FINE TUNING INDOBERT UNTUK ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI TIKET.COM DI GOOGLE PLAY STORE

Delvin Nuryadi, Farindika Metandi, Noor Alam Hadiwijaya,
M. Zainul Rohman, Subhan Hartanto, Agusdi Syafrizal, Erry Yadie
Teknologi Informasi, Politeknik Negeri Samarinda
Jl. Cipto Mangunkusumo, Samarinda Seberang, Kalimantan Timur, Indonesia
delvinnr19@gmail.com

ABSTRAK

Tiket.com merupakan salah satu platform pemesanan tiket perjalanan terbesar di Indonesia. Ulasan pengguna di Google Play Store dapat memberikan wawasan penting mengenai kepuasan pengguna terhadap layanan aplikasi ini. Namun, analisis manual terhadap ribuan ulasan tidak efisien, sehingga diperlukan pendekatan berbasis machine learning untuk mengklasifikasikan sentimen secara otomatis. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna Tiket.com dengan menggunakan model IndoBERT yang telah di fine-tune. Metode yang digunakan meliputi pengumpulan data ulasan melalui web scraping, preprocessing data, serta fine-tuning model IndoBERT menggunakan dataset SmSA dari IndoNLU. Evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score untuk mengukur performa klasifikasi sentimen positif, netral, dan negatif. Hasil menunjukkan bahwa model memiliki akurasi keseluruhan sebesar 91%, dengan F1-score 93% untuk sentimen positif, 94% untuk sentimen negatif, dan 77% untuk sentimen netral, meskipun recall pada kelas netral masih rendah (65%). Dari 10.000 ulasan, mayoritas bersentimen positif (5.271), diikuti negatif (4.022), dan netral (707). Penelitian ini membuktikan efektivitas IndoBERT dalam analisis sentimen berbahasa Indonesia dan dapat digunakan sebagai referensi dalam memahami opini pengguna serta meningkatkan layanan aplikasi.

Kata kunci : *analisis sentimen, indobert, fine-tuning, ulasan pengguna, google play store*

1. PENDAHULUAN

Industri perjalanan dan pariwisata di Indonesia telah mengalami perkembangan pesat seiring dengan kemajuan teknologi digital. Salah satu inovasi signifikan adalah kehadiran platform daring yang memudahkan akses pengguna terhadap layanan perjalanan, seperti Tiket.com. Sebagai salah satu agen perjalanan daring terbesar di Indonesia, Tiket.com menawarkan layanan pemesanan tiket penerbangan, kereta, akomodasi, serta tiket acara dan penyewaan kendaraan. Kehadiran aplikasi ini memberikan kemudahan bagi pengguna untuk melakukan berbagai transaksi langsung melalui perangkat mereka.

Untuk memberikan layanan yang lebih baik, Tiket.com menghadirkan fitur-fitur unggulan seperti pemberitahuan harga, penawaran eksklusif, dan antarmuka pengguna yang intuitif. Hal ini memungkinkan aplikasi untuk menjadi pilihan utama bagi banyak pengguna. Berdasarkan laporan YouGov pada kuartal pertama tahun 2024, Tiket.com menduduki posisi kedua sebagai agen perjalanan daring paling populer di Indonesia, dengan tingkat pengenalan mencapai 80%, setelah Traveloka (88%) . Popularitas ini juga terlihat dari volume ulasan pengguna yang terus meningkat, mencerminkan pengalaman mereka terhadap layanan yang disediakan.

Analisis sentimen terhadap ulasan pengguna menjadi semakin penting, mengingat ulasan tersebut dapat memberikan wawasan berharga bagi pengelola platform untuk memahami kebutuhan dan ekspektasi pengguna. Berbagai metode pembelajaran mesin telah digunakan untuk analisis sentimen, seperti Naive Bayes, Support Vector Machine (SVM), dan model berbasis deep learning seperti LSTM dan BERT [1]. BERT, khususnya IndoBERT, telah menunjukkan performa unggul dalam berbagai tugas Natural Language Processing (NLP), termasuk analisis sentimen dalam bahasa Indonesia [2].

Penelitian ini bertujuan untuk mengevaluasi performa model IndoBERT yang telah di-fine-tune pada dataset SmSA dari IndoNLU untuk menganalisis sentimen ulasan pengguna Tiket.com di Google Play Store. Dengan menggunakan pendekatan ini, diharapkan dapat diperoleh model yang mampu mengklasifikasikan sentimen positif, negatif, atau netral dengan tingkat akurasi tinggi, serta memberikan wawasan mengenai persepsi umum pengguna terhadap layanan Tiket.com.

2. TINJAUAN PUSTAKA

2.1. Penelitian terdahulu

Penelitian sebelumnya menganalisis sentimen ulasan aplikasi DLU Ferry di Google Play Store menggunakan model IndoBERT-base yang di fine-tuning. Dengan 1575 ulasan berisi sentimen positif, netral, dan negatif, serta hyperparameter batch size 32, learning rate 3e-6, dan epoch 5, penelitian ini mencapai akurasi 86%. Hasil ini menunjukkan efektivitas IndoBERT-base dalam klasifikasi sentimen aplikasi [2].

Analisis sentimen pada kendaraan listrik menggunakan IndoBERT. Penelitian ini membandingkan model yang dilatih dengan dan tanpa

data IndoNLU, dan menemukan bahwa model yang dilatih dengan data IndoNLU lebih akurat dan konsisten. Temuan ini memberikan wawasan baru mengenai pandangan masyarakat Indonesia terhadap kendaraan listrik, serta potensi untuk mendukung kebijakan dan inovasi teknologi ramah lingkungan [3].

2.2. Analisis sentimen

Analisis sentimen atau opinion mining adalah studi yang menganalisis opini, sikap, dan emosi manusia dari teks. Bidang ini sangat aktif dalam Natural Language Processing dan juga relevan dalam data mining, web mining, dan text mining. Penelitian ini kini meluas ke ilmu manajemen dan sosial, berkat peranannya yang penting bagi bisnis dan masyarakat, terutama seiring dengan berkembangnya media sosial seperti ulasan, forum, blog, dan jejaring sosial [4].

2.3. Machine Learning

Machine learning adalah paradigma komputasi di mana kemampuan untuk menyelesaikan masalah dibangun berdasarkan contoh-contoh sebelumnya. Ide dasarnya adalah case-based reasoning, yaitu menyelesaikan masalah dengan melihat kasus-kasus serupa yang telah terjadi sebelumnya. Contoh-contoh sebelumnya ini disebut contoh pelatihan, dan proses mempelajarinya disebut pembelajaran. Setelah pembelajaran, kemampuan untuk menyelesaikan masalah nyata disebut generalisasi [5].

2.4. Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah cabang dari artificial intelligence yang penuh dengan tugas-tugas yang rumit, canggih, dan menantang terkait bahasa, seperti machine translation, question answering, summarization, dan lain sebagainya. NLP melibatkan desain dan implementasi model, sistem, dan algoritma untuk menyelesaikan masalah praktis dalam memahami bahasa manusia [6].

2.5. Transformer

Transformer adalah arsitektur model neural yang telah menjadi dominan dalam Natural Language Processing (NLP). Keunggulannya terletak pada kemampuannya untuk diskalakan dengan data pelatihan dan ukuran model, serta kemampuannya untuk melatih secara paralel dengan efisien dan menangkap fitur urutan panjang. Transformer juga telah mengungguli model neural lain seperti jaringan neural konvolusional dan rekuren dalam tugas pemahaman dan pembuatan bahasa alami. Dengan kemajuan dalam arsitektur model dan teknik pra-pelatihan, Transformer memungkinkan pembangunan model dengan kapasitas lebih besar dan pemanfaatan kapasitas ini untuk berbagai tugas [7].

2.6. BERT

Bert merupakan singkatan dari Bidirectional Encoder Representations from Transformers. Berbeda dengan model representasi bahasa terbaru, BERT dirancang untuk melakukan pretrain representasi bidirectional yang dalam dari teks yang tidak berlabel dengan mengondisikan konteks kiri dan kanan secara bersama-sama di semua lapisan [8].

2.7. IndoBERT

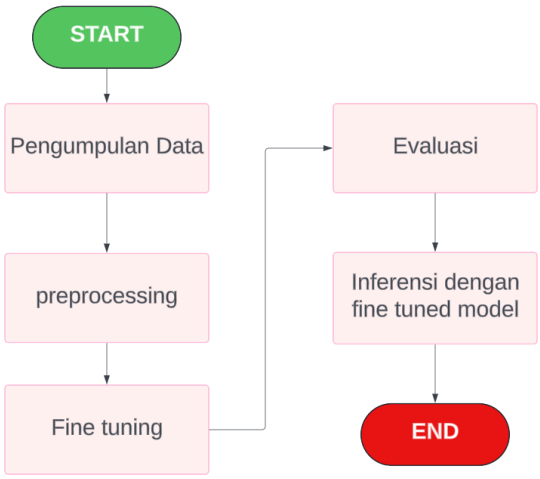
IndoBERT merupakan model BERT versi Indonesia, IndoBERT dibangun menggunakan sejumlah besar kosakata bahasa Indonesia dengan jumlah lebih dari 220 juta kata yang diperoleh dari sumber bahasa yang menggunakan bahasa Indonesia yang baik dan benar seperti dari koran online dan juga dari korpus web Indonesia dan sumber lainnya. Model IndoBERT dilatih dengan teknik transfer learning pada sebuah korpus teks besar dalam bahasa Indonesia, sehingga mampu mempelajari pola-pola bahasa Indonesia pada tingkat yang sangat tinggi [9].

2.8. Fine-Tuning

Fine-tuning mengacu pada proses mengambil model neural network yang telah dilatih sebelumnya dan menyesuaikannya untuk tujuan spesifik tugas baru. Idenya adalah memanfaatkan pengetahuan dan fitur yang dipelajari oleh model yang telah dilatih sebelumnya dan menyesuaikannya agar lebih sesuai dengan persyaratan tugas baru. Fine-tuning dapat menghemat waktu dan sumber daya yang signifikan dibandingkan dengan melatih model dari awal [10].

3. METODE PENELITIAN

Pada penelitian ini, proses analisis sentimen dilakukan melalui beberapa tahapan utama, mulai dari pengumpulan data hingga inferensi model. Berikut adalah tahapan dalam metode penelitian yang dilakukan.



Gambar 1. Alur proses fine tuning model

3.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dari dua sumber utama:

3.2. Dataset Latih (SmSA IndoNLU)

Dataset SmSA IndoNLU digunakan sebagai data latih untuk proses fine-tuning model IndoBERT. Dataset ini telah berisi label sentimen yang dikategorikan menjadi positif, netral, dan negatif, sehingga tidak memerlukan preprocessing tambahan. Data ini langsung dibagi menjadi data latih dan data uji dengan teknik train-test split untuk memastikan model dapat belajar dengan baik dan diuji secara objektif.

3.3. Data Ulasan Tiket.com (Scraping)

Selain dataset latih, data juga dikumpulkan melalui proses web scraping menggunakan pustaka google-play-scraper. Data yang dikumpulkan sebanyak 10.000 ulasan terbaru dari aplikasi Tiket.com hingga 24 Agustus 2024. Informasi yang diperoleh dari proses scraping meliputi, teks ulasan pengguna, versi aplikasi ,tanggal ulasan, dll. Data ini kemudian akan diproses lebih lanjut sebelum digunakan dalam tahap inferensi model.

3.4. Preprocessing

Proses preprocessing dilakukan untuk memastikan data siap digunakan dalam analisis sentimen.

3.5. Preprocessing dataset IndoNLU

Dataset SmSA IndoNLU tidak memerlukan preprocessing tambahan karena sudah dalam format bersih dan berlabel. Satu-satunya langkah yang dilakukan adalah pembagian dataset menggunakan train-test split untuk membagi data latih dan data uji.

3.6. Preprocessing untuk Data Ulasan Tiket.com

Data hasil scraping memerlukan preprocessing lebih lanjut untuk membersihkan teks sebelum digunakan dalam inferensi model. Langkah-langkah preprocessing meliputi, seleksi kolom,case folding yaitu mengubah semua huruf menjadi kecil agar seragam, penghapusan karakter non-ASCII yaitu menghilangkan simbol, emoji, dan karakter spesial yang tidak diperlukan. whitespace cleaning yaitu menghapus spasi berlebih dalam teks, penghapusan karakter berulang yaitu mengurangi redundansi dalam teks agar lebih terstruktur dan siap dianalisis.

3.7. Fine Tuning

Proses fine-tuning dilakukan pada model IndoBERT menggunakan dataset SmSA. Model ini disesuaikan dengan data melalui langkah tokenisasi teks menggunakan tokenizer dari pustaka transformers dan pengaturan hyperparameter seperti ukuran batch, learning rate (2e-5), serta jumlah epoch (3). Teknik fine-tuning memungkinkan model untuk menangkap pola bahasa spesifik dalam dataset dan menghasilkan klasifikasi sentimen yang lebih akurat [10].

3.8. Evaluasi

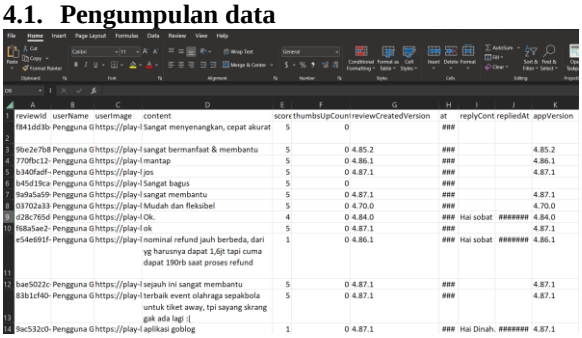
Kinerja model dievaluasi menggunakan data uji dari dataset SmSA. Metrik evaluasi meliputi akurasi, presisi, recall, dan F1-score untuk mengukur kemampuan model dalam mengklasifikasikan sentimen positif, negatif, atau netral [11].

3.9. Inferensi Model

Setelah melalui tahap evaluasi, model diterapkan pada data ulasan Tiket.com yang telah dikumpulkan. Hasil klasifikasi model berupa label sentimen positif, negatif, atau netral dianalisis untuk memberikan wawasan mengenai persepsi pengguna terhadap layanan aplikasi Tiket.com.

4. HASIL DAN PEMBAHASAN

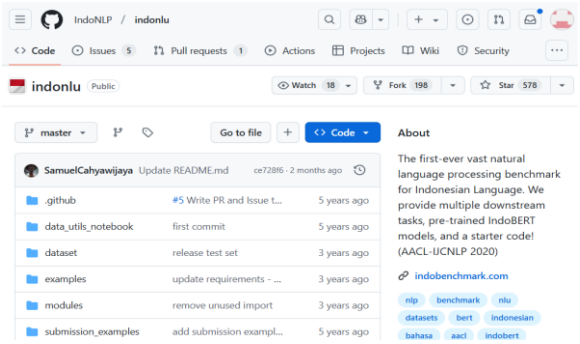
4.1. Pengumpulan data



reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
f841dd3b-	Pengguna G	https://play-l	Sangat menyenangkan, cepat akurat	5	0		###			
9be2e7b8	Pengguna G	https://play-l	sangat bermanfaat & membantu	5	0	4.85.2	###			4.85.2
770fbc12-	Pengguna G	https://play-l	mantap	5	0	4.86.1	###			4.86.1
b340fadf-	Pengguna G	https://play-l	jos	5	0	4.87.1	###			4.87.1
b45d19ca-	Pengguna G	https://play-l	Sangat bagus	5	0		###			
9a9a5a59-	Pengguna G	https://play-l	sangat membantu	5	0	4.87.1	###			4.87.1
03702a33-	Pengguna G	https://play-l	Mudah dan fleksibel	5	0	4.70.0	###			4.70.0
d28c765d-	Pengguna G	https://play-l	Ok.	4	0	4.84.0	###	Hai sobat	########	4.84.0
f68a5ae2-	Pengguna G	https://play-l	ok	5	0	4.87.1	###			4.87.1
e54e691f-	Pengguna G	https://play-l	nominal refund jauh berbeda, dari yg harusnya dapat 1,6jt tapi cuma dapat 190rb saat proses refund	1	0	4.86.1	###	Hai sobat	########	4.86.1
bae5022c-	Pengguna G	https://play-l	sejauh ini sangat membantu	5	0	4.87.1	###			4.87.1
83b1cf40-	Pengguna G	https://play-l	terbaik event olahraga sepakbola untuk tiket away, tpi sayang skrang gak ada lagi :(	5	0	4.87.1	###			4.87.1
9ac532c0-	Pengguna G	https://play-l	aplikasi goblog	1	0	4.87.1	###	Hai Dinah,	########	4.87.1

Gambar 2. Data ulasan Tiket.com

Penelitian ini melakukan pengumpulan data dengan teknik scraping pada ulasan pengguna aplikasi Tiket.com di Google Play Store. Sebanyak 10 ribu data ulasan terbaru per tanggal 24 Agustus 2024 dikumpulkan, dengan rentang waktu ulasan dimulai dari 9 November 2021. Proses crawling dilakukan menggunakan library Python, yaitu google_play_scraper, untuk memastikan akurasi dan kelengkapan data yang diperoleh. Selain itu, penelitian ini juga memanfaatkan dataset SMSA (Social Media Sentiment Analysis) dari IndoNLU, sebuah kumpulan dataset bahasa Indonesia yang dirancang untuk tugas-tugas Natural Language Processing (NLP). Dataset SMSA ini telah dilabeli dengan kategori sentimen positif, netral, dan negatif, yang berasal dari berbagai platform media sosial. Dataset tersebut terbagi menjadi tiga bagian, yaitu dataset latih, validasi, dan uji, yang diakses melalui repositori GitHub IndoNLU. Penggunaan dataset ini bertujuan untuk mendukung pengembangan model analisis sentimen yang lebih akurat dan relevan dengan konteks bahasa Indonesia.



Gambar 3. Github page dataset IndoNLU

4.2. Eksplorasi Data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   reviewId              10000 non-null  object
 1   userName              10000 non-null  object
 2   userImage             10000 non-null  object
 3   content               10000 non-null  object
 4   score                 10000 non-null  int64
 5   thumbsUpCount         10000 non-null  int64
 6   reviewCreatedVersion  7725 non-null   object
 7   at                    10000 non-null  object
 8   replyContent          3313 non-null   object
 9   repliedAt             3313 non-null   object
 10  appVersion            7725 non-null   object
dtypes: int64(2), object(9)
memory usage: 859.5+ KB
```

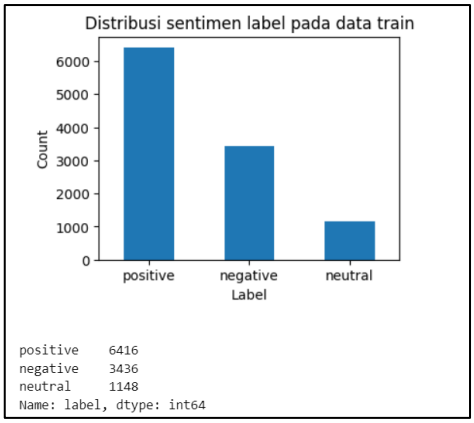
Gambar 4. Informasi data ulasan setelah di scraping

Data ulasan Tiket.com yang diambil melalui scraping mencakup 11 kolom dengan berbagai informasi. Beberapa kolom utama antara lain reviewId (ID unik ulasan), userName (nama pengguna), content (teks ulasan), dan score (rating 1-5). Ada juga kolom seperti thumbsUpCount (jumlah like), at (waktu ulasan dibuat), dan replyContent (balasan developer) jika ada. Beberapa kolom, seperti reviewCreatedVersion dan appVersion, mencatat versi aplikasi, meskipun tidak semua ulasan menyertakan informasi ini. Data ini memberikan gambaran lengkap tentang pengalaman pengguna dan interaksi mereka dengan aplikasi Tiket.com.

Selanjutnya dataset yang diunduh dari IndoNLU terdiri dari tiga bagian utama, yaitu dataset untuk training, validasi, dan testing. Pada dataset training, terdapat dua kolom: text yang berisi teks untuk dianalisis dan label yang menunjukkan kategori sentimen (positif, netral, atau negatif). Dataset ini terdiri dari 11.000 baris data dengan distribusi sentimen sebagai berikut: 6.416 data positif, 1.148 data netral, dan 3.436 data negatif. Dataset training ini digunakan untuk melatih model dalam proses fine-tuning agar dapat mengenali pola sentimen dengan lebih baik.

Untuk dataset validasi, terdapat dua kolom yang sama, yaitu text dan label, dengan total 1.260 baris data. Distribusi sentimen pada dataset validasi adalah 735 data positif, 131 data netral, dan 394 data negatif. Dataset ini berfungsi untuk mengevaluasi performa model selama proses pelatihan dan memastikan model tidak mengalami overfitting.

Sementara itu, dataset testing juga terdiri dari dua kolom (text dan label) dengan total 500 baris data. Distribusi sentimen pada dataset testing adalah 208 data positif, 88 data netral, dan 204 data negatif. Dataset ini digunakan untuk menguji performa akhir model setelah proses pelatihan dan validasi selesai, sehingga dapat memberikan gambaran akurat tentang kemampuan model dalam menganalisis sentimen.



Gambar 5. Distribusi label pada dataset train

4.3. Preprocessing Data

Tahap preprocessing dilakukan untuk membersihkan dan mempersiapkan data ulasan pengguna aplikasi Tiket.com yang diperoleh melalui teknik scraping. Tujuannya adalah memastikan data dalam format yang siap untuk diolah.

Proses preprocessing dimulai dengan seleksi kolom, di mana hanya kolom content yang berisi teks ulasan dipilih untuk analisis. Hal ini dilakukan untuk memfokuskan proses pada data yang relevan dan meningkatkan efisiensi dalam tahap selanjutnya.

Selanjutnya, dilakukan case folding untuk mengubah seluruh teks menjadi huruf kecil. Tujuannya adalah menghindari perbedaan antara kata dengan huruf kapital dan huruf kecil, sehingga analisis menjadi lebih konsisten.

Tabel 1. casefolding

Sebelum Case Folding	Sesudah Case folding
Aplikasi terbaikkk! Saya sukaaa sekali menggunakannya setiap hari 🎉✨	aplikasi terbaikkk! saya sukaaa sekali menggunakannya setiap hari 🎉✨

Karakter non-ASCII, seperti simbol dan emoticon, dihapus untuk memastikan data hanya berisi huruf, angka, dan tanda baca standar. Langkah ini membantu mengurangi noise dalam data yang dapat mengganggu proses analisis.

Tabel 2. hapus karakter non-ascii

Sebelum hapus karakter non-ascii	Sesudah hapus karakter non-ascii
aplikasi terbaikkk! saya sukaaa sekali menggunakannya setiap hari 🎉✨	aplikasi terbaikkk! saya sukaaa sekali menggunakannya setiap hari

Proses dilanjutkan dengan whitespace cleaning, di mana spasi ganda serta spasi kosong di awal atau akhir kalimat dihilangkan. Hal ini membuat teks lebih terstruktur dan bebas dari spasi yang tidak perlu.

Tabel 3. Whitespace cleaning

Sebelum whitespace cleaning	Sesudah whitespace cleaning
aplikasi terbaikkk! saya sukaaa sekali menggunakannya setiap hari	aplikasi terbaikkk! saya sukaaa sekali menggunakannya setiap hari

Terakhir, karakter berulang dalam kata (misalnya, "baaaik" menjadi "baik") dihilangkan untuk menyederhanakan kata ke bentuk dasarnya. Langkah ini mempermudah analisis dengan mengurangi variasi kata yang tidak perlu.

Tabel 4. Hapus karakter berulang

Sebelum menghapus karakter berulang	Sesudah menghapus karakter berulang
aplikasi terbaikkk! saya sukaaa sekali menggunakannya setiap hari	aplikasi terbaik! saya suka sekali mengunakanya setiap hari

Hasil dari preprocessing ini adalah data yang lebih bersih, terstruktur, dan siap untuk diolah dalam tahap analisis lebih lanjut.

4.4. Fine Tuning Model

Tahap fine-tuning dimulai dengan mengimpor library dan dataset yang diperlukan. Library seperti pandas, transformers (BERT), TensorFlow, dan scikit-learn digunakan untuk pengolahan data, tokenisasi, pelatihan model, dan evaluasi. Dataset SMSA dari IndoNLU diimpor dan diproses dengan memberikan header pada kolom text dan label, serta mengubah label kategoris menjadi numerik menggunakan LabelEncoder.

	text	label
0	warung ini dimiliki oleh pengusaha pabrik tahu...	2
1	mohon ulama lurus dan k212 mmbri hujjah partai...	1
2	lokasi strategis di jalan sumatera bandung . t...	2
3	betapa bahagia nya diri ini saat unboxing pake...	2
4	duh . jadi mahasiswa jangan sombong dong . kas...	0

Gambar 6. Dataset SMSA dari IndoNLU

Proses tokenisasi teks dilakukan menggunakan BertTokenizer untuk mengonversi teks menjadi format yang kompatibel dengan model BERT, menghasilkan input_ids, token_type_ids, dan attention_mask. Label kemudian dikonversi ke format tensor menggunakan TensorFlow untuk memastikan kompatibilitas dengan proses pelatihan. Dataset diubah ke format TensorFlow Dataset untuk memudahkan pengambilan data dalam batch dan pengaturan seperti shuffle.

Hyperparameter tuning dilakukan dengan memuat model BERT (TFBertF or Sequence Classification) dan mengatur optimizer Adam dengan learning rate 2e-5. Fungsi loss Sparse Categorical Crossentropy digunakan, dan model dikompilasi dengan metrik akurasi. Pelatihan model dilakukan dengan mengacak data (shuffle) dan menggunakan ukuran batch 16 untuk menyeimbangkan efisiensi dan penggunaan memori. Hasilnya adalah model yang siap untuk evaluasi lebih lanjut.

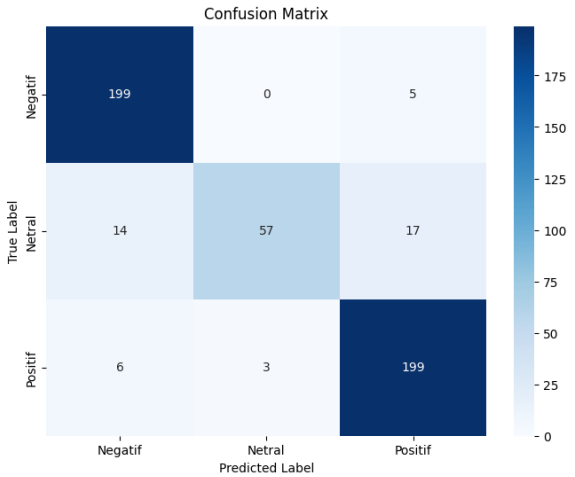
```
input_ids ->

tf.Tensor(
[    2  6540    92  2970   213  4259  3553   899    34   259  5590   262
  2558   386   899  1687    26  1574 30470   899  3310 30468 22130 30360
  6123  6368 30468 22130 30360  2652  1746 30468  8869  6540    34  6315
  1622  1256  8949   899 30468  4222  1622   752   245   295  2083 30470
  2346  7107   300 30470   405   724  5189 30470   843 17464   899   540
 10989  3331  1107 30468   119  3221    79    34  2170    98  9167 30457
     3     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0], shape=(128,), dtype=int32)
```

Gambar 7. input_ids setelah di tokenizing

4.5. Evaluasi Model

Hasil evaluasi confusion matrix menunjukkan performa model dalam mengklasifikasikan sentimen pada data uji. Untuk kelas negatif, model berhasil memprediksi 199 dari 204 sampel dengan benar, dengan 5 sampel salah diklasifikasikan sebagai positif. Pada kelas netral, dari 88 sampel, hanya 57 yang berhasil diprediksi dengan benar, sementara 14 sampel salah diklasifikasikan sebagai negatif dan 17 sebagai positif, menunjukkan bahwa kelas netral adalah yang paling sulit diidentifikasi. Untuk kelas positif, model menunjukkan performa yang sangat baik dengan 199 dari 208 sampel diprediksi benar, meskipun ada 6 sampel yang salah diklasifikasikan sebagai negatif dan 3 sebagai netral. Secara keseluruhan, model memiliki akurasi tinggi untuk kelas negatif dan positif, namun menghadapi tantangan dalam mengidentifikasi kelas netral, yang mungkin memerlukan penambahan data atau penyesuaian model untuk meningkatkan sensitivitas terhadap fitur-fitur kelas netral.



Gambar 8. Performa model dalam confusion matrix

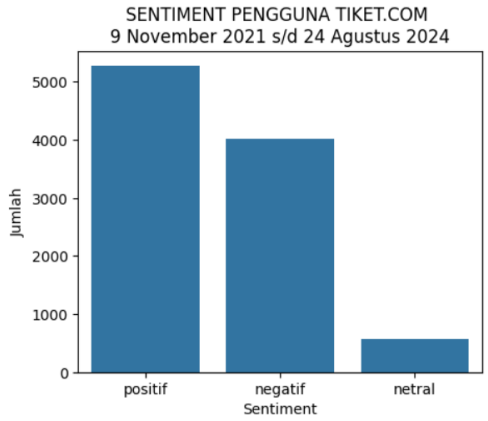
Berdasarkan evaluasi menggunakan library scikit-learn terhadap performa model yang telah di-fine-tuning, diperoleh akurasi keseluruhan sebesar 91%. Hal ini menunjukkan bahwa model mampu mengklasifikasikan data dengan benar pada 91% dari keseluruhan dataset. Kinerja model pada masing-masing kelas juga dievaluasi secara terpisah. Untuk

kelas negatif, model mencapai precision sebesar 91%, recall sebesar 98%, dan F1-score sebesar 94%, menunjukkan kemampuan yang sangat baik dalam mendeteksi dan mengklasifikasikan data negatif. Pada kelas netral, meskipun precision mencapai 95%, recall hanya sebesar 65%, dengan F1-score 77%, mengindikasikan bahwa model kurang efektif dalam mendeteksi data netral dibandingkan kelas lainnya. Sementara itu, untuk kelas positif, model mencapai precision 90%, recall 96%, dan F1-score 93%, yang menunjukkan performa yang sangat baik dalam mengidentifikasi data positif. Secara keseluruhan, model ini memiliki akurasi yang tinggi, meskipun masih terdapat ruang untuk peningkatan, terutama dalam meningkatkan recall untuk kelas netral.

	precision	recall	f1-score	support
0	0.91	0.98	0.94	204
1	0.95	0.65	0.77	88
2	0.90	0.96	0.93	208
accuracy			0.91	500
macro avg	0.92	0.86	0.88	500
weighted avg	0.91	0.91	0.91	500

Gambar 9. Classification report scikit-learn

4.6. Inferensi model terhadap ulasan Tiket.com



Sentiment	Jumlah
Positif	5271
Negatif	4022
Netral	574

Gambar 10. Hasil inferensi model

Setelah diterapkan pada data ulasan Tiket.com, model menghasilkan distribusi sentimen dengan mayoritas ulasan menunjukkan sentimen positif sebanyak 5.271 ulasan (52,7%). Sentimen negatif tercatat sebanyak 4.022 ulasan (40,2%), sementara sentimen netral hanya berjumlah 707 ulasan (7,1%). Dominasi sentimen positif mencerminkan bahwa sebagian besar pengguna merasa puas dengan layanan yang diberikan. Namun, proporsi sentimen negatif yang cukup signifikan memberikan wawasan penting terkait area yang memerlukan perbaikan, seperti masalah teknis pada aplikasi atau keluhan terhadap layanan pelanggan.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menganalisis sentimen pengguna aplikasi Tiket.com melalui ulasan di Google Play Store menggunakan model IndoBERT yang di-fine-tune pada dataset SmSA dari IndoNLU, dengan akurasi 91% dan F1-score tinggi untuk sentimen positif dan negatif, meskipun performa pada sentimen netral masih memerlukan perbaikan. Analisis menunjukkan mayoritas ulasan bernada positif (52,7%), mencerminkan kepuasan pengguna, sementara 40,2% ulasan bernada negatif mengindikasikan adanya masalah seperti kendala teknis dan layanan pelanggan. Penelitian ini menegaskan efektivitas IndoBERT dalam analisis sentimen berbahasa Indonesia dan memberikan kontribusi dalam memahami pola sentimen pengguna, sekaligus menjadi acuan untuk pengembangan NLP dan perbaikan layanan aplikasi berbasis ulasan pengguna.

DAFTAR PUSTAKA

[1] I. F. Hanif, I. R. Affandi, F. N. Hasan, E. Sinduningrum, and Z. Halim, "Analisis Sentimen Opini Masyarakat Terkait Penyelenggaraan Sistem Elektronik Menggunakan Metode Logistic Regression," *Jurnal Linguistik Komputasional*, vol. 5, no. 2, pp. 77–84, Nov. 2022, doi: 10.26418/jlk.v5i2.103.

[2] E. P. A. Akhmad, "Analisis Sentimen Ulasan Aplikasi DLU Ferry Pada Google Play Store Menggunakan Bidirectional Encoder Representations from Transformers," *JURNAL APLIKASI PELAYARAN DAN KEPELABUHANAN*, vol. 13, no. 2, pp. 104–112, Mar. 2023, doi: 10.30649/japk.v13i2.94.

[3] R. Merdiansah, S. Siska, and A. Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 1, pp. 221–228, Mar. 2024, doi: 10.55338/jikomsi.v7i1.2895.

[4] B. Liu, *Sentiment Analysis and Opinion Mining*. in Synthesis Lectures on Human Language Technologies. Springer International Publishing, 2022. [Online]. Available: https://books.google.co.id/books?id=xYhyEAAAQBAJ

[5] T. Jo, *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*. Springer International Publishing, 2021. [Online]. Available: https://books.google.co.id/books?id=0egdEAAAQBAJ

[6] I. Lauriola, A. Lavelli, and F. Aiolli, "An introduction to Deep Learning in Natural Language Processing: Models, techniques, and

tools," *Neurocomputing*, vol. 470, pp. 443–456, Jan. 2022, doi: 10.1016/j.neucom.2021.05.103.

[7] T. Wolf *et al.*, "Transformers: State-of-the-Art Natural Language Processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 38–45. doi: 10.18653/v1/2020.emnlp-demos.6.

[8] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2019, Accessed: Jul. 19, 2024. [Online]. Available: https://github.com/tensorflow/tensor2tensor

[9] S. T. M. Yessy Asri and M. K. Dr. Dra. Dwina Kuswardani, *MACHINE LEARNING & DEEP LEARNING: Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi*. Uwais Inspirasi indonesia, 2024. [Online]. Available: https://books.google.co.id/books?id=Yu7uEAAAQBAJ

[10] R. Botwright, *Deep Learning: Computer Vision, Python Machine Learning And Neural Networks*. Rob Botwright, 2024. [Online]. Available: https://books.google.co.id/books?id=KpbtEAAAQBAJ

[11] K. M. Ting, "Confusion Matrix," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 209–209. doi: 10.1007/978-0-387-30164-8_157.