

**KOMPARASI AKURASI EKSTRAKSI FITUR METODE GLOVE DAN
FASTTEXT MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK
PADA ANALISIS SENTIMEN MENGENAI PEMINDAHAN IBU KOTA
NEGARA INDONESIA**

SKRIPSI



Oleh:

DESYA RISTYA PUTRI

20081010240

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2024

**LEMBAR PENGESAHAN
SKRIPSI**

**Judul : KOMPARASI AKURASI EKSTRAKSI FITUR METODE
GLOVE DAN FASTTEXT MENGGUNAKAN
CONVOLUTIONAL NEURAL NETWORK PADA ANALISIS
SENTIMEN MENGENAI PEMINDAHAN IBU KOTA
NEGARA INDONESIA**

**Oleh : DESYA RISTYA PUTRI
NPM : 20081010240**

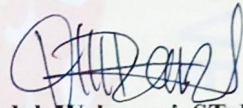

**Telah Diseminarkan Dalam Ujian Skripsi Pada :
Hari Kamis, Tanggal 4 Juli 2024**

Mengetahui

Dosen Pembimbing

Dosen Penguji

1.



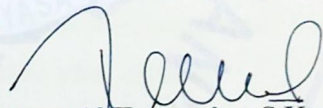
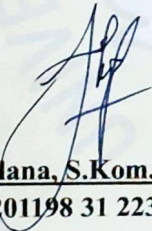
Eva Yulia Puspaningrum, S.Kom., M.Kom.

Henni Endah Wahanani, ST. M.Kom.

NIP : 19890705 2021212 002

NIP : 19780922 2021212 005

2.



Hendra Maulana, S.Kom., M.Kom.

M. Muharrom Al Haromainy, S.Kom., M.Kom.

NPT : 201198 31 223248

NIP : 19950601 202203 1 006

Menyetujui

Dekan

Koordinator Program Studi

Fakultas Ilmu Komputer

Informatika



Prof. Dr. Ir. Novirina Hendrasarie, M.T.

Fetty Tri Anggraeny, S.Kom. M.Kom.

NIP : 196811261994032001

NIP : 19820211 2021212 005

SURAT PERNYATAAN BEBAS DARI PLAGIASI

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

Nama : Desya Ristya Putri

NPM : 20081010240

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul:

“KOMPARASI AKURASI EKSTRAKSI FITUR METODE GLOVE DAN FASTTEXT MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK PADA ANALISIS SENTIMEN MENGENAI PEMINDAHAN IBU KOTA NEGARA INDONESIA”

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/Penelitian orang lain dan juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan 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 di Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.

Surabaya, 10 Juli 2024

Hormat saya,



The stamp is yellow with a red Garuda emblem. It contains the text: 'SIKIPULIH BEBAS PLAGIASI 10000', 'ALTAIR', 'ITEN BEL', and 'F2AKX561990410'.

Desya Ristya Putri

NPM. 20081010240

KOMPARASI AKURASI EKSTRAKSI FITUR METODE GLOVE DAN FASTTEXT MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK PADA ANALISIS SENTIMEN MENGENAI PEMINDAHAN IBU KOTA NEGARA INDONESIA

Nama Mahasiswa : Desya Ristya Putri

NPM : 20081010240

Program Studi : Informatika

Dosen Pembimbing : Eva Yulia Puspaningrum, S.Kom, M.Kom.

Hendra Maulana, S.Kom, M.Kom.

ABSTRAK

Saat ini, media sosial telah menjadi platform komunikasi virtual yang populer di kalangan masyarakat. Di dalamnya, terdapat beragam informasi yang seringkali mencakup opini-opini. Namun, hal ini sering kali memunculkan ketidakseimbangan antara informasi yang disampaikan dan bagaimana informasi tersebut diinterpretasikan oleh pembaca. Tak jarang, informasi yang tersebar juga berupa hoaks yang memperumit pemahaman akan sentimen yang sebenarnya ingin disampaikan. Terutama dalam beberapa tahun terakhir, dimana informasi yang akurat mengenai topik pemerintahan cukup menjadi pembicaraan. Penelitian ini bertujuan untuk menilai keakuratan percobaan pembelajaran mesin dengan menerapkan metode word embedding GloVe dan FastText dalam menganalisis sentimen terhadap komentar pengguna YouTube mengenai pemindahan ibu kota negara Indonesia. Metode tersebut diuji dengan menerapkan algoritma Convolutional Neural Network (CNN) pada 44957 data komentar yang dibagi menjadi data pelatihan dan data pengujian dengan perbandingan 70:30. Hasil percobaan menunjukkan bahwa penggunaan GloVe-CNN menghasilkan akurasi sebesar 76.1%, sedangkan FastText-CNN menghasilkan akurasi sebesar 75.5%. Dengan demikian, penelitian ini menyoroti pentingnya ekstraksi fitur dan analisis sentimen menggunakan metode seperti GloVe dan FastText dalam klasifikasi.

Kata kunci: *word embedding, GloVe, FastText, CNN, analisis sentimen*

KATA PENGANTAR

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa atas petunjuk, pertolongan, dan nikmat yang diberikan-Nya, penulis dapat menyelesaikan laporan skripsi ini yang berjudul “KOMPARASI AKURASI EKSTRAKSI FITUR METODE GLOVE DAN FASTTEXT MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK PADA ANALISIS SENTIMEN MENGENAI PEMINDAHAN IBU KOTA NEGARA INDONESIA”.

Bantuan dan dukungan dari berbagai pihak sangat berperan penting selama proses penelitian hingga penulisan laporan skripsi ini selesai. Penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dan berkontribusi dalam penyusunan laporan skripsi ini dari awal hingga akhir.

Penulis menyadari bahwa laporan skripsi ini mungkin masih memiliki kekurangan karena keterbatasan pengetahuan dan kemampuan. Oleh karena itu, penulis sangat menghargai masukan, kritik, dan saran yang konstruktif untuk memperbaiki dan mengembangkan laporan ini.

Surabaya, Juli 2024

Penulis

UCAPAN TERIMA KASIH

Dalam penyusunan dan penulisan skripsi ini, saya tidak lepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, saya ingin mengucapkan terima kasih kepada Tuhan Yang Maha Esa yang selalu memberikan kelancaran, kemudahan, kekuatan, dan sumber dari segala pencerahan. Saya juga mengucapkan terima kasih kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Fetty Anggraini, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Eva Yulia Puspaningrum, S.Kom, M.Kom. selaku dosen pembimbing I dan Bapak Hendra Maulana, S.Kom, M.Kom. selaku dosen pembimbing II, yang dengan sabar dan tulus meluangkan waktu, tenaga, serta pikiran untuk memberikan bimbingan, motivasi, arahan, dan saran-saran berharga selama penyusunan skripsi ini.
4. Staf dan dosen Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah berbagi ilmu, arahan, dan pengalaman selama perkuliahan.
5. Kedua orang tua saya, A. Haris Sudjoko dan Listyawati, yang selalu memberikan semangat, motivasi, serta doa untuk kebaikan, keberhasilan, dan kesuksesan saya.
6. Kakak saya, Octgi Ristya Perdana dan Nadia Ristya Dewi, yang selalu mendoakan dan mendukung hingga penyusunan skripsi ini selesai.
7. Teman-teman Informatika angkatan 2020, yang telah bersedia membantu, mengisi, dan menjadi bagian dalam perjalanan hidup dan perkuliahan saya.
8. Semua pihak yang tidak dapat saya sebutkan satu per satu, yang juga turut mendoakan dan memberikan semangat dalam menyelesaikan skripsi ini.

Terima kasih atas segala bantuan, dukungan, dan doa-doanya. Semoga Tuhan Yang Maha Esa memberikan balasan yang setimpal kepada semua pihak

yang telah membantu sehingga saya mampu menyelesaikan penelitian dan penyusunan laporan skripsi ini dengan baik dan tepat waktu.

Surabaya, Juli 2024

Penulis

DAFTAR ISI

ABSTRAK	iii
KATA PENGANTAR	iv
UCAPAN TERIMA KASIH	v
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
BAB II	5
TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu	5
2.2 Dasar Teori	6
2.3 Preprocessing untuk Data Text	7
2.4 Labelling	9
2.5 Word Embedding	10
2.6 Convolutional Neural Network	11
2.7 Confusion Matrix	13
BAB III	15
METODOLOGI	15

3.1	Tahapan Pengerjaan	15
3.2	Studi Literatur.....	16
3.3	Analisis dan Desain	16
3.4	Memperoleh Data.....	17
3.5	Preprocessing Data.....	21
3.6	Labelling	26
3.7	Ekstraksi Fitur	31
3.8	Klasifikasi Data	40
3.9	Skenario Pengujian	42
3.10	Pengujian Hasil	44
BAB IV		45
HASIL DAN PEMBAHASAN		45
4.1	Proses Memperoleh Data.....	45
4.2	Preprocessing Data.....	49
4.3	Labelling	54
4.4	Ekstraksi Fitur	59
4.5	Split Train dan Test Data.....	64
4.6	Klasifikasi	66
4.7	Pengujian Hasil	68
BAB V.....		75
PENUTUP.....		75
5.1	Kesimpulan	75
5.2	Saran	75
DAFTAR PUSTAKA.....		xi

DAFTAR GAMBAR

Gambar 2.1	Arsitektur DoubleMax Convolutional Neural Network	12
Gambar 3.1	Desain Sistem	16
Gambar 3.2	Tahap Memperoleh Data	17
Gambar 3.3	Filter Video YouTube	18
Gambar 3.4	Lima Video YouTube Teratas	18
Gambar 3.5	Crawl Data Komentar YouTube	19
Gambar 3.6	Data Sebelum Preprocessing	21
Gambar 3.7	Tahap Preprocessing Data	22
Gambar 3.8	File CSV Hasil Preprocessing	26
Gambar 3.9	Alur Proses Pelabelan Data	26
Gambar 3.10	Alur Metode GloVe	32
Gambar 3.11	Alur Metode FastText	37
Gambar 3.12	Ilustrasi Cara Kerja Fasttext	38
Gambar 3.13	Tahap Klasifikasi	40
Gambar 3.14	Arsitektur DoubleMax CNN	41
Gambar 4.1	YouTube API pada Google Developer Console	45
Gambar 4.2	Bentuk Data Hasil Crawl	48
Gambar 4.3	Data dalam Bentuk List	50
Gambar 4.4	Hasil Proses Case Folding	51
Gambar 4.5	Hasil Proses Tokenization	52
Gambar 4.6	Daftar Kata Stopwords	52

Gambar 4.7 Hasil Proses Stopwords Removal	53
Gambar 4.8 Hasil Proses Stemming	53
Gambar 4.9 File CSV Hasil Preprocessing	54
Gambar 4.10 Hasil Menerjemahkan Teks Sentimen	55
Gambar 4.11 Compound Score Setiap Teks Sentimen	56
Gambar 4.12 Hasil Labelling	57
Gambar 4.13 Dataset Hasil Labelling	58
Gambar 4.14 Persebaran Data Berdasarkan Label	59
Gambar 4.15 Vektor Kata GloVe	61
Gambar 4.16 Format File untuk Pelatihan Model FastText	62
Gambar 4.17 Vektor Kata FastText	63
Gambar 4.18 Data Teks Hasil Diterapkannya Padding	64
Gambar 4.19 Label Encoding	65
Gambar 4.20 Jumlah Data Hasil Split	65
Gambar 4.21 Persebaran Data Hasil Split Berdasarkan Labelnya	66
Gambar 4.22 Classification Report Skenario Terbaik GloVe-CNN	71
Gambar 4.23 Classification Report Skenario Terbaik FastText-CNN	72
Gambar 4.24 Confusion Matrix GloVe-CNN	72
Gambar 4.25 Confusion Matrix FastText-CNN	73

DAFTAR TABEL

Tabel 2.1 Case Folding	7
Tabel 2.2 Tokenization	8
Tabel 2.3 Stopwords Removal	8
Tabel 2.4 Stemming	9
Tabel 2.5 Nilai Confusion Matrix	13
Tabel 3.1 Alur Pengerjaan Tugas Akhir	15
Tabel 3.2 Video YouTube yang Digunakan	19
Tabel 3.3 Jumlah Data yang Digunakan	20
Tabel 3.4 Komentar Sebelum Dilakukan Preprocessing	22
Tabel 3.5 Data Teks Setelah Dilakukan Case Folding	23
Tabel 3.6 Data Teks Setelah Dilakukan Tokenization	24
Tabel 3.7 Data Teks Setelah Dilakukan Stopwords Removal	24
Tabel 3.8 Data Teks Setelah Dilakukan Stemming	25
Tabel 3.9 Karakteristik GloVe	31
Tabel 3.10 Karakteristik FastText	31
Tabel 3.11 Contoh Bentuk Matriks Co-Occurrence Glove	34
Tabel 3.12 Contoh Bentuk Matriks Joint Probability Glove (Pecahan)	35
Tabel 3.13 Contoh Bentuk Matriks Joint Probability Glove (Desimal)	36
Tabel 3.14 Contoh Representasi Vektor Sub-Kata FastText 2 Dimensi	39
Tabel 3.15 Contoh Representasi Vektor Kata FastText 2 Dimensi	39
Tabel 3.16 Skenario Pengujian	43

Tabel 4.1 Waktu Komputasi Crawling Data Komentar Youtube	48
Tabel 4.2 Hasil Akurasi Seluruh Skenario Uji	69
Tabel 4.3 Label Eksperimen Klasifikasi Dengan & Tanpa Preprocessing	70
Tabel 4.4 Akurasi Eksperimen Klasifikasi Dengan & Tanpa Preprocessing ..	71
Tabel 4.5 Rata-Rata Hasil Skenario Terbaik	74