

**PENGGUNAAN DEEP LEARNING UNTUK LEXICON-BASED
SENTIMENT ANALYSIS TRAGEDI KANJURUHAN PADA
MEDIA SOSIAL TWITTER**

SKRIPSI



Oleh:

ARIF WIDIASAN SUBAGIO

19081010065

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2024

**PENGGUNAAN DEEP LEARNING UNTUK LEXICON-BASED
SENTIMENT ANALYSIS TRAGEDI KANJURUHAN PADA
MEDIA SOSIAL TWITTER**

SKRIPSI

Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam Menempuh Gelar Sarjana
Komputer Program Studi Informatika



Oleh:

ARIF WIDIASAN SUBAGIO

19081010065

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2024

LEMBAR PENGESAHAN SKRIPSI

Judul : PENGGUNAAN DEEP LEARNING UNTUK LEXICON-BASED SENTIMENT ANALYSIS TRAGEDI KANJURUHAN PADA MEDIA SOSIAL TWITTER

Oleh : Arif Widiason Subagio

NPM : 19081010065

**Telah Diseminarkan Dalam Ujian Skripsi Pada :
Hari Jumat, Tanggal 5 Januari 2024**

Mengetahui

1. **Dosen Pembimbing**

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT
NPT. 222198 60 816400

1. **Dosen Penguji**

Dr. I Gede Susrama Mas Diyasa, ST., MT
NIP. 19700619 2021211 009

2.

Andreas Nugroho Sihananto, S.Kom., M.Kom
NPT. 211199 00 412271

2.

Afina Lina Nurlaili, S.Kom., M.Kom
NIP. 1993121 3202203 2010

Menyetujui

**Dekan
Fakultas Ilmu Komputer**

Prof. Dr. Ir. Noyirina Hendrasarie, MT
NIP. 19681126 199403 2 001

**Koordinator Program Studi
Informatika**

Fetty Tri Anggrachy, S.Kom., M.Kom
NIP. 19820211 2021212 005

SURAT PERNYATAAN ORISINALITAS

Saya mahasiswa Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : ARIF WIDIASAN SUBAGIO

NPM : 19081010065

Dengan ini menyatakan bahwa judul skripsi atau tugas akhir yang saya ajukan dan kerjakan, yang berjudul:

“PENGUNAAN DEEP LEARNING UNTUK LEXICON-BASED SENTIMENT ANALYSIS TRAGEDI KANJURUHAN PADA MEDIA SOSIAL TWITTER”

Bukan merupakan plagiat dari skripsi atau tugas akhir maupun penelitian orang lain dan juga bukan merupakan produk atau software yang saya beli dari pihak lain. Saya juga menyatakan bahwa skripsi ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam daftar pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun institusi pendidikan lainnya.

Jika ternyata kemudian hari pernyataan ini terbukti tidak benar, maka Saya siap menerima segala konsekuensinya.

Surabaya, 5 Januari 2024

Hormat Saya,



Arif Widiaskan Subagio
NPM. 19081010065

PENGGUNAAN DEEP LEARNING UNTUK LEXICON-BASED SENTIMENT ANALYSIS TRAGEDI KANJURUHAN PADA MEDIA SOSIAL TWITTER

Nama Mahasiswa : Arif Widiawan Subagio

NPM : 19081010065

Program Studi : Informatika

Dosen Pembimbing : Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.

Andreas Nugroho Sihananto, S.Kom., M.Kom.

ABSTRAK

Twitter menjadi salah satu platform media sosial yang cukup berpengaruh bagi penggunaannya di Indonesia. Banyak pengguna Twitter di Indonesia memposting kicauan atau *tweet* mengenai perasaan atau pendapat mereka terhadap suatu hal yang terjadi di Indonesia, salah satunya adalah peristiwa tragedi Kanjuruhan. Pada tugas akhir ini memiliki tujuan untuk melakukan penerapan *deep learning* dalam melakukan *lexicon-based sentiment analysis* tragedi Kanjuruhan pada media sosial Twitter. Metode yang digunakan adalah *Multilayer Perceptron* (MLP) dan *Convolutional Neural Network* (CNN).

Uji coba yang dilakukan melewati beberapa alur seperti pengambilan data di Twitter lalu dibersihkan. Dalam pembuatan model MLP dan CNN terdapat berbagai variasi konfigurasi untuk mendapatkan model yang terbaik sehingga hasil yang diberikan nanti akan lebih akurat.

Hasil dari uji coba ini menunjukkan bahwa model metode CNN memiliki performa lebih baik daripada metode MLP dengan tingkat akurasi metode CNN sekitar 87.77%, sedangkan metode MLP memiliki tingkat akurasi sekitar 83.9%. Metode CNN terbukti dapat dijadikan salah satu opsi untuk melakukan *lexicon-based sentiment analysis*.

Kata Kunci : Twitter, Tragedi Kanjuruhan, Lexicon, Analisis Sentimen, CNN, MLP

KATA PENGANTAR

Dengan penuh rasa syukur, penulis ingin mengucapkan terima kasih kepada Allah SWT atas berkat dan rahmat-Nya yang telah memungkinkan penyelesaian skripsi berjudul "Penggunaan *Deep Learning* untuk *Lexicon-Based Sentiment Analysis* Tragedi Kanjuruhan pada Media Sosial Twitter."

Laporan skripsi ini disusun dalam rangka memenuhi persyaratan mata kuliah skripsi dan sebagai salah satu langkah penting dalam menyelesaikan pendidikan tingkat Sarjana (S1) di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Penulis sadar akan adanya keterbatasan dalam penulisan skripsi ini, dan dengan rendah hati menerima saran serta kritik yang membangun. Semoga skripsi ini dapat memberikan manfaat yang signifikan bagi semua pihak, baik bagi pembaca maupun untuk penelitian mendatang.

Suarabaya, 5 Januari 2024

Penulis

UCAPAN TERIMA KASIH

Penulisan laporan skripsi ini tidak dapat terwujud tanpa bantuan, motivasi, dan dukungan dari berbagai pihak. Oleh karena itu, dengan tulus, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Dr. Ir. Novirina Hendrasarie, M.T., selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom., selaku Koordinator Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur dan Dosen Wali penulis.
4. Ibu Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT., selaku Dosen Pembimbing I yang meluangkan waktu, tenaga, serta pikiran untuk membimbing dan mengarahkan penulis selama proses penyelesaian skripsi ini.
5. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom., selaku Dosen Pembimbing II yang telah memberikan arahan, dukungan, serta saran kepada penulis sehingga penulis dapat menyelesaikan penyusunan skripsi ini.
6. Seluruh Dosen Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memberikan ilmunya kepada penulis selama perkuliahan.
7. Kedua orang tua penulis, Bapak Saeku Subagio dan Ibu Widiyati yang selalu memberikan dukungan dan motivasi kepada penulis selama perkuliahan.
8. Muhammad Dafa Ardiansyah, Dionisius Damarta Yapenrui, Joni Bastian, dan seluruh teman – teman dari program studi Informatika angkatan 2019 yang telah memberikan bantuan dan dukungan kepada penulis dalam menempuh perkuliahan dan menyelesaikan skripsi.

Penulis hanya bisa berharap, semoga Tuhan Yang Maha Esa senantiasa memberi perlindungan dan membalas semua kebaikan yang telah diberikan.

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
SURAT PERNYATAAN ORISINALITAS	ii
ABSTRAK	iii
KATA PENGANTAR	iv
UCAPAN TERIMA KASIH.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	x
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan.....	5
1.4. Manfaat.....	5
1.5. Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	6
2.1. Penelitian Terdahulu.....	6
2.2. Analisis Sentimen.....	8
2.3. Twitter	8
2.4. Pandas.....	8
2.5. Scikit-learn	9
2.6. Tensorflow.....	9
2.7. <i>Text Mining</i>	9
2.8. <i>Text Preprocessing</i>	9
2.8.1. <i>Cleansing</i>	10
2.8.2. <i>Case-folding</i>	10
2.8.3. <i>Tokenizing</i>	11
2.8.4. <i>Stopword Removal</i>	11
2.8.5. <i>Stemming</i>	12
2.8.6. <i>Word Cloud</i>	12
2.9. <i>Lexicon-based</i>	12
2.10. <i>Neural Network</i>	13

2.11.	<i>Deep Learning</i>	15
2.11.1.	<i>Multilayer Perceptron</i>	16
2.11.2.	<i>Convolutional Neural Network</i>	17
2.12.	Fungsi Aktivasi	17
2.12.1.	Fungsi <i>ReLU</i>	18
2.12.2.	Fungsi <i>Softmax</i>	18
2.13.	<i>Optimizer Adam</i>	19
2.14.	<i>Confusion Matrix</i>	19
BAB III METODOLOGI PENELITIAN		22
3.1.	Deskripsi Studi Kasus	22
3.2.	Tahapan Penelitian	22
3.3.	Analisa dan Desain	23
3.4.	Akuisisi Data	24
3.5.	Text Preprocessing	25
3.5.1.	Cleansing.....	26
3.5.2.	Case-folding	26
3.5.3.	Tokenizing.....	27
3.5.4.	Stopword Removal.....	27
3.5.5.	Stemming	28
3.6.	<i>Lexicon-based Sentiment Analysis</i>	28
3.7.	Klasifikasi Data	30
3.7.1.	<i>Multilayer Perceptron</i>	31
3.7.2.	<i>Convolutional Neural Network</i>	32
3.8.	Pembuatan Model.....	32
3.8.1.	Set Split Data	33
3.8.2.	Set Layer Model.....	33
3.8.3.	Set Epoch	33
3.8.4.	Set Learning Rate.....	34
3.8.5.	Pelatihan dan Pengujian Model.....	34
3.8.6.	Evaluasi Model.....	34
3.9.	Skenario Uji Coba	34

BAB IV	HASIL DAN PEMBAHASAN	36
4.1.	Proses Akuisisi Data.....	36
4.2.	Text Preprocessing	37
4.2.1.	Cleansing.....	38
4.2.2.	Case-folding	40
4.2.3.	Tokenize, Stopword Removal, dan Stemming.....	41
4.2.4.	Analisa Kata/ <i>Word Cloud</i>	43
4.2.5.	Penanganan Data Kosong dan Duplikat.....	46
4.3.	Lexicon-Based Sentiment Analysis.....	46
4.4.	Balancing Dataset.....	48
4.5.	Skema Uji Coba.....	49
4.5.1.	Skenario.....	49
4.5.2.	Skenario Lanjutan	57
4.6.	Hasil Pelatihan dan Pengujian Model	57
4.6.1.	<i>Split Data</i>	57
4.6.2.	<i>Set Hidden Layer</i>	59
4.6.3.	<i>Epoch</i>	60
4.6.4.	<i>Learning Rate</i>	62
4.7.	Hasil dan Analisis Uji.....	63
BAB V	PENUTUP	68
5.1.	Kesimpulan.....	68
5.2.	Saran	69
DAFTAR PUSTAKA	70
Lampiran I	74
Lampiran II	83
Lampiran III	90
Lampiran IV	97
Lampiran V	122
Biodata Penulis	144

DAFTAR GAMBAR

Gambar 2.1 Struktur <i>Neuron</i> dari <i>Neural Network</i>	13
Gambar 2.2 Model <i>Neuron</i>	14
Gambar 2.3 Arsitektur <i>Neural Network</i>	15
Gambar 2.4 Perbedaan dari NN dan <i>Deep Learning</i> NN.....	15
Gambar 2.5 Ilustrasi Arsitektur MLP	16
Gambar 2.6 Arsitektur CNN	17
Gambar 2.7 Grafik Aktivasi ReLU	18
Gambar 2.8 Grafik Aktivasi <i>Softmax</i>	19
Gambar 3.1 Desain Alur Sistem.....	23
Gambar 3.2 <i>Flowchart</i> Akuisisi Data	24
Gambar 3.3 <i>Flowchart</i> <i>Preprocessing</i> Data.....	26
Gambar 3.4 <i>Flowchart</i> Pembobotan <i>Lexicon-based</i>	29
Gambar 3.5 <i>Flowchart</i> Klasifikasi Data	30
Gambar 3.6 <i>Flowchart</i> MLP	31
Gambar 3.7 <i>Flowchart</i> CNN.....	32
Gambar 3.8 <i>Flowchart</i> Pembuatan Model.....	32
Gambar 4.1 Visualisasi <i>Word Cloud</i>	45
Gambar 4.2 Plot Grafik Pelatihan	55

DAFTAR TABEL

Tabel 2.1 <i>Cleansing</i>	10
Tabel 2.2 <i>Case-folding</i>	10
Tabel 2.3 <i>Tokenizing</i>	11
Tabel 2.4 <i>Stopword Removal</i>	11
Tabel 2.5 <i>Stemming</i>	12
Tabel 2.6 Indonesia <i>Sentiment Lexicon</i>	13
Tabel 2.7 <i>Confusion Matrix</i>	19
Tabel 3.1 Tahapan Penelitian	22
Tabel 3.2 Contoh <i>Dataset</i>	25
Tabel 3.3 Hasil Proses <i>Cleansing</i>	26
Tabel 3.4 Hasil Proses <i>Case-folding</i>	27
Tabel 3.5 Hasil Proses <i>Tokenizing</i>	27
Tabel 3.6 Hasil Proses <i>Stopword Removal</i>	28
Tabel 3.7 Contoh <i>Lexicon</i> Positif.....	29
Tabel 3.8 Contoh <i>Lexicon</i> Negatif	29
Tabel 3.9 Contoh <i>Tweet</i> Beserta <i>Score</i> Sentimen	30
Tabel 3.10 <i>Split</i> Data.....	33
Tabel 3.11 <i>Set Layer</i> dan <i>Neuron</i>	33
Tabel 3.12 <i>Epoch</i>	33
Tabel 3.13 <i>Tuning Hyperparameter</i>	34
Tabel 4.1 Informasi Data <i>Tweet</i>	36
Tabel 4.2 Isi <i>Dataframe</i>	37
Tabel 4.3 Data Setelah Proses <i>Cleansing</i>	39
Tabel 4.4 Data Setelah Proses <i>Case-folding</i>	40
Tabel 4.5 Data Setelah Proses <i>Preprocess Text</i>	42
Tabel 4.6 Data Setelah <i>Split</i> Token.....	44
Tabel 4.7 Informasi Data Setelah <i>Preprocessing</i>	46
Tabel 4.8 Hasil <i>Lexicon-based Sentiment Analysis</i>	48
Tabel 4.9 Jumlah <i>Dataset</i> Tiap Kelas Sentimen Setelah <i>Rebalancing</i>	48
Tabel 4.10 Skema <i>Hyperparameter</i> Model.....	49
Tabel 4.11 Pelatihan MLP	52

Tabel 4.12 <i>Summary</i> Arsitektur Model CNN	54
Tabel 4.13 Pelatihan CNN	55
Tabel 4.14 <i>Classification Report</i> Hasil Uji MLP.....	56
Tabel 4.15 <i>Classification Report</i> Hasil Uji CNN.....	57
Tabel 4.16 Hasil Pelatihan dan Pengujian MLP pada Perbandingan <i>Split Data</i> ..	58
Tabel 4.17 Hasil Pelatihan dan Pengujian CNN pada Perbandingan <i>Split Data</i> ..	58
Tabel 4.18 Hasil Pelatihan dan Pengujian MLP pada <i>Set Hidden Layer</i>	59
Tabel 4.19 Hasil Pelatihan dan Pengujian CNN pada <i>Set Hidden Layer</i>	60
Tabel 4.20 Hasil Pelatihan dan Pengujian MLP pada <i>Epoch</i>	61
Tabel 4.21 Hasil Pelatihan dan Pengujian CNN pada <i>Epoch</i>	61
Tabel 4.22 Hasil Pelatihan dan Pengujian MLP pada <i>Learning Rate</i>	62
Tabel 4.23 Hasil Pelatihan dan Pengujian CNN pada <i>Learning Rate</i>	63
Tabel 4.24 Parameter Terbaik Model MLP dan CNN	63
Tabel 4.25 Perbandingan Hasil Uji Model MLP dan CNN	64
Tabel 4.26 <i>Classification Report</i> Pengujian MLP Dengan Parameter Terbaik	64
Tabel 4.27 <i>Classification Report</i> Pengujian CNN Dengan Parameter Terbaik....	64
Tabel 4.28 Total <i>Predicted Data</i> Pengujian MLP dan CNN.....	65