

**ANALISIS SENTIMEN KOMENTAR INSTAGRAM
MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE*
(SVM)**

SKRIPSI

Oleh:

**SONTIAR ESERIA TAMPUBOLON
NIM. 221112223**



**PROGRAM STUDI S-1 TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2026**

INSTAGRAM COMMENT SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE (SVM) ALGORITHM

FINAL RESEARCH

By:

**SONTIAR ESERIA TAMPUBOLON
ID NUMBER. 221112223**



**MAJOR OF S-1 INFORMATICS ENGINEERING
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2026**

LEMBARAN PENGESAHAN

**ANALISIS SENTIMEN KOMENTAR INSTAGRAM MENGGUNAKAN
ALGORITMA *SUPPORT VECTOR MACHINE* (SVM)**

SKRIPSI

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Sarjana
Program Studi S-1 Teknik Informatika

Oleh:

SONTIAR ESERIA TAMPUBOLON
NIM. 221112223

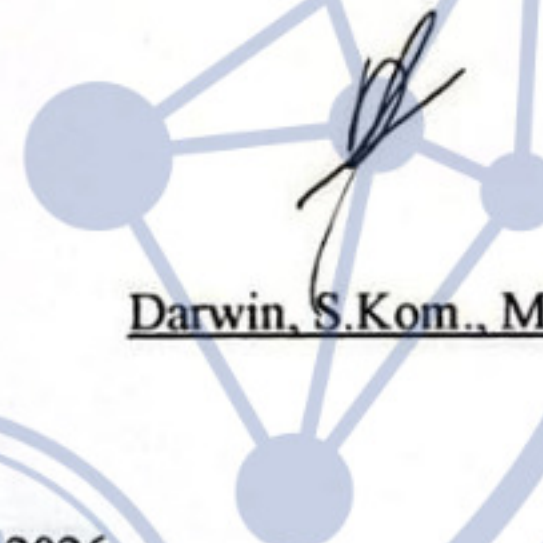
Disetujui Oleh:

Dosen Pembimbing I,

Dosen Pembimbing II,



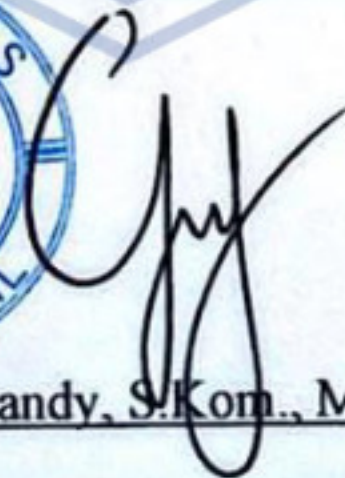
Mirza Ilhami, S.Kom., M.TI



Darwin, S.Kom., M.Kom

Medan, 23 Juli 2026
Diketahui dan Disahkan Oleh:

Ketua Program Studi
S-1 Teknik Informatika



Carles Juliandy, S.Kom., M.Kom

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Sontiar Eseria Tampubolon

NIM : 221112223

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Analisis Sentimen Komentar Instagram Menggunakan Algoritma Support Vector Machine (SVM)

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 1 juli 2026

Saya yang membuat pernyataan,


BREA2AOX047544249

Sontiar Eseria Tampubolon

ANALISIS SENTIMEN KOMENTAR INSTAGRAM MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

Abstrak

Instagram merupakan media sosial dengan volume interaksi komentar yang sangat tinggi, sehingga rentan menjadi wadah munculnya komentar bersifat cyberbullying. Sifat komentar yang tidak terstruktur, beragam, dan subjektif membuat analisis manual menjadi tidak efisien, sehingga dibutuhkan pendekatan otomatis melalui analisis sentimen. Penelitian ini bertujuan mengukur performa algoritma Support Vector Machine (SVM) dalam mengklasifikasikan komentar cyberbullying pada Instagram berbahasa Indonesia, serta menganalisis pengaruh kombinasi Word TF-IDF dan Character TF-IDF dengan delapan variasi nilai parameter C terhadap performa model. Dataset yang digunakan berjumlah 650 komentar dari Kaggle dengan proporsi seimbang antara komentar non-bullying dan bullying. Tahapan penelitian meliputi preprocessing, ekstraksi fitur N-gram yang dibobotkan dengan kombinasi Word dan Character TF-IDF, kemudian diklasifikasikan menggunakan SVM kernel linear dan dievaluasi dengan 5-Fold Cross Validation. Hasil eksperimen menunjukkan bahwa parameter $C = 0,5$ menghasilkan performa terbaik dengan rata-rata akurasi 86,77%, precision 86,81%, recall 86,77%, dan F1-score 86,76%. Hasil ini membuktikan bahwa kombinasi Word dan Character TF-IDF mampu menangkap variasi penulisan bahasa tidak baku, sehingga model SVM yang dibangun terbukti andal dan stabil dalam mendeteksi cyberbullying di Instagram.

Kata Kunci: Analisis Sentimen, Cyberbullying, Support Vector Machine, TF-IDF, Instagram

Abstract

Instagram is a social media platform with an extremely high volume of comment interactions, making it prone to cyberbullying comments. Because such comments are unstructured, diverse, and subjective, manual analysis becomes inefficient, requiring an automated approach through sentiment analysis. This research aims to measure the performance of the Support Vector Machine (SVM) algorithm in classifying cyberbullying comments on Indonesian-language Instagram, and to analyze the effect of combining Word TF-IDF and Character TF-IDF with eight variations of parameter C on model performance. The dataset consists of 650 comments from Kaggle with a balanced distribution between non-bullying and bullying comments. The research stages include preprocessing, N-gram feature extraction weighted using a combination of Word and Character TF-IDF, followed by classification using a linear-kernel SVM evaluated with 5-Fold Cross Validation. The results show that parameter $C = 0.5$ produced the best performance, achieving an average accuracy of 86.77%, precision of 86.81%, recall of 86.77%, and F1-score of 86.76%. These findings prove that combining Word and Character TF-IDF effectively captures non-standard language variations, making the resulting SVM model reliable and stable in detecting cyberbullying on Instagram.

Keywords: Sentiment Analysis, Cyberbullying, Support Vector Machine, TF-IDF, Instagram

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, atas segala berkat, rahmat, dan penyertaan-Nya yang tiada henti, sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan tepat waktu. Perjalanan menyusun skripsi ini bukanlah hal yang mudah, penuh dengan pasang surut, air mata, rasa lelah, dan juga rasa syukur di setiap prosesnya. Namun berkat dukungan dari banyak pihak, akhirnya penulis mampu melalui semuanya dan menuntaskan tanggung jawab ini.

Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan Strata Satu (S-1) pada Program Studi Teknik Informatika, Fakultas Informatika, Universitas Mikroskil Medan.

Pada kesempatan yang berharga ini, dengan penuh ketulusan hati, penulis ingin menyampaikan rasa terima kasih yang sedalam-dalamnya kepada:

1. Bapak **Mirza Ilhami, S.Kom., M.TI**, selaku Dosen Pembimbing I, yang telah dengan sabar meluangkan waktu, tenaga, dan pikiran untuk membimbing, mengarahkan, serta memberikan saran dan masukan yang sangat berharga kepada penulis selama proses penyusunan skripsi ini.
2. Bapak **Darwin, S.Kom., M.Kom.**, selaku Dosen Pembimbing II, yang telah dengan sabar meluangkan waktu, tenaga, dan pikiran untuk membimbing, mengarahkan, serta memberikan saran dan masukan yang sangat berharga kepada penulis selama proses penyusunan skripsi ini.
3. Bapak **Kristian Telaumbanua, S.T., M.T.**, selaku Dosen Penguji I, yang telah memberikan kritik, saran, dan masukan yang sangat berharga demi penyempurnaan skripsi ini.
4. Bapak **Kelvin, S.Kom., M.Kom.**, selaku Dosen Penguji II, yang telah memberikan kritik, saran, dan masukan yang sangat berharga demi penyempurnaan skripsi ini.
5. Bapak **Hardy, S.Kom., M.Sc., Ph.D., M.Kom.**, selaku Rektor Universitas Mikroskil Medan.
6. Bapak **Sunaryo Winardi, S.Kom., M.T.**, selaku Dekan Fakultas Informatika Universitas Mikroskil Medan.
7. Bapak **Carles Juliandy, S.Kom., M.Kom.**, selaku Ketua Program Studi S-1 Teknik Informatika Fakultas Informatika Universitas Mikroskil Medan.

8. Seluruh dosen dan staf Fakultas Informatika Universitas Mikroskil Medan yang telah memberikan ilmu pengetahuan, bimbingan, serta pelayanan akademik selama penulis menempuh pendidikan.
9. Kedua orang tua, adik-adik, sahabat sekaligus saudara penulis, serta diri penulis sendiri yang telah menjadi sumber doa, dukungan, semangat, kasih sayang, dan kekuatan dalam setiap proses hingga skripsi ini dapat terselesaikan.

Penulis menyadari bahwa skripsi ini masih memiliki berbagai keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan sebagai bahan perbaikan di masa mendatang. Akhir kata, penulis berharap skripsi ini dapat memberikan manfaat serta menambah wawasan bagi pembaca yang membutuhkan.

Medan, 1 juli 2026

Penulis,

Sontiar Eseria Tampubolon

DAFTAR ISI

ABSTRAK.....	i
KATA PENGANTAR	ii
DAFTAR ISI	iv
DAFTAR GAMBAR	vii
DAFTAR TABEL.....	viii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan.....	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
BAB II KAJIAN LITERATUR.....	5
2.1 Hasil Penelitian Terdahulu.....	5
2.2 Analisis Sentimen	6
2.2.1 Tingkat Analisis Sentimen	6
2.2.2 Kategori Sentimen.....	7
2.2.3 Pendekatan Analisis Sentimen	7
2.3 Instagram	8
2.3.1 Karakteristik Komentar Instagram.....	9
2.3.2 <i>Dataset</i> Komentar Instagram	10
2.4 <i>Text Mining</i>	10
2.5 <i>Preprocessing</i>	10
2.6 <i>Ekstraksi Fitur</i>	12
2.6.1 TF-IDF (Term Frequency – Inverse Document Frequency).....	12
2.6.2 Normalisasi Vektor TF-IDF	14
2.6.3 N-gram (Unigram dan Bigram).....	14

2.7. K-Fold Cross Validation	15
2.8 <i>Support Vector Machine</i> (SVM).....	15
2.8.1 Cara Kerja SVM.....	16
2.8.2 Fungsi Kernel SVM	17
2.8.3 Parameter SVM.....	18
2.9 Pustaka Sastrawi	19
2.10 Kamus <i>Slang</i> (Normalisasi Kata Tidak Baku).....	20
2.11 Char TF-IDF (<i>Character-level</i> TF-IDF).....	20
2.12 <i>Confusion Matrix</i>	21
BAB III METODOLOGI PENELITIAN.....	24
3.1 Jenis Penelitian	24
3.2 Data dan Sumber Data	24
3.3 Teknik Pengumpulan Data.....	24
3.4 Variabel Penelitian.....	25
3.5 Pemetaan Label <i>Dataset</i> dan Statistik Data.....	26
3.6 Teknik Analisis Data	27
3.6.1 Case folding.....	31
3.6.2 Data Cleaning.....	31
3.6.3 Tokenizing.....	32
3.6.4. Standardizing.....	32
3.6.5. Stopword Removal.....	33
3.6.6. Stemming	34
3.7 Ekstraksi Fitur.....	35
3.7.1 Pembentukan Fitur <i>N-Gram</i> (<i>Unigram</i> dan <i>Bigram</i>).....	35
3.7.2 TF-IDF	37
3.7.3 Char TF-IDF.....	44
3.8 <i>Testing</i>	45

3.9 Evaluasi Model	48
3.9.1 Confusion Matrix	49
BAB IV HASIL DAN PEMBAHASAN	52
4.1 Data Penelitian	52
4.2 <i>Preprocessing</i>	54
4.2.1 Case Folding.....	54
4.2.2 Data Cleaning.....	55
4.2.3 Tokenizing.....	56
4.2.4 Standardizing.....	57
4.2.5 Stopword Removal.....	58
4.2.6 Stemming	59
4.3.1 N-gram (Unigram dan Bigram).....	62
4.3.2 Representasi Biner Fitur.....	63
4.3.2 Term Frequency (TF).....	64
4.3.3 Inverse Document Frequency (IDF)	66
4.3.4 TF-IDF	67
4.3.5 Normalisasi TF-IDF	68
4.4 Sebaran Kata Menggunakan WordCloud	69
4.5 Parameter C.....	71
4.5.1 K-Fold Cross Validation	72
4.6 Confusion Matrix.....	74
4.7 Evaluasi Metrik Performa Akhir.....	76
BAB V KESIMPULAN DAN SARAN	79
5.1 Kesimpulan	79
5.2 Saran	79
DAFTAR PUSTAKA	80
DAFTAR RIWAYAT HIDUP.....	87

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Hyperplane dan Margin pada SVM [46]	17
Gambar 2. 2 Perbandingan SVM Linear dan Non-Linear (Kernel Trick) [51]	18
Gambar 2. 3 Confusion Matrix Klasifikasi 2 Kelas Sentimen [61]	22
Gambar 3. 1 Halaman Awal Kaggle	25
Gambar 3. 2 Flowchart proses (a) training, (b) testing	28
Gambar 3. 3 Flowchart tahapan prapemrosesan data	30
<u>Gambar 4 1 Grafik Distribusi Label Dataset Kategori Komentar</u>	53
Gambar 4 2 <i>Source Code Preprocessing Case Folding</i>	55
Gambar 4 3 <i>Source Code Preprocessing Data Cleaning</i>	56
Gambar 4 4 <i>Source Code Preprocessing Tokenizing</i>	57
Gambar 4 5 <i>Source Code Preprocessing Standardizing</i>	58
Gambar 4 6 <i>Source Code Preprocessing Stopword Removal</i>	59
Gambar 4 7 <i>Source Code Preprocessing Stopword Removal</i>	60
Gambar 4 8 Grafik Perbandingan Komponen Kata Sebelum dan Sesudah Preprocessing .	61
Gambar 4 9 Visualisasi WordCloud per Kelas Sentimen	70
Gambar 4 10 Grafik Perbandingan Akurasi 8 Variasi Nilai C	72
Gambar 4 11 Grafik Tren Stabilitas Performa Akurasi Model SVM	74
Gambar 4 12 Heatmap Visualisasi Confusion Matrix Agregat	76
Gambar 4 13 Grafik Fitur Kata Paling Berpengaruh pada Klasifikasi SVM	78

DAFTAR TABEL

Tabel 2. 1 Tabel Penelitian Terdahulu.....	5
Tabel 2. 2 Confusion Matrix.....	21
Tabel 3. 1 Statistik Dataset yang Digunakan.....	27
Tabel 3. 2 Data Mentah.....	29
Tabel 3. 3 Data yang telah diberi label	29
Tabel 3. 4 Tahapan Case folding	31
Tabel 3. 5 Tahapan Data Cleaning.....	31
Tabel 3. 6 Tahapan Tokenizing	32
Tabel 3. 7 Tahapan Standardizing	33
Tabel 3. 8 Tahapan Stopword Removal.....	33
Tabel 3. 9 Tahapan Stemming.....	34
Tabel 3. 10 Hasil Prapemrosesan.....	35
Tabel 3. 11 Hasil Unigram dan Bigram.....	36
Tabel 3. 12 Hasil Perhitungan TF.....	37
Tabel 3. 13 Hasil perhitungan IDF	39
Tabel 3. 14 Hasil Ekstraksi Fitur menggunakan TF-IDF	40
Tabel 3. 15 Hasil Normalisasi dari Ekstraksi Fitur TF-IDF	41
Tabel 3. 16 Data yang digunakan untuk Training.....	42
Tabel 3. 17 Representasi Vektor Data Training.....	44
Tabel 3. 18 Contoh Char N-Gram (ngram_range = 2,5)	45
Tabel 3. 19 Contoh Data Testing	46
Tabel 3. 20 Data Testing Hasil Prapemrosesan	46
Tabel 3. 21 Hasil Ekstraksi Fitur TF-IDF Data Testing	46
Tabel 3. 22 Hasil Normalisasi TF-IDF Data Testing.....	47
Tabel 3. 23 Hasil Klasifikasi Testing.....	48
Tabel 3. 24 Contoh Confusion Matrix Hasil Evaluasi Model SVM.....	49
Tabel 4. 1 Distribusi Label Kategori Dataset.....	52
Tabel 4. 2 Sampel Data Komentar Instagram.....	53
Tabel 4. 3 Aturan Pemetaan Label Kategori (Label Encoding).....	54
Tabel 4. 4 Hasil Preprocessing Case Folding	55
Tabel 4. 5 Hasil Preprocessing Data Cleaning.....	56
Tabel 4. 6 Hasil Preprocessing Tokenizing	57

Tabel 4. 7 Hasil Preprocessing Standardizing	58
Tabel 4. 8 Hasil Preprocessing Stopword Removal.....	59
Tabel 4. 9 Hasil Preprocessing Stemming	60
Tabel 4. 10 Rekapitulasi Hasil Keseluruhan Preprocessing Teks.....	62
Tabel 4. 11 Hasil Ekstraksi Fitur Unigram dan Bigram	63
Tabel 4. 12 Hasil Output Representasi Biner Fitur	65
Tabel 4. 13 Hasil Perhitungan Nilai Term Frequency	66
Tabel 4. 14 Hasil Perhitungan Pembobotan Nilai IDF	67
Tabel 4. 15 Hasil Akhir Pembobotan Nilai TF-IDF	68
Tabel 4. 16 Hasil Normalisasi TF-IDF (L2-Norm) (D1 s/d D3)	70
Tabel 4. 17 Hasil Eksperimen 8 Variasi Nilai Parameter C	71
Tabel 4. 18 Hasil Evaluasi Model SVM dengan 5-Fold Cross Validation ($C = 0,5$)	73
Tabel 4. 19 Confusion Matrix Agregat (5-Fold CV, $n = 650$ data)	74
Tabel 4. 20 Ringkasan Metrik Evaluasi Performa Model SVM Akhir.....	76

