

**ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER
TERHADAP ISU GANGGUAN DEPRESI DENGAN
MENGUNAKAN METODE NAÏVE BAYES**

SKRIPSI

**Diajukan untuk memenuhi persyaratan
dalam memperoleh gelar Sarjana Komputer
Program Studi Sistem Informasi**

Disusun Oleh:



NUR LICKHA LAVENIA

19082010038

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2023**

SKRIPSI

**ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP
ISU GANGGUAN DEPRESI DENGAN MENGGUNAKAN METODE
NAÏVE BAYES**

Disusun Oleh:

NUR LICKHA LAVENIA

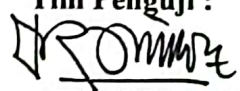

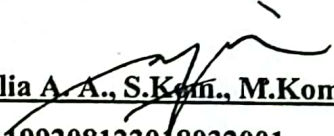
19082010038

**Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi
Program Studi Sistem Informasi Fakultas Ilmu Komputer
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Pada Tanggal 13 Juli 2023**

Pembimbing :

1. 
Dhian Satria Y. K., S.Kom., M.Kom
NPT. 201198 60 522249
2. 
Reisa Permatasari, S.T., M.Kom.,
NIP. 19920514 202203 2007

Tim Penguji :

1. 
M. Irwan A., S.T., M.Sc.
NIP. 19760718 2021211 003
2. 
Doddy R., S.Kom., M.Kom.
NPT. 378050702181
3. 
Amalia A. A., S.Kom., M.Kom.
NIP. 199208122018032001

Mengetahui,

Dekan Fakultas Ilmu Komputer

Universitas Pembangunan Nasional "Veteran" Jawa Timur


Dr. Ir. Noyirina Hendrasarie, MT.

NIP. 19681126 199403 2 001

LEMBAR PENGESAHAN

**ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP
ISU GANGGUAN DEPRESI DENGAN MENGGUNAKAN METODE
NAÏVE BAYES**

Disusun Oleh:

NUR LICKHA LAVENIA

19082010038

**Telah disetujui mengikuti Ujian Negara Lisan Gelombang Juli
Periode 2023 pada Tanggal 13 Juli 2023**

Menyetujui,

Dosen Pembimbing 1



Dhian Satria Y. K., S.Kom., M.Kom

NPT. 201198 60 522249

Dosen Pembimbing 2



Reisa Permatasari, S.T., M.Kom.,

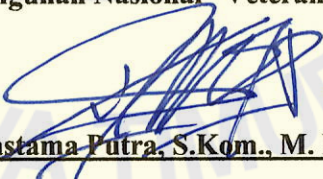
NIP. 19920514 202203 2007

Mengetahui,

Ketua Program Studi Sistem Informasi

Fakultas Ilmu Komputer

Universitas Pembangunan Nasional "Veteran" Jawa Timur



Agung Brastama Putra, S.Kom., M. Kom.

NIP. 19851124 2021211 003

UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA TIMUR
FAKULTAS ILMU KOMPUTER

KETERANGAN REVISI

Kami yang bertanda tangan di bawah ini menyatakan bahwa mahasiswa berikut:

Nama : Nur Lickha lavenia

NPM : 19082010038

Program Studi : Sistem Informasi

Telah mengerjakan revisi Ujian Negara Lisan Skripsi pada tanggal 20 Juli 2023 dengan judul:

ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP ISU GANGGUAN DEPRESI DENGAN MENGGUNAKAN METODE NAÏVE BAYES

Oleh karenanya mahasiswa tersebut diatas dinyatakan bebas revisi Ujian Negara Lisan Skripsi dan diijinkan untuk membukukan laporan SKRIPSI dengan judul tersebut.

Surabaya, 29 Juli 2023

Dosen penguji yang memeriksa revisi:

1. M. Irwan Afandi, S.T., M.Sc.

NIP. 19760718 2021211 003

2. Doddy Ridwandono, S.Kom., M.Kom.

NPT. 378050702181

3. Amalia A. A., S.Kom., M.Kom.

NIP. 199208122018032001

Mengetahui,

Pembimbing Utama

Pembimbing Pendamping

Dhian Satria Y. K., S.Kom., M.Kom

NPT. 2011986 0522 249

Reisa Permatasari, S.T., M.Kom.,

NIP. 19920514 2022032 007

LEMBAR PERNYATAAN ORISINILITAS

Berdasarkan Undang-Undang Nomor 19 Tahun 2002 tentang Hak Cipta dan Pemendiknas Nomor 17 Tahun 2010 tentang pencegahan dan penanggulangan plagiat di Perguruan Tinggi, maka saya yang bertanda tangan di bawah ini:

Nama : Nur Lickha Lavenia
NPM : 19082010038
Program Studi : Sistem Informasi
Tahun Akademik : 2019/2020

Menyatakan bahwa dalam penulisan dan pengerjaan Skripsi/Tugas Akhir yang saya ajukan, yang berjudul:

“ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP ISU GANGGUAN DEPRESI DENGAN MENGUNAKAN METODE NAÏVE BAYES”

Bukan merupakan plagiat dari Skripsi/Tugas Akhir/Penelitian orang lain dan juga bukan merupakan produk dan atau software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN “Veteran” Jawa Timur maupun di institusi Pendidikan lain. Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka saya siap menerima segala konsekuensinya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya.

Surabaya, 24 Januari 2023



Nur Lickha Lavenia
NPM. 19082010038

Judul : ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER
TERHADAP ISU GANGGUAN DEPRESI DENGAN
MENGUNAKAN METODE NAÏVE BAYES

Pembimbing 1 : Dhian Satria Yudha Kartika, S.Kom, M.Kom

Pembimbing 2 : Reisa Permatasari, S.T, M.Kom

ABSTRAK

Gangguan depresi adalah isu serius dalam kesehatan mental global. Penelitian ini menganalisis sentimen terkait gangguan depresi di Twitter menggunakan metode Naïve Bayes. Hal ini penting karena media sosial, terutama Twitter, menjadi platform penting dalam berbagi perasaan dan emosi serta mempengaruhi kesejahteraan mental masyarakat.

Penelitian ini melibatkan pengumpulan dan pemrosesan data *tweet* terkait depresi dengan menggunakan metode *snsrape*. Tiga metode *Naïve Bayes* (*Multinomial*, *Gaussian*, *Bernoulli*) dibandingkan untuk mengklasifikasikan sentimen positif, negatif, atau netral pada *tweet* terkait gangguan depresi di Twitter.

Hasil pengujian menunjukkan bahwa metode *Multinomial Naïve Bayes* memiliki tingkat akurasi terbaik dalam analisis sentimen terkait gangguan depresi di Twitter, dengan akurasi sebesar 90,13%. Rekomendasi yang diusulkan adalah menggunakan metode *Multinomial Naïve Bayes* sebagai pendekatan yang efektif dalam mengklasifikasikan sentimen gangguan depresi di media sosial..

Kata kunci: gangguan depresi, kesehatan mental, sentiment, twitter, klasifikasi, Naïve Bayes

Judul : ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER
TERHADAP ISU GANGGUAN DEPRESI DENGAN
MENGUNAKAN METODE NAÏVE BAYES

Pembimbing 1 : Dhian Satria Yudha Kartika, S.Kom, M.Kom

Pembimbing 2 : Reisa Permatasari, S.T, M.Kom

ABSTRACT

Depression disorder is a serious issue in global mental health. This research analyzes sentiments related to depression disorder on Twitter using the Naïve Bayes method. It is important because social media, especially Twitter, has become a significant platform for sharing feelings and emotions, as well as influencing the mental well-being of society.

The study involves collecting and processing tweet data related to depression using the sncrape method. Three Naïve Bayes methods (Multinomial, Gaussian, Bernoulli) are compared to classify positive, negative, or neutral sentiments in tweets related to depression disorder on Twitter.

The test results indicate that the Multinomial Naïve Bayes method has the highest accuracy in sentiment analysis related to depression disorder on Twitter, with an accuracy of 90.13%. The recommended approach for effectively classifying sentiments related to depression disorder on social media is to use the Multinomial Naïve Bayes method.

Keywords: *depression disorder, mental health, sentiment, Twitter, classification, Naïve Bayes.*

KATA PENGANTAR

Puji dan syukur kehadiran ALLAH SWT, atas limpahan Rahmat serta Kasih Sayang-Nya sehingga penulis dapat menyelesaikan Skripsi yang merupakan persyaratan dalam menyelesaikan Program Studi S1/Sistem Informasi di Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Dalam penyusunan Skripsi ini tidak terlepas dari bantuan berbagai pihak, dan dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU., selaku Rektor UPN “Veteran” Jawa Timur
2. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur
3. Agung Bramasta Putra, S.Kom., M. Kom., selaku Ketua Program Studi Sistem Informasi Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur
4. Dhian Satria Yudha Kartika, S.Kom., M.Kom., selaku dosen pembimbing I yang telah memberikan bimbingan dan arahan dalam penyelesaian skripsi ini.
5. Reisa Permatasari, S.T., M.Kom., selaku pembimbing II yang telah memberikan bimbingan dan arahan dalam penyelesaian skripsi ini.
6. Seluruh dosen dan staf pendidikan UPN “Veteran” Jawa Timur atas ilmu pengetahuan yang telah diberikan dan membantu selama proses perkuliahan.
7. Teristimewa untuk orangtua tercinta yang telah memberikan kasih sayang dan cinta kasih dalam mendidik, membesarkan, mendampingi dengan dukungan moril dan kepercayaan, serta do’a restu dalam setiap langkah penulis
8. Seluruh rekan-rekan pendidikan S1 Sistem Informasi UPN “Veteran” Jawa Timur 2019 yang telah berjuang Bersama.
9. Seluruh rekan-rekan di Pondok Pesantren Mahasiswa Khoirul Huda Surabaya yang telah berjuang dan tidak menyerah Bersama dalam kuliah-mondok.

Penulis menyadari bahwa dalam laporan Skripsi ini, namun penulis berharap semoga Skripsi ini dapat ikut menunjang perkembangan ilmu pengetahuan, khususnya ilmu komputer.

Surabaya, 07 Juli 2023

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR PERNYATAAN	ii
LEMBAR PENGESAHAN	ii
KETERANGAN REVISI.....	ii
LEMBAR PERNYATAAN ORISINILITAS	v
ABSTRAK	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	1
BAB I.....	2
1.1 Latar Belakang	2
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan.....	5
1.5 Sistematika Penulisan.....	5
BAB II.....	7
2.1 Dasar Teori	7
2.1.1 Kesehatan Mental	7
2.1.2 Gangguan Depresi	8
2.1.3 Text Mining	9
2.1.4 <i>Preprocessing Text</i>	9
2.1.5 <i>Exploratory Data Analysis (EDA)</i>	12

2.1.6	Sentiment Analysis	12
2.1.7	Twitter.....	13
2.1.8	<i>Social Network Scraping</i> (SNScrap)	14
2.1.9	Klasifikasi Teks	14
2.1.10	Algoritma Naive Bayes.....	15
2.1.11	<i>Term Frequency-Inverse Document Frequency</i> (TF-IDF) ...	17
2.1.12	<i>Confusion Matrix</i>	18
2.1.13	<i>Python</i>	21
2.2	Penelitian Terdahulu.....	22
BAB III	25
3.1	Studi Literatur.....	25
3.2	Analisis Kebutuhan	26
3.2.1	Kebutuhan Data	26
3.2.2	Kebutuhan Hardware dan Software.....	26
3.3	Perancangan Model	26
3.3.1	Pengumpulan Data.....	27
3.3.2	Penyaringan Data.....	28
3.3.3	Pelabelan Data	28
3.3.4	<i>Text Preprocessing</i>	29
3.3.5	Exploratory Data Analysis (EDA).....	31
3.3.6	Pembagian Data	31
3.3.7	Penambahan Data Validasi	31
3.3.8	Pembobotan TF-IDF.....	32
3.3.9	Klasifikasi Naïve Bayes.....	32
3.3.10	Evaluasi Model Klasifikasi	32
3.4	Visualisasi	33

BAB IV	36
4.1 Implementasi Kebutuhan.....	36
4.1.1 Kebutuhan Data	36
4.1.2 Kebutuhan Hardware dan Software.....	36
4.2 Pembangunan Model	37
4.2.1 Pengumpulan Data.....	37
4.2.2 Penyaringan Data.....	41
4.2.3 Pelabelan Data	42
4.2.4 <i>Text Preprocessing</i>	44
4.2.5 <i>Exploratory Data Analysis (EDA)</i>	84
4.2.6 Pembagian Data	93
4.2.7 Penambahan Data Validasi	94
4.2.8 Pembobotan TF-IDF	95
4.2.9 Klasifikasi <i>Naïve bayes</i>	96
4.2.10 Evaluasi Model Klasifikasi	101
4.3 Visualisasi	107
BAB V.....	114
5.1 Kesimpulan.....	114
5.2 Saran	115
DAFTAR PUSTAKA	116
LAMPIRAN	119

DAFTAR TABEL

Table 2.1 Rumus Confusion Matrix (Sumber: Ferdinandus, 2015).....	20
Table 2.2 Rumus Multiclass Confusion Matrix	20
Tabel 4. 1 Perangkat keras (hardware).....	36
Tabel 4. 2 Perangkat lunak (software)	36
Tabel 4. 3 Contoh tweet yang diperoleh	39
Tabel 4. 4 Hasil penyaringan data.....	41
Tabel 4. 5 Hasil pelabelan data	43
Tabel 4. 6 Contoh hasil case folding	45
Tabel 4. 7 Contoh hasil menghilangkan URL situs lain dan URL gambar dan video	49
Tabel 4. 8 Contoh hasil menghilangkan mention atau tag nama akun.....	52
Tabel 4. 9 Contoh hasil menghilangkan hashtag atau tagar.....	55
Tabel 4. 10 Contoh hasil menghilangkan angka	57
Tabel 4. 11 Contoh hasil menghilangkan huruf belakang.....	60
Tabel 4. 12 Contoh hasil menghilangkan huruf tunggal	63
Tabel 4. 13 Contoh hasil menghilangkan spasi berlebih.....	66
Tabel 4. 14 Contoh hasil menghilangkan baris baru.....	69
Tabel 4. 15 Contoh hasil menghilangkan tanda baca.....	72
Tabel 4. 16 Contoh hasil menghilangkan simbol.....	75
Tabel 4. 17 Contoh hasil tokenizing	77
Tabel 4. 18 Contoh hasil stopword removal	80
Tabel 4. 19 Contoh hasil stemming.....	83
Tabel 4. 20 Contoh data validasi.....	94
Tabel 4. 21 Report klasifikasi secara keseluruhan	104

DAFTAR GAMBAR

Gambar 1. 1 Platform Media Sosial Yang Banyak Digunakan Di Indonesia Tahun 2022 (Sumber : We Are Social, 2022).....	2
Gambar 2. 1 Flowchart Algoritma Naïve Bayes (Sumber : Prasetyo, 2012).....	17
Gambar 3. 1 Alur Penelitian.....	25
Gambar 3. 2 Diagram Alir Perancangan Model Klasifikasi	27
Gambar 3. 3 Diagram Alir Text Preprocessing.....	29
Gambar 3. 4 Wireframe visualisasi web : halaman sentiment analysis	34
Gambar 3. 5 Wireframe visualisasi web : halaman word clouds.....	35
Gambar 3. 6 Wireframe visualisasi web : halaman depression trends.....	35
Gambar 4. 1 Source code pengumpulan data	37
Gambar 4. 2 Hasil pengambilan data	38
Gambar 4. 3 Source code menyimpan file ke dalam format .csv	38
Gambar 4. 4 Source code penghapusan kolom	39
Gambar 4. 5 Hasil pelabelan data	44
Gambar 4. 6 Source code case folding.....	45
Gambar 4. 7 Source code menghilangkan URL situs lain atau URL gambar dan video.....	48
Gambar 4. 8 Source code menghilangkan mention atau tag nama akun	51
Gambar 4. 9 Source code menghilangkan tagar atau hashtag.....	54
Gambar 4. 10 Source code menghilangkan angka.....	57
Gambar 4. 11 Source code menghilangkan huruf berulang.....	59
Gambar 4. 12 Source code menghilangkan huruf tunggal.....	62
Gambar 4. 13 Source code menghilangkan spasi berlebih	65
Gambar 4. 14 Source code menghilangkan baris baru.....	68
Gambar 4. 15 Source code menghilangkan tanda baca.....	71
Gambar 4. 16 Source code menghilangkan simbol.....	74
Gambar 4. 17 Source code tokenizing	76
Gambar 4. 18 Source code stopword removal	79
Gambar 4. 19 Source code stemming	82

Gambar 4. 20 Source code cek persebaran label dengan diagram lingkaran.....	84
Gambar 4. 21 Hasil persebaran label dengan diagram lingkaran.....	85
Gambar 4. 22 Source code plotting tweet.....	86
Gambar 4. 23 Hasil grafik plotting tweet.....	86
Gambar 4. 24 Source code mengetahui banyaknya panjang tweet.....	87
Gambar 4. 25 Hasil grafik banyaknya panjang tweet.....	87
Gambar 4. 26 Source code frekuensi kata digunakan.....	88
Gambar 4. 27 Hasil grafik frekuensi kata digunakan.....	89
Gambar 4. 28 Source code word cloud.....	90
Gambar 4. 29 Hasil word cloud.....	90
Gambar 4. 30 Source code word cloud berdasarkan class.....	91
Gambar 4. 31 Hasil word cloud class positif.....	91
Gambar 4. 32 Hasil word cloud class negatif.....	92
Gambar 4. 33 Hasil word cloud class netral.....	92
Gambar 4. 34 Source code pembagian data hold out.....	93
Gambar 4. 35 Source code pembobotan TF-IDF.....	95
Gambar 4. 36 Source code dan hasil klasifikasi multinomial naive bayes.....	97
Gambar 4. 37 Source code dan hasil klasifikasi bernoulli naive bayes.....	97
Gambar 4. 38 Source code dan hasil klasifikasi gaussian naive bayes.....	98
Gambar 4. 39 Source code dan hasil klasifikasi data validasi multinomial naive bayes.....	99
Gambar 4. 40 Source code dan hasil klasifikasi data validasi bernoulli naive bayes.....	99
Gambar 4. 41 Source code dan hasil klasifikasi data validasi gaussian naive bayes.....	99
Gambar 4. 42 Source code dan hasil confusion matrix dan f1-score multinomial naive bayes.....	101
Gambar 4. 43 Source code dan hasil confusion matrix dan f1-score bernoulli naive bayes.....	102
Gambar 4. 44 Source code dan hasil confusion matrix dan f1-score gaussian naive bayes.....	103

Gambar 4. 45 Source code dan hasil pengecekan overfitting multinomial naïve bayes.....	105
Gambar 4. 46 Source code dan hasil pengecekan overfitting bernoulli naïve bayes	106
Gambar 4. 47 Source code dan hasil pengecekan overfitting gaussian naïve bayes	106
Gambar 4. 48 Tampilan Visualisasi Sentimen Analysis	108
Gambar 4. 49 Visualisasi grafik.....	108
Gambar 4. 50 Visualisasi pie chart	109
Gambar 4. 51 Visualisasi diagram batang.....	110
Gambar 4. 52 Visualisasi data tweet hasil klasifikasi	110
Gambar 4. 53 Fitur filter	111
Gambar 4. 54 Tampilan Visualisasi Word Clouds.....	112
Gambar 4. 55 Fitur selection.....	112
Gambar 4. 56 tampilan visualisasi depression trends	113

DAFTAR LAMPIRAN

- I. Source code visualisasi web : halaman sentiment analysis
- II. Source code visualisasi web : halaman word clouds
- III. Source code visualisasi web : halaman kata yang sering muncul