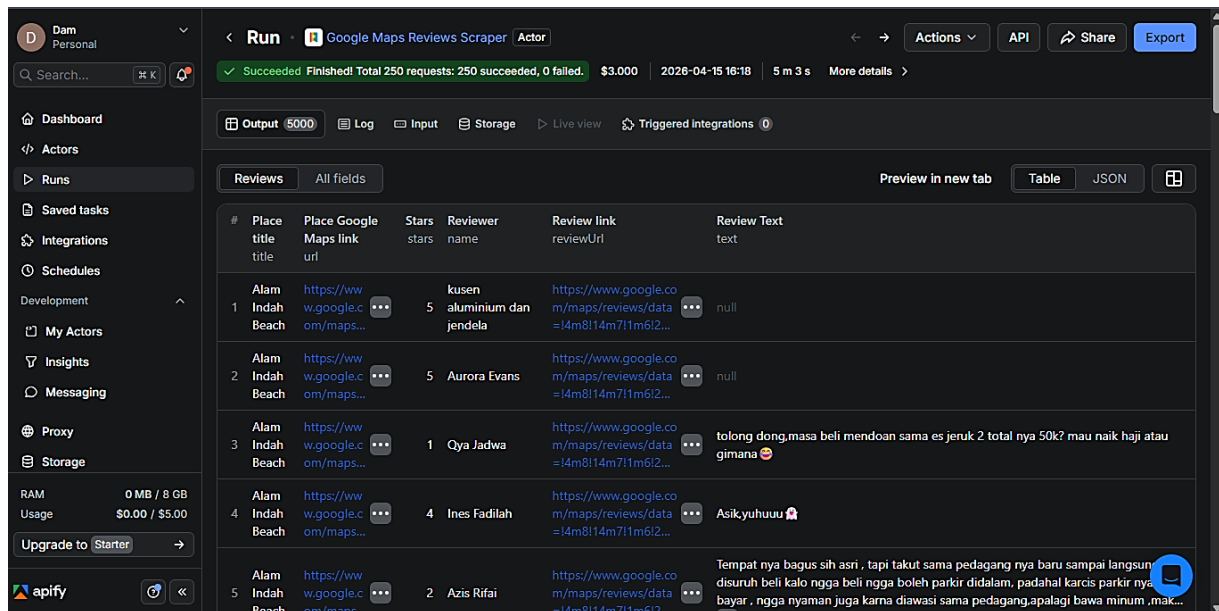
Nilai dari setiap metrik tersebut digunakan untuk membandingkan performa kedua model pada data uji yang sama. Namun, interpretasi hasil evaluasi tidak hanya didasarkan pada nilai *accuracy*, tetapi juga mempertimbangkan *precision*, *recall*, *F1-score*, dan *confusion matrix* agar kemampuan model dalam mengenali setiap kelas sentimen dapat dianalisis secara lebih seimbang.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian yang diperoleh dari proses pengolahan ulasan Pantai Alam Indah. Pembahasan mencakup hasil pengumpulan dan seleksi data, distribusi sentimen, pengelompokan aspek ulasan, serta perbandingan kinerja model *LSTM* dan *BiLSTM*. Selain menampilkan nilai evaluasi model, bagian ini juga menjelaskan aspek yang paling banyak memperoleh tanggapan positif maupun negatif dari pengunjung. Pemanfaatan ulasan *Google Maps* sebagai sumber data dalam analisis sentimen wisata sejalan dengan penelitian sebelumnya yang menggunakan ulasan *Google Maps* untuk mengetahui kecenderungan opini pengunjung terhadap objek wisata (Maulida et al., 2026).

3.1 Analisis Data

Data penelitian diperoleh dari ulasan pengunjung Pantai Alam Indah yang tersedia pada *Google Maps*. Proses pengambilan data dilakukan menggunakan *Apify* melalui *Actor Google Maps Reviews Scraper*. Data yang berhasil dikumpulkan meliputi teks ulasan, rating, nama tempat, nama pengguna, serta tautan ulasan. Tampilan hasil pengambilan data menggunakan *Apify* disajikan pada Gambar 2.



#	Place title	Place Google Maps link	Stars	Reviewer name	Review link	Review Text
1	Alam Indah Beach	https://www.google.com/maps...	5	kusen aluminium dan jendela	https://www.google.com/maps/reviews/data=!4m8!14m7!1m6!2...	null
2	Alam Indah Beach	https://www.google.com/maps...	5	Aurora Evans	https://www.google.com/maps/reviews/data=!4m8!14m7!1m6!2...	null
3	Alam Indah Beach	https://www.google.com/maps...	1	Qya Jadwa	https://www.google.com/maps/reviews/data=!4m8!14m7!1m6!2...	tolong dong,masa beli mendoan sama es jeruk 2 total nya 50k? mau naik haji atau gimana 🤔
4	Alam Indah Beach	https://www.google.com/maps...	4	Ines Fadilah	https://www.google.com/maps/reviews/data=!4m8!14m7!1m6!2...	Asikyuhuuu 🙌
5	Alam Indah Beach	https://www.google.com/maps...	2	Azis Rifai	https://www.google.com/maps/reviews/data=!4m8!14m7!1m6!2...	Tempat nya bagus sih asri , tapi takut sama pedagang nya baru sampai langsung disuruh beli kalo ngga beli ngga boleh parkir didalam, padahal karcis parkir nya bayar , ngga nyaman juga karna diawasi sama pedagang.apalagi bawa minum ,mak...

Gambar 2. Hasil Scraping Ulasan Google Maps Menggunakan Apify

Berdasarkan Gambar 2, data hasil *scraping* masih berupa data mentah sehingga belum dapat langsung digunakan dalam proses analisis. Data tersebut masih memuat ulasan kosong, simbol, emoji, tanda baca, angka, dan beberapa teks yang tidak sesuai dengan kebutuhan penelitian. *Dataset* awal yang diperoleh berjumlah 5.609 ulasan. Setelah melalui proses seleksi dan prapemrosesan, jumlah data yang digunakan dalam penelitian ini menjadi 2.702 ulasan. Hasil ini menunjukkan bahwa ulasan dari platform digital perlu dibersihkan dan diseleksi terlebih dahulu agar layak digunakan sebagai masukan dalam analisis sentimen. Penelitian lain pada ulasan *Google Maps* juga menunjukkan bahwa data ulasan wisata umumnya perlu melalui proses pembersihan sebelum digunakan dalam analisis sentimen (Maulida et al., 2026).

3.2 Hasil Prapemrosesan Data

Prapemrosesan diterapkan untuk mengubah ulasan mentah menjadi teks yang lebih teratur dan siap digunakan dalam pemodelan. Tahapan yang dilakukan terdiri dari *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*. Melalui proses ini, elemen yang tidak diperlukan dapat dikurangi, bentuk kata menjadi lebih seragam, dan teks hasil ulasan dapat diproses lebih mudah oleh model *LSTM* maupun *BiLSTM*. Contoh perubahan data pada setiap tahap prapemrosesan ditampilkan pada Tabel 6.

Tabel 6. Tahapan Prapemrosesan Data

Tahapan	Data Sebelum Prapemrosesan	Data Sesudah Prapemrosesan
<i>Case folding</i>	Pantainya bagus bgt dan asyik buat liburan trs pemandangan pas sunsetnya juga gokil abis!! ☐	pantainya bagus bgt dan asyik buat liburan trs pemandangan pas sunsetnya juga gokil abis!! ☐
<i>Cleaning</i>	pantainya bagus bgt dan asyik buat liburan trs pemandangan pas sunsetnya juga gokil abis!! ☐	pantainya bagus bgt dan asyik buat liburan trs pemandangan pas sunsetnya juga gokil abis
<i>Tokenizing</i>	pantainya bagus bgt dan asyik buat liburan trs pemandangan pas sunsetnya juga gokil abis	['pantainya'], ['bagus'], ['bgt'], ['dan'], ['asyik'], ['buat'], ['liburan'], ['trs'], ['pemandangan'], ['pas'], ['sunsetnya'], ['juga'], ['gokil'], ['abis']
<i>Stopword removal</i>	pantainya, bagus, bgt, dan, asyik, buat, liburan, trs, pemandangan, pas, sunsetnya, juga, gokil, abis	pantainya, bagus, bgt, asyik, liburan, trs, pemandangan, sunsetnya, gokil, abis
<i>Stemming</i>	pantainya, bagus, bgt, asyik, liburan, trs, pemandangan, sunsetnya, gokil, abis	pantai bagus bgt asyik libur trs pandang sunset gokil abis

Berdasarkan Tabel 6, ulasan yang semula masih mengandung tanda baca, kata tidak baku, dan kata umum dapat diubah menjadi bentuk yang lebih sederhana. Karakter yang tidak dibutuhkan dihapus, sedangkan kata berimbuhan diproses menjadi bentuk dasar. Hasil akhir dari proses prapemrosesan disimpan pada kolom *stemming* dan digunakan sebagai data masukan dalam tahap pemodelan. Tahap prapemrosesan seperti *tokenisasi* dan *padding* juga digunakan



dalam penelitian *LSTM* berbasis ulasan *Mobile JKN* untuk menyiapkan data teks sebelum masuk ke model (Tamami et al., 2025).

3.3 Hasil Pelabelan Sentimen

Pelabelan sentimen dilakukan dengan mengacu pada rating yang diberikan pengunjung melalui *Google Maps*. Rating 1, 2, dan 3 dikategorikan sebagai sentimen negatif, sedangkan rating 4 dan 5 dikategorikan sebagai sentimen positif. Setelah proses pelabelan selesai, diperoleh distribusi sentimen sebagaimana ditunjukkan pada Tabel 7.

Tabel 7. Distribusi Label Sentimen

Label Sentimen	Jumlah Data	Persentase
Positif	2.024	74,91%
Negatif	678	25,09%
Total	2.702	100%

Berdasarkan Tabel 7, ulasan Pantai Alam Indah lebih banyak didominasi oleh sentimen positif. Dari total 2.702 ulasan, sebanyak 2.024 ulasan atau 74,91% masuk ke dalam kelas positif, sedangkan 678 ulasan atau 25,09% masuk ke dalam kelas negatif. Temuan ini menunjukkan bahwa sebagian besar pengunjung memberikan penilaian yang baik terhadap Pantai Alam Indah. Meskipun demikian, jumlah ulasan negatif tetap perlu diperhatikan karena menunjukkan adanya keluhan atau ketidakpuasan pengunjung terhadap aspek tertentu.

Distribusi data yang tidak seimbang dapat berpengaruh terhadap kemampuan model dalam mengenali kelas dengan jumlah data lebih sedikit. Pada penelitian sebelumnya, teknik *SMOTE* digunakan pada model *LSTM* untuk menangani ketidakseimbangan data ulasan *Mobile JKN*. Hal tersebut menunjukkan bahwa perbedaan jumlah data pada setiap kelas perlu dipertimbangkan saat menafsirkan hasil evaluasi model (Tamami et al., 2025).

3.4 Hasil Analisis Sentimen Berdasarkan Aspek

Selain mengklasifikasikan ulasan ke dalam sentimen positif dan negatif, penelitian ini juga melakukan pengelompokan ulasan berdasarkan aspek yang dibahas. Aspek yang digunakan terdiri dari kebersihan, fasilitas, harga, dan umum. Pengelompokan tersebut dilakukan dengan pendekatan *rule-based* berbasis kata kunci. Hasil distribusi sentimen pada setiap aspek ditampilkan pada Tabel 8.

Hasil pengelompokan aspek pada penelitian ini bersifat deskriptif karena aspek ditentukan menggunakan pendekatan *rule-based* berbasis kata kunci. Oleh karena itu, hasil aspek digunakan untuk melihat kecenderungan topik ulasan, bukan untuk menggantikan ekstraksi aspek otomatis berbasis model.

Tabel 8. Distribusi Sentimen Berdasarkan Aspek

Aspek	Negatif	Positif	Total
Fasilitas	92	151	243
Harga	59	133	192
Kebersihan	289	378	667
Umum	238	1.362	1.600
Total	678	2.024	2.702

Berdasarkan Tabel 8, aspek umum menjadi aspek dengan jumlah ulasan paling banyak, yaitu 1.600 ulasan. Dari jumlah tersebut, 1.362 ulasan termasuk sentimen positif dan 238 ulasan termasuk sentimen negatif. Kondisi ini menunjukkan bahwa sebagian besar pengunjung memberikan tanggapan positif terhadap pengalaman berkunjung secara umum di Pantai Alam Indah. Ulasan pada aspek ini berkaitan dengan suasana pantai, keindahan tempat, kenyamanan, serta pengalaman rekreasi secara keseluruhan.

Aspek kebersihan menjadi aspek dengan jumlah sentimen negatif paling tinggi, yaitu 289 ulasan negatif dari total 667 ulasan. Temuan ini menunjukkan bahwa kebersihan masih menjadi bagian yang cukup banyak dikeluhkan oleh pengunjung. Keluhan tersebut dapat berkaitan dengan keberadaan sampah, kondisi lingkungan yang kurang terawat, atau persepsi pengunjung terhadap kebersihan kawasan wisata. Penelitian sebelumnya, menunjukkan bahwa analisis ulasan wisata melalui *Google Maps* dapat membantu mengidentifikasi tema-tema dominan yang muncul dalam sentimen positif maupun negatif, termasuk isu yang berhubungan dengan kebersihan dan pengelolaan lingkungan wisata (Maulida et al., 2026).

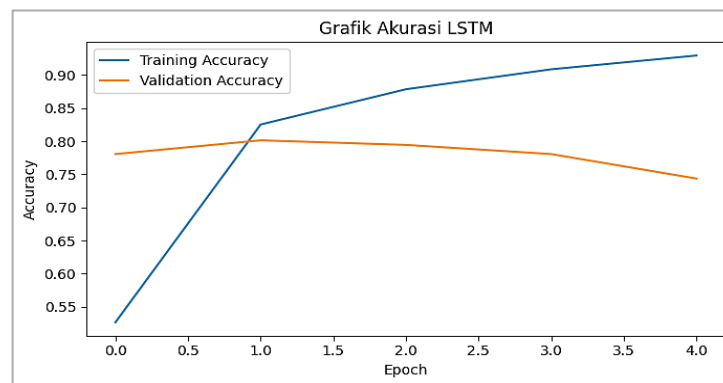
Pada aspek fasilitas, terdapat 151 ulasan positif dan 92 ulasan negatif. Hasil tersebut menunjukkan bahwa fasilitas di Pantai Alam Indah masih memperoleh tanggapan positif, meskipun beberapa pengunjung tetap menyampaikan keluhan terhadap sarana pendukung yang tersedia. Sementara itu, aspek harga memperoleh 133 ulasan positif dan 59 ulasan negatif. Hal ini menunjukkan bahwa sebagian besar pengunjung masih menilai biaya atau harga yang berkaitan dengan kunjungan ke Pantai Alam Indah sebagai hal yang dapat diterima. Secara keseluruhan, hasil pengelompokan aspek menunjukkan bahwa aspek umum perlu dipertahankan karena memperoleh banyak respons positif, sedangkan aspek kebersihan perlu menjadi prioritas perbaikan.

3.5 Hasil Pelatihan Model LSTM dan BiLSTM

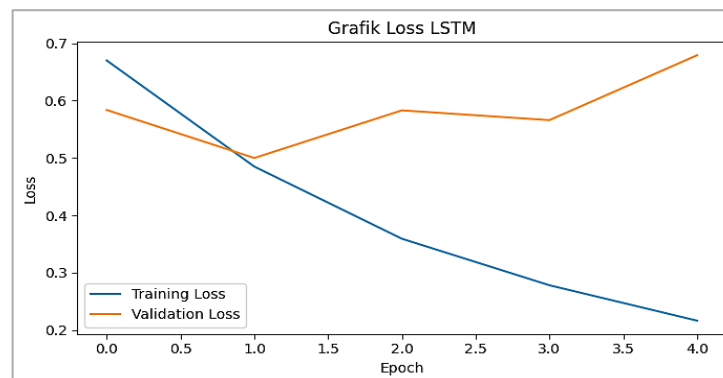
Dataset akhir yang digunakan dalam penelitian ini berjumlah 2.702 ulasan. Data tersebut kemudian dibagi menjadi 2.161 data latih dan 541 data uji. Data latih digunakan untuk membangun model *LSTM* dan *BiLSTM*, sedangkan data uji digunakan untuk mengukur performa model setelah proses pelatihan selesai. Sebelum masuk ke tahap pelatihan, teks hasil *stemming* terlebih dahulu dikonversi ke dalam bentuk numerik menggunakan *Tokenizer*. Selanjutnya, proses *padding* dilakukan agar seluruh ulasan memiliki panjang urutan yang seragam.

Model *LSTM* dan *BiLSTM* disusun dengan konfigurasi yang sama agar hasil perbandingan dapat dinilai secara seimbang. Kedua model menggunakan *embedding layer*, lapisan *LSTM* atau *BiLSTM*, *dropout layer*, *dense layer*, dan *output layer*. Proses pelatihan dilakukan dengan *optimizer Adam*, fungsi kerugian *binary_crossentropy*, *batch size* 32, jumlah *epoch* 10, dan *validation split* 0,2. Penggunaan parameter yang sama bertujuan agar perbedaan hasil evaluasi lebih menggambarkan kemampuan masing-masing arsitektur model, bukan disebabkan oleh perbedaan konfigurasi pelatihan.

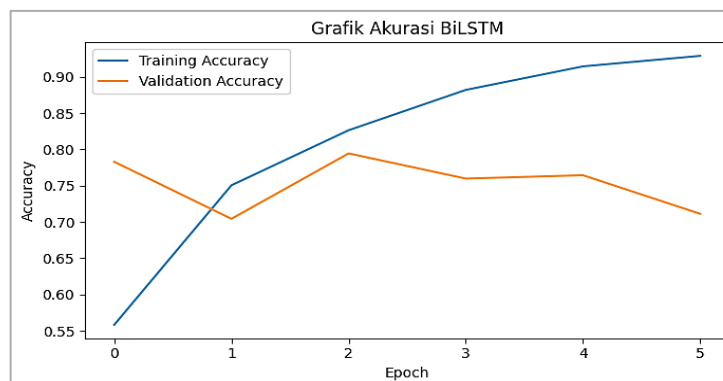
Hasil pelatihan divisualisasikan melalui grafik akurasi dan loss. Grafik akurasi digunakan untuk melihat peningkatan kemampuan model dalam mempelajari pola data selama proses pelatihan, sedangkan grafik loss digunakan untuk melihat besar kesalahan model pada setiap *epoch*. Visualisasi hasil pelatihan *LSTM* dan *BiLSTM* ditampilkan pada Gambar 3 sampai Gambar 6.



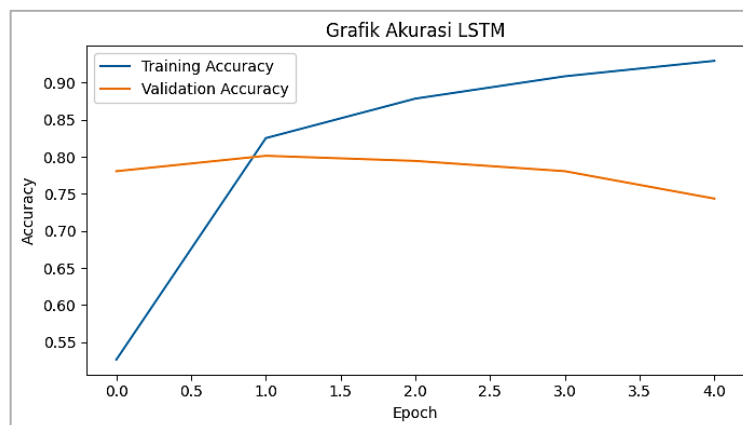
Gambar 3. Grafik Akurasi *Training* dan Validasi Model *LSTM*



Gambar 4. Grafik *Loss Training* dan Validasi Model *LSTM*



Gambar 5. Grafik Akurasi *Training* dan Validasi Model *BiLSTM*



Gambar 6. Grafik Loss Training dan Validasi Model *BiLSTM*

Berdasarkan Gambar 3 sampai Gambar 6, kedua model menunjukkan kemampuan dalam mempelajari pola sentimen dari data ulasan. Namun, *LSTM* dan *BiLSTM* memiliki perbedaan dalam cara memproses urutan teks. *LSTM* membaca urutan kata secara searah, sedangkan *BiLSTM* membaca teks dari dua arah sehingga dapat memperoleh konteks yang lebih lengkap. Kemampuan *BiLSTM* dalam mempelajari konteks teks secara sekuensial menjadikan model ini relevan digunakan pada analisis sentimen ulasan berbahasa Indonesia (Nasution et al., 2025).

Penggunaan *LSTM* dan *BiLSTM* dalam penelitian ini didukung oleh penelitian sebelumnya yang membandingkan kedua model tersebut pada data ulasan hotel berbahasa Indonesia. *LSTM* dan *BiLSTM* dinilai sesuai untuk memproses data teks karena mampu mengenali pola urutan kata serta hubungan konteks dalam kalimat. Oleh karena itu, penerapan kedua model pada ulasan Pantai Alam Indah memiliki dasar metodologis yang sesuai untuk analisis sentimen berbasis teks (Oktapiani et al., 2025).

3.6 Hasil Evaluasi Model *LSTM* dan *BiLSTM*

Evaluasi dilakukan menggunakan data uji sebanyak 541 ulasan. Pengukuran performa model menggunakan empat metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi kedua model disajikan pada Tabel 9.

Tabel 9. Perbandingan Hasil Evaluasi Model *LSTM* dan *BiLSTM*

Model	Accuracy	Precision	Recall	F1-score
<i>LSTM</i>	75,05%	75,48%	75,05%	75,25%
<i>BiLSTM</i>	75,79%	76,33%	75,79%	76,04%

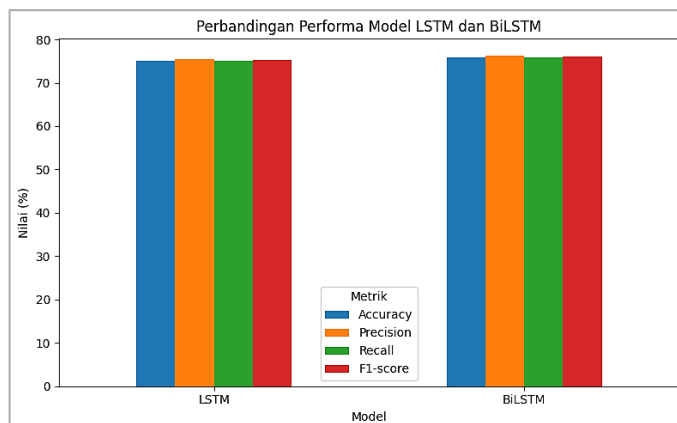
Berdasarkan Tabel 9, *Bidirectional Long Short-Term Memory* memperoleh nilai evaluasi sedikit lebih tinggi dibandingkan *Long Short-Term Memory* pada seluruh metrik. *Long Short-Term Memory* menghasilkan *accuracy* sebesar 75,05%, *precision* 75,48%, *recall* 75,05%, dan *F1-score* 75,25%. Sementara itu, *Bidirectional Long Short-Term Memory* memperoleh *accuracy* sebesar 75,79%, *precision* 76,33%, *recall* 75,79%, dan *F1-score* 76,04%. Selisih *accuracy* antara kedua model sebesar 0,74%, sedangkan selisih *F1-score* sebesar 0,79%. Selisih tersebut menunjukkan bahwa peningkatan performa *Bidirectional Long Short-Term Memory* relatif kecil, sehingga hasil perbandingan perlu ditafsirkan secara hati-hati dan tidak hanya berdasarkan *accuracy*.

Selain nilai evaluasi secara keseluruhan, hasil pengujian juga menunjukkan adanya perbedaan performa pada masing-masing kelas sentimen. *Long Short-Term Memory* memperoleh *F1-score* sebesar 0,52 pada kelas negatif dan 0,83 pada kelas positif. Sementara itu, *Bidirectional Long Short-Term Memory* memperoleh *F1-score* sebesar 0,53 pada kelas negatif dan 0,84 pada kelas positif. Hasil tersebut menunjukkan bahwa kedua model masih lebih kuat dalam mengenali kelas positif dibandingkan kelas negatif. Kondisi ini berkaitan dengan distribusi data yang didominasi oleh ulasan positif, meskipun *class weighting* telah diterapkan pada proses pelatihan.

Performa *BiLSTM* yang lebih tinggi dapat dikaitkan dengan mekanisme pemrosesan dua arah yang dimilikinya. Model ini tidak hanya membaca urutan kata dari awal ke akhir, tetapi juga dari akhir ke awal. Dengan cara tersebut, *BiLSTM* dapat menangkap informasi konteks sebelum dan sesudah kata secara lebih lengkap. Pada ulasan wisata yang cenderung singkat, tidak selalu baku, dan memiliki pola kalimat yang beragam, kemampuan memahami konteks dua arah dapat membantu model mengenali kecenderungan sentimen dengan lebih baik. Hasil ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa *BiLSTM* memperoleh akurasi lebih tinggi dibandingkan *LSTM* pada analisis sentimen multi-kelas media sosial *Twitter* (Asnawiyah & Putra, 2024).

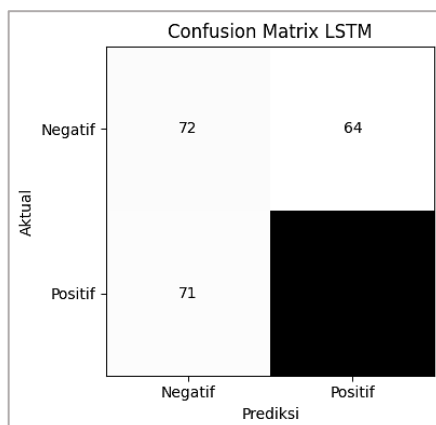
Temuan penelitian ini juga relevan dengan penelitian sebelumnya yang menerapkan *BiLSTM* pada analisis sentimen ulasan aplikasi *Mobile JKN*. Penelitian tersebut menggunakan tahapan *embedding*, *BiLSTM*, *dropout*, dan *dense layer*, serta mengevaluasi model menggunakan *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Meskipun objek penelitian berbeda, yaitu ulasan aplikasi layanan kesehatan, hasil tersebut tetap mendukung bahwa *BiLSTM* dapat digunakan untuk mengolah teks ulasan berbahasa Indonesia (Malik & Irawan, 2026).

Selain itu, penelitian sebelumnya menunjukkan bahwa *LSTM* dapat digunakan untuk analisis sentimen pada ulasan objek wisata pantai yang bersumber dari *Google Maps*. Hal tersebut menunjukkan bahwa *LSTM* masih relevan digunakan dalam analisis sentimen pada domain wisata. Namun, pada penelitian ini, hasil evaluasi menunjukkan bahwa performa *LSTM* masih berada sedikit di bawah *BiLSTM* (Oktaviana et al., 2021). Untuk memperjelas perbandingan performa kedua model, nilai evaluasi *LSTM* dan *BiLSTM* divisualisasikan dalam bentuk grafik pada Gambar 7.

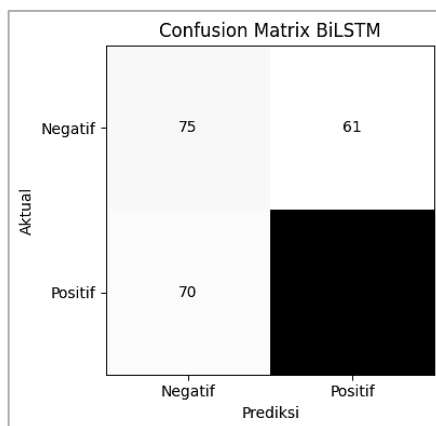


Gambar 7. Grafik Perbandingan Performa Model *LSTM* dan *BiLSTM*.

Berdasarkan Gambar 7, *BiLSTM* memperoleh nilai evaluasi sedikit lebih tinggi dibandingkan *LSTM* pada seluruh metrik. Namun, selisih performa kedua model relatif kecil, sehingga hasil perbandingan perlu ditafsirkan secara hati-hati. Grafik tersebut digunakan untuk memperjelas perbedaan nilai evaluasi kedua model yang diuji dengan *dataset*, skema pembagian data, parameter pelatihan, dan metrik evaluasi yang sama. Hasil *confusion matrix* masing-masing model ditunjukkan pada Gambar 8 dan Gambar 9.



Gambar 8. *Confusion Matrix* Model *LSTM*



Gambar 9. *Confusion Matrix* Model *BiLSTM*

Berdasarkan Gambar 8 dan Gambar 9, *confusion matrix* memperlihatkan jumlah prediksi benar dan salah pada kelas positif maupun negatif. Secara umum, kedua model masih dipengaruhi oleh distribusi data yang tidak seimbang. Jumlah ulasan positif yang lebih dominan membuat model cenderung lebih mudah mengenali kelas positif



dibandingkan kelas negatif. Oleh karena itu, interpretasi hasil evaluasi perlu mempertimbangkan karakteristik distribusi data. Perbedaan model dalam analisis sentimen dapat menghasilkan performa yang tidak sama karena dipengaruhi oleh karakteristik data ulasan yang digunakan (Ramadhan & Kusnawi, 2026).

3.7 Pembahasan

Hasil penelitian ini juga sejalan dengan penelitian sebelumnya yang menerapkan pengembangan *BiLSTM* berbasis *attention* dan *LoRA* pada analisis sentimen ulasan destinasi wisata. Meskipun penelitian tersebut menggunakan arsitektur *BiLSTM* yang telah dikembangkan, sedangkan penelitian ini menggunakan *LSTM* dan *BiLSTM* standar sebagai model pembandingan, keduanya sama-sama menunjukkan bahwa *BiLSTM* relevan digunakan dalam analisis sentimen pada domain wisata (Kusuma et al., 2026).

Secara keseluruhan, hasil penelitian menunjukkan bahwa *Bidirectional Long Short-Term Memory* memperoleh performa sedikit lebih tinggi dibandingkan *Long Short-Term Memory* pada *dataset* yang digunakan. Namun, selisih performa yang kecil menunjukkan bahwa perbedaan kedua model belum terlalu kuat. Hasil ini mengindikasikan bahwa pemrosesan dua arah pada *Bidirectional Long Short-Term Memory* dapat membantu model memahami konteks ulasan, tetapi karakteristik data yang tidak seimbang tetap memengaruhi kemampuan model dalam mengenali kelas negatif.

Dari sisi analisis aspek, ulasan positif paling banyak muncul pada aspek umum, sedangkan ulasan negatif paling banyak ditemukan pada aspek kebersihan. Temuan ini menunjukkan bahwa pengalaman umum pengunjung cenderung dinilai positif, tetapi kebersihan masih menjadi bagian yang perlu mendapat perhatian lebih dari pengelola destinasi wisata.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, analisis sentimen ulasan wisata dapat digunakan untuk mengetahui kecenderungan opini pengunjung sekaligus memetakan aspek yang banyak memperoleh tanggapan positif dan negatif. *Dataset* akhir yang digunakan berjumlah 2.702 ulasan, dengan distribusi sentimen yang tidak seimbang antara kelas positif dan negatif. Untuk mengurangi pengaruh ketidakseimbangan tersebut, pelatihan model menerapkan *class weighting*. Hasil evaluasi menunjukkan bahwa *Bidirectional Long Short-Term Memory* memperoleh performa sedikit lebih tinggi dibandingkan *Long Short-Term Memory* pada seluruh metrik evaluasi. *Bidirectional Long Short-Term Memory* memperoleh *accuracy* sebesar 75,79%, *precision* 76,33%, *recall* 75,79%, dan *F1-score* 76,04%, sedangkan *Long Short-Term Memory* memperoleh *accuracy* sebesar 75,05%, *precision* 75,48%, *recall* 75,05%, dan *F1-score* 75,25%. Namun, selisih performa kedua model relatif kecil, sehingga hasil tersebut perlu ditafsirkan secara hati-hati. Hasil evaluasi per kelas dan *confusion matrix* menunjukkan bahwa kedua model masih lebih kuat dalam mengenali kelas positif dibandingkan kelas negatif. Dari sisi aspek, ulasan positif paling banyak ditemukan pada aspek umum, sedangkan ulasan negatif paling banyak muncul pada aspek kebersihan. Temuan ini menunjukkan bahwa pengalaman umum pengunjung cenderung baik, tetapi kebersihan masih perlu menjadi prioritas perbaikan bagi pengelola destinasi wisata. Penelitian selanjutnya dapat mempertimbangkan teknik penyeimbangan data lain, ekstraksi aspek otomatis, serta model berbasis *attention* atau *transformer* untuk meningkatkan kemampuan klasifikasi, terutama pada kelas negatif.

REFERENCES

- Af'idah, D. I., Anggraeni, P. D., Rizki, M., Setiawan, A. B., & Handayani, S. F. (2023). Aspect-Based Sentiment Analysis for Indonesian Tourist Attraction Reviews Using Bidirectional Long Short-Term Memory. *JUITA: Jurnal Informatika*, 11(1), 27–36. <https://doi.org/10.30595/juita.v11i1.15341>
- Asnawiyah, & Putra, R. E. (2024). Perbandingan Algoritma LSTM dan BiLSTM untuk Analisis Sentimen Multi-Class Media Sosial Twitter. *Journal of Informatics and Computer Science (JINACS)*, 6(3), 778–786. <https://doi.org/10.26740/jinacs.v6n03.p778-786>
- Aufar, A. F., Rosid, M. A., Eviyanti, A., & Astutik, I. R. I. (2023). Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews. *JICTE (Journal of Information and Computer Technology Education)*, 7(2), 42–50. <https://doi.org/10.21070/jicte.v7i2.1650>
- Bahri, C. A., & Suadaa, L. H. (2023). Aspect-Based Sentiment Analysis in Bromo Tengger Semeru National Park Indonesia Based on Google Maps User Reviews. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 17(1), 79–90. <https://doi.org/10.22146/ijccs.77354>
- Cahyaningtyas, S., Fudholi, D. H., & Hidayatullah, A. F. (2021). Deep Learning for Aspect-Based Sentiment Analysis on Indonesian Hotels Reviews. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 6(3), 239–248. <https://doi.org/10.22219/kinetik.v6i3.1300>
- Cherradi, M., & El Haddadi, A. (2024). Comparative Analysis of Machine Learning Algorithms for Sentiment Analysis in Film Reviews. *Acadlore Transactions on AI and Machine Learning*, 3(3), 137–147. <https://doi.org/10.56578/ataiml030301>
- Febrianto, D. C., Fitriani, M. A., Afrad, M., & Khadija, M. A. (2024). Aspect Based Sentiment Analysis Menggunakan IndoBERT Model terhadap Review Pengunjung Objek Wisata Baturraden. *Melek IT: Information Technology Journal*, 10(2), 157–166. <https://doi.org/10.30742/melekitjournal.v10i2.358>
- Gunawan, A. R., & Aziza, R. F. A. (2025). Sentiment Analysis Using LSTM Algorithm Regarding Grab Application



- Services in Indonesia. *Journal of Applied Informatics and Computing*, 9(2). <https://doi.org/10.30871/jaic.v9i2.8696>
- Ipmawati, J., Saifulloh, & Kusnawi. (2024). Sentiment Analysis of Tourist Attractions Based on Reviews on Google Maps Using the Support Vector Machine Algorithm. *Analisis Sentimen Tempat Wisata Berdasarkan Ulasan Pada Google Maps Menggunakan Algoritma Support Vector Machine*, 4(1), 247–256. <https://doi.org/10.57152/malcom.v4i1.1066>
- Jayanto, R., Kusumaningrum, R., & Wibowo, A. (2022). Aspect-Based Sentiment Analysis for Hotel Reviews Using an Improved Model of Long Short-Term Memory. *International Journal of Advances in Intelligent Informatics*, 8(3), 391–403. <https://doi.org/10.26555/ijain.v8i3.691>
- Kusuma, T. B., Widiartha, I. M., Suhartana, I. K. G., & Putra, I. G. N. A. C. (2026). Implementasi Attention-Based BiLSTM dengan LoRA Parameter Tuning untuk Analisis Sentimen Ulasan Destinasi Wisata. *Techno.Com*, 25(1), 13–29. <https://doi.org/10.62411/tc.v25i1.15089>
- Malik, A., & Irawan, B. (2026). Penerapan Algoritma Bi-LSTM dengan Optimasi Threshold Adjustment untuk Analisis Sentimen Ulasan Aplikasi Mobile JKN. *JITET (Jurnal Informatika Dan Teknik Elektro Terapan)*, 14(1), 945–953. <https://doi.org/10.23960/jitet.v14i1.8879>
- Maulana, R., Voutama, A., & Ridwan, T. (2023). Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC. *Jurnal Teknologi Terpadu*, 9(1), 42–48. <https://doi.org/10.54914/jtt.v9i1.609>
- Maulida, A., Irawan, B., & Ramdhan, N. A. (2026). Analisis Sentimen Ulasan Wisata Alun-Alun Brebes pada Google Maps Menggunakan Support Vector Machine. *TIN: Terapan Informatika Nusantara*, 6(8), 1398–1406. <https://doi.org/10.47065/tin.v6i8.8883>
- Nasution, K., Saddami, K., Roslidar, R., Akhyar, A., Fathurrahman, F., & Aulia, N. (2025). Comparative Study of BiLSTM and GRU for Sentiment Analysis on Indonesian E-Commerce Product Reviews Using Deep Sequential Modeling. *Jurnal Teknik Informatika (JUTIF)*, 6(4), 1881–1896. <https://doi.org/10.52436/1.jutif.2025.6.4.4878>
- Oktapiani, S., Gunawansyah, & Aulia, B. E. (2025). Perbandingan Algoritma LSTM dan BiLSTM dalam Analisis Sentimen Data Ulasan Hotel. *Prosiding Seminar Sosial Politik, Bisnis, Akuntansi Dan Teknik (SoBAT) Ke-7*, 500–509. <https://doi.org/10.32897/sobat.2025.7.1.5285>
- Oktaviana, N., Rustamaji, H. C., & Sofyan, H. (2021). Sentiment Analysis on Reviews of Beach Tourism Objects on Google Maps Using Long-Short Term Memory Method. *Seminar Nasional Informatika (SEMNASIF)*, 133–143. <https://garuda.kemdiktisaintek.go.id/documents/detail/2708540>
- Ramadhan, T., & Kusnawi, K. (2026). Performance Comparison of IndoBERT and Bi-LSTM Models for Sentiment Analysis of Shopee App Users in Indonesia. *Journal of Applied Informatics and Computing*, 10(2), 2076–2085. <https://doi.org/10.30871/jaic.v10i2.10996>
- Sudiatmika, I. P. G. A., Saputra, P. S., Rahardian, R. L., & Dewi, K. H. S. (2025). Sentiment Analysis of Tourist Reviews on Google Maps for Pura Besakih Using Machine Learning Algorithms. *Jurnal Mandiri IT*, 14(1), 149–158. <https://doi.org/10.35335/mandiri.v14i1.449>
- Suhaeni, C., Kamila, S. A., Fahira, F., & Yusran, M. (2024). Sentiment Classification on the 2024 Indonesian Presidential Candidate Dataset Using Deep Learning Approaches. *Indonesian Journal of Statistics and Its Applications*, 8(4), 567–580. <https://doi.org/10.29244/ijsa.v8i4.1259>
- Tamami, G., Triyanto, W. A., & Muzid, S. (2025). Sentiment Analysis Mobile JKN Reviews Using SMOTE Based LSTM. *Indonesian Journal of Computing and Cybernetics Systems*, 19(1), 13–24. <https://doi.org/10.22146/ijccs.101910>