

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 (Atsqalani 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 (Atsqalani 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

yang memotret kesadaran masyarakat secara real-time untuk mendukung kebijakan kesehatan mental preventif (Pandia dkk., 2021).

Hubungan sebab-akibat antara rendahnya *awareness* dan derajat keparahan gangguan mental semakin jelas: minim literasi mengakibatkan penundaan pencarian bantuan, stigma internal, dan prognosis yang lebih buruk. Sebaliknya, kampanye literasi yang tepat sasaran memicu deteksi dini, meningkatkan kepercayaan pada layanan profesional, dan menurunkan beban ekonomi keluarga (Arifin dkk., 2021). Intervensi berbasis data dapat mengidentifikasi populasi berisiko tinggi sehingga sumber daya kesehatan dialokasikan lebih efektif (Wahdi dkk., 2023). Karenanya, penelitian ini fokus mengukur kesadaran sebagai variabel penentu jalur intervensi kesehatan mental di Indonesia (Sulistiowati dkk., 2020).

Media sosial dipilih sebagai sumber data karena menyediakan aliran teks real-time yang memantulkan persepsi publik tanpa filter institusional (Ardiansyah dkk., 2025). Karakter big-data-nya memungkinkan penerapan proses *Knowledge Discovery in Databases* (KDD) untuk mengekstraksi pola sentimen pada skala populasi secara berkelanjutan (Rakajati & Hidayat, 2024). Selain itu, interaktivitas platform menghadirkan konteks emosional yang kaya, memperkuat validitas analisis perspektif (Andriani dkk., 2024). Dengan demikian, media sosial menjadi laboratorium natural untuk mempelajari kesadaran kesehatan mental masyarakat digital Indonesia (Bur R dkk., 2023).

TikTok dipilih karena 60% penggunanya berasal dari Generasi Z dan rata-rata waktu penggunaan mencapai 31,5 jam per bulan di Indonesia (IDN Media, 2024). Format video pendek dan algoritma personalisasi mendorong pengguna mengekspresikan pengalaman kesehatan mental secara autentik (Krestina Huwae & Azis, 2024). Data.AI melaporkan 64,8 miliar jam dihabiskan warga Indonesia di TikTok sepanjang 2023, melampaui platform lain dalam hal keterlibatan. Volume dan kedalaman interaksi tersebut menjadikan TikTok sumber data ideal untuk mengukur kesadaran kesehatan mental Gen Z .

Generasi Z merupakan digital native yang pola perilakunya sangat dipengaruhi tren media sosial (Andriani dkk., 2024). Studi kualitatif menunjukkan Gen Z Indonesia lebih terbuka membahas depresi, kecemasan, dan coping di TikTok

dibandingkan platform lain (Kustiawan dkk., 2025). Keterbukaan ini menghasilkan korpus komentar yang kaya nuansa kesadaran, dukungan sebaya, tetapi juga potensi misinformation (Krestina Huwae & Azis, 2024). Oleh sebab itu, analisis perspektif pada konten TikTok relevan untuk menangkap realitas psikososial Gen Z Indonesia secara akurat.

Pandemi COVID-19 memperburuk kesehatan mental Gen Z melalui isolasi sosial, pembelajaran daring, dan ketidakpastian ekonomi (Pandia dkk., 2021). Penelitian di Indonesia mengonfirmasi peningkatan kecemasan tertinggi pada remaja sosio-ekonomi rendah, diperparah oleh paparan hoaks kesehatan di media sosial (Krestina Huwae & Azis, 2024). TikTok, meski berfungsi sebagai kanal dukungan, juga menyebarkan rumor yang menambah beban psikologis. Fokus penelitian pada dampak pandemi memungkinkan perumusan intervensi berbasis bukti untuk subkelompok paling rentan.

Penelitian ini menyempitkan ruang lingkup pada komentar TikTok berbahasa Indonesia guna mengevaluasi pola sentimen *aware* versus *not aware* terkait kesehatan mental. Pendekatan ini memungkinkan identifikasi isu spesifik—misalnya stigma terhadap terapi, pencarian self-help, atau normalisasi gangguan—yang mempengaruhi keputusan bantuan Gen Z. Hasil klasifikasi dapat dijadikan dasar merancang kampanye digital bertarget tinggi presisi (Armanda & Tobing, 2024). Dengan demikian, intervensi berbasis bukti dapat ditembakkan langsung pada titik-titik percakapan daring yang paling berpengaruh.

Metodologi KDD dimulai dari seleksi data komentar relevan, lalu praproses teks—case-folding, tokenisasi, stop-word removal, normalisasi—untuk menghasilkan korpus bersih (Rakajati & Hidayat, 2024). Tahap transformasi mengonversi teks menjadi representasi vektor menggunakan TF-IDF sehingga siap ditambang algoritma SVM dan LDA (Arifin dkk., 2021). Tahap penambangan data memproduksi model klasifikasi *awareness* serta topik laten yang mendasari perspektif pengguna (Arifin dkk., 2021). Terakhir, evaluasi akurasi dan coherence memastikan pola yang ditemukan valid dan bermanfaat bagi penyusunan strategi intervensi.

SVM dipilih karena kemampuannya memisahkan data di ruang berdimensi tinggi secara optimal dengan kompleksitas komputasi relatif rendah pada korpus menengah (Armanda & Tobing, 2024). Pengujian *k*-fold cross-validation ($k=10$) dalam penelitian terkait menunjukkan akurasi 96,48% untuk klasifikasi teks Indonesia menggunakan SVM-TF-IDF (Arifin dkk., 2021). Skema TF-IDF menekankan kata kunci diskriminatif dengan mempertimbangkan frekuensi lokal dan global, sehingga meningkatkan presisi hyperplane (Albin Pranata dkk., 2024). Kombinasi ini terbukti unggul dibandingkan *Naïve Bayes* atau jaringan saraf pada data singkat dan sarat variasi bahasa media sosial (Riadi dkk., 2023).

Latent Dirichlet Allocation (LDA) melengkapi analisis klasifikasi dengan mengungkap topik-topik laten yang dominan dalam diskusi kesehatan mental (Arifin dkk., 2021). Implementasi LDA pada 3.068 ulasan aplikasi Riliv berhasil mengidentifikasi empat klaster utama: *access support, counseling services, meditation features*, dan *user interface*. Dalam konteks TikTok, LDA diharapkan mengekstrak tema seperti self-diagnosis, stigma, atau tips coping yang berpotensi menjadi indikator kesadaran kolektif (Riadi dkk., 2023). Temuan topik dapat dipadukan dengan hasil SVM untuk memberikan gambaran komprehensif perspektif Gen Z (Arifin dkk., 2021).

Keberhasilan klasifikasi sangat bergantung pada praproses; pustaka Sastrawi menawarkan stemming bahasa Indonesia yang efektif mengurangi kata berimbuhan menjadi bentuk dasar (Junifer Pangaribuan & Putra Barus, 2023). Kombinasi Sastrawi dengan stop-word list terkurasi meningkatkan *term discrimination* pada fitur TF-IDF sehingga mendongkrak akurasi model hingga 5% (Prasetyo & Fahrurozi, 2023). Teknik normalisasi slang dan emotikon juga dibutuhkan karena komentar TikTok sarat ekspresi informal (Andriani dkk., 2024). Dengan pipeline praproses komprehensif, noise linguistik dapat diminimalkan dan kualitas pola yang diekstraksi meningkat signifikan.

Berdasarkan uraian tersebut, riset ini mengintegrasikan tahapan KDD, SVM-TF-IDF, LDA, dan praproses Sastrawi untuk menyusun kerangka analisis kesadaran kesehatan mental Gen Z di TikTok (Prasetyo & Fahrurozi, 2023). Pendekatan terintegrasi ini diharapkan menghasilkan deteksi cepat, hemat biaya, dan akurat

atas perbedaan perspektif *aware vs not aware* di level komentar pengguna (Arifin dkk., 2021). Output model dapat menjadi dasar merancang kampanye digital serta kebijakan intervensi yang lebih presisi dan inklusif bagi remaja Indonesia . Pada akhirnya, penelitian ini berkontribusi pada literatur *digital mental-health analytics* dan mendukung transformasi kesehatan mental berbasis data di Indonesia.

I.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, penelitian ini diarahkan untuk menjawab permasalahan inti berikut:

1. Seberapa andal model klasifikasi berbasis SVM-TF-IDF dalam membedakan perspektif 'aware' versus 'not aware' terkait kesehatan mental pada komentar TikTok Gen Z Indonesia?
2. Bagaimana pola sentimen dan topik utama yang mendasari tingkat kesadaran kesehatan mental di kalangan pengguna TikTok Generasi Z?

Rumusan masalah tersebut menitikberatkan pada keandalan model, karakter pola perspektif, dan aspek linguistik yang menjadi tantangan teknis dalam pengklasifikasian kesadaran kesehatan mental Gen Z di TikTok.

I.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menguji daya klasifikasi model SVM-TF-IDF dalam membedakan komentar TikTok Gen Z berdasarkan tingkat kesadaran terhadap isu kesehatan mental.
2. Mengidentifikasi pola utama ekspresi kesadaran ('aware') dan ketidaksadaran ('not aware') terkait kesehatan mental pada komentar TikTok berbahasa Indonesia.

I.4 Manfaat Penelitian

Manfaat dari penelitian ini difokuskan secara spesifik sebagai berikut:

1. **Manfaat Akademis**
Memberikan contoh implementasi machine learning dan text mining pada korpus bahasa Indonesia di TikTok, khususnya analisis klasifikasi '*aware*' vs '*not aware*' untuk isu kesehatan mental.
2. **Manfaat Praktisi**
Memberikan wawasan bagi pengelola kampanye kesehatan mental digital terkait kecenderungan sentimen dan pola ekspresi Gen Z di TikTok agar intervensi komunikasi dapat lebih tertarget dan berbasis data aktual.
3. **Manfaat Metodologis**
Menyediakan panduan proses praproses hingga klasifikasi teks berbahasa Indonesia pada media sosial, yang dapat dijadikan referensi untuk analisis literasi kesehatan mental di platform lain dengan pendekatan serupa.

I.5 Batasan Penelitian

Agar permasalahan dapat mencapai tujuan yang telah ditentukan, maka dibuat batasan-batasan sebagai berikut:

1. Identifikasi analisis perspektif yang digunakan pada penelitian ini terbatas pada klasifikasi biner *aware* dan *not aware* terhadap isu kesehatan mental.
2. Data yang dianalisis merupakan hasil dari *web scraping* menggunakan *platform* Apify dari *platform* TikTok dengan fokus pada komentar video yang mengandung tagar relevan dengan kesehatan mental.
3. Dataset berupa kumpulan komentar yang diperoleh dari video-video yang berhubungan dengan kesehatan mental, khususnya yang menggunakan tagar seperti #mentalhealth, #kesehatanmental, #selfcare, dan #generasiz.
4. Pemodelan topik dilakukan dengan metode *Latent Dirichlet Allocation* (LDA) dan klasifikasi perspektif menggunakan algoritma *Support Vector Machine* (SVM) dengan ekstraksi fitur TF-IDF.
5. Proses *preprocessing* data meliputi pembersihan teks, normalisasi, *lower casing*, penghapusan *stopwords*, dan *stemming* menggunakan pustaka Sastrawi untuk bahasa Indonesia.

I.6 Sistematika Penulisan

Laporan ini disusun secara sistematis untuk memudahkan pembaca memahami setiap tahapan penelitian yang dilakukan. Berikut adalah sistematika penulisan laporan ini:

1. Bab I – Pendahuluan

Bab ini membahas latar belakang penelitian, perumusan masalah, tujuan penelitian, batasan penelitian, dan manfaat penelitian . Bagian ini menjelaskan alasan pemilihan topik serta pentingnya penelitian terhadap isu kesehatan mental di kalangan generasi Z melalui platform media sosial TikTok .

2. Bab II – Tinjauan Pustaka

Bab ini berisi kajian teori yang mendasari penelitian, meliputi konsep kesehatan mental, media sosial TikTok, analisis perspektif kesadaran, machine learning, Natural Language Processing (NLP), Knowledge Discovery in Databases (KDD), Support Vector Machine (SVM), Term Frequency-Inverse Document Frequency (TF-IDF), Latent Dirichlet Allocation (LDA), topic modelling, dan metode evaluasi menggunakan Confusion Matrix serta Receiver Operating Characteristic (ROC) . Selain itu, dibahas juga alasan pemilihan metode SVM dan LDA, serta state of the art terkait penelitian yang relevan .

3. Bab III – Metodologi Penelitian

Bab ini menjelaskan metode yang digunakan dalam penelitian, termasuk pengembangan model konseptual, sistematika penyelesaian masalah menggunakan kerangka kerja Knowledge Discovery in Databases (KDD), alasan pemilihan metode SVM dan LDA, pengumpulan data menggunakan web scraping, pengolahan data, dan metode evaluasi model . Proses preprocessing teks bahasa Indonesia menggunakan pustaka Sastrawi, representasi teks menggunakan TF-IDF dan CountVectorizer, serta penerapan model SVM untuk klasifikasi dan LDA untuk topic modelling dijelaskan secara rinci pada bagian ini .

4. Bab IV – Analisis dan Perancangan

Bab ini menjelaskan seluruh proses teknis yang dilakukan untuk membangun sistem klasifikasi komentar TikTok berdasarkan kesadaran terhadap isu kesehatan mental . Tahapan ini mencakup pengumpulan data menggunakan Apify TikTok Comment Scraper, preprocessing data teks (pembersihan data, lower casing, penghapusan stopwords, normalisasi, dan stemming menggunakan pustaka Sastrawi), pelabelan data menggunakan teknik keyword matching, tokenisasi dengan TF-IDF, serta implementasi model klasifikasi berbasis Support Vector Machine (SVM) dengan pembagian data rasio 80:20 . Selain itu, bab ini juga menjelaskan proses visualisasi data dan eksplorasi menggunakan topic modelling dengan metode Latent Dirichlet Allocation (LDA) untuk analisis deskriptif pola dan tema komentar berdasarkan label aware dan not aware .

5. Bab V – Hasil dan Pembahasan

Bab ini menyajikan hasil evaluasi dari model yang telah dibangun, meliputi hasil pelabelan data, visualisasi wordcloud, evaluasi skenario pengujian dengan berbagai konfigurasi hyperparameter SVM, dan hasil evaluasi menggunakan confusion matrix serta kurva Receiver Operating Characteristic (ROC) yang menunjukkan kemampuan model dalam membedakan antara komentar aware dan not aware . Selain itu, hasil dari topic modelling dengan pendekatan LDA dibahas secara mendalam, baik pada data aware maupun not aware, untuk mengidentifikasi tema-tema utama dalam wacana digital mengenai kesehatan mental melalui intertopic distance map, analisis top-30 most salient terms, dan visualisasi word cloud.

6. Bab VI – Kesimpulan dan Saran

Bab ini menyajikan kesimpulan dari penelitian yang telah dilakukan serta saran untuk pengembangan lebih lanjut . Kesimpulan mencakup ringkasan hasil analisis dan temuan utama dari implementasi SVM dan LDA dalam menganalisis perspektif kesehatan mental di media sosial TikTok, sementara saran memberikan rekomendasi untuk penelitian di masa depan terkait analisis perspektif kesehatan mental di media sosial .