



SKRIPSI

**KLASIFIKASI EMOSI KOMENTAR YOUTUBE
PADA LAGU KARYA IDGITAF MENGGUNAKAN
METODE *COST-SENSITIVE SUPPORT VECTOR
MACHINE* (CS-SVM)**

DESI DAOMARA SITANGGANG
NPM 22083010027

DOSEN PEMBIMBING

Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom.
Muhammad Nasrudin, M.Stat.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

KLASIFIKASI : EMOSI KOMENTAR YOUTUBE PADA LAGU KARYA IDGITAF MENGGUNAKAN METODE *COST-SENSITIVE SUPPORT VECTOR MACHINE (CS-SVM)*

Oleh:
DESI DAOMARA SITANGGANG
NPM 22083010027

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi
Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan
Nasional Veteran Jawa Timur pada 26 Agustus 2026:

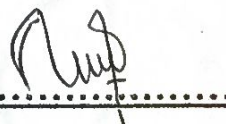
Menyetujui,

Wahyu Syaifullah J.S.,
S.Kom., M.Kom.
NIP 19860825 202121 1 003



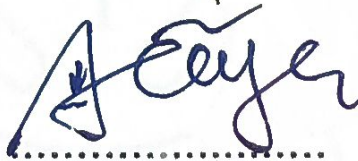
(Pembimbing I)

Muhammad Nasrudin,
M.Stat
NIP 19960909 202406 1 002



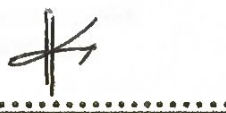
(Pembimbing II)

Amri Muhaimin, S.Stat.,
M.Stat., M.S.
NIP 19950723 202406 1 002



(Ketua Penguji)

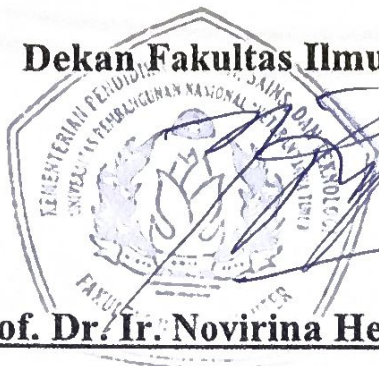
Kartika Maulida
Hindrayani, S.Kom.,
M.Kom.
NIP 19920909 202203 2 009



(Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.

NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

KLASIFIKASI EMOSI KOMENTAR YOUTUBE PADA LAGU KARYA DIGITAL MENGGUNAKAN METODE COST-SENSITIVE SUPPORT VECTOR MACHINE (CS-SVM)

Oleh:
DESI DAOMARA SITANGGANG
NPM 22083010027

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

**Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer**



Amri Muhaimin S.Stat., M.Stat., M.S

NIP 199507232024061002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Desi Daomara Sitanggang
NPM : 22083010027
Program : Sarjana (S1)
Program Studi : Sains Data
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 di kemudian 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, 3 September 2026
Yang Membuat Pernyataan,



DESI DAOMARA SITANGGANG
NPM 22083010027

ABSTRAK

Nama Mahasiswa / NPM : Desi Daomara Sitanggang / 22083010027
Judul Skripsi : Klasifikasi Emosi Komentar YouTube pada Lagu Karya Idgitaf Menggunakan Metode *Cost-Sensitive Support Vector Machine* (CS-SVM)
Dosen Pembimbing : 1. Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom.
2. Muhammad Nasrudin, M.Stat

YouTube merupakan salah satu platform media sosial yang memungkinkan pengguna mengekspresikan opini dan emosi melalui kolom komentar. Komentar pada lagu “Takut” dan “Satu-Satu” karya Idgitaf mengandung berbagai ekspresi emosi yang dapat dianalisis menggunakan teknik *Natural Language Processing* (NLP). Namun, klasifikasi emosi pada data teks menghadapi permasalahan ketidakseimbangan distribusi kelas yang dapat memengaruhi kemampuan model dalam mengenali kategori emosi dengan jumlah data yang lebih sedikit. Penelitian ini bertujuan untuk mengklasifikasikan emosi pada komentar YouTube lagu “Takut” karya Idgitaf menggunakan metode *Cost-Sensitive Support Vector Machine* (CS-SVM) dengan ekstraksi fitur Word2Vec serta membandingkan performa empat jenis kernel, yaitu *Linear*, *Polynomial*, *Radial Basis Function* (RBF), dan *Sigmoid*. Data penelitian diperoleh melalui *crawling* menggunakan YouTube Data API v3 sebanyak 20.644 komentar. Tahapan penelitian meliputi *preprocessing* yang terdiri atas *cleaning*, *case folding*, *tokenizing*, *normalisasi*, *filtering*, dan *stemming*, kemudian pelabelan emosi dilakukan menggunakan NRC Emotion Lexicon. Data yang sudah memiliki label selanjutnya direpresentasikan dalam bentuk vektor menggunakan Word2Vec dan diklasifikasikan menggunakan CS-SVM dengan pembobotan kelas untuk menangani ketidakseimbangan data. Evaluasi model dilakukan menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa kategori *trust* merupakan emosi yang paling dominan, diikuti oleh *fear* dan *anticipation*, sedangkan *surprise* merupakan kategori dengan jumlah data paling sedikit. Hasil perbandingan empat kernel menunjukkan bahwa kernel RBF memberikan performa terbaik dengan akurasi sebesar 70,87%, dibandingkan kernel Linear sebesar 70,63%, *Polynomial* sebesar 68,43%, dan *Sigmoid* sebesar 63,98%. Analisis visualisasi menunjukkan adanya tumpang tindih antarkategori emosi pada ruang fitur Word2Vec, sedangkan hasil *WordCloud* menunjukkan bahwa kategori *fear*, *sadness*, dan *joy* memiliki kata dominan yang relatif sesuai dengan karakteristik emosinya. Model CS-SVM dengan kernel RBF kemudian diimplementasikan ke dalam *Graphical User Interface* (GUI) berbasis Flask untuk melakukan klasifikasi emosi terhadap komentar baru secara interaktif.

Kata Kunci: Klasifikasi Emosi, *NRC Emotion Lexicon*, *Word2Vec*, *Cost-Sensitive Support Vector Machine (CS-SVM)*

ABSTRACT

Student Name / NPM : Desi Daomara Sitanggang / 22083010027
Undergraduate thesis : *Emotion Classification of YouTube Comments on the Song by Idgitaf Using a Cost-Sensitive Support Vector Machine (CS-SVM)*
Advisors : 1. Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom.
2. Muhammad Nasrudin, M.Stat.

YouTube is a social media platform that enables users to express their opinions and emotions through the comment section. Comments on the song “Takut” and “Satu-Satu” by Idgitaf contain various emotional expressions that can be analyzed using Natural Language Processing (NLP) techniques. However, emotion classification in text data faces the challenge of class imbalance, which can affect the model’s ability to recognize emotion categories with fewer samples. This study aims to classify emotions in YouTube comments on the song “Takut” by Idgitaf using the Cost-Sensitive Support Vector Machine (CS-SVM) method with Word2Vec feature extraction, while also comparing the performance of four kernel types: Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid. The research data were collected through crawling using the YouTube Data API v3, resulting in 20,644 comments. The research stages included preprocessing steps consisting of cleaning, case folding, tokenizing, normalization, filtering, and stemming, followed by emotion labeling using the NRC Emotion Lexicon. The labeled data were then represented as vectors using Word2Vec and classified using CS-SVM with class weighting to address data imbalance. Model evaluation was conducted using accuracy, precision, recall, and F1-score. The results showed that trust was the most dominant emotion category, followed by fear and anticipation, while surprise was the category with the smallest amount of data. The comparison of the four kernels indicated that the RBF kernel achieved the best performance, with an accuracy of 70.87%, compared to the Linear kernel at 70.63%, the Polynomial kernel at 68.43%, and the Sigmoid kernel at 63.98%. Visualization analysis showed overlap among emotion categories in the Word2Vec feature space, while the WordCloud results indicated that the fear, sadness, and joy categories had dominant words that were relatively consistent with their emotional characteristics. The CS-SVM model with the RBF kernel was then implemented in a Flask-based Graphical User Interface (GUI) to classify emotions in new comments interactively.

Keywords: *Emotion Classification, NRC Emotion Lexicon, Word2Vec, Cost Sensitive Support Vector Machine (CS-SVM)*

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yesus Kristus atas kasih, penyertaan-Nya serta berkat-Nya yang begitu berlimpah kepada penulis sehingga skripsi dengan judul **“Klasifikasi Emosi Komentar YouTube pada Lagu Karya Idgitaf Menggunakan Metode *Cost-Sensitive Support Vector Machine* (CS-SVM)”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Bapak Wahyu Syaifullah J. S, S.Kom., M.Kom. selaku Dosen Pembimbing I dan Muhammad Nasrudin, M.Stat. selaku Dosen Pembimbing 2 yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Selain itu, selama penyusunan proposal skripsi penulis juga banyak menerima bantuan dari berbagai pihak. Untuk itu penulis mengucapkan terima kasih kepada:

1. Diri sendiri, terimakasih atas usaha dan keteguhan untuk menyelesaikan skripsi ini. Terimakasih untuk bahu yang masih bertahan hingga saat ini dan terimakasih untuk tidak memenuhi kedagingan yang ditawarkan oleh dunia. Terimakasih sudah tetap berjalan di jalan Tuhan di tengah kepahitan yang menekan untuk menyerah.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Bapak Amri Muhaimin, S.Stat., M.Stat., M.S., selaku Koordinator Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur.
4. Bapak Trimono, S.Si., M.Si., selaku PIC Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur.
5. Idgitaf selaku pemilik karya lagu takut, serta karya-karya lainnya yang menemani penulis dalam pengerjaan skripsi.
6. Orang tua dan keluarga yang selalu mendoakan dan memberikan dukungan untuk menjaga kesehatan mental penulis.

7. Teman-teman sains data angkatan 22 yang tidak dapat penulis sebutkan satu per satu atas bantuan, dukungan, dan kebersamaan selama penyusunan skripsi ini.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, Agustus 2026

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN	iv
SURAT PERNYATAAN BEBAS PLAGIASI.....	v
ABSTRAK	vi
<i>ABSTRACT</i>	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xiii
DAFTAR NOTASI.....	xiv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	7
1.3. Batasan Masalah.....	8
1.4. Tujuan Penelitian	9
1.5. Manfaat Penelitian	9
BAB II TINJAUAN PUSTAKA.....	11
2.1. Penelitian Terdahulu	11
2.2. Kerangka Teori.....	22
2.2.1. YouTube.....	22

2.2.2. Klasifikasi Emosi.....	23
2.2.3. Text Mining	25
2.2.4. Preprocessing.....	26
2.2.5. Pelabelan.....	27
2.2.6. Ekstraksi Fitur	28
2.2.7. <i>Support Vector Machine</i> (SVM)	31
2.2.8. Evaluasi Model.....	37
2.2.9. Word Cloud	38
2.2.10. Graphical User Interface (GUI).....	39
BAB III METODOLOGI PENELITIAN.....	41
3.1. Variabel Penelitian dan Sumber Data.....	41
3.1.1. Variabel Penelitian	41
3.1.2. Sumber Data	42
3.2. Langkah Analisis	42
3.3. Desain Sistem	51
BAB IV HASIL DAN PEMBAHASAN.....	53
4.1. Hasil Penelitian.....	53
4.1.1. Pengumpulan Data.....	53
4.1.2. Preprocessing Data	54
4.1.3. Pelabelan Data	64
4.1.4. Ekstraksi Fitur	66
4.1.5. Splitting Data.....	71
4.1.6. Klasifikasi Model CS-SVM	71
4.1.7. Evaluasi Model.....	78
4.1.8. Visualisasi WordCloud.....	88
4.2. Tampilan Antar Muka (<i>Graphical User Interface</i>).....	95
BAB V PENUTUP	101

5.1. Kesimpulan	101
5.2. Saran Pengembangan	102
DAFTAR PUSTAKA	103
LAMPIRAN.....	107

DAFTAR GAMBAR

Gambar 2. 1. Hyperplane terbaik dan margin maksimum.....	32
Gambar 3. 1. Diagram alir.....	43
Gambar 4. 1. Confusion matrix	80
Gambar 4. 2. Visualisasi MDS	82
Gambar 4. 3 Wordcloud kategori anger	88
Gambar 4. 4. Wordcloud kategori anticipation	89
Gambar 4. 5. Wordcloud kategori disgust.....	90
Gambar 4. 6. Wordcloud kategori fear	91
Gambar 4. 7. Wordcloud kategori joy	92
Gambar 4. 8. Wordcloud kategori sadness	92
Gambar 4. 9. Tampilan pada halaman Home	96
Gambar 4. 10. Tampilah pada halaman Home bagian bawah.....	97
Gambar 4. 11. Tampilan pada halaman dataset.....	97
Gambar 4. 12. Tampilan pada halaman preprocessing.....	98
Gambar 4. 13. Tampilan pada halaman pelabelan	99
Gambar 4. 14. Tampilan pada halaman modeling.....	99
Gambar 4. 15. Tampilan pada halaman visualisasi	100

DAFTAR TABEL

Tabel 2. 1. Penelitian Terdahulu	11
Tabel 2. 2. Penjelasan Setiap Kategori Emosi.....	24
Tabel 2. 3. Confusion matrix.....	37
Tabel 3. 1. Variabel Penelitian	41
Tabel 3. 2. Struktur data	42
Tabel 4. 1. Atribut Data Hasil Crawling	53
Tabel 4. 2. Dataset.....	54
Tabel 4.3. Hasil Data Cleaning	57
Tabel 4.4. Hasil <i>Case Folding</i>	58
Tabel 4.5. Hasil Normalisasi	59
Tabel 4.6. Hasil Tokenizing	60
Tabel 4.7. Hasil <i>Stopword Removal</i>	62
Tabel 4.8. Hasil Stemming Data	64
Tabel 4.9. Hasil Pelabelan Data	65
Tabel 4.10. Jumlah Data per Kategori.....	66
Tabel 4.11. Hasil Word2Vec	68
Tabel 4. 12. Hasil TF-IDF	70
Tabel 4.13. Jumlah Data Splitting	71
Tabel 4.14. Hasil Grid Search menggunakan Word2Vec	73
Tabel 4. 15. Hasil Grid Search menggunakan TF-IDF	74
Tabel 4. 16. Bobot kelas dan nilai C setiap kategori.....	76
Tabel 4. 17. Evaluasi model CS-SVM	78
Tabel 4. 18. Perbandingan Evaluasi Model.....	83
Tabel 4. 19. Perbandingan Precision per kategori Emosi	84
Tabel 4. 20. Perbandingan Recall per Kategori Emosi	84
Tabel 4. 21. Perbandingan F1-Score per Kategori Emosi.....	85
Tabel 4. 22. Perbandingan Word2Vec dengan TF-IDF	86

DAFTAR LAMPIRAN

Lampiran 1. LoA.....	107
Lampiran 2. Code dan Dashboard	109

DAFTAR NOTASI

T	:	<i>Total kata dalam korpus</i>
c	:	<i>Ukuran jendela konteks (window size)</i>
w_t	:	<i>Kata target pada posisi t</i>
w_{t+j}	:	<i>Kata konteks</i>
v_{w_I}	:	<i>Vektor input (embedding) dari kata target w_I</i>
v'_{w_O}	:	<i>Vektor output dari konteks w_O</i>
V	:	<i>Ukuran total kosakata (vocabulary size) di dalam korpus</i>
w	:	<i>Parameter hyperplane yang dicari</i>
b	:	<i>Nilai bias pada hyperplane</i>
n	:	<i>Jumlah data</i>
α_i	:	<i>Lagrange multiplier yang diperoleh dari optimasi fungsi SVM</i>
x_i^+	:	<i>Support vector kelas positif</i>
x_i^-	:	<i>Support vector kelas negatif</i>
d	:	<i>Jumlah dimensi total fitur yang ada dalam dataset</i>
$f(x)$	:	<i>Nilai fungsi keputusan klasifikasi</i>
k	:	<i>Indeks kelas emosi target</i>
$\frac{1}{2} \ w\ ^2$	:	<i>Batas margin yang ingin dimaksimalkan</i>
ξ_i	:	<i>Tingkat toleransi kesalahan untuk data ke-i yang diprediksi di sisi margin yang salah</i>
C^+	:	<i>Parameter pinalti biaya untuk salah klasifikasi pada kelas minoritas</i>
C^-	:	<i>Parameter pinalti biaya untuk salah klasifikasi pada kelas mayoritas</i>

x_i	:	<i>Vektor representasi teks word2vec dari komentar ke i</i>
y_i	:	<i>label actual dari emosi target</i>
$K(X_i, X_j)$	:	<i>Fungsi kernel</i>
X	:	<i>Data yang akan diklasifikasikan</i>
p	:	<i>Parameter Degree</i>
e	:	<i>Bilangan Euler</i>
γ	:	<i>Parameter Gamma</i>
$\tanh$	:	<i>Fungsi tangen</i>
TP (True Positive)	:	<i>Data positif yang dikelompokkan dengan benar sebagai positif</i>
TN (True Negative)	:	<i>Data negatif yang dikelompokkan dengan benar sebagai negatif</i>
FP (False Positive)	:	<i>Data negatif yang dikelompokkan dengan salah sebagai positif</i>
FN (False Negative)	:	<i>Data positif yang dikelompokkan dengan salah sebagai negatif.</i>