

EKSPERIMEN MODEL LSTM PADA KLASIFIKASI SENTIMEN PUBLIK TENTANG VAKSIN ASTRAZENECA

Diterima Redaksi: 1 November 2025; Revisi Akhir: 10 Desember 2025; Diterbitkan Online: 31 Mei 2026

Hardian Oktavianto ¹⁾, Henny Wahyu Sulistyo ²⁾, Taufiq Timur Warisaji ³⁾, dan Guruh Wijaya ⁴⁾

^{1, 2, 3, 4)} Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

^{1, 2, 3, 4)} Jalan Karimata 49 Kecamatan Sumbersari, Kabupaten Jember, Jawa Timur, 68121

e-mail: hardian@unmuhjember.ac.id¹⁾, henny.sulistyo@unmuhjember.ac.id²⁾, taufiqtimur@unmuhjember.ac.id³⁾,
guruh.wijaya@unmuhjember.ac.id⁴⁾

Abstrak: Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap vaksin AstraZeneca menggunakan metode Long Short-Term Memory. Data penelitian diperoleh dari media sosial X dengan rentang waktu 1 Februari 2024 hingga 15 Mei 2024, menghasilkan 251 data tweet yang telah melalui proses pembersihan, normalisasi, stopwords removal, stemming, tokenisasi, dan padding. Model dibangun menggunakan arsitektur Sequential API TensorFlow Keras dengan lapisan Embedding, SpatialDropout1D, LSTM, dan Dense beraktivasi sigmoid. Evaluasi dilakukan pada tiga jenis distribusi data: original, oversampled, dan SMOTE. Hasil menunjukkan bahwa akurasi model pada ketiga kondisi data relatif sama, yaitu sebesar 65%, dengan nilai loss berkisar antara 64% hingga 67%. Hal ini mengindikasikan bahwa teknik penyeimbangan data seperti oversampling dan SMOTE tidak memberikan peningkatan performa signifikan dan bahkan dapat menimbulkan overfitting. Secara keseluruhan, penelitian ini membuktikan bahwa model LSTM mampu menangani ketidakseimbangan data dengan baik serta memberikan gambaran kuantitatif terhadap persepsi publik mengenai vaksin AstraZeneca.

Kata Kunci—analisis sentimen; LSTM; media sosial; vaksin AstraZeneca

Abstract: This study aims to analyze public sentiment towards the AstraZeneca vaccine using the Long Short-Term Memory method. The research data was obtained from social media X, spanning the period from February 1, 2024, to May 15, 2024, resulting in 251 tweets that have undergone cleaning, normalization, stopwords removal, stemming, tokenization, and padding processes. The model was built using the TensorFlow Keras Sequential API architecture with Embedding, SpatialDropout1D, LSTM, and Dense layers with sigmoid activation. Evaluation was conducted on three types of data distributions: original, oversampled, and SMOTE. The results show that the model accuracy across all three data conditions is relatively similar, at 65%, with loss values ranging from 64% to 67%. This indicates that data balancing techniques such as oversampling and SMOTE do not provide significant performance improvements and can even lead to overfitting. Overall, this study proves that the LSTM model is able to handle data imbalance well and provides a quantitative overview of public perception towards the AstraZeneca vaccine.

Keywords—AstraZeneca, LSTM, sentiment analysis, social media

Penulis Korespondensi:

Hardian Oktavianto

Program Studi Teknik Informatika

Universitas Muhammadiyah Jember

Email: hardian@unmuhjember.ac.id

Orchid ID: <https://orcid.org/0000-0002-2446-7401>

I. PENDAHULUAN

ANALISIS sentimen adalah salah satu cabang ilmu dari penambangan teks atau *text mining* yang memiliki fokus pada pemahaman dan pengelompokan opini atau emosi yang tertulis maupun tersirat dalam suatu teks [1]. Di era sosial media yang berkembang secara cepat dan pesat seperti saat ini, banyak informasi berbentuk opini pengguna yang dapat dianalisis untuk mendapatkan wawasan publik [2]. Salah satu metode di bidang atau lingkup jaringan saraf tiruan yang banyak digunakan dalam melakukan analisis sentimen adalah *Long Short-Term Memory* (LSTM), dimana metode ini adalah lanjutan atau pengembangan lanjut dari jaringan saraf berulang (*Recurrent Neural Network* atau RNN) [3]. Model ini mampu memecahkan masalah *vanishing gradient* dan lebih baik dalam mempelajari konteks jangka panjang dari teks [4].

Pandemi COVID-19 beberapa waktu lalu telah menimbulkan berbagai reaksi dan opini di masyarakat, terutama terkait program vaksinasi [5]. Berkaitan dengan program vaksinasi ini dan telah ditemukannya beberapa jenis vaksin, AstraZeneca adalah salah satu jenis vaksin COVID-19 yang digunakan secara luas di Indonesia [6]. Komentar – komentar masyarakat mengenai vaksin tersebut bervariasi mulai dari dukungan terhadap efektivitas vaksin hingga kekhawatiran terhadap efek sampingnya. Data opini publik ini menjadi sumber yang sangat berharga untuk memahami persepsi masyarakat terhadap vaksinasi dan dapat dijadikan dasar dalam menilai tingkat penerimaan serta efektivitas komunikasi publik yang dilakukan pemerintah [7].

Analisis terhadap opini masyarakat tersebut dapat dilakukan dengan pendekatan analisis sentimen, yaitu teknik berbasis penambangan teks yang digunakan untuk melakukan identifikasi dan melakukan pengelompokan opini menjadi sentimen positif, sentimen negatif, atau sentimen netral [8]. Dalam konteks vaksin AstraZeneca, analisis sentimen berperan penting dalam memetakan persepsi publik, mendeteksi isu-isu yang berpotensi menimbulkan keraguan terhadap vaksinasi, serta membantu pengambil kebijakan dalam merancang strategi komunikasi yang lebih efektif [9]. Dengan memahami pola sentimen yang muncul di media sosial, pihak terkait dapat mengambil keputusan yang lebih tepat berbasis data untuk meningkatkan kepercayaan masyarakat terhadap vaksin [10].

Dalam melakukan analisis sentimen, diperlukan model yang mampu menangkap konteks makna dari urutan kata dalam kalimat [11]. Model *Long Short-Term Memory* merupakan salah satu model arsitektur pengembangan dari jaringan saraf tiruan yang terbukti efektif dalam mengolah data teks karena kemampuannya mengenali hubungan dependensi jangka panjang antar kata [12]. Berbeda dengan metode pembelajaran mesin tradisional seperti *Naïve Bayes* atau *Support Vector Machine* (SVM), LSTM tidak hanya memperhatikan frekuensi kata, tetapi juga urutan dan konteksnya [13]. Oleh karena itu, penerapan model LSTM dalam analisis sentimen komentar masyarakat tentang vaksin AstraZeneca diharapkan dapat menghasilkan klasifikasi yang lebih akurat dan representatif terhadap opini yang sebenarnya.

Kombinasi antara dataset opini masyarakat, metode analisis sentimen, dan pendekatan jaringan saraf tiruan berbasis LSTM menjadikan penelitian ini memiliki nilai ilmiah dan praktis yang signifikan. Dari sisi sosial, penelitian ini memberikan gambaran kuantitatif mengenai persepsi masyarakat terhadap vaksin AstraZeneca, sedangkan dari sisi teknis, penelitian ini menguji efektivitas LSTM dalam pemrosesan bahasa alami berbahasa Indonesia. Sehingga penelitian ini diharapkan dapat berkontribusi dalam pengembangan model analisis sentimen serta memberikan manfaat nyata dalam memahami dinamika opini publik.

Sebagai rumusan masalah yang diangkat pada penelitian ini adalah, bagaimana proses pra-pemrosesan data teks untuk analisis sentimen berbahasa Indonesia, bagaimana arsitektur model LSTM dapat diterapkan pada data opini masyarakat tentang vaksin AstraZeneca, bagaimana kinerja atau performa model LSTM dalam melakukan klasifikasi data sentimen positif dan sentimen negatif. Tujuan dari penelitian ini adalah melakukan implementasi model LSTM untuk analisis sentimen pada komentar masyarakat tentang vaksin AstraZeneca, mengevaluasi performa model LSTM dalam mengklasifikasikan sentimen positif dan sentimen negatif, memberikan penjelasan secara kuantitatif mengenai persepsi masyarakat terhadap vaksin AstraZeneca berdasarkan opini di media sosial X (media sosial ini sebelumnya dikenal dengan nama Twitter dan sekarang berganti nama menjadi X, sumber: <https://en.wikipedia.org/wiki/Twitter>). Hasil penelitian ini diharapkan dapat menjadi referensi pengembangan model analisis sentimen berbasis jaringan saraf tiruan untuk bahasa Indonesia, sebagai

sumber informasi bagi pemerintah dan lembaga kesehatan dalam memahami opini publik terhadap vaksinasi, sebagai kontribusi ilmiah dalam penerapan *deep learning* untuk analisis teks dan opini masyarakat.

II. TINJAUAN PUSTAKA

Model *Long Short-Term Memory* adalah bentuk khusus dari RNN yang memiliki kemampuan dalam mempelajari *long-term dependencies* [14]. Pada model ini sebuah *encoder* diterapkan pada pernyataan tersembunyi terakhir dari LSTM. Arsitektur LSTM ditunjukkan pada Gambar 1, dan persamaan lengkap untuk *update* pengetahuan adalah sebagai berikut [15]:

$$f_{m+1} = \sigma(\theta^{(h \rightarrow f)} h_m + \theta^{(x \rightarrow f)} x_{m+1} + b_f) \text{ forget gate} \quad (1)$$

$$i_{m+1} = \sigma(\theta^{(h \rightarrow i)} h_m + \theta^{(x \rightarrow i)} x_{m+1} + b_i) \text{ input gate} \quad (2)$$

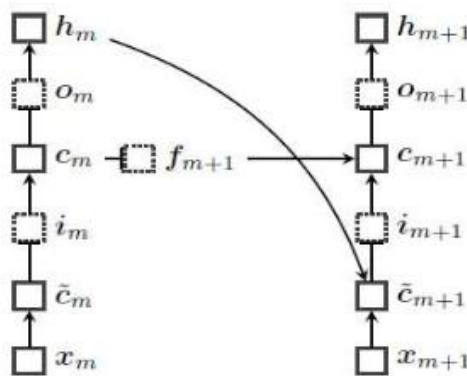
$$\tilde{c}_{m+1} = \tanh(\theta^{(h \rightarrow c)} h_m + \theta^{(w \rightarrow c)} x_{m+1}) \text{ update candidate} \quad (3)$$

$$c_{m+1} = f_{m+1} \odot c_m + i_{m+1} \odot \tilde{c}_{m+1} \text{ memory cell update} \quad (4)$$

$$o_{m+1} = \sigma(\theta^{(h \rightarrow o)} h_m + \theta^{(x \rightarrow o)} x_{m+1} + b_o) \text{ Output gate} \quad (5)$$

$$h_{m+1} = o_{m+1} \odot \tanh(c_{m+1}) \text{ Output} \quad (6)$$

Operator $\odot$ adalah produk elemen demi elemen. Forget Gate sesuai formula (1) memutuskan informasi mana dari memori sebelumnya, kemudian Input Gate yang diwakili formula (2) menambah informasi baru ke cell state melalui formula (3), Cell State diperbarui berdasarkan gabungan informasi lama + baru sesuai dengan formula (4), dan Output Gate menghasilkan output seperti pada formula (5). Model LSTM adalah hasil penambahan status tersembunyi yang dinyatakan dalam h_m dengan sel memori direpresentasikan sebagai c_m . Nilai sel memori pada setiap waktu ke- m adalah bentuk gerbang yang merupakan penjumlahan dari dua kuantitas: nilai sel memori sebelumnya yang dinyatakan dengan c_{m-1} , dan nilai sel memori yang telah diperbarui yang dinyatakan dengan c_m , yang dihitung dari masukan sebelumnya dalam bentuk x_m saat ini dan status tersembunyi sebelumnya dalam bentuk h_{m-1} . Status selanjutnya adalah h_m yang diwakili oleh formula (7) dihitung dari memori sel. Karena sel memori tidak melewati jalur fungsi non-linier selama pembaruan, informasi dapat diperoleh melalui jaringan jarak jauh.



Gambar 1. Arsitektur LSTM

Setiap gerbang dikendalikan oleh vektor yang memiliki bobot, yang menentukan status tersembunyi sebelumnya (misalnya, $\theta^{(h \rightarrow f)}$) dan arus masukan (misalnya, $\theta^{(x \rightarrow f)}$), ditambah *offset* vektor (misalnya, b_f). Operasi keseluruhan dapat diringkas secara informal sebagai $(h_m, c_m) = \text{LSTM}(x_m, (h_{m-1}, c_{m-1}))$, dengan (h_m, c_m) mewakili status LSTM setelah membaca token m . LSTM mengungguli jaringan saraf tiruan standar dalam berbagai masalah. Gerbang ditampilkan dalam kotak dengan tepi putus-putus. Dalam model bahasa LSTM, setiap h_m akan digunakan untuk memprediksi kata berikutnya w_{m+1} . LSTM mengungguli jaringan saraf rekuren standar dalam berbagai masalah, misalnya dalam masalah pemodelan bahasa.

Jaringan saraf berulang (RNN) diperkenalkan sebagai teknik pemodelan bahasa, di mana konteks dalam token m (di mana token berhenti) serta token M (nilai token M yang ditentukan) diringkas oleh vektor yang diperbarui berulang kali,

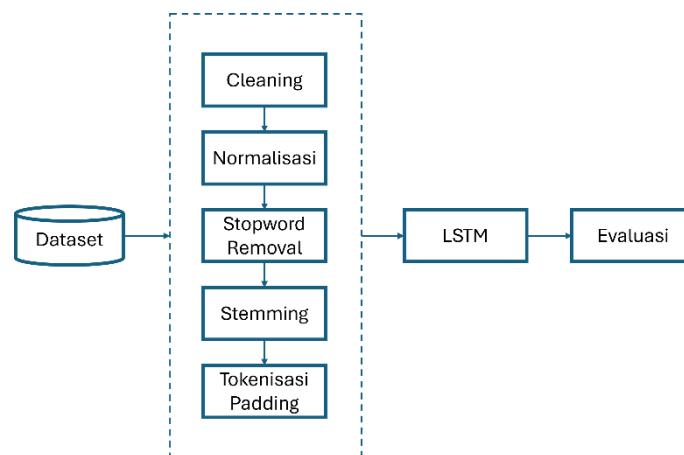
$$h_m = g(x_m, h_{m-1}), m = 1, 2, \dots, M, \quad (7)$$

dengan x_m adalah vektor tersimpan dari token w_m dan fungsi g mendefinisikan *loop*. Kondisi awal h_0 adalah parameter tambahan model. Model LSTM (Memori Jangka Panjang dan Pendek) memiliki *loop* yang lebih kompleks, di mana sel-sel memori melewati serangkaian gerbang, menghindari penerapan non-linearitas yang berulang. Gerbang adalah fungsi masukan dan merupakan status tersembunyi sebelumnya. Bentuknya dihitung dari aktivasi sigmoid elemen demi elemen, $\sigma(x) = (1 + \exp(-x))^{-1}$, yang memastikan bahwa nilainya akan berada dalam rentang $[0, 1]$. Oleh karena itu, bentuk ini dapat dilihat sebagai gerbang logika terdiferensiasi.

III. METODE PENELITIAN

A. Desain Penelitian

Studi ini mengaplikasikan pendekatan kuantitatif yang menggunakan metode eksperimen dengan acuan *machine learning* dan *deep learning*. Fokus utama penelitian adalah membangun dan mengevaluasi model *Long Short-Term Memory* (LSTM) untuk melakukan klasifikasi sentimen masyarakat terhadap vaksin AstraZeneca berdasarkan data komentar dalam bentuk teks di media sosial.



Gambar 2. Rancangan Penelitian

Pendekatan ini dipilih karena mampu memberikan hasil yang terukur dan dapat diuji secara empiris terhadap data nyata yang bersumber dari opini publik. Gambar 2 merepresentasikan rancangan penelitian yang dilakukan.

B. Dataset Penelitian

Pada tahap pengumpulan data, penelitian ini melakukan proses akuisisi data berupa *tweet* yang diambil dari media sosial X. Pemilihan *platform* tersebut didasarkan pada pertimbangan bahwa media sosial X merupakan salah satu sumber informasi publik yang banyak digunakan untuk mengekspresikan opini, pandangan, dan tanggapan terhadap isu-isu kesehatan, termasuk terkait vaksin dan produk farmasi [16].

Proses pengumpulan data dilakukan dengan menggunakan kata kunci “AstraZeneca”, yang dipilih karena secara langsung merepresentasikan topik penelitian mengenai persepsi publik terhadap vaksin atau perusahaan terkait. Rentang waktu pengambilan data ditetapkan dari 1 Februari 2024 hingga 15 Mei 2024. Potongan dataset yang diperoleh ditunjukkan pada gambar 3.

Melalui proses pencarian dan penyaringan data menggunakan kriteria tersebut, diperoleh sebanyak 460 *tweet*. Data yang telah terkumpul kemudian akan digunakan sebagai bahan analisis lebih lanjut untuk menjawab rumusan masalah dan mencapai tujuan penelitian yang telah ditetapkan.

Dalam penelitian ini, proses pelabelan data dilakukan secara manual oleh peneliti untuk memahami konteks, nuansa bahasa, dan makna implisit dalam setiap *tweet* secara lebih mendalam [17]. Label yang dipakai adalah ‘Positif’ dan ‘Negatif’. Pendekatan ini dipilih guna memastikan akurasi klasifikasi, mengingat *tweet* atau komentar pengguna cenderung mengandung bahasa informal, bahasa singkatan, atau sarkasme yang sulit dikenali oleh sistem otomatis [18]. Untuk menjamin validitas dan reliabilitas data, hasil pelabelan kemudian divalidasi oleh seorang ahli bahasa agar setiap label sesuai dengan makna dan konteks *tweet* [19]. Setelah tahap pelabelan ini dataset berkurang menjadi 251 data.

C. Pra-pemrosesan Data

Sebelum dilakukan pemodelan, dataset diolah melalui *preprocess* data untuk mengurangi *noise* dan menyeragamkan format data. Tahapan yang dilakukan meliputi:

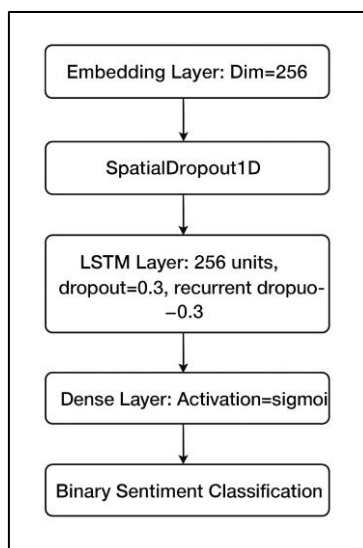
1. Pembersihan teks (*text cleaning*): Merupakan proses menghapus karakter – karakter non-alfabet, tanda baca, dan simbol – simbol yang tidak relevan.
2. Normalisasi: Mengubah seluruh huruf pada dataset menjadi huruf kecil agar konsisten.
3. *Stopword removal*: Proses melakukan penghapusan kata umum yang tidak memiliki makna sentimen signifikan dengan menggunakan pustaka Sastrawi.
4. *Stemming*: Proses merubah kata menjadi bentuk dasarnya (contoh: “diberikan” → “beri”) untuk mengurangi variasi kata yang berlebihan.
5. Tokenisasi dan *Padding*: Teks diubah menjadi urutan bilangan (*token sequence*) menggunakan *Tokenizer* dari Keras, dan diseragamkan panjangnya hingga 100 token per teks dengan fungsi *pad_sequences*.

Tahapan ini memastikan bahwa model LSTM dapat menerima input yang bersih, seragam, dan representatif terhadap makna semantik dari teks.

```
Tweets,Label
"Alasan Sosiolog Unair Sebut Penarikan Vaksin AstraZeneca Bisa Memicu Kecemasan Publik",Negatif
". . menderita luka serius setelah menerima vaksin COVID ini. AstraZeneca mengakui di pengadilan bahwa suntikan COVID-19 mereka berpotensi menyebabkan efek samping yang mengancam jiwa seperti pembekuan darah. Pemerintah Inggris memiliki perjanjian hukum dengan AstraZeneca",Negatif
"Raksasa farmasi Inggris-Swedia AstraZeneca telah mulai mencabut persetujuan penjualan dan penggunaan vaksin COVID-19 di seluruh dunia. Sejak tahun lalu telah ada gugatan class action yg diajukan terhadap AstraZeneca oleh keluarga dari mereka yang meninggal atau...",Negatif
"Sosiolog Unair Sebut Penarikan Vaksin AstraZeneca Perburuk Situasi. Sosiolog Unair Septi Ariadi menyebut penarikan vaksin AstraZeneca oleh pemerintah dapat memperburuk situasi dan memicu",Negatif
"Semoga kalian yang menganjurkan BISA BERTANGGUNG JAWAB atas KORBAN yang sudah berjatuh akibat PUCK_SIN BERACUN AstraZeneca dkk",Negatif
"Sehat sehat dan panjang umur untuk aku dan kalian semua yang pernah di vaksin AstraZeneca",Positif
"Astrazeneca dan efek samping? Kalau ngomongin ini dizaman kopid dulu bisa diomelin paramedik dadakan dan paramedik betulan",Negatif
"apakah ak udah terkena efek astrazeneca .bulan kemaren baru sembuh flu masa sekarang udah gejala flu lagi",Negatif
"kepala rasa mcm kena cucuk cucuk eh dua tiga harini. effect astrazeneca kepa'ku",Negatif
```

Gambar 3. Dataset

D. Pembangunan Model LSTM



Gambar 4 Arsitektur LSTM

Model atau arsitektur LSTM seperti direpresentasikan oleh gambar 4, dibangun menggunakan *Sequential API* dari *TensorFlow Keras* [20] dengan arsitektur sebagai berikut:

1. *Embedding Layer*: Dimensi 256, untuk merepresentasikan kata dalam bentuk vektor numerik.
2. *SpatialDropout1D*: Mencegah *overfitting* dengan mengabaikan sebagian unit selama pelatihan.
3. *LSTM Layer*: 256 unit *neuron* dengan *dropout* dan *recurrent dropout* sebesar 0.3, untuk menangkap konteks temporal dalam teks.
4. *Dense Layer*: Aktivasi sigmoid untuk mengklasifikasikan sentimen *biner* (positif/negatif).

Model ini kemudian dikompilasi menggunakan *optimizer Adam* dengan fungsi *cost binary_crossentropy* dan fungsi metrik *accuracy*. Proses pelatihan dilakukan selama 10 *epoch* dengan ukuran *batch* 32.

E. Evaluasi Model

Untuk menguji performa model, dataset dibagi menjadi 80% untuk data latih dan 20% untuk data uji menggunakan metode *train_test_split*. Selain itu, dilakukan *oversampling* dengan *RandomOverSampler* dan SMOTE untuk menyeimbangkan distribusi kelas positif dan negatif agar model tidak bias terhadap salah satu kategori. Evaluasi performa menggunakan metrik akurasi dan loss, serta dilakukan analisis terhadap hasil prediksi untuk menilai kemampuan model dalam mengenali sentimen publik secara efektif.

IV. HASIL DAN PEMBAHASAN

A. Analisis Distribusi Data

Distribusi awal dataset menunjukkan ketidakseimbangan kelas, di mana jumlah komentar negatif lebih banyak dibandingkan komentar positif. Komentar positif 92 buah dan komentar negatif sejumlah 159.

Langkah penyeimbangan data dilakukan dengan *oversampling* dan SMOTE, penyeimbangan data dilakukan dengan menggunakan dan memanfaatkan *library* dari python, sehingga distribusi menjadi seimbang, masing – masing label kelas mempunyai jumlah sama yaitu 126 buah. Model LSTM akan dilakukan pada 3 jenis distribusi data, *Original*, *Oversampled*, dan SMOTE. Tabel 1 menyajikan distribusi dari hasil klasifikasi menggunakan LSTM.

Tabel 1. Distribusi Data

Label	Original	Oversampled	SMOTE
Positif	92	126	126
Negatif	159	126	126

B. Hasil Kinerja Model

Setelah melalui tahapan pelatihan, model LSTM menunjukkan performa yang baik dalam mengklasifikasikan komentar masyarakat terhadap vaksin AstraZeneca.

Tabel 2 menyajikan hasil perbandingan akurasi dan *loss* dari penelitian ini. Pada data *original* akurasi yang diperoleh sebesar 65% dan loss 64%, pada data *oversampled* akurasi sebesar 65% dan loss 64%, dan pada data SMOTE juga sama yaitu akurasi sebesar 65% dan loss 67%.

Hal ini menunjukkan bahwa penerapan model LSTM pada dataset ini, penyeimbangan data menggunakan SMOTE tidak meningkatkan akurasi pengujian dan bahkan mengakibatkan *loss* pada pengujian yang lebih besar dibandingkan dengan ketika menggunakan data *original* atau data *oversampled*.

Tabel 2. Perbandingan Akurasi dan Loss

	Original	Oversampled	SMOTE
Akurasi	65%	65%	65%
Loss	64%	64%	67%

C. Pembahasan

Berdasarkan hasil uji pada bagian B yang telah dilakukan, model LSTM mampu mengklasifikasikan sentimen komentar masyarakat terhadap vaksin AstraZeneca dengan tingkat akurasi yang konsisten pada ketiga jenis distribusi data, yaitu data *original*, *oversampled*, dan SMOTE. Meskipun dilakukan upaya penyeimbangan data menggunakan teknik *oversampling* dan SMOTE untuk mengatasi ketidakseimbangan kelas, hasil akurasi yang diperoleh tidak mengalami peningkatan signifikan, yaitu tetap berada pada kisaran 65%. Bahkan, pada data hasil SMOTE, nilai *loss* justru mengalami sedikit kenaikan menjadi 67%, yang menunjukkan adanya potensi *overfitting* atau *noise* akibat penambahan data sintesis.

Fenomena ini dapat terjadi karena model LSTM, yang secara alami memiliki kemampuan menangkap pola sekuensial dalam data teks, sudah mampu mempelajari representasi penting dari data asli meskipun terdapat ketidakseimbangan kelas. Oleh karena itu, proses penyeimbangan data dengan menambahkan sampel sintesis melalui SMOTE tidak selalu memberikan keuntungan dalam konteks pemrosesan bahasa alami (NLP), terutama ketika dataset awal masih memiliki jumlah data yang relatif terbatas. Secara keseluruhan, hasil ini menegaskan bahwa efektivitas metode penyeimbangan data sangat bergantung pada karakteristik dataset serta jenis model yang digunakan.

V. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini adalah, bahwa model LSTM menunjukkan performa yang baik dalam mengklasifikasikan sentimen komentar masyarakat terhadap vaksin AstraZeneca, dengan akurasi yang stabil pada berbagai kondisi distribusi data yaitu 65% akurasi. Meskipun terdapat ketidakseimbangan kelas pada data awal, penerapan metode *oversampling* maupun SMOTE tidak memberikan peningkatan yang signifikan terhadap akurasi model. Hal ini mengindikasikan bahwa model LSTM sudah mampu menangkap pola semantik dari data asli tanpa perlu dilakukan modifikasi distribusi yang kompleks. Namun demikian, kenaikan nilai *loss* pada data hasil SMOTE menunjukkan bahwa penambahan data sintesis justru dapat menurunkan generalisasi model.

Sebagai saran, penelitian selanjutnya dapat mempertimbangkan beberapa hal untuk meningkatkan performa model. Pertama, memperbanyak jumlah dataset agar model dapat mempelajari representasi sentimen dengan lebih baik. Kedua, mengeksplorasi arsitektur lain seperti *Bidirectional LSTM* atau GRU yang berpotensi menangkap konteks dua arah dalam teks. Ketiga, melakukan *fine-tuning* pada *hyperparameter* seperti jumlah *neuron*, tingkat *dropout*, dan *learning rate* untuk mengoptimalkan hasil pelatihan. Selain itu, penggunaan teknik *embedding* yang lebih canggih seperti *Word2Vec*, *GloVe*, atau BERT juga disarankan agar model dapat memahami konteks linguistik dengan lebih mendalam dan menghasilkan klasifikasi sentimen yang lebih akurat.

DAFTAR RUJUKAN

- [1] N. A. M. Razali *et al.*, “Opinion mining for national security: techniques, domain applications, challenges and research opportunities,” *J. Big Data*, vol. 8, no. 1, p. 150, 2021, doi: 10.1186/s40537-021-00536-5.
- [2] L. Belcastro, R. Cantini, and F. Marozzo, “Knowledge Discovery from Large Amounts of Social Media Data,” *Applied Sciences*, vol. 12, no. 3, 2022. doi: 10.3390/app12031209.
- [3] R. Refianti, A. B. Mutiara, and R. A. Putra, “A Lexicon-Based Long Short-Term Memory (LSTM) Model for Sentiment Analysis to Classify Halodoc Application Reviews on Google Playstore,” *J. Appl. Data Sci. Vol 5, No 1 JANUARY 2024 DO - 10.47738/jads.v5i1.160*, Jan. 2024, [Online]. Available: <https://bright-journal.org/Journal/index.php/JADS/article/view/160>
- [4] G. Van Houdt, C. Mosquera, and G. Nápoles, “A review on the long short-term memory model,” *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5929–5955, 2020, doi: 10.1007/s10462-020-09838-1.
- [5] Akshay Kaushal, Anandadeep Mandal, Diksha Khanna, and Animesh Acharjee, “Analysis of the opinions of individuals on the COVID-19 vaccination on social media,” *Digit. Heal.*, vol. 9, p. 20552076231186250, Jan. 2023, doi: 10.1177/20552076231186246.
- [6] R. Ida, S. E. Kinasih, S. N. Febriyanti, R. Puspa, M. Saud, and M. S. A. Bakar, “COVID-19 vaccine rumour and resistance in Indonesia: Analysing vaccine hesitancy and cultural myths circulated on online platforms,” *Vacunas*, vol. 25, no. 3, pp. 313–322, 2024, doi: <https://doi.org/10.1016/j.vacun.2024.03.002>.

- [7] Y.-J. Choi, J. Lee, and S. Y. Paek, “Public Awareness and Sentiment toward COVID-19 Vaccination in South Korea: Findings from Big Data Analytics,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 16. 2022. doi: 10.3390/ijerph19169914.
- [8] K. L. Tan, C. P. Lee, and K. M. Lim, “A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research,” *Appl. Sci.*, vol. 13, no. 7, 2023, doi: 10.3390/app13074550.
- [9] W. Aljedaani, E. Saad, F. Rustam, I. de la Torre Díez, and I. Ashraf, “Role of Artificial Intelligence for Analysis of COVID-19 Vaccination-Related Tweets: Opportunities, Challenges, and Future Trends,” *Mathematics*, vol. 10, no. 17. 2022. doi: 10.3390/math10173199.
- [10] H. Wang *et al.*, “Patterns of diverse and changing sentiments towards COVID-19 vaccines: a sentiment analysis study integrating 11 million tweets and surveillance data across over 180 countries,” *J. Am. Med. Informatics Assoc.*, vol. 30, no. 5, pp. 923–931, May 2023, doi: 10.1093/jamia/ocad029.
- [11] V. B. Vaghela, Z. Y. Noorani, K. Patel, P. G. Patel, H. D. Rajput, and M. Shah, “Aspect Based Sentiment Analysis Using Self-Attention Based LSTM Model with Word Embedding,” *J. Comput. Sci.*, vol. 20, no. 10, 2024, doi: 10.3844/jcssp.2024.1195.1202.
- [12] Aditi Chauhan, Yashika Kukkar, Preeti Nagrath, Kirti Gupta, and Jude D Hemanth, “Integrating LSTM and NLP techniques for essay generation,” *Intell. Decis. Technol.*, vol. 18, no. 1, pp. 571–584, Feb. 2024, doi: 10.3233/IDT-230424.
- [13] Y. I. Alzoubi, A. E. Topcu, and A. E. Erkaya, “Machine Learning-Based Text Classification Comparison: Turkish Language Context,” *Applied Sciences*, vol. 13, no. 16. 2023. doi: 10.3390/app13169428.
- [14] H. Wang and L. Li, “Study on the Application of LSTM in Meteorological Element Prediction based on Computer Software,” *J. Phys. Conf. Ser.*, vol. 1744, no. 2, p. 22034, 2021, doi: 10.1088/1742-6596/1744/2/022034.
- [15] P. Anki, A. Bustamam, H. S. Al-Ash, and D. Sarwinda, “Intelligent Chatbot Adapted from Question and Answer System Using RNN-LSTM Model,” *J. Phys. Conf. Ser.*, vol. 1844, no. 1, p. 12001, 2021, doi: 10.1088/1742-6596/1844/1/012001.
- [16] F. Cascini *et al.*, “Social media and attitudes towards a COVID-19 vaccination: A systematic review of the literature,” *eClinicalMedicine*, vol. 48, Jun. 2022, doi: 10.1016/j.eclinm.2022.101454.
- [17] A. A. Jamali, C. Berger, and R. J. Spiteri, “Momentary Depressive Feeling Detection Using X (Formerly Twitter) Data: Contextual Language Approach,” *JMIR AI*, vol. 2, p. e49531, 2023, doi: 10.2196/49531.
- [18] O. Narushynska, V. Teslyuk, A. Doroshenko, and M. Arzubov, “Data Sorting Influence on Short Text Manual Labeling Quality for Hierarchical Classification,” *Big Data and Cognitive Computing*, vol. 8, no. 4. 2024. doi: 10.3390/bdcc8040041.
- [19] M. Francisco and J. L. Castro, “A Methodology to Quickly Perform Opinion Mining and Build Supervised Datasets Using Social Networks Mechanics,” *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 9, pp. 9797–9808, 2023, doi: 10.1109/TKDE.2023.3250822.
- [20] A. B. Amendolara, D. Sant, H. G. Rotstein, and E. Fortune, “LSTM-based recurrent neural network provides effective short term flu forecasting,” *BMC Public Health*, vol. 23, no. 1, p. 1788, 2023, doi: 10.1186/s12889-023-16720-6.