



**PERBANDINGAN METODE WORD EMBEDDING UNTUK KLASIFIKASI  
GEJALA COVID-19 DARI PESAN DI MEDIA SOSIAL DENGAN LONG  
SHORT-TERM MEMORY**

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh**

**RAISA AMALIA**

**NIM 1911016220023**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**JUNI 2023**



**PERBANDINGAN METODE WORD EMBEDDING UNTUK KLASIFIKASI  
GEJALA COVID-19 DARI PESAN DI MEDIA SOSIAL DENGAN LONG  
SHORT-TERM MEMORY**

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh**

**RAISA AMALIA  
NIM 1911016220023**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**JUNI 2023**

# SKRIPSI

## PERBANDINGAN METODE *WORD EMBEDDING* UNTUK KLASIFIKASI GEJALA COVID-19 DARI PESAN DI MEDIA SOSIAL DENGAN *LONG SHORT-TERM MEMORY*

Oleh:

**RAISA AMALIA**  
**NIM. 1911016220023**

Telah dipertahankan di depan Dosen Penguji pada tanggal 21 Juni 2023.

Susunan Dosen Penguji:

**Pembimbing I**



M. Reza Faisal, S.T., M.T., Ph.D

NIP. 197612202008121001

**Dosen Penguji I**



Muhammad Itqan Mazdadi, S.Kom., M.Kom.

NIP. 199006122019031013

**Pembimbing II**



Fatma Indriani, S.T., M.I.T., Ph.D

NIP. 198404202008122004

**Dosen Penguji II**



Irwan Budiman, S.T., M.Kom.

NIP. 197703252008121001



## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 21 Juni 2023

Yang Menyatakan,



Raisa Amalia

NIM.1911016220023

## ABSTRAK

### PERBANDINGAN METODE *WORD EMBEDDING* UNTUK KLASIFIKASI GEJALA COVID-19 DARI PESAN DI MEDIA SOSIAL DENGAN *LONG SHORT-TERM MEMORY*

(Oleh : Raisa Amalia; Pembimbing: M. Reza Faisal, S.T., M.T., Ph.D dan Fatma Indriani S.T., M.I.T., Ph.D; 2023; 60 halaman)

Kasus COVID-19 pertama kali muncul di Indonesia pada bulan Maret Tahun 2020. Gejala umum penderitaan COVID-19 antara lain demam, batuk, kelelahan, dan kehilangan rasa pada indera penciuman. Sejak pandemi dimulai, beberapa pengguna Twitter menyampaikan keresahan mereka akan gejala yang mereka alami melalui tweet dan beberapa secara jelas menyatakan apakah mereka terinfeksi virus tersebut atau tidak. Tweet yang berhubungan dengan COVID-19 beberapa kali menjadi sumber penelitian terkait analisis sentimen, diantaranya tweet sentimen tentang varian baru dari virus, kebijakan pemerintah dalam menghadapi pandemi, dan sentimen terhadap proses vaksinasi, sedangkan penelitian untuk mendeteksi gejala COVID-19 dari pesan di media sosial masih belum banyak dilakukan di Indonesia. Hasil deteksi untuk menunjukkan apakah tweet tersebut dibuat oleh seseorang yang terindikasi positif COVID-19 atau negatif. Metode *Long Short-Term Memory* (LSTM) merupakan bagian dari jaringan saraf berulang dan pengembangan dari metode RNN dengan kelebihan mampu memproses data teks yang relatif panjang. Metode *Word Embedding* yang digunakan adalah Word2Vec, FastText, dan GloVe untuk membuat model dan membandingkan metode yang memiliki hasil akurasi dan prediksi yang baik. Pada penelitian ini diusulkan metode gabungan dari ketiga *word embedding*. Hasil yang diperoleh yaitu metode LSTM dengan *Word Embedding* Word2Vec mendapatkan akurasi 87%, hasil akurasi dengan GloVe 84.5%, hasil akurasi dengan FastText 86%, dan gabungan dari ketiga *word embedding* yang diusulkan nilai akurasinya sebesar 89%.

**Kata kunci:** Gejala COVID-19, Twitter, Klasifikasi, *Long Short-Term Memory*, *Word Embedding*, Ekstraksi Fitur

## **ABSTRACT**

### **COMPARISON OF WORD EMBEDDING METHOD FOR CLASSIFICATION OF COVID-19 SYMPTOMS FROM SOCIAL MEDIA MESSAGES WITH LONG SHORT-TERM MEMORY**

*(By : Raisa Amalia; Supervisor: M. Reza Faisal, S.T., M.T., Ph.D and Fatma Indriani S.T., M.I.T., Ph.D; 2023; 60 pages)*

*The first case of COVID-19 appeared in Indonesia in March 2020. Common symptoms of suffering from COVID-19 include fever, cough, fatigue, and loss of sense of smell. Since the pandemic began, several Twitter users have expressed their dismay about their symptoms through tweets and some have clearly stated whether they are infected with the virus or not. Tweets related to COVID-19 have several times become a source of research related to sentiment analysis, including sentiment tweets about new variants of the virus, government policies in dealing with pandemics, and sentiments about the vaccination process, while research to detect symptoms of COVID-19 from messages on social media still not much done in Indonesia. Detection results to show whether the tweet was made by someone who indicated positive for COVID-19 or negative. The Long Short-Term Memory (LSTM) method is part of a recurrent neural network and the development of the RNN method with the advantage of being able to process relatively long text data. The Word Embedding methods used are Word2Vec, FastText, and GloVe to create models and compare methods that have good accuracy and prediction results. In this study, a combined method of the three-word embedding is proposed. The results obtained by the method are LSTM with Word Embedding Word2Vec getting 87% accuracy, accuracy with GloVe 84.5%, accuracy with FastText 86%, and a combination of the three word embeddings which claimed an accuracy value of 89%.*

**Keywords:** *Symptoms of COVID-19, Twitter, Classification, Long Short-Term Memory, Word Embedding, Feature Extraction*

## PRAKATA

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “PERBANDINGAN METODE WORD EMBEDDING UNTUK KLASIFIKASI GEJALA COVID-19 DARI PESAN DI MEDIA SOSIAL DENGAN LONG SHORT-TERM MEMORY” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Keluarga besar terutama orangtua dan saudara yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini
2. Bapak M. Reza Faisal, S.T., M.T., Ph.D. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Ibu Fatma Indriani, S.T., M.I.T., Ph.D. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman, S.T., M.Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
7. Ucapan terima kasih kepada rekan mahasiswa Ilmu Komputer Angkatan 2019 terkhusus untuk Muhammad Nouval Rahman, Fadhila Nisrina Saumy, Alya Firyal Nuha, Siti Roziana Azizah, Almira Syahadati Arsyah, Dina Arifah, Ajwa

Helisa, Anggota Grup (Apa Aja, HOPE), dan teman-teman yang memberikan dukungan dalam proses mengerjakan skripsi.

8. Sahabat yang telah membantu dalam menunjang aktivitas penulisan skripsi secara fisik dan juga mental seperti Safina Salsabila Putri, Miranda Putri Kalsum, Nurushofiah Safarina, Ratu Embun Permata Sari, Rizqiqa Harini, Sekar Vanida Sari, Tsania Rizqa Weninda, Lanjar Yoga Rahmawati, Dinda Galuh Wulandini, Indira Shafa Widiani, Nouvaliza Aisy Akmalia, Oktavina Yoka Pratami, dan Rosari Rizqiya Farhana.
9. Mental support yang menemani dan menghibur selama masa pembuatan skripsi seperti Monsta X terutama Chae Hyungwon, Suguru, Oikawa, Kaeya, Alhaitham, dan teman-teman online yang menemani yaitu Aufa Zahra Andrian, Hanan Afifah Rachmadini, Nabila Safa Athaya, dan Muhammad Devanda Ghoziandi.
10. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan serupa berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 21 Juni 2023



Raisa Amalia



## DAFTAR ISI

	<b>Halaman</b>
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN .....	iii
ABSTRAK.....	iv
<i>ABSTRACT</i> .....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	xi
DAFTAR LAMPIRAN.....	xii
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah .....	4
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>5</b>
2.1 Kajian Terdahulu .....	5
2.2 Landasan Teori .....	12
2.2.1 COVID 19 .....	12
2.2.2 Twitter.....	12
2.2.3 <i>Text Mining</i> .....	13
2.2.4 <i>Natural Language Processing (NLP)</i> .....	13
2.2.5 <i>Text Preprocessing</i> .....	14
2.2.6 Ekstraksi Fitur .....	14
2.2.7 Klasifikasi .....	15
2.2.8 <i>Long Short-Term Memory</i> .....	15
2.2.9 Word2Vec .....	17
2.2.10 Glove.....	20
2.2.11 FastText.....	20

2.2.12	Metode Gabungan <i>Word Embedding</i> .....	23
2.2.13	<i>Confusion Matrix</i> .....	23
<b>BAB III</b>	<b>METODE PENELITIAN .....</b>	<b>25</b>
3.1	Bahan Penelitian .....	25
3.2	Alat Penelitian .....	25
3.3	Variabel Penelitian.....	25
3.4	Prosedur Penelitian .....	26
<b>BAB IV</b>	<b>HASIL DAN PEMBAHASAN .....</b>	<b>29</b>
4.1	Hasil.....	29
4.1.1	Pengumpulan Dataset.....	29
4.1.2	<i>Preprocessing Data</i> .....	30
4.1.3	Pembagian <i>Data Training</i> dan <i>Data Testing</i> .....	35
4.1.4	Ekstraksi Fitur .....	35
4.1.5	Klasifikasi .....	40
4.1.6	Evaluasi.....	43
4.2	Pembahasan .....	46
<b>BAB V</b>	<b>PENUTUP .....</b>	<b>54</b>
5.1	Kesimpulan .....	54
5.2	Saran .....	54
	DAFTAR PUSTAKA .....	55
	LAMPIRAN.....	61

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
Tabel 1. Keaslian Penelitian .....	8
Tabel 2. Perancangan Penelitian.....	11
Tabel 3. <i>Confusion Matrix</i> .....	23
Tabel 4. Data Tweet dan Label.....	29
Tabel 5. Jumlah Distribusi Kelas.....	30
Tabel 6. Teks sebelum dan sesudah tahapan <i>cleaning</i> .....	30
Tabel 7. Teks sebelum dan sesudah proses <i>case folding</i> .....	32
Tabel 8. Tabel <i>word index</i> .....	33
Tabel 9. Teks hasil tokenisasi.....	34
Tabel 10. Teks dengan Vektor Word2Vec .....	36
Tabel 11. Teks dengan Vektor GloVe .....	37
Tabel 12. Teks dengan Vektor FastText.....	37
Tabel 13. <i>Confusion Matrix</i> model pertama.....	41
Tabel 14. <i>Confusion Matrix</i> model kedua .....	41
Tabel 15. <i>Confusion Matrix</i> model ketiga .....	42
Tabel 16. <i>Confusion Matrix</i> model keempat .....	43
Tabel 17. <i>Confusion Matrix</i> model kelima.....	43
Tabel 18. Hasil akurasi, presisi, dan <i>recall</i> model pertama.....	44
Tabel 19. Hasil akurasi, presisi, dan <i>recall</i> model kedua .....	44
Tabel 20. Hasil akurasi, presisi, dan <i>recall</i> model ketiga.....	45
Tabel 21. Hasil akurasi, presisi, dan <i>recall</i> model keempat .....	45
Tabel 22. Hasil akurasi, presisi, dan <i>recall</i> model kelima.....	46
Tabel 23. Nilai <i>padding length</i> .....	47
Tabel 24. Perbandingan akurasi, presisi, dan <i>recall</i> antara panjang <i>padding</i> .....	48
Tabel 25. Hasil akurasi, presisi, dan <i>recall</i> dengan <i>word embedding</i> gabungan vektor 100 .....	48
Tabel 26. Hasil akurasi, presisi, dan <i>recall</i> dengan <i>word embedding</i> gabungan vektor 300 .....	49

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
Gambar 1. Tahapan <i>Text Mining</i> .....	13
Gambar 2. Arsitektur <i>Long Short-Term Memory</i> .....	16
Gambar 3. Contoh Model Vektor Word2Vec .....	17
Gambar 4. Perbedaan Metode CBOW dan Skip-Gram .....	18
Gambar 5. Cara Kerja Metode CBOW .....	18
Gambar 6. Cara Kerja Metode Skip-Gram .....	19
Gambar 7. Ilustrasi Skip-Gram .....	19
Gambar 8. Ilustrai cara kerja <i>n-grams</i> pada FastText .....	21
Gambar 9. Struktur Data 1D .....	22
Gambar 10. Struktur Data 2D .....	22
Gambar 11. Data 2D dari kombinasi 3 <i>word embedding</i> .....	23
Gambar 12. Alur Penelitian .....	26
Gambar 13. Plot model dari <i>word embedding</i> Word2Vec .....	36
Gambar 14. Tampilan <i>model summary</i> dari metode gabungan .....	38
Gambar 15. Tampilan plot model dari metode LSTM dengan <i>word embedding</i> gabungan .....	39
Gambar 16. Data dengan kombinasi 3 <i>word embedding</i> .....	39
Gambar 17. Tampilan plot model dari metode LSTM dengan <i>word embedding</i> gabungan .....	40
Gambar 18. Perbandingan rata-rata presisi sesuai metode <i>word embedding</i> .....	50
Gambar 19. Perbandingan rata-rata <i>recall</i> sesuai metode <i>word embedding</i> .....	51
Gambar 20. Perbandingan rata-rata akurasi sesuai metode <i>word embedding</i> .....	52
Gambar 21. Perbandingan rata-rata akurasi sesuai <i>padding</i> .....	53

## DAFTAR LAMPIRAN

### Lampiran

Lampiran 1. *Source Code Import Dataset*

Lampiran 2. *Source Code Label Encoder*

Lampiran 3. *Source Code Tokenisasi*

Lampiran 4. *Source Code Pembagian Dataset*

Lampiran 5. *Source Code load model Word2Vec*

Lampiran 6. *Source Code load model FastText*

Lampiran 7. *Source Code load model GloVe*

Lampiran 8. *Source Code LSTM dengan satu ekstraksi fitur*

Lampiran 9. *Source Code LSTM dengan ekstraksi fitur gabungan vektor 100*

Lampiran 10. *Source Code LSTM dengan ekstraksi fitur gabungan vektor 300*

Lampiran 11. Contoh Hasil Prediksi Model LSTM dengan Word2Vec

Lampiran 12. Contoh Hasil Prediksi Model LSTM dengan FastText

Lampiran 13. Contoh Hasil Prediksi Model LSTM dengan GloVe

Lampiran 14. Contoh Hasil Prediksi Model LSTM dengan Gabungan Vektor 100

Lampiran 15. Contoh Hasil Kata Terdekat yang ada Pada Korpus