

**ANALISIS SENTIMEN PENGGUNA SOSIAL MEDIA TWITTER  
TERHADAP PEROKOK DI INDONESIA**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**DEWI SETIYAWATI**

**19.11.2658**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER  
TERHADAP PEROKOK DI INDONESIA**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**DEWI SETIYAWATI**

**19.11.2658**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2023**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**ANALISIS SENTIMEN PENGGUNA SOSIAL MEDIA TWITTER  
TERHADAP PEROKOK DI INDONESIA**

yang disusun dan diajukan oleh

**Dewi Setiyawati**

**19.11.2658**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 24 Maret 2023

**Dosen Pembimbing,**



**Nuri Cahyono, M.Kom**  
**NIK. 190302278**

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**ANALISIS SENTIMEN PENGGUNA SOSIAL MEDIA TWITTER**  
**TERHADAP PEROKOK DI INDONESIA**

yang disusun dan diajukan oleh

**Dewi Setiyawati**

**19.11.2658**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 24 Maret 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

Arif Dwi Laksito, M.Kom  
NIK. 190302150

Ainul Yaqin, M.Kom  
NIK. 190302255

Nuri Cahyono, M.Kom  
NIK. 190302278



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 24 Maret 2023

**DEKAN FAKULTAS ILMU KOMPUTER**



Hanif Al Fatta, S.Kom., M.Kom.  
NIK. 190302096

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dewi Setiyawati  
NIM : 19.11.2658

Menyatakan bahwa Skripsi dengan judul berikut:

### **Analisis Sentimen Pengguna Sosial Media Twitter Terhadap Perokok Di Indonesia**

Dosen Pembimbing : Nuri Cahyono, M.Kom

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, 24 Maret 2023

Yang Menyatakan,



Dewi Setiyawati

## HALAMAN PERSEMBAHAN

Alhamdulillah segala puji bagi Allah SWT yang telah memberikan rahmat, hidayah, kemudahan, dan kelancaran kepada saya dalam pengerjaan skripsi ini, sehingga dapat terselesaikan dengan baik. Saya berterima kasih kepada semua pihak yang telah membantu dan mendukung saya, sehingga mampu menyelesaikan studi untuk meraih gelar sarjana. Skripsi ini saya persembahkan kepada:

1. Kedua orang tua saya, Bapak Ngadnan dan Ibu Melati, yang selalu mendukung, mendoakan, memberi kasih sayang yang luar biasa, serta atas seluruh kesabaran dalam menghadapi saya.
2. Adik saya tersayang, Diky Darmawan, terima kasih telah membuat saya selalu semangat dalam mencapai impian.
3. Stray Kids sebagai penyemangat dikala saya merasa lelah serta terima kasih telah memberikan motivasi tinggi dalam mencapai banyak impian.
4. Rifka Annisa, sebagai sahabat seperjuangan saya, terima kasih selalu mengajak diri saya dalam kebaikan.
5. Angga Reni Dwi Astuti dan Diah Ayu Nur Wulan yang telah selalu ada saat saya kesulitan dalam mengerjakan skripsi.
6. Teman-teman yang selalu membantu saya selama perkuliahan, Erika Yosita Putri, Herjuno Dwi Nugroho, Raynaldo Wira, Habib Hisbangun, M. Agung Wicaksonosidi, Budi Prasetyo, Pipin Ristriani, Budi Setiawan, dan Aini Nur Khasanah.
7. Teman-teman ketika susah dan senang saya, Avis Priyati, Galih Wahyu Sriagung, Fajar Ramadhanie, Nurul Hidayati, Nur Halisa Putri Eryanto, Qaeda Putri Rana Pahlevi, dan Prima Wijanarko.
8. Teman-teman kelas 19 Informatika 01 yang telah berbagi ilmu dan menemani semasa perkuliahan.
9. Teman-teman AMCC yang selalu memberikan saya masukan dan saran.
10. Seluruh pihak yang tidak bisa saya sebutkan satu per satu. Terima kasih atas segala dukungan, motivasi, dan doa sehingga saya bisa berjuang sampai titik ini.

## KATA PENGANTAR

Puji syukur kita panjatkan kepada Allah SWT atas limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini tepat pada waktunya. Skripsi dengan judul “Analisis Sentimen Pengguna Sosial Media Twitter Terhadap Perokok di Indonesia” yang merupakan syarat untuk menyelesaikan program Strata-I Informatika di Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Penulis menyadari bahwa penyusunan skripsi ini tidak akan selesai tanpa adanya bimbingan, dukungan, bantuan, serta nasehat dari berbagai pihak. Dalam kesempatan kali ini penulis menyampaikan terima kasih yang sebesar-besarnya terkhusus kepada:

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya Duhita, M.Kom., selaku Kepala Program Studi Informatika Universitas Amikom Yogyakarta.
4. Bapak Nuri Cahyono, M.Kom., selaku Dosen Pembimbing penulis, yang telah meluangkan waktu ditengah kesibukan beliau, memberikan kritik, saran, dan arahan selama penyusunan skripsi.
5. Bapak dan Ibu Dosen Penguji. Terima kasih atas saran yang diberikan sehingga penelitian ini menjadi lebih baik.
6. Bapak dan Ibu Dosen Universitas Amikom Yogyakarta yang telah memberikan ilmu pengetahuan selama menempuh Pendidikan di Universitas Amikom Yogyakarta.
7. Semua pihak yang telah membantu, memberikan motivasi, dan mendukung secara langsung maupun tidak langsung.

Semoga Tuhan memberikan balasan yang lebih kepada semua pihak yang telah membantu penulis sehingga dapat menyelesaikan skripsi ini. Penulis

menyadari bahwa kesempurnaan hanya milik Allah SWT, sehingga dalam penyusunan skripsi ini masih banyak kekurangan dan kesalahan, oleh karena itu segala kritik dan saran yang akan menyempurnakan dalam skripsi ini penulis terima dengan senang hati dan rasa terima kasih. Semoga skripsi ini dapat memberikan manfaat bagi penulis dan para pembaca.

Yogyakarta, 3 Maret 2023

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 .....	viii
DAFTAR TABEL .....	xi
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN .....	xiv
DAFTAR LAMBANG DAN SINGKATAN .....	xv
DAFTAR ISTILAH .....	xvi
INTISARI .....	xvii
ABSTRACT .....	xviii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah .....	4
1.4 Tujuan Penelitian .....	4
1.5 Manfaat Penelitian .....	5
1.6 Sistematika Penulisan .....	5
BAB II TINJAUAN PUSTAKA .....	7
2.1 Studi Literatur .....	7
2.2 Dasar Teori .....	17
2.2.1 Text mining .....	17
2.2.2 Machine Learning .....	17
2.2.3 Analisis sentimen .....	18
2.2.4 Twitter .....	18
2.2.5 TF-IDF .....	19
2.2.6 Naive bayes .....	19
2.2.7 Support Vector Machine (SVM) .....	19
2.2.8 Decision Tree .....	20
2.2.9 Google Colab .....	20
2.2.10 Python .....	20
BAB III METODE PENELITIAN .....	21

3.1	Alur Penelitian .....	21
3.1.1	Pengumpulan Data .....	22
3.1.2	Pelabelan Data .....	23
3.1.3	Pre-processing Data .....	26
3.1.3.1	Cleansing .....	27
3.1.3.2	Case Folding .....	28
3.1.3.3	Tokenizing .....	29
3.1.3.4	Stopword .....	31
3.1.3.5	Stemming .....	32
3.1.4	Pembobotan .....	34
3.1.5	Penerapan Metode .....	34
3.2	Alat dan Bahan .....	35
3.2.1	Data Penelitian .....	35
3.2.2	Alat Penelitian .....	36
3.2.3	Pandas .....	36
3.2.4	Numpy .....	36
3.2.5	Matplotlib .....	37
3.2.6	Seaborn .....	37
3.2.7	Sklearn .....	37
3.2.8	NLTK .....	37
3.2.9	Wordcloud .....	37
3.2.10	Transformers .....	38
3.2.11	Translator .....	38
3.2.12	Snsrape .....	38
<b>BAB IV</b>	<b>HASIL DAN PEMBAHASAN .....</b>	<b>39</b>
4.1	Visualisasi Data .....	39
4.2	Penerapan Metode .....	39
4.2.1	Metode Naïve Bayes .....	40
4.2.1.1	Menggunakan 2.772 data terlabel manual .....	40
4.2.1.2	Menggunakan 2.772 data terlabel dengan library transformers .....	42
4.2.1.3	Menggunakan 6.300 data terlabel dengan library transformers .....	43
4.2.1.4	Hasil Pengujian Algoritma Naïve Bayes .....	45
4.2.2	Metode Support Vector Machine .....	46
4.2.2.1	Menggunakan 2.772 data terlabel manual .....	46
4.2.2.2	Menggunakan 2.772 data terlabel dengan library transformers .....	47
4.2.2.3	Menggunakan 6.300 data terlabel dengan library transformers .....	49
4.2.2.4	Hasil Pengujian Algoritma Support Vector Machine .....	50
4.2.3	Metode Decision Tree .....	51
4.2.3.1	Menggunakan 2.772 data terlabel manual .....	51
4.2.3.2	Menggunakan 2.772 data terlabel dengan library transformers .....	53
4.2.3.3	Menggunakan 6.300 data terlabel dengan library transformers .....	54
4.2.3.4	Hasil Pengujian Algoritma Decision Tree .....	56
4.3	Validasi Data .....	57
4.3.1	Data Testing 20% Metode Naïve Bayes .....	57
4.3.2	Data Testing 20% Metode Support Vector Machine .....	58
4.3.3	Data Testing 20% Metode Decision Tree .....	60

BAB V PENUTUP .....	62
5.1 Kesimpulan .....	62
5.2 Saran .....	62
REFERENSI .....	64
LAMPIRAN .....	71



## DAFTAR TABEL

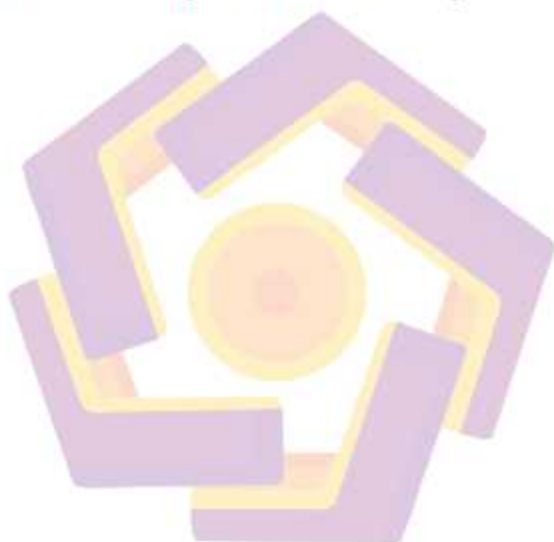
Tabel 2.1 Keaslian penelitian	10
Tabel 3.1 Jumlah data tidak dipakai dari 2992 data mentah	23
Tabel 3.2 Jumlah data tidak dipakai dari 7512 data mentah	23
Tabel 3.3 Hasil pelabelan manual	24
Tabel 3.4 Hasil pelabelan otomatis	25
Tabel 3.5 Cleaning data	27
Tabel 3.6 Case folding	28
Tabel 3.7 Tokenizing	30
Tabel 3.8 Stopword	31
Tabel 3.9 Stemming	32
Tabel 3.10 Confusion matrix	34
Tabel 3.11 Alat penelitian	36
Tabel 4.1 Pemecahan data	40
Tabel 4.2 Confusion matrix naïve bayes pengujian I dengan testing 20%	40
Tabel 4.3 Confusion matrix naïve bayes pengujian I dengan testing 40%	41
Tabel 4.4 Confusion matrix naïve bayes pengujian I dengan testing 50%	41
Tabel 4.5 Confusion matrix naïve bayes pengujian I dengan testing 60%	41
Tabel 4.6 Confusion matrix naïve bayes pengujian I dengan testing 80%	41
Tabel 4.7 Confusion matrix naïve bayes pengujian II dengan testing 20%	42
Tabel 4.8 Confusion matrix naïve bayes pengujian II dengan testing 40%	42
Tabel 4.9 Confusion matrix naïve bayes pengujian II dengan testing 50%	42
Tabel 4.10 Confusion matrix naïve bayes pengujian II dengan testing 60%	43
Tabel 4.11 Confusion matrix naïve bayes pengujian II dengan testing 80%	43
Tabel 4.12 Confusion matrix naïve bayes pengujian III dengan testing 20%	43
Tabel 4.13 Confusion matrix naïve bayes pengujian III dengan testing 40%	44
Tabel 4.14 Confusion matrix naïve bayes pengujian III dengan testing 50%	44
Tabel 4.15 Confusion matrix naïve bayes pengujian III dengan testing 60%	44
Tabel 4.16 Confusion matrix naïve bayes pengujian III dengan testing 80%	45
Tabel 4.17 Hasil pengujian algoritma naïve bayes	45
Tabel 4.18 Confusion matrix SVM pengujian I dengan testing 20%	46
Tabel 4.19 Confusion matrix SVM pengujian I dengan testing 40%	46
Tabel 4.20 Confusion matrix SVM pengujian I dengan testing 50%	46
Tabel 4.21 Confusion matrix SVM pengujian I dengan testing 60%	47
Tabel 4.22 Confusion matrix SVM pengujian I dengan testing 80%	47
Tabel 4.23 Confusion matrix SVM pengujian II dengan testing 20%	47
Tabel 4.24 Confusion matrix SVM pengujian II dengan testing 40%	48
Tabel 4.25 Confusion matrix SVM pengujian II dengan testing 50%	48
Tabel 4.26 Confusion matrix SVM pengujian II dengan testing 60%	48
Tabel 4.27 Confusion matrix SVM pengujian II dengan testing 80%	49
Tabel 4.28 Confusion matrix SVM pengujian III dengan testing 20%	49
Tabel 4.29 Confusion matrix SVM pengujian III dengan testing 40%	49
Tabel 4.30 Confusion matrix SVM pengujian III dengan testing 50%	50
Tabel 4.31 Confusion matrix SVM pengujian III dengan testing 60%	50

Tabel 4.32 Confusion matrix SVM pengujian III dengan testing 80%	50
Tabel 4.33 Hasil pengujian algoritma support vector machine	50
Tabel 4.34 Confusion matrix decision tree pengujian I dengan testing 20%	51
Tabel 4.35 Confusion matrix decision tree pengujian I dengan testing 40%	52
Tabel 4.36 Confusion matrix decision tree pengujian I dengan testing 50%	52
Tabel 4.37 Confusion matrix decision tree pengujian I dengan testing 60%	52
Tabel 4.38 Confusion matrix decision tree pengujian I dengan testing 80%	52
Tabel 4.39 Confusion matrix decision tree pengujian II dengan testing 20%	53
Tabel 4.40 Confusion matrix decision tree pengujian II dengan testing 40%	53
Tabel 4.41 Confusion matrix decision tree pengujian II dengan testing 50%	53
Tabel 4.42 Confusion matrix decision tree pengujian II dengan testing 60%	54
Tabel 4.43 Confusion matrix decision tree pengujian II dengan testing 80%	54
Tabel 4.44 Confusion matrix decision tree pengujian III dengan testing 20%	54
Tabel 4.45 Confusion matrix decision tree pengujian III dengan testing 40%	55
Tabel 4.46 Confusion matrix decision tree pengujian III dengan testing 50%	55
Tabel 4.47 Confusion matrix decision tree pengujian III dengan testing 60%	55
Tabel 4.48 Confusion matrix decision tree pengujian III dengan testing 80%	56
Tabel 4.49 Hasil pengujian algoritma decision tree	56



## DAFTAR GAMBAR

Gambar 3.1 Diagram alir penelitian	21
Gambar 4.1 Tahapan pre-processing data	27
Gambar 4.2 Visualisasi data sentimen negatif dan positif	39
Gambar 4.3 Jumlah data testing 20% dari metode naïve bayes	57
Gambar 4.4 Nilai data testing 20% dari metode naïve bayes	57
Gambar 4.5 Jumlah data testing 20% dari metode naïve bayes	58
Gambar 4.6 Nilai data testing 20% dari metode naïve bayes	59
Gambar 4.7 Jumlah data testing 20% dari metode naïve bayes	60
Gambar 4.8 Nilai data testing 20% dari metode naïve bayes	60




## DAFTAR LAMPIRAN

Lampiran 1. Data mentah 2.772 tweet	66
Lampiran 2. Data mentah 6.300 tweet	72



## DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
KNN	K-Nearest Neighbor
NBC	Naïve Bayes Clasifier
SRM	Structural Risk Minimization
NLTK	Natural Language ToolKit
NLP	Natural Language Processing
API	Appication Programming Interface
HTML	Hyper Text Markup Language
DPR	Dewan Perwakilan Rakyat
Kemendikbud	Kementerian Pendidikan dan Kebudayaan





## DAFTAR ISTILAH

Database	Sekumpulan informasi yang tersimpan
Query	Informasi yang diterima atau diambil dari database
Netizen	orang yang menggunakan internet
Matrix	susunan bilangan dalam baris dan kolom
Tweet	pesan text yang dikirim di aplikasi Twitter
Gitbash	aplikasi yang menyediakan tiruan layer untuk command line
Library	kumpulan kode yang telah dikompilasi
LaTex	Bahasa system penyimpanan dokumen dalam membuat formula matematika



## INTISARI

Salah satu alat yang digunakan pengguna web untuk mengakses, berbagi, dan mendiskusikan subjek yang sedang diminati adalah media sosial. Salah satu situs jejaring sosial, Twitter, sering digunakan secara real time untuk berkomunikasi ini. Karena dampak negatifnya yang signifikan baik bagi kesehatan maupun ekonomi, merokok masih menjadi topik perdebatan dan perdebatan rutin di Indonesia. Penelitian ini dilakukan untuk menilai sentimen terhadap perokok dan membedakan antara emosi positif dan negatif. Data yang digunakan dalam penelitian ini diperoleh dengan melakukan crawling pada jaringan media sosial Twitter. Tiga teknik, yaitu Naïve Bayes (NB), Support Vector Machine (SVM), dan Logistic Regression digunakan dalam penelitian ini. Dalam penelitian ini, pengguna Twitter lebih banyak tidak setuju dengan keberadaan perokok di Indonesia. Metode Support Vector Machine digunakan dalam penelitian ini, memberikan nilai akurasi tertinggi 90% dengan menggunakan 20% data latih dan 80% data uji.

**Kata kunci:** Analisis Sentimen, Perokok, Naïve Bayes, Support Vector Machine, Decision Tree.



## ABSTRACT

One of the tools web users use to access, share, and discuss subjects of interest is social media. One social networking site, Twitter, is often used in real time to communicate this. Due to its significant negative impacts on both health and the economy, smoking is still a topic of regular debate and debate in Indonesia. This research was conducted to assess sentiment towards smokers and differentiate between positive and negative emotions. The data used in this study were obtained by crawling the Twitter social media network. Three techniques, namely Naïve Bayes (NB), Support Vector Machine (SVM), and Logistic Regression are used in this study. In this study, Twitter users disagreed with the existence of smokers in Indonesia. The Support Vector Machine method was used in this study, giving the highest accuracy value 90% using 20% training data and 80% test data.

**Keyword:** Sentiment Analysis, Smokers, Naïve Bayes, Support Vector Machine, Decision Tree.

