

**ANALISIS SENTIMEN TERHADAP TAGAR KABUR AJA
DULU DI MEDIA SOSIAL X MENGGUNAKAN SUPPORT
VECTOR MACHINE**



SKRIPSI

**MUHAMMAD BAYU PRATAMA
00000057853**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**ANALISIS SENTIMEN TERHADAP TAGAR KABUR AJA
DULU DI MEDIA SOSIAL X MENGGUNAKAN SUPPORT
VECTOR MACHINE**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

MUHAMMAD BAYU PRATAMA

00000057853

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Muhammad Bayu Pratama

Nomor Induk Mahasiswa : 00000057853

Program Studi : Informatika

Skripsi dengan judul:

Analisis Sentimen terhadap Tagar Kabur Aja Dulu di Media Sosial X Menggunakan Support Vector Machine

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 03 Juli 2025



(Muhammad Bayu Pratama)

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN TERHADAP TAGAR KABUR AJA DULU DI MEDIA SOSIAL X MENGGUNAKAN SUPPORT VECTOR MACHINE

oleh

Nama : Muhammad Bayu Pratama
NIM : 00000057853
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Selasa, 15 Juli 2025

Pukul 15.00 s/d 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang

Penguji

(Dr. Maria Irmira Prasetyowati, S.Kom.,
M.T.)

NIDN: 0725057201

(Marlinda Vasty Overbeek, S.Kom.,
M.Kom.)

NIDN: 0818038501

Pembimbing

(Eunike Endariahna Surbakti, S.Kom., M.T.I.)

NIDN: 0322099401

Ketua Program Studi Informatika,

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)

NIDN: 0315109103

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Muhammad Bayu Pratama
NIM : 00000057853
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Analisis Sentimen terhadap Tagar
Kabur Aja Dulu di Media Sosial
X Menggunakan Support Vector
Machine

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.

Tangerang, 03 Juli 2025

Yang menyatakan



Muhammad Bayu Pratama

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

HALAMAN PERSEMBAHAN / MOTTO

"With great power comes great responsibility."

Uncle Ben (In Every Spider-Man Universe)



KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala limpahan rahmat, karunia, dan kekuatan yang telah diberikan sehingga penulis dapat menyelesaikan penyusunan Tugas Akhir ini dengan judul "Analisis Sentimen Terhadap Tagar Kabur Aja Dulu di Sosial Media X Menggunakan Algoritma Support Vector Machine". Tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.

Mengucapkan terima kasih :

1. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
2. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
3. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
4. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
5. Ibu Eunike Endariahna Surbakti, S.Kom., M.T.I., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
6. Teman-teman saya, baik yang hingga saat ini masih kebersamai maupun yang pernah singgah dalam perjalanan ini, atas segala bentuk dukungan, semangat, dan pelajaran berharga yang telah diberikan selama proses penyusunan tugas akhir ini.'

Semoga karya ilmiah ini dapat memberikan kontribusi yang berarti dalam pengembangan ilmu pengetahuan, khususnya di bidang informatika dan machine learning, serta menjadi referensi bagi penelitian lanjutan di masa mendatang. Lebih dari itu, penulis berharap agar hasil penelitian ini juga dapat menjadi bahan refleksi bagi pihak-pihak yang berwenang dalam membenahi sistem pemerintahan dan kebijakan publik, dengan mempertimbangkan suara dan kebutuhan masyarakat

secara lebih adil, agar tidak semakin banyak individu yang merasa harus meninggalkan tanah air secara tidak resmi karena keterdesakan keadaan.

Tangerang, 03 Juli 2025



Muhammad Bayu Pratama



ANALISIS SENTIMEN TERHADAP TAGAR KABUR AJA DULU DI MEDIA SOSIAL X MENGGUNAKAN SUPPORT VECTOR MACHINE

Muhammad Bayu Pratama

ABSTRAK

Peningkatan aktivitas digital telah mendorong opini publik semakin luas tersebar di media sosial, termasuk topik-topik sensitif seperti keresahan sosial melalui frasa “*kabur aja dulu*”. Penelitian ini bertujuan untuk mengklasifikasikan sentimen dari teks berbahasa Indonesia di media sosial X terkait fenomena tersebut menggunakan algoritma *Support Vector Machine (SVM)*. Data dikumpulkan melalui *web crawling* pada periode Januari hingga April 2025, menghasilkan 4.236 *tweet*. Analisis dilakukan dengan empat skenario yang memvariasikan penerapan *stemming*, teknik penyeimbangan data *SMOTE oversampling*, dan konfigurasi *parameter grid*. Dataset dibagi dalam tiga rasio (60:20:20, 70:15:15, dan 80:10:10), dengan ekstraksi fitur menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)*, serta parameter terbaik ditentukan melalui *Grid Search*. Hasil terbaik diperoleh pada skenario dengan *stemming* tanpa *SMOTE*, parameter *sempit*, dan data split 80:10:10, menghasilkan nilai *precision*, *recall*, dan *f1-score* masing-masing 0,85 serta akurasi 85%, pada konfigurasi $C = 15$, *class_weight = None*, dan *max_iter = 5000*. Penelitian ini menunjukkan bahwa kombinasi *preprocessing* teks yang tepat dengan pemilihan parameter efisien mampu meningkatkan performa klasifikasi tanpa bergantung pada metode penyeimbangan data sintetik. Hasil klasifikasi juga menunjukkan dominasi sentimen negatif sebanyak 2.974 *tweet* (70,2%), diikuti netral 699 *tweet* (16,5%), dan positif 563 *tweet* (13,3%). Ini mencerminkan persepsi publik yang cenderung kritis atau pesimistis, sehingga menjadi indikator penting bagi evaluasi kebijakan terhadap faktor sosial dan ekonomi yang mempengaruhi munculnya narasi tersebut.

Kata kunci: #KaburAjaDulu, Analisis Sentimen, *Support Vector Machine*, TF-IDF.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

SENTIMENT ANALYSIS ON THE "KABUR AJA DULU" TOPIC ON SOCIAL MEDIA X USING SUPPORT VECTOR MACHINE

Muhammad Bayu Pratama

ABSTRACT

The increase in digital activities has amplified public opinions spread widely across social media, including sensitive topics such as social unrest expressed through phrases like "kabur aja dulu". This study aims to classify sentiments from Indonesian text data on the social media platform X related to this phenomenon using the Support Vector Machine (SVM) algorithm. Data were collected through web crawling from January to April 2025, resulting in 4,236 tweets. The analysis involved four scenarios varying the application of stemming, SMOTE oversampling techniques for data balancing, and configurations of the parameter grid. The dataset was divided into three data splits (60:20:20, 70:15:15, and 80:10:10), with feature extraction using the Term Frequency–Inverse Document Frequency (TF-IDF) method, and the best parameters were selected through a Grid Search. The optimal performance was achieved in the scenario with stemming without SMOTE, using a narrow parameter configuration and an 80:10:10 data split, yielding precision, recall, and f1-score values of 0.85 each, and an accuracy of 85%, with parameters set at $C = 15$, $class_weight = None$, and $max_iter = 5000$. This study demonstrates that the combination of appropriate text preprocessing and efficient parameter selection significantly improves classification performance without relying on synthetic data balancing methods. The classification results also revealed a predominance of negative sentiments in 2,974 tweets (70.2%), followed by neutral sentiments in 699 tweets (16.5%), and positive sentiments in 563 tweets (13.3%). This indicates a predominantly critical or pessimistic public perception, serving as an important indicator for policy evaluations concerning social and economic factors influencing the emergence of such narratives.

Keywords: #KaburAjaDulu, Sentiment Analysis, Support Vector Machine, TF-IDF

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR KODE	xiv
DAFTAR RUMUS	xv
DAFTAR LAMPIRAN	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Permasalahan	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	6
2.1 #KaburAjaDulu	6
2.2 Media Sosial X	6
2.3 Natural Language Processing	7
2.4 Analisis Sentimen	8
2.5 Term Frequency - Inverse Document Frequency (TF-IDF)	8
2.6 Support Vector Machine	9
BAB 3 METODOLOGI PENELITIAN	12
3.1 Literature Study	13
3.2 Data Scraping	13
3.3 Text Preprocessing	14
3.4 Data Labeling	16
3.5 Split Dataset	16
3.6 Feature Extraction using TF-IDF	17
3.7 SMOTE Oversampling	18
3.8 Sentiment Modeling with Support Vector Machine (SVM)	19
3.9 Testing Scenario	21
3.10 Data Visualization	21
BAB 4 HASIL DAN DISKUSI	23
4.1 Spesifikasi Sistem	23
4.2 <i>Data Scraping</i>	23
4.3 Text Preprocessing	25
4.3.1 Clear Coloumn	25
4.3.2 Install all library	26
4.3.3 Explorasi Data Awal	27
4.3.4 Case Folding	29
4.3.5 Cleansing	29

4.3.6	Slang Normalization	31
4.3.7	Tokenization	32
4.3.8	Stemming	33
4.4	Labeling with lexicon	34
4.5	Split Dataset	41
4.6	Feature Extraction using TF-IDF	43
4.7	SMOTE Oversampling	45
4.8	Evaluation SVM Model	46
4.9	Testing Scenario & Evaluation Model	51
4.9.1	Skenario 1 : Dengan <i>Stemming</i> dan Tanpa <i>SMOTE</i> <i>Oversampling</i>	53
4.9.2	Skenario 2 : Stemming & SMOTE Oversampling	54
4.9.3	Skenario 3 : Tanpa Stemming & Tanpa SMOTE Oversampling	56
4.9.4	Skenario 4 : Tanpa Stemming & Menggunakan SMOTE Oversampling	58
4.9.5	Kesimpulan Evaluasi Model	60
4.10	Gambaran Analisis Sentimen	61
4.10.1	Visualisasi Sentimen Kata Negatif	62
4.10.2	Visualisasi Sentimen Kata Positif	64
4.10.3	Visualisasi Sentimen Kata Netral	65
4.10.4	Uji Kata Deteksi Sentimen	66
BAB 5	SIMPULAN DAN SARAN	68
5.1	Simpulan	68
5.2	Saran	69
DAFTAR PUSTAKA		70



DAFTAR TABEL

Tabel 4.1	Evaluasi Model : Skenario 1 (Parameter Luas)	53
Tabel 4.2	Evaluasi Model : Skenario 1 (Parameter Sempit)	54
Tabel 4.3	Evaluasi Model : Skenario 2 (Parameter Luas)	55
Tabel 4.4	Evaluasi Model : Skenario 2 (Parameter Sempit)	55
Tabel 4.5	Evaluasi Model : Skenario 3 (Parameter Luas)	56
Tabel 4.6	Evaluasi Model : Skenario 3 (Parameter Sempit)	57
Tabel 4.7	Evaluasi Model : Skenario 4 (Parameter Luas)	58
Tabel 4.8	Evaluasi Model : Skenario 4 (Parameter Sempit)	59
Tabel 4.9	Perbandingan Performa Terbaik Setiap Skenario	61



DAFTAR GAMBAR

Gambar 3.1	Alur Metodologi Penelitian	12
Gambar 3.2	Alur <i>Data Scraping</i>	14
Gambar 3.3	Alur <i>Text Preprocessing</i>	15
Gambar 3.4	Alur <i>Feature Extraction using TF-IDF</i>	18
Gambar 3.5	<i>Flowchart SVM</i>	19
Gambar 3.6	<i>Flowchart SVM with Grid Search</i>	20
Gambar 4.1	Tampilan data setelah konversi ke Excel (.xlsx)	26
Gambar 4.2	Visualisai hasil <i>labeling</i> dengan <i>lexicon</i>	40
Gambar 4.3	Distribusi Data Setelah Pembagian	43
Gambar 4.4	Dimensi hasil vektorisasi	44
Gambar 4.5	Sebelum dan Sesudah <i>SMOTE Oversampling</i>	46
Gambar 4.6	Hasil evaluasi model SVM)	50
Gambar 4.7	Hasil evaluasi model SVM)	52
Gambar 4.8	Grafik Perbandingan Evaluasi Model : Skenario 1 (Parameter Luas)	53
Gambar 4.9	Grafik Perbandingan Evaluasi Model : Skenario 1 (Parameter Sempit)	54
Gambar 4.10	Grafik Perbandingan Evaluasi Model : Skenario 2 (Parameter Luas)	55
Gambar 4.11	Grafik Perbandingan Evaluasi Model : Skenario 2 (Parameter Sempit)	56
Gambar 4.12	Grafik Perbandingan Evaluasi Model : Skenario 3 (Parameter Luas)	57
Gambar 4.13	Grafik Perbandingan Evaluasi Model : Skenario 3 (Parameter Sempit)	58
Gambar 4.14	Grafik Perbandingan Evaluasi Model : Skenario 4 (Parameter Luas)	59
Gambar 4.15	Grafik Perbandingan Evaluasi Model : Skenario 4 (Parameter Sempit)	60
Gambar 4.16	<i>Pie chart visualization</i>	62
Gambar 4.17	<i>Word Cloud Negative Sentiment</i>	64
Gambar 4.18	<i>Word Cloud Positive Sentiment</i>	65
Gambar 4.19	<i>Word Cloud Neutral Sentiment</i>	66
Gambar 4.20	Kalimat hasil prediksi setimen	67

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR KODE

Kode 4.1	<i>Authorized token</i>	24
Kode 4.2	<i>Import Phyton package and Install Node.js</i>	24
Kode 4.3	<i>Crawling Data Code</i>	25
Kode 4.4	<i>Install and Import All Library</i>	27
Kode 4.5	<i>Mount Google Drive</i>	27
Kode 4.6	Membaca dataset dari Google Drive	28
Kode 4.7	Mengubah Nama Kolom Menjadi tweet	28
Kode 4.8	Pemeriksaan Nilai Kosong	28
Kode 4.9	Pemeriksaan Data Duplikat	28
Kode 4.10	Pemeriksaan Ukuran Dataset	29
Kode 4.11	<i>Case Folding</i>	29
Kode 4.12	<i>Cleansing</i>	30
Kode 4.13	<i>Load slang dictionary</i>	31
Kode 4.14	<i>Normalize slang</i>	31
Kode 4.15	<i>Tokenization</i>	32
Kode 4.16	<i>Stemming</i>	33
Kode 4.17	<i>Read lexicon</i>	34
Kode 4.18	<i>Load lexicon</i>	36
Kode 4.19	<i>Load emoticon lexicon</i>	36
Kode 4.20	<i>Lexicon sentimen from tokens</i>	37
Kode 4.21	<i>Cek token data</i>	39
Kode 4.22	<i>Labeling</i>	39
Kode 4.23	<i>Labeling Visualization</i>	40
Kode 4.24	<i>Manual audit</i>	40
Kode 4.25	Konversi <i>string literal</i> menjadi <i>list token</i> Python	41
Kode 4.26	Penggabungan <i>token</i> menjadi <i>string</i> utuh	41
Kode 4.27	Pemisahan fitur dan label	41
Kode 4.28	Split pertama: <i>training</i> dan <i>temporary</i> (70:30)	42
Kode 4.29	Split kedua: <i>validation</i> dan <i>test</i> dari data sementara	42
Kode 4.30	Pengecekan distribusi hasil <i>split</i> data	42
Kode 4.31	Inisialisasi dan transformasi data menggunakan TF-IDF	43
Kode 4.32	Penerapan dan Hasil <i>SMOTE Oversampling</i>	45
Kode 4.33	SVM Training Model	47
Kode 4.34	Parameter Grid Luas	51
Kode 4.35	Parameter Grid Sempit	51
Kode 4.36	<i>Pie chart visualization</i>	62
Kode 4.37	<i>Word Cloud Negative</i>	63
Kode 4.38	<i>Word Cloud Positive</i>	64
Kode 4.39	<i>Word Cloud Neutral</i>	65
Kode 4.40	<i>Sentiment Prediction</i>	66

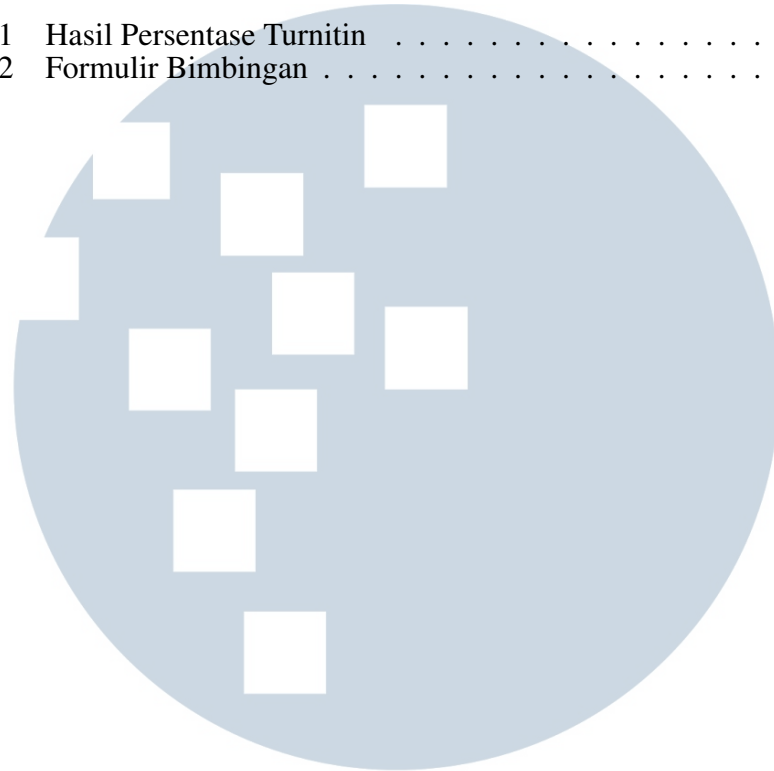
DAFTAR RUMUS

Rumus 2.1	Rumus TF-IDF	9
Rumus 2.2	Rumus SVM 1	10
Rumus 2.3	Rumus SVM 2	10
Rumus 2.4	Rumus SVM 3	10
Rumus 2.5	Rumus SVM 4	10
Rumus 2.6	Rumus SVM 5	11
Rumus 2.7	Rumus SVM 6	11



DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin	73
Lampiran 2	Formulir Bimbingan	74



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA