

**ANALISIS BIG DATA UNTUK MENDETEKSI EMOSI PENGGUNA
MELALUI MEDIA SOSIAL**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh

DEWI NUR ISTIKHOMAH

21.83.0662

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**ANALISIS BIG DATA UNTUK MENDETEKSI EMOSI PENGGUNA
MELALUI MEDIA SOSIAL**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh

DEWI NUR ISTIKHOMAH

21.83.0662

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS BIG DATA UNTUK MENDETEKSI EMOSI PENGGUNA MELALUI MEDIA SOSIAL

yang disusun dan diajukan oleh

Dewi Nur Istikhomah

21.83.0662

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 29 Juli 2025

Dosen Pembimbing,



Tonny Hidayat, S.Kom., M.Kom, Ph.D.

NIK. 190302182

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS BIG DATA UNTUK MENDETEKSI EMOSI PENGGUNA MELALUI MEDIA SOSIAL

yang disusun dan diajukan oleh

Dewi Nur Istikhomah

21.83.0662

Telah dipertahankan di depan Dewan Penguji
pada tanggal 29 Juli 2025

Susunan Dewan Penguji

Nama Penguji

Wahid Miftahul Ashari, S.Kom., M.T.
NIK. 190302452

Muhammad Koprari, S.Kom., M.Eng.
NIK. 19030245

Tonny Hidayat, S.Kom., M.Kom, Ph.D.
NIK. 190302182

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 29 Juli 2025

DEKAN FAKULTAS ILMU KOMPUTER

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106



HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dewi Nur Istikhomah
NIM : 21.83.0662

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS BIG DATA UNTUK MENDETEKSI EMOSI PENGGUNA
MELALUI MEDIA SOSIAL**

Dosen Pembimbing : Tonny Hidayat, S.Kom., M.Kom, Ph.D.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 29 Juli 2025

Yang Menyatakan,



Dewi Nur Istikhomah

HALAMAN PERSEMBAHAN

Skripsi ini dipersembahkan dengan rasa hormat dan penuh rasa syukur kepada:

1. Kedua orang tua
Atas cinta, dukungan, dan doa yang tak pernah henti mengiringi setiap langkah saya.
2. Dosen
Yang telah memberikan ilmu, bimbingan dan inspirasi selama saya kuliah di Universitas Amikom Yogyakarta
3. Pemilik NIM 21.83.0736 yang sudah mensupport dan menemani bimbingan selama pengerjaan skripsi ini.
4. Diri sendiri
Yang sudah berjuang sampai saat ini.

KATA PENGANTAR

Puji Syukur saya panjatkan ke hadirat Allah SWT yang telah memberikan rahmat dan kaunia sehingga saya dapat menyelesaikan skripsi ini dengan baik skripsi yang berjudul “Analisis Big Data Untuk Mendeteksi Emosi Pengguna Melalui Media Sosial” ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik di Universitas Amikom Yogyakarta.

Saya mengucapkan terima kasih yang sebesar besarnya kepada:

1. Tonny Hidayat, S.Kom., M.Kom, Ph.D., selaku Dosen Pembimbing, yang dengan sabar memberikan arahan, bimbingan, dan motivasi selama proses penyusunan skripsi ini.
2. Tim Dosen Penguji, yang telah memberikan masukan konstruktif dan kritik yang membangun untuk perbaikan kualitas skripsi ini.
3. Kedua orang tua saya, yang selalu memberikan dukungan moral, doa, dan kasih sayang sepanjang perjalanan studi ini.
4. Pemilik NIM 21.83.0736, yang telah mau menemani dari awal masuk kuliah sampai saat ini.

Akhir kata, saya berharap skripsi ini dapat memberikan manfaat bagi semua pihak yang berkepentingan, terutama dalam, mendeteksi emosi. Saya juga menyadari bahwa skripsi ini masih jauh dari sempurna, oleh karena itu saya sangat mengharapkan kritik dan saran yang membangun demi perbaikan di masa mendatang.

Yogyakarta, 11 Juli 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	9
2.2.1 Emosi	9
2.2.2 Media Sosial	9
2.2.3 Twitter	10
2.2.4 Big Data.....	10
2.2.5 <i>Text Preprocessing</i>	10
2.2.6 <i>Confusion Matrix</i>	11
2.2.7 Algoritma TF-IDF	13

2.2.8	Navie Bayes	14
2.2.9	Training Klasifikasi Navie Bayes	15
2.2.10	Testing Klasifikasi Navie Bayes	16
BAB III METODE PENELITIAN		17
3.1	Objek Penelitian.....	17
3.2	Alur Penelitian	17
3.2.1	Flowchart Alur Penelitian.....	17
3.2.2	Flowchart Alur Kerja Sistem.....	18
3.3	Analisis Kebutuhan.....	20
3.3.1	Kebutuhan fungsional.....	20
3.3.2	Kebutuhan Non Fungsional.....	21
3.3.3	<i>Human Resource</i>	22
3.4	Alat dan Bahan	22
BAB IV HASIL DAN PEMBAHASAN		25
4.1	Implementasi	25
4.2	Pengumpulan Data.....	25
4.2.1	Perlabelan Data.....	26
4.2.2	<i>Data Preprocessing</i>	27
4.2.2.1	<i>Case Folding</i>	28
4.2.2.2	<i>Tokenizing</i>	29
4.2.2.3	<i>Stopword Removal</i>	31
4.2.2.4	Normalisasi.....	33
4.2.2.5	Stemming	34
4.2.2.6	Ekstraksi Fitur	35
4.2.2.7	Pembobotan TF-IDF	37
4.2.2.8	Klasifikasi.....	39
4.2.2.9	Evaluasi model Navie Bayes.....	40
BAB V PENUTUP		53
5.1	Kesimpulan	54
5.2	Saran	54
REFERENSI		55
LAMPIRAN		57

DAFTAR TABEL

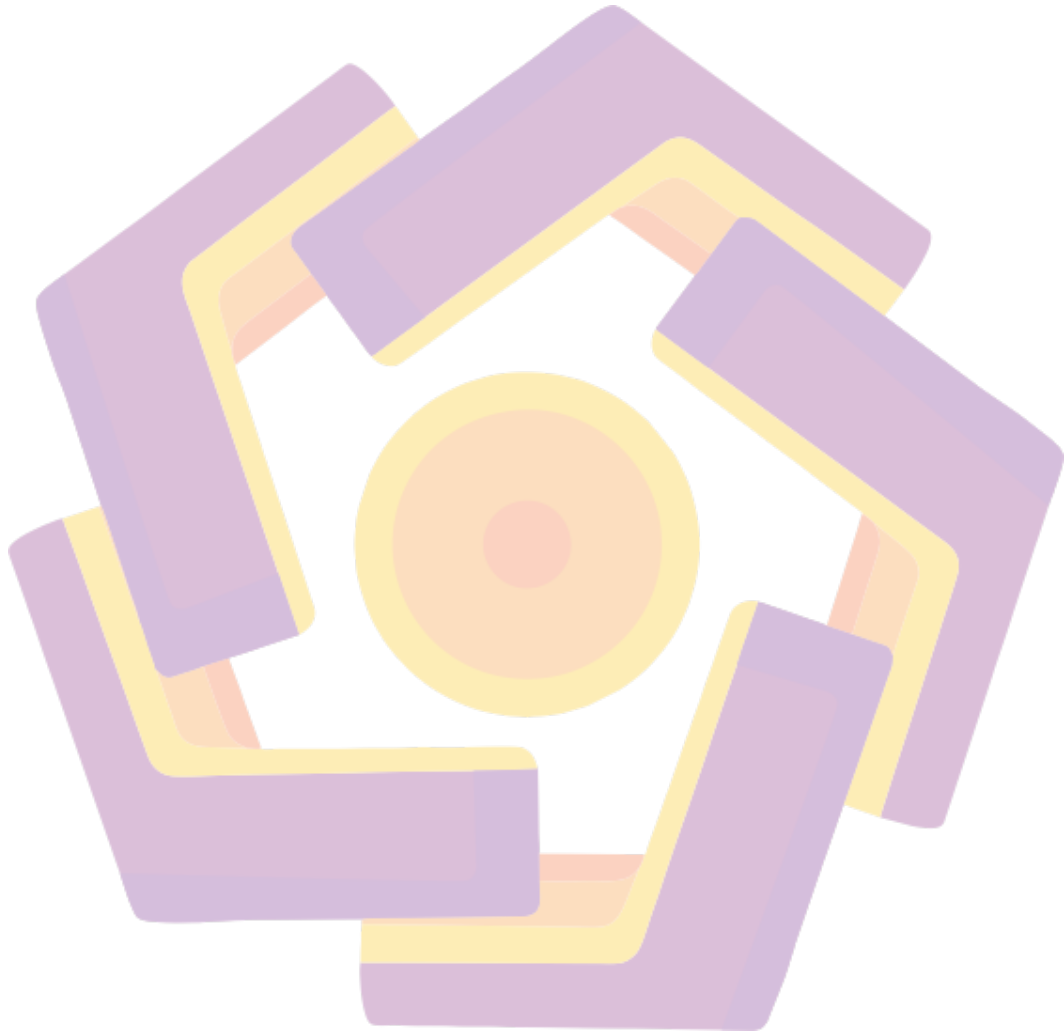
Tabel 2.1	Keaslian Penelitian	7
Tabel 2.2	Confusion Matrix	12
Tabel 3.1	Perangkat Lunak	21
Tabel 3.2	Human Resource	22
Tabel 4.1	Pelabelan Dataset	27
Tabel 4.2	Hasil Proses Case Folding	28
Tabel 4.3	Hasil Proses Tokenizing	30
Tabel 4.4	List Stopword Removal	31
Tabel 4.5	Hasil Proses Stopword Removal	32
Tabel 4.6	List Normalisasi Text	34
Tabel 4.7	Hasil Normalisasi Kata	34
Tabel 4.8	Hasil Proses Stemming	35
Tabel 4.9	Beberapa data ulasan	36
Tabel 4.10	Beberapa hasil pembobotan TF-IDF	37
Tabel 4.11	Sampel dan ilustrasi TF - IDF	38
Tabel 4.12	Rasio Pembagian Dataset	39
Tabel 4.13	Hasil Confusion Matrix Skrenario 1	41
Tabel 4.14	Hasil Confusion Matrix Skrenario 2	43
Tabel 4.15	Hasil Confusion Matrix Skrenario 3	46
Tabel 4.16	Rangkuman Evaluasi Performa Navie Bayes	49

DAFTAR GAMBAR

Gambar 3.1	Alur Penelitian	18
Gambar 3.2	Flowchart Alur Kerja Sistem	20
Gambar 4.1	Repositori GitHub	25
Gambar 4.2	Hasil repositori GitHub	26
Gambar 4.3	Hasil Pelabelan Data	26
Gambar 4.4	Diagram pie pelabelan	27
Gambar 4.5	Script Case Folding	28
Gambar 4.6	Library Pada Proses Tokenizing	29
Gambar 4.7	Script Tokenizing	30
Gambar 4.8	Import Library Stopword Removal	31
Gambar 4.9	Script Stopword Removal	32
Gambar 4.10	Script Normalisasi Kata	33
Gambar 4.11	Script Stemming Data	35
Gambar 4.12	Script proses ekstrasi fitur	36
Gambar 4.13	Visualisasi Wordcloud	38
Gambar 4.14	Klasifikasi menggunakan algoritma Navie Bayes	39
Gambar 4.15	Script python dari model Navie Bayes	40
Gambar 4.16	Visualisasi Heatmap Confusion Matrix Scenario 1	41
Gambar 4.17	Script menghitung nilai akurasi, recall	42
Gambar 4.18	Hasil script	43
Gambar 4.19	Visualisasi Heatmap Confusion Matrix Scenario 2	44
Gambar 4.20	Script menghitung nilai akurasi, recall	45
Gambar 4.21	Hasil Script	45
Gambar 4.22	Visualisasi Heatmap Confusion Matrix Scenario 3	47
Gambar 4.23	Script menghitung nilai akurasi, recall	47
Gambar 4.24	Hasil Script	48
Gambar 4.25	Grafik Perbandingan Accuracy, Precision, recall	50

DAFTAR LAMPIRAN

Lampiran 1. Source Code Preprocessing Data.....	57
Lampiran 2. Source Code Tokenisasi dan Stopword Removal	57
Lampiran 3. Source Code Ekstraksi Fitur TF-IDF	57
Lampiran 4. Source Code Pelatihan Model Naive Bayes.....	58



DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
API	Application Programming Interface
IDF	Inverse Document Frequency
NLTK	Natural Language Toolkit (Library Python)
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negatif

DAFTAR ISTILAH

Preprocessing	Proses pembersihan dan persiapan data sebelum dilakukan analisis.
Case Folding	Mengubah seluruh huruf dalam teks menjadi huruf kecil (lowercase).
Tokenisasi	Memecah teks menjadi potongan-potongan kata (token).
Stopword- Removal	Menghapus kata-kata umum yang tidak bermakna penting dalam analisis.
Normalisasi	Mengubah kata tidak baku menjadi kata baku.
Stemming	Mengubah kata berimbuhan ke bentuk dasar.

INTISARI

Perkembangan pesat teknologi informasi dan komunikasi telah menjadikan media sosial sebagai platform utama bagi individu untuk berbagi pemikiran, perasaan, dan pengalaman. Emosi yang diekspresikan melalui teks media sosial twitter, merupakan sumber data yang kaya untuk dianalisis. Deteksi emosi ini penting seperti pemasaran, psikologi dan analisis sosial. Dengan volume yang terus bertambah, diperlukan pendekatan komprehensif seperti pembelajaran mesin dan pemrosesan bahasa alami (NLP).

Penelitian ini bertujuan untuk menganalisis dan mendeteksi emosi pengguna melalui data teks yang diambil dari github dengan memanfaatkan pendekatan Big Data. Metode Navie Bayes dipilih sebagai algoritma klasifikasi utama karena kesederhanaan dan kemampuannya dalam memberikan hasil yang akurat pada klasifikasi text. Tahapan prapemrosesan meliputi, stopwords removal, stemming, dan pembobotan kata menggunakan Term Frequency-Inverse Document Frequency (TF-IDF), yang merupakan Langkah penting untuk meningkatkan akurasi model deteksi emosi.

Hasil penelitian ini diharapkan dapat memberikan pengalaman yang lebih mendalam mengenai bagaimana emosi diekspresikan dalam komunikasi digital, serta menunjukan efektivitas algoritma Navie Bayes dalam menganalisis emosi tersebut secara otomatis kontribusi penelitian ini diharapkan dapat memperdalam bidang analisis emosi, khususnya dalam konteks media sosial dan menjadi dasar untuk pengembangan aplikasi yang canggih di masa mendatang.

Kata kunci: Deteksi Emosi, Big Data, Media Sosial, Navie Bayes.

ABSTRACT

The rapid development of information and communication technology has made social media a primary platform for individuals to share thoughts, feelings, and experiences. Emotions expressed through Twitter texts constitute a rich source of data for analysis. Detecting these emotions is crucial for fields such as marketing, psychology, and social analytics. With the ever-increasing volume, comprehensive approaches such as machine learning and natural language processing (NLP) are required.

This research aims to analyze and detect user emotions through text data retrieved from GitHub using a Big Data approach. The Navie Bayes method was chosen as the primary classification algorithm due to its simplicity and ability to produce accurate results in text classification. Preprocessing steps include stopword removal, stemming, and word weighting using Term Frequency-Inverse Document Frequency (TF-IDF), which are crucial steps to improve the accuracy of emotion detection models.

The results of this study are expected to provide a deeper understanding of how emotions are expressed in digital communication and demonstrate the effectiveness of the Navie Bayes algorithm in automatically analyzing these emotions. This research is expected to deepen the field of emotion analysis, particularly in the context of social media, and serve as a foundation for the development of sophisticated applications in the future.

Keyword: Emotion Detection, Big Data, Social Media, Naive Bayes.