



**PERBANDINGAN PERFORMA *CONVOLUTIONAL NEURAL NETWORK* 1D DAN 2D DENGAN FITUR *WORD EMBEDDINGS* PADA  
KLASIFIKASI GEJALA COVID-19 DARI PESAN SOSIAL MEDIA**

**SKRIPSI**

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

**Oleh**

**Muhammad Khairie**

**NIM 1911016310017**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU  
JUNI 2023**

**SKRIPSI**  
**PERBANDINGAN PERFORMA *CONVOLUTIONAL NEURAL NETWORK* 1D DAN 2D DENGAN FITUR *WORD EMBEDDINGS* PADA  
KLASIFIKASI GEJALA COVID-19 DARI PESAN SOSIAL MEDIA**

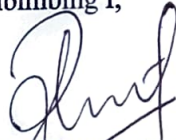
Oleh:

**MUHAMMAD KHAIRIE**  
**1911016310017**

Telah dipertahankan di depan Dosen Penguji pada tanggal 26 Juni 2023

Susunan Penguji :

Pembimbing I,



M. Reza Faisal, S.T., M.T., PhD  
NIP. 197612202008121001

Dosen Penguji I,



Irwan Budiman, S.T., M.Kom  
NIP. 197703252008121001

Pembimbing II,



Rudy Herteno, S.Kom., M.Kom.  
NIP. 198809252022031003

Dosen Penguji II,



Friska Abadi, S.Kom., M.Kom.  
NIP. 19880913201612110001



Banjarbaru, 26 Juni 2022  
Koordinator Program Studi Ilmu Komputer

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, 26 Juni 2023



Muhammad Khairie  
NIM. 1911016310017

## ABSTRAK

**PERBANDINGAN PERFORMA CONVOLUTIONAL NEURAL NETWORK 1D DAN 2D DENGAN FITUR WORD EMBEDDINGS PADA KLASIFIKASI GEJALA COVID-19 DARI PESAN SOSIAL MEDIA** (Oleh: Muhammad Khairie; Pembimbing: Mohammad Reza Faisal, S.T., M.T., Ph.D dan Rudy Herteno, S.Kom, M.Kom; 2023; 63 Halaman)

Media sosial berperan penting sebagai sumber informasi yang paling banyak diakses oleh masyarakat untuk mencari informasi terkait COVID-19. Pemanfaatan pesan sosial media tersebut nantinya dapat membantu dalam hal-hal seperti deteksi gejala COVID-19. Hasil deteksi gejala COVID-19 dari pesan sosial media dapat dijadikan tolak ukur oleh pihak terkait untuk dapat membuat kebijakan yang dapat berguna bagi masyarakat umum. Namun, data pesan sosial media memiliki banyak ragam topik sehingga diperlukan proses klasifikasi dengan berbagai macam pendekatan agar bisa mendapatkan hasil seakurat mungkin. Klasifikasi dilakukan dengan menggunakan variasi word embedding (Word2vec, Glove, fastText, dan gabungan Word2vec, Glove, dan fastText) dan variasi word padding (max, mean, median, dan mode) pada arsitektur CNN 1D dan 2D (single-channel dan multi-channel). Hasil dengan akurasi terbaik diperoleh dari kombinasi penggunaan word embedding Word2vec dan word padding max pada 1D single-channel yaitu 93%. Hasil perbandingan menunjukkan bahwa penggunaan Word2vec dengan *word padding* Max memiliki konsistensi dalam memberikan performa yang baik pada berbagai arsitektur yang digunakan.

**Kata kunci:** gejala covid-19, klasifikasi, word embedding, convolutional neural network, media sosial

## ***ABSTRACT***

**COMPARISON OF PERFORMANCE 1D AND 2D CONVOLUTIONAL NEURAL NETWORKS WITH WORD EMBEDDINGS FEATURES IN COVID-19 SYMPTOMS CLASSIFICATION FROM SOCIAL MEDIA MESSAGES** (By: Muhammad Khairie; Supervisor: Mohammad Reza Faisal, S.T., M.T., Ph.D dan Rudy Herteno, S.Kom, M.Kom; 2023; 63 Pages)

Social media plays a crucial role as the most accessed source of information by the public in seeking COVID-19 related information. The utilization of social media messages can assist in various aspects, such as COVID-19 symptom detection. The detection results of COVID-19 symptoms from social media messages can serve as a benchmark for relevant parties to formulate policies that are beneficial to the general public. However, social media data encompasses diverse topics, necessitating a classification process employing various approaches to achieve the most accurate results. The classification is performed using different variations of word embedding (Word2vec, Glove, fastText, and the combination of Word2vec, Glove, and fastText) and word padding techniques (max, mean, median, and mode) on 1D and 2D CNN architectures (single-channel and multi-channel). The best accuracy is obtained by employing Word2vec word embedding and max word padding on the 1D single-channel architecture, achieving a rate of 93%. Comparative results demonstrate that the utilization of Word2vec with max word padding consistently delivers excellent performance across the various architectures utilized.

**Keywords:** gejala covid-19, klasifikasi, word embedding, convolutional neural network, media sosial

## **PRAKATA**

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Performa Convolutional Neural Network 1D dan 2D dengan Fitur Word Embeddings pada Klasifikasi Gejala COVID-19 Dari Pesan Sosial Media” untuk memenuhi syarat dalam menyelesaikan pendidikan program S-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga yaumul qiyamah.

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

1. Kedua orang tua dan keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
2. Bapak M. Reza Faisal, S.T., M.T., PhD selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Rudy Herteno, S.Kom., M.Kom. 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 staf 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. 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 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, 26 Juni 2023

Muhammad Khairie

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK .....	iv
<i>ABSTRACT</i> .....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xi
DAFTAR LAMPIRAN.....	xiii
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
<b>BAB II TINJAUAN PUSTAKA</b>	
2.1 Kajian Terdahulu .....	4
2.2 Keaslian Penelitian .....	6
2.3 Kajian Pustaka .....	9
2.3.1 COVID-19.....	9
2.3.2 Twitter.....	10
2.3.3 Word Embedding .....	10
2.3.4 Word2Vec .....	11
2.3.5 GloVe .....	11
2.3.6 fastText.....	12
2.3.7 Word Padding .....	12
2.3.8 Klasifikasi Teks.....	13



2.3.9	Convolutional Neural Network (CNN)	13
2.3.10	Matriks Evaluasi	14
<b>BAB III METODE PENELITIAN</b>		
3.1	Alat Penelitian	16
3.2	Bahan Penelitian	16
3.3	Variabel Penelitian	16
3.4	Alur Penelitian	16
<b>BAB IV HASIL DAN PEMBAHASAN</b>		
4.1	Hasil	20
4.1.1	Pengumpulan Data	20
4.1.2	Preprocessing Data	21
4.1.3	Word Padding	25
4.1.4	Membangun Model Word Embedding	26
4.1.5	Pembagian Data	28
4.1.6	Tokenisasi	28
4.1.7	Pembuatan Model Klasifikasi	28
4.1.8	Klasifikasi dan Evaluasi	32
4.2	Pembahasan	47
<b>BAB V PENUTUP</b>		
5.1	Kesimpulan	62
5.2	Saran	62
DAFTAR PUSTAKA		
LAMPIRAN		

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
Tabel 1. Keaslian Penelitian.....	6
Tabel 2. Rancangan Penelitian.....	9
Tabel 3. Confusion matrix.....	14
Tabel 4. Contoh data tweet.....	20
Tabel 5. Data Cleaning.....	21
Tabel 6. Proses Case Folding.....	23
Tabel 7. Proses <i>label encoding</i> .....	24
Tabel 8. Word padding.....	26
Tabel 9. Vektor kata Word2vec.....	26
Tabel 10. Vektor kata Glove.....	26
Tabel 11. Vektor kata fastText.....	27
Tabel 12. Vektor kata yang memiliki kemiripan.....	27
Tabel 13. Hasil pengujian klasifikasi pada arsitektur 1D single-channel.....	32
Tabel 14. Hasil pengujian klasifikasi pada arsitektur 1D multi-channel.....	36
Tabel 15. Hasil pengujian klasifikasi pada arsitektur CNN 2D single-channel.....	40
Tabel 16. Hasil pengujian klasifikasi pada arsitektur CNN 2D multi-channel.....	44
Tabel 17. Matriks evaluasi pada arsitektur 1D single-channel.....	48
Tabel 18. Matriks evaluasi pada arsitektur 1D multi-channel.....	49
Tabel 19. Matriks evaluasi pada arsitektur 2D single-channel.....	50
Tabel 20. Matriks evaluasi pada arsitektur 2D multi-channel.....	51

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
Gambar 1. Struktur utama CNN 2D multi-channel .....	14
Gambar 2. Alur Penelitian.....	17
Gambar 3. Arsitektur CNN single-channel yang digunakan .....	31
Gambar 4. Arsitektur multi-channel yang digunakan .....	31
Gambar 5. Kinerja penggunaan <i>word padding</i> max pada 1D single-channel.....	33
Gambar 6. Kinerja penggunaan <i>word padding</i> mean pada 1D single-channel...	34
Gambar 7. Kinerja penggunaan <i>word padding</i> median pada 1D single-channel .....	35
Gambar 8. Kinerja penggunaan <i>word padding</i> mode pada 1D single-channel..	35
Gambar 9. Kinerja penggunaan <i>word padding</i> max pada 1D multi-channel.....	37
Gambar 10. Kinerja penggunaan <i>word padding</i> mean pada 1D multi-channel.	38
Gambar 11. Kinerja penggunaan <i>word padding</i> median pada 1D multi-channel .....	38
Gambar 12. Kinerja penggunaan <i>word padding</i> mode pada 1D multi-channel.	39
Gambar 13. Kinerja penggunaan <i>word padding</i> max pada 2D single-channel..	41
Gambar 14. Kinerja penggunaan <i>word padding</i> mean pada 2D single-channel	42
Gambar 15. Kinerja penggunaan <i>word padding</i> median pada 2D single-channel .....	42
Gambar 16. Kinerja penggunaan <i>word padding</i> mode pada 2D single-channel	43
Gambar 17. Kinerja penggunaan <i>word padding</i> max pada 2D multi-channel...	45
Gambar 18. Kinerja penggunaan <i>word padding</i> mean pada 2D multi-channel.	46
Gambar 19. Kinerja penggunaan <i>word padding</i> median pada 2D multi-channel .....	46
Gambar 20. Kinerja penggunaan <i>word padding</i> mode pada 2D multi-channel.	47
Gambar 21. Perbandingan rata-rata akurasi berdasarkan <i>word embedding</i> pada 1D single-channel.....	52
Gambar 22. Perbandingan rata-rata akurasi berdasarkan <i>word embedding</i> pada 1D multi-channel.....	53

Gambar 23. Perbandingan rata-rata akurasi berdasarkan <i>word embedding</i> pada 2D single-channel.....	53
Gambar 24. Perbandingan rata-rata akurasi berdasarkan <i>word embedding</i> pada 2D multi-channel.....	54
Gambar 25. Perbandingan rata-rata akurasi berdasarkan <i>word embedding</i> .....	55
Gambar 26. Perbandingan rata-rata akurasi berdasarkan <i>word padding</i> pada 1D single-channel .....	56
Gambar 27. Perbandingan rata-rata akurasi berdasarkan <i>word padding</i> pada 1D multi-channel.....	57
Gambar 28. Perbandingan rata-rata akurasi berdasarkan <i>word padding</i> pada 2D single-channel .....	57
Gambar 29. Perbandingan rata-rata akurasi berdasarkan <i>word padding</i> pada 2D multi-channel.....	58
Gambar 30. Perbandingan rata-rata akurasi berdasarkan <i>word padding</i> .....	59
Gambar 31. Perbandingan rata-rata akurasi berdasarkan arsitektur CNN .....	60
Gambar 32. Perbandingan akurasi tertinggi dari setiap arsitektur .....	61

## DAFTAR LAMPIRAN

### Lampiran

Lampiran 1. *Source code import dataset dan preprocessing*

Lampiran 2. *Label encoder, pembagian data, dan Tokenisasi*

Lampiran 3. *Source code perhitungan word padding dan pembuatan matriks*

Lampiran 4. *Source code load model word2vec dan embedding matrix word2vec*

Lampiran 5. *Source code load model glove dan embedding matrix glove*

Lampiran 6. *Source code load model fasttext dan embedding matrix fasttext*

Lampiran 7. *Source code CNN 1D*

Lampiran 8. *Source code CNN 2D*

Lampiran 9. *Source Code Compile model*

Lampiran 10. *Source Code Fit model*