

Perbandingan Model BERT dan RNN-LSTM pada Analisis Sentimen Aplikasi BRI Mobile

Dea Yuliana Ayu Ningrum^{1a)}, Erna Daniati^{*2a)}, M. Najibulloh Muzaki^{3a)}

^{a)} Sistem Informasi,
Fakultas Teknik dan Ilmu Komputer,
Universitas Nusantara PGRI Kediri

Author Emails

^{2,a)} Corresponding author: ernadaniati@unpkediri.ac.id

^{1,a)} deayulianaayuningrum@gmail.com

^{3,a)} m.n.muzaki@gmail.com

Abstract. This study seeks to evaluate and contrast the effectiveness of two deep learning architectures—BERT (Bidirectional Encoder Representations from Transformers) and RNN-LSTM (Recurrent Neural Network – Long Short-Term Memory)—in the task of sentiment analysis for user feedback on the BRI Mo mobile application. The dataset, comprising 10,000 user reviews, was extracted using web scraping techniques from Google Play via random sampling out of an approximate pool of one million available reviews. The initial preprocessing steps included text cleaning, removal of symbols, numbers, URLs, and tokenization. Initial evaluation showed BERT achieved an accuracy of 54%, while RNN-LSTM reached 53%. Further experiments were conducted by removing the additional tokenization during preprocessing. The results indicated a significant improvement, with BERT reaching 73% accuracy and RNN-LSTM 70%. These findings suggest that redundant tokenization can degrade input quality. Overall, BERT demonstrated superior performance in understanding the linguistic context of the Indonesian language, especially in handling ambiguity and complex sentence structures in user-generated review texts.

Keywords: *Sentiment analysis; BERT; RNN-LSTM; BRI Mobile.*

Abstrak. Penelitian ini dimaksudkan untuk mengevaluasi serta membandingkan performa dari dua arsitektur deep learning, yakni BERT (Bidirectional Encoder Representations from Transformers) dan RNN-LSTM (Recurrent Neural Network – Long Short-Term Memory), dalam mengklasifikasikan sentimen pada tanggapan pengguna aplikasi BRI Mo. Kumpulan data diperoleh melalui teknik web scraping di platform Google Play dengan metode pengambilan acak (random sampling), sehingga terkumpul 10.000 ulasan dari total sekitar satu juta ulasan yang tersedia. Proses preprocessing awal mencakup pembersihan teks, penghapusan simbol, angka, URL, serta tokenisasi. Evaluasi awal menunjukkan bahwa model BERT memiliki akurasi sebesar 54%, sedangkan RNN-LSTM memperoleh akurasi 53%. Selanjutnya, dilakukan eksperimen lanjutan dengan menghilangkan proses tokenisasi tambahan pada preprocessing. Hasilnya, akurasi meningkat secara signifikan menjadi 73% untuk BERT dan 70% untuk RNN-LSTM. Peningkatan ini menunjukkan bahwa tokenisasi ganda dapat menurunkan kualitas input ke dalam model. Secara keseluruhan, model BERT terbukti lebih unggul dalam memahami konteks linguistik dalam bahasa Indonesia, terutama dalam menangani ambiguitas dan struktur kalimat kompleks dalam teks ulasan pengguna aplikasi.

Kata kunci: Sentimen analisis; BERT; RNN-LSTM; BRI Mobile

PENDAHULUAN

BRImo merupakan platform aplikasi seluler yang dirancang oleh Bank Rakyat Indonesia (BRI) guna memberikan kemudahan bagi pengguna dalam menjalankan beragam aktivitas perbankan[1]. Menurut ketentuan Bank Indonesia Nomor 9/15/PBI/2007, layanan perbankan digital adalah fasilitas yang diberikan oleh bank guna mempermudah nasabah dalam mengakses data, berkomunikasi, serta melaksanakan transaksi perbankan melalui berbagai media elektronik, seperti ATM, layanan telepon perbankan, transfer dana secara elektronik, internet *banking*, dan telepon genggam[1]. Dalam proses pembuatan aplikasi, perusahaan perlu memahami bagaimana reaksi pengguna terhadap layanan BRImo. Salah satu metode yang diterapkan adalah analisis sentimen pada komentar pengguna. Analisis ini bertujuan untuk mengetahui apakah pandangan yang diungkapkan bernada positif atau negatif, sehingga hasilnya bisa dijadikan bahan evaluasi demi peningkatan mutu layanan BRImo di masa mendatang.

Pandangan pengguna menyebar luas di berbagai *platform* media sosial, di mana setiap individu mengungkapkan pikiran dan emosinya lewat komentar daring dan aplikasi digital. Pandangan tersebut dapat dibagi menjadi tiga kategori utama, yakni positif, negatif, dan netral. Pembagian jenis pandangan ini dikenal dengan istilah analisis sentimen[2][3]. Dalam kajian sentimen, metode *Deep Learning* dianggap sangat ampuh dalam mengatasi permasalahan rumit, sekalipun diaplikasikan pada data yang sederhana. Keunggulan ini muncul dari mekanisme berlapis-lapis yang mampu mengolah informasi nonlinier, sehingga model dapat mengekstrak ciri-ciri penting, mendeteksi pola, dan mengklasifikasikan dengan tingkat presisi yang lebih tinggi[4].

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa dua arsitektur *deep learning*, yaitu BERT (*Bidirectional Encoder Representations from Transformers*) dan RNN-LSTM (*Recurrent Neural Network – Long Short-Term Memory*), dalam mengelompokkan sentimen ke dalam kategori positif, negatif, maupun netral. Pengelompokan tersebut merupakan tahap identifikasi pola atau fungsi spesifik yang mampu membedakan serta merepresentasikan masing-masing kelas data secara tepat[5]. Selain itu, penelitian ini juga mempertimbangkan tantangan seperti *overfitting* dan mengoptimalkan *hyperparameter* agar model dapat memberikan hasil yang lebih relevan dalam merespon umpan balik pengguna secara efektif dengan menggabungkan dua model canggih, yaitu BERT dan RNN-LSTM.

BERT memiliki beragam model *pre-trained*, termasuk BERT Multilingual yang dirancang untuk berbagai bahasa[6]. Namun, model ini kurang optimal untuk tugas yang berfokus pada satu bahasa, karena tidak memiliki mekanisme untuk secara spesifik mengenali atau memilih bahasa tertentu. Akibatnya, *tokenizer* dapat secara tidak sengaja mencampurkan kata dari berbagai bahasa, yang dapat memengaruhi akurasi analisis[7]. Tak hanya itu, RNN konvensional memiliki keterbatasan seperti masalah *exploding* dan *vanishing gradient*. LSTM, sebagai salah satu varian dari RNN, dirancang untuk mengatasi kelemahan tersebut melalui penggunaan mekanisme gerbang. Oleh karena itu, LSTM lebih optimal dibandingkan RNN konvensional[8].

Penelitian ini bertujuan menghasilkan klasifikasi sentimen pengguna serta membandingkan performa dua model *deep learning*, yakni BERT dan RNN-LSTM. Perbandingan dilakukan untuk mengidentifikasi model yang paling unggul dalam mengolah analisis sentimen berdasarkan akurasi dan hasil evaluasi model, untuk mengidentifikasi kinerja unggulan dari model dalam mengolah analisis sentimen pada *feedback* pengguna aplikasi BRI Mobile.

TINJAUAN PUSTAKA

Analisis Sentimen

Analisis sentimen adalah metode untuk mengidentifikasi tipe pendapat atau emosi yang terkandung dalam suatu kalimat, cuitan, atau sekumpulan teks. Proses ini diterapkan pada konten media sosial guna menilai serta mengkaji seberapa besar kepuasan konsumen terhadap produk yang ditawarkan oleh sebuah perusahaan. Ini membuat perusahaan tertarik untuk menggunakan hasil analisis ini dalam menentukan kebijakan sehingga mempengaruhi jalannya bisnis dan mendukung tujuan organisasi [3].

Sentimen analisis adalah rangkaian langkah yang bertujuan mengidentifikasi serta mengklasifikasikan perasaan atau pandangan dalam sebuah tulisan, baik bersifat positif, negatif, maupun netral, dengan menggunakan teknik pengolahan teks. [9]. Sentimen analisis dapat digunakan untuk menyimpulkan konten teks dari setiap opini dalam sebuah ulasan, apakah pengguna menyukai atau tidak menyukai pengalaman mereka saat menggunakan aplikasi. Dengan memahami preferensi pengguna melalui ulasan, pengembang dapat meningkatkan aspek kegunaan aplikasi

untuk memperkuat keberhasilan bisnis. [10]. Analisis sentimen biasanya dilakukan dengan memanfaatkan Natural Language Processing (NLP) [11].

BERT

BERT (*Bidirectional Encoder Representations from Transformers*) adalah sebuah model pembelajaran mendalam yang dirancang untuk menangkap arti kata melalui konteksnya dalam tahap pelatihan awal, guna mendukung berbagai aplikasi pemrosesan bahasa alami (NLP). Model ini dibangun dengan pelatihan menggunakan kumpulan data Wikipedia yang mencakup 104 bahasa berbeda. Dikembangkan oleh *Google* dan pertama kali dikenalkan pada 2018, proses pelatihan BERT mengadopsi metode *Masked Language Model* (MLM) dan menggunakan arsitektur transformer yang beroperasi secara dua arah (*bidirectional*) [7].

Bidirectional Encoder Representation from Transformers (BERT) adalah teknik pembelajaran mesin berbasis *transformator* yang dikembangkan oleh *Google* [12]. Tahap pemecahan token dalam BERT mencakup penyisipan token khusus seperti [CLS] yang berfungsi untuk klasifikasi dan [SEP] sebagai pemisah, masing-masing memiliki peran spesifik. Proses tokenisasi ini diterapkan pada level kata maupun karakter, sehingga model mampu menangkap makna konteks serta hubungan antar karakter dalam teks. Pendekatan ini mendukung BERT dalam mengelola kata-kata yang rumit atau beragam, sekaligus mempertahankan struktur informasi yang esensial [13].

Fine Tuning BERT

Penyempurnaan mendalam pada BERT menjadi langkah krusial dalam merancang model yang lebih akurat guna keperluan analisis sentimen [14]. Studi ini menggunakan model BERT *base-uncased* yang telah melewati fase pra-pelatihan. Struktur *fine-tuning* yang diterapkan terdiri dari beberapa elemen, termasuk lapisan input, lapisan transformer yang terbentuk dari beberapa blok, serta lapisan *output*. Dalam proses *fine-tuning*, sejumlah parameter seperti laju pembelajaran, jumlah *epoch*, dan ukuran batch diatur ulang guna memaksimalkan performa model [15].

RNN

Arsitektur *Recurrent Neural Network* (RNN) yang dipadukan dengan *Long Short Term Memory* (LSTM) adalah salah satu teknik *deep learning* efektif untuk pengklasifikasian sentimen. Struktur ini dibuat khusus untuk memproses data berurutan, seperti teks, suara, dan video. [16]. RNN, atau jaringan *neural* umpan balik, adalah tipe jaringan neural yang memiliki koneksi *berloop* dalam strukturnya. Hal ini memungkinkan *output* dari jaringan untuk dipakai kembali sebagai input guna menghasilkan *output* berikutnya. RNN dibuat khusus untuk memproses data yang memiliki sifat berurutan atau sekuensial [17]. Berikut adalah persamaan dari RNN untuk memproses data sekuensial dengan mempertimbangkan hubungan temporal antara elemen data dalam urutan tertentu, misalnya dalam sentiment analysis [18]:

$$h_q = \tanh(W_{hx}x_q + W_{hh}h_{q-1} + b_h) \quad (1)$$

$$o_q = W_{oh}h_q + b_o \quad (2)$$

LSTM

Long Short Term Memory (LSTM) LSTM adalah sebuah penyempurnaan dari algoritma Recurrent Neural Network (RNN) yang mengadaptasi arsitektur RNN dengan memasukkan komponen sel memori atau unit penyimpanan khusus. Pendekatan ini memungkinkan LSTM menyimpan informasi yang diperoleh dalam rentang waktu yang lebih panjang secara efektif. [19]. Sebagai evolusi dari Recurrent Neural Network (RNN), LSTM unggul dalam mengatasi tantangan ketergantungan jangka panjang. Pertama kali diperkenalkan oleh Sepp Hochreiter dan Jurgen Schmidhuber pada tahun 1997, model ini mengadopsi serangkaian sel memori eksklusif yang menggantikan neuron tradisional pada lapisan tersembunyi di RNN. [20]. Rumus LSTM dibagi menjadi 4 [21]:

Forget Gate

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + W_{fc}c_{t-1} + b_f) \quad (3)$$

Input Gate

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + W_{ic}c_{t-1} + b_i) \quad (4)$$

Memory Update

$$c_t = f_t \odot c_{t-1} + i_t \odot \phi(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \quad (5)$$

Output Gate

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + W_{oc}c_t + b_o) \quad (6)$$

$$h_t = o_t \odot \phi(c_t) \quad (7)$$

BRI Mobile

Dalam rangka memperbaiki kualitas layanan kepada para pelanggan, Bank BRI meluncurkan sebuah aplikasi berbasis ponsel yang dinamakan *BRI Mobile*. Platform ini menawarkan beragam fungsi praktis bagi nasabah, seperti pengecekan saldo serta pengiriman dana antar rekening bank. Meskipun demikian, sejumlah pengguna yang sudah memakai aplikasi ini masih menyampaikan beberapa keluhan[22].

Confusion Matrix

Confusion matrix Merupakan sebuah matriks yang memperlihatkan kuantitas data uji yang berhasil diprediksi dengan benar beserta data yang salah dikategorikan[23]. Di bawah ini adalah ilustrasi *confusion matrix* yang menggambarkan klasifikasi sebagaimana tercantum pada Tabel 1.

TABEL 1. Confusion Matrix

		Kelas Prediksi	
		1	0
Kelas sebenarnya	11	TP	FN
	00	FP	TN

Keterangan:

TP (*True Positive*) = banyaknya data dari kelas 1 yang berhasil tepat diidentifikasi sebagai kelas 1.

TN (*True Negative*) = jumlah data dari kelas 0 yang secara akurat dikategorikan sebagai kelas 0.

FP (*False Positive*) = kuantitas data kelas 0 yang keliru diberi label sebagai kelas 1.

FN (*False Negative*) = banyak data kelas 1 yang salah diklasifikasikan menjadi kelas 0.

Rumus *accuracy*, *precision*, dan *recall* pada *confusion matrix* seperti berikut[23]:

Accuracy

$$accuracy = \frac{TP+TN}{Total} \quad (8)$$

Precision

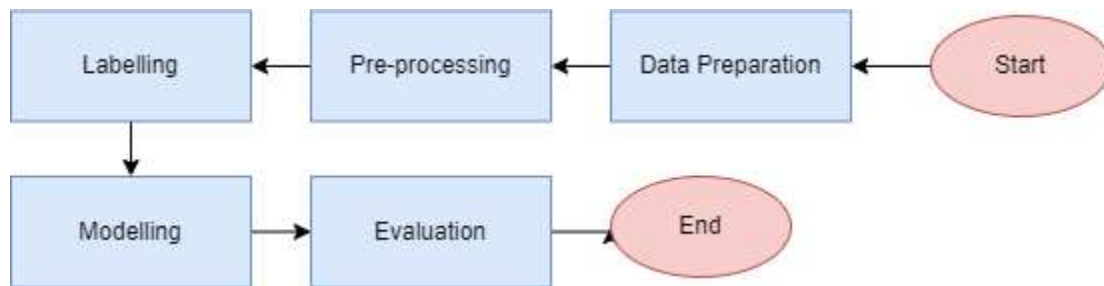
$$Precision = \frac{TP}{TP+FP} \quad (9)$$

Recall

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

METODE PENELITIAN

Dalam penelitian ini, proses dilakukan dengan memanfaatkan pustaka (*library*) yang tersedia di *Google Colaboratory*. Terdapat lima tahapan utama dalam penelitian ini, yaitu: 1) Persiapan Data yang meliputi pengumpulan data secara otomatis (*scraping*), 2) Tahap pra-pemrosesan data, 3) Pemberian label (*labeling*), 4) Pembuatan model (*modelling*), dan 5) Pengujian atau evaluasi. Metodologi yang diaplikasikan dapat dilihat secara visual pada Gambar 1



GAMBAR 1. Metode Penelitian

Data Preparation

Dalam studi ini, pengumpulan data dilakukan dengan teknik *scraping* dari ulasan pengguna aplikasi BRImo yang tersedia di platform *Google Play Store*. Proses pengambilan data menggunakan pustaka *Python* seperti *Google-Play-Scraper*, dengan data yang bersifat publik dan layak digunakan untuk tujuan penelitian. Sumber data berasal dari *Google Play Store* pada alamat <https://play.google.com/store/search?q=brimo&c=apps>. *Scraping* merupakan metode untuk mengekstraksi data atau informasi dari situs web tertentu. Data yang diperoleh bisa berupa teks, link, video, audio, atau dokumen[24]. Parameter *id.co.bri.brimo* adalah nama paket aplikasi BRImo. Total data yang diambil berjumlah 10.000 ulasan dari keseluruhan sekitar 1 juta ulasan. Setelah ulasan berhasil dikumpulkan dan disimpan dalam variabel *result*, tahap berikutnya adalah menampilkan data mentah hasil *scraping* guna melihat contoh isi dari data ulasan yang telah dikumpulkan.

Pre-processing

Pre-processing Merupakan langkah permulaan dalam proses analisis sentimen yang fokus pada penyaringan data, agar informasi yang dipakai terhindar dari komponen-komponen yang tidak berkaitan atau yang bisa menurunkan ketepatan hasil pemodelan. Pembersihan data adalah tahapan dalam menilai mutu data melalui penyesuaian, perubahan, atau penghilangan elemen data yang dianggap kurang relevan atau memiliki bentuk yang tidak sesuai standar[25]. Setelah proses ini, dilakukan tahap penghapusan *stopword*, yaitu kata-kata umum seperti "dan", "atau", atau "di" yang tidak memberikan kontribusi makna signifikan dalam analisis teks. *Stop words* merupakan istilah umum seperti 'dan', 'atau', 'di', yang biasanya tidak menyumbang makna signifikan dalam analisis teks[26]. Selanjutnya, teks diproses melalui tahap *tokenizing*, yaitu pemecahan kalimat menjadi bagian-bagian kecil yang disebut token, biasanya berupa kata, untuk pemecahan sebuah teks menjadi segmen-segmen mini, lazimnya berupa kata-kata, yang dinamakan token. Langkah ini bertujuan untuk mengidentifikasi setiap elemen kata penyusun dalam teks tersebut[27].

Labeling

Informasi yang terkumpul sudah mencakup skor atau nilai rating, yang kemudian menjadi landasan dalam proses pemberian label untuk mengkategorikan ulasan ke dalam tiga kelas sentimen, yakni negatif, netral, dan positif. Pemberian label ini dilakukan secara manual berdasarkan jumlah bintang yang diberikan oleh pengguna, dengan aturan bahwa rating 1 sampai 2 tergolong sentimen negatif, rating 3 dianggap netral, dan rating 4 sampai 5 masuk dalam sentimen positif. Tahapan ini dijalankan menggunakan mekanisme pengulangan, dan demi memudahkan

analisis, setiap label diubah menjadi kode numerik: negatif = 2, netral = 1, dan positif = 0. Penentuan label tersebut kemudian direview kembali untuk memastikan keakuratan klasifikasi sekaligus meningkatkan mutu data yang akan diolah.

Modelling

Di fase ini, proses pemodelan serta pengelompokan sentimen dijalankan menggunakan dua pendekatan *deep learning*, yakni *Bidirectional Encoder Representations from Transformers* (BERT) dan *Recurrent Neural Network Long Short-Term Memory* (RNN-LSTM). Kedua teknik tersebut dipilih karena keunggulannya dalam menangkap konteks serta pola bahasa alami, khususnya pada teks berbahasa Indonesia.

BERT adalah sebuah model berbasis transformer yang dipelajari secara dua arah (*bidirectional*), sehingga mampu menangkap hubungan antar kata dari kedua sisi kalimat secara bersamaan. Pada studi ini, digunakan versi *pre-trained* BERT Multilingual yang sudah dilatih pada berbagai bahasa, termasuk Bahasa Indonesia. Selanjutnya, model ini diadaptasi (*fine-tuned*) dengan dataset ulasan aplikasi BRImo untuk mengklasifikasikan sentimen ke dalam tiga kategori: positif, negatif, dan netral. Proses *fine-tuning* meliputi tahap tokenisasi memakai tokenizer BERT, pengkodean teks, serta pelatihan model dengan pengaturan parameter yang telah dioptimasi[28].

Di sisi lain, RNN-LSTM adalah sebuah lapisan dalam jaringan saraf berulang (RNN) yang dirancang khusus untuk mengatasi tantangan *vanishing gradient* saat memproses data berurutan. LSTM mengandung struktur memori internal yang memungkinkan model menyimpan informasi krusial dari rangkaian kata dalam sebuah kalimat. Dalam penelitian ini, model RNN-LSTM digunakan sebagai pembanding model BERT dalam mengklasifikasikan sentimen. Arsitektur RNN-LSTM terdiri dari beberapa bagian utama. Pertama, teks yang sudah dibersihkan dan ditokenisasi dimasukkan ke dalam *Embedding Layer* untuk mengubah setiap kata menjadi vektor angka. Kemudian, vektor tersebut diproses oleh satu LSTM Layer yang berfungsi untuk memahami hubungan antar kata dalam kalimat. Untuk mencegah *overfitting*, digunakan *Dropout Layer*, lalu dilanjutkan ke *Dense Layer* dengan fungsi untuk menghasilkan klasifikasi sentimen ke dalam tiga kategori: positif, netral, dan negatif. Pelatihan model dilakukan selama 10 *epoch* menggunakan *loss function categorical cross-entropy* dan algoritma optimisasi Adam. Selanjutnya, model dilatih untuk mengidentifikasi pola sentimen yang terkandung dalam ulasan pengguna. Kedua model kemudian diuji dan dibandingkan performanya dalam klasifikasi menggunakan dataset yang sama, dengan atau tanpa tambahan pra-pemrosesan (mengandalkan tokenisasi bawaan model). [8].

Evaluation

Tahapan penilaian ini dirancang untuk mengukur kinerja model klasifikasi sentimen yang telah dibangun memakai BERT dan RNN-LSTM. Proses pengujian dilakukan dengan membandingkan prediksi model terhadap label asli pada dataset pengujian. Metode evaluasi yang dipakai meliputi confusion matrix dan F1-Score, yang umum digunakan untuk tugas klasifikasi dengan banyak kelas (multi-class classification).

Matriks kebingungan berfungsi untuk menampilkan kuantitas prediksi yang tepat serta keliru pada tiap kategori (positif, negatif, dan netral). Tabel ini membantu mempermudah evaluasi pola kesalahan model saat mengelompokkan data. Melalui matriks kebingungan, bisa dihitung metrik seperti presisi, sensitivitas (recall), dan skor F1 untuk setiap kategori secara terpisah.

Matriks kebingungan digunakan untuk memperlihatkan jumlah prediksi yang benar dan salah pada masing-masing kelas (positif, negatif, dan netral). Tabel ini memudahkan proses analisis kesalahan model dalam mengklasifikasikan data. Dengan memanfaatkan matriks kebingungan, dapat dihitung metrik-metrik seperti presisi, sensitivitas (recall), dan skor F1 untuk setiap kelas secara individual[23].

Penilaian dilakukan pada dua skenario berbeda: satu menggunakan model dengan langkah preprocessing tambahan, dan satu lagi menggunakan model tanpa preprocessing, hanya mengandalkan tokenisasi standar. Perbandingan hasil dari kedua metode ini digunakan untuk mengukur seberapa besar pengaruh preprocessing terhadap akurasi klasifikasi sentimen yang dijalankan oleh masing-masing model. Confusion matrix disajikan dalam tabel 2.

TABEL 2. Confusion matrix

		Kelas Prediksi	
		1	0
Kelas sebenarnya	11	TP	FN
	00	FP	TN

HASIL DAN PEMBAHASAN

Melalui metode pengumpulan data menggunakan random sampling, sebanyak 10.000 ulasan berhasil dikumpulkan dari total sekitar satu juta ulasan yang tersedia. Setelah proses pengambilan data selesai dan disimpan dalam variabel `result`, langkah berikutnya adalah menampilkan data mentah untuk mengamati contoh isi ulasan yang diperoleh. Visualisasi data tersebut disajikan dalam bentuk gambar guna memberikan gambaran awal mengenai struktur dan isi ulasan hasil pengumpulan data. Gambar 2 adalah data hasil scraping yang ditampilkan.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	aj
0	24868505-ec59-4b9f-9814-652642d8cc5b	Pengguna Google	lh.googleusercontent.com/EGemoI2N...	Hasil Update hanya mengutamakan tampilan saja ..	1	7	2.1.0	2021-01-13 22:05:02	Hai, Sobat BRI. Mohon maaf atas kendala yang d...	2025-01-12 11:10:06	
1	c600cd41-bc11-4593-b631-d11b18ddebdb8	Pengguna Google	lh.googleusercontent.com/EGemoI2N...	bikin username aja ribetnya minta ampun. sudah...	1	118	2.84.0	2025-06-07 00:03:21	Hai Sobat BRI, mohon maaf atas ketidaknyamanan...	2025-06-07 07:37:26	
2	6d8cfbcc-ff04-4393-9513-3d24e0b0467a	Pengguna Google	lh.googleusercontent.com/EGemoI2N...	Sudah digunakan 7 bulan di device yang sama, t...	1	128	2.84.0	2025-06-05 01:25:11	Hai Sobat BRI, mohon maaf atas ketidaknyamanan...	2025-06-05 03:32:11	
3	05fdaa08-7eb7-400e-be14-bfd8d5e4b20a	Pengguna Google	lh.googleusercontent.com/EGemoI2N...	saya sudah terdaftar di apk nrimo.tpi pas lori	1	0	2.84.0	2025-06-15 17:27:16	Hai Sobat BRI, mohon maaf atas ketidaknyamanan...	2025-06-15 17:11:55	

GAMBAR 2. Data hasil scraping

Gambar 2 menunjukkan dataset ulasan pengguna aplikasi BRImo yang diambil melalui *Google Play Store* menggunakan teknik web *scraping*. Setiap baris mewakili satu ulasan, berisi informasi seperti ID ulasan (*reviewId*), nama pengguna, isi ulasan (*content*), skor rating (*score*), jumlah likes (*thumbsUpCount*), versi aplikasi saat ulasan dibuat, serta tanggapan dari pihak BRI (*replyContent*) beserta waktu balasannya. Data ini menjadi dasar dalam analisis sentimen yang dilakukan pada penelitian. Dalam proses pra-pemrosesan teks, yang mencakup pembersihan data, penghilangan kata-kata umum (*stopword*), serta tokenisasi, dengan hasil yang tersaji pada tabel 3.

TABEL 3. Pre-processing

Pre-Processing	Contoh data
Data asli	<code>https://play-lh.googleusercontent.com/EGemoI2N... Hasil Update hanya mengutamakan tampilan saja ..</code>
Data <i>cleaning</i>	<code>hasil update hanya mengutamakan tampilan saja ..</code>
<i>Stopword</i>	<code>hasil update mengutamakan tampilan ..</code>
<i>Tokenize</i>	<code>['hasil', 'update', 'mengutamakan', 'tampilan' ..]</code>

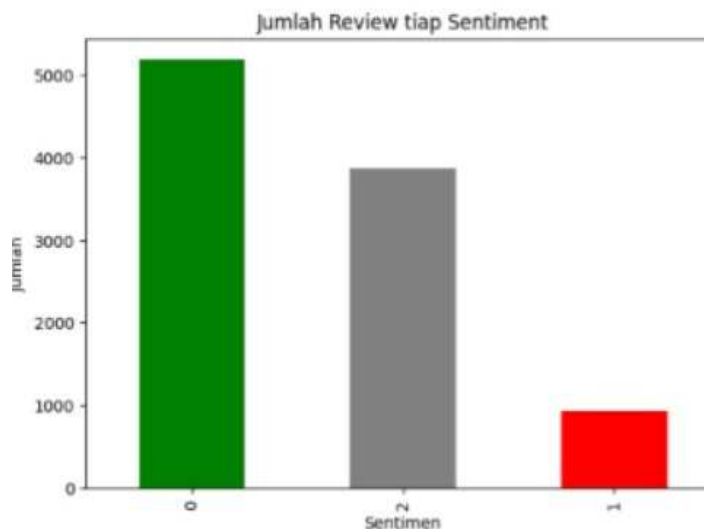
Setelah tahap pengumpulan data melalui teknik scraping selesai, langkah berikutnya adalah pemberian label, yakni proses mengidentifikasi sentimen pada setiap ulasan. Dalam penelitian ini, ulasan-ulasan tersebut dikelompokkan ke dalam tiga jenis sentimen, yaitu positif, negatif, dan netral. Gambar 3 memperlihatkan hasil dari proses pelabelan tersebut.

Gambar 3 menunjukkan bagian dari dataset yang telah melalui tahap pelabelan sentimen. Memuat isi ulasan pengguna aplikasi BRImo beserta label sentimen yang sudah dikategorikan ke dalam tiga kelas, yaitu positif, negatif, dan netral. Label ini diberikan berdasarkan konteks dari isi ulasan, yang ditunjukkan oleh angka skor (1 untuk positif, 0 untuk netral, dan -1 untuk negatif) serta label teks yang sesuai. Data berlabel ini digunakan sebagai input utama dalam proses pelatihan model BERT dan RNN-LSTM untuk melakukan klasifikasi sentimen secara otomatis.

	content	sentiment	label
9272	aplikasinya bermanfaat sekali namun selalu te...	2	Positif
5867	Mantap, gampang gunainnya, dan loginnya sangat...	2	Positif
7089	setiap tahun pasti ada pembaharuan aplikasi.se...	2	Positif
4906	aplikasinya sering error, padahal sudah berhasi...	0	Negatif
7038	Mohon untuk ditingkatkan. Secara UI tampilan s...	2	Positif
6038	BRImo dimohon untuk dibuatkan password ke 2 ya...	0	Negatif
7246	HATI-HATI, BISA TERDEBET 2 KALI pengguna aplik...	0	Negatif
7845	Aplikasi belum bener udah di rilis, niatnya ba...	0	Negatif
9740	Setiap selesai menggunakan aplikasi ini harus ...	0	Negatif
6506	Kenapa Setiap Kali Sy akan Log in,selalu Tidak...	0	Negatif
5932	mohon lbh dimaksimalkan lagi app nya. aplikasi...	0	Negatif
9363	aplikasinya sering error, baru siap kita downl...	0	Negatif
6645	aplikasinya bagus.. tapi bisa dipake sekali aj...	1	Netral
7549	Sangat membantu dalam aktivitas, akan tetapi d...	2	Positif

GAMBAR 3. Labeling

Sedangkan pada gambar 4 menampilkan diagram batang yang menunjukkan distribusi jumlah ulasan pengguna aplikasi BRImo berdasarkan sentimen. Terdapat tiga kategori sentimen yang ditampilkan: negatif (-1) ditandai dengan warna hijau, positif (1) dengan warna abu-abu, dan netral (0) dengan warna merah. Dari grafik terlihat bahwa jumlah ulasan dengan sentimen negatif merupakan yang terbanyak, disusul oleh sentimen positif, sementara sentimen netral jumlahnya paling sedikit. Distribusi yang tidak seimbang ini penting untuk diperhatikan karena dapat memengaruhi performa model klasifikasi, terutama dalam mendeteksi sentimen dengan jumlah data yang lebih sedikit seperti netral.



GAMBAR 4. Hasil labeling

Setelah seluruh proses pra-pemrosesan rampung, tahap berikutnya adalah penerapan model BERT (Bidirectional Encoder Representations from Transformers). Awalnya, data dibagi menjadi dua bagian, yakni 80% untuk pelatihan dan 20% untuk pengujian. Format input untuk model BERT disiapkan dengan memanfaatkan fungsi `encode_plus` dari tokenizer BERT, yang berfungsi mengolah teks sebelum masuk ke dalam model. Tokenisasi memakai tokenizer dari model pra-latih `IndoBERT-base-p2`. Pada fase ini, setiap kalimat diberi token khusus, yakni `[CLS]` di awal dan `[SEP]` di akhir, sebagai penanda struktur input. Jika kalimat melebihi panjang maksimal (`max_length`), maka akan dipangkas, sedangkan kalimat yang lebih pendek akan diisi padding menggunakan token `[PAD]` agar panjang input seragam.

Arsitektur RNN-LSTM dibangun dengan metode berurutan yang terdiri dari beberapa lapisan utama. Proses dimulai pada lapisan Embedding yang berfungsi mengonversi setiap token kata menjadi representasi vektor berdimensi 128, dengan cakupan kosakata terbatas pada 10.000 kata dan panjang input maksimal 100 token. Selanjutnya, model ini mengintegrasikan dua lapisan LSTM bertingkat, masing-masing berisi 128 dan 64 unit, guna menangkap pola ketergantungan dalam data berurutan. Untuk mengurangi risiko overfitting, diterapkan lapisan Dropout dengan tingkat 0,5 setelah setiap lapisan LSTM. Pada fase akhir, dipakai lapisan Dense yang memiliki tiga unit dengan fungsi aktivasi softmax untuk memproduksi prediksi klasifikasi sentimen dalam tiga kategori: positif, negatif, dan netral. Model ini disusun menggunakan fungsi loss categorical crossentropy, dioptimalkan dengan algoritma Adam, serta dievaluasi menggunakan metrik akurasi. Perbandingan hasil akurasi dari kedua model dapat dilihat pada tabel 4

TABEL 4. Perbandingan akurasi

	Bert	RNN-LSTM
Akurasi	54%	53%

Setelah mendapatkan tingkat akurasi awal sebesar 54% untuk BERT dan 53% untuk RNN-LSTM, dilakukan eksperimen lanjutan dengan menghilangkan tahap tokenisasi dalam proses pra-pemrosesan. Keputusan ini diambil karena kedua model, BERT dan RNN-LSTM, sudah memiliki sistem tokenisasi bawaan yang secara otomatis mengelola input teks. Dengan menghindari proses tokenisasi berulang, diharapkan data masukan ke dalam model menjadi lebih efisien dan sesuai dengan format yang dibutuhkan oleh masing-masing arsitektur. Hal ini bertujuan untuk mengurangi risiko degradasi kualitas input akibat pemrosesan yang tumpang tindih. Tabel 5 adalah perbandingan akurasi yang dirangkum.

TABEL 5. Perbandingan akurasi dengan tambahan tokenisasi dan tidak menggunakan FL-score

	BERT	RNN-LSTM
AKURASI sebelum	54%	53%
AKURASI sesudah	73%	69%

Hasil pengujian model RNN-LSTM menunjukkan bahwa model ini mampu mengklasifikasikan sentimen dengan akurasi sebesar 69%. Performa terbaik dicapai pada kelas negatif dan positif, namun model kesulitan dalam mengenali sentimen netral dengan baik. Jika dibandingkan dengan model BERT yang mencapai akurasi 73%, RNN-LSTM menunjukkan performa yang lebih rendah.

KESIMPULAN

Dari hasil penelitian yang dilakukan, dapat disarikan bahwa proses pengumpulan data berhasil mengakumulasi sebanyak 10.000 ulasan aplikasi BRImo melalui metode random sampling dari total sekitar satu juta ulasan yang tersedia di Google Play. Data tersebut kemudian dianalisis secara visual guna memberikan gambaran awal mengenai struktur teks ulasan. Pada tahap awal, dilakukan preprocessing dengan membersihkan teks dan menerapkan tokenisasi eksternal. Hasil akurasi awal memperlihatkan model BERT mencapai 54%, sementara RNN-LSTM memperoleh 53%. Namun, setelah percobaan lanjutan tanpa proses tokenisasi ganda dalam preprocessing, akurasi meningkat secara signifikan. Model BERT mencatatkan akurasi hingga 73%, sedangkan RNN-LSTM naik menjadi 69%. Hal ini mengindikasikan bahwa penerapan tokenisasi berlapis justru dapat menurunkan kualitas input dan berdampak negatif pada performa model secara keseluruhan. Berdasarkan perbandingan tersebut, dapat ditarik kesimpulan bahwa model BERT memiliki keunggulan dalam menangkap konteks bahasa Indonesia, terutama setelah dilakukan penyesuaian.

Penelitian lanjutan dianjurkan untuk mengupayakan penyetelan hiperparameter secara lebih rinci, mencakup penyesuaian learning rate, ukuran batch, jumlah epoch, serta penerapan strategi warm-up guna meningkatkan efisiensi model. Selain itu, evaluasi dengan metrik tambahan seperti AUC dan Macro-F1 Score sangat direkomendasikan agar kinerja model terhadap data yang tidak seimbang dapat terukur dengan lebih akurat.

DAFTAR PUSTAKA

- [1] M. K. Insan, U. Hayati, and O. Nurdian, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di," *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 478–483, 2023.
- [2] E. Daniati and H. Utama, "Analisis Sentimen Dengan Pendekatan Ensemble Learning Dan Word Embedding Pada Twitter," *J. Inf. Syst. Manag.*, vol. 4, no. 2, pp. 125–131, 2023, doi: 10.24076/joism.2023v4i2.973.
- [3] E. Daniati and H. Utama, "TOPSIS in Decision-Making Framework Based on Twitter Sentiment Analysis," *ICOIACT 2021 - 4th Int. Conf. Inf. Commun. Technol. Role AI Heal. Soc. Revolut. Turbul. Era*, pp. 268–273, 2021, doi: 10.1109/ICOIACT53268.2021.9564015.
- [4] M. Z. Rahman, Y. A. Sari, and N. Yudistira, "Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5120–5127, 2021.
- [5] R. W. Putri, A. Ristyawan, and M. N. Muzaki, "Comparison Performance of K-NN and NBC Algorithm for Classification of Heart Disease," *JTECS J. Sist. Telekomun. Elektron. Sist. Kontrol Power Sist. dan Komput.*, vol. 2, no. 2, p. 143, 2022, doi: 10.32503/jtecs.v2i2.2708.
- [6] E. Daniati, *The Relevance of Andersen 's Children 's Stories to the Growth and Development of Indonesian Children*, no. Ichss 2024. Atlantis Press SARL, 2025. doi: 10.2991/978-2-38476-352-8.
- [7] N. Husin, "Komparasi Algoritma Random Forest, Naïve Bayes, dan Bert Untuk Multi-Class Classification Pada Artikel Cable News Network (CNN)," *J. Esensi Infokom J. Esensi Sist. Inf. dan Sist. Komput.*, vol. 7, no. 1, pp. 75–84, 2023, doi: 10.55886/infokom.v7i1.608.
- [8] F. I. Komputer, U. Dian, and N. Semarang, "Metode Recurrent Neural Network (Rnn) Dengan Arsitektur Lstm Untuk Analisis Sentimen Opini Publik Terkait Vaksin Covid-19," *J. Inform. Upgris*, vol. 8, no. 1, pp. 44–48, 2022.
- [9] S. Lestari and S. Saepudin, "Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes," *Semin. Nas. Sist. Inf. dan Manaj. Inform.*, pp. 163–170, 2021.
- [10] A. S. P. Braja and A. Kodar, "Implementasi Fine-Tuning BERT untuk Analisis Sentimen terhadap Review Aplikasi PUBG Mobile di Google Play Store," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 7, no. 3, p. 120, 2023, doi: 10.51213/jimp.v7i3.779.
- [11] R. Parlika, S. Ilham Pradika, A. Muhammad Hakim, and K. Rachman N.M, "Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python TextBlob," *J. Ilm. Teknol. Inf. dan Robot.*, vol. 2, no. 2, pp. 33–37, 2020, doi: 10.33005/jifti.v2i2.22.
- [12] N. Putu, V. D. Saraswati, N. Yudistira, and P. P. Adikara, "Analisis Sentimen terhadap Perundungan Siber pada Twitter menggunakan Algoritma Bidirectional Encoder Representations from Transformer (BERT)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 2, pp. 909–916, 2023.
- [13] A. A. Mudding, "Mengungkap Opini Publik: Pendekatan BERT-based-caused untuk Analisis Sentimen pada Komentar Film," *J. Syst. Comput. Eng.*, vol. 5, no. 1, pp. 36–43, 2024, doi: 10.61628/jsce.v5i1.1060.
- [14] E. Daniati, A. Prasetya, W. Sakti, G. Irianto, and A. Ghosh, "Analyzing event relationships in Andersen ' s Fairy Tales with BERT and Graph Convolutional Network (GCN)," vol. 5, no. 1, pp. 40–59, 2024.
- [15] D. F. Sjoraida, B. W. K. Guna, and D. Yudhakusuma, "Analisis Sentimen Film Dirty Vote Menggunakan BERT (Bidirectional Encoder Representations from Transformers)," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 8, no. 2, pp. 393–404, 2024, doi: 10.35870/jtik.v8i2.1580.
- [16] R. Cahyadi *et al.*, "Recurrent Neural Network (Rnn) Dengan Long Short Term Memory (Lstm) Untuk Analisis Sentimen Data Instagram," *J. Inform. dan Komput.*, vol. 5, no. 1, pp. 1–9, 2020.
- [17] D. Tarkus, S. R. U. A. Sompie, and A. Jacobus, "Implementasi Metode Recurrent Neural Network pada Pengklasifikasian Kualitas Telur Puyuh," *J. Tek. Inform.*, vol. 15, no. 2, pp. 137–144, 2020.
- [18] S. Saadah, P. Eko Yunanto, and C. Author, "Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) Methods to Forecast Daily Turnover at BM Motor Ngawi," *Indones. J. Artif. Intell. Data Min.*, vol. 7, no. 1, pp. 141–147, 2024.
- [19] K. S. Witanto, N. A. Sanjaya ER, A. E. Karyawati, I. G. A. G. A. Kadyanan, I. K. G. Suhartana, and L. G. Astuti, "Implementasi LSTM Pada Analisis Sentimen Review Film Menggunakan Adam Dan RMSprop Optimizer," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 10, no. 4, p. 351, 2022, doi: 10.24843/jlk.2022.v10.i04.p05.
- [20] R. Akbar, R. Santoso, and B. Warsito, "Prediksi Tingkat Temperatur Kota Semarang Menggunakan Metode Long Short-Term Memory (Lstm)," *J. Gaussian*, vol. 11, no. 4, pp. 572–579, 2023, doi: 10.14710/j.gauss.11.4.572-579.
- [21] M. Rizki, S. Basuki, and Y. Azhar, "Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory(LSTM) Untuk Prediksi Curah Hujan Kota Malang," *J. Repos.*, vol. 2, no. 3, pp. 331–338, 2020, doi: 10.22219/repositor.v2i3.470.
- [22] F. A. Hariansyah, N. H. Wardani, and A. D. Herlambang, "Analisis Pengaruh Kualitas Layanan Mobile Banking Terhadap Kepuasan dan Loyalitas Nasabah Pada Pengguna Layanan BRI Mobile Bank Rakyat Indonesia di Kantor Cabang Cirebon," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 5, p. 9, 2020.
- [23] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [24] S. Kusumo, "Penerapan Web Scraping Deskripsi Produk Menggunakan Selenium Python Dan Framework Laravel,"

-
- JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 4, pp. 3426–3435, 2022, doi: 10.35957/jatisi.v9i4.2727.
- [25] A. R. Raharja, Jayadi, A. Pramudianto, and Y. Muchsam, “Penerapan Algoritma Decision Tree dalam Klasifikasi Data ‘Framingham’ Untuk Menunjukkan Risiko Seseorang Terkena Penyakit Jantung dalam 10 Tahun Mendatang,” *Technol. J.*, vol. 1, no. 1, 2024, doi: 10.62872/cwgzp962.
- [26] H. Utama, E. Daniati, and A. Masruro, “Weak Supervision Dengan Pendekatan Labeling Function Untuk Analisis Sentimen Pada Twitter,” *Indones. J. Comput. Sci. Res.*, vol. 3, no. 1, pp. 49–57, 2024, doi: 10.59095/ijcsr.v3i1.93.
- [27] T. N. Wijaya, R. Indriati, and M. N. Muzaki, “Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter,” *Jambura J. Electr. Electron. Eng.*, vol. 3, no. 2, pp. 78–83, 2021, doi: 10.37905/jjee.v3i2.10885.
- [28] Z. W. Farida and N. Rochmawati, “Analisis Sentimen Masyarakat terhadap Fenomena Childfree Menggunakan Metode Long Short Term Memory dan Bidirectional Encoder Representations from Transformers di Twitter,” *J. Informatics Comput. Sci.*, vol. 5, no. 03, pp. 369–376, 2024.