

BAB II

TINJAUAN PUSTAKA

2.1 Pustaka Terdahulu

Berdasarkan hasil kajian dan analisa beberapa pustaka telah menjadi rujukan dan tumpuan dalam melakukan seluruh proses penelitian. Daftar pustaka terdahulu dapat dilihat pada Tabel 1.

Tabel 1. Pustaka Terdahulu

No	Penulis (Tahun)	Kontribusi	Metode
1	Dwijayanti, et al. (2024) [10]	Judul: Klasifikasi Tweets Gangguan Kesehatan Mental di Indonesia Menggunakan LDA dan Cosine Similarity. Dataset: tweet atau re-tweet dalam Bahasa Indonesia menggunakan kata kunci “gangguan kesehatan mental”. Hasil: Klasifikasi tweets terkait gangguan kesehatan mental di Indonesia yang termasuk kedalam 5 kategori dengan akurasi 94%.	LDA dan Cosine Similarity
2	Damayanti, et al. (2020) [11]	Judul: ELIMINASI DATA NON-TOPIC MENGGUNAKAN PEMODELAN TOPIK UNTUK PERINGKASAN OTOMATIS DATA TWEET DENGAN KONTEKS COVID-19. Dataset: Twitter Api. Hasil: Hasil pengujian peringkasan dengan indikator ROUGE-L dari metode pemodelan topik LSI memiliki performa paling baik.	LDA, LSI dan HDP
3	Kevin, et al. (2022) [15]	Judul: Analisis Topik Tagar Covidindonesia pada Instagram Menggunakan Latent Dirichlet Allocation. Dataset: caption pada postingan di instagram dengan hashtag covidindonesia Hasil: Hasil dari nilai analisis menunjukkan bahwa 3 model lebih baik dari pada 4 atau 5 yang dibuktikan dengan nilai <i>perplexity & coherence</i> .	LDA
4	Diandra, et al. (2022) [16]	Judul: PEMODELAN TOPIK PADA KELUHAN PELANGGAN MENGGUNAKAN ALGORITMA LATENT DIRICHLET ALLOCATION DALAM MEDIA SOSIAL TWITTER. Dataset: tweet para pengguna twitter dengan	LDA

		<p>“@IndiHomCare” dan “@FirstMediaCares”.</p> <p>Hasil: Jumlah topik yang optimal digunakan yaitu 10 dan 11 untuk menghasilkan akurasi sebesar 70% dan 81,81% pembahasan topik pada hastag.</p>	
5	<p>Khairul, et al. (2021) [17]</p>	<p>Judul: SISTEM DETEKSI TOPIK POLITIK PADA TWITTER MENGGUNAKAN ALGORITMA LATENT DIRICHLET ALLOCATION.</p> <p>Dataset: tweet-tweet terkait peristiwa politik di Indonesia pada tahun 2019.</p> <p>Hasil: Hasil penelitian menunjukkan bahwa dengan pengujian 3 kali menggunakan jumlah data masing-masing 100, 1000, dan 6000 tweet, serta menggunakan setingan LDA bawaan dari library Genism dengan jumlah topik 10, menghasilkan rata-rata nilai kebenaran 90%.</p>	LDA

2.2 Definisi Kesehatan Mental

Kesehatan mental, sebagaimana didefinisikan oleh Organisasi Kesehatan Dunia (WHO), adalah suatu kondisi kesejahteraan mental yang membantu orang mengelola stresor kehidupan, mengenali potensi mereka, berhasil dalam belajar dan bekerja, dan berkontribusi pada komunitas mereka [1].

2.3 X (Twitter) dan Kesehatan Mental

Pengaruh media sosial pada masa kini membawa pengaruh besar terhadap berbagai aspek kehidupan tak terkecuali pada kesehatan mental. Media sosial, seperti X (Twitter), memiliki dampak signifikan pada kesehatan mental penggunaannya. Beberapa studi menunjukkan bahwa penggunaan media sosial X (Twitter) dapat meningkatkan kesadaran akan masalah kesehatan mental dan keterampilan dalam berinteraksi sosial [18]. Namun di sisi lain, penggunaan media sosial X (Twitter) berlebihan dapat meningkatkan risiko gangguan mental pada remaja, seperti depresi, kecemasan, dan stres [19].

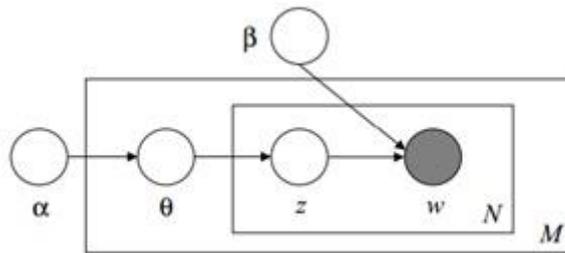
2.4 Label Kesehatan Mental

Penelitian ini menggunakan tujuh label yang didefinisikan oleh Pavlova

dan Berkers [20] untuk menganalisis berbagai aspek kesehatan mental yang dibahas di sosial media yang meliputi *'Awareness'* mencakup kesadaran individu mengenai kesehatan mental, termasuk berbagai penyakit mental. *'Feelings and Problematization'* terdiri dari tweet yang mengekspresikan perasaan seperti kecemasan dan stres, serta membahas berbagai masalah kesehatan mental. *'Classification'* mengelompokkan tweet berdasarkan masalah kesehatan mental seperti stres, depresi, dan kecemasan. *'Accessibility and Funding'* mencakup tweet yang membahas aksesibilitas dan pendanaan layanan kesehatan mental. *'Stigma'* berisi tweet yang membahas stigma masyarakat terhadap masalah kesehatan mental. *'Service'* mencakup tweet yang berkaitan dengan layanan kesehatan mental, termasuk mencari atau menawarkan bantuan. *'Youth'* terdiri dari tweet yang membahas masalah kesehatan mental pada remaja. Penggunaan ketujuh label ini membantu penelitian mengidentifikasi dan menganalisis berbagai aspek kesehatan mental yang dibahas di media sosial, khususnya Twitter, dan memberikan wawasan komprehensif tentang persepsi dan pengalaman individu terkait kesehatan mental.

2.5 Latent Dirichlet Allocation (LDA)

Pada LDA mengasumsikan bahwa setiap dokumen terdiri dari campuran topik, dan setiap topik terdiri dari kombinasi kata. Algoritma ini bekerja dengan menganalisis frekuensi kata dalam dokumen untuk menentukan topik yang paling mungkin terdapat dalam dokumen tersebut [12]. LDA memiliki tiga tingkat representasi, dengan parameter α dan β yang berfungsi sebagai parameter corpus atau kumpulan kata yang diasumsikan diambil sampelnya satu kali dalam proses pembuatan corpus. Selanjutnya, variabel θ_d merupakan variabel tingkat dokumen yang diambil sampelnya satu kali untuk setiap dokumen. Sementara itu, variabel z_{dn} dan w_{dn} adalah variabel kata yang diambil sampelnya satu kali untuk setiap kata dalam setiap dokumen. Representasi grafis LDA dapat dilihat pada Gambar 1. Kotak-kotak dalam diagram tersebut menunjukkan replikasi, di mana kotak luar menggambarkan dokumen, sementara kotak dalam menunjukkan pemilihan topik dan kata yang berulang dalam dokumen tersebut [10].



Gambar 1. Representasi Grafis LDA

2.6 Latent Semantic Indexing (LSI)

Latent Semantic Indexing adalah metode pengindeksan dan pencarian yang menggunakan matematika yang disebut Singular Value Decomposition (SVD) untuk mengidentifikasi pola hubungan antara istilah dan konsep-konsep yang terkandung dalam sebuah koleksi teks yang tidak terstruktur [21]. Dalam SVD, jika A adalah matriks dengan m baris yang menunjukkan fitur/term (kata) unik dan n kolom yang mewakili dokumen, setiap sel dalam matriks menyatakan frekuensi kata dalam dokumen. LSI kemudian melakukan perhitungan Singular Value Decomposition (SVD) pada matriks A , menghasilkan tiga matriks. Matriks pertama merepresentasikan baris entity sebagai vektor dari nilai ortogonal, matriks kedua adalah matriks diagonal yang terdiri dari nilai skala dari ketiga komponen matriks, dan matriks ketiga menjelaskan kolom original. Jika A berukuran $m \times n$ dengan $m \geq n$, maka SVD (A) didefinisikan sebagai:

$$\begin{array}{cccc}
 (A) & = & (U) & (S) & (V^t) \\
 m \times n & & m \times n & n \times n & n \times n
 \end{array}$$

SVD melibatkan pemfaktoran A ke dalam hasil kali USV^t , dimana :

U = matriks kolom orthogonal $m \times n$ atau disebut juga dengan left singular vectors

S = matriks diagonal $n \times n$

V = matriks orthogonal $n \times n$, atau disebut juga dengan right singular vectors

V^T = matriks transpose dari matriks V

2.7 Hierarchical Dirichlet Process (HDP)

HDP merupakan algoritma yang memiliki pendekatan Bayesian non-parametrik untuk data terkluster. Ini berfungsi sebagai model keanggotaan campuran yang kuat untuk pengelompokan data tanpa pengawasan. Tidak seperti LDA dan LSI yang memerlukan penentuan jumlah topik sebelumnya, HDP mampu menentukan jumlah topik secara otomatis berdasarkan data yang dianalisis [14].

