

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

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]. Definisi ini menggarisbawahi bagian positif dari kesehatan mental, menyoroti peran pentingnya dalam memfasilitasi individu untuk menikmati kehidupan yang memuaskan dan berinteraksi secara produktif dengan anggota komunitas mereka. Namun, pandemi COVID-19 menyebabkan peningkatan masalah kesehatan mental seperti depresi dan kecemasan di seluruh dunia [2]. Tren ini juga terlihat jelas di Indonesia, dimana tingkat masalah kesehatan mental di kalangan pengguna pemeriksaan mandiri meningkat dari 70,7% pada tahun 2019 menjadi 80,4% pada tahun 2020, dan selanjutnya menjadi 82,5% pada tahun 2022 [3]. Hal ini juga dibuktikan dengan survei mengenai kesehatan mental melalui swaperiksa yang dihimpun oleh Perhimpunan Dokter Spesialis Kedokteran Jiwa Indonesia (PDSKJI) yang dilakukan secara daring menjelaskan bahwa sebanyak 63 persen responden mengalami cemas dan 66 persen responden mengalami depresi akibat pandemi COVID-19 [4]. Kesehatan mental di Indonesia juga ramai kembali dengan adanya peristiwa pemilu 2024 yang banyak menimbulkan Masyarakat yang mengalami gangguan Kesehatan mental. Dalam menangani fenomena tersebut, diperlukan metode yang komprehensif untuk mendeteksi masalah kesehatan mental. Salah satu solusinya adalah penggunaan *Machine Learning (ML)* untuk mendeteksi dan menganalisis topik terkait kesehatan mental, seperti pola perilaku, pemicu, dan pengobatan.

*Machine Learning (ML)* adalah cabang dari kecerdasan buatan yang berkaitan dengan pengembangan sistem komputer yang mampu belajar dari data tanpa perlu diprogram secara eksplisit. Prinsip dasar *ML* adalah

menggunakan algoritma untuk menganalisis data, mengidentifikasi pola, dan membuat prediksi atau keputusan berdasarkan pola-pola tersebut. Dalam dua dekade terakhir, *ML* menjadi terkenal karena efektivitasnya dalam mengatasi berbagai masalah pendeteksian. Pemanfaatan *ML* di bidang kesehatan mental telah melonjak untuk tugas-tugas seperti mengidentifikasi perilaku individu yang membantu memahami gejala kesehatan mental dan faktor risiko [5], serta mendeteksi dan mendiagnosis kondisi kesehatan mental melalui analisis data pasien [6]. Pesatnya perluasan platform media sosial telah memfasilitasi pendeteksian masalah kesehatan mental, dan X (Twitter) menjadi platform populer untuk tujuan ini. Menganalisis masalah kesehatan mental di Twitter menawarkan beberapa keuntungan, antara lain (a) memprediksi dan memantau masalah kesehatan mental berdasarkan aktivitas online [7]; (b) pengambilan data secara *real-time*, memungkinkan pelacakan tren kesehatan mental [8]; dan (c) kemampuan menerapkan teknik *Natural Language Processing* (NLP) untuk menganalisis data Twitter, membantu mengidentifikasi pola terkait kesehatan mental [9]. Mempertimbangkan faktor-faktor tersebut, tidak mengherankan jika berbagai penelitian yang berfokus pada masalah kesehatan mental memanfaatkan data Twitter.

Dilihat pada penelitian sebelumnya terdapat penelitian yang menggunakan *Topic Modeling* pada bidang kesehatan mental yaitu pada penelitian yang mengklasifikasikan tweet berdasarkan topik kesehatan mental namun hanya menggunakan algoritma LDA yang dapat dikategorikan kedalam 5 kategori yaitu ‘depresi’, ‘gangguan bipolar’, ‘Skizofrenia’, ‘Demensia’ dan ‘PTSD’ [10]. Selanjutnya pada penelitian yang membuat peringkasan otomatis menggunakan pemodelan topik dengan data tweet namun penelitian ini berfokus pada pembahasan Covid-19. Algoritma yang digunakan yaitu *Latent Dirichlet Allocation* (LDA), *Latent Semantic Indexing* (LSI), dan *Hierarchical Dirichlet Process* (HDP) dengan hasil algoritma LSI yang memiliki performa paling baik [11].

Dalam penelitian ini menggunakan tiga algoritma yaitu LDA, LSI, dan HDP. 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]. LSI adalah teknik yang mengidentifikasi topik dengan menganalisis hubungan antara kata-kata dan dokumen-dokumen serta mengidentifikasi pola kejadian umum. Dengan menggunakan LSI dapat mengurangi dimensi data dan menemukan struktur mendasar yang menghubungkan kata-kata dengan dokumen-dokumen tertentu [13]. HDP, di sisi lain, adalah pendekatan Bayesian non-parametrik untuk data terklaster. Ini berfungsi sebagai model keanggotaan campuran yang kuat untuk pengelompokan data tanpa pengawasan. Tidak seperti LDA yang memerlukan penentuan jumlah topik sebelumnya, HDP mampu menentukan jumlah topik secara otomatis berdasarkan data yang dianalisis [14]. Ini sangat berguna dalam konteks di mana jumlah topik tidak diketahui sebelumnya dan dapat berubah-ubah, seperti dalam analisis media sosial di mana topik pembicaraan dapat bervariasi dari waktu ke waktu. Dengan menggunakan ketiga algoritma ini, penelitian dapat lebih efektif dalam mengidentifikasi dan memahami topik-topik yang mendasari berbagai jenis data teks, sehingga memberikan wawasan yang lebih dalam tentang pokok pembicaraan dalam suatu konteks.

Namun demikian, hasil review pada bidang ini terkait penggunaan *Topic Modeling* pada platform media sosial menunjukkan adanya beberapa gap penelitian, yaitu: **(i) minim atau sedikitnya penelitian yang menerapkan *topic modeling* di bidang kesehatan mental, meskipun isu kesehatan mental populer; (ii) penelitian saat ini cenderung terbatas pada penggunaan satu algoritma pemodelan topik saja, yaitu LDA dan hanya menggunakan skenario supervised (menggunakan data yang berlabel).** Maka dari itu belum ada penelitian yang melakukan ekstraksi kata-kata kunci yang merepresentasikan setiap jenis kategori kesehatan mental di Indonesia dengan membandingkan berbagai algoritma *Topic Modeling*. Daftar kata kunci ini penting untuk diketahui untuk mengidentifikasi, memahami, dan mengelompokkan topik dari isu yang dibahas dalam teks cuitan. Sehingga, penelitian ini bertujuan untuk mengambil pendekatan tersebut dengan mengimplementasikan algoritma seperti *Latent Dirichlet Allocation* (LDA), *Latent Semantic Indexing* (LSI), dan *Hierarchical Dirichlet Process* (HDP)

dengan kerangka kesehatan mental yang digunakan terdiri dari 7 label utama yaitu “*Awareness*”, “*Feelings and Problematization*”, “*Classification*”, “*Accessibility and Funding*”, “*Stigma*”, “*Service*”, dan “*Youth*”. Diharapkan melalui penelitian ini, dapat dihasilkan data yang terkategori dengan tepat sesuai dengan kerangka analisis yang dapat digunakan untuk memahami lebih dalam isu Kesehatan Mental di Indonesia.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka dapat diuraikan rumusan masalah yaitu sebagai berikut:

- a. Bagaimana pengambilan dan *preprocessing* data twitter dengan topik kesehatan mental di Indonesia.
- b. Bagaimana penerapan skenario deteksi supervised dan unsupervised pada data Twitter (X) tentang Kesehatan Mental.
- c. Bagaimana implementasi algoritma *Topic Modeling* dalam mengekstraksi kata kunci data twitter (X) tentang kesehatan mental.
- d. Apa saja kata kunci yang dihasilkan dari *Topic Modeling* dalam isu kesehatan mental di Indonesia.
- e. Evaluasi algoritma *Latent Dirichlet Allocation* (LDA), *Latent Semantic Indexing* (LSI), dan *Hierarchical Dirichlet Process* (HDP) dalam menghasilkan data berupa kata kunci untuk isu kesehatan mental di Indonesia

## 1.3 Tujuan Penelitian

Tugas akhir ini bertujuan untuk mendekteksi topik pembicaraan pada platform X (Twitter) terkait dengan isu kesehatan mental.

- a. Mengekstraksi kata kunci utama pada isu kesehatan mental pada data Twitter (X).
- b. Implementasi algoritma *Topic Modeling* menggunakan *Latent Dirichlet Allocation* (LDA), *Latent Semantic Indexing* (LSI), dan *Hierarchical Dirichlet Process* (HDP).

## 1.4 Batasan Masalah

Terdapat beberapa batasan permasalahan pada penelitian ini, diantaranya adalah:

- a. Dataset yang digunakan menggunakan data Twitter (X) dari mulai 2020 – 2023 dengan kata kunci “mental health” dan “kesehatan mental”.
- b. Penggunaan algoritma pada penelitian ini hanya menggunakan *Latent Dirichlet Allocation* (LDA), *Latent Semantic Indexing* (LSI), dan *Hierarchical Dirichlet Process* (HDP).

