

**ANALISIS DAN KLASIFIKASI TEKS BERINDIKASI
KESEHATAN MENTAL MENGGUNAKAN MODEL
MACHINE LEARNING**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

MUCH ARIEF RE FADILLA

23.22.2425

Kepada

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

**ANALISIS DAN KLASIFIKASI TEKS BERINDIKASI
KESEHATAN MENTAL MENGGUNAKAN MODEL
MACHINE LEARNING**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh
MUCH ARIEF RE FADILLA
23.22.2425

Kepada

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

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS DAN KLASIFIKASI TEKS BERINDIKASI KESEHATAN
MENTAL MENGGUNAKAN MOD EL MACHINE LEARNING**

yang disusun dan diajukan oleh

Much Arief Re Fadilla

23.22.2526

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 17 April 2025

Dosen Pembimbing,



Yoga Pristyanto, S.Kom., M.Eng
NIK. 190302412

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS DAN KLASIFIKASI TEKS BERINDIKASI KESEHATAN
MENTAL MENGGUNAKAN MODEL MACHINE LEARNING**

yang disusun dan diajukan oleh

Much Arief Re Fadilla

23.22.2526

Telah dipertahankan di depan Dewan Penguji
pada tanggal 17 April 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ika Nur Fajri, S.Kom., M.Kom.
NIK. 190302268

Arif Nur Rohman, M.Kom
NIK. 190302684

Yoga Pristyanto, S.Kom., M.Eng.
NIK. 190302412



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 17 April 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Much Arief Re Fadilla

NIM : 23.22.2526

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis dan Klasifikasi Teks Berindikasi Kesehatan Mental Menggunakan Model Machine Learning

Dosen Pembimbing : Yoga Pristyanto, S.Kom., M.Eng.

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, 17 April 2025

Yang Menyatakan,



Much Arief Re Fadilla

HALAMAN PERSEMBAHAN

Bismillahirrahmanirrahim, puji syukur kepada Allah SWT, yang telah memberikan kekuatan, kesabaran, dan pertolongan-Nya dalam setiap proses penelitian dan pengerjaan naskah skripsi ini, sehingga dapat diselesaikan dengan baik dan tepat waktu. Tentunya, pengerjaan naskah skripsi ini bukan hanya hasil usaha saya seorang, namun juga merupakan hasil dari campur tangan dan kontribusi dari beberapa pihak lain. Maka dari itu, saya persembahkan skripsi ini sebagai wujud dedikasi saya kepada :

1. Tuhan yang maha esa, karena dengan izin serta kehendak-Nya, naskah skripsi ini dapat diselesaikan dengan baik.
2. Kedua orang tua, Ibu Parni dan Bapak Darmono yang senantiasa memberi dukungan moril maupun materil.
3. Adik saya, Muhammad Alvin Faiz serta seluruh keluarga saya yang selalu memberikan motivasi dan dukungan emosional.
4. Bapak Yoga Pristyanto, S.Kom., M.Eng., selaku dosen pembimbing saya, yang dengan penuh kesabaran memberikan kepercayaan, motivasi, dan bimbingan sehingga penelitian ini dapat diselesaikan.
5. Bapak Yogi Wijaya dan seluruh keluarga besar perusahaan Kinycurly yang telah memberikan dukungan materil, toleransi dan pengertian sehingga saya dapat menyelesaikan masa studi.
6. Sahabat, teman serta rekan seperjuangan baik di lingkup lingkungan kampus maupun diluar kampus yang telah memberikan semangat, dan motivasi, baik secara langsung maupun tidak langsung.
7. Serta pihak-pihak lain yang tidak bisa saya sebutkan satu persatu.

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Alhamdulillahirabbil'alamin, segal puji kehadiran Allah SWT atas rahmat, berkah, dan hidayah-Nya. Shalawat serta salam senantiasa tercurah kepada Nabi Muhammad SAW, yang telah menjadi teladan bagi umat manusia, sehingga penelitian dan skripsi dengan judul "Analisis dan Klasifikasi Teks Berindikasi Kesehatan Mental Menggunakan Model Machine Learning" ini dapat diselesaikan dengan baik. Penulis menyadari bahwa penyusunan skripsi dan penelitian ini tidak luput dari kekurangan serta kesalahan. Maka dari itu, penulis mengharapkan saran serta masukan konstruktif dari berbagai pihak. Izinkan penulis mengucapkan terima kasih sebesar-besarnya kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta
2. Ibu Prof. Dr. Kusriani, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta
3. Bapak Anggit Dwi Hartanto, M.Kom. selaku Ketua Prodi SI Sistem Informasi Universitas Amikom Yogyakarta
4. Bapak Yoga Pristyanto, S.Kom., M.Eng. selaku dosen pembimbing, yang dengan sabar telah memberikan kepercayaan, motivasi dan bimbingan sehingga penulis dapat menyelesaikan skripsi ini.
5. Serta pihak-pihak lain yang telah berkontribusi dan tidak dapat penulis sebutkan satu-persatu.

Sekali lagi dengan segala kerendahan hati penulis mengucapkan terima kasih yang sebesar-besarnya, penulis mengharapkan semoga skripsi ini dapat bermanfaat bagi khalayak ramai.

Wassalamu'alaikum Wr. Wb.

Yogyakarta, 14 Maret 2025

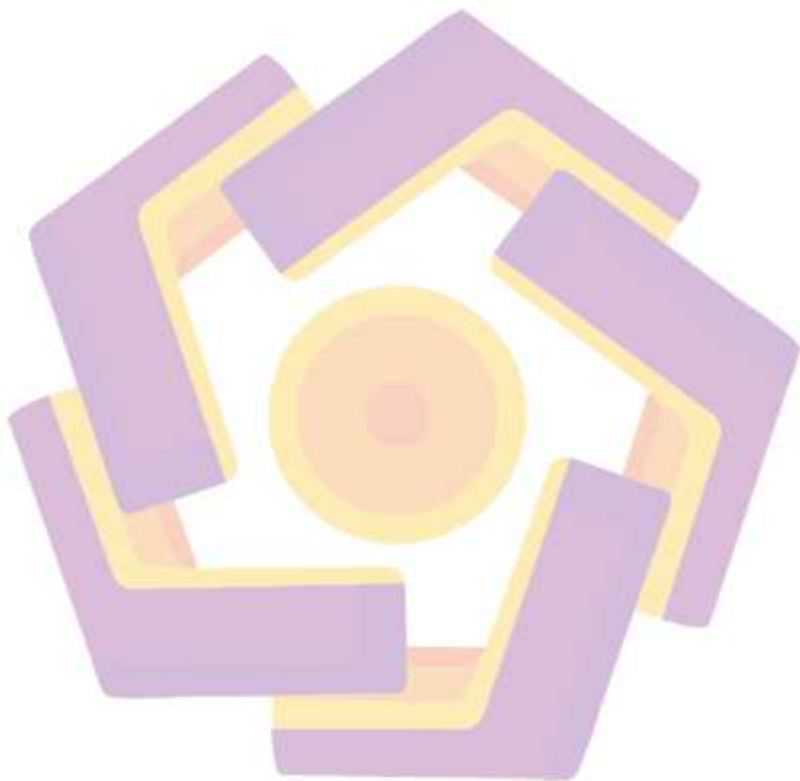
Penulis

DAFTAR ISI

HALAMAN JUDUL	1
HALAMAN PERSETUJUAN	2
HALAMAN PENGESAHAN	3
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	4
HALAMAN PERSEMBAHAN	5
KATA PENGANTAR	6
DAFTAR ISI	7
DAFTAR TABEL	10
DAFTAR GAMBAR	11
DAFTAR LAMBANG DAN SINGKATAN	12
DAFTAR ISTILAH	13
INTISARI	14
ABSTRACT	15
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II	5
TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	11
2.2.1. Bunuh Diri	11
2.2.2. Machine Learning	12
2.2.3. NLP (Natural Language Processing)	12
2.2.4. Dataset Collection	12
2.2.5. Anotation/Labeling	13
2.2.6. Pre-processing	13
2.2.6.1. Stopwords Removal	13
2.2.6.2. Casefolding	14
2.2.6.3. Word Normalization	14
2.2.6.4. Cleansing	14
2.2.6.5. Stemming	14

2.2.7. Resampling	15
2.2.8. Feature Engineering	15
2.2.9. Feature Extraction	16
2.2.9.1. Tokenizing	16
2.2.9.2. N-gram	16
2.2.9.3. TF-IDF	16
2.2.10. Feature Selection	18
2.2.10.1. Chi-Square	18
2.2.11. Modelling	19
2.2.11.1. Naïve Bayes	20
2.2.12. Confusion Matriks	20
2.2.13. Streamlit	22
BAB III	
METODE PENELITIAN	23
3.1. Objek Penelitian	23
3.2. Alur Penelitian	23
3.2.1. Data Collection	25
3.2.2. Preprocessing Pertama	25
3.2.3. Anotasi/Labeling	25
3.2.4. Preprocessing Kedua	26
3.2.5. Resampling	27
3.2.6. Feature Engineering	28
3.2.7. Feature Extraction	28
3.2.1. Feature Selection	29
3.2.2. Modelling & Evaluasi	29
3.2.3. Deployment	29
3.3. Alat dan Bahan	30
BAB IV	
HASIL DAN PEMBAHASAN	32
4.1. Data Collection	32
4.2. Preprocessing pertama	36
4.3. Anotasi/Labeling	38
4.4. Preprocessing kedua	41
4.5. Resampling	45
4.6. Feature Engineering	48
4.7. Feature Extraction	51
4.8. Feature Selection	68
4.9. Modelling & Evaluasi	78

4.1. Deployment	82
BAB V	
PENUTUP	85
5.1 Kesimpulan	85
5.2 Saran	86
REFERENSI	88



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	7
Tabel 2.2 Tabel kontingensi	18
Tabel 4.1 Jumlah data tiap dataset	33
Tabel 4.2 Nama kolom dan tipe datanya	34
Tabel 4.3 Hasil preprocessing pertama	37
Tabel 4.4 Perbandingan <i>preprocessing</i> pertama	37
Tabel 4.5 Data anotator	39
Tabel 4.6 Data yang sudah dilabeli	39
Tabel 4.7 Perbandingan <i>cleansing</i> kedua	42
Tabel 4.8 Perbandingan <i>word normalization</i>	42
Tabel 4.9 Perbandingan <i>stopwords removal</i>	43
Tabel 4.10 Perbandingan <i>stemming</i>	43
Tabel 4.11 Perbandingan hasil <i>preprocessing</i> kedua	43
Tabel 4.12 Contoh data fitur	48
Tabel 4.13 Contoh data target	48
Tabel 4.14 Contoh hasil pengelompokkan	51
Tabel 4.15. Fitur Kata	52
Tabel 4.16 Contoh dokumen TF-IDF	53
Tabel 4.17 Tabel TF	55
Tabel 4.18 Tabel $TF+df$	57
Tabel 4.19 Tabel normalisasi TF	61
Tabel 4.20 Tabel IDF	62
Tabel 4.21 Tabel TF-IDF	64
Tabel 4.22 Tabel kontingensi fitur "mati"	68
Tabel 4.23 Tabel kontingensi fitur "capek"	69
Tabel 4.24 Tabel kontingensi fitur "hidup"	69
Tabel 4.25 Tabel urutan nilai Chi-Square	75
Tabel 4.26 Tabel data split	78
Tabel 4.27 Hasil Evaluasi	80

DAFTAR GAMBAR

Gambar 2.1. Confussion Matrix	21
Gambar 3.1 Alur Penelitian	24
Gambar 4.1 Proses Data Colletion	32
Gambar 4.2 Proses Data Colletion 2	33
Gambar 4.3 Hasil Data Colletion	33
Gambar 4.4 Kolom dan tipe data dataset	35
Gambar 4.5 Perbandingan hasil <i>preprocessing</i> 1	38
Gambar 4.6 Kolom <i>clean_text</i> dan <i>label</i>	41
Gambar 4.7 Kolom <i>full_text</i> , <i>label</i> dan <i>clean_text</i>	45
Gambar 4.8 Diagram batang jumlah selisih data	46
Gambar 4.9 Diagram pie hasil <i>oversampling</i>	47
Gambar 4.10 Diagram pie hasil <i>undersampling</i>	48
Gambar 4.11 <i>Feature</i>	50
Gambar 4.12 <i>Label</i>	51
Gambar 4.13 <i>Wordeloud</i> fitur kata	67
Gambar 4.14 Diagram batang fitur kata	67
Gambar 4.15 Jumlah fitur kata setelah <i>selection</i>	76
Gambar 4.16 Diagram batang <i>Chi-square</i> fitur kata	77
Gambar 4.17 Fitur dan nilai <i>Chi-Square</i> nya	77
Gambar 4.18 Jumlah data training dan testing	78
Gambar 4.19 Input berindikasi	79
Gambar 4.20 Output berindikasi	79
Gambar 4.21 Input tidak berindikasi	79
Gambar 4.22 Output tidak berindikasi	80
Gambar 4.23 Confusion matriks	80
Gambar 4.24 <i>Interface Streamlit</i>	83
Gambar 4.25 Test <i>Streamlit</i> 1	83
Gambar 4.26 Test <i>Streamlit</i> 2	84

DAFTAR LAMBANG DAN SINGKATAN



WHO	World Health Organization
AI	Artificial Intelligence
ML	Machine Learning
NLP	Natural Language Processing
CNB	Complement Naïve Bayes
SVM	Support Vector Machines
KNN	K-Nearest Neighbors
CBT	Cognitive Behavioral Therapy
API	Application Programming Interface
URL	Uniform Resource Locator
Σ	Sigma
log	Logaritma
TF	Term Frequency
IDF	Inverse Document Frequency
DF	Document Frequency
TF-IDF	Term Frequency - Inverse Document Frequency
GPU	Graphics Processing Unit
TPU	Tensor Processing Unit
NLTK	Natural Language Toolkit
SMOTE	Synthetic Minority Over-sampling Technique
ID	Identifier

DAFTAR ISTILAH

Algoritma	Langkah/formula sistematis untuk menyelesaikan masalah.
Tweet	Postingan singkat di Twitter/X.
Psikososial	Faktor psikologis dan sosial yang memengaruhi individu.
Cyberbullying	Perundungan melalui internet.
Cyber Harassment	Pelecehan online.
Polaritas	Sentimen teks (<i>positif, negatif, netral</i>).
Bahasa gaul/slang	Kata informal dalam komunikasi.
Emote	Simbol ekspresi dalam teks digital.
Noise	Data yang mengganggu analisis.
Vektor	Representasi data dalam bentuk angka.
Chi-Square	Uji statistik hubungan antar variabel.
Kontingensi	Tabel dalam uji Chi-Square.
Observasi	Unit data yang diamati.
Variabel	Elemen yang dapat berubah dalam data.
Probabilitas	Peluang suatu peristiwa terjadi.
Prior Probability	Probabilitas awal sebelum data baru.
Posterior Probability	Probabilitas setelah data tambahan.
Matriks (Matrix)	Tabel dengan baris dan kolom.
Deployment	Menjalankan model dalam penerapan machine learning.
Ensemble Bagging	Gabungan model untuk hasil lebih akurat.
Trial and Error	Mencoba berbagai solusi hingga berhasil.
Web Scraping	Mengambil data dari web secara otomatis.
Consumer Key	Kunci API untuk otorisasi.
Consumer Secret	Kode rahasia untuk API.
Access Token	Token akses API atas nama pengguna.
Access Token Secret	Kode rahasia untuk keamanan API.
Framework	Kerangka kerja pengembangan perangkat lunak.
Open Source	Kode sumber bebas diakses dan dimodifikasi.
Data Science	Analisis data menggunakan ilmu komputer & statistik.
Keyword	Kata kunci dalam pencarian/analisis teks.
Data Anotator	Pemberi label pada data untuk pelatihan AI.
Dataframe	Struktur data berbentuk tabel (pandas).
Feature	Atribut penting dalam analisis machine learning.
Vocabulary	Kumpulan kata unik dalam dataset teks.
Python's Pickle	Modul Python untuk menyimpan/memuat objek.
Select KBest	Seleksi fitur terbaik untuk model.
Model	Representasi matematis untuk prediksi/klasifikasi.
Crawling	Mengumpulkan data dari internet secara otomatis.

INTISARI

Kesehatan mental belakangan ini menjadi isu kritis di Indonesia, dengan jumlah kasus yang terus meningkat setiap tahunnya, seiring dengan meningkatnya kasus bunuh diri di Indonesia. Bunuh diri dan percobaan bunuh diri tidak hanya berdampak pada individu, tetapi juga memiliki dampak mendalam pada keluarga dan masyarakat. Meskipun demikian, deteksi dini kecenderungan bunuh diri tetap menjadi tantangan, terutama karena perilaku ini sering kali diekspresikan secara tidak langsung di platform media sosial seperti *Twitter*. Dalam penelitian ini, model machine learning menggunakan algoritma *Naïve Bayes* yang dikembangkan untuk melakukan klasifikasi tweet yang mengindikasikan kecenderungan bunuh diri. Data yang diberi label oleh ahli digunakan untuk melakukan pelatihan model, dan hasil evaluasi menunjukkan akurasi yang tinggi sebesar 88%, dengan skor presisi dan recall masing-masing 95% dan 80% untuk kelas negatif, serta 83% dan 96% untuk kelas positif. Temuan ini menunjukkan bahwa model secara efektif mengidentifikasi teks yang berkaitan dengan bunuh diri, menawarkan potensi sebagai alat untuk pencegahan bunuh diri dini. Peningkatan di masa depan dapat mencakup eksplorasi algoritma alternatif, penyempurnaan metode pra-pemrosesan, dan perluasan dataset untuk hasil yang lebih baik.

Kata kunci: Bunuh diri, Machine Learning, Model, Klasifikasi, Twitter.

ABSTRACT

Mental health has recently become a critical issue in Indonesia, with the number of cases increasing each year, alongside the rising suicide rates in the country. Suicide and suicide attempts not only impact individuals but also have profound effects on families and communities. However, early detection of suicidal tendencies remains a challenge, especially since such behavior is often expressed indirectly on social media platforms like Twitter. In this study, a machine learning model using the Naïve Bayes algorithm was developed to classify tweets indicating suicidal tendencies. Expert-labeled data was used to train the model, and the evaluation results showed a high accuracy of 88%, with precision and recall scores of 95% and 80% for the negative class and 83% and 96% for the positive class, respectively. These findings suggest that the model effectively identifies text related to suicide, offering potential as a tool for early suicide prevention. Future improvements may include exploring alternative algorithms, refining preprocessing methods, and expanding the dataset for better results.

Keyword: *Classification, Machine Learning, Model, Suicide, Twitter*