

1. Pendahuluan

1.1 Latar Belakang

COVID-19 telah berkembang menjadi pandemi global yang mengubah kehidupan masyarakat di seluruh dunia. Selama pandemi, Twitter menjadi salah satu media utama untuk menyebarkan informasi terkini maupun berbagi ekspresi. Namun, pandemi COVID-19 ini menyebabkan banyak penyebaran informasi palsu dan berita bohong yang mengkhawatirkan. Hal ini berdampak pada penerimaan masyarakat terhadap klaim dan informasi palsu yang berpotensi merugikan, membuat masyarakat merasakan berbagai emosi, merasa kebingungan, panik, serta merasakan ketidakpastian dari informasi yang beredar. Maka dari itu, diperlukan adanya pendekatan yang efektif untuk mengidentifikasi dan mengatasi masalah tersebut, dengan melakukan penelitian terhadap respon yang diberikan masyarakat terhadap informasi yang beredar, sehingga dari hasil penelitian yang dilakukan dapat membantu pemerintah dalam mengatasi informasi dan berita yang beredar di media sosial, terutama yang berkaitan dengan pandemi COVID-19.

Analisa emosi pada teks atau postingan di media sosial menjadi salah satu aspek dalam memahami pandangan serta respon masyarakat terhadap pandemi COVID-19. Identifikasi emosi dapat memberikan pemahaman terkait sentimen dan reaksi masyarakat terhadap berbagai isu dan perkembangan terkini terkait pandemi. Beberapa penelitian seperti yang sebelumnya sudah dilakukan oleh Purba & Wijaya (2022) yang mengidentifikasi dan menganalisis emosi masyarakat yang paling mendominasi di media sosial Twitter selama masa pandemi COVID-19 pada tahun 2019. Penelitian ini menggunakan model emosi Ekman dan algoritma pembelajaran mendalam *Long Short Term Memory* (LSTM) dengan *Word Embedding Fasttext*. Hasil penelitian menunjukkan bahwa emosi bahagia adalah emosi yang paling dominan di kalangan masyarakat dalam situasi pandemi ini. Model LSTM-FastText yang dikembangkan juga menunjukkan kinerja yang sangat baik, dengan akurasi 99.24% dan loss 0.0264 dalam pengklasifikasian *data train*. Model ini dianggap dapat digunakan dengan baik dalam menganalisis emosi masyarakat selama pandemi COVID-19 di platform Twitter. Terdapat juga penelitian dengan menggunakan metode lain yang dilakukan Ferdiana, dkk., (2019) tentang *dataset* teks berbahasa Indonesia yang disiapkan untuk keperluan analisis teks. *Dataset* ini terdiri dari 10.806 baris data berbahasa Indonesia dari media sosial Twitter, yang telah diberi label positif, negatif, dan netral, serta 454.559 baris data mentah. Data yang sudah dilabeli telah melewati proses pembersihan untuk menghapus elemen-elemen pengganggu seperti simbol atau tautan web. *Dataset* ini telah diuji menggunakan model analisis sentimen sederhana dan dibandingkan dengan *dataset* lain untuk memastikan kecocokannya. Pengujian ini melibatkan algoritma *Support Vector Machine*, *K-Nearest Neighbor*, dan *Stochastic Gradient Descent*, dan hasilnya menunjukkan bahwa *dataset* ini cocok untuk digunakan dalam pemodelan analisis sentimen dengan akurasi yang baik. Namun, untuk mendapatkan hasil yang lebih baik, perlu untuk terus dilakukan penelitian untuk mengembangkan dan mengangkat metode terbaru yang mampu menangani berbagai hal terutama dalam mengidentifikasi emosi dari teks.

Untuk mengatasi masalah informasi palsu pada Twitter, model *transformer* yang ada dalam pemrosesan bahasa alami dapat menjadi solusi. Di antara model yang ada, model XLNet merupakan model yang unggul dalam menangani pengidentifikasian melalui teks. Metode *deep learning* seperti XLNet dapat digunakan dalam konteks ini untuk mengenali emosi dalam *tweet* terkait COVID-19 di platform media sosial seperti Twitter. Menurut Zihang Dai, Zhilin Yang, Yiming Yang, William W. Cohen (2019) XLNet adalah algoritma *deep learning* yang dapat memberikan hasil yang baik pada berbagai data teks, termasuk *tweet* terkait COVID-19 dan XLNet merupakan metode yang lebih baik daripada BERT karena mengatasi masalah ketergantungan posisi-*mask* pada BERT. Pada BERT, input diberi posisi-*mask* secara acak dan posisi-*mask* tersebut mengandung ketergantungan antara satu sama lain. Hal ini dapat menyebabkan kehilangan informasi penting dalam representasi teks. Sehingga, XLNet menggunakan metode *autoregressive* untuk memodelkan konteks secara *bidireksional* dan menghindari ketergantungan posisi-*mask* [12]. Selain itu, XLNet juga menggunakan faktorisasi permutasi untuk memaksimalkan informasi dari semua urutan faktorisasi, yang tidak dilakukan oleh BERT [18]. Integrasi ide dari model Transformer-XL juga memberikan keunggulan pada XLNet [27]. Hasil pengujian menunjukkan bahwa XLNet lebih unggul dari BERT dalam berbagai tugas seperti analisis sentimen dan inferensi bahasa alami [18].

Berdasarkan dari penjelasan-penjelasan yang sudah dijabarkan sebelumnya, dalam penelitian ini, dipilih metode XLNet sebagai pendekatan utama untuk mengidentifikasi emosi dalam *tweet-tweet* yang terkait dengan pandemi COVID-19 dalam bahasa Indonesia, serta nilai akurasi yang didapatkan dari metode XLNet (*Extended Transformer-Based Language Model*) yang akan digunakan untuk penelitian ini. Teks yang dikumpulkan didapatkan dari Twitter dengan menggunakan *Twint Python*, serta memberikan beberapa batasan seperti kata kunci dan jua jangka waktu data yang akan dikumpulkan.

1.2 Topik dan Batasannya

Berdasarkan latar belakang yang telah dijabarkan sebelumnya, maka dapat disimpulkan beberapa permasalahan yaitu cara mengimplementasikan metode XLNet dalam mengidentifikasi emosi terkait dengan COVID-19 dalam data teks yang berasal dari media sosial Twitter, lalu bagaimana hasil dari performansi XLNet

dalam mengidentifikasi emosi masyarakat dalam membahas topik COVID-19 di platform Twitter, dan juga bagaimana hasil analisis performansi metode XLNet dalam mengidentifikasi emosi pada topik COVID-19.

Terdapat beberapa batasan masalah dari penelitian yang dilakukan, batasan-batasan tersebut yaitu pengkategorian emosi dibagi menjadi beberapa jenis yaitu *Joy*, *Fear*, *Anger*, *Sadness*, *Surprise*, *Disgust* serta *dataset* yang digunakan merupakan *dataset* yang dikumpulkan dari *tweet* yang dipublikasikan oleh pengguna Twitter menggunakan Bahasa Indonesia berjumlah 12.000 data dengan batasan waktu Mei hingga Juli Tahun 2020.

1.3 Tujuan

Pada penelitian ini memiliki beberapa tujuan, yaitu: menganalisis bagaimana metode XLNet bekerja jika digunakan untuk mengidentifikasi emosi dalam membahas topik COVID-19 di *platform* Twitter, memperoleh hasil performansi XLNet dalam mengidentifikasi emosi masyarakat dalam membahas topik COVID-19 di *platform* Twitter, serta memperoleh hasil analisis dari performansi penelitian yang dilakukan.

1.4 Organisasi Tulisan

Penelitian ini dibagi menjadi beberapa bagian, yaitu pendahuluan, studi terkait, sistem yang dibangun, evaluasi, dan kesimpulan. Pada bagian pendahuluan menjelaskan pengenalan tentang penelitian yang dilakukan, studi terkait berisi tentang materi pendukung untuk topik penelitian yang sedang dikerjakan, sistem yang dibangun menjelaskan tentang gambaran alur sistem yang digunakan dalam penelitian, evaluasi menjelaskan tentang hasil pengujian dalam melakukan penelitian, serta kesimpulan yang memaparkan tentang inti dari penelitian yang dilakukan.