Rancang Bangun Aplikasi Sentimen Pendapat Masyarakat Indonesia Terhadap Vaksin COVID-19 Dengan Menggunakan Algoritma Long Short-Term Memory Berbasis Web

Amanda Amalia[1], Dewi Rahmawati[2], Mohammad Sholik[3]
[1]Program Studi Rekayasa Perangkat Lunak, Fakultas Teknologi Informasi dan Bisnis, Institut Teknologi Telkom Surabaya
[2]Program Studi Rekayasa Perangkat Lunak, Fakultas Teknologi Informasi dan Bisnis, Institut Teknologi Telkom Surabaya
[3]Program Studi Rekayasa Perangkat Lunak, Fakultas Teknologi Informasi dan Bisnis, Institut Teknologi Telkom Surabaya
Email: [1]amanda@student.ittelkom-sby.ac.id, [2]dewirahmawati@ittelkom-sby.ac.id, [3]mohammadsholik@ittelkom-sby.ac.id

Abstract. *Active users of Twitter social media in Indonesia are considered as one of the countries with the most growth of daily active users of Twitter social media based on Twitter's 2019 financial report which is used to communicate various information with other users. Twitter is used as a source of health-related information, given the large amount of information, news, and opinions shared by both citizens and official sources. In this paper, we use sentiment analysis to explore a large number of tweets in Indonesia. We use 4,000 tweets in Indonesian. The purpose of this study is to see the tendency of a person's opinion or opinion on a topic with apply the word embedding method and Long Short-Term Memory (LSTM) algorithm for analysis of sentiment regarding the COVID-19 vaccine originating from social media Twitter. The experiment using LSTM-Word Embedding resulted in an accuracy of 74,46% with the final result obtained being 425 data with positive sentiments and 507 data with negative sentiments. This is important to do to understand public reactions, disseminate information on social media, and make the right strategy that the Indonesian government should follow.*
Keywords: *Sentiment Analysis, COVID-19 vaccine, Twitter, Long Short-Term Memory (LSTM)*

Abstrak. *Pengguna aktif media sosial Twitter di Indonesia dinilai sebagai salah satu negara dengan pertumbuhan pengguna aktif harian media sosial Twitter paling banyak berdasarkan laporan finansial Twitter tahun 2019 yang digunakan untuk berkomunikasi berbagai informasi dengan pengguna lain. Twitter digunakan sebagai sumber informasi yang berhubungan dengan kesehatan, mengingat banyaknya informasi, berita, dan juga opini yang disebarluaskan oleh warga negara dan juga sumber resmi. Dalam makalah ini, kami menggunakan analisis sentimen untuk mengeksplorasi sejumlah besar tweet di Indonesia. Kami menggunakan 4.000 tweet dalam bahasa Indonesia. Tujuan dari penelitian ini yaitu untuk melihat kecenderungan pendapat atau opini seseorang terhadap sebuah topik dengan menerapkan metode word embedding dan algoritma Long Short-Term Memory (LSTM) untuk analisis terhadap sentimen mengenai vaksin COVID-19 yang berasal dari media sosial Twitter. Percobaan menggunakan LSTM-Word Embedding menghasilkan akurasi sebesar 74,46% dengan hasil akhir yang diperoleh yaitu sebanyak 425 data bersentimen positif dan 507 data bersentimen negatif. Hal ini penting dilakukan untuk memahami reaksi publik, penyebaran informasi pada media sosial, dan membuat strategi yang tepat yang selanjutnya harus dilakukan oleh pemerintah Indonesia.*
Kata Kunci: *Analisis Sentimen, vaksin COVID-19, Twitter, Long Short-Term Memory (LSTM)*

1. Pendahuluan

Sebagian besar masyarakat Indonesia pengguna layanan digital mengakses informasi mengenai COVID-19 melalui media sosial (Tiana dan Amalia, 2021). Sebanyak 81,5% diantaranya masih bersinggungan dengan berbagai bentuk postingan-postingan yang memuat teori konspirasi mengenai vaksin COVID-19. Media sosial dan kebebasan berpendapat membuat sebuah informasi yang belum tentu benar lebih cepat menyebar di lingkungan masyarakat. Hal tersebutlah yang menimbulkan pro dan kontra di antara masyarakat (Layla, 2021). Mayoritas masyarakat Indonesia menyakini bahwa vaksin COVID-19 dibuat untuk mendapatkan keuntungan pihak farmasi ataupun untuk memasukkan microchip dalam tubuh manusia (Tiana

dan Amalia, 2021). Kemudian pada akhir maret 2022, Presiden Joko Widodo menetapkan wajib untuk mendapatkan vaksin booster untuk dapat melakukan mudik lebaran membuat polemik ditengah masyarakat. Walaupun hal ini merupakan stategi pemerintah agar masyarakat Indonesia segera mendapatkan vaksin COVID-19 secara lengkap, banyak masyarakat yang merasa terbebani dengan keputusan yang dibuat oleh pemerintah (Uswah, 2022). Masyarakat beranggapan bahwa vaksin booster sebagai syarat mudik ini bersifat tidak adil terhadap masyarakat, hal ini dikarenakan banyak kegiatan yang menimbulkan kerumunan seperti acara konser, liburan tahun baru, dan pagelaran MotoGP yang diadakan di Mandalika, dan tidak ada persyaratan untuk vaksinasi kedua. Apalagi saat ini, penerima vaksin juga belum merata dan masih banyak masyarakat yang baru menyelesaikan vaksin kedua. Agar kehadiran vaksin COVID-19 dapat diterima oleh seluruh masyarakat, maka Pemerintah perlu memperkenalkan dan penyuluhan secara intens dan menyeluruh mengenai pentingnya vaksin COVID-19 bagi kesehatan (Amin dan Azhari, 2019). Oleh karena itu, perlu dilakukan analisis sentimen pendapat masyarakat dan mengklasifikasikannya kedalam kelas positif, negatif, atau netral.

Untuk melihat opini masyarakat terhadap vaksin COVID-19 dapat dilihat dari berbagai media salah satunya yaitu media sosial Twitter (Kurniawandanarissy dan Sutan, 2021). Isi tweet ini dapat dijadikan sebagai sumber data pendapat dan sentimen masyarakat terhadap vaksin COVID-19 yang dapat digunakan untuk analisis dan bahan evaluasi bagi pemerintah Indonesia dalam memberikan edukasi kepada masyarakat mengenai program vaksin COVID-19..

Pada penelitian yang berjudul "LSTM (Long Short Term Memory) for Sentiment COVID-19 Vaccine Classification on Twitter", oleh Miftahul Ihsan, Benny Sukma Negara, dan Surya Agustian mengenai pendapat mengenai vaksin COVID-19 pada media sosial Twitter dengan menerapkan metode *Long Short-Term Memory* (LSTM) dengan model *word embeddings* Word2Vec (Ihsan, 2022) untuk menganalisis sentimen tweet yang berkaitan dengan kegiatan vaksinasi COVID-19. Data yang diambil dengan menggunakan bahasa pemrograman python dan Twitter API, setelah data yang ganda (tweet yang sama) dihapus, data akhir yang diperoleh yaitu sebanyak 13.115 tweet dari tanggal 23 Maret sampai 20 April 2021. Dari penelitian yang dilakukan, didapatkan hasil akurasi sebesar 66% untuk penggunaan metode LSTM, 61% untuk penggunaan Naïve Bayes, dan 65% untuk penggunaan Support Vector Machine (SVM).

Penelitian yang dilakukan oleh Nadia Ristya Dewi, Eva Yulia Puspanigrum, dan Hendra Maulana dengan judul "Analisis Sentimen Tweet Vaksinasi COVID-19 Menggunakan RNN Dengan Metode TF-IDF Dan Word2Vec" (Nadia, dkk., 2022). Penelitian ini membahas mengenai pendapat masyarakat Indonesia mengenai vaksin COVID-19 dengan menggunakan data tweet dari media sosial Twitter. Pengambilan data dilakukan dengan 2 metode yaitu metode crawling data dengan hashtag #vaksin dan #vaksinasi dari tanggal 5 Desember 2021 sampai dengan 20 Desember 2021 dengan memperoleh sebanyak 3000 data dan data kedua yaitu data tweet berbahasa Indonesia yang diambil dari *website* kaggle sebanyak 3000 data dari rentang tanggal 10 Januari 2021 sampai dengan 11 Februari 2021. Hasil akhir penelitian ini membuktikan bahwa penggunaan metode *word embedding* Word2Vec lebih baik dibandingkan dengan TF-IDF. Hal ini dapat dilihat dari nilai akurasi dengan metode *word embedding* Word2Vec dengan RNN sebesar 53%, sedangkan perolehan nilai akurasi dengan metode RNN dengan TF-IDF menghasilkan nilai akurasi sebesar 51%. Berdasarkan latar belakang tersebut, diusulkan metode *Long Short-Term Memory* (LSTM) dengan *word embedding* untuk menganalisis sentimen masyarakat Indonesia terhadap vaksin COVID-19. Penelitian menggunakan pencarian pada Twitter dengan kata kunci "vaksin" dan "vaksin covid". Data yang diperoleh dari proses *crawling* tersebut akan diklasifikasikan menjadi kelas positif, negatif, atau netral. Algoritma *machine learning Long Short-Term Memory* (LSTM) yang digunakan pada penelitian ini diharapkan mampu memberikan hasil analisis data yang sesuai dengan akurasi yang terbaik dan lebih baik dari penelitian sebelumnya.

2. Tinjauan Pustaka

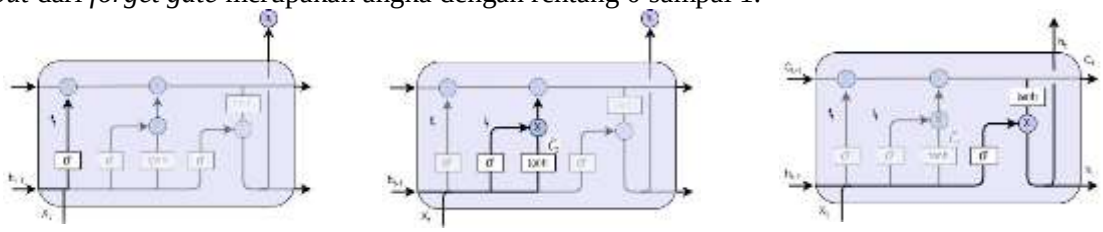
2.1. Analisis Sentimen

Analisis sentimen adalah sebuah studi yang digunakan untuk menganalisis/mempelajari suatu objek, dimana objek tersebut berisi opini positif, negatif, atau netral (Amin dan Azhari, 2019). Opini publik sering dijumpai di media sosial contohnya twitter, tempat orang-orang menuliskan opini mereka terhadap sebuah objek, dan isi dari opini tersebut ada yang bersifat aktual maupun tidak dan bersifat positif ataupun negatif terhadap suatu objek (Munasatya dan Novianto, 2020). Menurut KBBI

(Kamus Besar Bahasa Indonesia) sentimen positif merupakan reaksi atau pendapat yang setuju dan sependapat sehingga meningkatkan nilai sesuatu atau seseorang. Sentimen negatif merupakan reaksi atau pendapat yang tidak setuju atau menentang yang dapat menurunkan nilai seseorang atau sesuatu. Sedangkan sentimen netral merupakan sentimen yang tidak berpihak kepada apapun, baik mendukung ataupun tidak mendukung (KBBI Daring, 2022).

2.2. Long-Short Term Memory (LSTM)

Long Short-Term Memory (LSTM) merupakan varian dari *Recurrent Neural Network* (RNN) yang bertujuan untuk menghindari masalah RNN mengenai masalah *vanishing gradien* dengan mengganti *node* tersembunyi RNN dengan sel LSTM yang telah dirancang untuk menyimpan informasi sebelumnya. LSTM terdiri dari tiga komponen *gates* atau gerbang yang berguna untuk mengontrol aliran informasi yaitu *input gate*, *output gate*, dan *forget gete* yang melingdungi dan mengontrol cell state (Amin dan Azhari, 2019). Dalam LSTM, *forget gate* merupakan gerbang pertama yang dioperasikan. *Forget gate* berguna untuk menentukan informasi yang harus diperhatahan atau dibuang dari cell state. Gerbang ini menerima dua *input* yaitu dari h_{t-1} dan X_t. Dimana h_{t-1} merupakan keluaran pada time step $t-1$ dari proses LSTM sedangkan X_t merupakan *input* pada *time step* t. *Output* dari *forget gate* merupakan angka dengan rentang 0 sampai 1.



(a) (b) (c)

Gambar 1. (a) *Forget Gate,* (b) *Input Gate,* dan (3)*Output Gate*

Perhitungan nilai *forget gate* dan *input gate* dilakukan dengan persamaan (1) (2), dan (3).

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (2)$$

$$\widetilde{c_t} = tanh(W_c.[h_{t-1}, w_t] + b_c) \quad (3)$$

Dimana,

f_t = *forget gate*

σ = *sigmoid function*

W_f = bobot pada *forget gate*

h_{t-1} = hasil dari *output* pada *time step* $t-1$

x_t = *input* pada *time step* t

b_f = bias pada *forget gate*

i_t = *input gate*

$tanh$ = *tanh function*

W_c = bobot pada operasi *cell state* yang baru

W_i = bobot pada *input gate*

x_t = *input* pada *time step* t

b_i = bias pada *input gate*

b_c = bias pada *cell state* baru

$\widetilde{c_t}$ = kandidat nilai *cell state* yang baru yang akan ditambahkan ke C_{t-1}

Perhitungan *output gate* ditunjukkan dengan persamaan (4), (5), dan (6).

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \quad (4)$$

$$O_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = O_t * \tanh(C_t) \quad (6)$$

dimana,

O_t = *Output gate*

h_{t-1} = Hasil *output* pada *time step* $t-1$

C_t = *Cell state* pada *time step t*
C_{t-1} = *Cell state* pada *time step* $t-1$
h_t = output pada *time step* t
$\widetilde{C_t}$ = Nilai *cell state* baru hasil dari operasi perhitungan pada *input gate*
x_t = *Input pada time step t*
b_o = Bias pada *output gate*
W_o = Bobot pada *input gate*
σ = *sigmoid function*

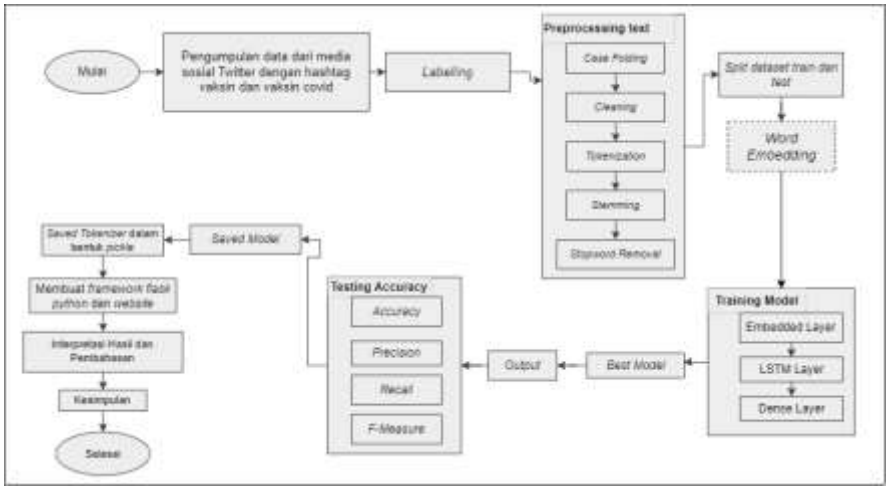
2.3 Text Mining

Text mining merupakan proses yang mendefinikasikan pengetahuan informasi dimana pengguna dapat berinteraksi dengan dokumen dari waktu ke waktu menggunakan alat dan suatu metode tertentu untuk melakukan analisis (Amin dan Azhari, 2019). Agar mendapatkan hasil yang akurat, pada tahapan pre-processing dalam text mining ini sangat penting dikarenakan mengekstraksi dan identifikasi fitur-fitur yang mewakili setiap kata pada sebuah dokumen. Tahapan umum pada *pre-processing* yaitu *case folding, tokenizing, filtering,* dan *stopword removal* (Amin dan Azhari, 2019). Text mining memiliki tujuan yaitu untuk mendapat informasi dari data yang sedang diolah dan memiliki tugas yaitu pengelompokan teks (*text clustering*) dan pengkategorisasian teks (*text categorization*).

3. Metode Penelitian

3.1.Pengumpulan Data

Data yang dikumpulkan yakni data *tweet* dari tanggal 20 Maret 2022 – 3 April 2022 dengan cara crawling menggunakan bahasa pemrograman *python* dan Twitter API. Data yang dikumpulkan yaitu data tweet berbahasa Indonesia. Kata kunci atau hashtag yang digunakan yaitu "vaksin" dan "vaksin COVID-19", yang mengarah ke topik kegiatan vaksinasi COVID-19. Dari pengumpulan data yang dilakukan dan telah melalui penghapusan data yang ganda (*duplicate tweet*) dihapus, menghasilkan sebanyak 3072 *tweet* yang sebelum penghapusan sebanyak 4000 *tweet*. Kemudian dilakukan proses pemberian label terhadap *tweet* kedalam 3 kelas yang telah ditentukan yaitu sentimen positif, negatif atau netral.



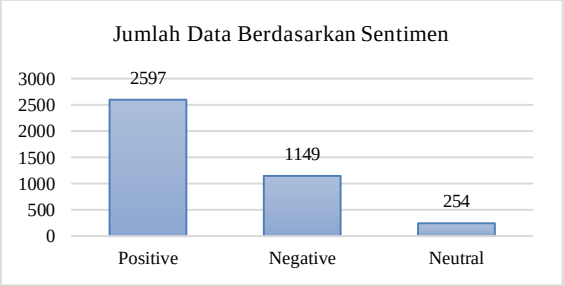
Gambar 2. Metodologi Penelitian

Tabel 1. Pelabelan Data Tweet

Tweet	*Sentimen*
@theycalledmemel @DisharahRahat1 Masih percaya vaksin. Silakan. Standby dos 4,5,6 dan seterusnya. Orang yg dah vaksin pun masih kena lagi covid kan?	Negatif
Banyak yg protes klo vaksin 3 booster jadi syarat mudik lebaran	Positif
Padahal niat pemerintahnya baik biar pada di vaksin biar terhindar dr covid pas mudik, pada ngeluh ae padahal vaksin juga gratis kenapa bilang pemerintah bisnis, yg ada keluar biaya bestie.	

Tweet	***Sentimen***
@hnurwahid Tingkat vaksin ke 2 mencapai 78% artinya kekebalan masy. thd covid sdh terbentuk. Buktinya tkt Booster hanya 8% tp positif covid sdh melandai hanya 5000 org. Jadi pemerintah tdk perlu memaksakan diri buat persyaratan booster bagi pemudik lebaran hanya utk cuci gudang stock vaksin.	Negatif
Kalo lagi flu boleh vaksin enggak ya?	Netral

Berikut adalah hasil dari pelabelan data.



Gambar 3. Metodologi Penelitian

3.2. *Text Preprocessing*

Pada penelitian ini dilakukan beberapa tahapan *preprocessing* yang meliputi *case folding, cleaning, tokenization, stemming, dan stopword removal*.

Tabel 2. Tahap *Case Folding*

Sebelum	***Sesudah***
Nonton â • ¦@MotoGPâ • © cuma 2 kali vaksin. Bahkan ada yg lewat begitu saja tak perlu discan barkod di pedulilindungi https://t.co/y2o90QkglG	nonton â • ¦@motogpâ • © cuma 2 kali vaksin. bahkan ada yg lewat begitu saja tak perlu discan barkod di pedulilindungi https://t.co/y2o90qkglg

Pada tahap case folding teks yang dihapus seperti tanda baca, *link*, hashtags, mentions, angka, *retweet*, dan *punctuation*.

Tabel 3. Tahap *Cleaning*

Sebelum	***Sesudah***
nonton â • ¦@motogpâ • © cuma 2 kali vaksin. bahkan ada yg lewat begitu saja tak perlu discan barkod di pedulilindungi https://t.co/y2o90qkglg	nonton cuma kali vaksin bahkan ada yg lewat begitu saja tak perlu discan barkod di pedulilindungi

Selanjutnya tahap *tokenization* atau *tokenizing* yaitu proses mengubah kalimat menjadi deretan potongan kata tunggal.

Tabel 4. Tahap *Tokenization*

Sebelum	***Sesudah***
nonton cuma kali vaksin bahkan ada yg lewat begitu saja tak perlu discan barkod di pedulilindungi	['nonton', 'cuma', 'kali', 'vaksin', 'bahkan', 'ada', 'yg', 'lewat', 'begitu', 'saja', 'tak', 'perlu', 'discan', 'barkod', 'di', 'pedulilindungi']

Tahap *stemming* yaitu proses untuk mencari kata dasar dari sebuah kata dengan menghapus semua kata imbuhan baik awalan, sisipan, maupun akhiran.

Tabel 5. Tahap *Stemming*

Sebelum	***Sesudah***
['nonton', 'cuma', 'kali', 'vaksin', 'bahkan', 'ada', 'yg', 'lewat', 'begitu', 'saja', 'tak', 'perlu', 'discan', 'barkod', 'di', 'pedulilindungi']	nonton kali vaksin yg discan barkod pedulilindungi

 ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

Kamus *stopword* yang digunakan pada penelitian ini ada dua yaitu kamus *stopword* yang ada pada pustaka *python* dan membuat list kamus *stopword* secara manual untuk menghilangkan kata yang tidak ada dikamus *stopword*.

Tabel 6. Tahap *Stopword Removal*

Sebelum	Sesudah
nonton kali vaksin yg discan barkod pedulilindungi	discan barkod pedulilindungi

3.3. Percancangan Model LSTM

Word embedding merupakan sebuah pendekatan yang berguna untuk merepresentasikan *vector* kata atau *array* dari suatu bilangan *riil* dan perkembangan dari komputasi permodelan kata-kata yang sederhana seperti perhitungan menggunakan jumlah dan frekuensi kemunculan sebuah kata didalam suatu dokumen (Malik, dkk., 2020). Pada penelitian ini, proses *word embedding* melatih *word embedding* menggunakan *framework* yang ada pada *python* yang bernama *keras*.

3.3.1. Pemberian *Indeks* Kalimat

Pada tahap ini, digunakan *parameter num_words* yang berfungsi untuk mengatur ukuran *vocabulary* yang akan digunakan. Pada penelitian in, *parameter num_words* yang digunakan yaitu sejumlah 4500 kata berdasarkan jumlah data yaitu sebanyak 4500 data.

Tabel 7. *Indeks* Setiap Kata

Indeks Setiap Kata
{'-': 1, 'booster': 2, 'mudik': 3, 'vaksinasi': 4, 'dosi': 5, 'vaksinboost': 6, 'proke': 7, 'orang': 8, 'syarat': 9, 'masyarakat': 10, 'pemerintah': 11, 'anak': 12, 'lawancovid': 13, 'kesehatan': 14, 'indonesia': 15, 'lebaran': 16, 'pandemi': 17, 'wajib': 18, 'viru': 19, 'warga': 20, 'lengkap': 21, 'x': 22, 'masker': 23, 'puasa': 24, 'newnorm': 25, 'sehat': 26, 'melaksanakan': 27, 'pcr': 28, 'sakit': 29, 'kak': 30, 'ketiga': 31, 'antigen': 32, 'kegiatan': 33, 'ramadhan': 34, 'negara': 35, 'negatif': 36, 'abi': 37, 'kota': 38, 'maret': 39, 'te': 40, 'salah': 41, 'terawan': 42, 'nusantara': 43, dan seterusnya ... 'balapan': 9479, 'pikirmu': 9480, 'rank': 9481, 'akang': 9482, 'teteh': 9483, 'landasan': 9484, 'chao': 9485, 'shg': 9486, 'sembarang': 9487, 'kesengaraan': 9488, 'sirna': 9489, 'sepuluh': 9490, 'davo': 9491, 'big': 9492, 'pharma': 9493, 'ngajakin': 9494}

3.3.2. *Pad Sequences*

Paramater yang digunakan selanjutnya adalan *input_length* yang berguna untuk mengatur urutan vektor. Jika terdapat kalimat yang pendek, maka makin banyak pula angka 0 pada *feature vector* kalimat tersebut. Karena setiap kalimat memiliki jumlah data yang berbeda-beda, maka untuk mengisi panjang vektor agar semua panjang vektor kalimat sama digunakan *pad_sequence* yang ada pada *keras.preprocessing.sequence* untuk mengisi urutan kata yang kosong dengan angka 0 secara otomatis. Pada penelitian ini jumlah *max_len* yang digunakan adalah 100.

Tabel 8. *Vector Pad Sequences*

Vector Pad_Sequences
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 49, 1786, 3751, 3752, 837, 967, 572, 1786, 631, 108, 572, 1413, 108 572, 258, 968, 3753, 3754], dan seterusnya ... [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 623, 396, 1202, 3661, 520, 1600, 1,1,1, 1405, 520, 3661, 1])

3.3.3. Lapisan *Word Embedding*

Lapisan *word embedding* atau *layer embedding* merupakan tahap masuk lapisan pertama untuk model yang akan dibangun. Lapisan ini akan menggunakan *parameter input_dim* dan *output_dim.* Pada penelitian ini *input_dim* yang digunakan yaitu sejumlah 4500, sedangkan untuk

parameter output_dim yang digunakan yaitu sejumlah 100. Parameter vektor *embedding* dengan jumlah *output_dim* sebanyak 100 dengan *max_len* berjumlah 30. Sehingga didapatkan yaitu *variable* input x_t berupa vektor *word embedding tensor* menjadi 3D ukuran (4000, 30, 100) yang selanjutnya akan digunakan oleh LSTM untuk melakukan *training* menggunakan data sentimen.

3.4. Proses *Training* LSTM

Setelah menerima *output* dari proses *embedding*, proses selanjutnya yaitu melakukan pelatihan model LSTM. Model pelatihan dalam penelitian terdiri dari empat lapisan yaitu lapisan *Embedding*, lapisan LSTM, satu lapisan *dense* dengan berbagai fitur *input*. Secara umum, arsitektur jaringan dibentuk agar dapat menghasilkan akurasi yang optimal.

4. Hasil dan Pembahasan

4.1. Hasil Analisis Dengan *Long Short-Term Memory* (LSTM)

Tabel 9. Hasil Klasifikasi Model LSTM

	0	*1*	*2*
0	0.012558	0.012248	0.975194
1	0.72847	0.078789	0.192741
2	0.527535	0.109965	0.362501
3	0.052292	0.033571	0.914137
4	0.007938	0.010395	0.981667
5	0.016458	0.01543	0.968111
6	0.324525	0.089002	0.586473
7	0.016533	0.013302	0.970165
8	0.044941	0.031235	0.923824
9	0.618186	0.112272	0.269542
10	0.009775	0.009767	0.980459
11	0.026742	0.025113	0.948144
12	0.539731	0.126679	0.33359
13	0.341744	0.107292	0.550964

Hasil akhir analisis ditunjukkan pada tabel 10 pada kolom result. Kolom result didapatkan dari hasil perumusan yang dibuat pada Microsoft Excel dengan menggunakan rumus 7 dan data pada kolom 0, kolom 1, dan kolom 2.

$$= IF(B2 > C2, IF(B2 > D2, \$B\$1, \$D\$1), IF(C2 > D2, \$C\$1, \$D\$1)) \quad (7)$$

Dari formulas 7, dimana jika nilai yang dicari lebih besar dari kolom yang digunakan sebagai acuan yaitu semua kolom baik kolom 0, kolom 1, dan kolom 2 maka hasil yang diperoleh apabila hasil tersebut terdapat pada kolom 0 maka dinyatakan sebagai sentimen negatif, apabila hasil tersebut terdapat pada kolom 1 dinyatakan sebagai sentimen netral, dan apabila hasil tersebut terdapat pada kolom 2 dinyatakan sebagai sentimen positif.

Tabel 10. Hasil Klasifikasi Beserta Dengan Sentimen

	0	*1*	*2*	*result*	*Sentimen*
0	0.012558	0.012248	0.975194	2	Positif
1	0.72847	0.078789	0.192741	0	Negatif
2	0.527535	0.109965	0.362501	0	Negatif
3	0.052292	0.033571	0.914137	2	Positif
4	0.007938	0.010395	0.981667	2	Positif
5	0.016458	0.01543	0.968111	2	Positif
6	0.324525	0.089002	0.586473	2	Positif
	0	1	2	result	Sentimen

ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

7	0.016533	0.013302	0.970165	2	Positif
8	0.044941	0.031235	0.923824	2	Positif
9	0.618186	0.112272	0.269542	0	Negatif
10	0.009775	0.009767	0.980459	2	Positif
11	0.026742	0.025113	0.948144	2	Positif
12	0.539731	0.126679	0.33359	0	Negatif

Tabel 11. Jumlah *Epoch* dan Hasil Percobaan

Nilai Epoch	*Akurasi*	*Precision*	*Recall*	*F-Measure*
25	71.65%	86%	70%	78%
24	70.69%	76%	80%	78%
23	72.38%	83%	75%	79%
22	71.41%	81%	78%	79%
21	72.22%	82%	77%	79%
15	70.53%	82%	74%	77%
10	66.02%	68%	86%	76%
5	57.72%	58%	100%	73%

Hasil akurasi yang dihasilkan oleh *scenario* algoritma LSTM dengan jumlah *epoch* yaitu 23 sebesar 72.38%, dengan nilai *precision* sebesar 83%, *recall* sebesar 75%, dan *f-measure* sebesar 79%. Percobaan dilakukan dengan *splitting data training* dan data testing yang dilakukan dengan perbandingan 70:30 dengan nilai akurasi sebesar 74,46%, *precision* sebesar 79%, *recall* sebesar 83%, dan nilai *f-measure* sebesar 81%. Berikut *classification report* yang menampilkan nilai dari *accuracy, presisi, recall, dan f1-measure* kedua *scenario* yang menghasilkan akurasi tertinggi yaitu percobaan LSTM-*Word Embedding* pada Tabel 12.

Tabel 12. *Classification Report* LSTM-*Word Embedding*

		precision	*recall*	*f1-measure*	*support*
	0	0.64	0.84	0.73	323
	1	0	0	0	66
	2	0.83	0.78	0.80	543
accuracy				0.74	932
macro avg		0.49	0.54	0.51	932
weighted avg		0.71	0.74	0.73	932

Tabel 13. Hasil Klasifikasi dengan LSTM

	LSTM
Positif	425
Negatif	507
Neutral	-
Total	932 (data testing)

Pada penelitian ini akurasi yang dimaksud adalah ukuran kemampuan model klasifikasi untuk melakukan prediksi terhadap kelas benar (*TruePositif*, *TrueNegatif*, dan *TrueNetral*) dan pada model LSTM menghasilkan nilai akurasi sebesar 74%. Presisi menentukan kelas benar positif yang diprediksi benar dan pada model LSTM menghasilkan nilai 64%, 0%, dan 83% untuk setiap kelas negatif, netral, dan positif. *Recall* menyajikan prediksi yang aktualnya benar positif diukur untuk semua prediksi positif, pada model LSTM menghasilkan nilai 84%, 0%, dan 78% untuk setiap kelas negatif, netral, dan positif. Sedangkan *f-masure* merupakan rata-rata dari nilai presisi dan *recall* yang digunakan untuk mengukur performa program klasifikasi jika data *false* tidak saling mendekat.

4.2. Implementasi Website

Implementasi website yang dibuat yaitu menggunakan algortima *flask python* dan bahasa pemrograman PHP dengan cara menyimpan model LSTM yang telah dibuat.

Gambar 2. Halaman Utama Website

Gambar 3. Halaman Antarmuka Website

Gambar 4. Halaman Klasifikasi Kalimat

Gambar 5. Halaman Tampilan Hasil Klasifikasi

4.3 Rencana Pengujian

Pada penelitian ini dilakukan test case pada website dengan menggunakan metode *blackbox testing* dengan teknik *Equivalence Partitioning*. Tahap pertama pengujian perangkat lunak pada penilitian ini yaitu dengan membuat *Test Case website* dan hasil yang diharapkan yang diberikan oleh *website* yang dapat dilihat pada tabel 14.

Tabel 14. Rencana Pengujian

ID	*Test Case*	*Hasil Pengujian*	*Hasil yang Diharapkan*
A.1-A01	*User* menekan tombol panah arah selanjutnya pada *website*.	Sistem menerima permintaan user dan ditampilkan halaman selanjutnya.	Sistem menerima permintaan *user* dan menampilkan halaman selanjutnya.
A.1-A02	*User* tidak menekan tombol panah arah selanjutnya pada *website*.	Sistem menampilkan halaman website yang sama.	Sistem menampilkan halaman *website* yang sama.
A.2-B01	*User* memasukkan kalimat yang akan diklasifikasikan sentimen dari kalimat tersebut, kemudian menekan tombol klasifikasi.	Sistem menerima permintaan user dan ditampilkan hasil klasifikasi sentimen pada halaman selanjutnya.	Sistem menerima permintaan *user* dan menampilkan hasil klasifikasi sentimen pada halaman selanjutnya.
A.2-B02	*User* mengosongkan input kalimat dan menekan tombol klasifikasi.	Sistem menolak permintaan user dan tidak menampilkan hasil klasifikasi apapun.	Sistem menolak permintaan *user* dan tidak menampilkan hasil klasifikasi apapun.

Tabel 15. Kesimpulan Hasil Pengujian Website

No.	*ID*	*Kesimpulan*
1.	A.1-A01	Berhasil
2.	A.1-A02	Berhasil
3.	A.2-B01	Berhasil
4.	A.2-B02	Berhasil

4. Kesimpulan

Gambaran umum data tweet pada media sosial Twitter yang diambil berdasarkan tanggal 20 Maret 2022 sampai dengan tanggal 3 April 2022 setelah melalui proses labelling, didapatkan jumlah sentimen terbanyak yaitu positif sebanyak 2597 tweet, sentimen negatif sebanyak 1149 tweet, dan sentimen netral sebanyak 254 tweet. Jadi sebanyak 64,925% dari 4000 tweet merupakan sentimen positif, 28,725% merupakan sentimen negatif, dan 6,35% merupakan sentimen netral. Berdasarkan hal

tersebut, dapat disimpulkan bahwa 64,925% orang di Indonesia cenderung bersifat mendukung dan menyetujui adanya program vaksinasi COVID-19 yang diwajibkan oleh pemerintah. Sedangkan sebanyak 28,725% dari 4000 tweet bersifat menentang adanya program vaksinasi, dan sisanya yaitu sebanyak 6,35% dari 4000 tweet yang diambil bersifat netral tidak memihak pihak yang mendukung ataupun menentang adanya program vaksinasi COVID-19. Dalam makalah ini, kami memiliki beberapa percobaan dengan menggunakan algoritma LSTM dengan melakukan hyperparameter tunning pada bagian epoch yang sangat cocok untuk menghasilkan hasil yang baik dalam teknik klasifikasi dengan akurasi sebesar 74%. Dari hasil percobaan yang telah dilakukan, dapat disimpulkan bahwa algoritma LSTM memiliki performa yang baik untuk dataset tweet mengenai vaksin COVID-19 pada media sosial Twitter.

Daftar Pustaka

A. K. Ingason, S. Helgadóttir, H. Loftsson, and E. Rögnvaldsson, "A mixed method lemmatization algorithm using a Hierarchy of Linguistic Identities (HOLI)," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 5221 LNAI, pp. 205–216, 2008, doi: 10.1007/978-3-540-85287-2_20.

A. Susilo, C. M. Rumende, C. W. Pitoyo, W. D. Santoso, M. Yulianti, Herikurniawan, R. Sinto, G. Singh, L. Nainggolan, E. J. Nelwan, L. K. Chen, A. Windhani, E. Wijaya, B. Wicaksana, M. Maksum, F. Annisa, C. O. Jasirwan, and E. Yunihastiti, "Coronavirus Disease 2019: Tinjauan Literatur Terkini," Jurnal Penyakit Dalam Indonesia, vol. 7, no. 1, p. 45, 2020, doi: 10.7454/jpdi.v7i1.415.

C. K. N. Paputungan and A. Jacobus, "Sentiment Analysis of Social Media Users Using Long-Short Term Memory Method," Jurnal Teknik Elektro dan Komputer, vol. 10, no. 2, pp. 99–106, 2021.

D. Kurniawandanarissy and J. Sutan, "Penggunaan Sosial Media Dalam Menyebarkan Program Vaksinasi Covid-19 Di Indonesia," Kebijakan Publik, vol. 12, no. 1, pp. 27–34, 2021.

Dr. G. S. N. Murthy, Shanmukha Rao Allu, Bhargavi Andhavarapu, and Mounika Bagadi, Mounika Belusonti, "Text based Sentiment Analysis using LSTM," International Journal of Engineering Research and, vol. V9, no. 05, pp. 299–303, 2020, doi: 10.17577/ijertv9is050290.

G. Wiro Sasmito, "Penerapan Metode Waterfall Pada Desain Sistem Informasi Geografis Industri Kabupaten Tegal," Jurnal Informatika:Jurnal Pengembangan IT (JPIT), vol. 2, no. 1, pp. 6–12, 2017.

K. Malik, M. Jasri, and A. S. Mashuri, "Information Retrieval (Ir) Pencarian Ide Pokok Dalam Teks Artikel Olahraga Berbahasa Inggris Menggunakan Metode Mmr (Maximum …," COREAI: Jurnal Kecerdasan …, vol. 1, no. 1, pp. 18–24, 2020, [Online]. Available: https://ejournal.unuja.ac.id/index.php/core/article/view/1641.

K. Relan, Building REST APIs with Flask. 2019. doi: 10.1007/978-1-4842-5022-8.

M. A. Nurrohmat and A. SN, "Sentiment Analysis of Novel Review Using Long Short-Term Memory Method," IJCCS (Indonesian Journal of Computing and Cybernetics Systems), vol. 13, no. 3, p. 209, 2019, doi: 10.22146/ijccs.41236.

N. Buslim and R. P. Iswara, "Pengembangan Algoritma Unsupervised Learning Technique Pada Big Data Analysis di Media Sosial sebagai media promosi Online Bagi Masyarakat," Jurnal Teknik Informatika, vol. 12, no. 1, pp. 79–96, 2019, doi: 10.15408/jti.v12i1.11342.

N. Mohamed Ali, M. M. A. el Hamid, and A. Youssif, "Sentiment Analysis for Movies Reviews Dataset Using Deep Learning Models," International Journal of Data Mining & Knowledge Management Process, vol. 09, no. 03, pp. 19–27, 2019, doi: 10.5121/ijdkp.2019.9302.

N. R. Dewi, E. Y. Puspaningrum, and H. Maulana, "Analisis Sentimen Tweet Vaksinasi Covid-19 Menggunakan Rnn Dengan Metode Tf-Idf Dan Word2Vec," vol. 3, no. 1, pp. 56–65, 2022.x

S. N. Z. Baiq, "Analisis Perancangan dan Pengembangan Sistem Manajemen Informasi Kesiswaan di SMK Negeri 2 Gerung", Universitas Negeri Yogyakarta, pp. 1-90, 2016.

S. T. ind, Karambir, "A Simulation Model for the Spiral Software Development Life Cycle," International Journal of Innovative Research in Computer and Communication Engineering, vol. 03, no. 05, pp. 3823–3830, 2015, doi: 10.15680/ijircce.2015.0305013.

U. Rofiqoh, R. S. Perdana, and M. A. Fauzi. (2017). "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya. Vol. 1. No. 12. pp. 1725–1732.

Y. Astari and R. Wahib, "Analisis Sentimen Multi-Class pada Sosial Media Menggunakan Metode Long Short-Term Memory(LSTM), "Jurnal Linguistik Komputasional, vol.4, no.1, pp. 8-12, 2021, [Online]. Available: http://inacl.id/journal/index.php/jlk/article/view/43.