



Jurnal Teknologi Informasi dan Komunikasi

Vol: 13 No 1 2022

E-ISSN: 2477-3255

Diterima Redaksi: 26-04-2022 | Revisi: 12-05-2022 | Diterbitkan: 31-05-2022

LSTM (Long Short Term Memory) for Sentiment COVID-19 Vaccine Classification on Twitter

Miftahul Ihsan¹, Benny Sukma Negara², Surya Agustian³

^{1,2,3,c}Teknik Informatika, Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim,
^{1,2,3}Jl. HR. Soebrantas RW 15, Simpang Baru, Pekanbaru, Riau, telp. (0761) 562223
e-mail: 111751100570@students.uin-suska.ac.id, ²bsnegara@uin-suska.ac.id,
³surya.agustian@uin-suska.ac.id

Abstract

The implementation of the Covid-19 vaccination carried out by Indonesian government was ignited pros and contras among the public. Certainly, there will be pros and cons about the vaccination from the community. This attitude of pros and cons, which is also called sentiment, can influence people to accept or refuse to be vaccinated. Today, people express their sentiment in social media in comments, post, or status. One of the methods used to detect sentiment on social media, whether positive or negative, is through a categorisation of text approach. This research provides a deep learning technique for sentiment classification on Twitter that uses Long Short Term Memory (LSTM), for positive, neutral and negative classes. The word2vec word embeddings was used as input, using the pretrained Bahasa Indonesia model from Wikipedia corpus. On the other hand, the topic-based word2vec model was also trained from the Covid-19 vaccination sentiment dataset which collected from Twitter. The data used after balanced is 2564 training data, 778 data validation data, and 400 test data with 1802 neutral data, 1066 negative data, and 566 positive data. The best results from various parameter processes give an F1-Score value of 54% on the test data, with an accuracy of 66%. The result of this research is a model that can classify sentiments with new sentences.

Keywords: LSTM, Covid-19 vaccine, word2vec, word embeddings, sentiment classification

Metode LSTM (Long short term memory) untuk Klasifikasi Sentimen Vaksin Covid-19 pada Twitter

Abstrak

Pelaksanaan vaksinasi Covid-19 yang dilakukan pemerintah Indonesia, pada awalnya telah memicu pro kontra dalam masyarakat. Perbuatan pro kontra yang disebut juga sentiment tersebut, dapat mempengaruhi orang untuk menerima atau menolak divaksin. Saat ini, sentiment dapat diungkapkan dalam media sosial dalam bentuk posting, komentar, maupun status. Salah satu cara untuk mendeteksi sentimen di media sosial apakah positif atau negatif, adalah pendekatan klasifikasi teks. Penelitian ini mengusulkan teknik deep learning dengan Long short term memory (LSTM) untuk klasifikasi sentimen pada Twitter, untuk kelas positif, netral dan negatif. Model word embeddings word2vec digunakan sebagai input, menguji penggunaan pretrained model Bahasa Indonesia dari korpus Wikipedia. Di samping itu model

word2vec berbasis topik juga dilatih dari dataset sentimen vaksinasi Covid-19 yang dikumpulkan dari Twitter. Data yang digunakan setelah dilakukan penyeimbangan adalah 2563 data latih, 778 data validasi dan 400 data uji dengan 1802 data netral, 1066 data negatif dan 566 data positif. Hasil terbaik dari berbagai proses optimasi parameter memberikan nilai F1-Score 54% dengan akurasi 66%. Hasil dari penelitian ini menghasilkan model yang dapat mengklasifikasi sentimen dengan kalimat yang baru.

Kata kunci: LSTM, Vaksin Covid-19, word2vec, word embeddings, klasifikasi sentimen

1. Pendahuluan

Pelaksanaan vaksinasi Covid-19 sudah dilakukan pemerintah Indonesia di 34 Provinsi dengan total sasaran sebanyak 181,5 juta jiwa. Pada januari 2021 Indonesia melalui gugus tugas COVID 19 melaporkan bahwa terdapat 27.203 kematian yang diakibatkan oleh COVID-19 dan menjadi yang ketiga di Asia [1]. Program vaksinasi ini diharapkan mampu untuk mengurangi tingkat kematian yang terjadi di indonesia. Namun, Proses vaksinasi ini banyak mendapat penentangan dari masyarakat. Penolakan banyak disampaikan melalui media sosial dan disebarluaskan melalui pesan berantai pada grup *chat*.

Pada awal program vaksinasi tersebut, survei dilakukan oleh WHO yang bermitra dengan kementerian Kesehatan Indonesia, NITAG, dan UNICEF, pada 112.888 responden seluruh provinsi di Indonesia. Alhasil sekitar 65% responden menyatakan kesiapannya untuk menerima vaksin yang diberikan oleh pemerintah Indonesia, sementara 8% responden menyatakan mereka tidak akan menerimanya. Lebih dari 27% sisanya menyatakan kekhawatiran tentang niat pemerintah untuk mengelola vaksin. Sekitar 30% responden menyatakan bahwa mereka yang memiliki teman dekat, anggota keluarga, kolega, atau tetangga yang pernah terkena Covid-19 ingin segera mendapatkan vaksin. Responden juga menanyakan tentang keamanan dan kemanjuran dari vaksin dan merasa kurang percaya terhadap vaksin dan keraguan tentang halal-haram vaksin tersebut [1].

Selanjutnya, sikap pro (mendukung) dan kontra (menolak) dikalangan masyarakat dapat berkembang menjadi sentimen. Sentimen merupakan padangan, perasaan, pendapat atau penilaian seseroang terhadap suatu produk, jasa atau peristiwa [2]. Analisis sentimen merupakan kegiatan menganalisis pendapat warga, tingkah laku, penilaian, serta emosi pada suatu produk, organisasi, perhimpunan, seorang tokoh, pelayanan, serta kasus maupun permasalahan yang terjadi di lingkungan masyarakat itu sendiri [3].

Pesatnya penggunaan internet dan perangkat gawai pribadi saat ini, menyebabkan media sosial yang langsung diakses dari gawai pribadi menjadi tempat untuk mengeskpresikan diri, pikiran dan perasaan pribadi. Sentimen yang dilontarkan oleh para *influencer* media sosial dapat mempengaruhi sikap masyarakat dalam menerima program vaksin. Langkah-langkah strategis perlu diambil oleh pemerintah dan Lembaga terkait untuk mengatasi penolakan tersebut, termasuk juga menyusun cara persuasif untuk membujuk dan meyakinkan masyarakat akan pentingnya program vaksinasi Covid-19. Analisis sentimen masyarakat menjadi satu metode yang dilakukan untuk mengetahui respon masyarakat, agar program vaksinasi dapat berjalan dengan lancar dan mencapai target.

Tugas analisis sentimen pada media sosial ini sulit dilakukan secara manual, karena jumlah posting dan komentar di media social sangat banyak jumlahnya. Sehingga, pendekatan yang dilaksanakan salah satunya yakni mengumpulkan dan mengelompokkan teks secara otomatis dengan algoritma tertentu. Pada twitter misalnya, analisis sentimen dilakukan untuk melihat apakah suatu *tweet* bernilai positif, negatif atau netral [4].

Teknik analisa sentimen pada teks atau dokumen, umumnya menggunakan pendekatan klasifikasi dan deteksi. Beberapa yang sudah dikenal luas di antaranya adalah *rule base*, *machine learning* (*deep learning* dan *conventional machine learning*) serta *hybrid*. Metode *deep learning* telah membawa banyak kemajuan pada *Natural Language Processing* (NLP), khususnya tugas klasifikasi teks.

Metode *rule base* adalah salah satu metode yang berbasis *rule* (aturan), umumnya digunakan pada sistem pakar. Metode ini terdiri dari sekumpulan aturan yang disusun secara manual, dan diterapkan berulang ulang pada suatu fakta [5]. Contoh implementasinya adalah klasifikasi penamaan buah mangga yang menggunakan ciri ciri buah mangga sebagai basis aturannya [6]. Untuk kasus pengelompokan dokumen teks, metode berbasis aturan sangat tidak efisien karena dimensi teks dan kompleksitasnya yang sangat besar.

Berbeda dari *rule base*, *machine learning* adalah metode pembelajaran mesin (komputer) yang menggunakan algoritma tertentu untuk melakukan tugas tanpa perlu menerapkan rule secara manual. Sejumlah data yang telah diberikan label diberikan sebagai input untuk computer menemukan pola (model) yang kemudian akan digunakan terhadap data baru. Berdasarkan perkembangannya, *machine learning* juga dapat dikelompokkan menjadi dua yaitu *conventional machine learning* dan *deep learning* [7]. *Conventional Machine Learning* memiliki struktur yang simple seperti linear regression atau decision tree sedangkan *deep learning* memiliki struktur yang kompleks seperti jaringan saraf tiruan .

Metode machine learning konvensional yang cukup efektif antara lain adalah *support vector machine* serta *naïve bayes*. Metode *Naïve bayes* yang digunakan pada [8] untuk tugas klasifikasi sentimen vaksin Covid-19 pada twitter. *Tuning parameter*, kombinasi teks *preprocessing* serta penyeimbangan jumlah data latih pada dataset yang tidak seimbang. Menghasilkan akurasi tertinggi *F1-score* 57.15% dengan akurasi 61%. Metode *support vector machine* dengan input *Bag of Word* berdasarkan TF-IDF pada [9] dengan dataset yang sama, menghasilkan *F1-score* 56.81% dan akurasi 65% setelah tuning parameter, dan kombinasi setup eksperimen yang paling optimal.

Metode machine learning konvensional memiliki masalah yaitu sulitnya untuk menetukan *feature extraction* yang harus dimasukkan pada model. Jika *feature* hilang atau belum komplit maka model akan memberi keluaran yang tidak sempurna. Seiring dengan perkembangan pemodelan dengan menggunakan Jaringan Syaraf Tiruan (JST), pada tahun 1997 muncul model baru yang disebut *Long Short Term Memory* (LSTM). LSTM sangat baik dalam menangani klasifikasi teks karna model LSTM dirancang untuk bekerja pada untuk menangani masalah ketergantungan jangka panjang dan merupakan unit spesial dari *Recurrent Neural Network* (RNN) [10].

Long short term memory (LSTM) yang digunakan untuk klasifikasi sentimen pada media sosial multiclass [11]. Hasil akurasi tertinggi diperoleh dari uji coba sebanyak 5 kali, adalah 91,9%. Word2vec sebagai input diterapkan pada sentimen ulasan hotel [12], dengan membandingkan model *Continous Bag of Word* (CBOW) serta word2vec Skip-gram. Hasil akurasi tertinggi 85.96% diperoleh pada model skip-gram.

Paper ini membahas metode LSTM dengan input *word embeddings* *word2vec* yang bertujuan untuk mengklasifikasi sentimen masyarakat pada vaksin Covid-19 di Twitter dan model dapat memprediksi kata kata baru. Eksperimen dilakukan terhadap dataset yang digunakan pada [8], dan [9] dan teks preprocessing yang sama untuk mengevaluasi apakah metode Long Short Term Memory dengan menggunakan Word2vec sebagai word embedding dapat memiliki performa yang lebih baik dari kedua metode *machine learning* konvensional itu.

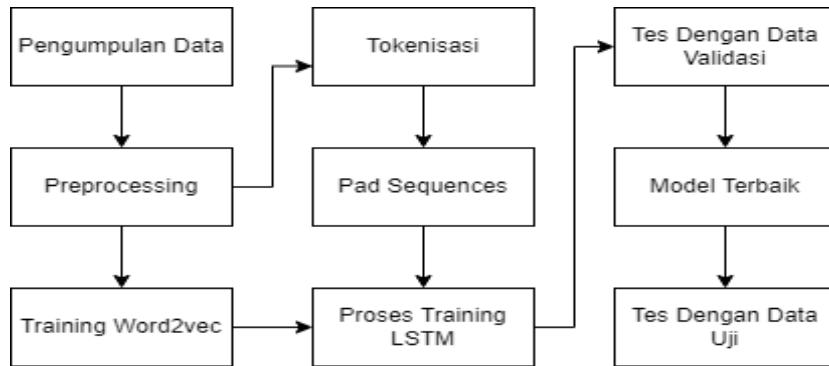
2. Metodologi Penelitian

Tujuan penelitian ini guna mengklasifikasi sentimen masyarakat di media sosial Twitter terkait vaksin Covid 19. Secara garis besar penelitian ini terdiri dari 9 langkah penelitian, yaitu pengumpulan dan konstruksi *dataset* , *preprocessing*, *Training word2vec*, tokenisasi, *pad sequences*, proses latih LSTM, optimisasi parameter, pemilihan model terbaik, dan pengujian akhir/evaluasi seperti pada gambar 1.

2.1. Pengumpulan dan Kontruksi Dataset

Data yang dikumpulkan yakni data tweet dari tanggal 23 Maret sampai 20 April 2021 (sekitar 1 bulan) dengan cara *crawling* menggunakan bahasa pemrograman *python* dan *Twitter API*. Data yang diperoleh yakni tweet Bahasa Indonesia telah difilter dari regional “Indonesia”

saja. Kata kunci dipilih sedemikian sehingga mengarah ke topik kegiatan vaksinasi Covid-19, walaupun tidak mengandung kata-kata terkait langsung seperti “covid”, “Sinovac” dan seterusnya. Dari daftar kata kunci yang digunakan sebagaimana tabel 1, setelah data yang ganda (tweet yang sama) dihapus, menghasilkan sebanyak 13.115 tweet.



Gambar 1. Metodologi Penelitian

Kemudian dilakukan proses pemberian label untuk tweet yang mengandung sentimen positif, negatif. Label adalah anotasi yang diberikan untuk setiap tweet, untuk diklasifikasikan ke dalam kelas positif, negatif atau netral. Label positif biasanya mengandung kata-kata puji, saran, masukan, dan refleksi atas emosi positif seperti kegembiraan, kesenangan, dan kebahagiaan. Label negatif dapat berisi keluhan, ungkapan sindiran, kritik, dan refleksi emosi negatif seperti frustasi, ketidakpuasan, dan kekecewaan. Sedangkan di luar itu, dapat dikategorikan sebagai netral. Dari jumlah label tersebut, tugas klasifikasi yang akan dilakukan adalah klasifikasi multilabel, yang memiliki tantangan lebih sulit daripada hanya menggunakan 2 label (*binary classification*).

Proses labeling dilakukan secara *crowdsourcing*, setiap tweet dianotasi oleh tiga orang annotator [13]. Label final untuk setiap tweet adalah majority vote dari tiga annotator. Labeling dilakukan oleh 12 orang yang dibagi menjadi 4 kelompok yang mana 1 kelompok terdiri atas 3 orang. Setiap satu orang akan melabeli sebanyak 3000 data. Contoh anotasi label terdapat pada tabel 2. Dari proses labeling ini, beberapa tweet yang memiliki anotasi positif, negatif dan netral sekaligus, dibuang dari dataset.

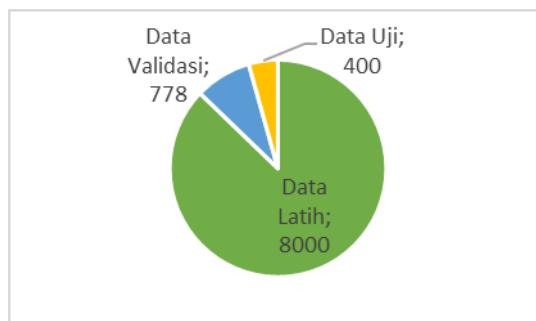
Tabel 1 Kata Kunci Pencarian

Sangat relevan dengan topik sentimen vaksin Covid-19	Tidak spesifik mengandung kata “covid” atau nama vaksin, tapi dapat mengarah ke topik vaksinasi covid
<ul style="list-style-type: none"> • Vaksin Corona, • Vaksin Covid, • Vaksin Covid Jadi Sehat, • Vaksin nusantara, • Vaksin Sinovac, • Vaksin Covid Gratis, • Vaksin Gotong Royong, • Vaksin Sinovac aman • Vaksin Sinovac mantap • Vaksin Sinovac bagus • Positif covid setelah vaksin • Vaksin Sinovac lemah • Vaksin Sinovac gagal 	<ul style="list-style-type: none"> • Vaksin Tanpa Efek Samping, • Vaksin Murah, • Vaksin Nakes, • Vaksin Guru, • Vaksin Sukarela” • Vaksinasi Indonesia, • Vaksin berhasil • Optimis vaksin aman • Vaksin lemes • Vaksin gagal • Vaksin bulan puasa • Sakit habis vaksin

Tabel 2 Pelabelan Dataset

Sentimen	Tweet
Negatif	@501Awani Bodoh. SOP Bodoh, Mentri Bodoh, Vaksin Bodoh
Positif	@kei_aria ga perlu ragu bro,vaksin aztrazeneca itu halal dan aman,MUI udh kelurkan fatwanya kok apalagi vaksinnya sudah lolos uji BPOM
Netral	6 juta vaksin kembali tiba di Indonesia http://t.co/fEoHkc1r7

Dari data yang sudah valid, dilakukan pembagian data ke dalam 3 dataset (Gambar 3): *latih* atau data latih, *development* atau data validasi dan *testing* atau data uji. Data *latih* terdiri atas 8000 tweet, data *development* 778 tweet, seperti komposisi pada Gambar 3. Data *testing* terdiri atas 400 data (komposisi yang relatif seimbang antara jumlah data positif, negatif dan netral), adalah data yang tidak terlihat selama proses latih metode ML yang diusulkan, seperti digunakan pada [8] [9]. *Dataset* ini merupakan *dataset* yang sama yang digunakan oleh [8] [9].



Gambar 2 Hasil Pelabelan Data

2.2. Text Preprocessing

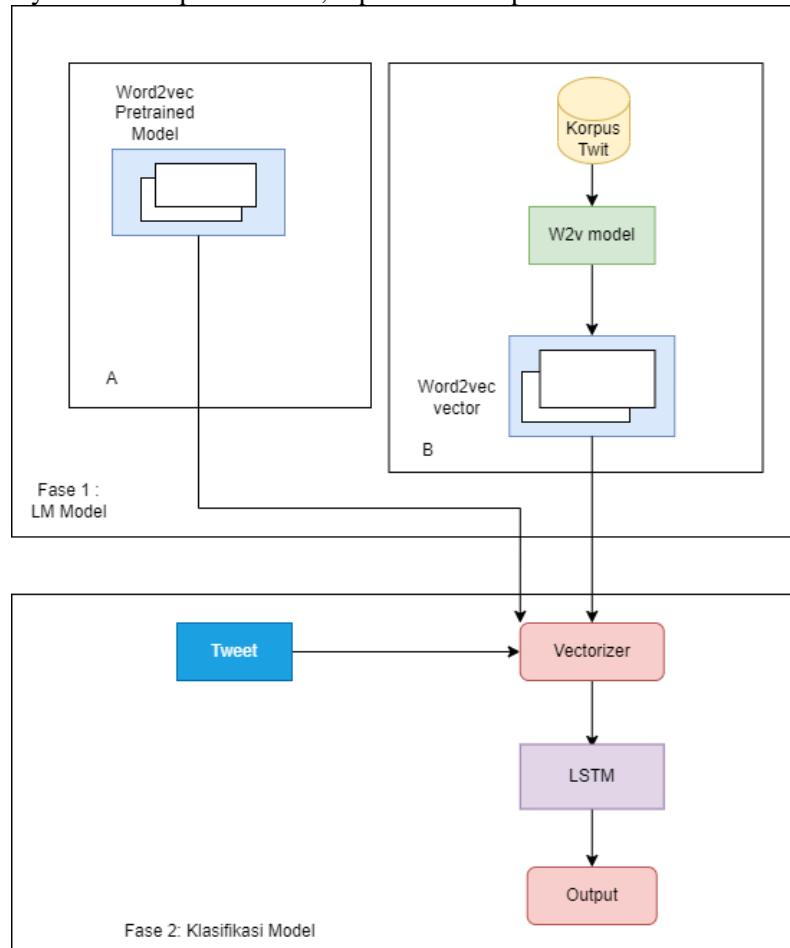
Tahap *text preprocessing* bertujuan untuk menyiapkan data sebelum proses klasifikasi [14]. Pada penelitian ini dilakukan beberapa tahapan preprocessing, yaitu *hyperlink removal* (menghapus *hyperlink*) dan pengubahan setiap *mention* (misalnya “@budi0232”, “@supermania”, dan lainnya) menjadi token “@USER”. Selanjutnya penghapusan angka dan *emoticon*, lalu *case folding* (mengubah isi teks menjadi huruf kecil semuanya), dan *punctuation removal* (menghapus tanda baca). Proses *stemming* (pengembalian kata ke bentuk akar kata) dan *stopword removal* (menghapus kata-kata yang tidak penting) dilakukan dengan menggunakan menggunakan library sastrawi. Komposisi preprocessing yang digunakan ini mengacu pada pengalaman hasil optimal yang dicapai pada [8][9].

2.3. Training Word2vec

Word2vec diperkenalkan oleh Mikolov, Corado & Chen [12]. *Word2vec* membantu sistem untuk mengenal kata yang direpresentasikan dalam bentuk vektor (*word embeddings*). *Word embeddings* sangat bermanfaat dalam tugas-tugas NLP dalam memperkecil dimensi vektor kata dibandingkan menggunakan model *bag of words*, agar lebih mudah dan efisien saat komputasi.

Penelitian menyelidiki penggunaan *word2vec* dengan model *pretrained* Bahasa Indonesia (dilatih terlebih dahulu oleh periset lain). Model tersebut dibangkitkan dari hasil latih

word2vec pada arsip data *dump* wikipedia pada tahun 2020. Model ini menjadi input untuk vektor kata-kata penyusun tweet pada LSTM, seperti terlihat pada Gambar 4 blok A.



Gambar 3 Proses Klasifikasi dengan model word embeddings

Di samping itu *word embeddings* yang berasal dari dataset tweet sentiment Covid-19 juga dilatih dengan *word2vec* (Gambar 3 blok B fase 1) untuk input LSTM dalam skema eksperimen kedua. Dimensi vektor *word embeddings* dari kedua *language model* yang digunakan adalah 100.

2.4. Tokenisasi

Proses tokenisasi yang dilakukan (Tabel 3) adalah memecah-mecah teks (*tweet*) menjadi *token* (kata). Setiap kata akan menjadi input pada setiap *node* di dalam LSTM, seperti ilustrasi cara kerja LSTM pada gambar 4.

2.5. Pad Sequences

Dalam arsitektur *deep learning*, dimensi input harus konsisten untuk seluruh *tweet* di dalam dataset. Karena panjang tweet tidak sama, maka diperlukan suatu teknik untuk memenuhi jumlah node input pada LSTM. *Pad Sequences* adalah teknik yang biasa digunakan untuk kebutuhan tersebut.

Suatu array dengan dimensi n dibentuk untuk menampung kata-kata pada tweet yang diwakili dengan nomor ID kata. Apabila jumlah kata lebih kecil dari n , maka node yang kosong akan diisi dengan nilai nol, sedangkan apabila panjang tweet melebihi n kata, maka sisanya akan dibuang [15]. Nilai nol dapat diisi pada bagian awal (*padding* awal), dan ID kata-kata *tweet* pada bagian akhir, atau sebaliknya. Penelitian ini menggunakan metode *padding* awal. Tabel 3 menggambarkan contoh proses *pad sequences* untuk suatu *tweet* menggunakan teknik *padding* awal.

Tabel 3 Proses pad sequences

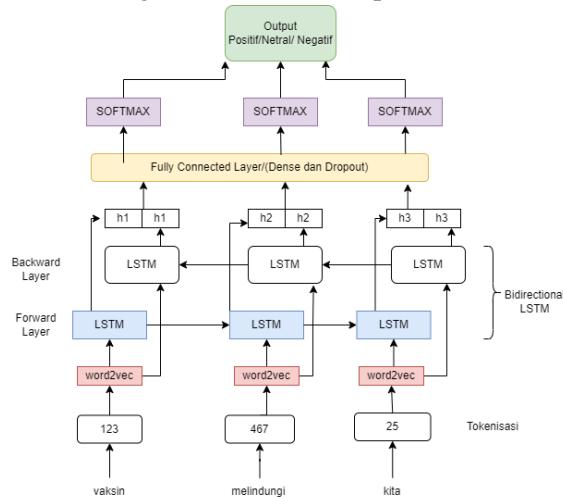
Langkah Proses	Hasil
Output dari preprocessing dimensi(1x20)	ga perlu ragu bro,vaksin aztrazeneca itu halal dan aman,MUI udh kelurkan fatwanya kok apalagi vaksinnya sudah lolos uji BPOM
Hasil Tokenisasi dimensi(1x20)	[[7, 109, 371, 463, 1, 572, 712, 493, 95, 557, 49, 29, 467, 202, 1098, 83, 75, 81, 19, 266, 31, 99]]
Hasil Pad Sequences dimensi(1x100)	array([[0,0,0,..., 7, 109, 371, 463, 1, 572, 712, 493, 95, 557, 49, 29, 467, 202, 1098, 83, 75, 81, 19, 266, 31, 99]], dtype=int32)

2.5. Proses Training LSTM

Metode LSTM yang digunakan adalah *Bidirectional LSTM* (Gambar 4). *Bidirectional LSTM* memanfaat konteks sebelumnya (*Forward Layer*) dan konteks setelahnya (*Backward Layer*) dengan memproses data dari dua arah dengan *hidden layer* terpisah.

Proses training LSTM dimulai dari vektor input dari hasil sub bab 3.3 yang diubah ke dalam bentuk *vektor word embeddings* oleh blok *vectorizer* (fase model klasifikasi pada Gambar 3). *Training* dilakukan terhadap data latih, dengan validasi menggunakan dataset validasi. Proses latih dilakukan dengan *epoch* sebanyak 50 dan *batch* 64. *Parameter tuning* dilakukan untuk menemukan model klasifikasi yang paling optimal. Dalam penelitian ini, diambil metrik akurasi untuk menemukan model optimal dari klasifikasi menggunakan LSTM ini.

Proses LSTM (Gambar 4) dimulai dari memecahkan kalimat menjadi kata dan dijadikan sebagai indeks (tokenisasi). Kata yang telah ditoken selanjutnya bentuk indeks tadi akan dijadikan bentuk vektor untuk menjadi *inputan* pada LSTM dengan word2vec. Selanjutnya, vektor word2vec masuk kedalam proses *Bidirectional LSTM* yaitu *forward* dan *backward layer* untuk memproses daa dari dua arah dengan *hidden layer* terpisah.



Gambar 4 Proses Training LSTM

3. Hasil dan Pembahasan

Tahap ini adalah tahap pencarian model optimal. Tahap ini memiliki tiga langkah yaitu : data *balancing* (penyeimbangan data), *parameter tuning* dan pengujian dengan data uji

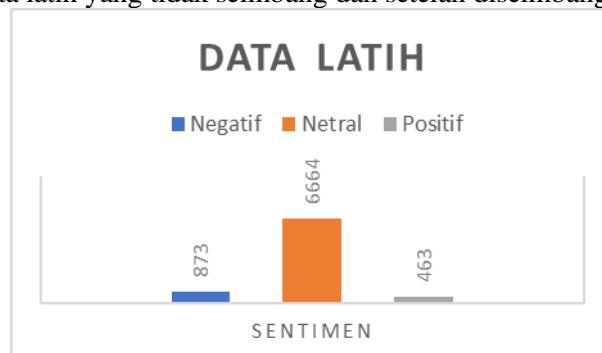
3.1 Data Balancing

Dari proses anotasi pemberian label yang dilakukan secara crowdsourcing, diperoleh kelas data yang tidak seimbang, antara positif, negatif dan netral. Pada dataset latih terdapat 6664 kelas netral, 873 kelas negatif dan 463 kelas positif. Sistem *machine learning* pada umumnya hanya akan berhasil mendeteksi suatu *tweet* sebagai netral (kelas yang terbanyak), dan gagal mendeteksi adanya sentiment positif dan negatif. dari data validasi untuk setiap kelas, setelah system dilatih menggunakan dataset latih yang tidak seimbang ini. Dari hasil tersebut, diperoleh F1-score keseluruhan adalah 30%.

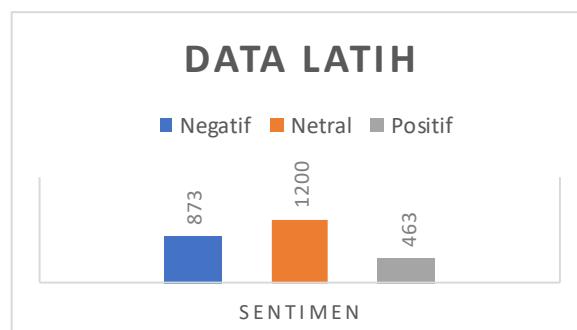
Tabel 4. Performa model dengan data latih awal (8000 Tweet) terhadap data validasi.

Metrik	Precision	Recall	F1 Score
Kelas			
Negatif	0.0	0.0	0.0
Netral	0.83	1.0	0.91
Positif	0.0	0.0	0.0

Langkah pertama dari pencarian model optimal ini adalah dengan melakukan penyeimbangan (balancing) pada data latih. Cara yang dilakukan adalah dengan mengurangi jumlah data pada kelas netral. Penentuan besar pengurangan data dilakukan secara empiris, dengan rationale bahwa distribusi netral akan sebanding dengan gabungan kelas positif dan negative. Gambar 5 dan 6 menunjukkan data latih yang tidak seimbang dan setelan diseimbangkan



Gambar 5. Komposisi data latih yang tidak seimbang



Gambar 6. Komposisi data latih diseimbangkan (balanced dataset)

Setelah langkah penyeimbangan data latih ini, proses pelatihan LSTM dengan data yang baru memberikan hasil *F1-score* yang meningkat signifikan, khususnya pada kelas negatif dan positif, seperti terlihat pada Tabel 5. Terlihat bahwa saat ini sistem telah dapat mengenali atau mendeteksi sentiment postif dan negatif. Perbandingan akurasi antara proses validasi untuk model awal dan model dengan *balanced dataset*, dapat dilihat pada gambar 7 berikut ini.

Selanjutnya adalah pencarian model optimal dari hasil saat ini sebagai dasar (*baseline*) dengan penyetelan parameter (parameter tuning).

Tabel 5 Performa balanced dataset terhadap data validasi

Kelas	Metrik	Precision	Recall	F1 Score
Negative		0.27	0.59	0.37
Netral		0.89	0.79	0.84
Positif		0.60	0.13	0.22

		Predicted labels		
		NEGATIVE	NETRAL	POSITIF
True labels	NEGATIVE	0	85	0
	NETRAL	1	647	0
	POSITIF	0	45	0

		Predicted labels		
		NEGATIVE	NETRAL	POSITIF
True labels	NEGATIVE	50	35	0
	NETRAL	130	514	4
	POSITIF	8	31	6

Gambar 7. Akurasi (dalam confusion matrix) terhadap pengujian data validasi (A) menggunakan data latih awal (B) menggunakan balanced train dataset

3.2 Parameter Tuning

Penyetelan parameter LSTM (*parameter tuning*) dilakukan untuk menemukan model yang paling optimal. Hal-hal yang di-tuning dalam penelitian ini adalah *learning rate* dan pemilihan fungsi aktivasi. *Learning rate* dipilih dari nilai {0.0001, 0.001, 0,01, dan 0,1}, sedangkan fungsi aktivasi yang dipilih adalah di antara fungsi *RELU* dan *leaky-RELU*. Selain itu, penelitian ini juga menginvestigasi model *word embedding* yang digunakan, yaitu *word embedding* dari *pre-trained Wikipedia corpus*, dan hasil *word embedding* yang dilatih sendiri menggunakan gabungan *tweet* pada dataset training dan validasi (8778 tweet).

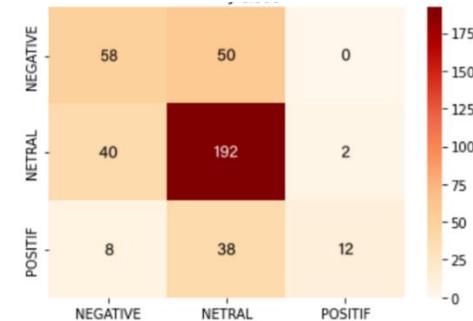
Proses pelatihan menggunakan mekanisme *early stopping* untuk mendeteksi *overfitting*. Pada saat terjadi *overfitting*, sistem akan memanggil kembali (*call-back*) model dengan hasil *epoch* sebelumnya yang nilai akurasi terhadap data validasinya lebih baik. Tabel 6 memperlihatkan hasil yang diperoleh dari tahap *parameter tuning*, terhadap data validasi. Berdasarkan tabel tersebut akurasi dan akurasi tertinggi diperoleh dengan komposisi fungsi aktivasi *Leaky-ReLU*, *learning rate* 10^{-4} , dan *word embedding* yang dibentuk dari dataset Covid, dengan hasil akurasi 64%. Model ini selanjutnya dipilih sebagai model final yang akan diterapkan pada data testing.

Tabel 6 tuning hyperparameter

Fungsi aktivasi	Learning rate	Word2vec pretrained model	Word2vec data train
Relu		Akurasi	Akurasi
	0.0001	49%	49%
	0.001	62%	62%
	0.01	45%	61%
	0.1	51%	51%
Leaky-Relu	0.0001	58%	64%
	0.001	45%	45%
	0.01	56%	56%
	0.1	50%	51%

3.3 Pengujian Data Uji

Model final yang diperoleh, diterapkan pada data testing, yang belum pernah terlihat sebelumnya pada saat proses training klasifikasi menggunakan LSTM maupun training pembentukan language model menggunakan word2vec. Hasil prediksi kelas positif, negative dan netral pada data uji yang terdiri atas 400 tweet tersebut, dapat dilihat pada gambar 8 dalam bentuk *confusion matrix*.



Gambar 8 . Hasil Confusion Matrix

Dari hasil tersebut, metode LSTM yang diusulkan memiliki performa yang kompetitif bila ditinjau dari *F1-score*, akurasi, precision dan recall, bila dibandingkan dengan penelitian sebelumnya yang menggunakan metode *Naïve Bayes* [8], dan *SVM* [9], sebagaimana dapat dicermati dalam tabel 7.

Tabel 7 Perbandingan Metode

Dari segi akurasi, LSTM memiliki nilai yang tertinggi yaitu 66% sedangkan *Naïve*

Metode	Accuracy	F1-score	Precision	Recall
Naïve Bayes[8]	0.61	0.57	0.58	0.60
Support Vector machine[9]	0.65	0.56	0.61	0.54
<i>Long short term memory</i> (Metode ini)	0.66	0.54	0.75	0.53

Bayes memiliki akurasi 61% dan *SVM* mendapatkan 65%. Namun pada nilai F1 score metode LSTM yang diusulkan masih lebih rendah, dengan nilai 56% dibandingkan dengan *Naïve Bayes* 57%. Walau demikian, untuk dataset training yang berukuran kecil, LSTM sebagai salah satu metode deep learning yang memiliki kompleksitas komputasi yang tinggi, telah menunjukkan performa yang cukup baik, karena dapat bersaing dengan metode machine learning konvensional seperti *SVM* dan *NB*.

Kurangnya jumlah data latih untuk kelas tweet yang mengandung sentimen, menyebabkan sistem kurang dapat mendeteksi kelas yang berisi sentimen positif dan negatif. Diperlukan lebih banyak kata-kata spesifik yang berisi sentimen untuk terlihat pada saat training, sehingga untuk ini, diperlukan lebih banyak data tweet berisi sentimen positif dan negatif, agar model hasil pelatihan lebih optimal lagi.

4. Kesimpulan

Hasil implementasi serta pengujian yang telah dilaksanakan, bahwasannya metode *Long Short Term Memory* dapat diterapkan pada klasifikasi sentimen pada media sosial Twitter. Parameter yang paling optimal didapatkan *leaky-relu* 10^4 dengan akurasi 64% pada data validasi dan 66% pada data uji. Pada kasus ini LSTM mempunyai akurasi lebih besar dibandingkan *Naïve Bayes* serta *SVM*. Namun, *F1-score* yang didapatkan tidak terlalu baik dikarenakan dataset yang sedikit dan tidak seimbang dan kasus yang dilakukan bukan klasifikasi biner, melainkan multilabel.

Saran untuk penelitian selanjutnya, adalah dengan menambah koleksi data tweet yang berisi sentimen positif dan negatif sehingga jumlah data yang banyak diharapkan dapat

meningkatkan hasil klasifikasi menggunakan metode deep learning LSTM. Selain itu, masih terbuka peluang untuk menyelidiki model arsitektur LSTM yang lebih kompleks dan melalukan parameter tuning untuk model tersebut.

Daftar Pustaka

- [1] Pristiyyono, M. Ritonga, M. A. Al Ihsan, A. Anjar, and F. H. Rambe, "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012045, 2021, doi: 10.1088/1757-899x/1088/1/012045.
- [2] V. N. T. Le, S. Ahderom, and K. Alameh, "Performances of the lbp based algorithm over cnn models for detecting crops and weeds with similar morphologies," *Sensors (Switzerland)*, vol. 20, no. 8, pp. 1–18, 2020, doi: 10.3390/s20082193.
- [3] A. L. Fairuz, R. D. Ramadhani, and N. A. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial," *J. DINDA*, vol. 1, no. 1, pp. 10–12, 2021.
- [4] M. A. Fauzi and S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking Image Processing View project Smart Wheelchair View project," *Researchgate.Net*, vol. 2, no. 8, pp. 2766–2770, 2018.
- [5] F. Gregorio, G. González, C. Schmidt, and J. Cousseau, "Internet of Things," in *Signals and Communication Technology*, 2020.
- [6] S. Hartanto, "Implementasi Fuzzy Rule Based System." in *Techsi*, vol 9 no 2, pp. 103-117, 2017
- [7] N. K. Chauhan and K. Singh, "A review on conventional machine learning vs deep learning," *Int. Conf. Comput. Power Commun. Technol. GUCON 2018*, pp. 347–352, 2019
- [8] P. Yohana, "Analisis Sentimen Vaksin Covid19 Menggunakan Naive Bayes," Skripsi, 2022.
- [9] M. Rizky, "Vaksin Covid-19 Menggunakan Metode Support Vector Machine Pada Media Sosial Twitter Covid-19", Skripsi, 2021.
- [10] A. Arfan and L. ETP, "Perbandingan Algoritma Long Short-Term Memory dengan SVR Pada Prediksi Harga Saham di Indonesia," *Petir*, vol. 13, no. 1, pp. 33–43, 2020,
- [11] Y. Astari and S. W. Rozaqi, "Analisis Sentimen Multi-Class pada Sosial Media menggunakan metode Long Short-Term Memory (LSTM)," vol. 4, no. 1, pp. 8–12, 2021.
- [12] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment Analysis Using Word2vec and Long Short-Term Memory (LSTM) for Indonesian Hotel Reviews," *Procedia Comput. Sci.*, vol. 179, pp. 728–735, 2021
- [13] R. Wardoyo, A. Musdholifah, G. Angga Pradipta and I. N. Hariyasa Sanjaya, "Weighted Majority Voting by Statistical Performance Analysis on Ensemble Multiclassifier," *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 2020, pp. 1-8, doi: 10.1109/ICIC50835.2020.9288552.
- [14] C. J. E. Munthe, N. A. Hasibuan, and H. Hutabarat, "Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace," vol. 2, no. 3, pp. 110–115, 2022.
- [15] Y. Xie, L. Le, Y. Zhou, and V. V. Raghavan, *Deep Learning for Natural Language Processing*, vol. 38. 2018.