

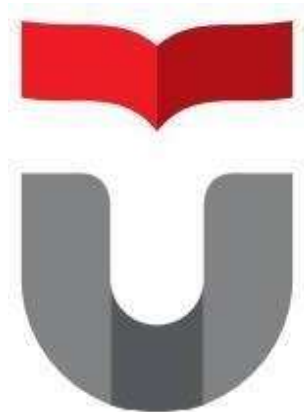
TUGAS AKHIR

**ANALISIS PERSPEKTIF PADA MEDIA SOSIAL TIKTOK
TERHADAP KESEHATAN MENTAL GENERASI Z DI
INDONESIA MENGGUNAKAN MODEL *SUPPORT VECTOR
MACHINE* (SVM)**

Oleh :

BECHAM NOVENTINUS PARSAORAN

1202213129



PROGRAM STUDI SISTEM INFORMASI 1202213129

FAKULTAS REKAYASA INDUSTRI

UNIVERSITAS TELKOM

2025

ABSTRAK

Isu kesehatan mental menjadi perhatian serius di kalangan Generasi Z, terutama di tengah tingginya interaksi mereka dengan media sosial seperti TikTok. Platform ini tidak hanya digunakan sebagai sarana hiburan, tetapi juga sebagai ruang untuk mengekspresikan perasaan dan pengalaman pribadi, termasuk yang berkaitan dengan kesehatan mental. Penelitian ini bertujuan untuk menganalisis perspektif kesadaran kesehatan mental Generasi Z di Indonesia melalui komentar pengguna di TikTok. Dengan menggunakan pendekatan berbasis *machine learning*, khususnya algoritma *Support Vector Machine* (SVM) dan representasi fitur teks *Term Frequency-Inverse Document Frequency* (TF-IDF), sistem klasifikasi otomatis dikembangkan untuk mengidentifikasi dua kategori perspektif: *aware* (sadar) dan *not aware* (tidak sadar).

Model SVM terbaik yang dibangun mencapai akurasi sebesar 92,06%, dengan nilai *precision* 96,59%, *recall* 86,55%, dan *F1-score* 91,32%. Evaluasi melalui *confusion matrix* dan ROC menghasilkan AUC sebesar 0,95, menunjukkan kemampuan model dalam membedakan kedua kelas dengan sangat baik. Dari total 13.331 komentar yang dianalisis, 53,2% termasuk kategori *not aware* dan 46,8% *aware*. Analisis *topic modelling* dengan metode *Latent Dirichlet Allocation* (LDA) mengungkap lima tema utama di masing-masing kelompok, yang menggambarkan spektrum kesadaran kesehatan mental Generasi Z.

Proses *preprocessing* teks menggunakan pustaka Sastrawi terbukti meningkatkan performa klasifikasi. Visualisasi *word cloud* dan kata penting menunjukkan bahwa kelompok *aware* menggunakan istilah seperti "psikolog", "kecemasan", dan "emosi", sementara kelompok *not aware* lebih didominasi kata-kata umum dan ambigu. Penelitian ini memberikan kontribusi terhadap pengembangan sistem pemantauan dan intervensi kesehatan mental digital yang kontekstual dan berbasis data.

Kata kunci : Generasi Z, kesehatan mental, LDA, SVM, TikTok

ABSTRACT

Mental health has become a serious concern among Generation Z, particularly due to their intense interaction with social media platforms like TikTok. This platform is not only used for entertainment but also serves as a space to express personal feelings and experiences, including those related to mental health. This study aims to analyze the mental health awareness perspective of Indonesian Generation Z through TikTok user comments. By using a machine learning approach—specifically the Support Vector Machine (SVM) algorithm and Term Frequency-Inverse Document Frequency (TF-IDF) feature representation—an automatic classification system was developed to identify two categories: aware and not aware.

The best SVM model achieved an accuracy of 92.06%, with a precision of 96.59%, recall of 86.55%, and F1-score of 91.32%. Evaluation using a confusion matrix and ROC curve produced an AUC of 0.95, confirming excellent discriminative capability. From a total of 13,331 analyzed comments, 53.2% were categorized as not aware and 46.8% as aware. Topic modelling using Latent Dirichlet Allocation (LDA) revealed five main themes within each group, reflecting the range of mental health awareness among Generation Z.

Text preprocessing using the Sastrawi library significantly improved classification performance. Word cloud visualization and salient term analysis showed that the aware group used specific terms such as “psychologist,” “anxiety,” and “emotion,” while the not aware group was dominated by general and ambiguous expressions. This research contributes to the development of contextual, data-driven digital mental health monitoring and intervention systems

Keywords — Generation Z, LDA, mental health, SVM, TikTok

LEMBAR PENGESAHAN

Tugas Akhir dengan judul :

**ANALISIS PERSPEKTIF PADA MEDIA SOSIAL TIKTOK
TERHADAP KESEHATAN MENTAL GENERASI Z DI
INDONESIA MENGGUNAKAN MODEL *SUPPORT VECTOR
MACHINE* (SVM)**

Telah disetujui dan disahkan pada Sidang Tugas Akhir

Program Studi Strata 1 Sistem Informasi

Fakultas Rekayasa Industri Universitas Telkom

Oleh :

BECHAM NOVENTINUS PARSAORAN

1202213129

Bandung, 31 Juli 2025

Disetujui oleh,

Pembimbing 1,



Rahmat Fauzi, S.T., M.T.
19000023

Pembimbing 2,



Riska Yanu Fa'rifah
NIP Pembimbing

LEMBAR PERNYATAAN ORISINALITAS



Nama : Becham Noventinus Parsaoran

NIM : 1202213129

Alamat : Jalan Mushola Ar Rahman No.102

No. Tlp : 087741221452

Menyatakan bahwa Tugas Akhir ini merupakan karya orisinal saya sendiri. Atas pernyataan ini, saya siap menanggung risiko atau sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran terhadap kejujuran akademik atau etika keilmuan dalam karya ini, atau ditemukan bukti yang menunjukkan ketidakaslian karya ini.

Bandung, 31 Juli 2025

Becham Noventinus Parsaroan

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Tuhan Yesus Kristus, Sang Juru Selamat dan sumber segala hikmat. Atas kasih karunia, penyertaan, dan rahmat-Nya, penulis dapat menyelesaikan tugas akhir ini dengan baik dan tepat waktu. Dalam segala keterbatasan dan tantangan, penulis percaya bahwa Tuhan telah menuntun setiap proses hingga laporan ini terselesaikan.

Tugas Akhir yang berjudul “ **ANALISIS PERSPEKTIF PADA MEDIA SOSIAL TIKTOK TERHADAP KESEHATAN MENTAL GENERASI Z DI INDONESIA MENGGUNAKAN MODEL SUPPORT VECTOR MACHINE**” ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan Strata 1 pada Program Studi Sistem Informasi, Fakultas Rekayasa Industri, Universitas Telkom.

Selama proses penyusunan tugas akhir ini, penulis mengalami berbagai kendala, baik teknis maupun non-teknis. Namun, semua kesulitan tersebut dapat teratasi berkat doa, dukungan, serta bantuan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Yang terkasih dan teristimewa, kedua orang tua penulis Bapak Venders Siregar, S.H. dan Ibu Berta Rouli Malau yang selalu memberikan cinta, doa, dan semangat yang tiada henti dalam setiap langkah kehidupan dan pendidikan penulis.
2. Ibu Riska Yanu Fa'rifah, S.Si., M.Si. selaku dosen wali yang telah membimbing dan memberikan arahan dengan penuh kesabaran selama masa perkuliahan.
3. Bapak Rahmat Fauzi, S.T., M.T. selaku dosen pembimbing satu dan Ibu Riska Yanu Fa'rifah, S.Si., M.Si. selaku dosen pembimbing dua, yang telah memberikan bimbingan, motivasi, serta masukan yang sangat berarti sepanjang proses pengerjaan tugas akhir ini.

4. Seluruh sahabat Putra Baraya dan Bronxxx seperjuangan, yang telah menjadi tempat berbagi semangat, tawa, dan dukungan dalam setiap tantangan yang dihadapi.
5. Semua pihak lainnya yang tidak dapat penulis sebutkan satu per satu, namun telah dengan tulus memberikan bantuan dan dukungannya dalam bentuk apapun selama proses penyusunan tugas akhir ini.

Penulis menyadari bahwa tugas akhir ini masih jauh dari kesempurnaan. Oleh karena itu, penulis dengan rendah hati memohon maaf atas segala kekurangan serta sangat mengharapkan kritik dan saran yang membangun demi perbaikan ke depannya.

Akhir kata, penulis berharap semoga laporan tugas akhir ini dapat memberikan manfaat dan menjadi referensi yang berguna bagi para pembaca di masa mendatang.

DAFTAR ISI

ABSTRAK	i
<i>ABSTRACT</i>	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN ORISINALITAS	iv
Kata Pengantar	v
Daftar Isi.....	vii
Daftar Gambar.....	xi
Daftar Tabel	xiii
Daftar Istilah.....	xiv
Bab I Pendahuluan	1
I.1 Latar Belakang	1
I.2 Perumusan Masalah.....	5
I.3 Tujuan Penelitian.....	5
I.4 Manfaat Penelitian.....	5
I.5 Batasan Penelitian	6
I.6 Sistematika Penulisan.....	7
Bab II Tinjauan Pustaka	10
II.1 Dasar Teori	10
II.1.1 Kesehatan Mental.....	10
II.1.2 Social Media.....	12
II.1.3 Machine Learning	14
II.1.4 Analisis Perspektif	16
II.1.5 Natural Language Processing.....	17
II.1.6 <i>Knowledge Discovery in Database</i>	22
II.1.7 Topic Modelling.....	24

II.1.8	TikTok.....	26
II.1.9	<i>Support Vector Machine</i>	29
II.1.10	<i>Latent Dirichlet Allocation</i>	32
II.1.11	Confusion Matrix	34
II.1.12	<i>Labelling</i>	37
II.2	Perbandingan Metode.....	38
II.3	State of The Art	43
Bab III	Metodologi Penelitian.....	47
III.1	Pengembangan Model Konseptual	47
III.1.1	Komponen Model Konseptual	48
III.2	Sistematika Penyelesaian Masalah	49
III.2.1	Seleksi Data.....	50
III.2.2	Pengumpulan Data	51
III.2.3	Pengolahan Data dan Pra-Pemrosesan	52
III.2.4	Transformasi Data.....	53
III.2.5	Penambahan Data.....	54
III.2.6	Penarikan Kesimpulan dan Rekomendasi.....	56
III.3	Alasan Pemilihan Metode.....	57
III.3.1	Kemampuan SVM dalam Klasifikasi Teks.....	58
III.3.2	Keunggulan LDA Dalam Topic Modeling	59
III.3.3	Kesesuaian Dengan Kebutuhan Dataset dari Media Sosial	60
III.3.4	Kemudahan Dalam Proses <i>Fine-Tuning</i>	61
III.3.5	Skalabilitas dan Dukungan Teknologi	62
III.3.6	Implikasi Metodologis	63
III.4	Pengumpulan Data.....	63
III.4.1	Sumber Data.....	64

III.4.2	Metode Pengumpulan Data	65
III.4.3	Penyarigan Data	65
III.5	Pengolahan Data	66
III.5.1	<i>Text Pre-processing</i>	67
III.5.2	Tokenisasi	67
III.5.3	Representasi Teks dengan	68
III.5.4	Penyesuaian Panjang Input	69
III.6	Metode Evaluasi	69
III.6.1	Metode Evaluasi Model	70
III.6.2	Validasi Model	71
III.6.3	Analisis Hasil Evaluasi	71
III.6.4	Evaluasi Visualisasi Data	72
Bab IV	Penyelesaian permasalahan	74
IV.1	Pengumpulan Data	74
IV.2	<i>Preprocessing Data</i>	77
IV.2.1	Pembersihan Data	78
IV.2.2	<i>Lower Casing</i>	81
IV.2.3	Penghapusan <i>Stopwords</i>	83
IV.2.4	Normalisasi Teks	85
IV.2.5	<i>Stemming</i>	87
IV.3	Labelling	89
IV.4	Tokenisasi	91
IV.5	<i>Data Mining</i>	93
IV.5.1	<i>Distribution of Data Training and Data Testing</i>	93
IV.5.2	<i>Support Vector System</i>	95
IV.5.3	Konfigurasi <i>Hyperparameter</i>	98

IV.5.4	<i>Predict Sentiment</i>	100
IV.5.5	Topic Modelling.....	101
IV.6	<i>Data Visualization</i>	105
IV.6.1	Top 20 Kata.....	106
IV.6.2	Distribusi Panjang Teks	108
IV.6.3	Distribusi Label.....	109
IV.6.4	<i>Word Cloud</i>	110
Bab V	Validasi, Analisis Hasil, dan Implikasi	113
V.1	Hasil Labelling Data.....	113
V.2	<i>Wordcloud</i>	114
V.3	Evaluasi Skenario Pengujian.....	115
V.4	Evaluasi Hasil Pengujian.....	118
V.4.1	<i>Confusion Matrix</i>	118
V.4.2	<i>Receiver Operating Characteristic</i>	123
V.5	<i>Topic Modelling</i>	128
V.5.1	<i>Topic Modelling Aware</i>	130
V.5.2	<i>Topic Modelling Tidak Aware</i>	142
V.6	Rekomendasi Setiap Topik.....	153
Bab VI	Kesimpulan dan Saran	156
VI.1	Kesimpulan	156
VI.2	Saran	157
Daftar Pustaka		159

DAFTAR GAMBAR

Gambar II-1. Tahapan <i>Natural Language Processing</i>	19
Gambar III-1. Model Konseptual	47
Gambar III-2. <i>Flowchart</i> Sistematika Penyelesaian Masalah.....	50
Gambar IV-1. Apify TikTok Comment Scraper	74
Gambar IV-2. Menghapus <i>Mention</i>	78
Gambar IV-3. Menghapus Emoji.....	79
Gambar IV-4. Menghapus Emoticon	80
Gambar IV-5. Mendeteksi Dominasi Karakter Khusus	81
Gambar IV-6. Membaca dan Menggabungkan Semua Dataset.....	82
Gambar IV-7. Menghapus Nilai Kosong dan Mengubah ke Huruf Kecil	82
Gambar IV-8. Menggabungkan Stopwords Sastrawi dan Kustom	84
Gambar IV-9. Menghapus <i>Stopwords</i> dari Kolom Teks.....	84
Gambar IV-10. Memuat Dataset dan Kamus Normalisasi.....	86
Gambar IV-11. Fungsi Normalisasi Teks	86
Gambar IV-12. Stemming	88
Gambar IV-13. Labelling	90
Gambar IV-14. Tokenisasi Menggunakan TF-IDF.....	92
Gambar IV-15. <i>Flow Chart Data Distribution</i>	94
Gambar IV-16. <i>Code Pipeline</i> untuk SVM dan TF-IDF.....	95
Gambar IV-17. Ekstraksi Informasi Model	98
Gambar IV-18. Pelatihan dan Prediksi menggunakan <i>Pipeline</i> SVM	100
Gambar IV-19. Evaluasi Hasil	100
Gambar IV-20. Kode <i>CountVectorizer</i> Untuk Representasi Teks	102
Gambar IV-21. Inisialisasi Model LDA.....	103
Gambar IV-22. Proses Pelatihan dan Evaluasi Model	104
Gambar IV-23. Top 20 Kata <i>Aware</i> dan <i>Not Aware</i>	107
Gambar IV-24. Distribusi Panjang Teks.....	108
Gambar IV-25. Distribusi Label	109
Gambar IV-26. <i>Word Cloud</i>	111
Gambar V-1. Distribusi Label <i>Bar Chart</i>	113
Gambar V-2. Hasil Wordcloud Dataset	114

Gambar V-3. 10 Hyperparameter Configurations	116
Gambar V-4. Learning Curve	117
Gambar V-5. Confusion Matrix Model Dasar	119
Gambar V-6. Confusion Matrix Model Terbaik	121
Gambar V-7. Receiver Operating Characteristic Model Dasar	124
Gambar V-8. Receiver Operating Characteristics Model Terbaik	126
Gambar V-9. Distribusi Kata	128
Gambar V-10. Intertopic Distance Map Aware	130
Gambar V-11. Top 30 Most Salient Terms Aware	133
Gambar V-12. Word Cloud Aware	135
Gambar V-13. Intertopic Distance Map Tidak Aware	142
Gambar V-14. Top-30 Most Salient Terms Tidak Aware	145
Gambar V-15. Word Cloud Tidak Aware	147

DAFTAR TABEL

Tabel II-1. <i>Confusion Matrix</i>	35
Tabel II-2. Perbandingan Metode	Error! Bookmark not defined.
Tabel II-3. <i>State of the Art</i>	43
Tabel IV-1. Kriteria.....	75
Tabel IV-2. Menghapus <i>Mention</i>	78
Tabel IV-3. Menghapus <i>Emoji</i>	79
Tabel IV-4. Menghapus <i>Emoticon</i>	80
Tabel IV-5. Menyaring Karakter.....	81
Tabel IV-6. Lower <i>Casing</i>	83
Tabel IV-7. Penghapusan Stopwords	85
Tabel IV-8. Normalisasi Teks.....	87
Tabel IV-9. Stemming.....	88
Tabel IV-10. Arsitektur Model SVM.....	96
Tabel IV-11. Konfigurasi <i>Hyperparameter</i>	98
Tabel V-1. Hasil <i>Confusion Matrix</i> Model Dasar	120
Tabel V-2. Hasil <i>Confusion Matrix</i> Model Terbaik.....	122
Tabel V-3. Topic Label <i>Aware</i>	132
Tabel V-4. <i>Top 5 Words</i> Pada Setiap Topik	139
Tabel V-5. Topik Tidak <i>Aware</i>	143
Tabel V-6. <i>Top 5 Words</i> Untuk Setiap Topik	149
Tabel V-7. Rekomendasi Setiap Aspek	153

DAFTAR ISTILAH

Istilah	Deskripsi	Halaman Pertama Kali Digunakan
<i>Confusion Matrix</i>	Matriks evaluasi performa model klasifikasi yang menunjukkan jumlah prediksi benar dan salah untuk setiap kelas.	28
Generasi Z	Generasi yang lahir sekitar tahun 1997–2012, tumbuh dengan teknologi digital, media sosial, dan internet sebagai bagian dari kehidupan sehari-hari.	9
Kesehatan Mental	Kondisi kesejahteraan emosional, psikologis, dan sosial yang memengaruhi cara berpikir, merasa, dan bertindak dalam kehidupan sehari-hari.	8
Label <i>Aware</i> / <i>Not Aware</i>	Kategori dalam penelitian ini untuk menunjukkan apakah komentar menunjukkan kesadaran terhadap isu kesehatan mental atau tidak.	67
<i>Latent Dirichlet Allocation</i> (LDA)	Algoritma <i>topic modeling</i> yang mengidentifikasi tema tersembunyi dalam kumpulan dokumen teks.	26
Pra-pemrosesan	Proses awal pengolahan teks seperti pembersihan, normalisasi, penghapusan simbol, dan <i>stemming</i> sebelum dianalisis.	49

Istilah	Deskripsi	Halaman Pertama Kali Digunakan
ROC Curve	Grafik yang menggambarkan kinerja model klasifikasi pada berbagai ambang batas klasifikasi.	98
Sentimen	Ekspresi atau sikap emosional dalam teks, baik positif, negatif, maupun netral, yang dianalisis untuk memahami persepsi pengguna.	14
Stemming	Proses mengubah kata ke bentuk dasarnya untuk menyederhanakan variasi kata.	65
Stopwords	Kata-kata umum yang diabaikan dalam analisis teks karena tidak membawa makna penting (contoh: "dan", "yang", "atau").	62
Support Vector Machine (SVM)	Algoritma klasifikasi dalam pembelajaran mesin yang mencari <i>hyperplane</i> optimal untuk memisahkan kelas data.	25
TF-IDF	Teknik representasi teks yang memberi bobot pada kata berdasarkan frekuensinya dalam dokumen dan seluruh korpus.	14
TikTok	Platform media sosial berbasis video pendek yang populer di kalangan Generasi Z, sering digunakan untuk berbagi pengalaman pribadi termasuk kesehatan mental.	10

Istilah	Deskripsi	Halaman Pertama Kali Digunakan
Tokenisasi	Proses pemecahan teks menjadi unit-unit kecil (<i>token</i>) seperti kata atau frasa untuk keperluan analisis teks.	50
<i>Web Scraping</i>	Teknik pengambilan data secara otomatis dari situs web, dalam penelitian ini digunakan untuk mengumpulkan komentar dari TikTok.	42
<i>Word Cloud</i>	Representasi visual kata-kata yang sering muncul dalam data teks, di mana ukuran kata menunjukkan frekuensi kemunculannya.	84

BAB I PENDAHULUAN

I.1 Latar Belakang

Kesehatan mental telah menjadi salah satu isu krusial dalam dekade terakhir karena well-being psikologis terbukti berkorelasi positif dengan produktivitas, kualitas hidup, dan pertumbuhan ekonomi (Atsqualani dkk., 2022). Namun, survei tradisional kerap terhambat bias respons, biaya tinggi, serta jeda waktu publikasi, sehingga kurang adaptif untuk memantau dinamika opini daring (Nanda dkk., 2022). Analisis perspektif berbasis Support Vector Machine (SVM) dengan skema Term Frequency-Inverse Document Frequency (TF-IDF) mampu mengotomatisasi klasifikasi teks tidak terstruktur dan mencapai akurasi 96,48% pada korpus bahasa Indonesia (Arifin dkk., 2021). Dengan demikian, SVM-TF-IDF menawarkan solusi lebih efisien dan reliabel dibandingkan metode survei konvensional dalam memetakan kesadaran kesehatan mental masyarakat digital (Armanda & Tobing, 2024).

Pemetaan tingkat *awareness* kesehatan mental penting karena pengetahuan memadai mendorong individu mengenali gejala gangguan lebih dini dan proaktif mencari bantuan profesional. Survei Indonesia National Adolescent Mental Health Survey (I-NAMHS) 2021 melaporkan 34,9% remaja mengalami masalah mental dalam 12 bulan terakhir, tetapi hanya 2,6% yang mengakses layanan resmi (Wahdi dkk., 2023). Disparitas ini menegaskan rendahnya *mental-health literacy* serta kuatnya stigma yang menghalangi perilaku mencari pertolongan. Oleh karena itu, peningkatan *awareness* menjadi prasyarat strategis untuk menekan beban penyakit dan biaya sosial-ekonomi di Indonesia (Sulistiowati, dkk., 2020).

Secara global, meta-analisis menunjukkan prevalensi depresi remaja naik 28,0% dan kecemasan 26,9% sepanjang pandemi COVID-19 (Atsqualani dkk., 2022a). Lonjakan ini memperlihatkan bahwa krisis kesehatan publik dapat memperparah kerentanan psikologis kelompok muda, terutama di negara berpendapatan menengah (Bilali dkk., 2025). *Awareness* tinggi terbukti menurunkan stigma dan mempercepat penanganan, sehingga menekan keparahan gangguan mental (Yani dkk., 2025). Kondisi tersebut menegaskan urgensi riset