



**KLASIFIKASI GENDER PENGGUNA PADA TWITTER
MENGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL
NETWORK**

Skripsi

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

**Oleh
Fitra Ahya Mubarok
NIM. 1811016210027**

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

JUNI 2023



**KLASIFIKASI GENDER PENGGUNA PADA TWITTER
MENGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL
NETWORK**

Skripsi

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

**Oleh
Fitra Ahya Mubarok
NIM. 1811016210027**

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

JUNI 2023

SKRIPSI
KLASIFIKASI GENDER PENGGUNA PADA TWITTER
MENGGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL
NETWORK

Oleh

Fitra Ahya Mubarak
NIM. 1811016210027

Telah dipertahankan di depan Dosen Penguji pada tanggal 27 Juni 2023

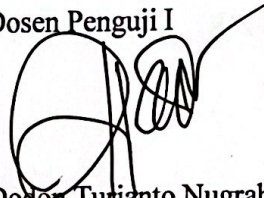
Susunan Dosen Penguji :

Pembimbing I



Mohammad Reza Faisal, S.T., M.T., PhD.
NIP. 19761220 2008121001

Dosen Penguji I



Dosen Turiyanto Nugrahadi S.Kom., M.Eng.
NIP. 198001122009121002

Pembimbing II

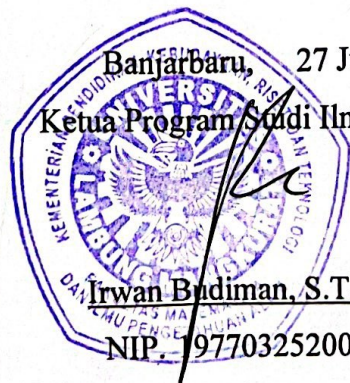


Dwi Kartini, S.Kom., M.Kom.
NIP. 198704212012122003

Dosen Penguji II



Triando Hamonangan Saragih S.Kom M.Kom
NIP. 199308242019031012



Banjarbaru, 27 Juni 2023
Ketua 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, 2023



Fitra Ahya Mubarok
NIM. 1811016210027

ABSTRAK

KLASIFIKASI *GENDER* PENGGUNA PADA TWITTER MENGGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK

(Oleh: Fitra Ahya Mubarak; Pembimbing: Mohammad Reza Faisal, S.T., M.T., PhD. dan Dwi Kartini, S.Kom., M.Kom.; 2023; 66 halaman)

Sosial media sudah menjadi tempat bagi para *sosial media analis* untuk mendapatkan data yang digunakan untuk mendapatkan wawasan dan pemahaman yang lebih dalam tentang perilaku pengguna, tren, opini publik, dan pola-pola yang terkait dengan penggunaan sosial media. *Twitter* adalah salah satu media sosial yang populer di mana pengguna dapat berbagi pesan atau "tweet" dalam format teks singkat. Namun, pada *twitter* informasi pengguna seperti *gender* tidak diperlihatkan tetapi tanpa disadari atau tidak, terdapat informasi mengenai hal tersebut secara tidak terstruktur. Untuk hal tersebut dilakukan klasifikasi *gender* dengan menggunakan pembentukan representasi vektor kata dengan menggunakan *word2vec*, dan penerapan model 2D *Convolutional Neural Network* (CNN). *Word2vec* digunakan untuk menghasilkan representasi vektor kata yang memperhitungkan konteks dan makna kata dalam teks. Model 2D *CNN* digunakan untuk mengekstraksi fitur dari representasi vektor kata dan melakukan klasifikasi *gender*. Pada penelitian ini dilakukan perbandingan antara data tweet, deskripsi, dan gabungan keduanya yaitu tweet dan deskripsi yang bertujuan untuk mencari akurasi terbaik. Hasil penelitian menunjukkan bahwa menggunakan data gabungan antara tweet dan deskripsi mendapatkan akurasi yang lebih baik daripada menggunakan data tweet atau deskripsi saja.

Kata kunci: 2D CNN, *word2vec*, klasifikasi *gender*, Twitter, analisis sosial media

ABSTRACT

GENDER CLASSIFICATION OF TWITTER USERS USING CONVOLUTIONAL NEURAL NETWORK ALGORITHM

(By: Fitra Ahya Mubarok; Advisors: Mohammad Reza Faisal, S.T., M.T., PhD. and Dwi Kartini, S.Kom., M.Kom.; 2023; 66 pages)

Social media has become a platform for social media analysts to gather data used to gain deeper insights and understanding of user behavior, trends, public opinions, and patterns related to social media usage. Twitter is one popular social media platform where users can share messages or "tweets" in short text format. However, on Twitter, user information such as gender is not explicitly displayed, but unintentionally, there is unstructured information regarding this. Therefore, gender classification is performed using the formation of word vector representations using word2vec and the application of a 2D Convolutional Neural Network (CNN) model. Word2vec is used to generate word vector representations that take into account the context and meaning of words in the text. The 2D CNN model is used to extract features from word vector representations and perform gender classification. This research compares tweet data, description data, and the combination of both (tweets and descriptions) to find the best accuracy. The results of the study indicate that using the combined data of tweets and descriptions achieves better accuracy compared to using tweet data or description data alone.

Keywords: *2D CNN, word2vec, gender classification, Twitter, social media analysis*

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 “Klasifikasi *Gender* Pengguna Pada Twitter Menggunakan Algoritma Convolutional Neural Network” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan sholawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiamah*.

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. Diri saya sendiri yang tidak pernah patah semangat walaupun banyak menemui kesulitan baik disebabkan oleh diri sendiri maupun dari hal lain.
2. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
3. Bapak Mohammad 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.
4. Ibu Dwi Kartini S.Kom, M.Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
5. Bapak Irwan Budiman, S.T., M.Kom selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
6. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
7. Venny yang selalu memberikan semangat, menemani, membantu dan juga mengingatkan untuk mengerjakan skripsi.

8. Teman-teman keluarga Ilmu Komputer angkatan 2018 yang telah lulus lebih dahulu dan juga yang belum sehingga memberikan motivasi untuk menyelesaikan skripsi ini dan juga tempat bertanya disaat kebingungan.
9. 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, 2023

Fitra Ahya Mubarok

DAFTAR ISI

PERNYATAAN.....	ii
ABSTRAK.....	iii
ABSTRACT.....	iv
PRAKATA.....	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Terdahulu.....	5
2.2 Twitter.....	7
2.3 Teks Mining.....	7
2.4 Word2Vec.....	8
2.5 Convolutional Neural Network (CNN).....	9
2.6 K-fold Cross Validation.....	11
2.7 Performa Klasifikasi.....	11
2.8 Keaslian Penelitian.....	13
BAB III METODE PENELITIAN.....	17
3.1 Alat Penelitian.....	17
3.2 Bahan Penelitian.....	17
3.3 Variabel Penelitian.....	17
3.4 Prosedur Penelitian.....	17
3.4.1 Pengumpulan Data.....	18
3.4.2 Preprocessing.....	20
3.4.3 Pembuatan Model Word2Vec.....	20
3.4.4 K-Fold Cross Validation.....	21

3.4.5	Ekstraksi Fitur Word2Vec.....	21
3.4.6	Pembuatan Model Klasifikasi	21
3.4.7	Evaluasi	21
BAB IV	HASIL DAN PEMBAHASAN	22
4.1	Hasil	22
4.1.1	Pengumpulan Data	22
4.1.2	Preprocessing Data.....	24
4.1.3	Pembuatan Model Word2Vec	33
4.1.4	K-Fold Cross Validation	34
4.1.5	Ekstraksi Fitur Word2Vec.....	34
4.1.6	Pembuatan Model Klasifikasi	35
4.1.7	Klasifikasi dan Evaluasi	38
4.2	Pembahasan	60
4.2.1	Pengaruh <i>Parameter</i> pada model pengujian	64
BAB V	PENUTUP.....	68
5.1	Kesimpulan	68
5.2	Saran.....	68
	Daftar Pustaka	69

DAFTAR TABEL

Tabel 1. Perbandingan data berdasarkan label	3
Tabel 2. Confusion Matrix	12
Tabel 3. Keaslian Penelitian.....	13
Tabel 4. Tabel Rancangan Penelitian.....	16
Tabel 5. Contoh data twitter dan deskripsi.....	19
Tabel 6. Contoh data twitter	23
Tabel 7. Jumlah pembagian data.....	24
Tabel 8. Data tweet yang telah diberi label.....	26
Tabel 9. Cleansing dan Case folding pada data	28
Tabel 10. Stemming dan Stopword Removal pada data	31
Tabel 11. Model pretrained Word2Vec	34
Tabel 12 Ekstraksi fitur dengan jumlah kata maksimal 30.....	40
Tabel 13 Ekstraksi fitur dengan jumlah kata 8.....	41
Tabel 14. Hasil kinerja CNN untuk klasifikasi deskripsi.....	41
Tabel 15. Hasil kinerja CNN untuk klasifikasi deskripsi.....	42
Tabel 16. Hasil kinerja CNN untuk klasifikasi deskripsi.....	43
Tabel 17. Hasil kinerja CNN untuk klasifikasi deskripsi.....	44
Tabel 18 Ekstraksi fitur dengan jumlah kata 12.....	47
Tabel 19. Hasil kinerja CNN untuk klasifikasi tweet	47
Tabel 20. Hasil kinerja CNN untuk klasifikasi tweet	48
Tabel 21. Hasil kinerja CNN untuk klasifikasi tweet	49
Tabel 22. Hasil kinerja CNN untuk klasifikasi tweet	50
Tabel 23. Hasil kinerja CNN untuk klasifikasi tweet dan deskripsi	53
Tabel 24 Hasil kinerja CNN untuk klasifikasi tweet dan deskripsi	54
Tabel 25. Hasil kinerja CNN untuk klasifikasi tweet dan deskripsi	55
Tabel 26. Hasil kinerja CNN untuk klasifikasi tweet dan deskripsi	56
Tabel 27. Hasil kinerja CNN untuk klasifikasi tweet dan deskripsi	56
Tabel 28. Hasil kinerja CNN untuk klasifikasi tweet dan deskripsi	57
Tabel 29. Hasil kinerja CNN untuk klasifikasi tweet dan deskripsi	58
Tabel 30. Hasil kinerja CNN untuk klasifikasi tweet dan deskripsi	59
Tabel 31 Hasil akurasi berdasarkan kernel size	64
Tabel 32. Hasil akurasi berdasarkan ukuran input.....	66

DAFTAR GAMBAR

Gambar 1 Vektor Word2Vec berdasarkan makna semantik.....	8
Gambar 2 Mutiple Context Words Skip-Gram (Rong, 2016).....	9
Gambar 3 Model Deep Convolutional Neural Network	11
Gambar 4 Alur Penelitian.....	18
Gambar 5 Tahapan pembuatan model word embedding Word2Vec.....	33
Gambar 6 Model Convolutional Neural Network menggunakan data deskripsi ...	39
Gambar 7 Model summary Convolutional Neural Network menggunakan data deskripsi.....	40
Gambar 8 Model Convolutional Neural Network menggunakan data tweet.....	45
Gambar 9 Model summary Convolutional Neural Network menggunakan data tweet.....	46
Gambar 10 Model Convolutional Neural Network menggunakan data deskripsi dan tweet.....	51
Gambar 11 Model summary Convolutional Neural Network menggunakan data tweet.....	52
Gambar 12 2 Channel CNN	52
Gambar 13 1 Channel CNN	53
Gambar 14 Grafik perbandingan akurasi pengujian antara deskripsi, tweet, dan deskripsi dan tweet dengan kernel size 5x1.....	60
Gambar 15 Grafik perbandingan akurasi pengujian antara deskripsi, tweet, dan deskripsi dan tweet dengan nilai input maksimal dan kernel size 5x1	61
Gambar 16 Grafik perbandingan akurasi pengujian antara deskripsi, tweet, dan deskripsi dan tweet dengan kernel size 5,5.....	62
Gambar 17 Grafik perbandingan akurasi pengujian antara deskripsi, tweet, dan deskripsi dan tweet dengan kernel size 5x5.....	63
Gambar 18 Hasil Pengaruh Kernel Size.....	65
Gambar 19 Hasil Akurasi Penambahan 5 Epoch	67