

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital di Indonesia menunjukkan peningkatan signifikan dalam penggunaan media sosial sebagai sarana komunikasi dan ekspresi diri masyarakat, khususnya generasi muda [1]. Generasi muda merupakan kelompok yang paling aktif memanfaatkan media sosial sebagai ruang interaksi sosial dan pembentukan identitas digital [2], [3]. Intensitas penggunaan media sosial yang tinggi berkorelasi dengan meningkatnya eksposur terhadap isu-isu psikologis, termasuk Kesehatan mental [4]. Diskursus mengenai Kesehatan mental di ruang digital semakin terbuka seiring meningkatnya literasi masyarakat terhadap isu psikologis [5]. Kondisi ini menjadikan media sosial sebagai sumber data potensial untuk analisis opini berbasis teks menggunakan pendekatan *Natural Language Processing* (NLP) [6].

Fenomena meningkatnya pembahasan mengenai stress, kecemasan, burnout, dan depresi di platform X (Twitter) menunjukkan adanya kecenderungan ekspresi emosional yang beragam dalam bentuk teks singkat [7]. Media sosial tidak hanya menjadi ruang berbagi pengalaman, tetapi juga refleksi kondisi psikososial generasi muda [8]. Namun, karakteristik Bahasa media sosial yang informal, mengandung singkatan, slang, dan campuran Bahasa menjadi tantangan dalam proses klasifikasi otomatis [6]. Oleh karena itu, diperlukan model pembelajaran mesin yang mampu memahami konteks semantik secara mendalam. Pendekatan berbasis deep learning dinilai lebih adaptif dalam menangani kompleksitas teks tidak terstruktur dibandingkan metode konvensional [9].

Penelitian terdahulu menunjukkan bahwa algoritma klasik seperti *Naïve Bayes* dan *Support Vector Machine* cukup efektif dalam teks berbahasa Indonesia, namun memiliki keterbatasan dalam memahami konteks kata secara dua arah [10]. Perkembangan model *transformer* seperti BERT menawarkan peningkatan akurasi

karena mampu memproses konteks secara *bidirectional*[11] . Adaptasi model tersebut dalam bentuk IndoBERT memungkinkan representasi Bahasa Indonesia yang lebih sesuai dengan struktur linguistic local[11]. Beberapa penelitian membuktikan bahwa IndoBERT menghasilkan performa lebih tinggi dibandingkan model *machine learning* tradisional pada tugas klasifikasi teks[12]. Meskipun demikian, kajian yang secara khusus menguji variasi *hyperparameter* seperti *learning rate* dan *optimizer* dalam analisis teks isu kesehatan mental di Indonesia masih terbatas[13].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis karakteristik dan kecenderungan opini kesehatan mental di Indonesia pada platform X menggunakan model IndoBERT [14]. Selain itu, penelitian ini menguji pengaruh variasi *learning rate* dan jenis *optimizer* terhadap kinerja model klasifikasi[15]. Evaluasi performa dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan F1-score guna memperoleh konfigurasi model yang optimal[16] . Dengan pendekatan tersebut, penelitian ini diharapkan memberikan kontribusi metodologis dalam pengembangan analisis teks bahasa Indonesia berbasis *transformer* sekaligus memberikan gambaran empiris mengenai ekspresi Kesehatan mental berbahasa Indonesia di ruang digital[17].

1.2 Rumusan Masalah

1. Bagaimana pengaruh variasi *hyperparameter*, khususnya *learning rate* dan jenis *optimizer*, terhadap kinerja model IndoBERT dalam analisis opini kesehatan mental di media social X ?
2. Sejauh mana performa algoritma IndoBERT dalam mengklasifikasikan teks kesehatan mental di media social X, ditinjau dari nilai akurasi, *precision*, *recall*, dan F1-score pada data uji?

1.3 Batasan Masalah

1. Objek penelitian dibatasi pada data teks yang berasal dari media social Twitter yang memuat pembahasan terkait isu kesehatan

mental di Indonesia.

2. Jenis analisis yang dilakukan terbatas pada analisis opini dengan dua kelas, yaitu opini positif dan opini negative, sementara opini netral tidak disertakan dalam proses analisis.
3. Metode yang digunakan dalam penelitian ini dibatasi pada algoritma IndoBERT dengan pendekatan *fine-tuning* untuk analisis teks, tanpa melakukan perbandingan dengan algoritma pembelajaran mesin atau *deep learning* lainnya.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai dalam penelitian ini adalah sebagai berikut :

1. Menganalisis pengaruh variasi *hyperparameter*, khususnya *learning rate* dan jenis *optimizer*, terhadap kinerja model IndoBERT dalam melakukan analisis opini terkait kesehatan mental Indonesia di media sosial X.
2. Mengevaluasi tingkat performa algoritma IndoBERT dalam mengklasifikasikan teks kesehatan mental Indonesia di media sosial X, dengan meninjau hasil metrik evaluasi yang meliputi *accuracy*, *precision*, *recall*, dan F1-score pada data uji.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan berbagai manfaat yang signifikan, baik secara teoritis dan praktis dengan rincian sebagai berikut:

1. Memberikan kontribusi akademik dalam pengembangan studi analisis teks berbasis model *transformer*, khususnya IndoBERT, pada teks berbahasa Indonesia. Melalui implementasi *fine-tuning* dan evaluasi kinerja model.
2. Penelitian ini memberikan rujukan empiris mengenai pengaruh variasi *learning rate* dan *optimizer* terhadap performa IndoBERT. Hasil *grid search* yang menemukan konfigurasi optimal dapat menjadi dasar metodologis bagi penelitian NLP berbasis *transformer* selanjutnya, khususnya dalam pengujian *hyperparameter*.

3. Hasil penelitian ini dapat menjadi dasar pengembangan system analisis otomatis untuk memantau opini public terkait kesehatan mental di media social. Model terbaik yang diperoleh berpotensi dijadikan prototipe awal sistem berbasis kecerdasan buatan guna mendukung pengambilan keputusan berbasis data.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN, berisi Latar belakang masalah, rumusan masalah, batasan masalah, tujuan masalah, manfaat penelitian baik secara teoritis maupun praktis.

BAB II TINJAUAN PUSTAKA, berisi tinjauan pustaka, dasar-dasar teori yang digunakan, dan sebagai landasan teoritis dan referensi ilmiah dalam penyusunan metode penelitian.

BAB III METODE PENELITIAN, di dalamnya menjelaskan objek yang akan menjadi acuan penelitian, alur penelitian serta alat dan bahan yang akan diterapkan selama penelitian.

BAB IV HASIL DAN PEMBAHASAN, bab ini menyajikan hasil implementasi metode, meliputi pengolahan dan distribusi data (EDA), pelatihan model serta variasi *hyperparameter*, dan evaluasi performa metrik.

BAB V PENUTUP, berisi kesimpulan hasil evaluasi dan saran pengembangan penelitian selanjutnya dari aspek dataset.