



SKRIPSI

**INTEGRASI BERT DAN NAMED ENTITY RECOGNITION
UNTUK ANALISIS EMOSI MULTILABEL KOMENTAR
PERS PRESIDEN DI YOUTUBE**

FAKHRI SABRAN YUNANSAH

NPM 22082010153

DOSEN PEMBIMBING

Eka Dyar Wahyuni, S.Kom., M.Kom.

Nambi Sembilu, S.Kom., M.Kom

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR**

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI SISTEM INFORMASI

SURABAYA

2026

LEMBAR PENGESAHAN

INTEGRASI BERT DAN *NAMED ENTITY RECOGNITION* UNTUK ANALISIS EMOSI MULTILABEL KOMENTAR PERS PRESIDEN DI YOUTUBE

Oleh :


FAKHRI SABRAN YUNANSAH

NPM. 22082010153


Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 8 Mei 2026

Menyetujui

Eka Dvar Wahyuni, S.Kom., M.Kom.
NIP. 198412012021212005


..... (Pembimbing I)


Nambi Sembilu, S.Kom., M.Kom.
NIP. 199005162024061003


..... (Pembimbing II)

Dr. Eng Agussalim, S.Pd., MT.
NIP. 198508112019031005


..... (Ketua Penguji)

Nur Cahyo Wibowo, S.Kom., M.Kom.
NIP. 197903172021211002


..... (Anggota Penguji I)

P.Karunia Farista Ananto, S.Kom.,
M.Kom., M.IM.
NIP. 19970704 202406 2001


..... (Anggota Penguji II)

Mengetahui,
Dekan Fakultas Ilmu Komputer,


.....
(Prof. Dr. Ir. Novirina Hendrasarie, MT.)
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

INTEGRASI BERT DAN *NAMED ENTITY RECOGNITION* UNTUK ANALISIS EMOSI MULTILABEL KOMENTAR PERS PRESIDEN DI YOUTUBE

Oleh :

FAKHRI SABRAN YUNANSAH

NPM. 22082010153

Menyetujui,

Koordinator Program Studi Sistem Informasi
Fakultas Ilmu Komputer



Siti Mukaromah, S.Kom, M.Kom.
NIP. 19851124 2021211 003

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertandatangan di bawah ini:

Nama Mahasiswa : Fakhri Sabran Yunansah
NPM : 22082010153
Program : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu Lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/Lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar Pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 18 Mei 2026

Yang Membuat Pernyataan,



Fakhri Sabran Yunansah
NPM. 22082010153

ABSTRAK

Nama Mahasiswa / NPM : Fakhri Sabran Yunansah / 22082010153
Judul Skripsi : Integrasi Bert Dan Named Entity Recognition
Untuk Analisis Emosi Multilabel Komentar Pers
Presiden Di Youtube

Dosen Pembimbing : 1. Eka Dyar Wahyuni, S.Kom, M.Kom
2. Nambi Sembilu, S.Kom, M.Kom

Perkembangan media sosial, khususnya YouTube, telah menjadikannya sebagai sarana utama bagi masyarakat dalam menyampaikan opini dan mengekspresikan emosi terhadap berbagai isu publik, termasuk pernyataan resmi pemerintah. Komentar pada video konferensi pers Presiden Republik Indonesia menunjukkan keberagaman emosi yang kompleks dan sering muncul secara bersamaan dalam satu teks. Hal ini memerlukan pendekatan analisis yang mampu menangani karakteristik multilabel serta memahami konteks bahasa yang dinamis. Penelitian ini bertujuan untuk menganalisis kinerja integrasi *Bidirectional Encoder Representations from Transformers* (BERT) dengan *Named Entity Recognition* (NER) dalam klasifikasi emosi multilabel pada komentar YouTube dengan lima kategori emosi, yaitu *love*, *happiness*, *anger*, *fear*, dan *sadness*. Metode yang digunakan adalah *pipeline-based integration*, di mana NER digunakan untuk memperkaya representasi teks sebelum diklasifikasikan menggunakan IndoBERT. Data diperoleh dari komentar publik pada video konferensi pers Presiden yang diunggah melalui kanal resmi YouTube Sekretariat Presiden Republik Indonesia. Hasil penelitian menunjukkan bahwa integrasi NER mampu meningkatkan performa model pada beberapa skenario. Model terbaik diperoleh pada augmentasi back-translation dengan integrasi NER pada pembagian data 80:20, menghasilkan F1-score (micro) sebesar 0,7843 dan Hamming Loss sebesar 0,1199. Model tersebut menjadi yang terbaik karena mampu menghasilkan keseimbangan precision dan recall yang optimal melalui augmentasi data dan informasi entitas. Integrasi NER dan augmentasi data mampu meningkatkan kualitas klasifikasi emosi pada komentar publik berbahasa Indonesia.

Kata kunci: *BERT*, *Named Entity Recognition*, Analisis Emosi, Multilabel *Classification*, YouTube

ABSTRACT

Nama Mahasiswa / NPM : Fakhri Sabran Yunansah / 22082010153
Judul Skripsi : Integrasi Bert Dan Named Entity Recognition
Untuk Analisis Emosi Multilabel Komentar Pers
Presiden Di Youtube
Dosen Pembimbing : 1. Eka Dyar Wahyuni, S.Kom, M.Kom
2. Nambi Sembilu, S.Kom, M.Kom

The rapid growth of social media, particularly YouTube, has made it a major platform for the public to express opinions and emotions regarding various public issues, including official government statements. Comments on Indonesian Presidential press conference videos reflect complex emotional expressions that often appear simultaneously within a single text. This condition requires an analytical approach capable of handling multilabel characteristics while understanding dynamic language contexts. This study aims to analyze the performance of integrating Bidirectional Encoder Representations from Transformers (BERT) with Named Entity Recognition (NER) in multilabel emotion classification on YouTube comments using five emotion categories: love, happiness, anger, fear, and sadness. The method used in this study is a pipeline-based integration approach, where NER is utilized to enrich text representation before classification using IndoBERT. The dataset was obtained from public comments on a Presidential press conference video uploaded to the official YouTube channel of the Indonesian Presidential Secretariat. The results show that NER integration improves model performance in several scenarios. The best model was achieved using back-translation augmentation with NER integration under the 80:20 data split scenario, resulting in an F1-score (micro) of 0.7843 and a Hamming Loss of 0.1199. The model achieved the best performance by providing a balanced precision and recall through data augmentation and entity information. NER integration and data augmentation improved emotion classification performance for Indonesian public comments.

Keywords: BERT, Named Entity Recognition, Emotion Analysis, Multilabel Classification, YouTube

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan judul “Integrasi BERT dan Named Entity Recognition untuk Analisis Emosi Multilabel Komentar Pers Presiden di YouTube”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa banyak pihak yang telah memberikan bantuan, dukungan, serta bimbingan baik secara langsung maupun tidak langsung. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua tercinta dan keluarga besar penulis yang senantiasa memberikan doa, dukungan baik secara moral maupun materi, serta kasih sayang yang tulus dalam setiap langkah, sehingga penulis dapat menyelesaikan skripsi ini.
2. Ibu Siti Mukaromah, S.Kom., M.Kom. selaku Koordinator Program Studi Sistem Informasi Fakultas Ilmu Komputer UPN Veteran Jawa Timur yang telah memberikan dukungan, arahan, serta fasilitas akademik selama proses perkuliahan hingga penyusunan skripsi ini.
3. Ibu Eka Dyar Wahyuni, S.Kom., M.Kom. dan Bapak Nambi Sembilu, S.Kom., M.Kom. selaku dosen pembimbing yang telah meluangkan waktu, tenaga, dan pikiran dalam memberikan bimbingan, arahan, serta masukan yang sangat berarti selama proses penyusunan skripsi ini.
4. Bapak dan Ibu dosen serta tenaga kependidikan Program Studi Sistem Informasi yang telah memberikan ilmu, bimbingan, serta pelayanan akademik yang baik selama masa perkuliahan, sehingga penulis dapat menyelesaikan studi dengan lancar.
5. Syuraini Noor Chamsyah yang selalu memberikan dukungan, perhatian, dan semangat yang tak pernah putus, serta menjadi tempat berbagi dalam setiap proses dan perjuangan penulis menyelesaikan skripsi ini.
6. Teman-teman grup KDRT yang telah memberikan dukungan, semangat, serta kebersamaan selama proses penyusunan skripsi ini.

7. Diri penulis sendiri yang telah berusaha dengan sungguh-sungguh, tetap bertahan dalam berbagai tantangan, serta tidak menyerah hingga akhirnya mampu menyelesaikan skripsi ini.
8. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan, dukungan, serta semangat dalam berbagai bentuk selama proses penyusunan skripsi ini. Terima kasih atas segala kontribusi yang telah menjadi bagian dari perjalanan penulis.

Penulis menyadari bahwa penyusunan skripsi ini masih memiliki keterbatasan dan belum sepenuhnya sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang bersifat membangun sebagai bahan perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca serta dapat dijadikan sebagai referensi untuk penelitian selanjutnya.

Surabaya, 4 Mei 2026

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN BEBAS PLAGIASI	vii
ABSTRAK	ix
<i>ABSTRACT</i>	xi
KATA PENGANTAR	xiii
DAFTAR ISI	xv
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxii
DAFTAR SINGKATAN, ISTILAH, DAN SIMBOL	xxv
DAFTAR LAMPIRAN	xxvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	6
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
1.6 Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA	9
2.1 Dasar Teori	9
2.1.1 Youtube	9
2.1.2 <i>Natural Language Processing (NLP)</i>	10
2.1.3 <i>Word Embedding</i>	11
2.1.4 BERT (<i>Bidirectional Encoder Representations from Transformers</i>)	13
2.1.5 NER (<i>Named Entity Recognition</i>)	14
2.1.6 <i>Pipeline Based Integration</i>	15
2.1.7 Analisis Emosi	16
2.1.8 Klasifikasi Multilabel	17
2.1.9 <i>Preprocessing Text</i>	18
2.1.10 <i>F1-Score (Micro-Averaged)</i>	19
2.1.11 Nilai Krippendof	20

2.2 Penelitian Terdahulu	21
BAB III METODOLOGI PENELITIAN	24
3.1 Alur Penelitian	25
3.2 Studi Literatur	25
3.3 Analisis Kebutuhan	25
3.3.1 Kebutuhan Data	26
3.3.2 Kebutuhan <i>Software</i> dan <i>Hardware</i>	26
3.4 Pengumpulan Data	26
3.5 Filtering Data	27
3.6 Pelabelan Data	27
3.7 Eksplorasi Data	31
3.8 Pembagian Data	32
3.9 Perancangan Model	34
3.9.1 <i>Data Preprocessing</i>	35
3.9.2 Model 1 <i>IndoBert Emotion Classification</i>	36
3.9.3 Model 2 <i>NER + IndoBert Emotion Classification</i>	36
3.9.4 Skenario Model	37
3.9.5 <i>Emotion Classification</i>	39
3.9.6 <i>Evaluation</i>	39
3.9.7 Hasil Analisis	40
3.10 Penerapan Model	41
3.11 Validasi Model	44
BAB IV HASIL DAN PEMBAHASAN	45
4.1 Hasil	45
4.1.1 Studi Literatur	45
4.1.2 Implementasi Kebutuhan	45
4.1.3 Pengumpulan Data	46
4.1.4 Filtering Data	48
4.1.5 Pelabelan Data	51
4.1.6 Eksplorasi Data	59
4.1.7 Pembagian Data	61
4.1.8 Perancangan Model	63

4.1.9 Penerapan Model	102
4.1.10 Validasi Model	105
4.2 Pembahasan	112
BAB V PENUTUP	122
5.1 Kesimpulan	123
5.2 Saran	123
DAFTAR PUSTAKA	125
LAMPIRAN	134

DAFTAR GAMBAR

Gambar 1.1 Statistik Pengguna Youtube di Dunia per Februari 2025	1
Gambar 1.2 Statistik <i>Screentime</i> Sosial Media di Indonesia	2
Gambar 1.3 Komentar Publik di YouTube	3
Gambar 3.1 Diagram Alir Metodologi Penelitian	25
Gambar 3.2 Perancangan Model	34
Gambar 3.3 Wireframe <i>Home</i> Dashboard	43
Gambar 3.4 <i>Wireframe</i> Analisis Dashboard	44
Gambar 4.1 Kode Scraping Komentar	47
Gambar 4.2 Hasil Scraping	48
Gambar 4.3 Kode Filtering 1-3	49
Gambar 4.4 Kode Filtering 4-6	50
Gambar 4.5 Kode Kripendorf Emosi	55
Gambar 4.6 Hasil Kripendorf Emosi	56
Gambar 4.7 Kode Kripendorf NER	57
Gambar 4.8 Hasil Kripendorf NER	58
Gambar 4.9 Distribusi Label	59
Gambar 4.10 Persebaran <i>Entity</i>	60
Gambar 4.11 <i>Word Cloud</i>	61
Gambar 4.12 Kode <i>Preprocessing</i>	63
Gambar 4.13 Kode <i>Emotion</i> Dataset	64
Gambar 4.14 Kode Konfigurasi Model Baseline	65
Gambar 4.15 Kode Konfigurasi Model NER	66
Gambar 4.16 Kode <i>Extract Entity</i>	67
Gambar 4.17 Kode <i>Entity</i> Target	68
Gambar 4.18 <i>Pipeline Based Integration</i>	69
Gambar 4.19 Kode <i>Run Scenario</i> 80:20 Original Baseline	70
Gambar 4.20 Hasil Skenario 80:20 Original Baseline	70
Gambar 4.21 Kode <i>Run Scenario</i> 80:20 Oversample Baseline	71
Gambar 4.22 Hasil Skenario 80:20 Oversample Baseline	72
Gambar 4.23 Kode <i>Run Scenario</i> 80:20 Undersample Baseline	73
Gambar 4.24 Hasil Skenario 80:20 Undersample Baseline	73

Gambar 4.25 Kode <i>Run Scenario</i> 80:20 Weighted Random Sampler Baseline	74
Gambar 4.26 Hasil Skenario 80:20 Weighted Random Sampler Baseline	74
Gambar 4.27 Kode <i>Run Scenario</i> 80:20 Augmentasi Baseline	75
Gambar 4.28 Hasil Skenario 80:20 Augmentasi Baseline	75
Gambar 4.29 Kode <i>Run Scenario</i> 70:30 Original Baseline	76
Gambar 4.30 Hasil Skenario 70:30 Original Baseline	77
Gambar 4.31 Kode <i>Run Scenario</i> 70:30 Oversample Baseline	78
Gambar 4.32 Hasil Skenario 70:30 Oversample Baseline	78
Gambar 4.33 Kode <i>Run Scenario</i> 70:30 Undersample Baseline	79
Gambar 4.34 Hasil Skenario 70:30 Undersample Baseline	79
Gambar 4.35 Kode <i>Run Scenario</i> 70:30 Weight Random Sampler Baseline	80
Gambar 4.36 Hasil Skenario 70:30 Weight Random Sampler Baseline	81
Gambar 4.37 Kode <i>Run Scenario</i> 70:30 Augmentasi Baseline	82
Gambar 4.38 Hasil Skenario 70:30 Augmentasi Baseline	82
Gambar 4.39 Kode <i>Run Scenario</i> 80:20 Original NER	83
Gambar 4.40 Hasil Skenario 80:20 Original NER	83
Gambar 4.41 Kode <i>Run Scenario</i> 80:20 Oversample NER	84
Gambar 4.42 Hasil Skenario 80:20 Oversample NER	85
Gambar 4.43 Kode <i>Run Scenario</i> 80:20 Undersample NER	86
Gambar 4.44 Hasil Skenario 80:20 Undersample NER	86
Gambar 4.45 Kode <i>Run Scenario</i> 80:20 Weight Random Sampler NER	87
Gambar 4.46 Hasil Skenario 80:20 Weight Random Sampler NER	87
Gambar 4.47 Kode <i>Run Scenario</i> 80:20 Augmentasi NER	88
Gambar 4.48 Hasil Skenario 80:20 Augmentasi NER	89
Gambar 4.49 Kode <i>Run Scenario</i> 70:30 Original NER	90
Gambar 4.50 Hasil Skenario 70:30 Original NER	90
Gambar 4.51 Kode <i>Run Scenario</i> 70:30 Oversample NER	91
Gambar 4.52 Hasil Skenario 70:30 Oversample NER	91
Gambar 4.53 Kode <i>Run Scenario</i> 70:30 Undersample NER	92
Gambar 4.54 Hasil Skenario 70:30 Undersample NER	93
Gambar 4.55 Kode <i>Run Scenario</i> 70:30 Weighted Random Sampler NER	94
Gambar 4.56 Hasil Skenario 70:30 Weighted Random Sampler NER	94

Gambar 4.57 Kode <i>Run Scenario</i> 70:30 Augmentasi NER	95
Gambar 4.58 Hasil Skenario 70:30 Augmentasi NER	95
Gambar 4.59 Contoh Hasil <i>Emotion Classification</i> Baseline	96
Gambar 4.60 Contoh Hasil <i>Emotion Classification</i> NER	97
Gambar 4.61 Kode <i>Backend</i> Flask	102
Gambar 4.62 Tampilan Dashboard Penerapan Model	103
Gambar 4.63 Tampilan Hasil Analisis Penerapan Model	104
Gambar 4.64 Tampilan Trend Emosi Penerapan Model	104
Gambar 4.65 Tampilan Dashboard Validasi Model	105
Gambar 4.66 Tampilan Dashboard Validasi Model Bagian 2	106
Gambar 4.67 Tampilan Dashboard Validasi Model Bagian 3	107
Gambar 4.68 Tampilan Hasil Analisis Validasi Model	108
Gambar 4.69 Tampilan Trend Emosi Validasi Model Bagian 1	109
Gambar 4.70 Tampilan Trend Emosi Validasi Model Bagian 2	110
Gambar 4.71 Tampilan Trend Emosi Validasi Model Bagian 3	111
Gambar 4.72 Hasil <i>Confusion Matrix</i> Per Emosi	112
Gambar 4.73 Visualisasi <i>Heatmap</i> Ko-eksistensi	113
Gambar 4.74 Hasil <i>Confusion Matrix</i> Antar 2 Label Bagian 1	114
Gambar 4.75 Hasil <i>Confusion Matrix</i> Antar 2 Label Bagian 2	115
Gambar 4.76 Grafik <i>Evaluation Matrix</i> Utama 80:20	118
Gambar 4.77 Grafik <i>Evaluation Matrix</i> Utama 70:30	120

DAFTAR TABEL

Tabel 2.1 Interpretasi Nilai <i>Krippendorff's Alpha</i>	21
Tabel 2.2 Penelitian Terdahulu	22
Tabel 3.1 Hierarki emosi [16]	28
Tabel 3.2 Hierarki Entity [69]	29
Tabel 3.2 Pedoman Pengujian Nilai <i>Krippendorff Alpha</i>	31
Tabel 3.3 Daftar Skenario Model	37
Tabel 4.1 Hasil Labeling Emosi	51
Tabel 4.2 Hasil Labeling NER	53
Tabel 4.3 Hasil <i>Majority Voting</i> Emosi	58
Tabel 4.4 Hasil <i>Majority Voting</i> NER	59
Tabel 4.5 Daftar Pembagian Data Skenario	62
Tabel 4.6 Rangkuman Seluruh Hasil Skenario Model	98
Tabel 4.7 Hasil Analisis 5 Skenario Terbaik	101

DAFTAR SINGKATAN, ISTILAH, DAN SIMBOL

SINGKATAN/ ISTILAH/ SIMBOL	KETERANGAN
α	: Koefisien tingkat kesepakatan Krippendorff's Alpha.
Y	: Urutan label yang diprediksi.
X	: Urutan kata/input observasi.
$P(Y X)$: Probabilitas label Y terhadap input X.
$Z(X)$: Fungsi normalisasi pada CRF.
exp	: Fungsi eksponensial.
k	: Indeks fitur.
f_k	: Fungsi fitur pada CRF.
λ_k	: Bobot parameter fitur.
y_{t-1}	: Label sebelumnya t.
$\sum_{k=1}^K$: Operasi penjumlahan seluruh fitur.
K	: Jumlah total fitur yang digunakan.
TP	: Jumlah data positif yang diprediksi benar sebagai positif.
TN	: Jumlah data negatif yang diprediksi benar sebagai negatif.
FP	: Jumlah data negatif yang diprediksi sebagai positif.
FN	: Jumlah data positif yang diprediksi sebagai negatif.

DAFTAR LAMPIRAN

Lampiran 1. Surat Izin Penelitian	135
Lampiran 2. Jadwal Penelitian	136
Lampiran 3. Permohonan Izin Penelitian	136
Lampiran 4. Berita Affan Kurniawan tewas usai dilindas mobil Brimob	137
Lampiran 5. Berita kenaikan tunjangan DPR	137
Lampiran 6. Bukti pengerjaan anotator manual 1	137
Lampiran 7. Berita pengerjaan anotator manual 2	138