

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa tahun terakhir, banyaknya aplikasi yang dapat digunakan, seperti analisis sentimen, pemantauan umpan balik pelanggan, analisis media sosial, dan penilaian kesehatan mental, telah membuat analisis dan deteksi emosi dalam teks menjadi topik yang sangat menarik. Meskipun sulit, mengidentifikasi emosi seseorang melalui dokumen teks yang mereka tulis adalah penting dalam berbagai situasi. Pengenalan konten emosional dalam teks sangat penting dalam interaksi manusia-komputer [1]. Ekspresi emosi dapat diungkapkan melalui kata-kata yang diucapkan, bahasa tertulis, dan ekspresi wajah. Namun, meskipun banyak penelitian telah berkonsentrasi pada pengenalan emosi dari ucapan dan ekspresi wajah, para peneliti masih membutuhkan lebih banyak perhatian untuk mendeteksi emosi melalui teks [2]. Dengan memahami emosi yang terkandung dalam teks, kita bisa mendapatkan pemahaman mendalam tentang preferensi, pendapat, dan keadaan mental pengguna. Hal ini sangat membantu perusahaan, peneliti, dan pembuat kebijakan membuat keputusan yang lebih baik [3].

Kondisi emosional manusia terdiri dari emosi seperti *happy*, *sad*, *angry*, *fearful*, takut, dan sebagainya. Ekman [4] membagi emosi menjadi enam kategori utama: *joy*, *sadness*, *fear*, *surprise*, *rage*, dan *disgust*. Secara umum, emosi seseorang dapat diekspresikan melalui gerak tubuh, ekspresi wajah, suara, dan kata-kata tertulis [5]. Namun, karena teks tertulis tidak memiliki nuansa suara, mereka sering kali tidak dapat mengekspresikan emosi dengan jelas. Teks yang kompleks dan ambigu membuatnya lebih sulit untuk membedakan emosi. Karena kata-kata memiliki banyak arti dan bentuk morfologis yang berbeda, sulit untuk menentukan suasana hati dalam teks. Oleh karena itu, fokus utama dalam bidang *natural language processing* (NLP) dan *artificial intelligence* (AI) adalah pembuatan model yang dapat diandalkan dan efektif untuk analisis dan deteksi emosi [6].

Artificial Neural Networks (ANN), *Natural Language Processing* (NLP), *Deep Learning* (DL), dan komputasi paralel adalah beberapa teknik yang telah digunakan untuk mengolah dan menganalisis sejumlah besar data, serta mengekstrak informasi

emosional darinya [7]. Salah satu aspek pengolahan bahasa natural adalah ekstraksi emosi dari teks, yang mencakup menemukan dan membagi konten emosional dalam data tertulis. Tujuannya adalah untuk mendapatkan interpretasi tentang perasaan, perasaan, atau kondisi efektif yang diungkapkan dalam teks. Bidang-bidang seperti kesehatan mental, layanan pelanggan, pemasaran, dan analisis media sosial sangat bergantung pada teknologi ini karena memberikan pemahaman yang lebih baik tentang bagaimana orang merasakan dan bereaksi dalam berbagai situasi. Banyak peneliti telah mengusulkan berbagai teknik untuk menemukan emosi dalam teks. Model berbasis kata kunci, afinitas leksikal, metode pembelajaran, dan kombinasi dari semua ini termasuk dalam kategori ini [8]. Afinitas leksikal dan kata kunci adalah dua metode utama yang digunakan dalam pendekatan berbasis aturan pada awalnya. Namun, dengan hasil yang lebih memuaskan dan akurasi yang lebih baik, pendekatan berbasis pembelajaran akhirnya mulai mendominasi. Metode ini mendeteksi emosi dalam teks dengan menggunakan berbagai model. Beberapa peneliti sekarang telah menggabungkan metode ini untuk membuat model hibrida yang lebih akurat dalam deteksi. Studi menunjukkan bahwa, untuk data teks dalam skala besar, model pembelajaran mendalam cenderung memberikan hasil yang lebih akurat dibandingkan dengan model pembelajaran mesin. Namun, model pembelajaran mesin seringkali lebih andal untuk data yang lebih kecil. Namun, belum ada metode yang secara umum berhasil untuk mengidentifikasi emosi secara menyeluruh dari teks [9].

Rumah pintar, kota pintar, dan analisis teks di internet adalah beberapa contoh aplikasi praktis yang menggunakan kecerdasan buatan [10]–[12]. Untuk mendeteksi dan menganalisis emosi dalam teks, teknik pembelajaran mesin konvensional seperti pohon keputusan, *naive Bayes*, dan *support vector machine* (SVM) telah digunakan [11]. Teknik-teknik ini mewakili teks masukan melalui *bag-of-words*, *n-grams*, dan leksikon sentimen yang dibuat secara manual. Namun, teknik ini seringkali gagal menangkap ketergantungan yang kompleks dan jangka panjang yang ada dalam bahasa alami, yang mengakibatkan kinerja yang kurang memuaskan. Teknik *deep learning* seperti *convolutional neural networks* (CNNs) dan *recurrent neural networks* (RNNs) telah digunakan untuk menganalisis dan mendeteksi emosi dalam teks untuk mengatasi masalah ini. Metode ini dapat

menemukan representasi teks yang bermakna secara otomatis dengan mempelajari atribut tingkat tinggi dari data masukan. Kemampuannya untuk menangkap ketergantungan jarak jauh dalam urutan data telah membuat jaringan *long-short-term memory* (LSTM), jenis khusus RNN, sangat populer di antara berbagai teknik *deep learning* [13]. Namun demikian, model LSTM konvensional biasanya memproses teks masukan secara searah, mengabaikan potensi pengaruh konteks masa depan terhadap keadaan emosional saat ini.

Beberapa penelitian telah menggunakan model *deep learning* untuk analisis emosi. Salah satunya menggunakan (CNN-BiLSTM) dengan data emosi berlabel *sadness, anger, fear, love, joy, dan surprise*. Metode ini juga memanfaatkan *word2vec* untuk merepresentasikan kata-kata, mencoba kinerja model dalam memanfaatkan representasi kata tersebut. Hasilnya menunjukkan akurasi mencapai 92.85% dengan *word2vec* dan 86.15% tanpa [14]. Sejalan dengan itu, artikel lain mengusulkan teknik analisis sentimen *CNN-BiLSTM* dengan mekanisme perhatian, menggabungkan CNNs dan Bi LSTM untuk penambahan informasi temporal dan semantik data dengan akurat. Verifikasi pada data opini publik Weibo menunjukkan akurasi lebih tinggi dibandingkan dengan model benchmark [15]. Penelitian lain mengusulkan model Bi-LSTM untuk analisis emosi, menggunakan RNNs untuk konteks masa lalu dan masa depan dalam teks. Model ini meningkatkan interpretasi dan kinerja, mengungguli beberapa metode terkini pada dataset deteksi emosi Kaggle, termasuk CNNs, LSTM, SVM, dan *naive Bayes* [16].

Penelitian ini mengusulkan model CNN-BiLSTM baru untuk analisis dan deteksi emosi dalam teks, yang mampu menangkap konteks masa lalu dan masa depan dari teks menggunakan kekuatan RNN. Model ini meningkatkan pemahaman tentang konten emosi dengan mengintegrasikan lapisan CNN dan Bi LSTM, efektif mempelajari representasi semantik kata-kata dan ketergantungan mereka dalam sebuah kalimat. Pendekatan ini memungkinkan model untuk menangkap ketergantungan yang kompleks dan jarak jauh dalam bahasa alami, meningkatkan kinerja dalam analisis dan deteksi emosi.

Untuk meningkatkan kinerja dan interpretasi model, peneliti memperkenalkan mekanisme *attention* [17] yang menimbang kata-kata secara asimetris. Komponen tambahan ini tidak hanya meningkatkan kinerja model tetapi juga memberikan

wawasan yang berharga tentang bagian-bagian penting. Dengan menambahkan label *neutral* sebagai bagian dari penelitian ini, tujuannya adalah untuk mengidentifikasi konten atau *respons* yang tidak menunjukkan emosi yang kuat seperti *sadness, anger, fear, love, joy, dan surprise*. Hal ini dapat membantu dalam memahami bagaimana pengguna merespons konten yang netral atau tidak mengandung emosi yang dominan. Dengan penambahan ini, penelitian dapat lebih mempertimbangkan semua aspek yang berbeda untuk mendapatkan pemahaman yang lebih lengkap dan mampu menangkap spektrum emosi yang lebih luas.

1.2 Rumusan Masalah

Penelitian ini difokuskan pada pembangunan dan evaluasi model menggunakan CNN-BiLSTM dan *attention layer* untuk mengklasifikasi tujuh label emosi: *sadness, anger, joy, surprise, fear, love, dan neutral*. Penambahan label "*neutral*" pada model bertujuan menciptakan model yang lebih inklusif dan mampu mengevaluasi konten dengan lebih komprehensif, termasuk dalam mengidentifikasi variasi dari emosi yang bersifat netral.

1.3 Manfaat Penelitian

Penelitian ini bertujuan untuk mengembangkan teknik analisis emosi dari teks sebagai alat deteksi yang dapat diterapkan secara praktis dalam bidang-bidang seperti kesehatan mental dan analisis media sosial.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini mencakup penggunaan sumber data dari Kaggle yang melibatkan tujuh label emosi sebagai fokus analisis. Penelitian ini terbatas pada penerapan CNN-BiLSTM dan Attention sebagai pendekatan utama untuk mengolah data teks guna klasifikasi emosi tersebut. Pengembangan model ini ditujukan untuk mengatasi tantangan dalam mengenali variasi emosi pengguna media sosial. Perlu diperhatikan bahwa penelitian ini masih berada dalam tahap pengembangan model awal dan belum mencapai tahap implementasi.

1.5 Sistematika Penulisan

Penulisan dalam sistem ini telah disesuaikan dengan standar yang telah ditetapkan, sehingga memuat 5 topik yang berisikan ringkasan bab yaitu :

BAB I Pendahuluan

Pendahuluan, mencakup gambaran umum masalah penelitian, latar belakang, rumusan masalah dari latar belakang tersebut, batasan masalah, tujuan penelitian, dan manfaatnya.

BAB II Tinjauan Pustaka

Tinjauan pustaka, menjelaskan secara rinci metode penelitian yang digunakan dan alur sistem yang diimplementasikan.

BAB III Metode Penelitian

Metode penelitian, menjelaskan secara rinci metode penelitian yang digunakan dan alur sistem yang diimplementasikan.

BAB IV Hasil dan Pembahasan

Hasil dan pembahasan, memuat analisis dan temuan dari pengujian model CNN-BiLSTM dan *Attention* terhadap data emosi.

BAB V Penutup

Penutup, berisi kesimpulan dari penelitian ini serta rekomendasi untuk penelitian mendatang guna perbaikan lebih lanjut.