

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Depresi merupakan salah satu gangguan kesehatan mental dengan prevalensi tinggi secara global dan memberikan dampak signifikan terhadap kualitas hidup penderitanya. Gangguan ini tidak hanya memengaruhi kondisi psikologis individu, tetapi juga berdampak pada aspek sosial, ekonomi, hingga fisik. Dalam kasus yang bersifat kronis atau berulang, depresi dapat menurunkan kemampuan fungsional secara drastis, meningkatkan risiko disabilitas jangka panjang, serta memperburuk angka kematian, baik secara langsung (misalnya karena bunuh diri) maupun tidak langsung [1]. Berdasarkan data dari *World Health Organization* (WHO) pada tahun 2021, lebih dari 280 juta orang di seluruh dunia mengalami depresi [2]. Dampak depresi tidak hanya terbatas pada penurunan kualitas hidup individu, melainkan juga dapat menyebabkan penurunan produktivitas serta meningkatkan risiko komplikasi kesehatan lainnya. Pada kelompok remaja, deteksi dini depresi sering kali terkendala oleh adanya stigma sosial serta keterbatasan akses terhadap layanan kesehatan mental [3]. Oleh karena itu, dibutuhkan pendekatan inovatif yang mampu membantu proses deteksi gangguan dini secara lebih efektif dan efisien.

Salah satu cara untuk mengatasi permasalahan dalam kesehatan mental adalah dengan menggunakan teknologi *Artificial Intelligent* (AI) khususnya dalam bidang *Natural Language Processing* (NLP) mulai digunakan untuk deteksi dini gangguan kesehatan mental [4]. Salah satu penerapan yang berkembang pesat adalah *chatbot* berbasis teks yang dinilai sebagai solusi praktis, aman, dan ekonomis untuk membantu individu menyampaikan keluhan emosional secara terbuka.

Model BERT (*Bidirectional Encoder Representations from Transformers*) merupakan salah satu model NLP yang telah menunjukkan performa tinggi dalam memahami konteks bahasa alami secara *bidirectional*. Model ini

banyak digunakan dalam berbagai tugas klasifikasi teks termasuk deteksi depresi [5]. Dataset DAIC-WOZ (*Distress Analysis Interview Corpus-Wizard of Oz*) menjadi salah satu sumber data standar yang banyak digunakan dalam penelitian terkait karena menyediakan transkrip wawancara klinis lengkap. Dataset ini terdiri dari 189 sesi wawancara klinis yang mencakup data audio, video, dan transkrip percakapan, serta dilengkapi dengan label tingkat depresi berdasarkan skor PHQ-8 [6].

Berbagai pendekatan telah dikembangkan untuk mengoptimalkan deteksi depresi berbasis teks. *Gavalan et al. (2023)* menerapkan metode *text summarization* berbasis KeyBERT untuk menyesuaikan panjang input terhadap batas BERT dan mencapai *F1-score* 67% pada dataset DAIC-WOZ [7]. *Burdisso et al. (2024)* menggunakan pendekatan berbasis teks untuk mengevaluasi peran pertanyaan dari *chatbot* Ellie dalam klasifikasi otomatis depresi dari wawancara klinis. Mereka menemukan bahwa kombinasi *prompt* dan respons menghasilkan *F1-score* hingga 88% pada dataset DAIC-WOZ, namun juga menyoroti potensi bias dari isi pertanyaan terapis [8]. *Zhang et al. (2025a)* lainnya menerapkan *multi-instance learning* dengan model MT5 dan RoBERTa (Multi-MTRB) untuk klasifikasi berbasis teks pada DAIC-WOZ dan mencapai *F1-score* 88% [9]. *Zhang et al. (2025b)* menggunakan pendekatan multimodal SMFL berbasis EmoBERTa, audio, dan video dan meraih *F1-score* 89% pada dataset DAIC-WOZ [10]. Sementara itu, *Anantharaman et al. (2022)* menggunakan BERT untuk klasifikasi multi-kelas depresi pada teks media sosial dan memperoleh *F1-score* 41,2% pada dataset LT-EDI [11].

Disisi lain, pendekatan berbasis *domain-specific model* seperti MentalBERT telah menunjukkan performa yang unggul dalam deteksi depresi. *Tavchioski et al. (2023)* memanfaatkan MentalBERT yang telah di *pretraining* pada data dari *subreddit* bertema kesehatan mental untuk mendeteksi depresi dalam postingan media sosial. Model ini diuji pada dua dataset, yaitu Reddit dan Twitter, dan mencapai *F1-score* sebesar 57,70% pada dataset Reddit dan 83,1% pada dataset Twitter [12]. *Shatte et al. (2023)* menggunakan

MentalBERT yang dilatih pada data kesehatan mental dari Reddit dan memperoleh *F1-score* 76% dalam mendeteksi gejala depresi dan ide bunuh diri yang menunjukkan keunggulan model yang dilatih khusus pada domain tersebut [13].

Meskipun model BERT telah banyak digunakan, terdapat keterbatasan dalam menangani input teks panjang seperti transkrip penuh dari sesi wawancara dalam DAIC-WOZ. Banyak studi sebelumnya memanfaatkan BERT untuk klasifikasi berbasis kalimat atau sesi pendek sehingga konteks percakapan yang lebih luas sering kali diabaikan [14]. Selain itu, belum ditemukan penelitian yang secara eksplisit menerapkan MentalBERT dalam skenario *fine-tuning* berbasis teks transkrip pada dataset DAIC-WOZ dengan pendekatan segmentasi yang mempertahankan makna semantik dari percakapan klinis. Dalam studi oleh *Archibald et al. (2024)* menunjukkan bahwa segmentasi semantik dapat secara signifikan meningkatkan performa model dalam klasifikasi dokumen panjang, terutama ketika digunakan sebagai tahap praproses sebelum ekstraksi fitur atau *embedding* [15]. Dengan memisahkan bagian dokumen yang bermakna, seperti teks tercetak, tulisan tangan, dan elemen struktur, model dapat lebih fokus pada fitur yang relevan dan mengabaikan *noise* yang tidak informatif.

Penelitian ini berfokus pada pengembangan sistem deteksi depresi berbasis teks dengan memanfaatkan transkrip teks lengkap dari dataset DAIC-WOZ. Seluruh respon partisipan dalam setiap sesi wawancara digabungkan kemudian disegmentasi secara semantik untuk membentuk unit teks bermakna yang tidak melampaui batas panjang token input model. Setiap segmen tersebut diproses menggunakan model *transformer*, khususnya MentalBERT yang telah dilakukan *pretrained* pada korpus percakapan kesehatan mental dari Reddit. Proses *fine-tuning* dilakukan menggunakan model MentalBERT untuk mengevaluasi efektivitas pendekatan segmentasi semantik dalam meningkatkan performa deteksi depresi [13].

Dengan mengeksplorasi penggunaan model BERT yang telah disesuaikan

dengan domain klinis serta menerapkan pendekatan segmentasi semantik sebagai solusi terhadap keterbatasan panjang input, penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan sistem deteksi depresi yang siap diintegrasikan ke dalam aplikasi praktis seperti *chatbot* atau sistem skrining awal.

1.2. Rumusan Masalah

Masalah utama yang dirumuskan dalam penelitian ini adalah:

1. Bagaimana performa model MentalBERT dalam mendeteksi indikasi depresi berbasis teks menggunakan transkrip wawancara dari dataset DAIC-WOZ?
2. Apakah segmentasi semantik terhadap transkrip wawancara dapat meningkatkan performa model deteksi depresi dibandingkan pendekatan baseline seperti kalimat tunggal atau keseluruhan sesi?
3. Bagaimana pengaruh kombinasi *fine-tuning* MentalBERT dengan teknik segmentasi semantik terhadap metrik evaluasi seperti *precision*, *recall*, dan *F1-score* dalam klasifikasi depresi?

1.3. Tujuan dan Manfaat

Penelitian ini bertujuan untuk mengembangkan dan menguji sistem deteksi depresi berbasis *fine-tuning* menggunakan model MentalBERT terhadap data percakapan yang diperoleh dari *chatbot* berbasis teks. Sistem ini dirancang untuk melakukan klasifikasi biner terhadap kondisi depresi, yaitu mengidentifikasi apakah seseorang termasuk dalam kategori depresi atau non-depresi berdasarkan pola bahasa dalam percakapannya. Tujuan dari penelitian ini meliputi:

1. Menganalisis kinerja model MentalBERT dalam tugas klasifikasi depresi berbasis teks menggunakan transkrip wawancara dari dataset DAIC-WOZ.
2. Mengkaji efektivitas segmentasi semantik terhadap transkrip wawancara dalam meningkatkan performa klasifikasi model

dibandingkan pendekatan baseline seperti per kalimat atau per sesi penuh.

3. Mengevaluasi pengaruh kombinasi antara *fine-tuning* MentalBERT dan segmentasi semantik terhadap metrik evaluasi, khususnya *precision*, *recall*, dan *F1-score* dalam tugas klasifikasi depresi berbasis teks.

Adapun manfaat dari penelitian ini di antaranya adalah:

1. Memberikan kontribusi terhadap pengembangan metode deteksi depresi berbasis NLP khususnya dalam pemanfaatan model BERT dan pendekatan segmentasi semantik.
2. Menambah kajian ilmiah mengenai efektivitas segmentasi semantik dalam meningkatkan representasi konteks percakapan dan akurasi klasifikasi pada model berbasis transformer.
3. Menjadi dasar bagi pengembangan sistem skrining depresi yang dapat diimplementasikan dalam layanan berbasis *chatbot* untuk membantu deteksi dini gangguan kesehatan mental.

Seluruh eksperimen dan evaluasi dalam penelitian ini dirancang untuk menjawab tujuan-tujuan tersebut, sehingga hasil yang diperoleh dapat digunakan sebagai dasar kesimpulan akhir yang relevan dan terukur.

1.4. Batasan Masalah

Dalam penelitian ini dilakukan beberapa penyederhanaan untuk memastikan ruang lingkup pekerjaan dapat diselesaikan dalam waktu dan sumber daya yang tersedia. Adapun batasan-batasan yang ditetapkan dalam penelitian ini adalah sebagai berikut:

1. Klasifikasi hanya dilakukan pada dua label (biner) yaitu depresi dan non-depresi, tanpa mengukur tingkat keparahan depresi secara bertingkat seperti pada PHQ-8. Pendekatan klasifikasi biner dipilih karena lebih sesuai untuk skenario deteksi awal dalam sistem *chatbot*

dimana identifikasi keberadaan gangguan menjadi prioritas utama sebelum mengarah ke asesmen lanjutan terkait tingkat keparahan.

2. Data yang digunakan berasal dari dataset transkrip teks DAIC-WOZ dengan total 189 partisipan. Pemilihan dataset ini dilakukan karena merupakan salah satu landasan dalam penelitian deteksi depresi berbasis teks meskipun jumlah partisipannya terbatas.
3. Model yang diuji dibatasi pada satu model *transformer* saja yaitu MentalBERT. Pemilihan model ini dilakukan untuk membandingkan kinerja model umum dan model yang telah dilatih khusus pada korpus kesehatan mental. Model lain seperti BERT, RoBERTa, DistilBERT, atau LLaMA tidak diuji karena tidak memiliki *pretraining* khusus pada data kesehatan mental dan diluar cakupan tujuan penelitian.
4. Ruang lingkup penelitian tidak mencakup pembangunan sistem *chatbot* interaktif. Istilah *chatbot* berbasis teks dalam penelitian ini merujuk pada konteks dan struktur percakapan dalam data DAIC-WOZ yang merepresentasikan interaksi antara pengguna dan agen virtual. Fokus penelitian diarahkan pada evaluasi model deteksi depresi dari segmen percakapan, bukan pada pengembangan atau pengujian langsung melalui integrasi *chatbot*.
5. Evaluasi model dibatasi pada pengukuran kinerja melalui metrik klasifikasi umum seperti akurasi, presisi, *recall*, dan *F1-score*. tanpa melibatkan uji coba berbasis interaksi manusia secara *real-time* atau implementasi dalam lingkungan produksi.

Batasan-batasan ini diambil untuk menyederhanakan proses penelitian agar dapat diselesaikan tepat waktu dan dapat menjadi dasar pengembangan lebih lanjut dalam tugas akhir atau penelitian berikutnya.

1.5. Metode Penelitian

Penelitian ini dilakukan dengan pendekatan studi literatur, pengolahan data, implementasi model *deep learning* berbasis BERT, serta evaluasi

pengujian untuk menghasilkan model deteksi depresi berbasis teks. Langkah-langkah penelitian dilakukan sebagai berikut:

1. Studi Literatur

Dilakukan kajian terhadap berbagai sumber ilmiah terkait deteksi depresi berbasis teks dikaji secara mendalam, mencakup arsitektur transformer seperti BERT dan varian khusus domain seperti MentalBERT. Penekanan diberikan pada teknik *fine-tuning*, segmentasi input teks panjang, serta metode penanganan ketidakseimbangan kelas. Hasil studi literatur menyediakan landasan teori dan justifikasi metodologis untuk menyusun eksperimen selanjutnya.

2. Pengumpulan dan Pra-Pemrosesan Data

Transkrip teks DAIC-WOZ diunduh dan dipilah berdasarkan partisipan. Setiap partisipan diwakili oleh satu paragraf panjang hasil dari penggabungan seluruh respons selama sesi wawancara. Selanjutnya dilakukan pembersihan teks, normalisasi huruf, serta tokenisasi dasar untuk menyiapkan data input. Untuk mempersiapkan input ke dalam model, teks setiap partisipan disegmentasi secara semantik ke dalam blok sepanjang maksimum 64 token. Pendekatan ini menjaga keterpaduan konteks percakapan tanpa melampaui batas panjang input model.

3. *Fine-tuning* Model

Tahap *fine-tuning* dilakukan hanya dengan menggunakan model *transformer* MentalBERT. Seluruh eksperimen difokuskan pada model ini tanpa melibatkan model lain untuk menjaga arah penelitian tetap konsisten dan efisien. Penyesuaian *hyperparameter* dilakukan untuk mencari konfigurasi terbaik dari *MentalBERT* dalam mendeteksi depresi berbasis teks.

4. Evaluasi Model

Setiap model diuji menggunakan metrik klasifikasi biner yaitu akurasi, presisi, *recall*, dan *F1-score* dengan fokus utama pada *macro F1*-

score dan performa prediksi kelas depresi. Evaluasi dilakukan dalam dua tingkat, yaitu per blok segmentasi (unit input model) dan per sesi partisipan (agregasi prediksi blok). Hasil dari seluruh kombinasi model dan skenario *loss function* dianalisis untuk mengidentifikasi konfigurasi terbaik terhadap skenario deteksi awal gejala depresi berbasis teks.

5. Analisis Hasil

Analisis hasil dilakukan dengan meninjau performa model MentalBERT berdasarkan evaluasi per blok dan per sesi partisipan. Temuan kemudian dikaitkan kembali dengan hasil studi literatur untuk memperkuat pemahaman serta menjadi dasar dalam merumuskan rekomendasi metodologis untuk penelitian selanjutnya dibidang deteksi depresi berbasis teks.

1.6. Jadwal Pelaksanaan

Berikut merupakan jadwal kegiatan yang dibuat berdasarkan metode penelitian.

Tabel 1. 1. Jadwal pelaksanaan Tugas Akhir.

No.	Deskripsi Tahapan	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6
1	Studi Literatur						
2	Pengumpulan dan Pra-pemrosesan Data						
3	<i>Fine-tuning</i> model						
4	Analisa dan Evaluasi Hasil						
5	Penyusunan Laporan/Buku TA						