

## ANALISIS SENTIMEN ULASAN PENGGUNA GRAB DI GOOGLE PLAY STORE MENGUNAKAN METODE LSTM DAN NAÏVE BAYES

### SENTIMENT ANALYSIS OF GRAB USER REVIEWS ON THE GOOGLE PLAY STORE USING LSTM AND NAÏVE BAYES METHODS

Mahima Cinta Hage<sup>1</sup>, Pratyaksa Ocsa Nugraha Saian<sup>2</sup>

Program Studi Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga<sup>1,2</sup>

672022169@student.uksw.edu

#### ABSTRACT

Along with the advancement of digital technology, the transportation sector has undergone significant changes through the adoption of online ride-hailing applications such as Grab. The use of these applications has generated numerous user reviews on the Google Play Store, which indirectly reflect the quality of the services provided. These reviews can be leveraged through sentiment analysis to evaluate user satisfaction levels. This study focuses on evaluating and comparing the performance of Long Short-Term Memory (LSTM) and Naïve Bayes algorithms in sentiment classification of Grab application reviews. The dataset consists of 5,000 reviews obtained via web scraping, which were then processed through preprocessing stages (cleaning, tokenization, stopword removal, and stemming), sentiment labeling, and TF-IDF weighting. Naïve Bayes was used as an efficient classical machine learning method, while LSTM, as a deep learning method, is capable of understanding context and word sequences. The results of the study show that LSTM has an accuracy of 91.45%, slightly higher than Naïve Bayes at 91.33%, and is superior in detecting negative sentiment based on recall and F1-score values. These findings are expected to serve as a reference in the development of sentiment analysis systems and the evaluation of service quality for online transportation apps.

**Keywords:** Sentiment Analysis, Grab, Naïve Bayes, Long Short-Term Memory (LSTM), Text Classification.

#### ABSTRAK

Seiring dengan berkembangnya teknologi digital, sektor transportasi mengalami perubahan melalui pemanfaatan aplikasi ojek online seperti Grab. Penggunaan aplikasi ini memunculkan berbagai ulasan dari pengguna di Google Play Store, yang secara tidak langsung mencerminkan tingkat kualitas layanan yang diberikan. Ulasan tersebut dapat dimanfaatkan melalui analisis sentimen untuk mengevaluasi tingkat kepuasan pengguna. Penelitian ini difokuskan pada evaluasi dan perbandingan performa algoritma Long Short-Term Memory (LSTM) dan Naïve Bayes dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Grab. Data yang digunakan sebanyak 5.000 ulasan yang diperoleh melalui web scraping, kemudian diproses melalui tahap preprocessing (cleaning, tokenizing, stopword removal, dan stemming), pelabelan sentimen, serta pembobotan TF-IDF. Naïve Bayes digunakan sebagai metode machine learning klasik yang efisien, sedangkan LSTM sebagai metode deep learning mampu memahami konteks dan urutan kata. Hasil penelitian menunjukkan bahwa LSTM memiliki akurasi sebesar 91,45%, sedikit lebih tinggi dibandingkan Naïve Bayes sebesar 91,33%, serta lebih unggul dalam mendeteksi sentimen negatif berdasarkan nilai recall dan F1-score. Temuan ini diharapkan dapat menjadi referensi dalam pengembangan sistem analisis sentimen dan evaluasi kualitas layanan aplikasi transportasi online.

**Kata Kunci:** Analisis Sentimen, Grab Naïve Bayes, Long Short-Term Memory (LSTM), Klasifikasi Teks.

#### PENDAHULUAN

Layanan ojek online mulai berkembang di Indonesia sejak 2012 dengan memanfaatkan aplikasi digital berbasis geolokasi, sistem pencocokan otomatis, dan pembayaran non-tunai, salah satunya adalah Grab [1]. Seiring perkembangan zaman, sektor transportasi mengalami transformasi, salah satunya melalui hadirnya layanan transportasi berbasis online atau ojek online yang

semakin banyak digunakan masyarakat [2]. Layanan ojek online mulai berkembang di Indonesia sejak 2012 dengan memanfaatkan aplikasi digital berbasis geolokasi, sistem pencocokan otomatis, dan pembayaran non-tunai, salah satunya adalah Grab [3]. Teknologi ini berkembang dari layanan antar-jemput menjadi platform multi-layanan yang mencakup pengiriman barang, makanan, dan pembayaran digital dalam satu aplikasi [4]. Pengembangan

layanan ojek online didorong oleh integrasi teknologi *machine learning*, *big data*, dan *Internet of Things* (IoT), yang berperan dalam menghadirkan layanan yang lebih personal, mampu memprediksi permintaan, serta meningkatkan efisiensi dan keselamatan perjalanan [5]. Perkembangan teknologi digital meningkatkan penggunaan aplikasi transportasi *online* seperti Grab yang banyak menerima ulasan pengguna di Google Play Store. Ulasan tersebut merupakan tanggapan publik yang memuat pengalaman, kelebihan, kekurangan, serta saran, sehingga dapat menjadi sumber informasi bagi calon pengguna dan bahan evaluasi bagi pengembang dalam meningkatkan kualitas layanan.

Analisis sentimen digunakan untuk mengidentifikasi tanggapan pengguna, di mana isi ulasan mencerminkan tingkat kepuasan maupun ketidakpuasan terhadap aplikasi [6]. Pada kasus ini, analisis sentimen membantu perusahaan memahami opini masyarakat dengan mengelompokkan ulasan berdasarkan kecenderungan pendapat, sehingga dapat diketahui pandangan pengguna terhadap produk atau layanan [7]. Klasifikasi ulasan pengguna di Google Play Store dilakukan untuk mengelompokkan ulasan berdasarkan isi komentar serta nilai rating yang diberikan oleh pengguna. Tujuannya agar lebih mudah mengetahui apakah ulasan tersebut bersifat positif, negatif, atau netral. Hal ini penting karena terkadang jumlah bintang yang diberikan tidak sesuai dengan isi komentarnya, misalnya memberi bintang tinggi tetapi isi ulasannya berupa keluhan. Dengan adanya klasifikasi, pengembang aplikasi dan pengguna lain dapat lebih mudah memahami pengalaman serta penilaian pengguna terhadap aplikasi [8].

Untuk melakukan klasifikasi sentimen ini, berbagai algoritma komputasi dapat diterapkan. Di satu sisi, terdapat metode *machine learning* klasik seperti *Naïve Bayes*, yang dikenal karena kesederhanaan, kecepatan komputasi, dan

kinerjanya yang baik pada tugas klasifikasi teks berbasis probabilitas. Di sisi lain, muncul metode *deep learning* yang lebih modern seperti *Long Short-Term Memory* (LSTM), yaitu jenis Jaringan Saraf Rekuren (RNN) yang dirancang khusus untuk memahami pola dan konteks dalam data sekuensial seperti kalimat. LSTM dianggap unggul karena mampu mengenali urutan kata dan dependensi jangka panjang dalam teks, sesuatu yang menjadi kelemahan *Naïve Bayes*. Perbedaan antara metode *Naïve Bayes* yang lebih sederhana dan metode LSTM yang lebih kompleks menimbulkan pertanyaan tentang metode mana yang paling tepat digunakan untuk menganalisis sentimen pada ulasan aplikasi Grab. Hal ini penting karena ulasan pengguna biasanya singkat, menggunakan bahasa sehari-hari, dan sering memakai kata yang tidak baku. Oleh karena itu, perlu dilakukan pengujian dengan membandingkan kedua metode tersebut pada data yang sama untuk mengetahui metode mana yang memberikan hasil paling baik.

## TINJAUAN PUSTAKA

Pendekatan analisis sentimen berbasis *machine learning* dan *deep learning* telah berkembang pesat dan banyak diterapkan dalam berbagai penelitian. Sejumlah studi menunjukkan bahwa algoritma *Long Short-Term Memory* (LSTM) dan *Naïve Bayes* sama-sama mampu menghasilkan performa klasifikasi yang baik, meskipun memiliki karakteristik yang berbeda [9]. LSTM cenderung lebih unggul dalam menangani data teks karena kemampuannya memahami konteks, sedangkan *Naïve Bayes* lebih sederhana dan efisien dalam proses komputasi. Pada beberapa penelitian, LSTM menunjukkan performa yang cukup baik, seperti pada analisis sentimen mobil listrik dengan nilai *F1-score* sebesar 0,64 serta analisis opini kebijakan publik dengan akurasi mencapai 80,42% [10][11]. Hal ini mengindikasikan bahwa LSTM efektif dalam menangani teks

berbahasa alami yang memiliki struktur kompleks.

Di sisi lain, *Naïve Bayes* tetap menjadi metode yang banyak digunakan dalam analisis sentimen karena kesederhanaannya. Beberapa penelitian menunjukkan bahwa algoritma ini mampu mencapai akurasi yang tinggi, misalnya sebesar 88,9% pada analisis kepuasan pengguna layanan transportasi *online* [12]. Kelebihan utama *Naïve Bayes* terletak pada proses komputasi yang cepat serta kemudahan dalam implementasi. Meskipun memiliki asumsi independensi antar fitur yang cukup sederhana, metode ini tetap mampu memberikan hasil yang kompetitif. Oleh karena itu, *Naïve Bayes* masih sering digunakan sebagai pembanding dalam berbagai penelitian klasifikasi teks.

Perbandingan antara LSTM dan *Naïve Bayes* menunjukkan bahwa keduanya memiliki keunggulan yang saling melengkapi. LSTM lebih baik dalam memahami hubungan antar kata dan konteks kalimat, sedangkan *Naïve Bayes* unggul dalam efisiensi dan kesederhanaan model. Meskipun demikian, studi yang secara spesifik membandingkan kedua metode pada ulasan aplikasi Grab di Google Play Store masih terbatas. Hal ini menunjukkan adanya peluang penelitian untuk mengeksplorasi performa kedua algoritma dalam domain tersebut. Oleh karena itu, penelitian ini dilakukan untuk menganalisis serta membandingkan kinerja LSTM dan *Naïve Bayes* dalam klasifikasi sentimen.

Secara konseptual, analisis sentimen merupakan bagian dari *text mining* dan *natural language processing* (NLP) yang bertujuan untuk mengidentifikasi dan mengelompokkan opini dalam bentuk teks ke dalam kategori tertentu, seperti positif, netral, dan negatif [13]. Proses ini tidak hanya melibatkan klasifikasi, tetapi juga tahapan lain seperti pengumpulan data, preprocessing, ekstraksi fitur, dan pembobotan kata. *Text mining* sendiri berperan dalam mengolah data teks yang tidak terstruktur menjadi informasi yang

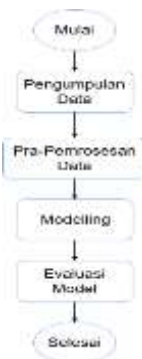
dapat dianalisis menggunakan pendekatan komputasi [14]. Dengan pendekatan tersebut, data ulasan yang sebelumnya tidak terstruktur dapat diubah menjadi informasi yang lebih bermakna. Oleh karena itu, analisis sentimen banyak digunakan untuk memahami persepsi pengguna terhadap suatu layanan.

Dalam proses pengumpulan data, teknik *web scraping* sering digunakan untuk mengambil data ulasan secara otomatis dari halaman *web*. Teknik ini memungkinkan pengambilan data dalam jumlah besar, termasuk data yang bersifat dinamis [15]. Pada tahap pemodelan, LSTM sebagai bagian dari *Recurrent Neural Network* (RNN) dirancang untuk memproses data sekuensial dan mampu mempertahankan informasi dalam jangka panjang [16]. Kemampuan tersebut membuat LSTM lebih efektif dalam memahami konteks kalimat yang kompleks. Sementara itu, *Naïve Bayes* merupakan algoritma berbasis probabilitas yang memanfaatkan *Teorema Bayes* dengan asumsi independensi antar fitur [17]. Dengan pendekatan tersebut, *Naïve Bayes* menjadi metode yang ringan dan efisien untuk klasifikasi teks.

Untuk mendukung proses klasifikasi, data teks biasanya diubah ke dalam bentuk numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) [18]. Metode ini memberikan bobot pada kata berdasarkan frekuensi kemunculan serta tingkat keunikannya dalam suatu dokumen. Sebelum proses pembobotan dilakukan, data terlebih dahulu melalui tahap *preprocessing* seperti *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Tahapan tersebut bertujuan untuk membersihkan data dan mengurangi *noise* sehingga kualitas data menjadi lebih baik. Dengan representasi numerik yang dihasilkan, algoritma klasifikasi dapat bekerja secara lebih optimal. Oleh karena itu, TF-IDF menjadi salah satu teknik yang umum digunakan dalam analisis sentimen berbasis teks.

## METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif yang berfokus pada analisis sentimen. Fokus utama penelitian adalah mengkaji kecenderungan opini pengguna terhadap aplikasi Grab melalui ulasan yang tersedia di Google Play Store dengan menggunakan algoritma LSTM dan *Naïve Bayes*. Penerapan kedua algoritma ini bertujuan untuk membandingkan kinerja dalam proses klasifikasi sentimen. Penelitian ini bersifat deskriptif kuantitatif karena hasil yang diperoleh disajikan dalam bentuk data numerik serta visualisasi grafis. Data yang digunakan adalah data sekunder, yaitu ulasan pengguna terhadap aplikasi Grab yang diambil dari Google Play Store. Data ini bersifat tekstual (berbentuk kalimat/paragraf), dan tersedia secara publik. Gambar 1 merupakan diagram tahapan rancangan penelitian yang akan dilakukan.



**Gambar 1. Diagram Alir Penelitian**

Data dikumpulkan melalui *web scraping* dengan menentukan URL target yaitu halaman aplikasi Grab pada Google Play Store, lalu menentukan jumlah ulasan yang akan diambil sekitar 5000 ulasan, kemudian menggunakan *Python* dan library *PlayScraper* untuk mengambil data ulasan, nama pengguna, rating, dan tanggal, kemudian menyimpan data hasil *scraping* dalam format CSV atau JSON. Jika sudah mengumpulkan data lanjut pra-pemrosesan data. Dalam proses pra-pemrosesan data teks, beberapa tahapan penting yang dilakukan, antara lain tahapan pertama adalah normalisasi kata dengan cara mengubah kata-kata yang tidak sesuai dengan Ejaan Yang Disempurnakan (EYD) pada Bahasa Indonesia menjadi bentuk

yang benar, dan dilanjutkan dengan *cleansing*, yaitu menghapus URL, tanda baca, angka, simbol, emoji, dan spasi berlebih yang tidak diperlukan dalam analisis. Setelah itu, dilakukan *case folding*, yaitu mengubah semua huruf dalam teks menjadi huruf kecil agar konsistensi data terjaga. Proses berikutnya adalah tokenisasi atau *tokenizing*, yaitu memecah kalimat menjadi kata-kata terpisah sehingga lebih mudah dianalisis. Kemudian, dilakukan *stopword removal* untuk menghapus kata-kata umum yang tidak memiliki makna penting, seperti "dan", "yang", atau "di". Terakhir, dilakukan *stemming*, yaitu mengubah kata-kata ke bentuk dasarnya menggunakan library NLP seperti Sastrawi untuk bahasa Indonesia. Seluruh tahapan ini bertujuan untuk menghasilkan data teks yang bersih dan siap untuk dianalisis lebih lanjut.

Kemudian dilakukan pelabelan data dengan menggunakan metode *Rule Based*, yang berdasarkan skor *rating* yang telah ada pada data ulasan tersebut, metode ini bekerja dengan mengubah skor numerik tersebut menjadi kelas sentimen baik positif, negatif, maupun netral. Kemudian *Term Frequency-Inverse Document Frequency* (TF-IDF) diolah secara matematis dan digunakan dalam model *machine learning*. Setelah data dibersihkan dan diberi label, dilakukan analisis sentimen dengan menggunakan metode LSTM dan *Naïve Bayes Classifier*, sebuah algoritma *machine learning* berbasis probabilitas yang sering digunakan untuk klasifikasi teks. Selanjutnya model akan dievaluasi dengan presisi atau *precision*, *recall*, *F1-score*, *support*, dan *confusion matrix*. Teknik visualisasi data akan ditampilkan dalam bentuk grafik distribusi sentimen pengguna (*bar chart* atau *pie chart*), *word cloud* untuk kata-kata dominan tiap kategori sentimen, grafik *loss* dan akurasi selama pelatihan (*training vs validation*), dan tabel *confusion matrix*.

**Tabel 1. Aturan untuk Tabel**

No	Elemen	Keterangan
----	--------	------------

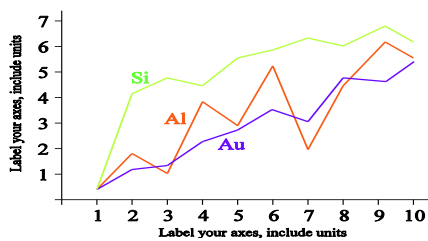
1	Penomoran	Dimulaidari 1, 2, dan 3, bukan 1.1, 1.2, dan 1.3
2	Nomordanjudultabel	Ditulis di bagianatastabel
3	Garis	Garis yang digunakanhanya horizontal
4	Judulkolom	Ditulisdengantebal

Untuk tabel, Peraturannya adalah (1) penomoran tabel dimulai dari 1, 2, 3, dan seterusnya, bukan 1.1, 1.2, dan seterusnya, (2) Nomor dan judul tabel ditulis pada bagian atas tabel berjarak 1 enter dari baris terakhir isi di atasnya dengan huruf times new roman, ukuran 10, tebal, sejajar dengan kiri tabel, dan setiap huruf pertama ditulis dengan huruf kapital dengan bentuk “Judul Tabel”, (3) Tabel hanya menggunakan garis horizontal, tanpa vertikal, (4) Judul kolom ditulis menggunakan huruf tebal, dan (5) Sumber tabel ditulis di bagian bawah tabel sejajar dengan bagian tepi kiri dari tabel menggunakan huruf times new roman ukuran 10.

**Tabel 2. Aturan Margin**

Margin	A4 Paper	US Letter Paper
Left	18.5 mm	14.5 mm
Right	18mm	13 mm (0.51

Sumber : ditulis miring dan ukuran 9



Sumber :

sumbergambarditulisdenganukuran 9

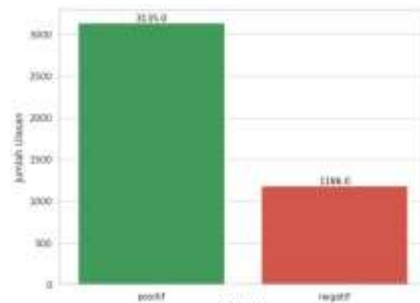
**Gambar 1. Warna Gambar Harus Kontras**

## HASIL DAN PEMBAHASAN

Berdasarkan hasil data yang digunakan 5000 data ulasan *real-time* dari pengguna aplikasi Grab yang diambil melalui *scraping data* pada Google Play Store, yang sebelumnya data awal masih belum diolah atau *raw* akan diolah dengan tahap pra-pemrosesan data yang meliputi

*cleaning data*, kemudian *tokenizing*, *stopwords*, *stemming*, dan *labeling data*. Data analisis sentimen tersebut diambil pada saat Natal dan Tahun Baru (Nataru) 2025-2026, untuk mengetahui opini masyarakat secara umum terhadap layanan aplikasi Grab pada periode tersebut. Pemilihan periode Nataru karena merupakan waktu interaksi publik yang tinggi, yang menghasilkan *volume data* opini yang signifikan dan relevan.

Setelah seluruh tahapan pra-pemrosesan dilakukan, data kemudian difilter lebih lanjut dengan membuang ulasan yang memiliki sentimen netral serta baris teks yang kosong. Proses ini dilakukan agar analisis sentimen yang dibangun lebih fokus pada dua polaritas utama, yaitu positif dan negatif, sehingga model dapat menghasilkan klasifikasi yang lebih jelas dan terarah. Hasil dari tahapan ini menunjukkan bahwa dari total 5.000 data awal, tersisa sekitar 4.321 ulasan yang layak digunakan untuk tahap analisis lebih lanjut. Data yang tersisa terdiri dari 3.135 ulasan positif dan 1.186 ulasan negatif, yang kemudian menjadi *dataset* akhir untuk proses evaluasi dan pemodelan sentimen. Analisis sentimen ulasan pengguna aplikasi Grab saat libur natal dan tahun baru periode 2025-2026 didominasi oleh ulasan positif dengan total 72,55% dari jumlah data ulasan yang digunakan. Beberapa ulasan yang ditangkap tidak sedikit yang mengulas pelayanan berdasarkan libur natal dan tahun baru yang ditandai dengan “libur”, dan “tahun baru”.



**Gambar 2. Perbandingan Jumlah Data Ulasan**

Hasil visualisasi diagram menunjukkan kombinasi kata yang paling sering muncul pada setiap kategori



Telpon	0	0	1	0	0	1195084	$-\log_5 = \frac{5}{1195084} = 6.846$	0	0	0.598	0	0
Aplikasi	0	0	0	1	0	35045	$-\log_5 = \frac{5}{35045} = 3.316$	0	0	0	0.599	0
Tanggung	0	0	0	0	1	919295	$-\log_5 = \frac{5}{919295} = 6.583$	0	0	0	0	0.636
Jajan	0	1	0	0	0	3983615	$-\log_5 = \frac{5}{3983615} = 8.050$	0	0.681	0	0	0
Fungsi	0	0	1	0	0	3983615	$-\log_5 = \frac{5}{3983615} = 8.050$	0	0	0.704	0	0
Baik	0	0	0	1	0	107664	$-\log_5 = \frac{5}{107664} = 4.439$	0	0	0	0.801	0

Setelah seluruh tahapan pra-pemrosesan dan pembobotan TF-IDF selesai dilakukan, langkah berikutnya adalah proses perbandingan antara metode *Naïve Bayes* dan LSTM, dengan memulai membangun model menggunakan masing-masing algoritma. Klasifikasi menggunakan algoritma *Naïve Bayes* dan LSTM dilakukan dengan memanfaatkan 4.321 data yang telah melewati pra-pemrosesan data dan *labeling* untuk membangun model prediksi sentimen. Selanjutnya setelah seluruh data berhasil dimuat dan melalui proses label *encoding* dengan fungsi *LabelEncoder* dalam *library scikit-learn*, dilakukan pembagian *dataset* menjadi data latih sebanyak 3.456 ulasan dan data uji sebanyak 865 ulasan. data pada Tabel 2 menunjukkan *confusion matrix* yang dihasilkan *Naïve Bayes*, sementara LSTM pada Tabel 3.

**Tabel 2. Hasil Confusion Matrix Naïve Bayes**

		Prediksi	
		Negatif	Positif
Sebenarnya	Negatif	188	49
	Positif	26	602

**Tabel 3. Hasil Confusion Matrix LSTM**

		Prediksi	
		Negatif	Positif
Sebenarnya	Negatif	203	34
	Positif	40	588

Sebenarnya	Negatif	203	34
	Positif	40	588

Pada Tabel 2, model *Naïve Bayes* menunjukkan kinerja klasifikasi dengan 602 ulasan positif dan 188 ulasan negatif yang berhasil diprediksi secara tepat, meskipun masih terdapat kesalahan berupa 26 *false negative* dan 49 *false positive*. Sementara itu, pada Tabel 3, model LSTM mampu mengklasifikasikan 588 ulasan positif dan 203 ulasan negatif dengan benar, dengan jumlah kesalahan sebesar 40 *false negative* dan 34 *false positive*. Berdasarkan evaluasi menggunakan *confusion matrix*, kedua model menunjukkan performa yang tinggi dalam klasifikasi sentimen ulasan memiliki keunggulan yang sedikit lebih baik dibandingkan *Naïve Bayes*, yang ditunjukkan oleh nilai akurasi sebesar 91,45%, lebih tinggi dibandingkan *Naïve Bayes* dengan akurasi 91,33%. Nilai akurasi tersebut merepresentasikan proporsi prediksi yang benar terhadap keseluruhan data uji, baik pada kelas positif maupun negatif.

$$Accuracy \ Naïve \ Bayes = \frac{TP + TN}{TP + FP + TN + FN} = \frac{602 + 188}{602 + 49 + 188 + 26} = \frac{790}{865} = 0,9133$$

$$Accuracy \ LSTM = \frac{TP + TN}{TP + FP + TN + FN} = \frac{588 + 203}{588 + 34 + 203 + 40} = \frac{791}{865} = 0,9145$$

Selanjutnya, presisi yaitu menguji seberapa akurat model algoritma *Naive Bayes* dan LSTM dalam mengklasifikasikan kelas tertentu. Fokus diletakkan pada kelas negatif karena keluhan merupakan sumber umpan balik yang dapat ditindaklanjuti untuk perbaikan layanan pada platform Grab, di mana akurasi prediksi sangat penting. Untuk kelas negatif, presisi dari *Naive Bayes* adalah 87.85%. Angka ini berarti, dari semua ulasan yang diprediksi negatif oleh model, sementara itu LSTM mendapatkan angka 83.54%. Presisi yang tinggi menunjukkan tingkat kesalahan tipe I (*false positive*) yang rendah, memastikan bahwa model jarang salah mengklasifikasikan ulasan positif sebagai keluhan yang perlu diselesaikan dengan solusi. Seperti salah satu contoh data ulasan yang sudah diolah “kecewa menit pesan grab tolak driver berbagai alasan” dikategorikan sebagai ulasan negatif karena secara eksplisit mengandung keluhan terhadap aspek layanan Grab yang lama, baik dari sisi pengemudi, sistem aplikasi.

Dalam konteks ini, presisi kelas negatif menunjukkan sejauh mana model *Naive Bayes* dan LSTM mampu secara konsisten mengklasifikasikan ulasan semacam ini sebagai negatif tanpa mencampurkannya dengan ulasan netral atau positif. Ketepatan klasifikasi tersebut penting agar pihak penyedia layanan tidak salah menginterpretasikan sentimen pengguna, sehingga rekomendasi perbaikan yang dihasilkan benar-benar didasarkan pada keluhan aktual yang dialami pengguna.

$$\text{Presisi (Negatif) Naive Bayes} = \frac{TN}{TN + FN} = \frac{188}{188 + 26} = \frac{188}{214} = 0,8785$$

$$\text{Presisi (Negatif) LSTM} = \frac{TN}{TN + FN} = \frac{203}{203 + 40} = \frac{203}{243} = 0,8354$$

*Recall* (atau sensitivitas) mengukur kemampuan model untuk menemukan semua kasus relevan. *Recall* memastikan bahwa keluhan penting tidak luput

terklasifikasi, meskipun bahasanya halus atau ambigu. Contoh dari data ulasan yang diambil "Cari kemudi dekat susah ya saat libur sekolah lempar lempar cari kemudinya temu kemudi lokasinya tolong baik layan grabnya" Pada periode Nataru, *recall* kelas negatif menunjukkan kemampuan model dalam menangkap seluruh keluhan pengguna yang meningkat akibat lonjakan permintaan dan kondisi lalu lintas. *Recall* yang rendah mengindikasikan masih adanya ulasan negatif, khususnya yang disampaikan secara implisit, yang gagal teridentifikasi oleh model. Untuk kelas negatif, *recall* dari *Naive Bayes* adalah 79.32%, dan LSTM 85,65%. Angka tersebut berarti dari semua ulasan yang sebenarnya negatif. Nilai *recall* yang tinggi sangat penting dalam analisis sentimen keluhan, karena menunjukkan risiko yang minim dalam melewatkan ulasan penting dalam Grab (*false negative*).

$$\text{Recall (negatif) Naive Bayes} = \frac{TN}{TN + FP} = \frac{188}{188 + 49} = \frac{188}{237} = 0,7932$$

$$\text{Recall (negatif) LSTM} = \frac{TN}{TN + FP} = \frac{203}{203 + 34} = \frac{203}{237} = 0,8565$$

*F1-score* digunakan untuk mengevaluasi keseimbangan antara presisi dan *recall* dalam mengklasifikasikan ulasan negatif Grab. Nilai *F1-score* yang tinggi menunjukkan bahwa model tidak hanya akurat dalam memprediksi keluhan pengguna, tetapi juga mampu menangkap sebagian besar ulasan negatif yang ada, khususnya pada periode dengan intensitas keluhan tinggi seperti Nataru. Pada *Naive Bayes* *F1-score* positif 94.14%, dan negatif 83.37%. Sedangkan pada LSTM *F1-score* positif 94.08%, dan negatif 84.58%. Angka ini memberikan gambaran tunggal yang kuat mengenai efektivitas model secara keseluruhan pada kelas tersebut, karena mempertimbangkan *trade-off* antara presisi dan *recall*.

$$\text{F1-score Naive Bayes} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{0,8785 \times 0,7932}{0,8785 + 0,7932} = 0,8337 \times 100\% =$$

83.37%

$$\begin{aligned}
 \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \\
 &= \frac{0,8354 \times 0,8565}{0,8354 + 0,8565} = 0,8458 \times 100\% = \\
 &84.58\%
 \end{aligned}$$

Dengan berikut, penelitian ini menunjukkan bahwa ulasan Grab yang dikumpulkan pada periode Nataru didominasi oleh sentimen positif, namun sentimen negatif merepresentasikan keluhan pengguna akibat lonjakan permintaan layanan, seperti keterlambatan pengemudi, susah mencari pengemudi, kenaikan tarif, dan pembatalan pesanan. Hasil evaluasi mengindikasikan bahwa *Naïve Bayes* memiliki presisi kelas negatif yang lebih stabil, sedangkan LSTM unggul dalam *recall* kelas negatif dengan kemampuan menangkap variasi keluhan yang lebih kompleks. Nilai *F1-score* menunjukkan bahwa LSTM memberikan keseimbangan yang lebih baik antara presisi dan *recall* dalam mengklasifikasikan ulasan negatif. Dengan demikian, fokus pada evaluasi kelas negatif memberikan kontribusi penting dalam mengidentifikasi keluhan yang relevan sebagai dasar perbaikan kualitas layanan Grab, khususnya pada periode Nataru.

## SIMPULAN

Berdasarkan hasil analisis sentimen terhadap ulasan pengguna aplikasi Grab di Google Play Store pada periode Natal dan Tahun Baru (Nataru), dapat disimpulkan bahwa algoritma *Long Short-Term Memory* (LSTM) dan *Naïve Bayes* sama-sama mampu mengklasifikasikan sentimen dengan tingkat kinerja yang tinggi. Hasil pengujian menunjukkan bahwa model LSTM memperoleh akurasi sebesar 91,45%, sedangkan *Naïve Bayes* mencapai 91,33%, sehingga perbedaan performa keduanya tergolong kecil dan menunjukkan hasil yang relatif sebanding. Meskipun demikian, dalam hal mendeteksi sentimen negatif, LSTM menunjukkan keunggulan dibandingkan *Naïve Bayes*. Hal ini terlihat

dari nilai *recall* negatif dan *F1-score* negatif yang lebih tinggi, yang mengindikasikan bahwa LSTM lebih efektif dalam mengidentifikasi ulasan yang mengandung keluhan atau kritik dari pengguna. Kemampuan ini menjadi penting karena ulasan negatif umumnya merepresentasikan permasalahan layanan yang perlu mendapatkan perhatian lebih lanjut.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa LSTM dapat menjadi pilihan metode yang lebih optimal dalam analisis sentimen, meskipun keunggulannya terhadap *Naïve Bayes* tidak terlalu signifikan. Oleh karena itu, untuk penelitian selanjutnya disarankan agar menggunakan *dataset* dengan jumlah yang lebih besar atau rentang waktu yang lebih panjang agar hasil yang diperoleh dapat merepresentasikan kondisi yang lebih luas. Selain itu, penambahan variasi algoritma lain dalam proses perbandingan juga dapat dilakukan untuk memperoleh gambaran performa model yang lebih komprehensif. Dengan demikian, hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan metode analisis sentimen yang lebih akurat dan aplikatif.

## DAFTAR PUSTAKA

- [1] M. Ratri, "Analisis Efektivitas Ojek Online Sebagai Pilihan Moda Transportasi Di Kecamatan Tamalanrea," vol. 4, pp. 2655–7266, 2022, [Online]. Available: <https://mail.jurnal.ft.umi.ac.id/index.php/JILMATEKS>
- [2] R. R. F. Tambunan, J. I. Sihotang, and J. Y. Mambu, "Analisa Tingkat Kepuasan Kerja Driver Maxim Terhadap Sistem Layanan Maxim Dengan Pieces Framework," *CogITO Smart J.*, vol. 7, no. 2, pp. 339–348, 2021, doi: 10.31154/cogito.v7i2.330.339-348.
- [3] H. Taufik, R. Setiawan, and Elianora, "Analisa Kelayakan Finansial Driver Ojek Online (Studi Kasus Grabbike

- Pekanbaru),” *Saintenk (E-Journal)*, vol. 11, no. 1, 2023.
- [4] M. Rizki, T. B. Joewono, and Y. O. Susilo, *The influence of app function evolution on transport SuperApp use behaviour over time*, no. 0123456789. Springer US, 2024. doi: 10.1007/s11116-024-10485-6.
- [5] E. Özer, M. İskefiyeli, and J. Azimjonov, “Toward lightweight intrusion detection systems using the optimal and efficient feature pairs of the Bot-IoT 2018 dataset,” *Int. J. Distrib. Sens. Networks*, vol. 17, no. 10, 2021, doi: 10.1177/15501477211052202.
- [6] A. Simanungkalit, J. P. P. Naibaho, and A. De Kweldju, “Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Shopee Menggunakan Algoritma Naïve Bayes,” *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, p. 659, 2024, doi: 10.35889/jutisi.v13i1.1826.
- [7] P. A. Kusnadi *et al.*, “FREE FLOW MENGGUNAKAN NAÏVE BAYES CLASSIFIER,” vol. 8, no. 4, pp. 7398–7404, 2024.
- [8] D. Fristtikasari, S. Alam, and I. Kurniawan, “Analisis Sentimen Pengguna Aplikasi Kitalulus pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes,” *J. Teknol. Inform. dan Komput.*, vol. 10, no. 2, pp. 458–473, 2024, doi: 10.37012/jtik.v10i2.2244.
- [9] A. R. Isnain, H. Sulistiani, B. M. Hurohman, A. Nurkholis, and S. Styawati, “Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen,” *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 2, p. 299, 2022, doi: 10.26418/jp.v8i2.54704.
- [10] A. S. Widagdo, K. N. Qodri, F. E. N. S, and N. A. R. P, “Analisis Sentimen Mobil Listrik di Indonesia Menggunakan Long-Short Term Memory ( LSTM ),” vol. 13, no. 3, pp. 416–423, 2023.
- [11] S. J. Pipin and H. Kurniawan, “Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat di Twitter Menggunakan LSTM,” vol. 23, no. 2, pp. 197–208, 2022.
- [12] R. Rahman and F. A. Sutanto, “Data Mining Untuk Memprediksi Tingkat Kepuasan Konsumen Gojek Menggunakan Algoritma Naive Bayes,” *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 18, no. 1, pp. 8–18, 2023, doi: 10.35969/interkom.v18i1.280.
- [13] H. Chyntia Morama, D. E. Ratnawati, and I. Arwani, “Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 1702–1708, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [14] M. N. Muttaqin and I. Kharisudin, “Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor,” *UNNES J. Math.*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [15] S. A. Kumar, M. M. Nasralla, I. García-Magariño, and H. Kumar, “A machine-learning scraping tool for data fusion in the analysis of sentiments about pandemics for supporting business decisions with human-centric AI explanations,” *PeerJ Comput. Sci.*, vol. 7, pp. 1–18, 2021, doi: 10.7717/PEERJ-CS.713.
- [16] L. M. Azizah, D. B. Ajipratama, N. A. R. Putri, and C. Damarjati, “Analisa Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 Di Indonesia Pada Twitter Menggunakan Algoritma LSTM La,” *J. IPTEKKOM J. Ilmu Pengetah.*

- Teknol. Inf.*, vol. 24, no. 2, pp. 161–172, 2022, doi: 10.17933/iptekkom.24.2.2022.161-172.
- [17] M. K. Insan, U. Hayati, and O. Nurdiawan, “Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di,” *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 478–483, 2023.
- [18] A. I. Tanggraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.