

## ***Tweets Classification of Mental Health Disorder in Indonesia Using LDA and Cosine Similarity***

Klasifikasi *Tweets* Gangguan Kesehatan Mental di Indonesia Menggunakan LDA dan *Cosine Similarity*

**Irmma Dwijayanti<sup>1\*</sup>, Muhammad Habibi<sup>2</sup>, Kartikadyota Kusumaningtyas<sup>3</sup>, Sujono Riyadi<sup>4</sup>**

<sup>1</sup> Sistem Informasi, Universitas Jenderal Achmad Yani Yogyakarta, Indonesia

<sup>2,3</sup> Informatika, Universitas Jenderal Achmad Yani Yogyakarta, Indonesia

<sup>4</sup> Keperawatan, Universitas Jenderal Achmad Yani Yogyakarta, Indonesia

<sup>1\*</sup> irmmadwijayanti@gmail.com, <sup>2</sup> muhammadhabibi17@gmail.com,

<sup>3</sup> kartikadyota@gmail.com, <sup>4</sup> sujono@gmail.com

\*: *Penulis korespondensi (corresponding author)*

### ***Abstract***

#### ***Informasi Artikel***

*Received: September 2023*

*Revised: November 2023*

*Accepted: December 2023*

*Published: February 2024*

*Purpose: Twitter can be an effective platform for conveying important information about mental health disorders to the public and health organizations on a large scale. However, there has been no adequate evaluation of tweet data related to mental health disorders. This research aims to classify tweet data to determine the most common mental health disorders in Indonesia based on the symptoms experienced. Methodology: The classification process is carried out using cosine similarity calculations between tweets data and keywords which are compiled based on theoretical studies and optimization of the LDA topic modeling results.*

*Findings/result: The classification results show that the most discussed issues on Twitter are depression, bipolar, schizophrenia, dementia, and PTSD. Based on these results it can be interpreted that the level of prevalence and public attention to depressive disorders is quite high compared to other disorders. From the results of the classification, it is also possible to identify the most discussed symptoms through the emergence of keywords from each category.*

*Originality: Classification is calculated based on the cosine similarity between tweets and keywords compiled from human judgement and enriched using the results of LDA topic modeling to improve classification performance.*

*Keywords: classification; cosine similarity; mental health disorder; LDA*

Kata kunci: *cosine similarity, gangguan kesehatan mental, klasifikasi, LDA*

## Abstrak

**Tujuan:** Twitter dapat menjadi platform yang efektif untuk menyampaikan informasi penting tentang gangguan kesehatan mental kepada masyarakat dan organisasi kesehatan dalam skala besar. Namun, belum ada evaluasi yang memadai terhadap data *tweets* terkait gangguan kesehatan mental. Penelitian ini bertujuan untuk melakukan klasifikasi data *tweets* untuk mengetahui gangguan kesehatan mental yang paling banyak terjadi di Indonesia berdasarkan gejala yang dialami.

**Metode:** Proses klasifikasi dilakukan menggunakan perhitungan *cosine similarity* antara data *tweets* dengan kata kunci yang disusun berdasarkan kajian teori dan optimasi dari hasil pemodelan topik LDA.

**Hasil:** Hasil klasifikasi menunjukkan isu paling banyak dibahas di Twitter secara berturut-turut adalah depresi, bipolar, skizofrenia, demensia, dan PTSD. Berdasarkan hasil tersebut dapat diartikan bahwa tingkat prevalensi dan perhatian masyarakat terhadap gangguan depresi cukup tinggi daripada gangguan lainnya. Dari hasil klasifikasi juga dapat diketahui gejala yang banyak dibahas melalui kemunculan *keyword* dari masing-masing kategori.

**Keaslian:** Klasifikasi dihitung berdasarkan nilai similaritas *cosine similarity* antara data *tweets* dan kata kunci yang disusun dari penilaian manusia dan diperkaya menggunakan hasil pemodelan topik LDA untuk meningkatkan performa klasifikasi.

## 1. Pendahuluan

Menurut WHO, kesehatan mental merupakan keadaan dimana seseorang memiliki kesejahteraan mental sehingga dapat mengatasi tekanan hidup, mampu menyadari kemampuan mereka, belajar dan bekerja dengan baik, serta berkontribusi pada komunitas mereka [1]. Kesehatan mental berkaitan dengan emosional, psikologis, dan kesejahteraan sosial. Kesehatan mental berperan penting dalam menunjang produktivitas dan kualitas kesehatan fisik seseorang.

Gangguan kesehatan mental atau kejiwaan dapat dialami semua orang. Berdasarkan Data Riset Kesehatan Dasar (Riskesdas) 2018, menunjukkan penduduk usia diatas 15 tahun mengalami gangguan mental emosional sebanyak 19 juta penduduk, dan mengalami depresi sebanyak 12 juta penduduk [2]. Pada tahun 2022, hasil survei Populix menunjukkan sebanyak 52% responden merasa memiliki masalah kesehatan mental, terutama di kalangan 18-24 tahun [3]. Gejala kesehatan mental yang dialami diantaranya perubahan suasana hati (26%), perubahan kualitas tidur dan nafsu makan (19%), kecemasan (18%), kelelahan berlebihan (10%), serta merasa bingung, pelupa, pemarah (8%). Data tersebut menunjukkan bahwa Indonesia belum dapat menyelesaikan permasalahan kesehatan mental secara tepat. Segala upaya dan diskusi keilmuan telah ditempuh berbagai organisasi untuk mengetahui penyebab gangguan mental guna

menurunkan jumlah penderita gangguan kesehatan mental ini. Pemahaman dan stigma negatif tentang gangguan mental serta kesadaran kesehatan mental yang rendah turut berperan dalam penyembuhan kesehatan mental. Media sosial sebagai saluran komunikasi yang dapat menjangkau populasi besar dengan cepat, murah, dan sedikit usaha sangat berpotensi untuk penyebaran literasi kesehatan mental [4], salah satunya adalah Twitter.

Twitter adalah platform interaksi dan komunikasi antar pribadi yang masif, populer, dan berskala kecil, dimana pengguna dapat saling bertukar pendapat tentang berbagai masalah seperti pemerintah, merek, kesehatan, kegiatan olahraga, dan sebagainya melalui suatu “*Tweet*”. Netray menemukan 14.8K total perbincangan pengguna Twitter terkait topik kesehatan mental. Jumlah tersebut semakin meningkat di masa pandemi Covid-19 hingga kini semakin banyak kasus bunuh diri disebabkan karena gangguan kesehatan mental. Konten-konten yang dihasilkan pengguna Twitter terkait kesehatan mental sangat berpotensi sebagai media untuk memberikan informasi penting bagi masyarakat dan organisasi kesehatan dalam skala besar. Namun, evaluasi data *tweet* terkait gangguan kesehatan mental belum ditemukan. Data *tweets* dapat dianalisis sehingga dapat mengetahui klasifikasi gangguan mental yang banyak terjadi di Indonesia berdasarkan gejala yang dialami.

Penelitian ini bertujuan untuk melakukan klasifikasi data *tweets* menjadi beberapa kategori gangguan kesehatan mental yang terjadi di Indonesia. Proses klasifikasi dilakukan menggunakan perhitungan similaritas antara data *tweets* dengan kata kunci yang disusun berdasarkan kajian teori dan optimasi dari hasil pemodelan topik. Hasil klasifikasi dapat digunakan untuk mengetahui gangguan kesehatan mental yang paling banyak terjadi di Indonesia berdasarkan gejala yang dialami. Dengan begitu, pihak berwenang dapat mengambil tindakan yang dibutuhkan untuk memutuskan kebijakan yang bertujuan untuk mengurangi jumlah penderita gangguan kesehatan mental.

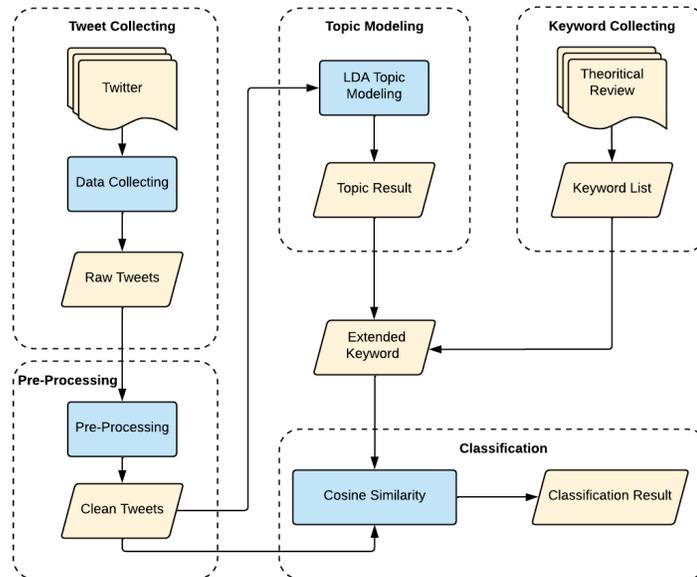
Penelitian terkait gangguan kesehatan mental yang menggunakan data dalam skala besar seperti Twitter atau media sosial lainnya sudah banyak dilakukan, seperti penelitian yang membangun model data analitik untuk mendeteksi tingkat kerentanan terhadap depresi dari pengguna media sosial [5]. Model *machine learning* yang dibangun menggunakan *Support Vector Machine* (SVM) dan *Naïve Bayes* dilatih untuk mengklasifikasikan ke dalam enam kriteria rentang depresi (dianggap normal, ringan, sedang, borderline, parah, dan ekstrim). Data media sosial yang dikumpulkan dianalisis dengan model yang dibangun dan memberikan label kepada pengguna sebagai depresi atau non-depresi. Pada tahun berikutnya, mendiagnosa depresi melalui media sosial juga dilakukan pada penelitian [6]. Penelitian lain menemukan bahwa penggunaan pemodelan topik dan analisis sentimen dalam mengelompokkan data tekstual, menemukan pengetahuan baru, dan mencapai wawasan dari sejumlah besar konten tekstual [7]. Seperti yang dilakukan pada penelitian [8] memodelkan topik yang berkaitan dengan GERMAS dari *caption* Instagram dengan tujuan mensukseskan program GERMAS di masyarakat.

Klasifikasi adalah salah satu teknik dalam penambangan data untuk mendapatkan informasi penting dan relevan tentang data dan metadata. Seperti penelitian [9] untuk mengetahui penggunaan media sosial oleh tenaga kesehatan sebagai media konsultasi dan informasi yang tidak bertentangan dengan etika profesionalisme. Penelitian dilakukan dengan dua tahapan yaitu membedakan kelas data kesehatan dengan yang bukan bidang kesehatan, dan berfokus pada bidang penelitian keparakan dalam kesehatan.

Pada tahun 2022, penelitian [10] mengusulkan suatu sistem prediksi kesehatan mental, dimana model yang dikembangkan dilatih menggunakan kumpulan data *Tweet*. Penelitian menggunakan LSTM dan NLP untuk mengetahui emosi dan perasaan siswa terutama ketika masa pandemi Covid-19. Penelitian lainnya [11] menggunakan pemodelan topik dan analisis sentimen untuk memperkirakan persepsi publik terkait masalah kesehatan mental, khususnya pada data Twitter. Data tersebut diidentifikasi dengan analisis sentimen melalui polaritas dan subjektivitas, sehingga dapat mengukur persepsi publik tentang kesehatan mental dengan mempertimbangkan keseluruhan emosi serta sejauh mana fakta atau opini. Sedangkan pemodelan topik dilakukan untuk mengidentifikasi topik utama yang paling banyak dibahas selama tujuh tahun terakhir. Peningkatan ini dilakukan untuk mendapatkan gambaran tentang tingkat kepentingan relatifnya sesuai pandangan orang tentang kesehatan mental setiap tahunnya.

## 2. Metode

Penelitian ini mencakup beberapa tahapan penelitian, diantaranya pengumpulan data *tweet* dari Twitter dan *keyword list* dari kajian teori, kemudian dilanjutkan *preprocessing* untuk data *tweet*. Tahapan berikutnya adalah melakukan pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA) untuk mendapatkan kata-kata penyusun topik-topik yang paling banyak dibahas yang selanjutnya digunakan untuk memperkaya *keyword list*. Tahapan terakhir melakukan perhitungan similaritas antara data *tweet* yang sudah dibersihkan dan perluasan kata kunci untuk mendapatkan klasifikasi jenis-jenis gangguan kesehatan mental. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

### 2.1. Pengumpulan Data

Tahap pengumpulan data terdiri dari dua langkah yaitu pengumpulan data *tweet* dan data *keyword list*.

### 2.1.1. Data Tweet

Data penelitian ini dikumpulkan dari hasil ekstraksi data *tweet* di Twitter menggunakan metode *web scraping* yaitu proses mengekstrak informasi dari situs web yang diproses menjadi struktur sederhana seperti .xlsx atau .csv [12]. Data yang diambil merupakan *tweet* atau *re-tweet* dalam Bahasa Indonesia menggunakan kata kunci “gangguan kesehatan mental”. Data yang digunakan dalam penelitian ini dikumpulkan dari tanggal 1 Januari 2021 sampai 31 Maret 2023 dengan perolehan data sebanyak 4.336 *tweet*.

### 2.1.2. Keyword List

Selain data *tweets*, penelitian ini juga mengumpulkan data *keyword* yang disusun berdasarkan gejala-gejala pada setiap kategori gangguan kesehatan mental. Gangguan kesehatan mental dapat diklasifikasikan menjadi lima kategori meliputi depresi, gangguan bipolar, skizofrenia, demensia, dan PTSD [13]. *Keyword* disusun berdasarkan gejala-gejala yang diperoleh dari teori dan studi literatur yang dikumpulkan dalam Bahasa Indonesia sesuai dengan bahasa yang digunakan data *tweets*. Hasil *keyword* yang disusun tersebut digunakan untuk proses klasifikasi menggunakan perbandingan similaritas. *Keyword list* untuk setiap kategori dapat dilihat pada Tabel 1.

Tabel 1. *Keyword list*

No.	Kategori	Keyword	Sumber
1.	Depresi	sedih, putus asa, pesimis, salah, konsentrasi, sulit, ingat, pikir, bunuh, diri, coba, hilang, trauma, stres, kosong, senang, singgung, hampa, mati, berat_badan, lelah, menangis, khawatir, acuh	[13], [14]
2.	Bipolar	mood, lebih, aktifitas, tingkat, tidur, turun, marah, sembron, distraksi, impulsif, suasana_hati, perilaku, ekstrem, kehidupan_sosial, gangguan_tidur, cemas, hiperaktif	[13], [15]
3.	Skizofrenia	persepsi, bahasa, emosi, perilaku, halusinasi, lihat, dengar, delusi, abnormal, aneh, liar, tertawa	[13], [16]
4.	Demensia	tua, lansia, memori, belajar, bahasa, keputusan, kendali, emosi, sosial, bingung, cemas	[13], [17]
5.	PTSD	trauma, ingat, mimpi_buruk, kejadian, peristiwa, ancaman, gangguan_jiwa, intuisi, hindar, rasa_bersalah, takut, ngeri, marah, malu, asing, waspada	[13], [18]

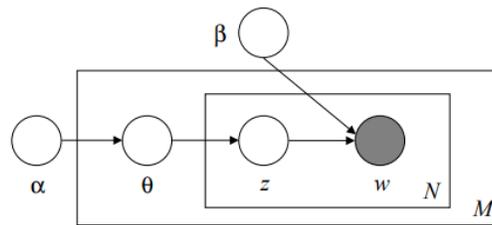
## 2.2. Preprocessing Data

Tahap berikutnya adalah melakukan *preprocessing* data. *Preprocessing* merupakan tahap awal dalam proses *text mining* yang bertujuan untuk membersihkan teks dari unsur-unsur yang tidak relevan seperti data tidak lengkap, *noise*, dan tidak konsisten [19]. Proses pada *preprocessing* dibagi menjadi ke dalam beberapa tahapan antara lain, *case folding* digunakan untuk mengubah teks ke dalam bentuk standar atau *lowercase*, *stopword removal* untuk menghilangkan kata hubung dan kata ganti, *tokenizing* digunakan proses pemotongan *string* tiap kata yang menyusunnya, dan tahap terakhir adalah *stemming* digunakan untuk proses mengubah kata ke dalam bentuk dasarnya.

## 2.3. Topic Modeling

Data *tweets* yang sudah dibersihkan melalui *preprocessing*, kemudian diolah pada tahap *topic modeling*. *Topic modeling* adalah metode pemodelan teks untuk mengeksplorasi kumpulan

dokumen dan merepresentasikannya dalam kumpulan *term* yang sering muncul bersamaan yang disebut sebagai topik [20]. Dalam penelitian ini menggunakan pemodelan topik LDA untuk menemukan *term* atau kata kunci yang paling banyak dibicarakan. Kata-kata tersebut digunakan untuk memperluas *keyword list* agar meningkatkan kinerja klasifikasi. *Topic modeling* merupakan pendekatan *unsupervised learning* yang tidak membutuhkan pelabelan dalam proses pembelajarannya. Dalam proses eksplorasi topik-topik dan identifikasi hubungan antar topik dilakukan berdasarkan penilaian manusia. Representasi grafis LDA dapat dilihat pada Gambar 2. *Plates* yang direpresentasikan dengan kotak-kotak mewakili replikasi, *plates* luar mewakili dokumen, sedangkan *plates* dalam mewakili pilihan topik dan kata yang berulang dalam dokumen.

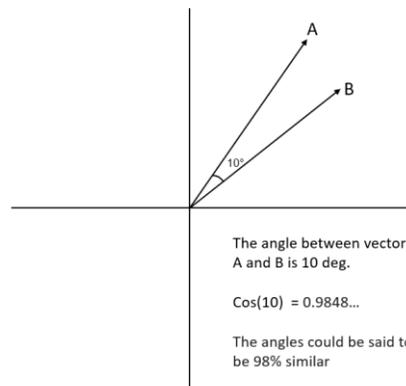


Gambar 2. Representasi model grafis LDA [21]

LDA memiliki tiga tingkat representasi, dimana parameter  $\alpha$  dan  $\beta$  merupakan parameter *corpus* atau kumpulan kata yang diasumsikan sampelnya diambil sekali dalam proses menghasilkan *corpus*. Kemudian variabel  $\theta_d$  merupakan variabel tingkatan dokumen yang diambil sampelnya satu kali setiap dokumen. Sedangkan variabel  $z_{dn}$  dan  $w_{dn}$  adalah variabel kata yang diambil sampelnya satu kali setiap kata pada setiap dokumen.

#### 2.4. Klasifikasi

Tahap berikutnya adalah klasifikasi data *tweets* ke dalam kategori gangguan kesehatan mental. Proses klasifikasi dilakukan dengan menghitung nilai similaritas antara *tweets* dengan *keyword list* menggunakan algoritma *cosine similarity*. *Cosine similarity* (*cosim*) merupakan pengukuran derajat similaritas berdasarkan korelasi dari dua teks melalui kosinus sudut antara dua vektor teks [22]. Ilustrasi derajat kosinus pada *cosim* dapat dilihat pada Gambar 3.



Gambar 3. Ilustrasi derajat *cosim* [23]

Pengukuran kesamaan dokumen dilakukan berdasarkan kata/istilah (*term*), yang direpresentasikan sebagai vektor. Sederhananya, untuk menemukan kesamaan antar teks yang ada dalam *keyword list* (A) dan data *tweet* (B), dapat dilakukan dengan membandingkan nilai kosinus sudut dari perkalian bobot kedua dokumen A dan B, dengan panjang vektor keduanya. Nilai cosin dibatasi pada rentang 0 dan 1, dimana dua vektor dinilai mirip apabila nilai sudut kosinusnya  $0^0 = 1$ . Semakin dekat jarak sudut kosinus kedua vektor maka semakin tinggi nilai kemiripannya. Penghitungan *cosine similarity* [24] dapat dilihat pada persamaan (1).

$$similarity = \cos(\theta) = \frac{A \cdot B}{|A| \cdot |B|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2 \sum_{i=1}^n B_i^2}} \quad (1)$$

Keterangan:

A = vektor (*keyword list*)

B = vektor 2 (*data tweet*)

$A_i$  = bobot *term* i dalam blok A<sub>i</sub>

$B_i$  = bobot *term* i dalam blok B

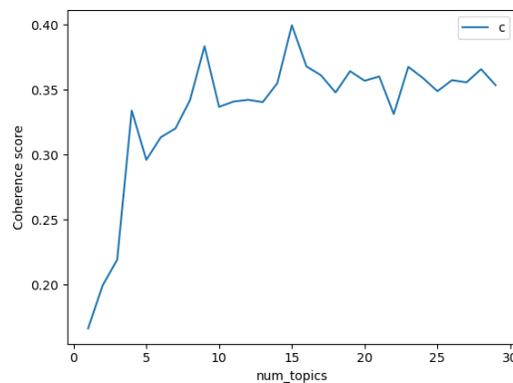
i = jumlah *term* dalam kalimat

n = jumlah vektor

### 3. Hasil dan Pembahasan

#### 3.1. Topic Modeling

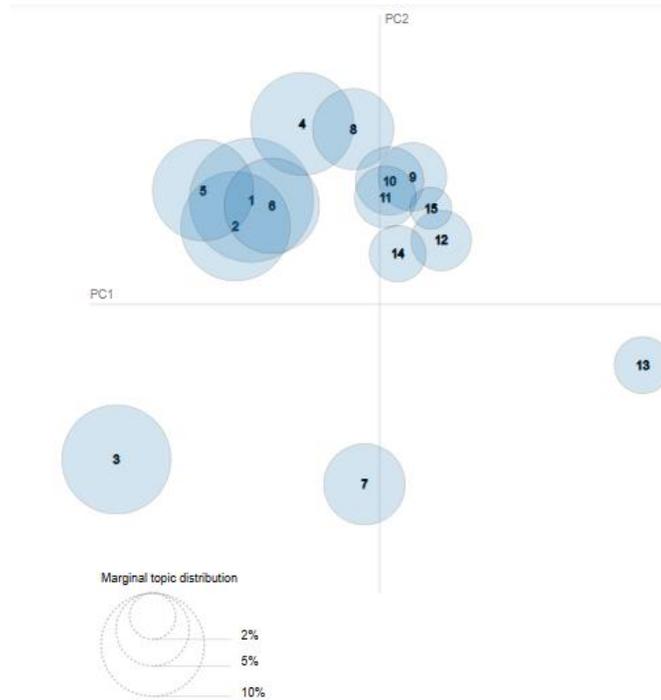
Pada tahap pemodelan topik LDA membutuhkan masukan berupa *dictionary*, *corpus*, dan jumlah topik. Penentuan jumlah topik pada penelitian ini menggunakan visualisasi pada grafik *coherence score*, yang ditunjukkan pada Gambar 4. *Coherence score* merupakan ukuran yang digunakan untuk mengevaluasi pemodelan topik, dimana semakin besar nilai skor koherensinya maka semakin baik hasil interpretasi pemodelan topik yang dihasilkan [25].



Gambar 4. Grafik hasil *coherence score*

Berdasarkan Gambar 4, menunjukkan bahwa topik 15 memiliki skor koheren tertinggi yaitu sebesar 0,399702 sehingga diambil menjadi jumlah topik terbaik untuk pemodelan topik. Gambar 5 merupakan visualisasi peta jarak antar topik yang dihasilkan dari 15 topik. Lingkaran biru merupakan representasi dari topik untuk menggambarkan *distance map*. Semakin dekat jarak antar topik berarti memiliki kata kunci penyusun yang sama sehingga memungkinkan

dapat dikategorikan dalam topik yang sama. Sebaliknya, jika jarak antar lingkaran semakin jauh maka kata kunci penyusunnya pasti sebagian besar berbeda.



**Gambar 5.** Peta jarak antar topik diproduksi untuk 15 topik

Tabel 2 menunjukkan hasil pemodelan 15 topik dengan masing-masing kata penyusun. Dalam tabel disajikan 30 kata kunci penyusun teratas dari setiap topik yang menggambarkan kata-kata yang paling banyak muncul atau sering digunakan.

**Tabel 2.** Hasil pemodelan 15 topik

Topik Ke-	Kata Kunci Penyusun
0	ganggu mental sehat alami cemas tanda takut bantu asa cenderung jiwa suhu gue disorder salah fisik sakit ga aktivitas sehari-hari depresi stres makan serang bangun psikiater anxiety khawatir pengaruh emosi
1	mental sehat tahu ganggu bpa bantu literasi tekanan ga manajemen dlm informasi jiwa yakin prevensi juta kaffah indikasi teliti generasi indonesia sekuler definisi millenials rekognisi islam hidup kapitalisme sadar remaja
2	mental sehat ganggu lapas kemenkumham perempuan alami depresi cemas atas cegah usia bina jiwa warga terapi candu jabar tangan palembang sumsel amp sebab judi online laksana aman kondusif dwi implementasi
3	mental sehat ganggu jiwa kondisi pandemi jaga derita darurat sadar sakit kuat naik remaja masyarakat tingkat dasar parah turun catat perilaku pengaruh cegah asa disorder rapuh suasana hati pikir indikasi
4	mental sehat ganggu alami jiwa warga bina rentan Indonesia pengaruh kemenkumham bikin lem hidup sakit ga tidur lu film sebab pasca kdrt rajin ngakak upload keluh baby kalo minimal raga
5	sehat mental ganggu idap puluh introvert balita mati stunting hamil ribu politik ideal rentan depresi alami ocd urban bogor mual tidur literasi cemas seru sakit asa jiwa kondisi berita oversharing

6	mental ganggu sehat hidup identitas gender nama buku cobain tulis dunia diagnosis picu tolak duga pria tumpu ujung jelas perspektif valid association psychological transgender organisasi daftar who nyata keluar hapus
7	mental ganggu sehat depresi sosial media cemas alami faktor stres gugat kondisi lengkap simak ancam solusi hidup tidur jiwa kenal muda terus mahasiswa fisik bantu pribadi illness sedih rentan derita
8	mental sehat ganggu remaja dampak milik alami simak derita transisi depresi lengkap cemas obat psikolog kondisi hamil janin rasa upaya tangan fisik psikiater keluarga lahir video jaga trauma gejala seksual
9	sehat ganggu alami gejala fisik stress deteksi sakit depresi faktor kerja keluarga diagnosis muncul ciri salah giat lindung muda apresiasi remaja rawat tanggung dasar hidup timbul tekan milik stres sosial
<b>Topik Ke-</b>	<b>Kata Kunci Penyusun</b>
10	mental sehat ganggu gejala alami anak bipolar kondisi tidur masyarakat atas milik remaja kenal dewasa buruk butuh tingkat faktor pikir depresi hidup risiko derita cerita konseling potensi makan diri insomnia
11	ganggu depresi cari malu milik stigma jiwa hilang asa sedih pasang aktivitas bunuh kualitas hidup turun sakit heroin tingkat fisik dasar kurang gila parah percaya pakar tidur rawat pergi bicara
12	jiwa ganggu sehat mental bebas tubuh fisik lengkap buka pikir cacat harap kuat kerja cepet stres tingkat psikologi bikin pola hidup komunikasi teliti takut bukti tuntutan temu asa sakit paham
13	ganggu mental sehat jiwa odgj pasung pasang sosial obat baik coba menteri kualitas sakit tidur cemas psikologi bebas stress milik depresi laku dekat daerah butuh usut insecure gelap ikat ruang
14	mental ganggu sehat remaja idap alami tidur tanda sakit rugi ekonomi angka tara anxiety attack duga stunting balita kelola sistem cegah dampak stres tingkat jiwa bukti buruk jauh mungkar tekan

Kata-kata penyusun setiap topik yang ditampilkan pada Tabel 2 dipetakan ke dalam lima kategori gangguan kesehatan mental untuk memperkaya *keyword list* yang telah disusun sebelumnya. Proses pemetaan kata melibatkan identifikasi kata-kata yang signifikan dan sering muncul yang belum dimasukkan ke dalam daftar kata kunci sebelumnya. Setelah itu, kata-kata tersebut akan ditetapkan ke dalam kategori gangguan kesehatan mental yang sesuai. Penetapan kategori ini didasarkan pada penilaian yang dilakukan para ahli. Tambahan kata kunci yang diperoleh digunakan untuk perluasan kata kunci sehingga dapat meningkatkan performa klasifikasi sehingga *keyword list* tidak hanya disusun berdasarkan kajian teori tetapi juga justifikasi ahli. Daftar perluasan kata kunci dapat dilihat pada Tabel 3.

**Tabel 3.** *Extended keyword*

No.	Kategori	Kata Kunci Tambahan	Extended Keyword
1.	Depresi	hidup, makan, insomnia, ego, kecewa, semangat, remaja	sedih, putus, asa, pesimis, salah, konsentrasi, sulit, ingat, pikir, bunuh, diri, coba, hilang, trauma, stres, kosong, senang, singgung, hampa, mati, berat_badan, lelah, menangis, khawatir, acuh, hidup, makan, insomnia, ego, kecewa, semangat, remaja
2.	Gangguan Bipolar	hiperaktif, insomnia, gembira, simpati, remaja	mood, lebih, aktifitas, tingkat, tidur, turun, marah, sembrono, distraksi, impulsif, suasana_hati, perilaku, ekstrem, kehidupan_sosial, gangguan_tidur, cemas, hiperaktif, singgung, insomnia, gembira, sedih, simpati, remaja

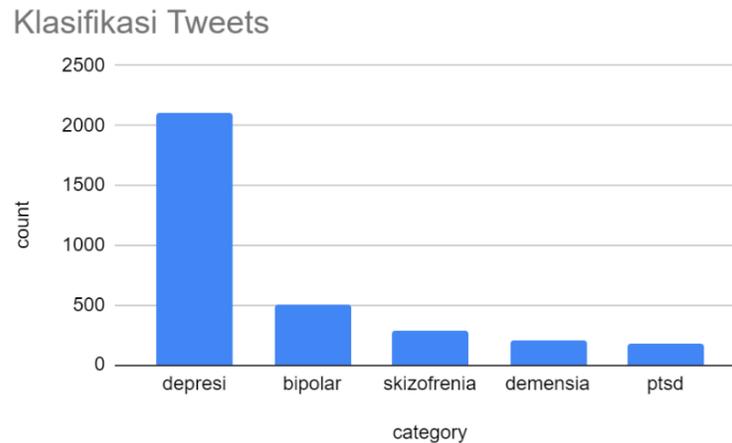
---

3.	Skizofrenia	nafsu_makan, insomnia, percaya, paranoid, curiga	persepsi, bahasa, emosi, perilaku, halusinasi, lihat, dengar, delusi, abnormal, aneh, liar, tertawa, nafsu_makan, insomnia, percaya, paranoid, curiga
4.	Demensia	kepribadian, kenal, ingat, gelisah, fokus, sadar	tua, lansia, memori, belajar, bahasa, keputusan, kendali, emosi, sosial, bingung, cemas
5.	PTSD	kejut, pendam, panik, insomnia, percaya, leceh	trauma, ingat, mimpi_buruk, kejadian, peristiwa, ancaman, gangguan_jiwa, intuisi, hindar, rasa_bersalah, takut, ngeri, marah, malu, asing, waspada, terkejut, pendam, panik, insomnia, percaya, leceh

---

### 3.2. Klasifikasi

Tahap berikutnya adalah melakukan klasifikasi data *tweets* ke dalam kategori gangguan kesehatan mental. Klasifikasi dilakukan dengan menghitung similaritas antara *tweets* dengan *keyword list* yang telah dikumpulkan menggunakan algoritma *cosine similarity*. Kriteria pengelompokan kategori adalah nilai tertinggi dari perhitungan similaritas. Metode *cosim* yang dibangun diuji terlebih dahulu menggunakan data uji yang sudah dilabeli secara manual ke dalam kategori-kategori gangguan kesehatan mental dan menghasilkan tingkat akurasi 94%. Hasil analisis jumlah *tweets* yang diklasifikasi berdasarkan kategori gangguan kesehatan mental dapat dilihat pada Gambar 6.



**Gambar 6.** Hasil jumlah klasifikasi *tweets*

Hasil klasifikasi *tweets* menunjukkan bahwa gangguan kesehatan mental kategori depresi menjadi isu yang paling banyak dibicarakan dalam media sosial Twitter yaitu sebanyak 2.908 *tweets*. Selain itu, gangguan lainnya jika diurutkan dari yang paling banyak yaitu bipolar sebanyak 500 *tweets*, skizofrenia 285 *tweets*, demensia 200 *tweets*, dan ptsd 171 *tweets*. Berdasarkan hasil tersebut dapat diartikan bahwa tingkat prevalensi dan perhatian terhadap gangguan kesehatan mental di masyarakat tentang depresi cukup tinggi dibandingkan dengan kategori lainnya. Hasil analisis *tweets* juga dapat digunakan untuk mengetahui gejala yang banyak dibahas melalui kata-kata yang digunakan masyarakat Indonesia mengekspresikan tentang gangguan kesehatan mental di Twitter. Gambar 7 menunjukkan visualisasi frekuensi kemunculan *keyword* yang merupakan representasi gejala dari lima kategori gangguan kesehatan mental.



**Gambar 7.** Visualisasi frekuensi gejala setiap kategori

Berdasarkan Gambar 7, dapat dilihat pada kategori depresi banyak menggunakan kata asa, hidup, stres, pikir, dan remaja untuk menceritakan tentang depresi. Kategori bipolar didominasi oleh kata kunci cemas, tingkat, remaja, tidur, dan sosial, kemudian pada skizofrenia banyak muncul kata remaja, perilaku, emosi, makan, dan percaya. Pembahasan tentang demensia banyak menggunakan kata sadar, cemas, kenal, sosial, emosi, dan tua. Sedangkan pada kategori PTSD banyak menggunakan kata trauma, hindar, malu, panik, dan waspada. Dari hasil tersebut dapat disimpulkan bahwa gangguan depresi, bipolar, dan skizofrenia banyak dikaitkan dengan remaja. Banyak remaja yang mengalami masalah di lingkungan sosial hingga mengalami stres, gangguan tidur, dan makan. Hal ini selaras dengan data Riskesdas [2] bahwa usia remaja memiliki presentase cukup tinggi sebesar 6,1% mengalami gejala-gejala kecemasan dan depresi. Dari hasil tersebut dapat disimpulkan bahwa klasifikasi *tweets* juga dapat digunakan untuk menemukan prevalensi masyarakat terkait gangguan kesehatan mental.

#### 4. Kesimpulan

Penelitian ini bertujuan untuk melakukan klasifikasi data *tweets* berdasarkan jenis-jenis gangguan kesehatan mental di Indonesia. Proses klasifikasi dilakukan menggunakan *cosine similarity* untuk menghitung similaritas antara data *tweets* dengan kata kunci yang disusun berdasarkan kajian teori dan optimasi dari hasil pemodelan topik LDA. Hasil klasifikasi menunjukkan isu paling banyak dibahas di Twitter secara berturut-turut adalah depresi, bipolar, skizofrenia, demensia, dan PTSD. Hal tersebut berarti tingkat prevalensi dan perhatian masyarakat terhadap gangguan depresi cukup tinggi daripada gangguan lainnya. Dari hasil klasifikasi juga dapat diketahui gejala yang banyak dibahas melalui kemunculan *keyword* dari masing-masing kategori. Hasil analisis data *tweets* diharapkan dapat dimanfaatkan bagi lembaga kesehatan yang berkaitan sebagai informasi tambahan untuk memutuskan kebijakan guna mengurangi jumlah penderita gangguan kesehatan mental dengan lebih tepat sasaran.

#### Daftar Pustaka

- [1] World Health Organization, “Mental Health,” *www.who.int*, 2022. <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response> (accessed Mar. 02, 2023).
- [2] Kementerian Kesehatan Republik Indonesia Badan Penelitian dan Pengembangan Kesehatan, “Laporan Nasional Riskesdas 2018,” *Kementerian Kesehatan RI*, 2019. <https://www.kemkes.go.id/article/view/19093000001/penyakit-jantung-penyebab-kematian-terbanyak-ke-2-di-indonesia.html> (accessed Jun. 12, 2023).
- [3] C. M. Annur, “Survei Populix: 1 dari 2 Penduduk Indonesia Punya Masalah Kesehatan Mental,” *databoks.katadata.co.id*, 2022. <https://databoks.katadata.co.id/datapublish/2022/10/27/survei-populix-1-dari-2-penduduk-indonesia-punya-masalah-kesehatan-mental> (accessed Mar. 02, 2023).
- [4] K. Saha, J. Torous, S. K. Ernala, C. Rizuto, A. Stafford, and M. De Choudhury, “A computational study of mental health awareness campaigns on social media,” *Transl. Behav. Med.*, vol. 9, no. 6, pp. 1197–1207, 2019, doi: 10.1093/tbm/ibz028.
- [5] N. Al Asad, M. A. Mahmud Pranto, S. Afreen, and M. M. Islam, “Depression Detection by Analyzing Social Media Posts of User,” *2019 IEEE Int. Conf. Signal Process. Information, Commun. Syst. SPICSCON 2019*, pp. 13–17, 2019, doi: 10.1109/SPICSCON48833.2019.9065101.
- [6] S. R. Kamite and V. B. Kamble, “Detection of Depression in Social Media via Twitter Using Machine learning Approach,” *Proc. 2020 Int. Conf. Smart Innov. Des. Environ. Manag. Plan. Comput. ICSIDEMPC 2020*, pp. 122–125, 2020, doi: 10.1109/ICSIDEMPC49020.2020.9299641.
- [7] N. S. Fal Dessai and J. A. Laxminarayanan, “A Topic Modeling based Approach for Mining Online Social Media Data,” *2019 2nd Int. Conf. Intell. Comput. Instrum. Control Technol. ICICICT 2019*, pp. 704–709, 2019, doi: 10.1109/ICICICT46008.2019.8993231.
- [8] M. Habibi, A. Priadana, A. B. Saputra, and P. W. Cahyo, “Topic Modelling of Germas Related Content on Instagram Using Latent Dirichlet Allocation (LDA),” vol. 34, no.

- Ahms 2020, pp. 260–264, 2021, doi: 10.2991/ahsr.k.210127.060.
- [9] P. I. Nainggolan, D. S. Prasvita, and D. S. Bukit, “Klasifikasi Informasi Kesehatan Pada Data Media Sosial Menggunakan Support Vector Machine dan K-Fold Cross Validation,” *Malikussaleh J. Mech. Sci. Technol.*, vol. 5, no. 2, pp. 34–38, 2021, [Online]. Available: <https://ojs.unimal.ac.id/mjmst/article/view/6317%0Ahttps://ojs.unimal.ac.id/mjmst/article/download/6317/3169>.
- [10] S. Gangbo and G. Shidaganti, “Classification of Student Mental Health Prediction Using LSTM,” *2022 IEEE 3rd Glob. Conf. Adv. Technol. GCAT 2022*, pp. 1–6, 2022, doi: 10.1109/GCAT55367.2022.9972061.
- [11] K. Reagan, A. S. Varde, and L. Xie, “Evolving Perceptions of Mental Health on Social Media and their Medical Impacts,” pp. 5328–5337, 2023, doi: 10.1109/bigdata55660.2022.10021013.
- [12] R. Diouf, E. N. Sarr, O. Sall, B. Birregah, M. Bousso, and S. N. Mbaye, “Web Scraping: State-of-the-Art and Areas of Application,” in *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 2019, pp. 6040–6042, doi: 10.1109/BIGDATA47090.2019.9005594.
- [13] World Health Organization, “Mental disorders,” *www.who.int*, Jun. 08, 2022. <https://www.who.int/en/news-room/fact-sheets/detail/mental-disorders> (accessed Jun. 14, 2023).
- [14] C. Sobin and H. A. Sackeim, “Psychomotor Symptoms of Depression,” *A m J Psychiatry*, vol. 154, no. January, pp. 4–17, 1997.
- [15] A. P. Association, “What are Bipolar Disorders?,” <https://www.psychiatry.org/>, 2023. <https://www.psychiatry.org/patients-families/bipolar-disorders/what-are-bipolar-disorders> (accessed Jul. 31, 2023).
- [16] N. C. Andreasen and M. Flaum, “Schizophrenia: The Characteristic Symptoms,” *Schizophr. Bull.*, vol. 17, no. 1, pp. 27–49, Jan. 1991, doi: 10.1093/SCHBUL/17.1.27.
- [17] World Health Organization, “Dementia,” <https://www.who.int/>, 2022. <https://www.who.int/news-room/fact-sheets/detail/dementia> (accessed Jul. 31, 2023).
- [18] M. L. Pacella, B. Hruska, and D. L. Delahanty, “The physical health consequences of PTSD and PTSD symptoms: A meta-analytic review,” *J. Anxiety Disord.*, vol. 27, no. 1, pp. 33–46, Jan. 2013, doi: 10.1016/j.janxdis.2012.08.004.
- [19] I. Hemalatha, D. G. P. S. Varma, and D. A. Govardhan, “Preprocessing The Informal Data for Efficient Sentiment Analysis,” *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 1, no. 2, p. 58, 2012, [Online]. Available: <http://ijettcs.org/Volume1Issue2/IJETTCS-2012-08-14-047.pdf>.
- [20] X. Sun, X. Liu, B. Li, Y. Duan, H. Yang, and J. Hu, “Exploring topic models in software engineering data analysis: A survey,” *2016 IEEE/ACIS 17th Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. SNPD 2016*, pp. 357–362, 2016, doi: 10.1109/SNPD.2016.7515925.

- [21] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” *Art Sci. Anal. Softw. Data*, vol. 3, pp. 993–1022, 2015, doi: 10.1016/B978-0-12-411519-4.00006-9.
- [22] M. Benard Magara, S. O. Ojo, and T. Zuva, “A comparative analysis of text similarity measures and algorithms in research paper recommender systems,” *2018 Conf. Inf. Commun. Technol. Soc. ICTAS 2018 - Proc.*, no. November 2019, pp. 1–5, 2018, doi: 10.1109/ICTAS.2018.8368766.
- [23] R. Alake, “Understanding Cosine Similarity and Its Applications,” *builtin.com*, 2023. <https://builtin.com/machine-learning/cosine-similarity> (accessed Jun. 19, 2023).
- [24] F. Fataruba, “Penerapan Metode Cosine Similarity Untuk Pengecekan Kemiripan Jawaban Ujian Siswa,” *J. Mhs. Tek. Inform.*, vol. 2, no. 2, pp. 88–95, 2018.
- [25] F. Alattar and K. Shaalan, “Emerging Research Topic Detection Using Filtered-LDA,” *AI*, vol. 2, no. 4, pp. 578–599, 2021, doi: 10.3390/ai2040035.