Jurnal Riset Mahasiswa Matematika
Volume 4 (5) (2025), Pages 204-217
p-ISSN: 2808-1552; e-ISSN: 2808-4926

Implementasi Bidirectional LSTM Dengan Word Embedding FastText Dalam Analisis Sentimen Ulasan Pengguna Aplikasi Maxim

Hanz Franklyn Bachruddin Wewengkang[1], Djihad Wungguli[2], Nisky Imansyah Yahya[2*], Isran K. Hasan[1], and Siti Nurmardia Abdussamad[1]

[1]*Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Gorontalo, Gorontalo, Indonesia*
[2]*Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Gorontalo, Gorontalo, Indonesia*

Abstrak

Aplikasi transportasi *online* kini menjadi bagian penting dalam kehidupan masyarakat Indonesia. Maxim, sebagai salah satu penyedia layanan, perlu memahami persepsi pengguna untuk meningkatkan kualitas layanannya. Penelitian ini menerapkan metode *Bidirectional Long Short-Term Memory* (BiLSTM) untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi Maxim di Google Play Store. Untuk memperkuat representasi kata, digunakan *word embedding FastText* yang mampu menangkap informasi *subword* secara lebih baik. Data penelitian diperoleh melalui scraping menggunakan *package* google-play-scraper pada Python. Model BiLSTM yang dilatih dengan konfigurasi hyperparameter optimal berhasil mengklasifikasikan sentimen ulasan secara efektif, dengan hasil *accuracy* 94%, *precision* 96%, *recall* 95%, dan *f1-score* 95%. Hasil ini menunjukkan bahwa kombinasi BiLSTM dan *FastText* mampu mendeteksi sentimen positif dan negatif secara akurat dan seimbang, serta relevan untuk mendukung evaluasi kualitas layanan berbasis opini pengguna.

Kata Kunci: Analisis Sentimen; Maxim; Word Embedding; FastText; Bidirectional Long Short-Term Memory

Abstract

Online transportation applications are now an important part of Indonesians' lives. Maxim, as one of the service providers, needs to understand users' perceptions to improve its service quality. This research applies the Bidirectional Long Short-Term Memory (BiLSTM) method to perform sentiment classification on Maxim app user reviews on Google Play Store. To strengthen word representation, FastText word embedding is used which is able to capture sub-word information better. The research data was obtained through scraping using the google-play-scraper package in Python. The BiLSTM model trained with the optimal hyperparameter configuration was able to classify the sentiment reviews effectively, with an accuracy of 94%, precision of 96%, recall of 95%, and f1-score of 95%. These results show that the combination of BiLSTM and FastText is able to detect positive and negative sentiments accurately and in a balanced manner, and is relevant to support user opinion-based service quality evaluation.

Keywords: Sentiment Analysis; Maxim; Word Embedding; FastText; Bidirectional Long Short-Term Memory

Copyright © 2025 by Authors, Published by JRMM Group. This is an open access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0)

*Corresponding author. E-mail: nisky@ung.ac.id

Submitted: 16 April 2025 Reviewed: 27 Mei 2025 Accepted: 16 Juni 2025
DOI: https://dx.doi.org/10.18860/jrmm.v4i5.33358

1 Pendahuluan

Perkembangan aplikasi-aplikasi *mobile* inovatif telah membawa perubahan di segala bidang, termasuk di bidang transportasi. Aplikasi layanan transportasi *online* menjadi salah satu aplikasi *mobile* yang paling banyak digunakan di Indonesia [1]. Berdasarkan hasil dari survei Goodstats pada Juni 2023, sebanyak 56,8% responden memilih menggunakan transportasi *online* karena alasan kemudahan dalam pemesanan dan penggunaan [2]. Perkembangan informasi dan teknologi yang semakin maju membuat pelanggan menjadi lebih selektif dalam memilih layanan transportasi *online* yang diinginkan. Sehingga, penyedia jasa transportasi *online* saling bersaing untuk menarik perhatian pelanggan untuk berlangganan layanan mereka [3].

Maxim sebagai salah satu penyedia jasa transportasi *online* telah berusaha bersaing dengan aplikasi transportasi *online* pendahulunya, seperti Gojek dan Grab. Penawaran tarif yang relatif lebih murah menarik perhatian penggunanya sebagai alternatif yang lebih ekonomis [4]. Namun, survei Goodstats pada tahun 2023 menunjukkan bahwa Maxim hanya memperoleh persentase pengguna sebesar 12,3%, jauh tertinggal dibandingkan dengan Gojek dengan 54,4% dan Grab dengan 29,9% [5]. Salah satu cara untuk meningkatkan daya saing adalah dengan memahami umpan balik pengguna yang tercermin dalam ulasan aplikasi di *platform* penyedia aplikasi seperti Google Play Store. Akan tetapi, analisis ulasan ini menghadapi berbagai tantangan, diantaranya jumlah ulasan yang sangat banyak, penggunaan bahasa yang tidak terstruktur, dan penggunaan kata-kata informal, sehingga sulit untuk menentukan sentimen yang dimiliki ulasan [6]. Oleh karena itu, diperlukan metode otomatis berbasis teknologi, seperti analisis sentimen untuk mengolah data ulasan secara efektif dan memberikan wawasan yang dapat digunakan untuk meningkatkan kualitas layanan Maxim.

Setiap ulasan membawa informasi yang berpotensi membentuk persepsi pengguna baru dan memengaruhi keputusan untuk menggunakan aplikasi tersebut. Hasil analisis sentimen yang diungkapkan dalam ulasan ini dapat memberikan wawasan yang berharga mengenai preferensi pengguna dan area yang perlu ditingkatkan [7]. Ulasan dapat dikategorikan menggunakan analisis sentimen ke dalam kelompok sentimen positif dan negatif [8]. Namun, jumlah ulasan yang sangat besar tidak memungkinkan diperiksa secara manual, karena akan menghabiskan banyak waktu, biaya, dan tenaga. Untuk mengatasi hal tersebut, diperlukan implementasi metode otomatis, seperti metode *deep learning*. Salah satu metode yang dapat digunakan adalah *Bidirectional Long Short-Term Memory* (*Bidirectional* LSTM atau BiLSTM).

BiLSTM adalah varian dari *Long Short-Term Memory* (LSTM) yang dikembangkan lebih lanjut dengan menerima *input* dari dua arah, yaitu dari depan dan belakang. Kemampuan pemrosesan dua arah ini memungkinkan model untuk memahami konteks tidak hanya dari masa lalu tetapi juga dari masa depan, sehingga meningkatkan akurasi dalam analisis sentimen [9]. Untuk meningkatkan performa model *deep learning*, diperlukan metode representasi data agar hasil klasifikasi akurat, salah satunya melalui *word embedding* [10]. *Word embedding* mengubah teks menjadi vektor numerik yang membantu model memahami hubungan semantik antar kata [11]. Di antara berbagai model, *FastText* menjadi pilihan unggul karena mempertimbangkan informasi *subword* dalam proses pembelajaran. Tidak seperti model lainnya yang hanya merepresentasikan kata sebagai satu entitas, FastText memecah kata menjadi *n-gram* sehingga dapat mengenali hubungan antar kata berdasarkan komponen penyusunnya. Keunggulan ini memungkinkan *FastText* menangani masalah kata *out-of-vocabulary* (OOV), yaitu kata-kata yang tidak ditemui selama proses pelatihan, yang sering muncul pada ulasan yang berbahasa informal atau tidak terstruktur [12]. Dengan menggunakan *FastText*, model *deep learning* dapat memanfaatkan informasi lebih kaya dari data teks, meningkatkan akurasi dalam analisis sentimen.

Beberapa penelitian sebelumnya telah menunjukkan keunggulan kombinasi model *deep learning* dengan *word embedding* dalam analisis sentimen. Penelitian terdahulu telah membandingkan *GloVe* dan *FastText* pada model LSTM untuk analisis sentimen di Twitter, dengan hasil bahwa FastText memberikan akurasi lebih tinggi [13]. Penelitian selanjutnya menunjukkan bahwa model BiLSTM lebih unggul dibandingkan LSTM dalam menganalisis ulasan aplikasi transportasi *online*, meskipun belum memanfaatkan *word embedding* untuk representasi kata yang lebih kaya [14]. Pada penelitian lain, digunakan kombinasi BiLSTM dan *word2Vec*, yang terbukti efektif dalam analisis sentimen ulasan aplikasi.

Namun, kombinasi ini bisa dikembangkan lebih lanjut dengan menggunakan *word embedding FastText* sebagai pengembangan dari *word2Vec* yang memanfaatkan informasi *subword* untuk menangani masalah kata *out-of-vocabulary* (OOV) [9].

Berdasarkan penelitian-penelitian sebelumnya, pendekatan yang diusulkan dalam penelitian ini adalah penerapan model BiLSTM yang dikombinasikan dengan *word embedding FastText* untuk analisis sentimen ulasan pengguna aplikasi Maxim. Pendekatan ini bertujuan mengatasi keterbatasan representasi kata dan menangani kata-kata baru (*out-of-vocabulary*) yang sering muncul dalam teks ulasan informal. Keterbaruan dari penelitian ini terletak pada penerapan kombinasi metode BiLSTM dan FastText secara spesifik pada studi kasus aplikasi Maxim, yang hingga kini belum banyak diteliti. Berbeda dari penelitian sebelumnya yang memanfaatkan metode representasi kata berbasis *word embedding Word2Vec* [15] atau *GloVe* [16] yang memperlakukan setiap kata sebagai satu unit dan kurang mampu menangani kata yang tidak ditemukan dalam pelatihan [17], *FastText* menawarkan keunggulan dalam menangkap informasi *subword*, sehingga lebih efektif untuk konteks bahasa tidak baku. Adapun kontribusi utama dari penelitian ini adalah menghasilkan pendekatan analisis sentimen yang lebih akurat dan dapat dimanfaatkan sebagai dasar pengambilan keputusan bagi pengembang layanan aplikasi transportasi *online*.

2 Metode

2.1 Pendekatan Penelitian

Pendekatan yang digunakan dalam penelitian ini adalah pendekatan studi kasus. Studi kasus dilakukan pada aplikasi Maxim, dengan pengumpulan data berupa ulasan pengguna aplikasi Maxim yang diperoleh melalui proses *scraping* dari Google Play Store. Data ulasan yang sudah terkumpul kemudian dianalisis, dengan fokus pada pemodelan analisis sentimen menggunakan model *Bidirectional Long Short-Term Memory* (BiLSTM) yang didukung oleh *word embedding FastText*.

2.2 Variabel Penelitian

Variabel dalam penelitian ini adalah hasil klasifikasi sentimen ulasan pengguna aplikasi Maxim. Hasil klasifikasi ini dinyatakan dalam bentuk angka, yaitu 0 untuk sentimen negatif dan 1 untuk sentimen positif. Klasifikasi dilakukan secara otomatis dengan bantuan module VADER dari package NLTK dalam Python, dengan definisi bahwa sentimen positif mencerminkan kepuasan atau apresiasi terhadap layanan aplikasi, sedangkan sentimen negatif mencerminkan keluhan atau ketidakpuasan terhadap aplikasi.

2.3 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder berupa teks ulasan pengguna aplikasi Maxim yang diperoleh dari Google Play Store. Data diperoleh melalui proses *scraping* menggunakan *package google-play-scraper* pada bahasa pemrograman *Python* dalam periode tertentu. Ulasan digunakan sebagai bahan analisis untuk mendukung penelitian analisis sentimen menggunakan model BiLSTM dengan *word embedding FastText*. Data ini mencerminkan opini pengguna secara publik dan digunakan untuk menganalisis persepsi pengguna terhadap aplikasi Maxim. Data yang diambil langsung dari Google Play Store hanya ulasan saja, tanpa mengambil informasi lain seperti *username*.

2.4 Analisis Data

Data yang digunakan dalam penelitian ini adalah data sekunder berupa ulasan pengguna aplikasi Maxim di Indonesia, yang dikumpulkan melalui proses *scraping* dari Google Play Store pada periode 12 Oktober 2024 hingga 30 Oktober 2024. Jumlah data yang berhasil dikumpulkan sebanyak 10.000 ulasan, yang hanya terdiri dari teks ulasan tanpa informasi pribadi lainnya dari pengguna. Langkah-langkah yang dilakukan pada penelitian ini untuk analisis sentimen dari pengguna aplikasi Maxim sebagai berikut.

2.4.1 Preprocessing Data

Preprocessing data dilakukan untuk menyiapkan data teks ulasan mentah ke dalam bentuk yang lebih sederhana dan dapat diproses oleh komputer. Tahap-tahap dari *preprocessing data* dan penjelasannya adalah sebagai berikut [18].

1. *Data cleaning*, yaitu proses menghapus tanda baca dan karakter yang tidak diperlukan seperti *emoji*, tautan, dan simbol.
2. *Case folding*, yaitu proses penyesuaian semua huruf dalam teks ulasan menjadi satu format, biasanya huruf kecil atau huruf besar, agar penulisan menjadi seragam.
3. Tokenisasi, yaitu proses memecah teks atau dokumen menjadi bagian-bagian kecil yang disebut token. Token merujuk pada kata-kata individual dalam suatu kalimat.
4. Penghapusan *stopword*, yaitu proses mengidentifikasi serta mengeliminasi kata-kata umum yang tidak memiliki kontribusi berarti terhadap analisis sentimen. Contoh *stop words* mencakup "yang", "adalah", "dan", "di", "itu", serta kata-kata lainnya.
5. *Stemming*, yaitu proses yang melibatkan penghilangan akhiran atau awalan kata-kata untuk menghasilkan bentuk dasar atau akar dari kata.

2.4.2 Pelabelan Awal Data

Dalam tahap ini, label diberikan secara otomatis menggunakan *module* VADER dari *package Natural Language Toolkit* (NLTK). Pelabelan dilakukan berdasarkan *compound score* yang berkisar antara -1 hingga 1. Jika nilai *compound score* kurang dari atau sama dengan -0,05, ulasan dikategorikan sebagai sentimen negatif. Ketika nilai *compound score* berada diantara -0,05 dan 0,05, ulasan dikategorikan sebagai sentimen netral. Dan jika nilai *compound score* lebih dari atau sama dengan 0,05, ulasan dikategorikan sebagai sentimen positif [19].

Ulasan dikategorikan menjadi dua, dan hasil klasifikasi ini dinyatakan dalam bentuk angka, yaitu 0 untuk sentimen negatif dan 1 untuk sentimen positif. [20]. Dalam penelitian ini, kelas netral sengaja dihilangkan agar fokus klasifikasi terbatas pada dua kelas utama, yaitu sentimen positif dan negatif. Hal ini dilakukan untuk menyederhanakan struktur model dan memfokuskan pelatihan model BiLSTM pada identifikasi opini yang eksplisit. Pendekatan ini juga diadopsi dalam beberapa studi sebelumnya pada analisis sentimen aplikasi dan media sosial [3], [21].

2.4.3 Pembagian Data

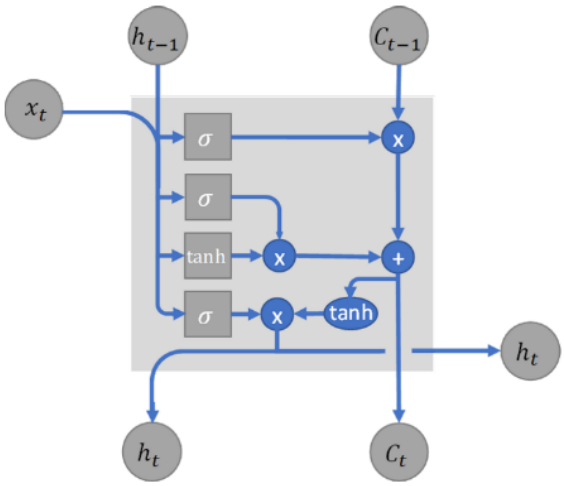
Data yang telah dilabeli akan dibagi menjadi tiga bagian, yaitu data latih, data validasi, dan data uji. Penelitian ini membandingkan tiga proporsi pembagian data yang berbeda, yaitu 60:20:20, 70:25:5, dan 80:10:10. Tujuan dari perbandingan ini adalah untuk mengidentifikasi proporsi pembagian yang menghasilkan kinerja terbaik berdasarkan metrik evaluasi, sehingga model yang dibangun dapat menggunakan konfigurasi data yang paling optimal dalam proses pelatihannya. Perbandingan ini dilakukan berdasarkan referensi penelitian terdahulu yang juga menerapkan model LSTM maupun BiLSTM pada analisis sentimen [14], [22], [23].

2.4.4 Word Embedding FastText

FastText adalah sebuah metode *word embedding* yang dikembangkan oleh tim *Facebook AI Research* (FAIR). *FastText* dipilih karena kemampuannya dalam menangkap informasi morfologis melalui representasi *subword*, yaitu *n-gram* karakter, sehingga tetap mampu menghasilkan representasi yang bermakna meskipun kata yang bersangkutan tidak terdapat dalam kosakata pelatihan (*out-of-vocabulary*) [24]. Sebuah *n-gram* karakter adalah himpunan dari n karakter berurutan dalam sebuah kata w.

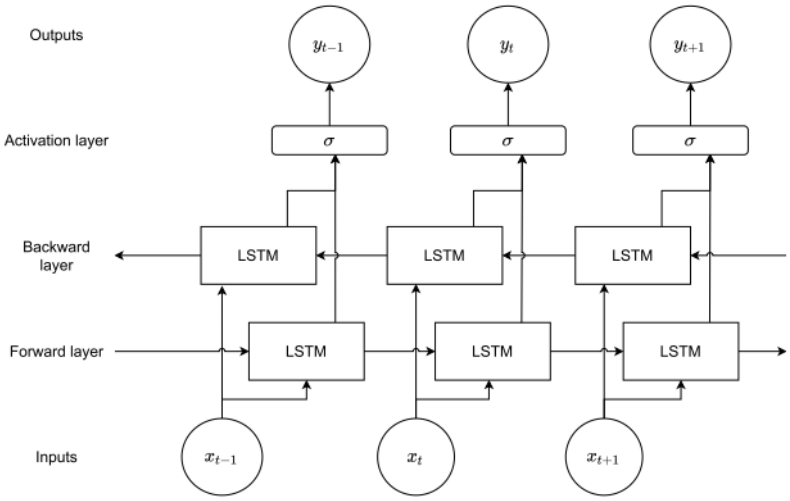
2.4.5 *Bidirectional Long Short-Term Memory (BiLSTM)*

Bidirectional Long Short-Term Memory (BiLSTM) merupakan pengembangan dari model *Long Short-Term Memory* (LSTM) yang memproses data dalam dua arah, sehingga mampu menangkap konteks dari kedua sisi dalam data sekuensial [10]. Komponen utama dari LSTM adalah *cell state* dan tiga jenis *gate*, yang secara selektif mengatur aliran informasi dari *cell state* di setiap waktu input. *Forget gate* berfungsi untuk menghapus informasi yang tidak relevan, *input gate* menentukan informasi baru yang akan ditambahkan, dan *output gate* menghasilkan *output* berdasarkan *input* saat ini dan memori dalam *cell* [15]. Gambaran arsitektur model LSTM dapat dilihat pada Gambar 1.



Gambar 1: Arsitektur model LSTM

Simbol $\mathbf{x}_t$ merupakan vektor input pada waktu t, $\mathbf{h}_t$ merupakan vektor *output* (*hidden state*) pada waktu t, dan $\mathbf{c}_t$ merupakan *cell state*. Sementara itu, σ dan *tanh* merupakan fungsi aktivasi yang digunakan dalam proses komputasi dalam unit LSTM. BiLSTM terdiri dari dua jalur LSTM paralel, LSTM maju yang memproses data dari kiri ke kanan dan LSTM mundur dari kanan ke kiri. Setiap *input* diteruskan secara simultan ke kedua jalur. *Output* dari keduanya kemudian digabungkan melalui *activation layer* menggunakan operasi konkatenasi. Arsitektur model BiLSTM dapat dilihat pada Gambar 2.



Gambar 2: Arsitektur model BiLSTM

2.4.6 *Pelatihan Model*

Pada tahap pelatihan, dilakukan proses *hyperparameter tuning* untuk memperoleh konfigurasi model BiLSTM yang optimal. *Hyperparameter* yang disesuaikan dalam penelitian ini mencakup jumlah unit

BiLSTM, *batch size*, *dropout rate*, serta *learning rate* [25]. Pemilihan kombinasi *hyperparameter* terbaik dilakukan menggunakan metode *grid search*, dimana semua konfigurasi yang mungkin terbentuk diuji untuk menentukan kombinasi dengan nilai evaluasi terbaik pada data validasi, terutama berdasarkan metrik *f1-score*, yang relevan untuk data dengan distribusi kelas yang tidak seimbang [26].

2.4.7 Klasifikasi dan Evaluasi Model

Tahap selanjutnya adalah melakukan klasifikasi pada data uji menggunakan model terbaik. Hasil klasifikasi ini akan menjadi bahan evaluasi kinerja model. Evaluasi model dilakukan untuk mengukur kinerjanya dalam melakukan klasifikasi, dengan menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score*. Untuk menghitung metrik-metrik tersebut, digunakan *confusion matrix* untuk memetakan hasil prediksi dari model terhadap data uji.

3 Hasil dan Pembahasan

3.1 *Preprocessing* Data

Tahapan persiapan yang perlu dilakukan yaitu *preprocessing* data dengan tujuan agar data teks dapat diproses pada tahap-tahap selanjutnya. Digunakan lima tahap *preprocessing* data, yaitu *data cleaning*, *case folding*, tokenisasi, penghapusan *stopword*, dan *stemming*. Setelah tahap ini, data yang tersisa sebanyak 9.709 ulasan. Hasil dari setiap tahap pada contoh ulasan dapat dilihat pada Tabel 1.

Tabel 1: Hasil *preprocessing* pada data ulasan

Tahap	*Input*	*Output*
Data Cleaning	Saya SEBAGAI CS KECEWA Pagi2 aplikasi eror TERPAKSA SY PINDAH KE IJO .... Gmn indo Mau Maju sekelas transpor aja gangguan mulu ,mikir dong...	Saya SEBAGAI CS KECEWA Pagi aplikasi eror TERPAKSA SY PINDAH KE IJO Gmn indo Mau Maju sekelas transpor aja gangguan mulu mikir dong
Case Folding	Saya SEBAGAI CS KECEWA Pagi aplikasi eror TERPAKSA SY PINDAH KE IJO Gmn indo Mau Maju sekelas transpor aja gangguan mulu mikir dong	saya sebagai cs kecewa pagi aplikasi eror terpaksa sy pindah ke ijo gmn indo mau maju sekelas transpor aja gangguan mulu mikir dong
Tokenisasi	saya sebagai cs kecewa pagi aplikasi eror terpaksa sy pindah ke ijo gmn indo mau maju sekelas transpor aja gangguan mulu mikir dong	'saya', 'sebagai', 'cs', 'kecewa', 'pagi', 'aplikasi', 'eror', 'terpaksa', 'sy', 'pindah', 'ke', 'ijo', 'gmn', 'indo', 'mau', 'maju', 'sekelas', 'transpor', 'aja', 'gangguan', 'mulu', 'mikir', 'dong'
Penghapusan *Stopword*	'saya', 'sebagai', 'cs', 'kecewa', 'pagi', 'aplikasi', 'eror', 'terpaksa', 'sy', 'pindah', 'ke', 'ijo', 'gmn', 'indo', 'mau', 'maju', 'sekelas', 'transpor', 'aja', 'gangguan', 'mulu', 'mikir', 'dong'	'cs', 'kecewa', 'pagi', 'aplikasi', 'eror', 'terpaksa', 'sy', 'pindah', 'ijo', 'gmn', 'indo', 'mau', 'maju', 'sekelas', 'transpor', 'aja', 'gangguan', 'mulu', 'mikir', 'dong'
Stemming	'cs', 'kecewa', 'pagi', 'aplikasi', 'eror', 'terpaksa', 'sy', 'pindah', 'ijo', 'gmn', 'indo', 'mau', 'maju', 'sekelas', 'transpor', 'aja', 'gangguan', 'mulu', 'mikir', 'dong'	'cs', 'kecewa', 'pagi', 'aplikasi', 'eror', 'paksa', 'sy', 'pindah', 'ijo', 'gmn', 'indo', 'mau', 'maju', 'kelas', 'transpor', 'aja', 'ganggu', 'mulu', 'mikir', 'dong'

3.2 Pelabelan Awal Data

Tahap selanjutnya adalah memberikan label pada data ulasan menggunakan package VADER (*Valence Aware Dictionary and sEntiment Reasoner*). Sebelum proses pelabelan dilakukan, data ulasan yang berbahasa Indonesia diterjemahkan terlebih dahulu ke dalam bahasa Inggris. Proses penerjemahan dilakukan untuk mempermudah VADER dalam mengklasifikasikan data ulasan karena VADER dirancang untuk bekerja pada teks berbahasa Inggris. Proses penerjemahan ini berpotensi mengubah konteks

atau nuansa semantik asli dari ulasan, terutama pada kata-kata informal atau berkonotasi lokal yang sulit untuk disepadankan. Meskipun pendekatan ini cukup praktis, hasil pelabelan hanya aproksimasi dari sentimen asli ulasan pengguna. Tabel 2 menampilkan contoh hasil penerjemahan data.

Tabel 2: Hasil penerjemahan data ulasan

Input	*Output*
'cs', 'kecewa', 'pagi', 'aplikasi', 'eror', 'paksa', 'sy', 'pindah', 'ijo', 'gmn', 'indo', 'mau', 'maju', 'kelas', 'transpor', 'aja', 'ganggu', 'mulu', 'mikir', 'dong'	'cs', 'disappointed', 'morning', 'application', 'error , 'forced', 'sy', 'moved', 'ijo', 'gmn', 'indo', 'want', 'forward', 'class', 'transport', 'just', 'disturb', 'first', 'think', 'dong'

Dalam proses pelabelan menggunakan VADER ini, digunakan nilai *compound score* yang berkisar antara -1 hingga 1. Pada penelitian ini tidak digunakan ulasan netral, sehingga ulasan yang termasuk kategori netral akan dihapus. Setelah proses pelabelan awal, beberapa data yang dilabeli netral dihapus dan diperoleh 7.848 data dengan 5.366 bersentimen positif dan 2.482 bersentimen negatif. Pada Tabel 3 adalah contoh hasil pelabelan yang dihasilkan menggunakan VADER.

Tabel 3: Hasil pelabelan data ulasan

Input	Label	*Compound Score*
'cs', 'disappointed', 'morning', 'application', 'error , 'forced', 'sy', 'moved', 'ijo', 'gmn', 'indo', 'want', 'forward', 'class', 'transport', 'just', 'disturb', 'first', 'think', 'dong'	Negatif	-0,8807

3.3 Pembagian Data

Tahap selanjutnya yaitu membagi data menjadi tiga bagian, yaitu data latih, data validasi, dan data uji. Dalam penelitian ini, digunakan beberapa proporsi pembagian data yang berbeda, yaitu proporsi 60:20:20, 70:25:5, serta 80:10:10 untuk data latih, data validasi, dan data uji berturut turut. Data yang berjumlah 7.848 ulasan kemudian dibagi secara acak sesuai dengan proporsi masing-masing. Hasil pembagian data disajikan dalam Tabel 4.

Tabel 4: Hasil pembagian data ulasan

Proporsi	Data Latih	Data Validasi	Data Uji
60:20:20	4709	1570	1569
70:25:5	5494	1962	392
80:10:10	6278	785	785

3.4 *Word Embedding*

Tahap selanjutnya yaitu pembangunan model *word embedding FastText* untuk mengubah teks ulasan yang telah diuraikan menjadi token pada tahap *preprocessing* data menjadi representasi numerik dalam bentuk vektor, sehingga dapat diproses oleh model untuk tahap-tahap selanjutnya. Penelitian ini menggunakan model *pre-trained word embedding FastText*. Pendekatan ini memungkinkan model mengenali dan merepresentasikan kata-kata baru yang tidak ditemui dalam data latih (*out-of-vocabulary*) serta imbuhan kata yang sering muncul dalam teks ulasan pengguna, yang tidak terstruktur secara baku. Jumlah dimensi fitur pada vektor model *FastText* adalah 300. Vektor kata dari sebagian token dalam data latih dapat dilihat pada Tabel 5.

Tabel 5: Vektor kata dari model *word embedding FastText*

Token	fitur_1	fitur_2	...	fitur_299	fitur_300
<unk>	0	0	...	0	0
<pad>	0	0	...	0	0
driver	0,0782	0,0211	...	0,0402	0,0913
nya	0,0895	-0,0885	...	-0,0151	0,0909
baik	-0,0184	-0,0478	...	0,0704	0,0729

3.5 *Hyperparameter Tuning*

Representasi vektor dari kata-kata yang telah diperoleh melalui *word embedding FastText* selanjutnya digunakan sebagai *input* untuk model klasifikasi BiLSTM. Dalam tahap pembangunan model ini dilakukan *hyperparameter tuning*, yaitu proses menentukan nilai nilai *hyperparameter* yang akan memengaruhi kinerja model. Nilai nilai ini akan diuji melalui pelatihan model dan dibandingkan untuk memperoleh konfigurasi model terbaik berdasarkan metrik evaluasi tertentu. Hyperparameter yang digunakan dalam penelitian ini jumlah unit BiLSTM, *batch size*, *dropout rate*, dan *learning rate*. Rentang nilai setiap *hyperparameter* yang digunakan pada penelitian ini dapat dilihat pada Tabel 6.

Tabel 6: Rentang nilai *hyperparameter* yang digunakan

Hyperparameter	Nilai
Unit LSTM	64, 128
Batch Size	32, 64
Dropout Rate	0,3, 0,5
Learning Rate	0,001, 0,0001

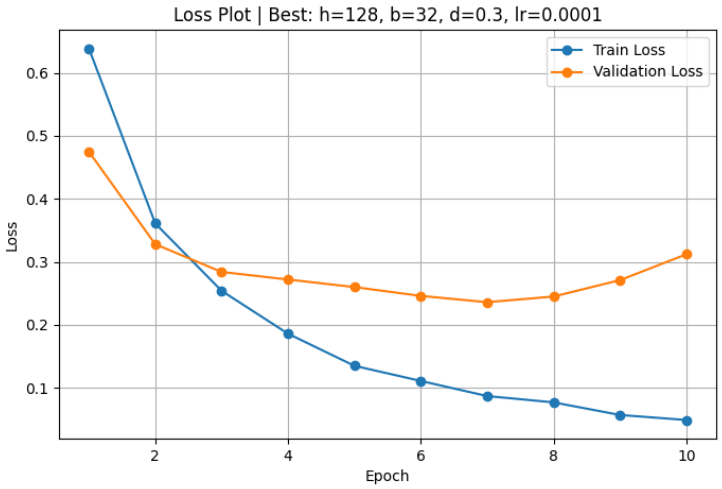
3.6 Pelatihan Model

Proses pelatihan model BiLSTM dilakukan dengan pendekatan *Grid Search*, di mana setiap rentang nilai *hyperparameter* yang telah ditentukan akan membentuk total 16 konfigurasi model. Masing-masing model dilatih selama 10 *epoch* untuk menguji performa model terhadap data validasi. Untuk mengevaluasi kinerja suatu konfigurasi model pada data validasi dan memilih konfigurasi terbaik, digunakan metrik *f1-score*. Model terbaik akan dipilih berdasarkan nilai *f1-score* tertinggi pada data validasi. Hasil pelatihan model BiLSTM untuk masing-masing pembagian data dapat dilihat pada Tabel 7.

Tabel 7: Hasil pelatihan model BiLSTM

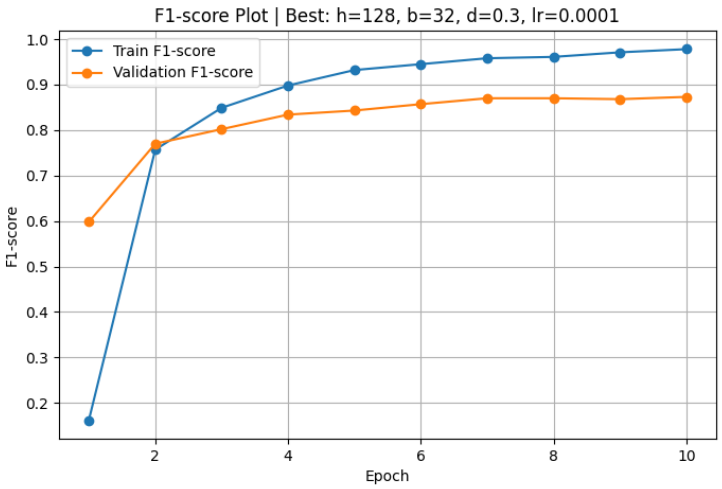
Hyperparameter				*F1-score*		
Unit BiLSTM	*Batch Size*	*Dropout Rate*	*Learning Rate*	60:20:20	70:25:5	80:10:10
64	32	0,3	0,001	0,843	0,837	0,800
			0,0001	0,847	0,854	0,870
		0,5	0,001	0,835	0,834	0,843
			0,0001	0,857	0,865	0,869
	64	0,3	0,001	0,816	0,857	0,836
			0,0001	0,841	0,846	0,862
		0,5	0,001	0,830	0,858	0,838
			0,0001	0,841	0,846	0,813
128	32	0,3	0,001	0,827	0,857	0,816
			0,0001	0,851	0,839	0,873
		0,5	0,001	0,847	0,859	0,835
			0,0001	0,843	0,869	0,856
	64	0,3	0,001	0,845	0,859	0,843
			0,0001	0,857	0,846	0,855
		0,5	0,001	0,811	0,847	0,842
			0,0001	0,852	0,863	0,853

Model terbaik adalah model yang menunjukkan *f1-score* tertinggi pada data validasi selama pelatihan. Sehingga berdasarkan Tabel 7, model terbaik yang akan digunakan untuk tahap klasifikasi adalah BiLSTM dengan 128 unit, *batch size* 32, *dropout rate* 0,3 dan *learning rate* 0,0001 pada proporsi pembagian data 80:10:10. Hasil dari proses pelatihan model terbaik selama setiap *epoch* kemudian dipetakan ke dalam grafik. Grafik *loss* dari hasil pelatihan model terbaik ditampilkan pada Gambar 3.



Gambar 3: Grafik *loss* model terbaik

Berdasarkan Gambar 3, terlihat bahwa nilai *loss* pada data latih mengalami penurunan secara gradual, sementara nilai *loss* pada data validasi juga menunjukkan penurunan awal yang konsisten hingga *epoch* ke 7, namun mulai mengalami peningkatan kembali setelahnya. Pola ini mengindikasikan kemungkinan awal terjadinya *overfitting*, di mana model terus membaik pada data latih namun mulai kehilangan generalisasi pada data validasi setelah titik tertentu. Oleh karena itu, model terbaik dipilih berdasarkan nilai *validation loss* terendah yang dicapai, yaitu pada *epoch* ke 7, untuk memastikan performa model yang optimal.



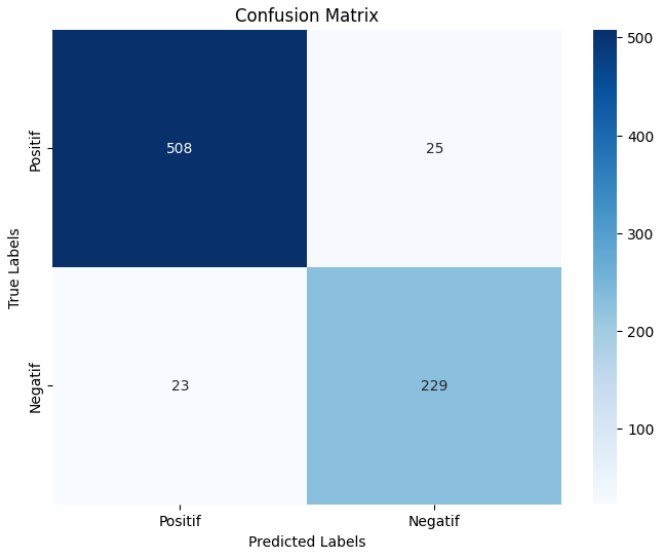
Gambar 4: Grafik *f1-score* model terbaik

Berdasarkan Gambar 4, *f1-score* pada data latih meningkat tajam sejak *epoch* pertama dan terus meningkat secara stabil mendekati 1 di *epoch* ke 10. Hasil ini menunjukkan bahwa model mampu mengenali dan mengklasifikasi data latih dengan sangat baik. Pada data validasi, *f1-score* juga mengalami peningkatan konsisten hingga sekitar *epoch* ke 7 dan tetap stabil setelahnya. Kenaikan *f1-score* yang

sejalan antara data latih dan validasi seiring dengan penurunan nilai *loss* menunjukkan bahwa pelatihan berjalan lancar dan model mampu melakukan generalisasi dengan baik terhadap data yang tidak terlihat selama pelatihan.

3.7 Klasifikasi dan Evaluasi Model

Tahap selanjutnya adalah melakukan klasifikasi terhadap data ulasan menggunakan model BiLSTM yang telah melalui proses pelatihan pada tahap sebelumnya. Pengujian dilakukan terhadap data uji untuk mengelompokkan ulasan pengguna aplikasi Maxim ke dalam dua kategori sentimen, yaitu positif dan negatif, yang bertujuan mendukung proses pengambilan keputusan. Dari total 785 data uji, sebanyak 533 berlabel asli sentimen positif dan 252 berlabel asli sentimen negatif. Untuk memberikan gambaran performa model yang lebih rinci dan sekaligus sebagai bahan untuk evaluasi, hasil klasifikasi disajikan dalam bentuk *confusion matrix*, yang merepresentasikan hubungan antara prediksi model dengan label asli. *Confusion matrix* dari hasil klasifikasi pada data uji ditampilkan pada Gambar 5.



Gambar 5: *Confusion matrix*

Berdasarkan Gambar 5, model BiLSTM menunjukkan kinerja klasifikasi yang sangat baik pada data uji, dengan tingkat akurasi tinggi pada kedua kelas sentimen. Sebanyak 508 dari 533 ulasan positif dan 229 dari 252 ulasan negatif berhasil diklasifikasikan dengan benar. Jumlah kesalahan klasifikasi juga relatif kecil dan seimbang, masing-masing 25 kesalahan untuk prediksi negatif terhadap ulasan positif, dan 23 kesalahan untuk prediksi positif terhadap ulasan negatif. Hasil ini mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik meskipun terdapat ketidakseimbangan label. *Confusion matrix* digunakan untuk menghitung metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *f1-score*. Metrik evaluasi dari hasil analisis sentimen ulasan pada data uji ditampilkan pada Tabel 8.

Tabel 8: Metrik evaluasi model BiLSTM

Sentimen	*Accuracy*	*Precision*	*Recall*	*F1-Score*	Jumlah Ulasan
Positif	0,94	0,96	0,95	0,95	533
Negatif	0,94	0,90	0,91	0,91	252

Tabel 8 menunjukkan bahwa model BiLSTM dengan *word embedding FastText* mencapai *accuracy* sebesar 94% pada data uji, yang menunjukkan tingkat ketepatan analisis sentimen yang tinggi. *Precision* sebesar 96% menandakan bahwa prediksi positif sangat akurat, sedangkan *recall* sebesar 95% menunjukkan model mampu mengenali sebagian besar ulasan positif secara tepat. *F1-score* sebesar

95% mencerminkan keseimbangan yang kuat antara *precision* dan *recall*, sehingga model tidak hanya akurat, tetapi juga andal dalam menjangkau ulasan dengan sentimen sebenarnya.

3.8 Interpretasi Model

Hasil klasifikasi sentimen terhadap data uji ulasan pengguna aplikasi Maxim di *Google Play Store* menunjukkan bahwa sebanyak 67,9% ulasan termasuk dalam kategori sentimen positif, sementara 32,1% lainnya merupakan sentimen negatif. Temuan ini mencerminkan bahwa secara umum mayoritas pengguna merasa puas dengan layanan yang ditawarkan oleh aplikasi Maxim. Ulasan positif umumnya memuji aspek tarif yang terjangkau serta kemudahan penggunaan, yang dinilai praktis dibandingkan aplikasi pesaing.

Sementara itu, proporsi sentimen negatif yang juga tidak sedikit mengindikasikan adanya aspek layanan yang perlu untuk diperbaiki. Keluhan paling umum dalam ulasan negatif berkaitan dengan lama waktu tunggu driver dan ketidakstabilan sistem aplikasi. Meskipun tidak dominan, sentimen negatif ini tetap memberikan informasi penting terkait hambatan yang dihadapi pengguna dalam menggunakan aplikasi.

Secara praktis, hasil klasifikasi ini memberikan gambaran umum persepsi publik terhadap layanan aplikasi Maxim. Informasi ini tidak hanya dapat digunakan untuk mengevaluasi kepuasan pelanggan, tetapi juga berfungsi sebagai masukan bagi pengembang aplikasi dalam mengidentifikasi kelemahan layanan dan menetapkan prioritas perbaikan yang tepat. Dengan memanfaatkan hasil analisis sentimen secara efektif, perusahaan dapat meningkatkan kualitas layanannya secara berkelanjutan dan memperkuat daya saing di pasar transportasi *online*.

4 Kesimpulan

Berdasarkan hasil pengujian, konfigurasi hyperparameter optimal diperoleh pada kombinasi 128 unit BiLSTM, *batch size* 32, *dropout rate* 0,3 dan *learning rate* 0,0001. Model menunjukkan performa tinggi dengan *accuracy* 94%, *precision* 96%, *recall* 95%, dan *f1-score* 95%. proporsi pembagian data 80:10:10 terbukti memberikan hasil terbaik. Analisis sentimen terhadap ulasan pengguna Maxim menunjukkan bahwa 68,37% ulasan tergolong positif dan 31,63% negatif, dengan sentimen positif didominasi oleh tarif murah dan kemudahan penggunaan, sementara keluhan umumnya berkaitan dengan waktu tunggu dan layanan. Penelitian ini berkontribusi dengan menerapkan kombinasi BiLSTM dan *FastText* yang efektif untuk menangani teks ulasan informal berbahasa Indonesia. Namun demikian, penelitian ini memiliki beberapa keterbatasan, antara lain penggunaan metode pelabelan otomatis dengan VADER yang memerlukan penerjemahan teks ke bahasa Inggris, yang dapat berisiko mengubah makna asli ulasan. Selain itu, cakupan data terbatas pada satu aplikasi dan periode waktu tertentu, sehingga generalisasi ke aplikasi lain atau kondisi waktu berbeda memerlukan validasi lebih lanjut.

Sebagai saran untuk penelitian selanjutnya, aspek yang dapat dieksplorasi meliputi perbandingan performa dengan model lain seperti CNN atau BiGRU, serta penggunaan alternatif word embedding seperti *GloVe* atau *Word2Vec* untuk mengidentifikasi kombinasi metode yang paling optimal dalam analisis sentimen.

Pernyataan Kontribusi Penulis (CRediT)

Hanz Franklyn Bachruddin Wewengkang: Konseptualisasi, Metodologi, Perangkat Lunak, Analisis Formal, Investigasi, Kurasi Data, Penulisan–Draf Awal, Penulisan-Telaah dan Penyuntingan, Visualisasi. **Djihad Wungguli:** Konseptualisasi, Metodologi, Penulisan–Draf Awal, Penulisan–Telaah dan Penyuntingan, Supervisi **Nisky Imansyah Yahya:** Konseptualisasi, Metodologi, Penulisan–Draf Awal, Penulisan–Telaah dan Penyuntingan, Supervisi **Isran K. Hasan:** Metodologi, Perangkat Lunak, Penulisan–Draf Awal, Penulisan–Telaah dan Penyuntingan, Supervisi **Siti Nurmardia Abdussamad:** Metodologi, Penulisan–Draf Awal, Penulisan–Telaah dan Penyuntingan, Supervisi

Deklarasi Konflik Kepentingan

Para penulis menyatakan tidak ada konflik kepentingan.

Ketersediaan Data dan Kode

Seluruh data dan kode program yang digunakan dalam penelitian ini tersedia secara terbuka dan dapat diakses oleh publik. Dataset terdiri dari ulasan pengguna aplikasi Maxim yang dikumpulkan dari Google Play Store dan tidak mengandung informasi pribadi. Dokumentasi dan tautan untuk mengakses data dan kode tersedia dalam repositori GitHub[1].

Daftar Pustaka

[1] M. Marlizar, W. Agustina, F. Bachtiar, and M. Mukhsinuddin, "Effect of Service Quality and Use of E-Service Technology on Customer Loyalty: A Case Study of Maxim in Aceh," *Budapest International Research and Critics Institute (BIRCI-Journal)*, vol. 4, no. 4, pp. 8002–8016, 2021. DOI: 10.33258/birci.v4i4.2717

[2] L. Liana, Y. Yinna, M. Chen, and F. Rahmah, "Analisis Kepuasan Pelanggan Kualitas Pelayanan pada Pengguna Transportasi Online Maxim," *Jurnal Ekonomi, Manajemen dan Akuntansi*, vol. 2, no. 1, pp. 9–19, 2024. DOI: 10.572349/neraca.v2i1.655

[3] E. Suryati, S. Styawati, and A. A. Aldino, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi Dan Sistem Informasi*, vol. 4, no. 1, pp. 96–106, 2023. DOI: 10.33365/jtsi.v4i1.2445

[4] K. Kapriani, A. Asmawiyah, S. Thaha, and H. Hariyanti, "Analisis Pengaruh Harga dan Kualitas Layanan terhadap Keputusan Konsumen Menggunakan Jasa Transportasi Online Maxim di Makassar," *Al-Buhuts*, vol. 17, no. 1, pp. 142–161, 2021. DOI: 10.30603/ab.v17i1.2251

[5] D. S. Barasa, M. Rizan, and T. A. Monoarfa, "Analisis Kepuasan Pengguna Layanan Ride Hailing Maxim Di Jabodetabek," *Jurnal Mufakat*, vol. 3, no. 1, pp. 413–421, 2024.

[6] I. M. M. Parwita and D. Siahaan, "Classification of Mobile Application Reviews using Word Embedding and Convolutional Neural Network," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, vol. 10, no. 1, p. 1, 2019. DOI: 10.24843/lkjiti.2019.v10.i01.p01

[7] L. S. Rao, S. Kumar, N. Reddy, J. Reddy, and B. Buvanne, "A Study of Sentiment Analysis on Mobile Applications," *International Journal of Research Publication and Reviews*, vol. 5, no. 5, pp. 8035–8038, 2024. DOI: 10.55248/gengpi.5.0524.1325

[8] F. Alshuwaier, A. Areshey, and J. Poon, "Applications and Enhancement of Document-Based Sentiment Analysis in Deep learning Methods: Systematic Literature Review," *Intelligent Systems with Applications*, vol. 15, p. 200 090, 2022. DOI: 10.1016/j.iswa.2022.200090

[9] R. Onsu, D. F. Sengkey, and F. D. Kambey, "Implementasi Bi-LSTM dengan Ekstraksi Fitur Word2Vec untuk Pengembangan Analisis Sentimen Aplikasi Identitas Kependudukan Digital," *Jurnal Teknologi Terpadu*, vol. 10, no. 1, pp. 46–55, 2024. DOI: 10.54914/jtt.v10i1.1225

[10] D. J. M. Pasaribu, K. Kusrini, and S. Sudarmawan, "Peningkatan Akurasi Klasifikasi Sentimen Ulasan Makanan Amazon dengan Bidirectional LSTM dan BERT Embedding," *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, vol. 10, no. 1, pp. 9–20, 2020. DOI: 10.35585/inspir.v10i1.2568

[1] https://github.com/hfbw03/maxim-BiLSTM

[11] E. Rudkowsky, M. Haselmayer, M. Wastian, M. Jenny, Š. Emrich, and M. Sedlmair, "More than Bags of Words: Sentiment Analysis with Word Embeddings," *Communication Methods and Measures*, vol. 12, no. 2, pp. 140–157, 2018. DOI: 10.1080/19312458.2018.1455817

[12] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017. DOI: 10.1162/tacl_a_00051

[13] A. Alfauzi and W. Maharani, "Sentiment Analysis of Forest Fires on Social Media Networks Twitter Using the Long Short Term Memory (LSTM) Method," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 1005–1012, 2023. DOI: 10.30865/klik.v3i6.866

[14] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 12, no. 2, pp. 89–99, 2022. DOI: 10.34010/jamika.v12i2.7764

[15] V. B. Lestari, E. Utami, and Hanafi, "Combining Bi-LSTM And Word2vec Embedding For Sentiment Analysis Models Of Application User Reviews," *Indonesian Journal of Computer Science*, vol. 13, no. 1, pp. 312–326, 2024. DOI: 10.33022/ijcs.v13i1.3647

[16] R. Refianti and N. Anggraeni, "Sentiment Analysis Using Convolutional Neural Network Method to Classify Reviews on Zoom Cloud Meetings Application Based on Reviews on Google Playstore," *International Journal of Engineering, Science and Information Technology*, vol. 3, no. 3, pp. 7–16, 2023. DOI: 10.52088/ijesty.v3i3.463

[17] A. E. Nanda, A. N. Sihananto, and A. M. Rizki, "Analisis Sentimen Pada Pembatalan Tuan Rumah Indonesia Di Piala Dunia U-20 Menggunakan Fasttext Embeddings Dan Algoritma Recurrent Neural Network," *SABER: Jurnal Teknik Informatika, Sains dan Ilmu Komunikasi*, vol. 2, no. 2, pp. 246–257, 2024. DOI: 10.59841/saber.v2i2.1000

[18] A. O. Praneswara and N. Cahyono, "Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes," *The Indonesian Journal of Computer Science*, vol. 12, no. 6, pp. 3925–3940, 2023. DOI: 10.33022/ijcs.v12i6.3473

[19] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, 2014, pp. 216–225. DOI: 10.1609/icwsm.v8i1.14550

[20] W. Wartumi, R. Kurniawan, and Y. A. Wijaya, "Analisis Data Sentimen Ulasan Pengguna Aplikasi Shopee di Google Play Store dengan Klasifikasi Algoritma Naive Bayes," *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 164–170, 2024. DOI: 10.36499/jinrpl.v6i1.10273

[21] N. D. Derra and D. Baier, "Working in detail: How lstm hyperparameter selection influences sentiment analysis results," *Archives of Data Science, Series A*, vol. 6, no. 1, pp. 10–22, 2020. DOI: 10.5445/KSP/1000098011/10

[22] A. P. Yunico, I. Cholissodin, and D. Kurnianingtyas, "Analisis Sentimen Pemindahan Ibu Kota Indonesia pada Media Sosial Twitter menggunakan Metode LSTM dan Word2Vec," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 5, pp. 2389–2397, 2023.

[23] F. Bouchra, I. M. Agus, D. Suarjaya, N. Kadek, and D. Rusjayanthi, "Analisis Sentimen Masyarakat terhadap Tayangan Televisi Nasional menggunakan Metode Deep Learning," *Jurnal Buana Informatika*, vol. 15, no. 2, pp. 89–99, 2024.

[24] S. F. Sabbeh and H. A. Fasihuddin, "A Comparative Analysis of Word Embedding and Deep Learning for Arabic Sentiment Classification," *Electronics*, vol. 12, no. 6, pp. 1–16, 2023. DOI: 10.3390/electronics12061425

[25] Z. Hameed and B. Garcia-Zapirain, "Sentiment Classification Using a Single-Layered BiLSTM Model," *IEEE Access*, vol. 8, pp. 73 992–74 001, 2020. DOI: 10.1109/ACCESS.2020.2988550

[26] I. N. Yulita et al., "Bidirectional Long Short-Term Memory for Analysis of Public Opinion Sentiment on Government Policy During the COVID-19 Pandemic," *International Journal of Advanced Computer Science & Applications*, vol. 14, no. 11, pp. 877–885, 2023. DOI: 10.14569/IJACSA.2023.0141189