

Analisis Sentimen Ulasan Aplikasi Duolingo di Google Play Store Menggunakan Algoritma *Bidirectional Long Short-Term Memory* (Bi-LSTM)

Muhammad Nasrullah ^{1*}, Abdul Azis ², Intan Mila Hakim ³

¹⁻³ Teknologi Informasi, Institut Teknologi Dan Kesehatan Aspirasi, Indonesia

*Penulis Korespondensi : muhammadnasrullahsib@gmail.com

Abstract. Reviews posted by users on the Google Play Store provide valuable feedback that can be utilized to measure user satisfaction and assess the quality of mobile applications. However, the growing volume of reviews makes manual evaluation increasingly impractical, highlighting the need for automated sentiment analysis techniques. This research proposes the use of the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm to classify the sentiment of Indonesian-language reviews for the Duolingo application. The review dataset was obtained through web scraping from the Google Play Store and underwent several preprocessing steps, including case folding, text cleaning, word normalization, tokenization, stopword removal, and stemming. After preprocessing, the data were divided into 80% training data and 20% testing data for model development and performance evaluation. The effectiveness of the model was measured using accuracy, precision, recall, and F1-score. The experimental results yielded an accuracy of 94.05%, precision of 88.46%, recall of 92.15%, and an F1-score of 91.17%. These findings indicate that the Bi-LSTM model is capable of capturing sentiment patterns with a high level of reliability, although its ability to classify negative reviews is still influenced by the imbalance between sentiment classes. Overall, the study confirms that Bi-LSTM is a suitable deep learning approach for sentiment classification of application reviews and offers meaningful insights that can support Duolingo developers in evaluating user opinions and enhancing application quality.

Keywords: Bi-LSTM ; Deep Learning; Duolingo; Sentiment Analysis; Text Mining.

Abstrak. Ulasan dari pengguna di Google Play Store menjadi salah satu bentuk umpan balik yang dapat dimanfaatkan untuk mengetahui tingkat kepuasan serta mengevaluasi kualitas suatu aplikasi. Banyaknya jumlah ulasan yang terus bertambah menjadikan proses analisis manual menjadi kurang efisien, sehingga diperlukan metode otomatis melalui analisis sentimen. Penelitian ini bertujuan menerapkan algoritma *Bidirectional Long Short-Term Memory* (Bi-LSTM) untuk mengklasifikasikan sentimen pada ulasan pengguna aplikasi Duolingo berbahasa Indonesia. Data penelitian dikumpulkan melalui teknik web scraping dari Google Play Store, kemudian diproses menggunakan tahapan *preprocessing* yang meliputi *case folding*, *cleaning*, normalisasi kata, tokenisasi, *stopword removal*, dan *stemming*. Dataset selanjutnya dibagi menjadi 80% data latih dan 20% data uji sebagai dasar dalam membangun serta mengevaluasi model. Evaluasi performa dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Berdasarkan hasil pengujian, model Bi-LSTM memperoleh nilai *accuracy* sebesar 94,05%, *precision* 88,46%, *recall* 92,15%, dan *F1-score* 91,17%. Hasil tersebut menunjukkan bahwa Bi-LSTM memiliki kemampuan yang baik dalam mengenali pola sentimen pengguna dengan tingkat ketepatan yang tinggi, meskipun kinerja klasifikasi pada sentimen negatif masih dipengaruhi oleh ketidakseimbangan distribusi data. Secara keseluruhan, hasil penelitian menunjukkan bahwa Bi-LSTM efektif digunakan untuk menganalisis sentimen ulasan aplikasi dan dapat dimanfaatkan sebagai sumber informasi bagi pengembang Duolingo dalam memahami persepsi pengguna serta mendukung peningkatan kualitas layanan.

Kata kunci: Analisis Sentimen; Bi-LSTM; Duolingo; Text Mining, Deep Learning.

1. LATAR BELAKANG

Pemanfaatan teknologi digital telah mengubah cara masyarakat mempelajari bahasa asing. Berbagai aplikasi pembelajaran berbasis mobile kini menjadi pilihan karena menawarkan fleksibilitas, kemudahan akses, dan metode belajar yang interaktif (Ginayah et al., 2024). Salah satu aplikasi yang sering digunakan adalah Duolingo, yaitu platform

pembelajaran bahasa yang menerapkan konsep gamification melalui sistem level, poin, dan penghargaan untuk meningkatkan motivasi belajar. Jumlah pengguna yang tinggi menghasilkan ribuan ulasan di Google Play Store yang berisi pengalaman, kritik, saran, maupun penilaian terhadap kualitas aplikasi.

Ulasan yang ditulis pengguna memberikan informasi yang penting bagi pengembang untuk mengetahui tingkat kepuasan, mengidentifikasi kendala, serta memperoleh masukan dalam pengembangan aplikasi. Namun, Tingginya jumlah ulasan yang tersedia mengakibatkan proses analisis secara manual menjadi lebih lambat dan kurang optimal karena memerlukan waktu yang lama dan rentan terhadap subjektivitas (Kevin et al., 2024). Oleh sebab itu, diperlukan teknik analisis sentimen yang mampu mengelompokkan opini pengguna secara otomatis sehingga informasi yang diperoleh dapat dimanfaatkan sebagai dasar pengambilan keputusan (Khoirul et al., 2023).

Analisis sentimen merupakan salah satu bidang Natural Language Processing (NLP) yang bertujuan mengidentifikasi dan mengklasifikasikan opini ke dalam kategori positif, negatif, maupun netral (Husna & Margareth, 2024). Berbagai penelitian telah menerapkan algoritma machine learning seperti Naïve Bayes, Support Vector Machine (SVM), dan Random Forest untuk klasifikasi sentimen (Maulana et al., 2024). Meskipun menghasilkan performa yang cukup baik, metode tersebut masih memiliki keterbatasan dalam memahami hubungan antar kata dan konteks kalimat karena menggunakan representasi bag-of-words, terutama pada ulasan berbahasa Indonesia yang banyak mengandung bahasa tidak baku, singkatan, dan campuran istilah asing (Wicaksono & Nastiti, 2024). Perkembangan deep learning memberikan alternatif yang lebih baik melalui algoritma Bidirectional Long Short-Term Memory (Bi-LSTM) (Sari et al., 2025). Berbeda dengan LSTM satu arah, Bi-LSTM memproses urutan kata dari arah maju (forward) dan mundur (backward) sehingga mampu menghasilkan representasi konteks kalimat yang lebih lengkap (Anas et al., 2023). Berbagai penelitian dalam sepuluh tahun terakhir menunjukkan bahwa Bi-LSTM memiliki performa yang lebih baik dibandingkan metode konvensional pada analisis sentimen media sosial, aplikasi digital, maupun e-commerce (Wangsajaya et al., 2025).

Meskipun demikian, penelitian mengenai analisis sentimen ulasan aplikasi Duolingo di Google Play Store, khususnya menggunakan ulasan berbahasa Indonesia, masih relatif terbatas. Selain itu, karakteristik ulasan yang didominasi bahasa informal, singkatan, serta campuran bahasa Indonesia dan Inggris menjadi tantangan tersendiri dalam proses klasifikasi. Kondisi tersebut menunjukkan adanya research gap yang menjadi dasar dilakukannya penelitian ini.

Kebaruan (novelty) penelitian terletak pada penerapan algoritma Bidirectional Long Short-Term Memory (Bi-LSTM) untuk mengklasifikasikan sentimen ulasan aplikasi Duolingo yang diperoleh melalui web scraping Google Play Store dengan tahapan preprocessing meliputi case folding, cleaning, normalisasi, tokenisasi, stopword removal, dan stemming (Yoga & Andreas, 2026). Penelitian ini bertujuan mengimplementasikan Bi-LSTM untuk mengklasifikasikan sentimen ulasan pengguna serta mengevaluasi kinerja model menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian diharapkan dapat memberikan kontribusi terhadap pengembangan analisis sentimen berbasis deep learning serta menjadi masukan bagi pengembang Duolingo dalam meningkatkan kualitas aplikasi berdasarkan persepsi pengguna.

2. KAJIAN TEORITIS

Analisis Sentimen

Analisis sentimen merupakan salah satu cabang Natural Language Processing (NLP) yang berfokus pada proses identifikasi dan pengelompokan opini atau penilaian seseorang yang terdapat dalam data teks. Tujuan utama analisis sentimen adalah mengetahui kecenderungan suatu opini ke dalam kategori positif, negatif, atau netral (Oktaria & Arif, 2021). Seiring meningkatnya penggunaan media digital, analisis sentimen banyak dimanfaatkan untuk mengolah ulasan pengguna, komentar media sosial, maupun survei daring sebagai bahan evaluasi terhadap suatu produk atau layanan.

Analisis sentimen pada penelitian ini diterapkan untuk mengidentifikasi persepsi pengguna terhadap aplikasi Duolingo berdasarkan ulasan yang dipublikasikan di Google Play Store. Hasil klasifikasi diharapkan dapat memberikan gambaran mengenai tingkat kepuasan pengguna serta menjadi sumber informasi yang bermanfaat dalam pengembangan aplikasi.

Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah cabang ilmu yang mengintegrasikan ilmu komputer, kecerdasan buatan, dan linguistik untuk memungkinkan komputer mengerti, mengolah, serta menganalisis bahasa manusia. NLP memiliki berbagai penerapan, seperti penerjemahan otomatis, chatbot, peringkasan dokumen, pengenalan suara, hingga analisis sentimen (Rivaldi & Wismarini, 2024).

Pada analisis sentimen, NLP berfungsi untuk mengubah data teks yang tidak terstruktur menjadi bentuk yang dapat diolah oleh model komputasi. Supaya informasi yang terdapat dalam teks dapat dikenali dengan lebih baik, diperlukan beberapa tahapan prapemrosesan (preprocessing) sebelum data digunakan dalam proses pelatihan model.

Web Scraping Data

Web scraping merupakan teknik pengumpulan data secara otomatis dari sebuah situs web menggunakan bantuan program atau library tertentu. Cara ini banyak digunakan dalam berbagai penelitian karena mampu mengumpulkan data dalam jumlah yang besar dengan waktu yang relatif singkat dibandingkan pengumpulan data secara manual. Pada penelitian analisis sentimen, web scraping sering digunakan untuk mengumpulkan ulasan dari berbagai platform digital, seperti Google Play Store, media sosial, maupun situs e-commerce (Alpin et al., 2023).

Dalam penelitian ini, data di dapatkan dengan proses web scraping pada halaman aplikasi Duolingo di Google Play Store dengan library google-play-scraper. Library tersebut memungkinkan pengambilan informasi ulasan, seperti isi komentar, rating, tanggal ulasan, dan informasi pendukung lainnya secara otomatis. Data yang berhasil dikumpulkan selanjutnya diseleksi dan dibersihkan sebelum memasuki tahap preprocessing dan proses klasifikasi sentimen. Penggunaan web scraping memberikan beberapa keuntungan, antara lain proses pengumpulan data yang lebih cepat, efisien, serta mampu menghasilkan dataset yang sesuai dengan kondisi aktual. Dengan demikian, data yang digunakan dalam penelitian dapat merepresentasikan opini pengguna secara lebih objektif dan mendukung proses analisis sentimen yang lebih akurat..

Preprocessing Teks

Preprocessing merupakan tahap penting dalam analisis sentimen karena kualitas data masukan akan memengaruhi hasil klasifikasi. Tahapan ini bertujuan mengurangi noise serta menyeragamkan bentuk teks agar lebih mudah dipelajari oleh model.

Beberapa tahapan preprocessing yang digunakan dalam Tahapan yang dilakukan dalam penelitian ini meliputi case folding, yaitu mengubah seluruh teks menjadi huruf kecil, serta cleaning yang bertujuan menghapus URL, angka, tanda baca, dan karakter lain yang tidak diperlukan, normalisasi, untuk mengganti kata tidak baku menjadi baku, tokenisasi, yaitu memecah kalimat menjadi kumpulan kata, stopwords removal, untuk menghilangkan kata-kata yang tidak memiliki makna penting serta stemming, yaitu mengembalikan kata berimbuhan ke bentuk dasar menggunakan pustaka PySastrawi (Kurniawan et al., 2023).

Bidirectional Long Short-Term Memory (Bi-LSTM)

Bidirectional Long Short-Term Memory (Bi-LSTM) adalah pengembangan dari algoritma Long Short-Term Memory (LSTM), yaitu salah satu arsitektur Recurrent Neural Network (RNN) yang dikembangkan untuk mengatasi permasalahan vanishing gradient pada data sekuensial. Berbeda dengan LSTM konvensional yang hanya memproses data dari satu arah, Bi-LSTM melakukan pemrosesan data dalam dua arah, yaitu dari awal ke akhir (forward)

dan dari akhir ke awal (backward) (Lubis et al., 2024). Pendekatan ini memungkinkan model memperoleh informasi konteks yang lebih lengkap sehingga hubungan antar kata dapat dipahami dengan lebih baik.

Kemampuan tersebut menjadikan Bi-LSTM banyak dimanfaatkan dalam berbagai tugas pengolahan bahasa alami, terutama analisis sentimen, karena model dapat mengenali makna suatu kata berdasarkan konteks kalimat secara keseluruhan.

Google Play Store sebagai Sumber Data

Google Play Store merupakan platform distribusi aplikasi berbasis Android yang menyediakan fasilitas bagi pengguna untuk memberikan penilaian (rating) dan ulasan setelah menggunakan suatu aplikasi. Ulasan tersebut mencerminkan pengalaman nyata pengguna sehingga menjadi sumber data yang bernilai dalam analisis sentimen. Selain membantu mengetahui tingkat kepuasan pengguna, hasil dari analisis ulasan juga dapat dimanfaatkan untuk mengidentifikasi, kekurangan aplikasi serta menentukan prioritas pengembangan fitur pada versi berikutnya.

Penelitian Terdahulu

Berbagai penelitian tentang analisis sentimen telah menerapkan beragam metode klasifikasi. Beberapa penelitian memanfaatkan algoritma Naïve Bayes, Support Vector Machine (SVM), dan Random Forest untuk mengklasifikasikan opini pengguna pada aplikasi maupun media sosial (Tarigan et al., 2025). Meskipun menghasilkan performa yang cukup baik, Metode tersebut masih mengalami kendala dalam memahami relasi antar kata karena menggunakan representasi bag-of-words (Mudya et al., 2024).

Seiring berkembangnya teknologi deep learning, algoritma LSTM dan Bidirectional LSTM (Bi-LSTM) mulai banyak diterapkan dalam analisis sentimen. Nurashila et al., (2023) menghasilkan bahwa LSTM dapat memperbaiki performa klasifikasi sentimen dibandingkan beberapa metode machine learning. Penelitian Alghifari et al., (2022) juga membuktikan bahwa Bi-LSTM menghasilkan performa yang lebih baik dari LSTM satu arah pada analisis sentimen layanan digital. Sementara itu, Wungguli et al., (2025) mengombinasikan FastText dan Bi-LSTM sehingga memperoleh akurasi klasifikasi yang tinggi pada analisis sentimen ulasan aplikasi.

Mengacu pada hasil penelitian tersebut, Bi-LSTM dapat disimpulkan memiliki kemampuan yang baik dalam memahami konteks bahasa sehingga sesuai digunakan untuk analisis sentimen. Namun, mayoritas penelitian terlalu berfokus pada media sosial, layanan transportasi daring, maupun aplikasi e-commerce. Penelitian yang mengkaji sentimen pengguna aplikasi Duolingo menggunakan ulasan berbahasa Indonesia pada Google Play Store

masih terbatas. Untuk menjawab kesenjangan tersebut, penelitian ini menerapkan algoritma Bi-LSTM pada data ulasan Duolingo dengan harapan dapat mendukung pengembangan analisis sentimen berbasis deep learning sekaligus menghasilkan informasi yang bermanfaat untuk developer aplikasi.

3. METODE PENELITIAN

Penelitian ini memanfaatkan metode analisis sentimen dalam mengevaluasi pendapat pengguna pada aplikasi Duolingo berdasarkan ulasan yang ada di Google Play Store. Proses klasifikasi dilakukan menggunakan algoritma Bidirectional Long Short-Term Memory (Bi-LSTM) yang mempunyai kelebihan dalam memahami konteks kalimat secara lebih baik. Dataset penelitian dikumpulkan melalui teknik web scraping dengan total 71.141 ulasan, sehingga menghasilkan data yang luas dan representatif. Jumlah data tersebut diharapkan mampu menggambarkan secara lebih komprehensif tentang pola sentimen pengguna dan mendukung peningkatan akurasi hasil klasifikasi.

Analisis sentimen terhadap ulasan pada aplikasi Duolingo pada Google Play Store diterapkan melalui pendekatan kuantitatif dengan metode eksperimen. Pendekatan ini bertujuan mengevaluasi kemampuan algoritma Bidirectional Long Short-Term Memory (Bi-LSTM) untuk klasifikasi sentimen pengguna secara otomatis. Pengumpulan data dilakukan menggunakan teknik web scraping melalui library google-play-scraper pada Python untuk memperoleh ulasan pengguna secara langsung dari Google Play Store. Selanjutnya, data yang berhasil dikumpulkan melalui proses seleksi dengan menghapus data duplikat, data kosong, serta ulasan yang tidak relevan sehingga diperoleh dataset yang siap digunakan pada tahap pengolahan berikutnya.

Tahap berikutnya adalah preprocessing teks yang bertujuan mempersiapkan data agar lebih siap digunakan dalam proses klasifikasi. Proses ini terdiri atas case folding dengan mengubah seluruh huruf menjadi huruf kecil, cleaning untuk menghilangkan URL, angka, tanda baca, dan karakter yang tidak diperlukan, normalisasi guna mengubah kata tidak baku menjadi bentuk baku, tokenisasi untuk memecah kalimat menjadi token-token kata, stopwords removal untuk menghapus kata yang tidak memberikan informasi penting, serta stemming menggunakan PySastrawi untuk mengembalikan kata imbuhan ke bentuk dasarnya. Setelah proses preprocessing selesai, data kemudian dikonversi ke bentuk numerik menggunakan Tokenizer dan Padding Sequence dari TensorFlow Keras agar dapat diproses oleh model deep learning.

Model klasifikasi dikembangkan dengan arsitektur Bidirectional Long Short-Term Memory (Bi-LSTM) yang terdiri atas Embedding Layer, Bidirectional LSTM, Dropout Layer, dan Dense Layer. Model dilatih dengan optimizer Adam dengan loss function `sparse_categorical_crossentropy`. Selanjutnya, dataset dibagi menjadi 80% data latih dan 20% data uji. Data latih dimanfaatkan untuk proses pembentukan model, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan sentimen pada data yang belum pernah digunakan selama proses pelatihan.

Tahapan terakhir dalam penelitian ini adalah mengevaluasi performa model menggunakan Confusion Matrix yang diukur berdasarkan metrik Accuracy, Precision, Recall, dan F1-Score. Keempat metrik tersebut digunakan untuk mengukur tingkat ketepatan model dalam mengklasifikasikan sentimen pada ulasan pengguna aplikasi Duolingo. Berdasarkan hasil evaluasi tersebut kemudian dilakukan analisis dan pembahasan untuk mengetahui performa algoritma Bi-LSTM dalam analisis sentimen serta menarik kesimpulan sesuai tujuan penelitian.

4. HASIL DAN PEMBAHASAN

Hasil Preprocessing Data

Tahap preprocessing merupakan proses penting pada analisis sentimen karena bertujuan meningkatkan kualitas data teks sebelum digunakan pada proses pelatihan model Bidirectional Long Short-Term Memory (Bi-LSTM). Ulasan pengguna yang didapatkan dari Google Play Store masih terdapat noise, seperti penggunaan huruf besar yang bervariasi atau tidak konsisten, tanda baca, angka, URL, emoji, kata tidak baku, singkatan, serta campuran bahasa Indonesia dan bahasa Inggris. Apabila data tersebut digunakan secara langsung, kinerja model dapat menurun karena banyaknya informasi yang tidak relevan.

Proses preprocessing pada penelitian ini dilakukan melalui beberapa tahapan, yaitu case folding, cleaning, normalisasi, tokenisasi, stopword removal, dan stemming. Proses case folding berfungsi menyamakan seluruh teks menjadi huruf kecil dengan makna yang sama tidak dianggap sebagai token yang berbeda. Selanjutnya, proses cleaning menghilangkan kata yang tidak dibutuhkan, seperti URL, angka, simbol, emoji, dan tanda baca sehingga hanya informasi penting yang dipertahankan.

Tahap normalisasi diterapkan untuk merubah kata-kata tidak baku, singkatan, maupun bahasa percakapan menjadi baku baku. Proses ini sangat penting karena ulasan pada Google Play Store umumnya menggunakan bahasa informal, misalnya kata "gk", "ga", atau "bgt" yang diubah menjadi bentuk baku agar lebih mudah dikenali oleh model. Setelah itu, tokenisasi

memecah kalimat menjadi kumpulan kata (token) sehingga setiap kata dapat diproses secara individual. Tahap stopwords removal kemudian menghapus kata-kata yang tidak mempunyai pengaruh besar pada penentuan sentimen, sedangkan stemming mengubah kata berimbuhan diubah ke bentuk dasar sehingga kata-kata yang mempunyai makna sama dapat direpresentasikan sebagai satu token.

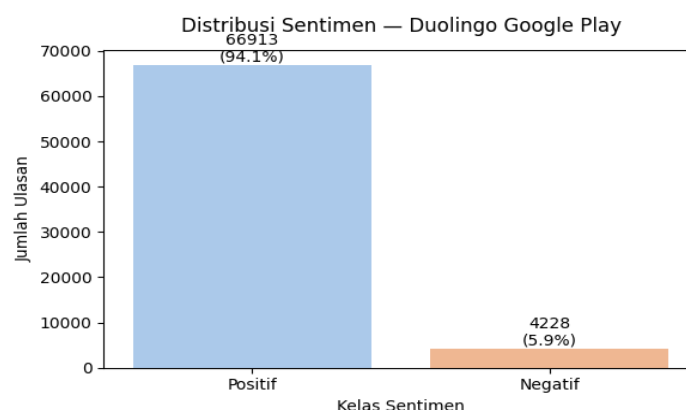
Secara keseluruhan, hasil preprocessing menghasilkan data yang lebih bersih, terstruktur, dan konsisten dibandingkan data mentah. Berkurangnya noise serta meningkatnya kualitas representasi teks memungkinkan model Bi-LSTM lebih mudah mempelajari pola hubungan antar kata dan konteks kalimat. Dengan demikian, proses preprocessing berperan penting dalam mendukung keberhasilan klasifikasi sentimen serta meningkatkan efektivitas proses pelatihan model pada penelitian ini. Dibawah ini merupakan visualisasi sebelum preprocessing dan setelah preprocessing.



Gambar 1. Visualisasi Sebelum dan Setelah Preprocessing.

Pelabelan Sentimen

Tahap pelabelan sentimen dilakukan untuk mengelompokkan ulasan pengguna aplikasi Duolingo ke dalam dua kategori, yaitu positif dan negatif. Berdasarkan hasil pelabelan, diperoleh 66.913 ulasan positif dan 4.228 ulasan negatif, sehingga total dataset yang digunakan berjumlah 71.141 ulasan.



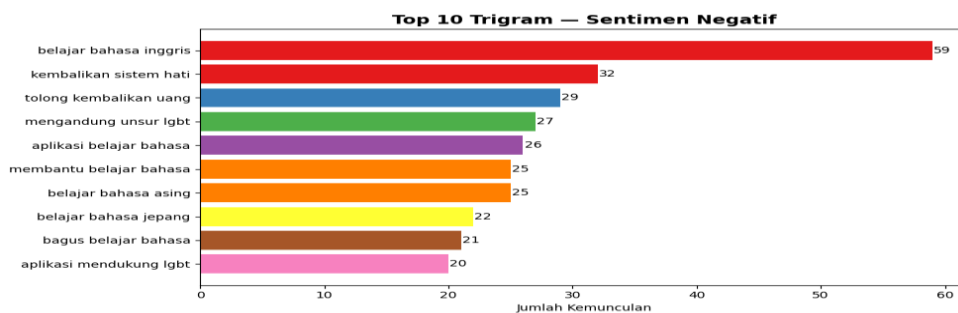
Gambar 2. Distribusi Pelabelan Sentimen Positif dan Negatif.

Distribusi tersebut memperlihatkan bahwa jumlah sentimen positif lebih dominan dibandingkan sentimen negatif. Hal ini menunjukkan bahwa sebagian besar pengguna memberikan penilaian yang baik pada aplikasi Duolingo, baik dari aspek kemudahan penggunaan, kualitas materi pembelajaran, maupun pengalaman belajar yang ditawarkan. Sementara itu, jumlah ulasan negatif yang relatif lebih sedikit menunjukkan bahwa hanya sebagian pengguna yang menyampaikan keluhan atau ketidakpuasan terhadap aplikasi.

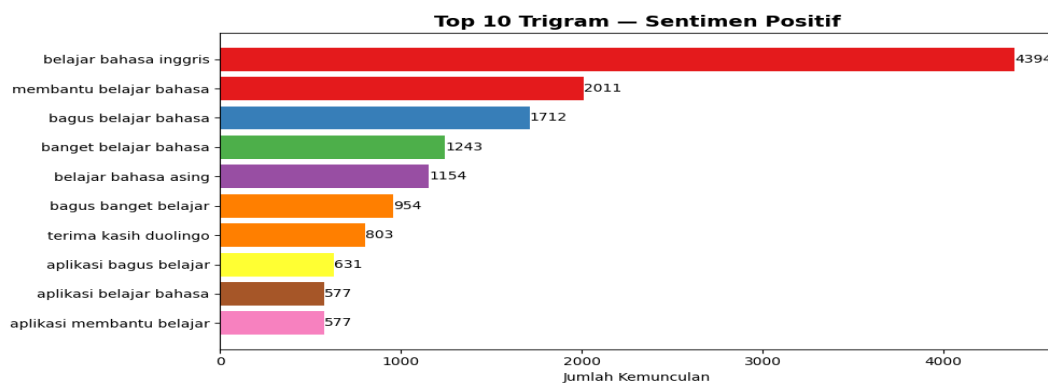
Di sisi lain, perbedaan jumlah data yang cukup besar antara kedua kelas menunjukkan adanya ketidakseimbangan data (class imbalance). Kondisi ini dapat memengaruhi proses pelatihan model karena algoritma cenderung lebih banyak mempelajari pola pada kelas mayoritas, yaitu sentimen positif. Oleh sebab itu, hasil evaluasi model perlu dianalisis secara menyeluruh menggunakan metrik seperti precision, recall, dan F1-score, sehingga kinerja klasifikasi tidak hanya dinilai berdasarkan nilai akurasi, maupun kemampuan model untuk mengenali pola kelas minoritas secara tepat.

Analisis N-Gram

Analisis N-Gram dilakukan untuk mengenali kata maupun kombinasi kata yang lebih sering muncul pada ulasan pengguna aplikasi Duolingo. Berbeda dengan analisis kata tunggal (unigram), N-Gram mampu menampilkan hubungan antar kata sehingga konteks yang terkandung dalam ulasan dapat dipahami dengan lebih baik.



Gambar 3. Top 10 Trigram Sentimen Positif.



Gambar 4. Top 10 Trigram Sentimen Negatif

Hasil Pengolahan Data Menggunakan Bi-LSTM

Salah satu tahapan utama dalam penelitian ini adalah melakukan pengujian terhadap model Bidirectional Long Short-Term Memory (Bi-LSTM) untuk mengetahui kemampuannya dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Duolingo di Google Play Store. Tahap pengujian menggunakan data uji yang telah melewati proses preprocessing dan pembagian dataset, kemudian dievaluasi menggunakan Confusion Matrix dengan metrik Accuracy, Precision, Recall, dan F1-Score. Hasil evaluasi tersebut memberikan gambaran mengenai kemampuan model dalam mengenali sentimen positif maupun negatif, sekaligus menunjukkan efektivitas algoritma Bi-LSTM dalam memahami konteks ulasan pengguna. Selanjutnya, hasil pengujian dianalisis untuk mengetahui kelebihan dan keterbatasan model dalam proses klasifikasi sentimen.

Tabel 1. Confusion Matrix Bi-LSTM.

Actual / Predicted	Positif	Negatif
Positif	13.267	115
Negatif	732	115

Berdasarkan confusion matrix, dari 13.382 ulasan positif, model berhasil mengklasifikasikan 13.267 ulasan dengan benar sebagai positif, sedangkan 115 ulasan salah diklasifikasikan sebagai negatif (false negative). Pada kelas negatif, dari 847 ulasan, model berhasil mengidentifikasi 115 ulasan sebagai negatif (true negative), sedangkan 732 ulasan yang salah diprediksi sebagai positif (false positive). Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang lebih baik dalam mengidentifikasi sentimen positif dibandingkan dengan sentimen negatif.

Tabel 2. Hasil Evaluasi.

Evaluasi	Persentase
Accuracy	94.05%
Precision	88.46%
Recall	92.15%
F1-Score	91.17%

Hasil pengujian menunjukkan bahwa algoritma Bidirectional Long Short-Term Memory (Bi-LSTM) mampu mengklasifikasikan sentimen ulasan pengguna aplikasi Duolingo dengan performa yang sangat baik. Hasil pengujian menunjukkan perolehan accuracy model sebesar 94,05%, yang menggambarkan bahwa mayoritas data uji telah diklasifikasikan dengan benar. Nilai akurasi yang tinggi mengindikasikan bahwa Bi-LSTM mampu mempelajari pola sentimen pada ulasan pengguna secara efektif setelah melalui tahapan preprocessing.

Selain itu, model menghasilkan precision sebesar 88,46%, yang menunjukkan bahwa mayoritas hasil prediksi sentimen yang dihasilkan model telah sesuai dengan kondisi sebenarnya. Nilai ini mengindikasikan bahwa model memiliki tingkat ketepatan yang baik dalam mengidentifikasi ulasan positif maupun negatif, meskipun masih ditemukan sejumlah data yang belum diklasifikasikan secara tepat.

Nilai recall sebesar 92,15% menunjukkan kemampuan model yang sangat baik dalam melakukan pengenalan terhadap data yang sesuai pada setiap kelas. Dengan nilai recall yang tinggi, sebagian besar ulasan yang termasuk dalam kelas sebenarnya berhasil dideteksi oleh model. Sementara itu, F1-Score sebesar 91,17% menunjukkan keseimbangan yang baik antara precision dan recall, sehingga model memiliki performa klasifikasi yang stabil dan tidak hanya bergantung pada satu metrik evaluasi.

Perbedaan kemampuan klasifikasi tersebut dipengaruhi oleh ketidakseimbangan jumlah data (class imbalance), di mana jumlah ulasan positif jauh lebih banyak dibandingkan ulasan negatif. Akibatnya, model cenderung mempelajari pola pada kelas mayoritas sehingga prediksi lebih sering mengarah pada sentimen positif. Meskipun demikian, nilai akurasi dan F1-Score yang tinggi menunjukkan bahwa algoritma Bi-LSTM tetap mampu menghasilkan performa klasifikasi yang baik dalam menganalisis sentimen ulasan aplikasi Duolingo.

Berdasarkan keseluruhan hasil penelitian, dapat diketahui bahwa algoritma Bidirectional Long Short-Term Memory (Bi-LSTM) efektif digunakan dalam analisis sentimen ulasan aplikasi pada Google Play Store. Kemampuan Bi-LSTM dalam memproses urutan kata pada dua arah (forward dan backward) memungkinkan model memahami konteks kalimat secara lebih baik, sehingga dapat menghasilkan klasifikasi sentimen dengan tingkat akurasi yang baik. Untuk penelitian selanjutnya, performa model masih dapat ditingkatkan melalui penyeimbangan dataset, hyperparameter tuning, maupun penggunaan word embedding yang lebih representatif, seperti FastText, Word2Vec, atau IndoBERT, agar kemampuan model dalam mengenali kelas minoritas menjadi lebih optimal.

5. KESIMPULAN DAN SARAN

Dari hasil penelitian yang telah dilakukan, algoritma Bidirectional Long Short-Term Memory (Bi-LSTM) mampu diterapkan untuk mengidentifikasi sentimen pada ulasan pengguna aplikasi Duolingo pada Google Play Store dengan performa yang sangat baik. Analisis confusion matrix memperlihatkan bahwa model mampu melakukan klasifikasi 13.267 dari 13.382 ulasan positif dengan benar, sedangkan pada kelas negatif model berhasil mengidentifikasi 115 dari 847 ulasan negatif. Evaluasi model menghasilkan accuracy sebesar

94,05%, precision 88,46%, recall 92,15%, dan F1-score 91,17%, yang menunjukkan bahwa Bi-LSTM memiliki kemampuan yang tinggi dalam memahami konteks ulasan dan mengklasifikasikan sentimen secara akurat. Meskipun demikian, kemampuan model dalam mengenali sentimen negatif masih dipengaruhi oleh ketidakseimbangan jumlah data (class imbalance), sehingga model cenderung lebih mudah mengenali kelas positif yang mendominasi dataset. Secara keseluruhan, hasil penelitian membuktikan bahwa algoritma Bi-LSTM efektif digunakan untuk analisis sentimen ulasan aplikasi Duolingo dan mampu memberikan informasi yang bermanfaat mengenai persepsi serta tingkat kepuasan pengguna.

Untuk meningkatkan performa model Pada penelitian mendatang, penggunaan dataset yang lebih besar disarankan agar lebih seimbang, menggunakan teknik penyeimbangan data seperti SMOTE, oversampling, atau undersampling, serta melakukan optimasi hyperparameter. Selain itu, pemanfaatan metode word embedding yang lebih representatif, seperti FastText, Word2Vec, atau IndoBERT, dapat dipertimbangkan agar kemampuan model dalam memahami konteks bahasa dan mengklasifikasikan sentimen, terutama pada kelas minoritas, menjadi lebih optimal.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan, bantuan, dan kontribusi selama proses penelitian ini, sehingga penelitian dapat diselesaikan dengan baik. Semoga hasil penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan dan menjadi referensi bagi penelitian selanjutnya.

DAFTAR REFERENSI

- Alghifari, D. R., Edi, M., & Firmansyah, L. (2022). Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia. *Jurnal Manajemen Informatika (JAMIKA)*, 12(2), 89–99. <https://doi.org/10.34010/jamika.v12i2.7764>
- Alpin, R. S., Alam, S., & Kurniawan, I. (2023). Analisis Sentimen Pengguna Aplikasi Jmo (Jamsostek Mobile) Pada Google Play Store Menggunakan Metode Naive Bayes. *Jurnal Ilmiah Teknik Dan Ilmu Komputer*, 2(3), 109–117. <https://doi.org/10.55123>
- Anas, F. H., Theopilus, B. S., & Arif, D. L. (2023). Perbandingan Kinerja LSTM, Bi-LSTM, dan GRU pada Klasifikasi Judul Berita Clickbait. *Indonesian Journal of Computer Science Attribution*, 12(4), 2136–2150. <https://doi.org/https://doi.org/10.33022/ijcs.v12i4.3281>
- Ginayah, R., Iswara, P. D., & Karlina, D. A. (2024). Pengembangan Media Interaktif Menggunakan Aplikasi Canva Untuk Keterampilan Membaca Permulaan. *Jurnal Pendidikan Madrasah Ibtidaiyah*, 8(4) 67-78, 1770. <https://doi.org/10.35931/am.v8i4.4066>

- Husna, J., & Margareth, D. A. W. (2024). Analisis Sentimen pada “Voices of History: 50 Iconic Speeches” Menggunakan Pendekatan Natural Language Processing Topic Modeling and Sentiment Analysis in “Voices of History: 50 Iconic Speeches” Using a Natural Language Processing Approach. *Jurnal Ilmiah Manajemen Informasi Dan Komunikasi*, 8(1), 15–24. <https://doi.org/10.20895/jimik.v8i1.page>
- Kevin, K., Enjeli, M., & Wijaya, A. (2024). Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes. *Jurnal Ilmiah Computer Science*, 2(2), 89–98. <https://doi.org/10.58602/jics.v2i2.24>
- Khoirul, M., Hayati, U., & Nurdyawan, O. (2023). Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes. *Jurnal Mahasiswa Teknik Informatika*, 2(7), 65–76. <https://doi.org/https://doi.org/10.36040/jati.v7i1.6373>
- Kurniawan, I., Lia. Hananto April, Shofia, H. S., Hananto, A., Priyatna, B., & Yuniar, R. A. (2023). Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter. *Jurnal Teknik Informatika Dan Sistem Informasi*, 10(1), 732–740. <https://doi.org/https://doi.org/10.35957/jatisi.v10i1.3582>
- Lubis, N., Siambaton, Mhd. Z., & Aulia, R. (2024). Implementasi Algoritma Deep Learning pada Aplikasi Speech to Text Online dengan Metode Recurrent Neural Network (RNN). *Sudo Jurnal Teknik Informatika*, 3(3), 113–126. <https://doi.org/10.56211/sudo.v3i3.583>
- Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N. (2024). Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 375–384. <https://doi.org/10.57152/malcom.v4i2.1206>
- Mudya Yolanda, A., & Tri Mulya, R. (2024). Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 6(2), 76–83. <https://doi.org/10.35580/variansiunm258>
- Nurashila, S. S., Hamami, F., & Kusumasari, T. F. (2023). Perbandingan Kinerja Algoritma Recurrent Neural Network (RNN) Dan Long Short-Term Memory (LSTM): Studi Kasus Prediksi Kemacetan Lalu Lintas Jaringan PT XYZ. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 8(3), 864–877. <https://doi.org/10.29100/jupi.v8i3.3961>
- Oktaria Sihombing, L., & Arif Dermawan, B. (2021). Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Pendidikan Informatika*, 5(2), 233–242. <https://doi.org/10.29408/edumatic.v5i2.4089>
- Rivaldi, R. C., & Wismarini, T. D. (2024). Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP). *Elkom: Jurnal Elektronika Dan Komputer*, 17(1), 120–128. <https://doi.org/10.51903/elkom.v17i1.1680>
- Sari, A. I., Hapsari, D. P., Wibowo, H. F. R., Putri, C. N., Lande, G. V. F., & Aldero, E. B. (2025). Implementasi Algoritma Pengklasifikasi Long Short-Term Memory (LSTM) untuk Data Time Series. *Prosiding Seminar Nasional Teknik Elektro, Sistem Informasi, Dan Teknik Informatika (SNESTIK) V*, 5, 653–666. Institut Teknologi Adhi Tama Surabaya. <https://doi.org/10.31284/p.snestik.2025.7034>
- Tarigan, D. A., Situmorang, Z., & Rosnelly, R. (2025). Analisis Sentimen Aplikasi Playstore Sirekap 2024 Pasca Pilpres Dengan Perbandingan Metode Support Vector Machine (SSVM), Naïve Bayes Classifier Dan Random Forest. *Jurnal Teknologi Informasi Dan*

Ilmu Komputer (JTIK), 11(3), 661–670.
<https://doi.org/https://doi.org/10.25126/jtiik.2025129608>

- Wangsajaya, Y. H. D., Setyowati, E., & Wibowo, A. (2025). Deteksi Emosi Teks X Berbahasa Indonesia Menggunakan Bi-LSTM dengan Seleksi Fitur Chi-Square. *Jurnal Algoritma*, 22(2), 546–555. <https://doi.org/10.33364/algoritma/v.22-2.2658>
- Wicaksono, B., & Nastiti, V. R. S. (2024). Analisis Sentimen dalam Opini Publik di Chanel Youtube Indonesia Lawyers Club Tentang Isu Populer dengan Menggunakan Metode LSTM dan Bi-LSTM. *Jurnal Algoritma*, 21(2), 241–251. <https://doi.org/10.33364/algoritma/v.21-2.1696>
- Wungguli, D., Yahya, N. I., Hasan, I. K., & Abdussamad, S. N. (2025). Implementasi Metode Bidirectional LSTM Dengan Word Embedding FastText Dalam Analisis Sentimen Ulasan Pengguna Aplikasi Maxim. *Jurnal Riset Mahasiswa Matematika*, 4(5), 204–217. <https://doi.org/10.18860/jrmm.v4i5.33358>
- Yoga, M. A. A., & Andreas, P. (2026). Optimasi Hyperparameter Bi-Directional Long Short Term Memory Menggunakan Particle Swarm Optimization Untuk Prediksi Saham BBRI. *Sienna*, 7(1), 21–36. <https://doi.org/10.47637/sienna.v7i1.2320>