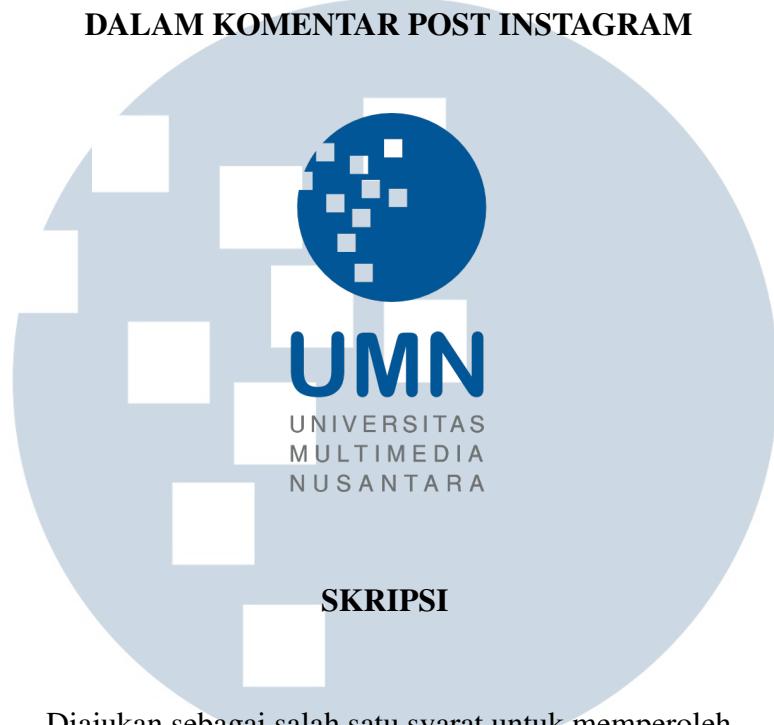


**IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK
UNTUK KLASIFIKASI JENIS CYBERBULLYING
DALAM KOMENTAR POST INSTAGRAM**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Syauqi Damario Djohan
00000022467

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA
PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2022

**IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK
UNTUK KLASIFIKASI JENIS CYBERBULLYING
DALAM KOMENTAR POST INSTAGRAM**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)



HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Syauqi Damario Djohan
Nomor Induk Mahasiswa : 00000022467
Program Studi : Informatika

Skripsi dengan judul:

Implementasi Convolutional Neural Network untuk Klasifikasi Jenis Cyberbullying dalam Komentar Post Instagram

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 17 Juni 2022




(Syauqi Damario Djohan)

UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK UNTUK KLASIFIKASI JENIS CYBERBULLYING DALAM KOMENTAR POST INSTAGRAM

oleh

Nama : Syauqi Damario Djohan
NIM : 00000022467
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Selasa, 21 Juni 2022

Pukul 10.00 s/d 12.00 dan dinyatakan

LULUS

Dengan susunan pengaji sebagai berikut

Ketua Sidang



(Yaman Khaeruzzaman, M.Sc.)

NIDN: 0413057104

Pembimbing I

(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)

NIDN: 0311106903

Pengaji

(Alethea Suryadibrata, S.Kom., M.Eng.)

NIDN: 0322099201

Pembimbing II



(Marlinda Vasty Overbeek, S.Kom., M.Kom.)

NIDN: 0818038501

Ketua Program Studi Informatika,



Digitally signed by
Marlinda Vasty
Overbeek
Date: 2022.07.01
05:18:27 +07'00'

(Marlinda Vasty Overbeek, S.Kom., M.Kom.)

NIDN: 0818038501

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama	:	Syauqi Damario Djohan
NIM	:	00000022467
Program Studi	:	Informatika
Fakultas	:	Teknik dan Informatika
Jenis Karya	:	Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada **Universitas Multimedia Nusantara** hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK UNTUK KLASIFIKASI JENIS CYBERBULLYING DALAM KOMENTAR POST INSTAGRAM

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 17 Juni 2022

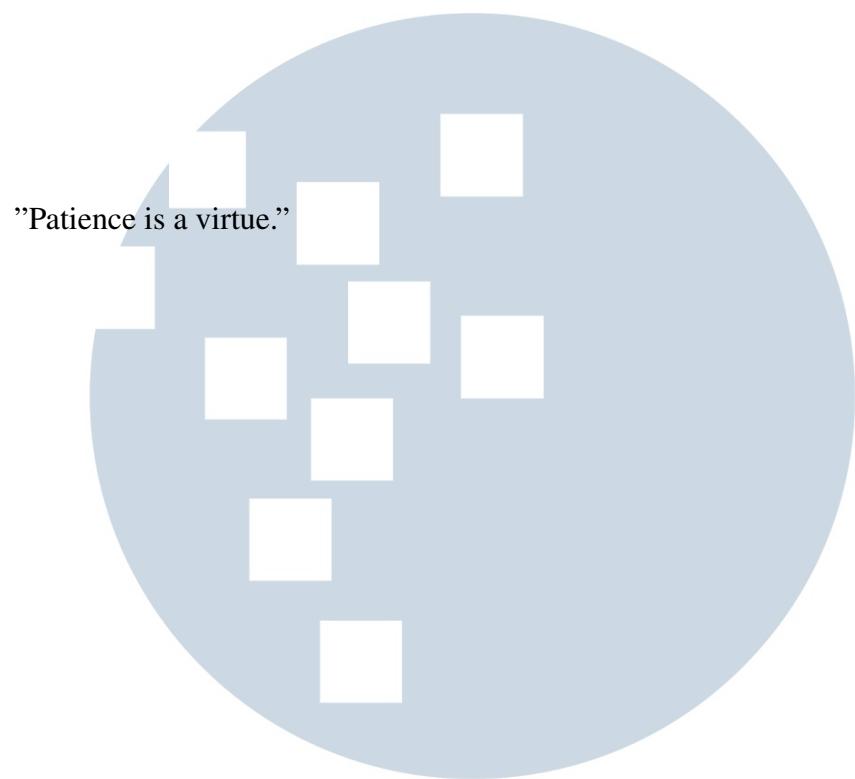
Yang menyatakan



Syauqi Damario Djohan

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Halaman Persembahan / Motto



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Implementasi Convolutional Neural Network untuk Klasifikasi Jenis Cyberbullying dalam Komentar Post Instagram dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

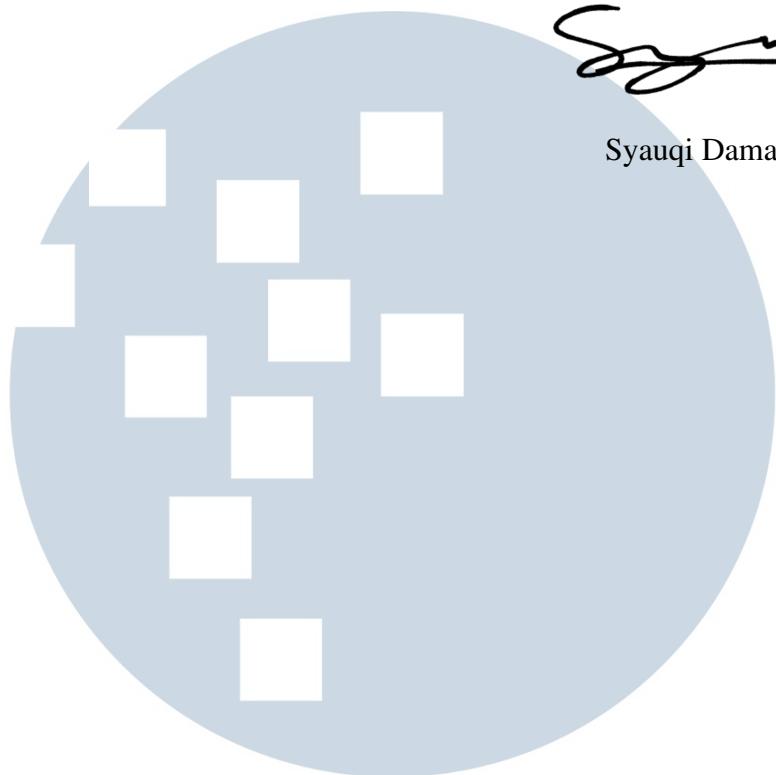
1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Moeljono Widjaja, B.Sc., M.Sc., Ph.D., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesaiannya Skripsi ini.
5. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., sebagai Pembimbing kedua yang telah banyak membantu dan memberikan bimbingan atas terselesaiannya Skripsi ini.
6. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan Skripsi ini.
7. Teman-teman, baik yang dari dan bukan dari Universitas Multimedia Nusantara yang satu program studi, serta senior-senior program studi Informatika yang telah memberi masukan-masukan.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 17 Juni 2022



Syauqi Damario Djohan



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

**IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK
UNTUK KLASIFIKASI JENIS CYBERBULLYING
DALAM KOMENTAR POST INSTAGRAM**

Syauqi Damario Djohan

ABSTRAK

Bullying merupakan suatu masalah yang memang sudah marak, akan tetapi masalah ini menjadi semakin marak di era digital. Munculnya *cyberbullying*, perilaku perundungan secara daring, membuat *bullying* bukan hanya terjadi secara fisik saja, tetapi juga melalui media sosial lewat berbagai bahasa dan ekspresi, dengan salah satu contohnya adalah lewat media sosial Twitter. Ekspresi *bullying* yang bermacam-macam bentuknya di Twitter ini harus diketahui makna dibalik perkataannya apa, sehingga dibutuhkan suatu cara untuk mengenali apa jenis *cyberbullying* tersebut lewat sentiment analysis. Dari itu, digunakanlah metode *convolutional neural network* sebagai metode *sentiment analysis* menggunakan *word embedding* lewat *pre-trained model* FastText. Hasil penelitian tersebut menunjukkan bahwa program ini memberikan akurasi sebesar 89% dalam mendekripsi *tweets* (cuitan dalam Twitter), dengan *loss* sebesar 29%. Penelitian ini menunjukkan bahwa *cyberbullying* yang diekspresikan lewat *tweets* dapat diklasifikasikan menggunakan *convolutional neural network* dengan baik, tetapi dapat ditingkatkan dengan metode-metode tambahan di masa mendatang untuk menghasilkan akurasi yang lebih tinggi dan *loss* yang lebih rendah.

Kata kunci: *analisis sentimen, cyberbullying, convolutional neural network, Twitter*



IMPLEMENTATION OF CONVOLUTIONAL NEURAL NETWORK FOR CLASSIFYING KINDS OF CYBERBULLYING ON INSTAGRAM POST COMMENTS

Syauqi Damario Djohan

ABSTRACT

Bullying is one of the biggest concerns in the world, however, this problem is becoming increasingly common in the digital era. The emergence of cyberbullying, defined as online bullying behavior, makes bullying not only happen physically or verbally but also through social media using various languages and expressions, in which Twitter is one of the social media platforms. The expressions of bullying which take multiple forms on Twitter's meanings behind the words have to be known, so it is necessary to have a way to identify what type of cyberbullying is through sentiment analysis. From that, the convolutional neural network method is used as a sentiment analysis method using word embedding via the pre-trained FastText model. The results of this study show this program provides an accuracy of 89% in detecting tweets (tweets on Twitter), with a loss of 29%. This study shows that cyberbullying expressed through tweets can be classified using a convolutional neural network well, but can be improved by additional methods in the future to produce higher accuracy and lower loss.

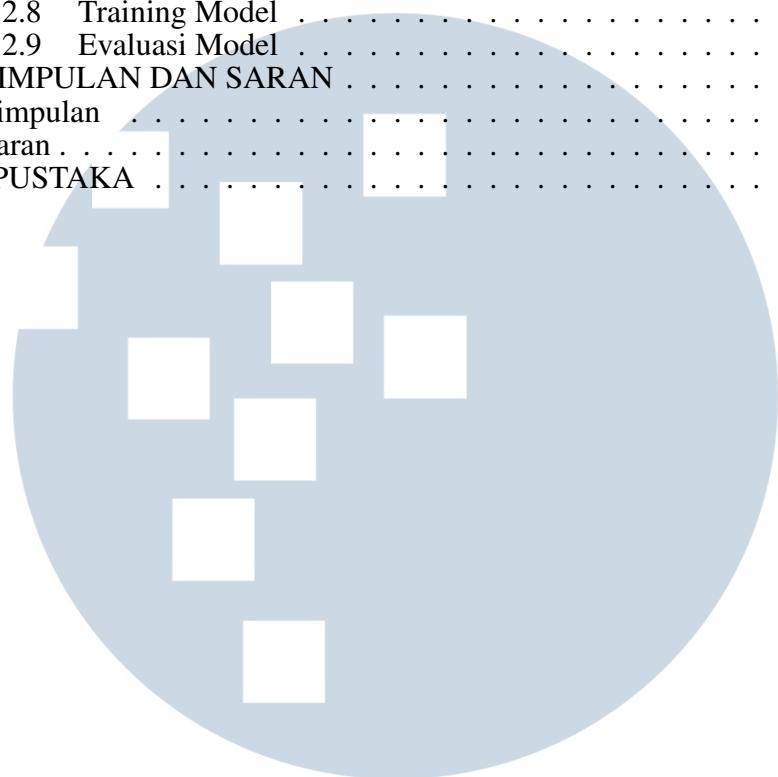
Keywords: *cyberbullying, convolutional neural network, sentiment analysis, Twitter*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR KODE	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Permasalahan	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB 2 LANDASAN TEORI	4
2.1 Twitter	4
2.2 Analisis Sentimen	4
2.3 Cyberbullying	4
2.4 Pembelajaran Mesin	6
2.5 Text Preprocessing	8
2.5.1 Tokenization	8
2.5.2 Stopword Removal	9
2.5.3 Case Folding	9
2.5.4 Noise Removal	9
2.5.5 Menghilangkan Angka	9
2.5.6 Stemming	10
2.6 Word Embedding	10
2.7 FastText	10
2.8 Multilayer Perceptron	11
2.9 Convolutional Neural Network	13
2.10 Evaluasi Model	14
BAB 3 METODOLOGI PENELITIAN	17
3.1 Metodologi Penelitian	17
3.2 Perancangan Sistem	18
BAB 4 HASIL DAN DISKUSI	24
4.1 Spesifikasi Perangkat	24
4.2 Implementasi Metode	25
4.2.1 Load dan Tampilkan Dataset	25
4.2.2 Preprocessing Dataset	30
4.2.3 Inisialisasi Variabel Fitur & Target	36
4.2.4 Word Embedding	36
4.2.5 Encoding Label	38

4.2.6	Pembagian Data	38
4.2.7	Arsitektur Model	39
4.2.8	Training Model	40
4.2.9	Evaluasi Model	44
BAB 5	SIMPULAN DAN SARAN	47
5.1	Simpulan	47
5.2	Saran	47
DAFTAR PUSTAKA		49



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

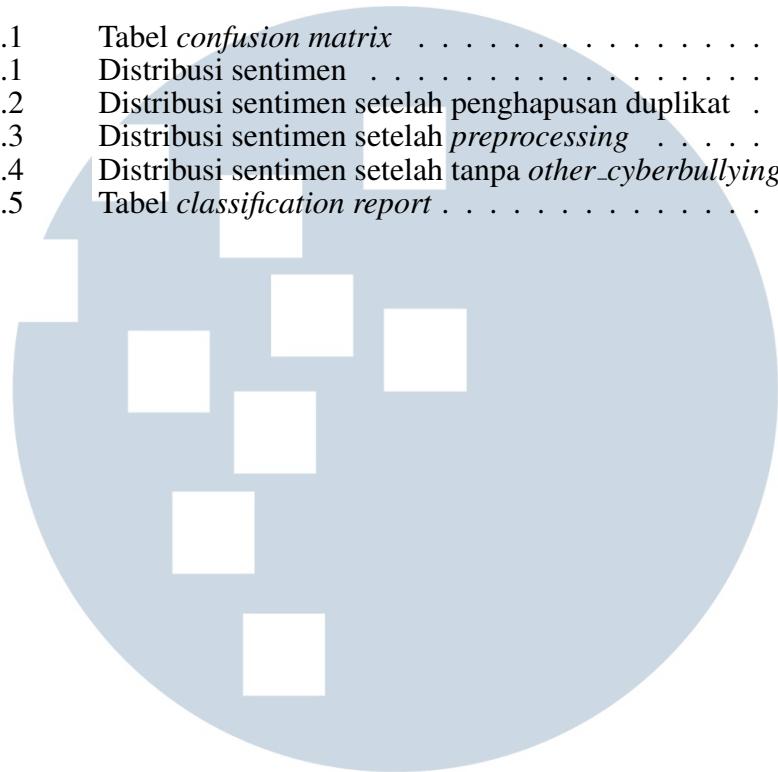
DAFTAR GAMBAR

Gambar 2.1	Langkah-langkah dalam pembelajaran mesin	6
Gambar 2.2	Ilustrasi langkah-langkah <i>text preprocessing</i>	8
Gambar 2.3	Ilustrasi proses <i>tokenization</i>	9
Gambar 2.4	Arsitektur MLP	11
Gambar 2.5	Arsitektur CNN	14
Gambar 3.1	<i>Flowchart</i> utama	18
Gambar 3.2	<i>Flowchart preprocessing</i>	19
Gambar 3.3	<i>Flowchart word embedding</i>	20
Gambar 3.4	<i>Flowchart</i> untuk <i>encoding</i> sentimen	20
Gambar 3.5	<i>Flowchart</i> pembagian data	21
Gambar 3.6	<i>Flowchart</i> pelatihan data	22
Gambar 3.7	<i>Flowchart</i> evaluasi model	23
Gambar 4.1	Lima data pertama dalam dataset	26
Gambar 4.2	Grafik distribusi sentimen <i>tweet</i>	27
Gambar 4.3	Grafik distribusi sentimen <i>tweet</i> setelah penghapusan duplicat	28
Gambar 4.4	<i>WordCloud</i> kalimat sentimen positif	29
Gambar 4.5	<i>WordCloud</i> kalimat sentimen negatif	30
Gambar 4.6	Grafik distribusi sentimen <i>tweet</i> tanpa <i>other_cyberbullying</i>	34
Gambar 4.7	<i>WordCloud</i> negatif baru	35
Gambar 4.8	<i>WordCloud</i> positif baru	35
Gambar 4.9	Diagram akurasi sampai 117 epoch	43
Gambar 4.10	Diagram <i>loss</i> sampai 117 epoch	44
Gambar 4.11	<i>Confusion Matrix</i> untuk lima kelas	46



DAFTAR TABEL

Tabel 2.1	Tabel <i>confusion matrix</i>	15
Tabel 4.1	Distribusi sentimen	27
Tabel 4.2	Distribusi sentimen setelah penghapusan duplikat	28
Tabel 4.3	Distribusi sentimen setelah <i>preprocessing</i>	33
Tabel 4.4	Distribusi sentimen setelah tanpa <i>other_cyberbullying</i>	34
Tabel 4.5	Tabel <i>classification report</i>	46



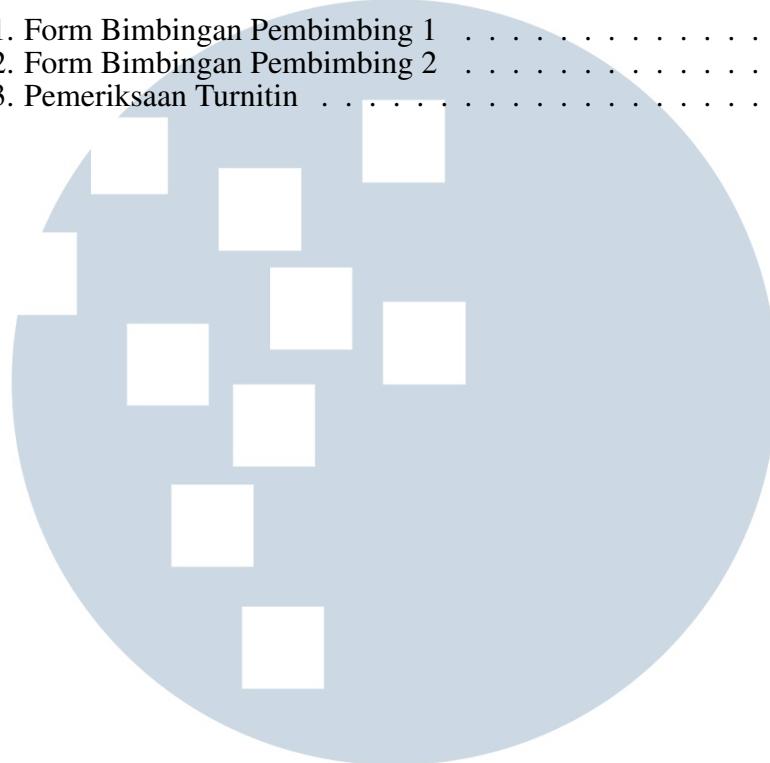
UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR KODE

4.1	Kode yang membaca dan menampilkan dataset	26
4.2	Kode untuk memeriksa data duplikat dan menghapus data duplikat	27
4.3	Kode untuk membedakan <i>tweet</i> sesuai sentimen dan membuat <i>WordCloud tweet negatif</i>	28
4.4	Potongan kode <i>function</i> untuk membersihkan emoji	30
4.5	Potongan kode <i>function</i> untuk membersihkan teks dari <i>noise</i>	30
4.6	Potongan kode <i>function</i> untuk menghilangkan singkatan-singkatan dari kata-kata	31
4.7	Potongan kode <i>function</i> untuk menghilangkan tagar/ <i>hashtag</i>	31
4.8	Potongan kode <i>function</i> untuk menghilangkan tanda-tanda spesial dan spasi berurutan yang berjumlah banyak	31
4.9	Potongan kode untuk melakukan <i>stemming</i>	32
4.10	<i>Function</i> untuk memanggil <i>function</i> lain yang telah dibuat untuk melakukan <i>preprocessing</i> dan menggabungkan kata-kata kembali menjadi <i>tweets</i>	32
4.11	Potongan kode untuk menyimpan teks yang sudah di- <i>preprocessing</i>	33
4.12	Potongan kode untuk memeriksa keberadaan data duplikat setelah <i>preprocessing</i>	33
4.13	Potongan kode untuk menghilangkan tipe <i>cyberbullying</i> lain	34
4.14	<i>Filtering</i> kata-kata dalam <i>tweet</i> setelah <i>preprocessing</i> dan menghilangkan <i>other_cyberbullying</i>	34
4.15	Kode untuk inisialisasi variabel fitur dan target	36
4.16	Potongan kode untuk menggunakan model <i>pre-trained</i> FastText untuk melakukan <i>word embedding</i>	36
4.17	Potongan kode untuk menggunakan objek Tokenizer()	37
4.18	Potongan kode untuk mengubah teks dalam objek Tokenizer menjadi <i>integer</i> sebelum diubah menjadi <i>list</i>	37
4.19	Potongan kode untuk melakukan <i>padding</i> pada <i>sequence</i>	37
4.20	Potongan kode untuk mengubah target menjadi <i>integer</i> menggunakan <i>function</i> LabelEncoder dan OneHotEncoder	38
4.21	Potongan kode pembagian data	38
4.22	Potongan kode perancangan arsitektur <i>Convolutional Neural Network</i>	40
4.23	Potongan kode kompilasi model	42
4.24	Potongan kode pembuatan <i>checkpoint</i> menggunakan <i>function</i> ModelCheckpoint()	42
4.25	Potongan kode implementasi <i>function</i> EarlyStopping()	42
4.26	Potongan kode untuk pelatihan model menggunakan 420 <i>epoch</i>	42
4.27	Potongan kode untuk menampilkan akurasi dan <i>loss</i> dalam <i>testing</i> dan <i>training</i>	44
4.28	Melakukan prediksi pada data	45
4.29	Potongan kode untuk menampilkan <i>classification report</i>	45
4.30	Potongan kode untuk menampilkan <i>confusion matrix</i>	45

DAFTAR LAMPIRAN

Lampiran 1. Form Bimbingan Pembimbing 1	52
Lampiran 2. Form Bimbingan Pembimbing 2	54
Lampiran 3. Pemeriksaan Turnitin	57



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA