

**ANALISIS KOMUNIKASI NEGATIF PADA SOSIAL MEDIA X
BERDASARKAN HATE SPEECH DAN ABUSIVE LANGUAGE
MENGUNAKAN ALGORITMA NAIVE BAYES**

SKRIPSI

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



disusun oleh

DHANI JAMIATUL HIKMAH

22.83.0805

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS KOMUNIKASI NEGATIF PADA SOSIAL MEDIA X
BERDASARKAN HATE SPEECH DAN ABUSIVE LANGUAGE
MENGUNAKAN ALGORITMA NAIVE BAYES**

SKRIPSI

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



disusun oleh

DHANI JAMIATUL HIKMAH

22.83.0805

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS KOMUNIKASI NEGATIF PADA SOSIAL MEDIA X
BERDASARKAN HATE SPEECH DAN ABUSIVE LANGUAGE
MENGUNAKAN ALGORITMA NAIVE BAYES**

yang disusun dan diajukan oleh

Dhani Jamiatul Hikmah

22.83.0805

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25-05-2026

Dosen Pembimbing,


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

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS KOMUNIKASI NEGATIF PADA SOSIAL MEDIA X
BERDASARKAN HATE SPEECH DAN ABUSIVE LANGUAGE
MENGGUNAKAN ALGORITMA NAIVE BAYES

yang disusun dan diajukan oleh

Dhani Jamiatul Hikmah

22.83.0805

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25-05-2026

Susunan Dewan Penguji

Nama Penguji

Melwin Syafrizal, S.Kom., M.Eng., Ph.D.
NIK. 190302105

Dr. Dony Ariyus, S.S., M.Kom.
NIK. 190302128

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

Tanda Tangan

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25-05-2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dhani Jamiatul Hikmah
NIM : 22.83.0805

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Komunikasi Negatif Pada Sosial Media X Berdasarkan Hate Speech Dan Abusive Language Menggunakan Algoritma Naive Bayes

Dosen Pembimbing : Wahid Miftahul Ashari, S.Kom., M.T.

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, 25-05-2026

Yang Menyatakan,

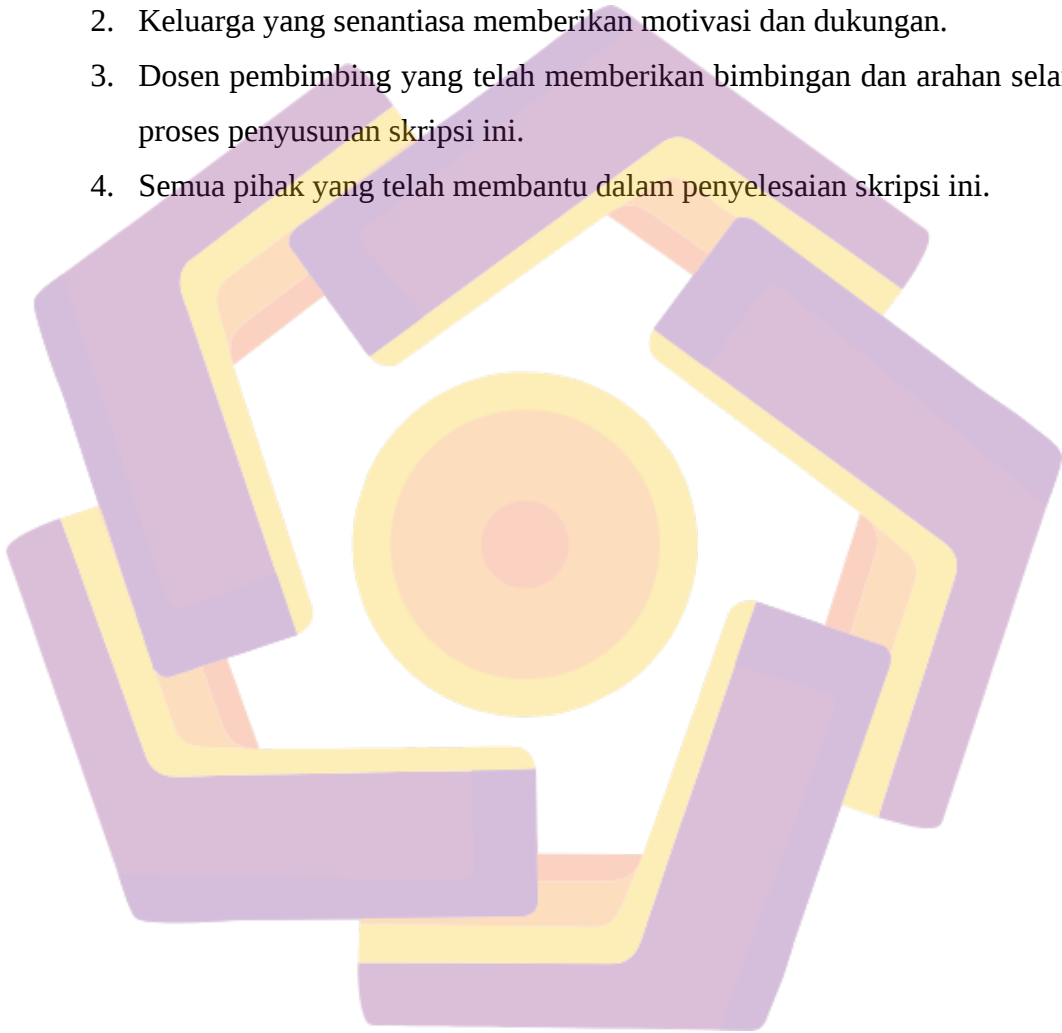


Dhani Jamiatul Hikmah

HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan kepada:

1. Kedua orang tua saya yang selalu memberikan doa, dukungan, dan semangat dalam setiap langkah.
2. Keluarga yang senantiasa memberikan motivasi dan dukungan.
3. Dosen pembimbing yang telah memberikan bimbingan dan arahan selama proses penyusunan skripsi ini.
4. Semua pihak yang telah membantu dalam penyelesaian skripsi ini.



KATA PENGANTAR

Puji syukur dipanjatkan kehadirat Allah SWT berkat rahmat, hidayah dan karunia-Nya sehingga penulis dapat menyelesaikan penulisan skripsi dengan judul **“Analisis Komunikasi Negatif Pada Sosial Media X Berdasarkan Hate Speech Dan Abusive Language Menggunakan Algoritma Naive Bayes”**. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Penulis menyadari bahwa dalam penyusunan skripsi ini terwujud berkat bantuan, arahan, bimbingan dan doa dari berbagai pihak. Oleh karena itu penulis menyampaikan terimakasih setulus-tulusnya kepada:

1. Bapak Wahid Miftahul Ashari, S.Kom., M.T. selaku dosen pembimbing yang membuat penulis dapat menyelesaikan skripsi ini dengan baik. Beliau adalah anutan penulis. Penulis mendapatkan semangat dan motivasi setiap kali berinteraksi dengan beliau dan tidak mau jika anak bimbingannya lama selesai atau wisuda. Hal ini tentunya menjadi sebuah keberuntungan dapat dibimbing oleh beliau. Penulis berharap semoga beliau sehat selalu, panjang umur, dan tetap di jalan yang benar.
2. Bapak Prof. Dr. M.Suyanto, M.M. selaku rektor Universitas Amikom Yogyakarta yang telah memberikan kesempatan kepada penulis untuk melaksanakan studi di Universitas Amikom Yogyakarta.
3. Ibu Prof. Dr. Kusri, S.Kom., M.Kom. selaku Dekan, Bapak Dr. Andi Sunyoto, M.Kom. selaku Wakil Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta yang telah memberikan pelayanan yang baik kepada semua mahasiswa untuk kelancaran studi.
4. Melwin Syafrizal, S.Kom., M.Eng., Ph.D. selaku penguji 1 yang sudah meluangkan waktunya untuk menguji penulis, menelaah, mengkritisi, dan memberikan masukan dalam skripsi ini.
5. Dr. Dony Ariyus, S.S., M.Kom. selaku dosen penguji 2 yang sudah meluangkan waktunya untuk menguji penulis, menelaah, mengkritisi, dan memberikan masukan dalam skripsi ini.

6. Dosen Fakultas Ilmu Komputer Universitas Amikom Yogyakarta yang tidak dapat penulis sebut satu persatu karena telah memberikan banyak ilmu dan masukan kepada penulis.
7. Teristimewa buat Kedua Orangtua tercinta tersayang, Bapak Jaharudin, S.E. dan Ibu Entin Suryati, S.Pd. yang telah memberikan kasih sayang, nasehat, motivasi, semangat, memenuhi kebutuhan penulis, dukungan serta Doa yang tiada henti kepada penulis selama melalui proses perkuliahan, penulisan skripsi ini sampai saat ini. Terima kasih pula atas cintanya yang tidak pernah berkurang dan dukungan yang tidak pernah berhenti menguatkan penulis sepanjang perjalanan pendidikan ini. Semoga Allah SWT membalas kebaikan dan ketulusan dengan keberkahan dan kesehatan sepanjang usia Aamiin
8. Kakak sekaligus paman Penulis, Dharman yang selalu memberikan dukungan, bantuan, dan semangat selama proses penyusunan skripsi ini. Tak lupa kepada adik penulis Aslan dan Azam. Terima kasih atas semangat dan kebersamaan selalu menjadi penghibur dengan tingkah lakunya. Dukungan kalian menjadi bagian penting dari terselesainya studi ini.
9. Nenek dan Bibi Penulis yang telah merawat dan membesarkan penulis dari kecil hingga dewasa, memberikan kasih sayang, perhatian dan doa, sehingga penulis dapat terus berjuang dalam meraih mimpi dan cita cita, Terimakasih.
10. Terimakasih Kepada seseorang yang penulis cintai Eka kurniati, terimakasih telah menemani penulis menyelesaikan skripsi ini. Telah berkontribusi banyak dalam penulisan skripsi ini, meluangkan baik tenaga, waktu dan pikiran kepada penulis dan senantiasa sabar menghadapi penulis. Terimakasih telah menjadi bagian dari perjalanan penulis hingga sekarang ini.
11. Kepada saudara-saudara sepupu penulis, terima kasih karena selalu memberikan dukungan yang tiada hentinya, yang selalu mengingatkan untuk terus semangat dalam menyelesaikan tugas yang sedang dijalani.

12. Kepada teman-teman penulis yang di Desa Jambu tak kalah penting kehadirannya, Terimakasih karena telah hadir dan menjadi bagian dari perjalanan hidup penulis, yang telah berkontribusi banyak dari awal hingga akhir penulisan, memberikan semangat, mendukung dan menghibur.
13. Mahasiswa Program Studi Teknik Komputer, teristimewa teman-teman angkatan 2022, terimakasih buat kebersamaan dan persahabatan kalian dalam suka dan duka selama di lapangan maupun kampus.
14. Semua pihak yang telah membantu hingga terselesaikannya pembuatan skripsi maupun dalam penyusunan yang tidak dapat disebutkan satu persatu.

Semoga skripsi ini dapat memberikan manfaat bagi pembaca pada umumnya Dan penulis khususnya. Kritik dan saran yang membangun sangat diharapkan dari pembaca untuk penyempurnaan skripsi ini.

Yogyakarta, 6 Mei 2026

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.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xvi
INTISARI	xviii
ABSTRACT.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	7
BAB II TINJAUAN PUSTAKA	8
2.1 Studi Literatur	8
2.2 Dasar Teori	15
BAB III METODE PENELITIAN	22
3.1 Objek Penelitian	22

3.2	Alur Penelitian.....	22
3.3	Alat dan Bahan	25
3.3.1	Data Penelitian	25
3.3.2	Alat/instrument	26
3.4	Metode Penelitian.....	27
3.5	Teknik Analisis Data	27
3.6	Variabel Penelitian	29
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Deskripsi Data Penelitian	30
4.2	Tahapan Preprocessing.....	31
4.2.1	Case Folding	31
4.2.2	Cleaning Text.....	32
4.2.3	Tokenizing	32
4.2.4	Stopword Removal.....	32
4.2.5	Stemming	33
4.3	Transformasi Data Menggunakan TF-IDF.....	33
4.4	Pembagian Data Training dan Testing.....	35
4.5	Penerapan Algoritma Naive Bayes	36
4.6	Hasil Evaluasi Dan Pengujian Model.....	38
4.7	Pembahasan.....	40
BAB V PENUTUP		42
5.1	Kesimpulan.....	42
5.2	Saran.....	43
DAFTAR PUSTAKA		44
LAMPIRAN.....		48

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu	14
Tabel 4.1. Distribusi Data Penelitian	30
Tabel 4.2. Hasil Case Folding	31
Tabel 4.3. Hasil Cleaning Text	32
Tabel 4.4. Hasil Tokenizing	32
Tabel 4.5 Hasil Stopword Removal	33
Tabel 4.6 Hasil Stemming	33
Tabel 4.7 Distribusi Data Training dan Testing	36
Tabel 4.8 Hasil Evaluasi Model	38

DAFTAR GAMBAR

Gambar 3.1. Alur Penelitian	26
Gambar 4.1 Proses Transformasi Data menggunakan TF-IDF	34
Gambar 4.2 Proses Pembagian Data	35
Gambar 4.3 Proses <i>Training</i> model <i>Naive Bayes</i>	37
Gambar 4.4 Proses Prediksi Data	37
Gambar 4.5 Proses Evaluasi Model	38
Gambar 4.6 Confusion Matrix	39



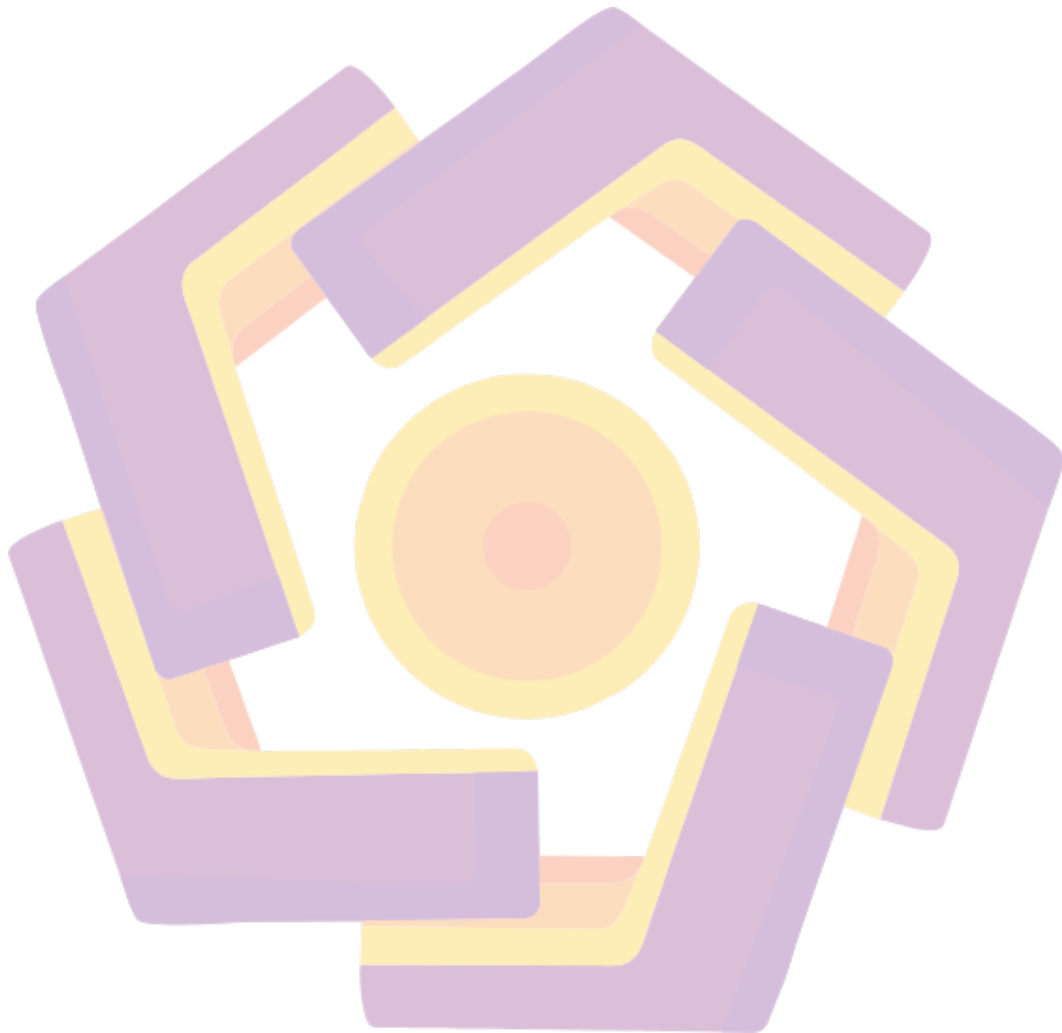
DAFTAR LAMPIRAN

Lampiran 1. Dataset Penelitian	50
Lampiran 2. Contoh Dataset	51
Lampiran 3. Hasil Preprocessing Data	53
Lampiran 4. Source Code Preprocessing	54
Lampiran 5. Source Code Tokenizing	55
Lampiran 6. Source Code Stopword Removal	56
Lampiran 7. Source Code Stemming	57
Lampiran 8. Transformasi TF-IDF	58
Lampiran 9. Proses Klasifikasi Naive Bayes	59
Lampiran 10. Prediksi dan Evaluasi Model	60
Lampiran 11. Confusion Matrix	61
Lampiran 12. Hasil Evaluasi Model	62
Lampiran 13. Hasil Distribusi Data	63

DAFTAR LAMBANG DAN SINGKATAN

Singkatan	Artinya
X	Variabel input berupa fitur teks hasil transformasi TF-IDF.
Y	Variabel output berupa label klasifikasi.
TF (Term Frequency)	Frekuensi kemunculan suatu kata dalam dokumen.
IDF (Inverse Document Frequency)	Ukuran pentingnya suatu kata dalam keseluruhan dokumen.
TF-IDF	Metode pembobotan kata
NB (Naive Bayes)	Algoritma klasifikasi
TP (True Positive)	Jumlah data positif yang diprediksi dengan benar.
TN (True Negative)	Jumlah data negatif yang diprediksi dengan benar.
FP (False Positive)	Jumlah data negatif yang salah diprediksi sebagai positif.
FN (False Negative)	Jumlah data positif yang salah diprediksi sebagai negatif.
Accuracy	Tingkat ketepatan model dalam melakukan klasifikasi.
Precision	Tingkat ketepatan prediksi positif oleh model.
Recall	Kemampuan model dalam menemukan data positif.
F1-Score	Nilai rata-rata harmonis antara precision dan recall.

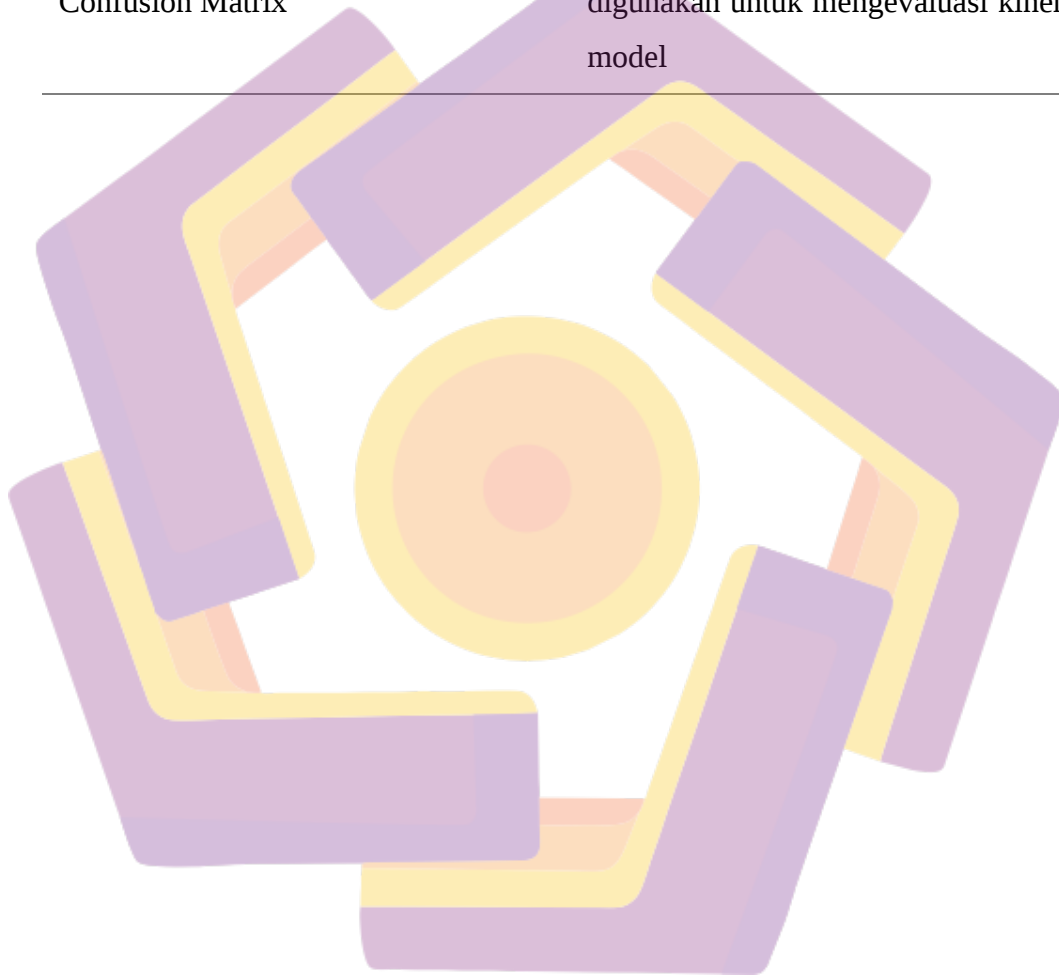
CM (Confusion Matrix)	Tabel evaluasi untuk membandingkan hasil prediksi dan data aktual.
ML (Machine Learning)	Metode pembelajaran mesin untuk membangun model prediksi dari data.



DAFTAR ISTILAH

Istilah	Artinya
Algoritma Naive Bayes	Metode klasifikasi berbasis probabilitas
Text Mining	Proses pengolahan dan analisis data teks
Preprocessing	pembersihan dan normalisasi data sebelum dianalisis
Cleaning Text	Proses membersihkan teks dari karakter yang tidak diperlukan
Tokenizing	Proses memecah teks menjadi unit-unit kecil
Stopword Removal	Proses menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis
Stemming	Proses mengubah kata menjadi bentuk dasar
TF-IDF	Metode pembobotan kata
Fitur	Representasi numerik dari data teks
Dataset	Kumpulan data yang digunakan dalam penelitian
Training Data	Data yang digunakan untuk melatih model
Testing Data	Data yang digunakan untuk menguji kinerja model
Label	Kategori atau kelas dari data, seperti negatif (1) dan non-negatif (0).
Klasifikasi	Proses mengelompokkan data ke dalam kategori tertentu.

Akurasi	tingkat ketepatan model
Precision	ketepatan prediksi positif
Recall	kemampuan model dalam menemukan semua data positif.
F1-Score	Nilai gabungan dari precision dan recall
Confusion Matrix	digunakan untuk mengevaluasi kinerja model



INTISARI

Perkembangan media sosial telah mendorong perubahan pola komunikasi masyarakat, di mana pengguna dapat menyampaikan pendapat secara bebas dan terbuka. Namun, kondisi ini juga memicu munculnya komunikasi negatif seperti *hate speech* dan *abusive language* yang dapat menimbulkan konflik sosial, menurunkan kualitas diskusi publik, serta berdampak negatif terhadap lingkungan digital. Tingginya volume data pada media sosial X (sebelumnya Twitter) menyebabkan proses identifikasi komunikasi negatif secara manual menjadi sulit dilakukan. Oleh karena itu, diperlukan suatu pendekatan otomatis yang mampu mengidentifikasi dan mengklasifikasikan komunikasi negatif secara efektif dan efisien.

Penelitian ini menggunakan metode kuantitatif dengan pendekatan text mining untuk menganalisis komunikasi negatif pada media sosial X. Data yang digunakan berupa dataset Indonesian *Abusive and Hate Speech Twitter Text* yang telah dianotasi. Tahapan penelitian meliputi preprocessing data yang terdiri dari *case folding*, *tokenizing*, *stopword removal*, dan *stemming*, kemudian dilanjutkan dengan proses klasifikasi menggunakan algoritma *Naive Bayes*. Data dibagi menjadi data training dan data testing untuk membangun dan menguji model. Selanjutnya, dilakukan evaluasi model menggunakan *confusion matrix* dengan metrik akurasi, *precision*, *recall*, dan *F1-score*.

Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* mampu mengklasifikasikan komunikasi negatif dengan baik berdasarkan kategori *hate speech* dan *abusive language*. Penelitian ini memberikan kontribusi dalam pengembangan analisis komunikasi digital berbasis bahasa Indonesia serta dapat dimanfaatkan oleh pengembang sistem, peneliti, dan pihak terkait dalam mendeteksi serta mengelola konten negatif di media sosial. Untuk penelitian selanjutnya, disarankan menggunakan metode yang lebih kompleks atau membandingkan beberapa algoritma guna meningkatkan akurasi model.

Kata kunci: komunikasi negatif, *hate speech*, *abusive language*, *naive bayes*, *text mining*.

ABSTRACT

The rapid development of social media has significantly changed communication patterns, allowing users to freely express their opinions in an open digital environment. However, this condition also triggers the emergence of negative communication such as hate speech and abusive language, which can lead to social conflict, reduce the quality of public discourse, and negatively impact the digital environment. The high volume of data on social media X (formerly Twitter) makes manual identification of negative communication difficult. Therefore, an automated approach is needed to effectively and efficiently identify and classify negative communication.

This study employs a quantitative method with a text mining approach to analyze negative communication on social media X. The dataset used is the Indonesian Abusive and Hate Speech Twitter Text, which has been annotated. The research process includes data preprocessing consisting of case folding, tokenizing, stopword removal, and stemming, followed by classification using the Naive Bayes algorithm. The dataset is divided into training and testing data to build and evaluate the model. Model evaluation is conducted using a confusion matrix with performance metrics including accuracy, precision, recall, and F1-score.

The results show that the Naive Bayes algorithm is capable of effectively classifying negative communication based on hate speech and abusive language categories. This study contributes to the development of digital communication analysis in the Indonesian language and can be utilized by system developers, researchers, and related stakeholders in detecting and managing negative content on social media. Future research is recommended to apply more advanced methods or compare multiple algorithms to improve model performance.

Keyword: negative communication, hate speech, abusive language, naive bayes, text mining