

BAB I PENDAHULUAN

I.1 Latar Belakang

Di era digital saat ini, kesehatan mental telah menjadi isu global yang semakin mendapatkan perhatian di seluruh dunia. Kesehatan mental adalah kondisi kesejahteraan di mana individu mampu menyadari kemampuan dirinya sendiri, mengatasi tekanan hidup, bekerja secara produktif, dan berkontribusi bagi komunitasnya (*World Health Organization*, 2022). Pentingnya kesehatan mental tidak hanya terbatas pada individu secara umum, tetapi juga memiliki dampak signifikan dalam konteks ibu hamil.

Dalam konteks ibu hamil, kesehatan mental menjadi sangat penting karena kondisi psikologis ibu tidak hanya memengaruhi kesejahteraan fisiknya tetapi juga perkembangan janin (Bedaso dkk., 2021). Salah satu gangguan kesehatan mental yang paling umum terjadi pada ibu hamil adalah *common mental disorders* (CMD), yang mencakup kondisi seperti depresi dan kecemasan (Ariasih dkk., 2024). CMD sering kali muncul selama kehamilan akibat perubahan hormonal, stres psikologis (Ariasih dkk., 2024), dan tekanan sosial (Bedaso dkk., 2021). Kondisi ini tidak hanya memengaruhi kesejahteraan ibu tetapi juga berpotensi mengganggu perkembangan janin, yang dapat berdampak pada kesehatan fisik dan psikologis anak di kemudian hari (Bränn dkk., 2024).

Di Indonesia, prevalensi CMD pada ibu hamil mencapai sekitar 12,6%, menunjukkan tingginya beban kesehatan mental pada kelompok rentan ini (Ariasih dkk., 2024). Angka ini menandakan bahwa satu dari delapan ibu hamil di Indonesia berisiko mengalami gangguan kesehatan mental yang membutuhkan perhatian khusus. Berbagai faktor sosial, ekonomi, dan budaya di Indonesia turut meningkatkan risiko CMD pada ibu hamil. Tekanan ekonomi seperti keterbatasan finansial dapat menyebabkan stres tambahan (Hou & Hou, 2022; McGowan dkk., 2023). Keterbatasan akses terhadap layanan kesehatan mental, terutama di daerah pedesaan, menghambat ibu hamil untuk mendapatkan bantuan profesional (Bedaso dkk., 2021). Jika CMD pada ibu hamil tidak ditangani dengan baik, dampaknya bisa sangat serius, baik bagi ibu maupun perkembangan janin. CMD yang tidak tertangani dapat menyebabkan risiko kelahiran prematur, berat badan

lahir rendah, dan gangguan perkembangan kognitif pada anak (Bränn dkk., 2024). Selain itu, ibu berisiko mengalami gangguan kesehatan mental jangka panjang (Howard & Khalifeh, 2020), termasuk kecenderungan menyakiti diri sendiri (Bedaso dkk., 2021; McGowan dkk., 2023). Oleh karena itu, diperlukan upaya untuk memahami dan mengatasi permasalahan kesehatan mental pada ibu hamil di Indonesia.

Urgensi penelitian ini juga ditinjau dari masih minimnya studi atau kajian akademik tentang analisis sentimen terhadap kesehatan mental ibu hamil di Indonesia melalui pendekatan *Natural Language Processing* (NLP). Hingga saat ini, sebagian besar kajian hanya berfokus pada CMD atau pendekatan klinis oleh ahli psikologi dan kesehatan (Ariasih dkk., 2024; Bedaso dkk., 2021; Bränn dkk., 2024), namun belum mengeksplorasi secara mendalam pemanfaatan data teks media sosial secara otomatis untuk mengidentifikasi kebutuhan emosional ibu hamil secara *real-time*. Oleh karena itu, penelitian ini tidak hanya mengisi celah akademik dalam bidang NLP berbahasa Indonesia, tetapi juga memberikan landasan ilmiah bagi pengembangan intervensi digital berbasis sentimen dan pemodelan topik yang lebih kontekstual dan relevan bagi ibu hamil di Indonesia.

Selanjutnya, Untuk memahami persepsi dan kebutuhan kesehatan mental ibu hamil, berbagai metode analisis sentimen telah diterapkan dalam penelitian. Beberapa model *Transformer*, seperti BERTweet, DistilBERT, dan Twitter RoBERTa, diterapkan untuk mengukur sentimen pada data media sosial secara otomatis, memungkinkan pengenalan pola-pola emosional dengan akurasi tinggi dalam konteks kesehatan mental (Zhao & Wong, 2024). Metode ini efektif dalam menangkap konteks dan nuansa bahasa, namun sebagian besar dikembangkan untuk Bahasa Inggris, sehingga adaptasinya ke Bahasa Indonesia menghadapi tantangan seperti perbedaan struktur bahasa dan keterbatasan data yang memadai.

Selain analisis sentimen, pemetaan topik penting untuk memperoleh gambaran utuh tentang isu kesehatan mental ibu hamil di media sosial. Metode klasik LDA masih banyak digunakan karena mampu memetakan sebaran topik pada kumpulan teks besar (Laureate dkk., 2023). Upaya meningkatkan koherensi topik mendorong pemanfaatan pendekatan berbasis matriks seperti LSI atau NMF, serta

model non-parametrik HDP yang secara otomatis menyesuaikan jumlah topik (Arias dkk., 2022). Untuk teks pendek dan penuh bahasa gaul, metode berbasis *embedding* lebih banyak dipilih misalnya Top2Vec yang menggabungkan doc2vec dan HDBSCAN (Egger & Yu, 2022) atau BERTopic yang memanfaatkan *embedding* BERT (Egger & Yu, 2022; Grootendorst, 2020).

Sementara itu, metode klasifikasi seperti *Support Vector Machine* (SVM), *Random Forest*, dan *Deep Neural Network* (DNN) telah terbukti efektif dalam mengkategorikan sentimen secara mendalam pada teks media sosial. Metode-metode ini menghasilkan akurasi yang tinggi dalam pengelompokan sentimen (Sarker dkk., 2017). Namun, metode ini memiliki keterbatasan dalam memahami hubungan konteks antar kata, terutama pada teks panjang yang kompleks seperti Bahasa Indonesia. Untuk mengatasi keterbatasan ini, model *transformer* seperti IndoBERT memberikan solusi yang lebih canggih. Dengan kemampuan memahami struktur dan konteks bahasa secara menyeluruh, IndoBERT menjadi pilihan yang lebih relevan dalam analisis sentimen media sosial, terutama untuk Bahasa Indonesia.

Dalam konteks Bahasa Indonesia, penggunaan IndoBERT dalam analisis sentimen memiliki beberapa keunggulan. IndoBERT mampu memahami struktur kalimat, idiom, dan nuansa Bahasa Indonesia, termasuk bahasa informal dan slang yang sering digunakan di media sosial (Nabiilah dkk., 2024). Hal ini meningkatkan akurasi dalam pengenalan sentimen dan memungkinkan analisis data *real-time* yang lebih autentik (Kartika dkk., 2023). Selain itu, IndoBERT dapat mengolah data dalam jumlah besar secara otomatis, menghemat waktu dan sumber daya dibandingkan dengan metode tradisional seperti wawancara individu (Jayadianti dkk., 2022).

Setelah sentimen komentar terklasifikasi, diperlukan algoritma *topic modeling* yang efektif untuk mengidentifikasi tema pada teks pendek dan informal seperti komentar TikTok. Metode klasik seperti LDA, LSI, NMF, dan HDP mampu menemukan topik tersembunyi, namun keterbatasan konteks pada pendekatan tersebut membuatnya kurang koheren pada teks pendek dan penuh slang (Arias dkk., 2022; Egger & Yu, 2022; Laureate dkk., 2023). Meskipun Top2Vec mampu

mengatasi isu konteks dengan pendekatan *embedding* berbasis doc2vec, metode ini memerlukan sumber daya tinggi, sensitif terhadap parameter, serta belum mendukung pemisahan sentimen (Egger & Yu, 2022). Sebaliknya, BERTopic dengan *embedding* IndoBERT, UMAP, dan HDBSCAN menghasilkan topik yang lebih koheren dan mudah diinterpretasi untuk data pendek dan informal (Egger & Yu, 2022; Grootendorst, 2020). Oleh karena itu, penelitian ini memilih BERTopic sebagai metode utama pemodelan topik.

Metode klasifikasi seperti SVM, *Random Forest*, dan DNN memiliki kinerja yang baik dalam pengkategorian sentimen dengan kompleksitas yang lebih rendah dibandingkan model *transformer* (Arias dkk., 2022). Kelebihannya termasuk kecepatan pelatihan dan interpretabilitas yang lebih baik. Namun, mereka kurang mampu menangkap hubungan konteks dan urutan kata dalam teks panjang, terutama dalam bahasa yang kompleks seperti Bahasa Indonesia.

Namun demikian, terdapat beberapa tantangan dalam penggunaan IndoBERT. Analisis sentimen mungkin tidak menggali wawasan mendalam yang bisa diperoleh melalui wawancara langsung, di mana peneliti dapat mengeksplorasi jawaban lebih lanjut. Selain itu, tidak semua ibu hamil aktif di media sosial, sehingga data mungkin tidak sepenuhnya mewakili populasi ibu hamil di Indonesia, mengakibatkan representasi populasi yang tidak merata. Kemungkinan bias juga dapat muncul jika data pelatihan tidak mencakup variasi bahasa dari berbagai daerah dan latar belakang.

Salah satu tantangan utama adalah bagaimana memastikan bahwa output dari analisis sentimen tidak bias. Bias data dapat terjadi jika data yang dikumpulkan didominasi oleh kelompok tertentu, misalnya ibu hamil dari daerah perkotaan yang lebih aktif di media sosial. Selain itu, IndoBERT mungkin kurang akurat dalam memahami bahasa daerah atau dialek yang tidak umum, serta perlu memahami konteks spesifik terkait kesehatan mental ibu hamil untuk memberikan interpretasi yang akurat. Untuk mengatasi masalah tersebut, penelitian ini akan:

1. Melakukan *fine-tuning* IndoBERT dengan data spesifik: Melatih ulang model IndoBERT menggunakan *dataset* yang dikumpulkan dari media sosial ibu

hamil di Indonesia, sehingga model lebih sensitif terhadap konteks dan bahasa yang digunakan.

2. Pengembangan Teknik *Preprocessing* yang ditingkatkan: Mengembangkan *preprocessing* teks yang mampu menangani bahasa informal, slang, dan istilah khusus yang digunakan oleh ibu hamil.
3. Integrasi dengan BERTopic untuk pemodelan topik: Menggunakan BERTopic untuk mengidentifikasi topik utama yang dibicarakan oleh ibu hamil, memperkaya analisis sentimen dengan pemahaman isu-isu yang relevan.

Berdasarkan permasalahan dan kesenjangan penelitian yang telah diuraikan, penelitian ini bertujuan untuk menganalisis sentimen terhadap kesehatan mental ibu hamil di Indonesia dengan mengembangkan metode yang lebih akurat dan relevan secara budaya. Hasil analisis ini akan digunakan untuk merekomendasikan fitur aplikasi yang mendukung kesehatan mental ibu hamil, sehingga berkontribusi pada peningkatan kesejahteraan ibu dan anak di Indonesia. Penelitian ini diharapkan memberikan kontribusi penting dalam memahami kesehatan mental ibu hamil, yang merupakan isu mendesak dalam konteks kesehatan masyarakat. Dengan mengidentifikasi isu utama dan kebutuhan spesifik ibu hamil, penelitian ini menyediakan landasan ilmiah untuk pengembangan aplikasi kesehatan mental yang relevan dan efektif. Aplikasi ini dirancang tidak hanya untuk mendukung kesejahteraan ibu hamil, tetapi juga untuk mengurangi dampak negatif Common Mental Disorders (CMD) pada ibu dan anak melalui solusi teknologi yang inovatif.

I.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, perumusan masalah dalam penelitian ini dapat dinyatakan sebagai berikut:

1. Belum adanya pemahaman yang memadai mengenai efektivitas penerapan analisis sentimen dan *topic modeling* dengan IndoBERT serta BERTopic dalam menghasilkan rekomendasi fitur aplikasi kesehatan mental yang relevan bagi ibu hamil di Indonesia.

2. Terbatasnya pengetahuan tentang isu dan kebutuhan utama terkait kesehatan mental ibu hamil di Indonesia yang dapat diungkap melalui analisis data media sosial.
3. Minimnya penelitian yang secara khusus mengkaji analisis sentimen dan identifikasi tema terkait kesehatan mental ibu hamil di Indonesia, khususnya pada data percakapan pendek dan informal di media sosial, sehingga diperlukan pengembangan metode berbasis IndoBERT dan BERTopic yang mampu memahami konteks dan nuansa bahasa Indonesia secara lebih akurat.

I.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Menggunakan hasil analisis sentimen dan *topic modeling* untuk merekomendasikan fitur aplikasi kesehatan mental bagi ibu hamil di Indonesia.
2. Mengidentifikasi isu dan kebutuhan utama yang memengaruhi kesehatan mental ibu hamil di Indonesia melalui analisis data dari media sosial.
3. Mengisi kekosongan penelitian terkait analisis sentimen dan identifikasi tema kesehatan mental ibu hamil di Indonesia, khususnya pada percakapan pendek dan informal di media sosial, melalui pengembangan dan penerapan metode berbasis IndoBERT dan BERTopic.

I.4 Batasan Penelitian

Berikut adalah batasan masalah dari penelitian saya

1. Data untuk penelitian analisis sentimen berupa postingan dan komentar teks yang berkaitan dengan kesehatan mental ibu hamil di Indonesia, hanya diambil dari *platform* media sosial TikTok.
2. Data dikumpulkan menggunakan alat pengumpulan data seperti Apify.com dan hanya mencakup konten berbahasa Indonesia.
3. Periode waktu pengambilan data adalah dari Januari 2022 hingga Desember 2024.
4. *Framework data mining* yang digunakan adalah *Knowledge Discovery in Databases* (KDD).

5. Sentimen yang dianalisis dalam penelitian ini dibatasi pada dua kelas, yaitu positif dan negatif.
6. Penelitian ini fokus pada identifikasi sentimen dan topik utama yang diekspresikan oleh ibu hamil di media sosial untuk merekomendasikan fitur aplikasi kesehatan mental yang sesuai.
7. Konten non-teks seperti gambar, video, dan audio tidak dianalisis dalam penelitian ini.
8. Penelitian tidak bertujuan untuk mengukur prevalensi gangguan kesehatan mental atau memberikan diagnosis klinis, tetapi fokus pada analisis sentimen untuk rekomendasi fitur aplikasi.
9. Aspek etika dan privasi diperhatikan dengan hanya menggunakan data publik dan menjaga anonimitas pengguna media sosial.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan, baik secara teoretis maupun praktis, bagi pengembangan layanan kesehatan mental ibu hamil di Indonesia. Secara teoretis, penelitian ini memberikan wawasan mendalam mengenai kondisi kesehatan mental ibu hamil, sehingga dapat membantu mengidentifikasi kebutuhan utama dan faktor-faktor yang memengaruhi kesejahteraan mereka. Selain itu, penelitian ini dapat menjadi referensi dalam penelitian serupa di masa depan, khususnya yang berkaitan dengan analisis sentimen dan pemodelan topik menggunakan algoritma IndoBERT dan BERTopic. Secara praktis, hasil penelitian ini diharapkan dapat menjadi acuan bagi pembuat kebijakan untuk mengembangkan layanan kesehatan mental yang lebih efektif dan responsif. Dengan rekomendasi berbasis data yang dihasilkan, penelitian ini juga diharapkan dapat meningkatkan kualitas hidup ibu hamil serta perkembangan anak yang dilahirkan melalui layanan kesehatan mental yang lebih baik dan tepat sasaran. mental yang lebih baik dan responsif terhadap kebutuhan mereka.

I.6 Sistematika Penulisan

BAB I Pendahuluan

Bab ini menyoroti tingginya gangguan kesehatan mental pada ibu hamil di Indonesia dan memanfaatkan data TikTok untuk menggalinya. Masalah utamanya ialah kebutuhan emosional bumil yang belum terpetakan; penelitian bertujuan membangun kerangka analisis sentimen–topik dengan IndoBERT dan BERTopic guna merekomendasikan fitur aplikasi kesehatan mental. Manfaatnya mengisi celah riset NLP berbahasa Indonesia sekaligus menyediakan panduan praktis bagi pengembang aplikasi. Ruang lingkupnya terbatas pada caption dan komentar TikTok 2022–2024, dengan satu *platform*, sentimen biner, dan asumsi bahwa ekspresi daring mencerminkan kondisi emosional.

BAB II Landasan Teori

Bab ini memuat konsep kesehatan mental perinatal, peran media sosial, dan teori *coping*. Di ranah NLP: *transformer* & IndoBERT, *semi-supervised learning* / *pseudo-Labeling*, BERTopic (*Embedding* + UMAP + HDBSCAN) berikut perbandingannya dengan LDA, serta teknik *fuzzy string-matching* untuk *aspect Mapping*. Bagian ini menjustifikasi pemilihan metode dan metrik (*accuracy*, *macro-F1*, *coherence*).

BAB III Metodologi Penyelesaian Masalah

Bab ini akan menjelaskan pendekatan yang digunakan dalam penelitian ini. Bab ini akan mencakup penjelasan mengenai metodologi konseptual menggunakan *Design Science in Information Systems Research*. Selain itu bab ini juga menjelaskan mengenai langkah-langkah sistematis dalam menyelesaikan masalah. Penjelasan akan memuat informasi mengenai *Knowledge Discovery in Databases* yang merupakan proses pengumpulan dan pengolahan data, serta metode evaluasi yang diterapkan dalam penelitian ini.

BAB IV Penyelesaian Masalah

Bab ini menguraikan rangkaian pengolahan data dan eksperimen yang dilakukan, dimulai dari pembersihan teks dan penyiapan korpus, pemisahan data latih–uji, hingga konversi teks ke bentuk vektor yang sesuai bagi model *transformer*.

Selanjutnya dijabarkan proses penyetelan IndoBERT untuk klasifikasi sentimen, diikuti pemetaan topik menggunakan BERTopic yang memecah percakapan ke dalam tema-tema kunci. Kata kunci tiap topik kemudian disejajarkan dengan empat belas aspek psikososial melalui teknik *fuzzy matching*. Tahapan tersebut dievaluasi menggunakan metrik akurasi, F1-makro, dan koherensi topik agar analisis persepsi dan kebutuhan emosional ibu hamil tersaji secara objektif dan terstruktur.

BAB V Validasi, Analisis, Hasil, Dan Implikasi

Bab ini menjelaskan pengujian keandalan model melalui validasi silang dan perbandingan dengan metode dasar, serta menampilkan hasil analisis pola sentimen dan topik utama. Temuan yang diperoleh dibahas sebagai kontribusi bagi riset NLP berbahasa Indonesia dan kesehatan mental perinatal, serta diterapkan dalam rancangan fitur aplikasi, seperti modul coping, konten spiritual, forum dukungan, alat keterlibatan pasangan, dan materi CBT otomatis saat terdeteksi sentimen negatif.

BAB VI Kesimpulan Dan Saran

Bab ini merangkum temuan pokok penelitian sekaligus menegaskan kesesuaian kerangka IndoBERT–BERTopic dengan rumusan masalah yang telah ditetapkan. Bab ini juga menghadirkan saran konseptual yang meliputi penerapan fitur aplikasi berdasarkan temuan, arah riset lanjutan seperti eksplorasi lintas-*platform*, deteksi multi-emosi dan juga augmentasi data lanjutan.