

BAB I PENDAHULUAN

1.1. Latar Belakang

Analisis sentimen telah menjadi salah satu aspek penting dalam *Natural Language Processing* (NLP) karena kemampuannya untuk memahami opini dan emosi yang terkandung dalam teks. Dalam era digital ini, dimana sebagian besar interaksi manusia dan informasi disampaikan melalui teks, analisis sentimen memainkan peran yang semakin signifikan. Teknologi ini telah diterapkan dalam berbagai bidang, seperti layanan pelanggan, pemasaran, kesehatan mental, hingga pengelolaan media sosial. Sebagai contoh, perusahaan dapat menggunakan analisis sentimen untuk mengevaluasi kepuasan pelanggan dari ulasan produk atau layanan mereka. Di bidang kesehatan mental, analisis teks dapat membantu mengidentifikasi potensi masalah emosional yang mungkin dialami individu, sehingga memungkinkan intervensi dini. Dalam konteks media sosial, analisis sentimen digunakan untuk memahami dinamika emosi masyarakat secara *real-time*, yang berguna dalam menangani krisis atau memantau opini publik terhadap isu tertentu [1][2].

Selain itu, analisis emosi dalam teks juga mulai menarik perhatian di sektor pendidikan. Penelitian menunjukkan bahwa emosi memengaruhi proses pembelajaran, sehingga pemahaman terhadap emosi siswa melalui analisis teks dapat menjadi alat yang bermanfaat untuk meningkatkan pengalaman belajar. Sebagai contoh, guru dapat menggunakan informasi dari analisis sentimen untuk menyesuaikan metode pengajaran sesuai dengan kebutuhan emosional siswa [3]. Hal ini menunjukkan bahwa penerapan analisis sentimen tidak terbatas pada sektor komersial atau sosial, tetapi juga memiliki potensi besar dalam mendukung pengembangan pendidikan yang lebih inklusif dan adaptif.

Dalam pengembangannya, analisis sentimen telah berevolusi dari sekadar klasifikasi polaritas (positif, negatif, netral) menuju analisis emosi yang lebih kompleks dan mendalam. Pada dasarnya, emosi merupakan respons emosional individu terhadap suatu topik atau isu tertentu yang dalam konteks ini diekspresikan melalui teks. Misalnya, model NLP modern kini mampu mendeteksi emosi spesifik seperti marah, senang, sedih, atau takut [4]. Penelitian mendasar oleh Shaver et al.

(2001) mengidentifikasi lima kategori dasar emosi dalam leksikon Bahasa Indonesia, yaitu cinta, kebahagiaan, kemarahan, ketakutan, dan kesedihan, yang menjadi landasan bagi banyak studi klasifikasi emosi [19]. Namun, pengembangan model analisis sentimen berbasis emosi dalam Bahasa Indonesia masih menghadapi banyak kendala. Salah satu tantangan utama adalah keterbatasan dataset yang representatif, yang dapat memengaruhi akurasi dan generalisasi model. Selain itu, struktur Bahasa Indonesia yang kaya akan variasi, termasuk penggunaan kata serapan, idiom, dan gaya bahasa informal, menambah kompleksitas dalam pengembangan model. Di sisi lain, kurangnya model yang dirancang khusus untuk Bahasa Indonesia membuat penerapan teknologi ini sering kali kurang optimal dibandingkan bahasa-bahasa lain seperti Inggris atau Mandarin [5].

Selain permasalahan terkait pengembangan model *Natural Language Processing* (NLP) untuk bahasa Indonesia yang kompleks, terdapat permasalahan terkait integrasi emoji dalam analisis teks. Emoji telah menjadi fenomena penting dalam komunikasi digital, berfungsi sebagai alat ekspresi emosional yang mampu memperkuat, mengubah, atau menciptakan ambiguitas makna. Penggunaannya lazim di platform media sosial seperti Twitter, Instagram, dan WhatsApp. Sebagai contoh, frasa "Aku terharu banget" yang disertai emoji menangis (😭) dapat diinterpretasikan secara ambigu bergantung pada konteks dan tujuan pengirim. Emoji tidak hanya berperan sebagai pelengkap teks, tetapi juga mampu memodifikasi makna emosional pesan. Namun, ketidaksesuaian antara emoji dan konteks teks dapat menyebabkan ketidakjelasan interpretasi. Analisis sentimen teks berbahasa Indonesia sering kesulitan memahami relasi antara teks dan emoji, terutama dalam konteks budaya yang beragam, sehingga menuntut pendekatan yang memadukan aspek linguistik, budaya, dan teknis.[6] Variasi interpretasi emoji juga dipengaruhi faktor budaya, seperti penggunaan emoji menangis (😭) yang tidak hanya merepresentasikan kesedihan, tetapi juga kebahagiaan ekstrem dalam situasi tertentu. Fenomena ini mempertegas perlunya pertimbangan konteks budaya dalam analisis emosi berbasis teks dan emoji, mengingat kompleksitas bahasa Indonesia yang kaya akan kata serapan, idiom, dan gaya bahasa informal yang sering dipadukan dengan emoji.

Berdasarkan kompleksitas yang telah diuraikan mulai dari keunikan Bahasa Indonesia, ambiguitas emoji, hingga perlunya fokus pada kategori emosi yang terdefinisi mendasari perlunya penetapan batasan masalah yang jelas dalam penelitian ini. Fokus penelitian pada teks Bahasa Indonesia dipilih karena adanya tantangan spesifik seperti struktur bahasa yang kompleks dan minimnya model yang dioptimalkan untuk bahasa ini, sehingga memerlukan pendekatan yang terspesialisasi [5]. Pemilihan lima kategori emosi (marah, takut, senang, cinta, dan sedih) didasarkan pada kerangka kerja leksikon emosi Bahasa Indonesia yang telah divalidasi oleh penelitian sebelumnya, yang juga digunakan sebagai label dalam dataset penelitian ini [18][19]. Sementara itu, penggunaan emoji yang terbatas pada daftar spesifik (pendukung dan bertentangan) merupakan langkah metodologis untuk mengontrol variabel. Mengingat makna emoji bisa sangat ambigu dan bergantung pada konteks, pembatasan ini memungkinkan model untuk mempelajari hubungan antara teks dan sinyal emosional dari emoji secara lebih terfokus tanpa terganggu oleh variasi interpretasi yang tidak terbatas [6].

Salah satu pendekatan yang menjanjikan untuk mengatasi tantangan ini adalah penggunaan algoritma hybrid CNN-BiLSTM. *Kombinasi Convolutional Neural Networks (CNN) dan Bidirectional Long Short-Term Memory (BiLSTM)* memungkinkan model untuk menangkap fitur lokal dalam teks sekaligus memahami hubungan temporal dan konteks kata. CNN unggul dalam mengekstrak pola-pola spesifik dalam data teks, sedangkan BiLSTM mampu mempertahankan informasi penting dalam jangka panjang. Integrasi kedua pendekatan ini telah terbukti meningkatkan akurasi dalam berbagai aplikasi analisis sentimen [7]. Dalam penelitian yang relevan, pendekatan ini telah menunjukkan hasil yang baik dalam menangkap hubungan temporal pada dataset yang kompleks [8]. Meskipun demikian, penerapan algoritma hybrid ini untuk teks dalam Bahasa Indonesia masih memerlukan eksplorasi lebih lanjut. Misalnya, Das dan Pradhan (2024) menyoroti pentingnya eksplorasi fitur lokal pada teks dengan pendekatan CNN-BiLSTM untuk meningkatkan akurasi analisis [9]. Di sisi lain, Kukreja dan Kundra (2024) menunjukkan bahwa model ini juga memerlukan optimalisasi parameter untuk meningkatkan performa dalam data yang berbasis media sosial [10].

Penelitian ini bertujuan untuk mengembangkan model NLP berbasis emosi untuk analisis sentimen teks dalam Bahasa Indonesia menggunakan algoritma hybrid CNN-BiLSTM. Melalui pendekatan ini, diharapkan dapat dihasilkan model yang tidak hanya memiliki akurasi tinggi, tetapi juga mampu mengakomodasi keunikan dan kompleksitas Bahasa Indonesia. Selain itu, hasil penelitian ini diharapkan dapat memberikan kontribusi signifikan pada pengembangan teknologi NLP lokal. Potensi penerapannya meliputi berbagai sektor, termasuk pendidikan, layanan pelanggan, kesehatan mental, dan media sosial. Dengan demikian, penelitian ini tidak hanya berkontribusi pada kemajuan teknologi tetapi juga memberikan dampak sosial yang lebih luas melalui pengembangan aplikasi berbasis NLP [11].

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas, diperoleh rumusan masalah sebagai berikut:

1. Bagaimana kemampuan algoritma hybrid CNN-BiLSTM dalam mengatasi tantangan unik struktur bahasa Indonesia?
2. Apa dataset yang representatif untuk Bahasa Indonesia, dan bagaimana mengatasi keterbatasan data?
3. Bagaimana performa model dalam mengklasifikasikan berbagai kategori emosi?

1.3. Tujuan Penelitian

Penelitian ini bertujuan untuk implementasi model analisis sentimen berbasis emosi untuk teks dalam Bahasa Indonesia menggunakan algoritma *hybrid* CNN-BiLSTM. Tujuan utama dari penelitian ini adalah untuk meningkatkan akurasi analisis sentimen emosi dengan memanfaatkan kemampuan CNN dalam mengenali pola-pola lokal dalam teks dan BiLSTM dalam memahami hubungan temporal dan konteks antar kata.

Secara lebih spesifik, penelitian ini memiliki beberapa tujuan sebagai berikut:

1. Implementasi Mengeksplorasi kemampuan algoritma hybrid CNN-BiLSTM dalam mengatasi tantangan unik struktur bahasa Indonesia, seperti kekayaan morfologi, sintaksis, dan semantik.

2. Mengidentifikasi dan memanfaatkan dataset yang representatif untuk Bahasa Indonesia, sekaligus mengatasi keterbatasan data dengan teknik augmentasi atau metode lain yang relevan.
3. Menguji dan mengevaluasi performa model yang diusulkan pada berbagai kategori emosi.

Melalui pencapaian tujuan-tujuan tersebut, penelitian ini diharapkan mampu menjadi solusi yang efektif dalam mendorong inovasi NLP untuk Bahasa Indonesia..

1.4. Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat signifikan baik dari segi teoretis maupun praktis, yang dapat dirinci sebagai berikut:

1.4.1. Manfaat Teoretis

Penelitian ini berkontribusi pada literatur ilmiah mengenai analisis sentimen berbasis emosi, terutama dalam pengembangan model NLP yang dirancang khusus untuk Bahasa Indonesia. Temuan dari penelitian ini akan memperkaya wawasan mengenai efektivitas pendekatan *hybrid* CNN-BiLSTM dalam menangani bahasa dengan struktur kompleks seperti Bahasa Indonesia. Model yang dihasilkan dapat menjadi referensi untuk penelitian lanjutan di bidang NLP, khususnya dalam pengembangan metode analisis sentimen berbasis emosi yang lebih presisi dan aplikatif.

1.4.2. Manfaat Praktis

Penelitian Penelitian ini memiliki manfaat praktis yang luas dalam berbagai bidang. Dalam layanan pelanggan, model yang diusulkan dapat digunakan untuk menganalisis kepuasan pelanggan secara otomatis, sehingga perusahaan dapat mengambil keputusan strategis berdasarkan opini pelanggan. Di bidang pendidikan, model ini dapat membantu pengajar memahami kondisi emosional siswa melalui analisis respons atau umpan balik, sehingga proses pembelajaran menjadi lebih personal dan efektif.

Dalam konteks kesehatan mental, model ini berpotensi diterapkan untuk mendeteksi emosi negatif seperti stres atau depresi dari teks yang dihasilkan oleh pengguna, misalnya di media sosial atau survei. Selain itu, pada ranah media sosial dan analisis sosial, model ini dapat digunakan untuk memahami

dinamika emosi masyarakat secara real-time, yang sangat berguna dalam penelitian sosial, pengelolaan krisis, atau bahkan pengembangan kebijakan publik.

1.4.3. Manfaat Lokal

Dengan berfokus pada Bahasa Indonesia, penelitian ini akan memberikan dampak langsung pada pengembangan teknologi yang lebih inklusif, relevan, dan sesuai dengan kebutuhan masyarakat Indonesia. Hal ini diharapkan dapat mempercepat adopsi teknologi NLP dalam berbagai sektor di Indonesia.

Secara keseluruhan, penelitian ini tidak hanya diharapkan meningkatkan akurasi analisis sentimen berbasis emosi, tetapi juga menjadi pijakan penting dalam pengembangan aplikasi berbasis NLP yang mendukung berbagai inovasi di Indonesia.

1.5. Batasan Masalah

Penelitian ini memiliki beberapa batasan yang harus diperhatikan, yaitu:

- a. Fokus Bahasa: Penelitian hanya berfokus untuk analisis emosi dalam teks Bahasa Indonesia. Hasil penelitian tidak akan bisa langsung diterapkan untuk bahasa lain tanpa penyesuaian lebih lanjut.
- b. Kategori Emosi: Emosi yang digunakan didalam penelitian ini hanya mencakup *Happy*, *Love*, *Anger*, *Sadness*, dan *Fear*. Emosi di luar ini akan dijadikan satu ke salah satu kategori emosi tersebut.
- c. Penggunaan Emoji Terbatas: Emoji yang digunakan dalam dataset telah dipilih secara spesifik (emoji pendukung dan bertentangan) untuk setiap kategori emosi. Interpretasi emoji yang ambigu atau penggunaan emoji di luar daftar yang ditentukan mungkin tidak sepenuhnya diakomodasi oleh model.