



**PERBANDINGAN TEKNIK PEMBOBOTAN FAST TEXT DAN
WORD2VEC UNTUK DETEKSI MULTICLASS EMOSI
MENGUNAKAN METODE CONV-LSTM**

SKRIPSI

SALWA ZIADA SALSABIILA

1910511094

**PROGRAM STUDI S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
2023**



**PERBANDINGAN TEKNIK PEMBOBOTAN FAST TEXT DAN
WORD2VEC UNTUK DETEKSI MULTICLASS EMOSI
MENGUNAKAN METODE CONV-LSTM**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Komputer**

SALWA ZIADA SALSABIILA

1910511094

**PROGRAM STUDI S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
2023**

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Salwa Ziada Salsabiila

NIM : 1910511094

Fakultas : Ilmu Komputer

Program Studi : S1 Informatika

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 16 Juli 2023

Yang Menyatakan,



Salwa Ziada Salsabiila

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK
KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini:

Nama : Salwa Ziada Salsabiila
NIM : 1910511094
Fakultas : Ilmu Komputer
Program Studi : S1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non Eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**PERBANDINGAN TEKNIK PEMBOBOTAN FAST TEXT DAN WORD2VEC
UNTUK DETEKSI MULTICLASS EMOSI MENGGUNAKAN METODE CONV-
LSTM**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih, media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 16 Juli 23

Yang Menyatakan,



(Salwa Ziada Salsabiila)

LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Salwa Ziada Salsabiila
NIM : 1910511094
Program Studi : S1 Informatika
Judul Tugas Akhir : Perbandingan Teknik Pembobotan Fast Text dan Word2Vec untuk Deteksi Multiclass Emosi menggunakan Conv-LSTM

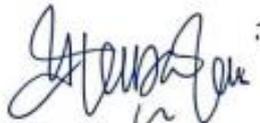
Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Helena Nurramdhani Irmanda, S.Pd, M.Kom
Pembimbing I



Artika Arista S.Kom, MMSI
Pembimbing II



Yuni Widiastiwi, S.Kom, M.Si
Penguji I



Jayanta, S.Kom., M.Si
Penguji II



Dr. Ermatita, S.Kom, M.Kom
Dekan



Dr. Widya Cholil, M.I.T
Kepala Program Studi

Ditetapkan di : Jakarta
Tanggal Ujian : 03 Juli 2023



KATA PENGANTAR

Alhamdulillah puji dan syukur saya panjatkan kepada Allah S.W.T atas rahmat dan karunia-Nya, saya sebagai penulis dapat menyelesaikan Skripsi ini sebagai salah satu persyaratan untuk memperoleh gelar sarjana. Judul yang dipilih dalam penelitian ini adalah **PERBANDINGAN TEKNIK PEMBOBOTAN FAST TEXT DAN WORD2VEC UNTUK DETEKSI MULTICLASS EMOSI MENGGUNAKAN METODE CONV-LSTM**. Laporan ini dapat diselesaikan tepat pada waktunya tidak terlepas dari bantuan dan dukungan dari berbagai pihak, Oleh karena itu penulis mengucapkan terima kasih kepada :

1. Allah S.W.T atas rahmat dan karunia-Nya kepada penulis sehingga penulis dapat menyelesaikan penyusunan Skripsi.
2. Ibu Dr. Ermatita, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
3. Ibu Dr. Widya Cholil, M.I.T selaku Ketua Program Studi S1 Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta
4. Ibu Helena Nurramdhani Irmanda, S.Pd, M. Kom. selaku Dosen Pembimbing Akademik dan Dosen Pembimbing penelitian yang telah membantu penulis dalam proses penyusunan Skripsi.
5. Ibu Artika Arista, S.Kom., MMSI selaku Dosen Pembimbing Akademik dan Dosen Pembimbing penelitian yang telah membantu penulis dalam proses penyusunan Skripsi.
6. Seluruh keluarga tercinta dan teman-teman yang selalu memberi motivasi dan dukungan doa sehingga penulis dapat menyelesaikan penyusunan Skripsi

Bogor, 06 Juni 2023

Salwa Ziada Salsabiila

Penulis

DAFTAR ISI

Abstrak	ii
Abstract	iii
KATA PENGANTAR	iv
DAFTAR ISI	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Ruang Lingkup Penelitian	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Luaran yang diharapkan	4
1.7 Sistematika Penulisan	4
BAB II	5
TINJAUAN PUSTAKA	5
2.1 Deteksi Emosi	5
2.2 Model Emosi Ekman	5
2.3 <i>Twitter</i>	5
2.4 <i>Twitter Scrapping</i>	5
2.5 <i>Python</i>	6
2.6 Dataset	6
2.7 <i>Preprocessing Data</i>	6
2.8 Pelabelan Data	6
2.9 Teknik Pembobotan Embedding	6
2.9.1 <i>Word2Vec</i>	7
2.9.1.1 <i>Continous Bag of Words</i>	7
2.9.2 <i>Fast Text</i>	10
2.10 Negative Sampling	12
2.10 Noise Distribution	14
2.11 Random State	15
2.12 Convolutional Long Short Term Memory (Conv-LSTM)	16
2.11 Evaluasi	25

2.12 Penelitian Terkait	27
BAB III	31
METODE PENELITIAN	31
3.1 Tahapan Penelitian.....	31
3.2 Alur Penelitian	32
3.2.1 Studi Literatur	32
3.2.2 Pengumpulan Dataset.....	32
3.2.3 Preprocessing Data.....	33
3.2.4 Pelabelan Data.....	38
3.2.5 Teknik Pembobotan Fast Text Embedding	41
3.2.6 Teknik Pembobotan Word2Vec	45
3.2.7 <i>Split Data</i>	51
3.2.8 Perhitungan Conv-LSTM.....	51
3.2.9 Evaluasi.....	57
3.3 Alat yang digunakan	57
3.4 Jadwal Penelitian.....	58
BAB IV	59
HASIL DAN PEMBAHASAN.....	59
4.1 Data.....	59
4.2 Preprocessing Data	59
4.2.1 Membersihkan Data	60
4.2.2 Menerjemahkan Data	60
4.2.3 Mengubah Slang menjadi Formal	61
4.2.4 Stopword	61
4.2.5 Stemming	62
4.2.6 Hapus Duplikasi Data	62
4.3 Pelabelan Data	63
4.4 Pembagian Data.....	64
4.5 Fast Text.....	64
4.5.1 Fast Text dengan Model Conv LSTM.....	64
1. Fast Text dengan Skipgram.....	65
2. Fast Text dengan CBOW	67
4.6 Word2Vec.....	69
4.6.1 Word2Vec	70
4.6.2 Word2Vec CBOW	73

4.8 Analisis Hasil Evaluasi Model.....	76
BAB V	81
PENUTUP	81
5.1 Kesimpulan	81
5.2 Saran	81
DAFTAR PUSTAKA	82

DAFTAR GAMBAR

Gambar 1 Continous Bag of Words (CBOW) (Rong, 2014)	8
Gambar 2 Model Skipgram (Rong, 2014)	9
Gambar 3 Arsitektur Model Fast Text Skipgram + Subword untuk kalimat dengan fitur ngram x_1, \dots, x_n . (Ganesan, n.d.)	11
Gambar 4 Random state bekerja (Modasiya, 2022)	16
Gambar 5 Perbedaan Max Pooling dan Average Pooling	19
Gambar 6 Perbedaan Global Max Pooling dan Global Average Pooling (Kim and Jeong, 2019)	19
Gambar 7 Dropout (Baeldung, 2023)	22
Gambar 8 Arsitektur LSTM	23
Gambar 9 Tahapan Penelitian	31
Gambar 10 Alur terjemahan data	34
Gambar 11 Alur ubah kata slang	36
Gambar 12 Algoritma kode Pelabelan	39
Gambar 13 Contoh Teks yang diproses	42
Gambar 14 Tokenisasi	43
Gambar 15 pembentukan vokabular	43
Gambar 16 Penyematan jumlah vektor contoh ngram 3	44
Gambar 17 kata konteks	44
Gambar 18 negative samples	44
Gambar 19 Fast Text Negative Sampling	45
Gambar 20 tahapan word2vec	45
Gambar 21 Contoh Teks yang diproses	46
Gambar 22 contoh inisialisasi skipgram	47
Gambar 23 word2vec Skipgram	47
Gambar 24 Memaksimalkan pasangan positif dan meminimalkan pasangan negatif	48
Gambar 25 inisialisasi contoh proses CBOW	49
Gambar 26 word2vec CBOW	49
Gambar 27 Memaksimalkan pasangan positif dan meminimalkan pasangan negatif	50
Gambar 28 Pembagian data	51
Gambar 29 Conv-LSTM Max Pooling (Zhou et al., 2015)	52
Gambar 30 Conv-LSTM Average Pooling (Zhou et al., 2015)	52
Gambar 31 Convolution Layer	54
Gambar 32 Model LSTM Layer	55
Gambar 33 Embedding Layer	55
Gambar 34 Concatenate / Penggabungan	56
Gambar 35 Hasil Scrapping Data Twitter	59
Gambar 36 Hasil Preprocessing data	60
Gambar 37 Menerjemahkan Data	60
Gambar 38 Stopword	61
Gambar 39 Stemming	62
Gambar 40 Parameter Fast Text Skipgram	65
Gambar 41 Fast Text Subword + Skipgram_Average Pooling	66

Gambar 42 Fast Text Subword + Skipgram_Max Pooling	67
Gambar 43 Parameter Fast Text CBOW	67
Gambar 44 Fast Text Subword + CBOW_Max Pooling	69
Gambar 45 Fast Text Subword + CBOW_Max Pooling	69
Gambar 46 Membangun Model Word2 Vec	70
Gambar 47 Hyperparameter Skipgram	70
Gambar 48 Word2vec Skipgram_Max Pooling Conv-LSTM	72
Gambar 49 Word2vec Skipgram_Average Pooling Conv-LSTM	73
Gambar 50 Hyperparameter CBOW	73
Gambar 51 Word2vec CBOW_Max Pooling	76
Gambar 52 Wor2Vec CBOW_Average Pooling	76
Gambar 53 Train Model ConvLSTM	77

DAFTAR TABEL

Table 1 Penjelasan Notasi	13
Table 2 Penjelasan Notasi	15
Table 3 Max Pooling dan Average Pooling	19
Table 4 Global Max Pooling dan Global Average Pooling	20
Table 5 Global Max Pooling dan Global Average Pooling (Brownlee, 2019)	21
Table 6 Contoh Confussion Matrix	25
Table 7 Predicted Class dan Actual Class	25
Table 8 Penelitian Terkait	27
Table 9 Pengumpulan dataset	33
Table 10 Contoh tidak diberikan spasi	34
Table 11 Contoh diberikan spasi	35
Table 12 contoh content	35
Table 13 tweet setelah diterjemahkan	36
Table 14 Duplikasi data	38
Table 15 Jumlah Pelabelan Data	39
Table 16 Notasi Kode Pelabelan	40
Table 17 Pembagian Kata dalam 6 emosi	41
Table 18 Tabel Parameter	42
Table 19 Parameter Model	46
Table 20 contoh kalimat tweet	47
Table 21 Word2vec Skipgram Negative Sampling	48
Table 22 Binomial Classification	48
Table 23 contoh kalimat tweet	49
Table 24 Word2vec Skipgram Negative Sampling	50
Table 25 Binomial Classification	50
Table 26 Model Convolutional	53
Table 27 Model LSTM	53
Table 28 Embedding Layer	54
Table 29 Contoh Output	56
Table 30 Alat yang digunakan	57
Table 31 Jadwal Penelitian	58
Table 32 Variabel Dataset	59
Table 33 Membersihkan data	60
Table 34 Menerjemahkan data	61
Table 35 Mengubah Slang	61
Table 36 Stopword	62
Table 37 Stemming	62
Table 38 Hapus Duplikat Data	62
Table 39 Hapus Duplikat Data	63
Table 40 Pelabelan Data	63
Table 41 Pengumpulan Data	64
Table 42 Penjelasan Parameter Fast Text Skipgram	65
Table 43 Evaluasi Fast Text Skipgram dengan Average Pooling dan Max Pooling	66
Table 44 Penjelasan Parameter Fast Text CBOW	67
Table 45 Akurasi Fast Text CBOW Conv LSTM – Average Pooling & Max Pooling	68

Table 46 Keterangan Hyperparameter	70
Table 47 Evaluasi Model	71
Table 48 Evaluasi Word2Vec Skipgram dengan Average Pooling	72
Table 49 Penjelasan Hyperparameter CBOW	74
Table 50 Evaluasi Model	75
Table 51 Evaluasi Word2Vec CBOW dengan Average Pooling	75
Table 52 Parameter Train Model	77