Pemanfaatan IndoBERT dan *Long Short-Term Memory* (LSTM) pada Analisis Sentimen Ulasan Aplikasi Depok *Single Window*

Rian Hidayat[1,*], Windu Gata [1]

Ilmu Komputer; Universitas Nusa Mandiri; Jalan Jatiwaringin Raya No. 02 RT 08 RW 013 Kelurahan Cipinang Melayu Kecamatan Makasar Jakarta Timur, Telp (021)28534471; e-mail: 14230027@nusamandiri.ac.id, windu@nusamandiri.ac.id

* Korespondensi: e-mail: 14230027@nusamandiri.ac.id

Diterima: 16 Januari 2025; Review: 03 Mei 2025; Disetujui: 05 Juni 2025

Cara sitasi: Hidayat R, Gata W. 2025. Pemanfaatan IndoBERT dan *Long Short-Term Memory* (LSTM) pada Analisis Sentimen Ulasan Aplikasi Depok *Single Window*. Information System for Educators and Professionals. Vol 10(1): 13-24.

Abstrak: Diskominfo Kota Depok membuat Aplikasi Depok *Single Window* (DSW), yang menawarkan akses terpadu ke layanan publik seperti berita lokal, jadwal sholat, dan layanan kesehatan. Penelitian ini menginvestigasi analisis sentiment pada DSW dengan menggunakan pelabelan IndoBERT-large-p2, sejumlah 1.010 ulasan pengguna DSW yang ditemukan di Google Play Store dengan menggunkan crawling. Data kemudian dilakukan preprocessing, termasuk case folding, tokenizing, stemming, dan removal of stopwords, sebelum dilakukan pelabelan dengan IndoBert. Klasifikasi sentiment menggunakan Model Long Short-Term Memory (LSTM) dengan empat rasio pembagian data: 90:10, 80:20, 70:30, dan 60:40. Hasil didapatkan rasio 90:10 memiliki akurasi tertinggi 97% pada kelas negative dimana mendeteksi sentimen negatif dengan sangat baik. Model menunjukkan kinerja sangat baik dalam mendeteksi kelas Negatif, dengan presisi 0.97, recall 1.00, dan F1-score 0.98. Ini berarti hampir semua prediksi Negatif benar, dan tidak ada data Negatif yang terlewat. Namun, model ini kurang cocok untuk mendeteksi sentimen positif dan netral. Hasil ini membantu pengembang meningkatkan fungsionalitas dan kepuasan pengguna aplikasi DSW.

Kata kunci: Analisis Sentimen, Depok Single Window, IndoBERT, *Long Short-Term Memory, Web Scraping*

Abstract: *The Department of Communication and Information Technology (Diskominfo) of Depok City developed the Depok Single Window (DSW) application, offering integrated access to public services such as local news, prayer schedules, and healthcare services. This study investigates sentiment analysis on DSW using the IndoBERT-large-p2 labeling model, based on 1,010 user reviews of DSW collected from the Google Play Store through crawling. The data underwent preprocessing, including case folding, tokenizing, stemming, and stopword removal, before being labeled with IndoBERT. Sentiment classification was performed using the Long Short-Term Memory (LSTM) model with four data split ratios: 90:10, 80:20, 70:30, and 60:40. The results showed that the 90:10 ratio achieved the highest accuracy of 97% on the negative class, indicating excellent detection of negative sentiment. The model demonstrated outstanding performance in detecting negative sentiment, with a precision of 0.97, recall of 1.00, and F1-score of 0.98. This means nearly all negative predictions were correct, and no negative data was missed. However, the model was less effective in detecting positive and neutral sentiments. These results help developers improve the functionality and user satisfaction of the DSW application.*

Keywords: *Sentiment Analysis, Web Scraping, Depok Single Window, IndoBERT, Long Short-Term Memory*

1. Pendahuluan

Aplikasi Depok Single Window (DSW), yang dibuat oleh Dinas Komunikasi dan Informatika (Diskominfo) Kota Depok, adalah platform layanan digital yang bertujuan untuk membuat masyarakat lebih mudah mendapatkan informasi dan layanan publik. Aplikasi ini memiliki banyak fitur yang dapat diakses melalui smartphone, seperti berita lokal, jadwal sholat, informasi kesehatan, lowongan kerja, harga barang, cuaca, dan layanan darurat [1]. Namun, berdasarkan 1.010 ulasan pengguna yang diposting di Google Play Store, aplikasi ini gagal memenuhi harapan pengguna, dengan banyak keluhan tentang kinerja dan fitur. Ulasan ini membantu memahami persepsi publik dan menentukan area perbaikan untuk meningkatkan layanan aplikasi [2]. Oleh karena itu, analisis ulasan pengguna sangat penting untuk mengklasifikasikan pendapat masyarakat ke dalam kategori positif, negatif, atau netral. Ini memungkinkan pengembang untuk membuat rencana perbaikan yang tepat [3]. Untuk analisis sentimen penelitian ini, kami menggunakan model IndoBERT-large-p2, sebuah modifikasi arsitektur BERT yang dimaksudkan untuk memproses bahasa Indonesia secara komprehensif dengan kemampuan untuk mengidentifikasi konteks dan karakteristik bahasa yang lebih dalam [4]. Dilatih dengan korpus teks yang luas, mulai dari artikel berita hingga postingan media sosial, IndoBERT cocok untuk tugas pemrosesan bahasa alami seperti analisis sentimen [5][6].

Selanjutnya berjalan pada jalur *transformer* sejalan dengan *Bidirectional Encoder Representations from Transformers* dikenal BERT, adaptasinya bahasa Indonesia membuat entitas baru yang dikenal IndoBERT [7]. Tidak hanya itu model ini dapat digunakan dalam berbagai analisa, selain telah dilatih dengan jutaan kata IndoBERT mudah untuk digunakan dalam analisa teks atau kalimat [8]. IndoBERT sendiri dirancang khusus menangani teks dan kalimat yang menggunakan bahasa Indonesia, membuatnya relevan digunakan untuk analisis sentimen berbahasa lokal [9]. Banyak penelitian yang berhasil menggunakan IndoBERT dengan segala kemampuan yang dimilikinya menangani bahsa yang kompleks dengan mudah dan hasil akurasinya tinggi jika dibandingkan lainnya [10][11].

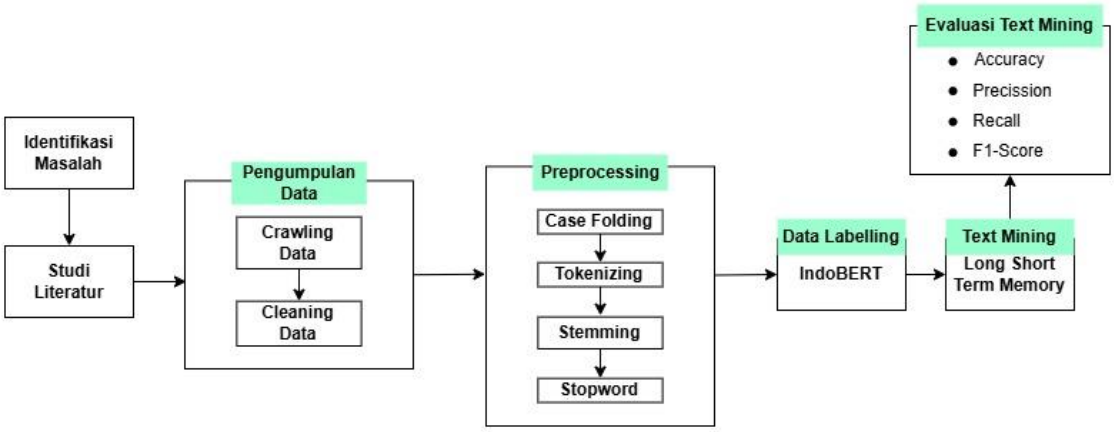
Hasil pengujiaan berbasis deep learning telah berkembang pesat dalam penelitian yang menggunakan komputerisasi, implementasi beragam kasus mulai dari analisis sentimen, pengolahan teks, gambar, suara dan video [12][13]. Pendekatan analisis sentimen berbasis teks seperti IndoBERT dirancang melabelkan suatu kalimat atau statement termasik dalam kategori buruk, netral atau baik, pada konteks ini negatif, netral dan positif [14][15]. *Long Short-Term Memory* termasuk dalam tipe arsitektur jaringan saraf yang dimaksudkan untuk meningkatkan akurasi prediksi untuk berbagai tujuan, seperti peramalan keuangan. Ini cocok untuk menganalisis tren keuangan masa lalu karena secara efektif menangkap ketergantungan jangka panjang pada data berurutan. Studi menunjukkan bahwa model LSTM dapat menghasilkan akurasi tinggi, terutama ketika digunakan bersama dengan *Convolutional Neural Networks* (CNN) [16][17][18].

Penelitian ini menggunakan IndoBERT untuk pelabelan otomatis ulasan pengguna aplikasi Depok Single Window (DSW). Selanjutnya, model Long Short-Term Memory (LSTM) digunakan untuk mengklasifikasikan sentimen dengan berbagai pembagian data agar hasilnya lebih akurat. Pendekatan ini menggabungkan teknologi terbaru dari IndoBERT dan LSTM untuk memahami bagaimana masyarakat menilai dan menggunakan DSW dalam mengakses informasi dan layanan publik di Depok. Inovasi dalam penelitian ini adalah penerapan gabungan IndoBERT dan LSTM dalam analisis sentimen aplikasi layanan publik di tingkat daerah, yang membantu pemerintah meningkatkan layanan digital berdasarkan masukan pengguna.

2. Metode Penelitian

Metode penelitian yang di usulkan untuk analisis sentimen ulasan pengguna aplikasi DSW, dapat dilihat pada gambar 1.

Vol. 10, No. 1, Juni 2025, 13 - 24



Gambar 1. Usulan Metode Penelitian

Identifikasi Masalah

Teknik ini bertujuan untuk mencari masalah yang mungkin ditemui oleh peneliti pada tahap selanjutnya. Salah satu masalah yang disebutkan adalah analisis sentimen ulasan DSW yang hanya dapat diakses melalui satu aplikasi smartphone, sehingga sulit untuk memastikan ulasan secara keseluruhan, apakah sebagian besar positif, negatif, atau netral [19].

Studi Literatur

Tahap ini melakukan tinjauan literatur terhadap artikel yang berkaitan dengan analisis sentimen. Kajian ini sudah lebih dulu digunakan para peneliti sebelumnya bahkan sudah mengembangkan model transformer seperti indoBERT untuk pelabelan sentimen apa yang diberikan untuk kalimat atau statement [20]. Untuk analisis sentimen ulasan pengguna aplikasi DSW, memanfaatkan deep learning untuk melatih dan menguji data untuk meningkatkan wawasan lebih mendalam.

Pengumpulan Data

Tahapan selanjutnya adalah proses metode pengumpulan data, dimana data yang digunakan adalah ulasan dari perangkat lunak di *marketplace google*. Teknik penggalian dengan memanfaatkan program *Python library Google Play Scraper* untuk mengakses dan mengumpulkan data. Temuan-temuan dari pengumpulan data kemudian proses *cleaning* dari *noise*, menghapus kolom yang tidak relevan pada konteks sentimen dan dicatat dalam format *file CSV*.

Preprocessing

Pada tahap ini, kata-kata diperbaiki untuk ulasan aplikasi yang ditulis oleh pengguna. *Cleansing, Tokenizing, Stopword,* dan *Stemming* adalah komponen dari tahapan ini. Penjelasan mendalam tentang hal-hal yang disebutkan di atas diberikan di bawah ini: a) *Case Folding*: langkah awal dalam analisis data, di mana kata-kata dengan huruf besar diubah menjadi huruf kecil; b) *Tokenizing*: memecah teks menjadi kata tunggal untuk memudahkan analisis. Ini bisa berupa kata, frasa, atau simbol dengan arti khusus.
c) *Stemming* adalah proses linguistik yang bertujuan untuk mengurangi kata ke bentuk dasarnya atau akarnya. Ini dilakukan dengan merubah teks ke awal kata untuk mengurangi variasi kata yang sama, seperti mengubah "daftar" menjadi "daftar". d) *Stopwords* menghapus kata-kata umum yang tidak signifikan dari data, seperti "dan", "atau", "tetapi". Ini adalah bagian penting dari proses preprocessing data teks, yang bertujuan untuk menghilangkan kata-kata umum yang tidak relevan untuk analisis.

Data Labelling

Penilaian pengguna bukanlah menjadi dasar pelabelan data, melainkan ulasan yang dipublikasikan. Teknik ini menggunakan model indobert sehingga hasil pengkategorian ulasan sebagai positif, negatif, atau netral dapat dilakukan secara mendalam [21]. Selain pelabelan penggunaan IndoBERT untuk memisahkan hasil ulasan, yaitu pelatihan untuk melatih sistem

agar dapat mendeteksi pola yang dicari, sedangkan pengujian untuk menguji hasil dari latihan yang telah dilakukan [22].

Text Mining

Metode text mining ini sama dengan istilah Penelitian tentang penambangan teks mencakup analisis data tidak terstruktur seperti file teks dengan menggunakan teknik pemrosesan bahasa alami, pembelajaran mesin, penambangan data, dan pengambilan informasi [23]. Digunakan untuk mendapatkan informasi berkualitas tinggi dari teks. Metode ini memungkinkan komputer untuk mengekstrak informasi dari berbagai sumber tertulis dan menemukan informasi baru [24]. Long Short-Term Memory, versi RNN yang dikembangkan oleh Hochreiter & Schmidhuber pada tahun 1997, digunakan dalam penelitian ini. Memperbarui arsitektur LSTM bahkan telah digunakan oleh peneliti dalam bidang lain seperti prediksi dan pengenalan suara.

Evaluasi Text Mining

Pada titik ini, *confusion matrix* digunakan untuk memeriksa kapasitas model dalam mengkategorikan data dan membantu dalam penentuan akurasi dan presisi. Klasifikasi dilakukan dengan membandingkan hasil yang diantisipasi oleh model dengan label sentimen. Setelah prosedur ini selesai, hasilnya dapat dievaluasi dalam hal akurasi, presisi, *recall*, dan total skor F1 untuk kemampuan model dalam menganalisis secara akurat, efisien, dan efektif [25].

3. Hasil dan Pembahasan

Pengumpulan Data

Pada pengumpulan data dengan *crawling data* didapatkan sebanyak 1.010 ulasan pengguna. Proses *cleaning data* kolom *reviewID, userImage, replyContent, dan appVersion* dihilangkan karena merupakan elemen yang tidak dibutuhkan dan tidak relevan pada penelitian ini, prosedur ini dilakukan agar data relevan dan dapat dilanjutkan proses analisis lanjutan.

Preprocessing

Setelah pengumpulan dan pembersihan data, tahap selanjutnya memastikan semua teks berubah bentuk menjadi lebih tersusun. Tahap ini dibantu *tools* yang populer mengolah data besar *IDE Python*, mulai dari tahap *case folding, tokenizing, stemming* dan *stopword removal.* Untuk fungsinya dijelaskan sebagai berikut.

a) Case folding

Salah satu fungu fungsi untuk mengubah semua huruf *capitalize* menjadi huruf *lower* supaya konsistensi.

```
    text = re.sub(r'RT[\s]+', '', text)
    # Menghapus URL
    text = re.sub(r'https?://\S+', '', text)
    # Menghapus karakter non-alfanumerik kecuali spasi
    text = re.sub(r'[^A-Za-z0-9 ]', '', text)
    # Menghapus spasi berlebih dan strip spasi di awal/akhir
    text = re.sub(r'\s+', ' ', text).strip()
    return text
# Membersihkan teks dalam kolom 'content' menggunakan fungsi clean_twitter_text
data['content'] = data['content'].apply(clean_twitter_text)
```

```
data['content'] = data['content'].str.lower()
data
```

	userName	content
0	Ina wirawan	mau daftar jadi pasien baru saja susah banget ...
1	Misbah Udlin	mau daftar ambil nomor antrian saja susah bang...
2	Mohamad Rival Jaya Saputra	banyak bug sering error padahal jaringan bagus...
3	Rahima Khalya Anugrah	aplikasi yang kurang membantu menurut saya har...
4	Anindya Nitisari	aplikasinya perlu diperbaiki dan dikembangkan ...
...	...	...
1003	Fahmi Adam	top aja
1004	Aznafatteh anakbunda	
1005	MASTER_BIJIX UMAM	
1006	Husein Barawas	

Gambar 2. Case Folding

b) Tokenizing

Berikutnya tokenizing Tokenizing adalah proses yang digunakan untuk membagi kata-kata yang termasuk dalam kalimat dalam data analisis sebelumnya menjadi satuan kata atau kata penyusun.

```
# Langkah preprocessing
data['normalized_content'] = data['content'].apply(normalize_text)
data['tokenized_content'] = data['normalized_content'].apply(tokenize_text)
data['filtered_content'] = data['tokenized_content'].apply(filter_stopwords)
data['stemmed_content'] = data['normalized_content'].apply(stem_text)
data['lemmatized_content'] = data['stemmed_content'].apply(lemmatize_text)

# Tampilkan hasil preprocessing (5 contoh pertama)
data[['content', 'normalized_content', 'tokenized_content', 'filtered_content', 'stemmed_content', 'lemmatized_content']].head()
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
```

	content	normalized_content	tokenized_content	filtered_content	stemmed_content	lemmatized_content
0	daftar pasien susah banget ngebug lurah gaada ...	daftar pasien susah sangat ngebug lurah gaada ...	[daftar, pasien, susah, sangat, ngebug, lurah,...	[daftar, pasien, susah, sangat, ngebug, lurah,...	daftar pasien susah sangat ngebug lurah gaada ...	daftar pasien susah sangat ngebug lurah gaada ...
1	daftar ambil nomor antri susah banget gilir da...	daftar ambil nomor antri susah sangat gilir da...	[daftar, ambil, nomor, antri, susah, sangat, g...	[daftar, ambil, nomor, antri, susah, sangat, g...	daftar ambil nomor antri susah sangat gilir da...	daftar ambil nomor antri susah sangat gilir da...
2	bug error jaring bagus gimana sih mudah masyar...	bug error jaring bagus gimana sih mudah masyar...	[bug, error, jaring, bagus, gimana, sih, mudah...	[bug, error, jaring, bagus, gimana, mudah, mas...	bug error jaring bagus gimana sih mudah masyar...	bug error jaring bagus gimana sih mudah masyar...
3	aplikasi bantu kali kali daftar ulang daftar n...	aplikasi bantu kali kali daftar ulang daftar n...	[aplikasi, bantu, kali, kali, daftar, ulang, d...	[bantu, daftar, ulang, daftar, nik, langsung, ...	aplikasi bantu kali kali daftar ulang daftar n...	aplikasi bantu kali kali daftar ulang daftar n...
4	aplikasi kembang masuk pas daftar tulis suka p...	aplikasi kembang masuk pas daftar tulis suka p...	[aplikasi, kembang, masuk, pas, daftar, tulis,...	[kembang, masuk, pas, daftar, tulis, suka, pen...	aplikasi kembang masuk pas daftar tulis suka p...	aplikasi kembang masuk pas daftar tulis suka p...

Gambar 3. Hasil preprocessing

Tabel 1. Hasil *Tokenizing*

Content	*Tokenizing*
mau daftar jadi pasien baru aja susah banget malah ngebug kelurahan juga gaada pilihannya giliran ngetik sndri disuruh pilih aneh banget padahal gaada pilihannya malah bikin sulid mau imunisasi anak jadi ketunda terus gara2 aplikasi doang!!!	['mau', 'daftar', 'jadi', 'pasien', 'baru', 'saja', 'susah', 'sangat', 'malah', 'ngebug', 'kelurahan', 'juga', 'gaada', 'pilihannya', 'giliran', 'ngetik', 'sndri', 'disuruh', 'pilih', 'aneh', 'sangat', 'padahal', 'gaada', 'pilihannya', 'malah', 'bikin', 'sulid', 'mau', 'imunisasi', 'anak', 'jadi', 'ketunda', 'terus', 'gara2', 'aplikasi', 'doang!!!']

c) Stemming

Pada tahap *stemming*, dapat dilihat pada gambar 3, imbuhan di awal, tengah, dan akhir kata dihilangkan untuk mengubah data ulasan menjadi kata dasar.

Tabel 2. Hasil *Stemming*

Content	*Stemming*
mau daftar jadi pasien baru aja susah banget malah ngebug kelurahan juga gaada pilihannya giliran ngetik sndri disuruh pilih aneh banget padahal gaada pilihannya malah bikin sulid mau imunisasi anak jadi ketunda terus gara2 aplikasi doang!!!	mau daftar jadi pasien baru saja susah sangat malah ngebug lurah juga gaada pilih gilir ngetik sndri suruh pilih aneh sangat padahal gaada pilih malah bikin sulid mau imunisasi anak jadi tunda terus gara2 aplikasi doang

d) Stopword Removal

Gambar 3 mengilustrasikan proses *stopword*. Meskipun penulis telah menghilangkan kata-kata yang tidak berguna, konten informasi kalimat tetap utuh.

Tabel 3. Hasil *Stopword Removal*

Content	*Stopword*
mau daftar jadi pasien baru aja susah banget malah ngebug kelurahan juga gaada pilihannya giliran ngetik sndri disuruh pilih aneh banget padahal gaada pilihannya malah bikin sulid mau imunisasi anak jadi ketunda terus gara2 aplikasi doang!!!	mau daftar jadi pasien baru saja susah sangat malah ngebug lurah juga gaada pilih gilir ngetik sndri suruh pilih aneh sangat padahal gaada pilih malah bikin sulid mau imunisasi anak jadi tunda terus gara2 aplikasi doang

Data Labelling

Labelling atau pelabelan pada data merupakan proses menentukan kategori atau kelas untuk setiap elemen dalam kumpulan data. Dalam penelitian ini, pelabelan menggunakan IndoBERT-large-p2 model yang di populerkan oleh Indobenchmark untuk mengkategorikan sentimen menjadi "positif", "negatif", atau "netral".

Tabel 4. Pelabelan *IndoBERT*

Content	*Indobert_label*	*Indobert_label_encode*
mau daftar ambil nomor antri saja susah banget gilir dateng langsung malah di suruh daftar lewat online gilir daftar aplikasi muter muter terus malah ngebug aplikasi masih mentah banyak kurang nya malah sulit warga	negative	0
waktu daftar dll lancar waktu mau daftar online buat ke poli umum pas suruh pilih pembayaran mau pake bpjs tidak bisa selalu loading muter terus lama banget waktu saya pencet bayar langsung bisa di konfirmasi maksudnya gimana ini tidak bisa pake bpjs kah	neutral	1
aplikasi ini sangat bantu jadi lebih mudah moga dapat tingkat lebih baik lagi	positive	2

Berdasarkan tabel 4, “mau daftar ambil nomor antrian saja susah banget giliran dateng langsung malah di suruh daftar lewat online giliran daftar aplikasi muter muter terus malah ngebug aplikasi masih mentah banyak kurang nya malah mempersulit warganya” indobert memberi label negative karena kalimat ini menggambarkan pengguna menggunakan aplikasi DSW sebagai pasien yang ingin mendaftar merasa kesulitan. Selanjutnya content “waktu daftar dll lancar waktu mau daftar online buat ke poli umum pas suruh pilih pembayaran mau pake bpjs tidak bisa selalu loading muter terus lama banget waktu saya pencet bayar langsung bisa di konfirmasi maksudnya gimana ini tidak bisa pake bpjs kah” ini menggambarkan aplikasi berjalan baik untuk beberapa menu, namun untuk menu yang diinginkan mendapatkan error sehingga belum maksimal membantu. Selanjutnya content “aplikasi ini sangat membantu saya sebagai warga depok lanjutkan” menggambarkan respon yang baik dari pengguna aplikasi DSW.

```
Some weights of BertForSequenceClassification were not initialized from the model checkpoint at indobenchmark/indobert-large-p2 and are newly initialized: ['classifier.bias', 'class
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
Proses analisis sentimen dimulai...
Proses analisis sentimen selesai!
Distribusi Sentimen:
indobert_label
negative    828
neutral      57
positive     35
Name: count, dtype: int64
```

Gambar 4. Hasil Pelabelan *IndoBERT-large-p2*

Dari hasil pengumpulan data 1010 data tersebut setelah di lakukan pre-processing menjadi 920 data. Hasil pelabelan menggunakan indoBERT-large-p2 ditemukan sebanyak 828 negative, 57 neutral dan 35 positif

Evaluasi Text Mining

Dari hasil pelabelan data, dapat dilihat hasil data analisis sentimen menggunakan indoBERT pada gambar 5. Distribusi sentimen histogram pada gambar 6. Kata-kata yang sering muncul pada ulasan yang divisualisasikan menggunakan library wordcloud. Visualisasi ulasan positif dapat dilihat pada gambar 7, ulasan netral pada gambar 8 dan ulasan negatif pada gambar 9.

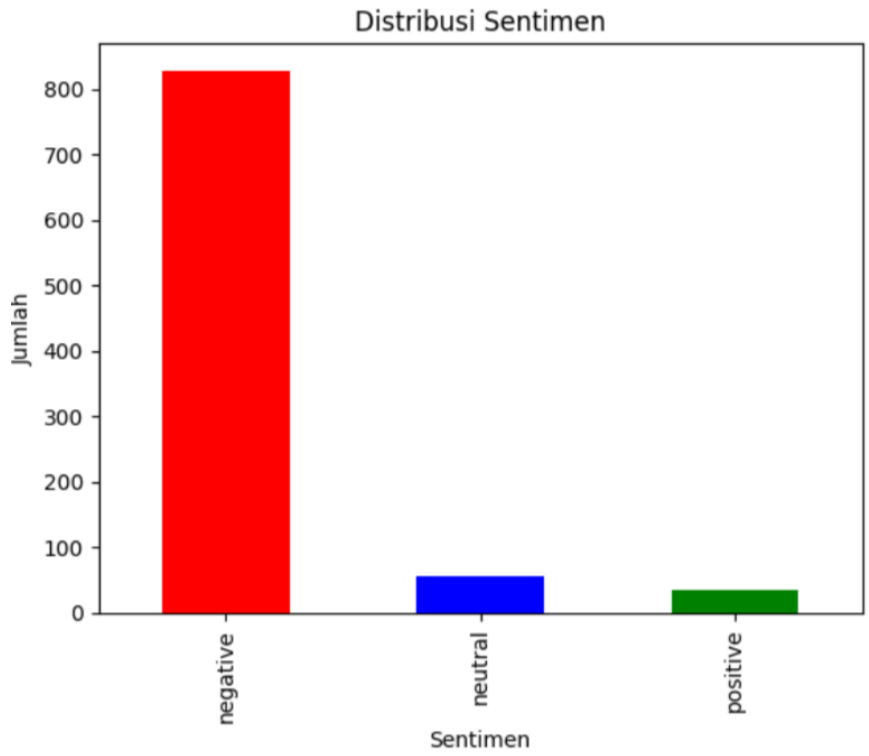
Vol. 10, No. 1, Juni 2025, 13 - 24

Contoh Data Hasil Analisis Sentimen:

	content	indobert_label	\
0	mau daftar jadi pasien baru saja susah banget ...	negative	
1	mau daftar ambil nomor antrian saja susah bang...	negative	
2	banyak bug sering error padahal jaringan bagus...	negative	
3	aplikasi yang kurang membantu menurut saya har...	positive	
4	aplikasinya perlu diperbaiki dan dikembangkan ...	negative	

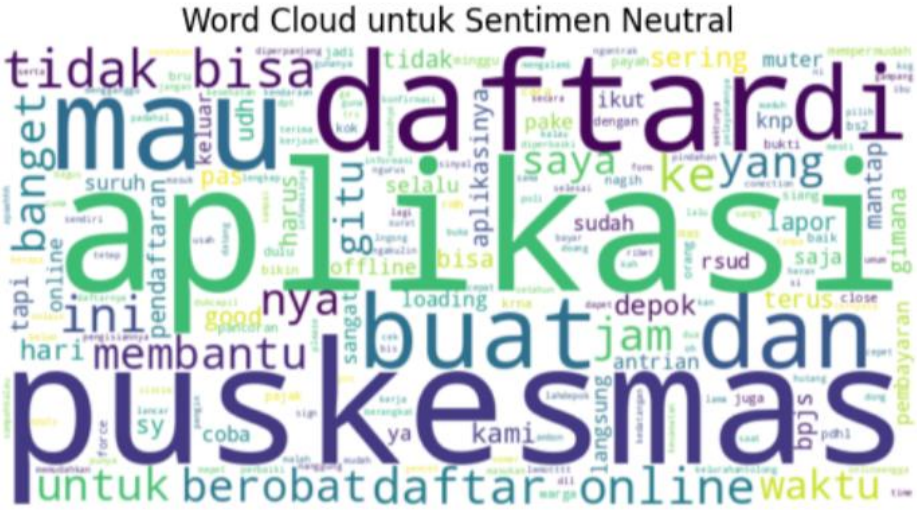
	indobert_label_encoded
0	0
1	0
2	0
3	2
4	0

Gambar 5. Hasil Analisis Sentimen

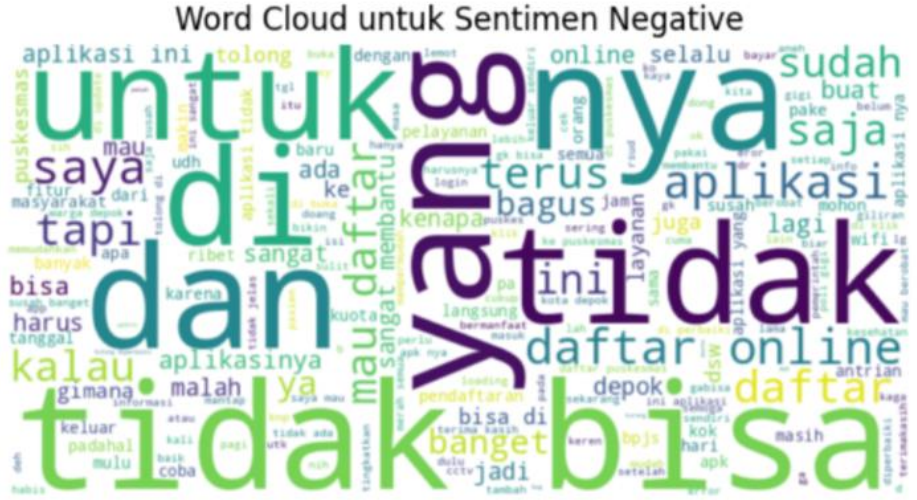


Gambar 6. Distribusi Analisis Sentimen

Gambar 7. Word Cloud Sentimen Positif



Gambar 8. Word Cloud Sentimen Netral



Gambar 9. Word Cloud Sentimen Negatif

Klasifikasi Deep Learning

Model pembelajaran mendalam Long Short Term Memory (LSTM) digunakan untuk evaluasi akhir, yang memerlukan data pelabelan indobert. Pengujian ini menentukan 10000 kata yang sering muncul dan disimpan dalam tokenizer. Untuk setiap teks dalam train dan test menjadi angka berdasarkan urutan yang dibuat oleh tokenizer. Lapisan LSTM dengan 64 unit dan lapisan dropout rate 0.5 untuk mencegah overfitting dengan membuang 50% neuron secara acak selama train. Lapisan dense 32 neuron dan activation ReLU, untuk lapisan output menggunakan softmax sehingga menghasilkan probabilitas 3 kelas yaitu negatif, netral dan positif. Optimizer Adam digunakan karena efisien dan sering digunakan. Ukuran epoch sebanyak 10 kali dengan beberapa skenario dan rasio pembagian data train dan test, dapat dilihat pada tabel 5.

Tabel 5. Skenario dan Rasio Pembagian Data

Skenario	Rasio Pembagian Data
1	90:10
2	80:20
3	70:30
4	60:40

Hasil Pengujian

Setelah data sudah dibagi sesuai dengan skenario yang ditentukan, mulai untuk proses pengujian menggunakan LSTM yang sudah dikonfigurasikan sesuai deksripsi tahap sebelumnya. Hasil ini akan dibagi sesuai dengan skenario yang telah ditentukan, yaitu.

a) Hasil Pengujian Skenario 1
Hasil pengujian data dengan rasio 90% train dan 10% test ditunjukan pada tabel dibawah ini dengan nilai akurasi 0.97.

Tabel 6. Hasil Skenario 90:10

Skenario	Label	Presisi	Recall	F1-Score	Support	Akurasi
90:10	Negatif	0.97	1.00	0.98	89	0.97
	Netral	0.00	0.00	0.00	1	
	Positif	0.00	0.00	0.00	2	

b) Hasil Pengujian Skenario 2
Hasil pengujian data dengan rasio 80% train dan 20% test ditunjukan pada tabel dibawah ini dengan nilai akurasi 0.93.

Tabel 7. Hasil Skenario 80:20

Skenario	Label	Presisi	Recall	F1-Score	Support	Akurasi
80:20	Negatif	0.93	1.00	0.96	171	0.93
	Netral	0.00	0.00	0.00	4	
	Positif	0.00	0.00	0.00	9	

c) Hasil Pengujian Skenario 3
Hasil pengujian data dengan rasio 70% train dan 30% test ditunjukan pada tabel dibawah ini dengan nilai akurasi 0.93.

Tabel 8. Hasil Skenario 70:30

Skenario	Label	Presisi	Recall	F1-Score	Support	Akurasi
70:30	Negatif	0.93	1.00	0.96	256	0.93
	Netral	0.00	0.00	0.00	9	
	Positif	0.00	0.00	0.00	11	

d) Hasil Pengujian Skenario 4
Hasil pengujian data dengan rasio 60% train dan 60% test ditunjukan pada tabel dibawah ini dengan nilai akurasi 0.91.

Tabel 9. Hasil Skenario 60:40

Skenario	Label	Presisi	Recall	F1-Score	Support	Akurasi
60:40	Negatif	0.91	1.00	0.95	336	0.91
	Netral	0.00	0.00	0.00	11	
	Positif	0.00	0.00	0.00	18	

Analisa Hasil

Berdasarkan pengujian menggunakan model deep learning LSTM, memiliki kinerja paling tinggi pada skenario 1 rasio 90:10. Hasil evaluasi label negatif dengan nilai akurasi 0.97, presisi 0.97, recall 1.00, f1-score 0.98 dan support 89. Artinya Model yang diuji memberikan hasil yang beragam untuk setiap kelas. Untuk kelas 'Negatif', model memiliki presisi 0,97, yang menyiratkan bahwa 97% dari prediksi 'Negatif' yang dihasilkan oleh model adalah benar. Selain itu, recall-nya adalah 1.00, menunjukkan bahwa model berhasil menemukan semua contoh yang benar-benar 'Negatif', dengan nilai F1-score 0.98, yang merupakan rata-rata harmonis dari akurasi dan recall. Terdapat 89 kejadian 'Negatif' dalam data pengujian.

Namun, untuk kelas 'Netral', model tidak memprediksi dengan baik. Presisi dan recall untuk kelas ini adalah 0.00, yang mengindikasikan bahwa model tidak dapat memprediksi atau menemukan kejadian 'Netral' sama sekali. Nilai F1-score juga 0.00, karena tidak ada prediksi yang valid untuk kelas ini, dengan hanya 1 kejadian 'Netral' pada data uji.

Demikian pula, untuk kelas 'Positif', model juga tidak berkinerja baik. Presisi dan recall juga 0.00, menyiratkan bahwa tidak ada kasus 'Positif' yang diprediksi dengan benar. Nilai F1-score untuk kelas ini juga 0.00, dan terdapat 2 kasus 'Positif' dalam data uji. Secara keseluruhan, model ini cukup efektif dalam mendeteksi kelas 'Negatif', tetapi tidak dapat mengenali kelas 'Netral' dan 'Positif'.

Kesimpulan

Setelah melalui berbagai tahapan untuk melakukan proses analisis terhadap ulasan pengguna Depok Single Window (DSW), berikut ini dapat disimpulkan:

1) Pelabelan IndoBERT sangat efisien dalam menentukan kalimat ini mengandung unsur negatif, netral atau positif. Pada modelnya dinyatakan bahwa IndoBERT model BERT yang dikembangakan oleh Indo Benchmark di Hugging Face terkini untuk bahasa Indonesia model yang sudah dilatih sebanyak 124.5 juta kata dengan menggunakan masked language modeling (MLM) dan next sentence prediction (NSP) objective.
2) Deep learning Long Short Term Memory (LSTM) dapat menguji hasil dari label analisis sentimen pada ulasan aplikasi DSW. Hasil pengujian dibagi menjadi 4 skenario dan memiliki hasil kinerja yang tinggi yaitu skenario 1 rasio pembagian model train dan test sebesar 90:10, nilai dari pengujian tersebut memiliki akurasi sebesar 0.97. Skenario 2 dan 3 yaitu rasio 80:20 dan 70:30 memiliki hasil 0.97, sedangkan skenario 4 yaitu rasio 60:40 memiliki hasil 0.91.

Referensi

[1] P. Arie Wijaya, “Analisis Sentimen Terhadap Review Pengguna Indrive Di Google Playstore Menggunakan Algoritma Support Vector Machine,” *J. Darma Agung*, vol. 31, no. 1, p. 1015, 2023, doi: 10.46930/ojsuda.v31i1.3078.

[2] I. T. Julianto, “Analisis Sentimen Terhadap Sistem Informasi Akademik Institut Teknologi Garut,” *J. Algoritm. Forum Ilm. bagi Ilmuwan dan Prakt. Tek. Inform.*, vol. 19, no. 1, pp. 449–456, 2022, doi: 10.33364/algoritma/v.19-1.1112.

[3] R. Merdiansah, S. Siska, and A. A. Ridha, “Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT,” *J. Ilmu Komput. Dan Sist. Inf.*, 2024, doi: 10.55338/jikomsi.v7i1.2895.

[4] A. Voutama and T. Ridwan, “Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC,” *J. Teknol. terpadu*, vol. 9, no. 1, pp. 42–48, 2023, doi: 10.54914/jtt.v9i1.609.

[5] W. D. E. Saputro, H. S. P. Yuana, and W. D. Puspitasari, “Analisis sentimen pengguna dompet digital dana pada kolom komentar google play store dengan metode klasifikasi support vector machine,” *JATI (Jurnal Mhs. Tek. Inform.*, 2023, doi: 10.36040/jati.v7i2.6842.

[6] W. A. Hidayat and V. R. S. Nastiti, “Perbandingan kinerja pre-trained indobert-base dan indobert-lite pada klasifikasi sentimen ulasan tiktok tokopedia seller center dengan model indobert,” *JSiI (Jurnal Sist. Informasi)*, vol. 11, no. 2, pp. 13–20, 2024, doi: 10.30656/jsii.v11i2.9168.

[7] F. T. Sanudin, B. Irawan, and A. Bahtiar, “Analisis sentimen ulasan aplikasi sapawarga di playstore menggunkan algoritma naÃ¯ve bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, 2024, doi: 10.36040/jati.v8i1.8300.

[8] H. Imaduddin, F. Y. Aâ€™la, and Y. S. Nugroho, “Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, 2023, doi: 10.14569/ijacsa.2023.0140813.

[9] R. Rahmadani, A. Rahim, and R. Rudiman, “Analisis sentimen ulasan â€œojol the gameâ€• di google play store menggunakan algoritma naive bayes dan model ekstraksi fitur tf-idf untuk meningkatkan kualitas game,” *JITET J. Inform. dan Tek. Elektro Terap.*, 2024, doi: 10.23960/jitet.v12i3.4988.

[10] W. Nurfitri and A. Chowanda, “Analisis Sentimen Pada Kasus Positif Covid-19 Berdasarkan Pemberitaan Media Di Indonesia Menggunakan Indobert,” *Progresif J. Ilm. Komput.*, 2024, doi: 10.35889/progresif.v20i1.1897.

[11] Anugerah Simanjuntak *et al.*, “Research and Analysis of IndoBERT Hyperparameter Tuning in Fake News Detection,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 13, no. 1, pp. 60–67, 2024, doi: 10.22146/jnteti.v13i1.8532.

[12] A. Simanjuntak *et al.*, “Research and Analysis of IndoBERT Hyperparameter Tuning in Fake News Detection,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 13, no. 1 SE-Articles, Feb. 2024, doi: 10.22146/jnteti.v13i1.8532.

[13] C. Bi and Y. Wang, “Long Short Term Memory Network (LSTM) Model Based on Neural Networks in Financial Forecasting,” pp. 1–5, 2024, doi: 10.1109/nmitcon62075.2024.10699249.

[14] S. Suleha, D. I. Inan, R. Juita, and V. A. lambadya Sirait, “Analisis Sentimen dan Topik Modeling Terhadap Aplikasi Mobile Detik.com,” *Technol. J. Ilm.*, vol. 15, no. 4, p. 828, 2024, doi: 10.31602/tji.v15i4.16363.

[15] R. Zulfiqri, B. N. Sari, and T. N. Padilah, “Analisis sentimen ulasan pengguna aplikasi media sosial instagram pada situs google play store menggunakan naÃ¯ve bayes classifier,” *JITET J. Inform. dan Tek. Elektro Terap.*, 2024, doi: 10.23960/jitet.v12i3.4995.

[16] X. Chun, M. Wang, Y. Ren, and S. Zhu, “Enhancing Aspect-based Sentiment Analysis in Tourism Using Large Language Models and Positional Information,” 2024, doi: 10.48550/arxiv.2409.14997.

[17] Q. Kang, D. Yu, K. H. Cheong, and Z. Wang, “Deterministic convergence analysis for regularized long short-term memory and its application to regression and multi-classification problems,” *Eng. Appl. Artif. Intell.*, vol. 133, p. 108444, 2024, doi: 10.1016/j.engappai.2024.108444.

[18] A. R. Mirza and A. K. Alâ€• Talabani, “Time Series-Based Spoof Speech Detection Using Long Short-Term Memory and Bidirectional Long Short-Term Memory,” *ARO. Sci. J. Koya Univ.*, vol. 12, no. 2, pp. 119–129, 2024, doi: 10.14500/aro.11636.

[19] G. R. Ramadhan and C. A. Sugianto, “Analisis sentimen ulasan aplikasi dana di google play store menggunakan algoritma naÃ¯ve bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 9849–9857, 2024, doi: 10.36040/jati.v8i5.10732.

[20] R. Fatmasari, W. Gata, N. Kusuma Wardhani, K. Prayogi, and M. Binti Husna, “Klasifikasi Sentimen Terhadap Kualitas Aplikasi Bahan Ajar Digital Akademik Universitas Terbuka di Google Play,” *J. SAINTEKOM*, vol. 14, no. 1, pp. 48–60, 2024, doi: 10.33020/saintekom.v14i1.591.

[21] D. Fristtikasari, S. Alam, and I. Kurniawan, “Analisis Sentimen Pengguna Aplikasi Kitalulus pada Ulasan Google Play Store Menggunakan Metode NaÃ¯ve Bayes,” *J. Teknol. Inform. Komput.*, vol. 10, no. 2, pp. 458–473, 2024, doi: 10.37012/jtik.v10i2.2244.

[22] I. K. Sendi Alpin Rizaldi Rizaldi, Syariful Alam, “Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) Pada Google Play Store Menggunakan Metode Naive Bayes,” *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, 2023, doi: 10.55123/storage.v2i3.2334.

[23] “Text Mining for Knowledge Discovery and Information Analysis,” pp. 1–5, 2024, doi: 10.1109/icccnt61001.2024.10725109.

[24] “Text Mining for Automated Data Analysis and Information Retrieval,” pp. 1–6, 2024, doi: 10.1109/icccnt61001.2024.10725784.
[25] “Sentiment Analysis and Text Mining in Environmental Sustainability and Climate Change,” *Pract. progress, Profic. Sustain.*, pp. 367–384, 2024, doi: 10.4018/979-8-3693-7230-2.ch020.