

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam peristiwa kerusuhan tragedi Mei 1998, khususnya yang melibatkan pemerkosaan massal atau kekerasan seksual yang merupakan sebuah tragedi paling traumatis di Indonesia [1]. Peristiwa tersebut terjadi di beberapa wilayah Indonesia, namun selama lebih dari dua *dekade* tidak ada penyelidikan secara resmi dan tuntas, baik terhadap pelaku maupun korban. Meskipun demikian, berbagai penelitian berusaha menganalisis berbagai faktor terkait isu tersebut dari segi peran negara dan militer sebagai aktor kerusuhan [2], serta kekerasan yang berbasis gender [3]. Seiring dengan minimnya penyelidikan dan pertanggungjawaban negara terhadap kejadian tersebut, media massa turut berperan penting dalam membentuk pemahaman publik dengan cenderung membingkai tragedi Mei 1998 sebagai krisis kemanusiaan dan kerusuhan sosial, dengan menonjolkan aspek kekacauan, korban, serta dampak setelah peristiwa [4]. Namun, isu pertanggungjawaban hukum dan kekerasan seksual tidak selalu memperoleh pemberitaan yang seimbang, sehingga narasi penderitaan korban sering kali *tereduksi*, akibatnya representasi media berkontribusi pada pembentukan ingatan *kolektif* yang *parsial* terhadap tragedi kekerasan seksual Mei 1998 [5].

Ketidakseimbangan representasi dalam media arus utama tersebut mendorong adanya pergeseran ekosistem media digital, pembentukan opini publik tidak lagi didominasi oleh media arus utama, tetapi juga berlangsung *intensif* di ruang publik digital. Salah satu platform yang berperan signifikan dalam proses ini adalah *YouTube*, yang berbasis konten video berdurasi panjang dan memungkinkan diskusi yang lebih mendalam [6]. Tingginya peran tersebut juga didukung oleh jumlah pengguna yang besar di Indonesia. Menurut laporan terbaru *We Are Social*, jumlah pengguna *YouTube* di Indonesia mencapai 139 juta per awal tahun 2024 [7]. Selain sebagai sarana distribusi konten, *YouTube* berfungsi sebagai ruang publik digital tempat wacana sosial dan politik diproduksi dan dinegosiasikan. Format video panjang seperti dokumenter dan diskusi isu sosial mendorong komentar yang

lebih *reflektif* dan emosional dibandingkan media teks atau video pendek [8]. Kondisi ini juga terlihat dalam diskusi mengenai tragedi Mei 1998 di *YouTube*. Respons publik terhadap isu kekerasan seksual dan pemerkosaan massal dalam tragedi Mei 1998 menunjukkan dinamika emosional yang kompleks, seperti marah, sedih, takut, bahagia dan cinta [9]. Dalam penelitian analisis teks, emosi umumnya dikategorikan ke dalam sentimen positif dan negatif [10]. Namun, pengelompokan tersebut memiliki keterbatasan dalam menangkap keragaman makna emosional yang muncul dalam respons publik. Oleh karena itu, penelitian ini menggunakan lima kategori emosi, yaitu marah, sedih, takut, bahagia, dan cinta, yang merepresentasikan emosi dasar dan sosial yang lazim muncul dalam respons terhadap isu kekerasan dan trauma kolektif. Pendekatan berbasis emosi *multi label* dinilai mampu menghasilkan analisis yang lebih mendalam dan *komprehensif* dibandingkan klasifikasi sentimen *biner* [11].

Untuk mengklasifikasi keragaman emosi tersebut secara sistematis, diperlukan pendekatan berbasis *machine learning* dalam pengolahan data teks. Pada tahap awal pelabelan data dilakukan melalui *clustering* menggunakan *Hierarchical Density-Based Spatial Clustering of Applications With Noise (HDBSCAN)*. Metode ini adalah pengembangan dari metode pengelompokan *DBSCAN* dengan algoritma *hierarkis* dan berbasis kepadatan. Algoritma ini memiliki keunggulan mampu menangani kluster dengan berbagai bentuk, dimensi tinggi, dan jumlah kluster yang tidak pasti. Kluster ini cocok untuk data spasial tanpa perlu menentukan jumlah parameter kluster optimal terlebih dahulu [12]. Dalam penelitian ini, metode pengelompokan *HDBSCAN* akan digunakan untuk mengelompokkan data yang akan dimanfaatkan sebagai data latih untuk mengevaluasi kinerja *BERT* dalam bahasa Indonesia dan *GRU*, pada tugas klasifikasi emosi komentar *YouTube* [13].

Metode yang digunakan dalam penelitian ini untuk mengklasifikasikan emosi adalah *Bidirectional Encoder Representations from Transformers (BERT)* dalam bahasa Indonesia, *Gated Recurrent Unit (GRU)*. Secara teoritis, *BERT* bahasa Indonesia merupakan model *pretrained* berskala besar berbasis *transformer*

yang menghasilkan representasi kontekstual dua arah melalui mekanisme *self-attention*, sehingga mampu menangkap hubungan antar kata dalam satu kalimat dan penting untuk memahami makna serta nuansa emosi pada teks digital [14][15]. Dalam analisis emosi teks, pemodelan *dependensi sekuensial* menjadi penting karena makna emosional suatu kalimat dipengaruhi oleh urutan kata dan konteks sebelumnya. Salah satu arsitektur *deep learning* yang efektif untuk data sekuensial adalah *Gated Recurrent Unit (GRU)*, yaitu varian dari *Recurrent Neural Network (RNN)* yang menggunakan mekanisme *gate* untuk mengatur aliran informasi. Arsitektur ini dirancang untuk memproses data sekuensial dengan memungkinkan informasi untuk dipertahankan atau dilupakan secara selektif seiring waktu [16].

1.2 Rumusan Masalah

Berdasar latar belakang masalah seperti di 1.1, Berikut adalah rumusan masalah untuk penelitian ini :

1. Bagaimana kualitas dari algoritma *HDBSCAN* dalam melakukan pelabelan emosi pada komentar *Youtube* tanpa menggunakan pelabelan awal, termasuk proporsi data yang teridentifikasi sebagai *noise* ?
2. Bagaimana perbandingan kinerja model algoritma *BERT* dalam bahasa Indonesia dan *GRU* dalam mengklasifikasi emosi publik pada komentar *YouTube* terkait peristiwa kekerasan seksual dalam Tragedi Mei 1998 ?

1.3 Batasan Masalah

Berdasarkan uraian latar belakang dan rumusan masalah, maka peneliti akan memfokuskan batasan masalah pada:

1. Penelitian ini mengkaji data berupa komentar publik dari platform *Youtube* yang berkaitan dengan peristiwa kekerasan seksual Tragedi Mei 1998 sebagai sumber utama dalam analisis emosi publik.
2. Analisis emosi dilakukan terhadap teks komentar, dengan penekanan pada makna kalimat yang terkandung di dalamnya.
3. Model klasifikasi emosi dalam penelitian ini menggunakan pendekatan

pembelajaran mesin dengan memanfaatkan model *BERT* berbahasa Indonesia serta *GRU* sebagai model perbandingan dalam pemrosesan teks.

4. Proses pengelompokan dan pelabelan awal di lakukan menggunakan algoritma *HDBSCAN* sebagai model clustering.
5. Kategori emosi yang digunakan mencakup emosi seperti marah, sedih, takut, bahagia, dan cinta, sebagai representasi emosi publik dalam teks komentar.

1.4 Tujuan Penelitian

Tujuan dalam penelitian ini adalah sebagai berikut:

1. Mengidentifikasi pelabelan awal menggunakan algoritma *HDBSCAN*. Melalui pendekatan ini, peneliti berusaha melihat bagaimana komentar-komentar tersebut dikelompokkan berdasarkan pola komentar dalam teks.
2. Melakukan perbandingan dalam menganalisis kemampuan model *BERT* dalam bahasa Indonesia dan *GRU* dalam mengenali dan mengklasifikasikan emosi publik pada komentar *YouTube* menggunakan pelabelan hasil dari *HDBSCAN* mengenai isu kekerasan seksual Mei 1998, melalui evaluasi beberapa metrik untuk mengevaluasi hasil model benar-benar akurat.
3. Memberikan gambaran yang lebih menyeluruh mengenai cara publik merespon isu kekerasan seksual atau pemerkosaan massa, dengan menggabungkan hasil model klasifikasi dan analisis pengelompokan.

1.5 Manfaat Penelitian

Manfaat penelitian dijabarkan menjadi dua yaitu manfaat teoritis dan manfaat praktis sebagai berikut:

1. Manfaat Teoritis

Penelitian ini diharapkan dapat bermanfaat bagi perkembangan *text mining*, terutama pada penelitian analisis sentimen emosi dalam teknik pembelajaran mesin. Prosesnya, model *BERT* dan *GRU* digunakan untuk mengklasifikasikan emosi, serta *HDBSCAN* untuk pelabelan dalam mengelompokkan emosi dari

data teks. Hasil penelitian dari penerapan klasifikasi dan *clustering* dalam teks komentar *YouTube* yang bersifat bebas dan tidak terstruktur dapat menambah pemahaman dan menjadi referensi dalam analisis teks yang lebih tepat.

2. Manfaat Praktis

- a. Mengetahui performa *BERT* dan *GRU* dalam menganalisis klasifikasi pada teks dalam bahasa Indonesia.
- b. Menemukan pola pelabelan awal teks menggunakan *HDBSCAN* dalam komentar dari *YouTube*.
- c. Hasil penelitian ini diharapkan bisa bermanfaat dalam perkembangan *text mining* dalam bidang klasifikasi emosi.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Berisi Latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian.

BAB II TINJAUAN PUSTAKA

Tinjauan pustaka berisi penelitian sebelumnya yang menjadi acuan dan teori-teori yang digunakan dalam penelitian ini.

BAB III METODE PENELITIAN

Metode penelitian menguraikan alur penelitian, data dan sumber data yang digunakan, teknik pengumpulan data, pelabelan awal, dan metode analisis yang diterapkan untuk membandingkan kinerja *IndoBERT* dan *GRU*.

BAB IV HASIL DAN PEMBAHASAN

Hasil dan pembahasan menyajikan hasil yang diperoleh dari penerapan algoritma *HDBSCAN* dan kedua model. Dilanjut dengan pembahasan mengenai hasil perbandingan kinerja antara *IndoBERT* dan *GRU*, serta faktor-faktor yang mempengaruhi algoritma.

BAB V PENUTUP

Penutup berisi kesimpulan yang diambil berdasarkan hasil penelitian dan saran untuk penelitian selanjutnya.

