

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kemajuan teknologi yang terjadi saat ini telah mempercepat pertumbuhan media sosial, menghasilkan beragam platform microblogging yang mendorong individu untuk mengekspresikan pandangan dan emosi mereka dengan bebas kapan saja[1]. Pada Januari 2022, Pengguna media sosial di Indonesia mencapai 191,4 juta orang, hal ini setara dengan 68,9 persen dari total populasi di Indonesia. Selama periode antara tahun 2021 dan 2022, 21 juta pengguna meningkat, mencapai pertumbuhan sebesar 12,5 persen[2]. Popularitas media sosial semakin meningkat karena pengguna memiliki kesempatan untuk menerima, menciptakan, dan berbagi informasi publik dengan mudah, di mana saja[3]. Banyak platform media sosial menawarkan beragam format data, seperti teks, suara, informasi geolokasi, gambar dan video[4]. Sifat terpisah dari media sosial cenderung membuat orang mengekspresikan diri secara terbuka lebih banyak media sosial daripada secara langsung[5]. Para pengembang telah membuat berbagai aplikasi media sosial untuk mengikuti perkembangan media sosial, dan salah satunya adalah Twitter.

Twitter telah menjadi sebuah platform media sosial di mana pengguna memanfaatkannya sebagai sarana untuk berbagi informasi secara real-time dengan pengguna lain[6]. Platform Twitter menjadi wadah bagi penggunanya untuk membagikan kisah, pendapat, dan aktivitas mereka secara luas melalui pengiriman cuitan[6]. Pada setiap tweet mengandung ekspresi emosional dari pengguna twitter, dan hal ini menjadi topik kasus untuk mendeteksi kasus depresi dan menilai tingkat emosi pengguna Twitter[7][8]. Depresi adalah kondisi yang memengaruhi lebih dari empat persen populasi di seluruh dunia. Depresi jangka panjang ini dimana kesehatan serius yang secara signifikan berpengaruh terhadap kemampuan seseorang untuk menjalani aktivitas sehari-hari mereka[9], sehingga diperlukan deteksi dini terhadap depresi.

Sangat sulit untuk mendeteksi depresi dan melakukan pencegahan, sehingga dalam dekade terakhir hal ini menjadi topik deteksi depresi ini sangat menarik. Penelitian lain telah memeriksa tweet di Twitter sebagai alat indentifikasi pengguna

media sosial yang mungkin sedang dalam kondisi depresi. Terdapat Penelitian tentang deteksi depresi pengguna Twitter berkesimpulan terdapat hubungan antara sentimen yang diungkapkan dalam tweet mereka dan kondisi depresi pengguna Twitter tersebut, yang dapat mencerminkan kondisi depresi mereka[7].

Penelitian ini berfokus pada klasifikasi teks emosi dalam data Twitter Indonesia dengan menggunakan dataset "Indonesian Twitter Emotion Dataset" yang terdiri dari 4.403 baris data tweet. Dalam penelitian ini, beberapa metode machine learning, seperti Regresi Logistik, SVM, dan Random forest, dibandingkan dengan penggunaan metode ekstraksi fitur lainnya, termasuk Bag-of-Words (BOW), Word2Vec (WV), dan Fast-Text (FT). Hasil penelitian menunjukkan bahwa model dapat mencapai skor-F1 sebesar 68,39%. Selanjutnya, penelitian ini juga mengusulkan tiga metode tambahan, yaitu Emotion Lexicons, Orthographic, dan POS Tag, yang berhasil meningkatkan hasil pengujian sebesar 26,64% dibandingkan dengan tidak menggunakan ketiga metode tersebut. Dengan penggunaan metode-metode tambahan ini, skor-F1 meningkat menjadi sebesar 69,73%. Hasil penelitian ini menunjukkan bahwa metode-metode tambahan dan penggunaan berbagai teknik ekstraksi fitur dapat meningkatkan kualitas klasifikasi teks emosi dalam data Twitter Indonesia[10].

Pada penelitian lain klasifikasi tweet dengan menggunakan Deep Learning LSTM-GloVe. Dengan dataset "Indonesian Twitter Emotion Dataset" dengan jumlah 4.403 baris data tweet. Setelah memasuki proses lapisan input, tweet di-tokenisasi dan dilakukan padding nol agar memiliki panjang yang konsisten. Selanjutnya, proses indeks embeddings dilakukan dengan mengambil indeks dari kamus GloVe. Beberapa parameter yang digunakan dalam penelitian ini mencakup learning rate (laju pembelajaran) dengan dua nilai, yaitu 0,1 (half) dan 0,001, dropout dengan dua nilai, yaitu 0,25 dan "half," optimizer (pengoptimal) dengan dua jenis, yaitu SGD dan Adam. Selain itu, penelitian ini melibatkan 50 epoch dan ukuran batch sebesar 100. Hasil dari penelitian ini menunjukkan bahwa tingkat presisi (precision) sebesar 33%, tingkat recall sebesar 38%, dan skor F1 sebesar 35%. Akurasi keseluruhan adalah 46%[11]. Penelitian ini menggambarkan bagaimana model dan parameter yang digunakan dalam pengolahan data tweet mempengaruhi kinerja klasifikasi teks dan kemampuan model untuk mengenali emosi dalam tweet.

Dalam penelitian lain yang berkaitan dengan klasifikasi tweet di Twitter menggunakan dataset "Indonesian Twitter Emotion Dataset" yang terdiri dari 4.403 baris data tweet, penelitian ini menerapkan metode CNN (Convolutional Neural Network) dengan sejumlah parameter tertentu. Hal ini mencakup penggunaan dropout sebesar 50%, penerapan regularisasi L2, dan penurunan learning rate untuk Adam Optimizer menjadi 0,0003. Hasil dari penelitian ini sangat mengesankan, terutama mengingat penggunaan dataset yang relatif kecil. Diperoleh tingkat presisi sebesar 90,1%, recall sebesar 90,3%, dan skor F1 sebesar 90,2%. Selain itu, akurasi dari model ini mencapai 89,8% [12].

Penelitian lain terkait klasifikasi komentar kasar di media sosial Twitter Indonesia dengan menggunakan dataset yang terdiri dari 13.169 baris data tweet, mengusulkan penggunaan gabungan antara Recurrent Neural Network (RNN) dan Long Short-Term Memory (LSTM) untuk mengklasifikasikan data teks. Hasil dari penelitian ini menunjukkan bahwa penggabungan RNN dan LSTM mampu menghasilkan hasil yang sangat memuaskan, dengan tingkat presisi sebesar 94%, recall sebesar 95%, dan skor F1 sebesar 94%. Hasil ini menunjukkan bahwa pendekatan ini sangat efektif dalam mengidentifikasi komentar kasar di platform media sosial Twitter dan dapat digunakan untuk memitigasi dampak negatif dari perilaku tidak pantas dalam interaksi online. [13].

Kontribusi serta tujuan dari penelitian ini adalah untuk mengidentifikasi data tweet depresi pengguna twitter dengan menggunakan kombinasi dua jenis arsitektur, yakni Long Short-Term Memory (LSTM) dan Recurrent Neural Network (RNN). Penggabungan dua arsitektur tersebut diharapkan dapat meningkatkan kinerja dan performa model dalam mengidentifikasi teks tweet depresi pengguna twitter. Kedua arsitektur nantinya akan saling melengkapi dalam mengidentifikasi teks tweet, dimana RNN memetakan urutan input ke vektor dengan ukuran tetap ke dalam hidden vektor, dan LSTM membantu mengatasi peningkatan vektor pada gradient vektor. Untuk mendukung arsitektur, diterapkan handling imbalance data untuk mengatasi data yang tidak seimbang jumlahnya. Sehingga pada proses training model tidak terjadi overfit setelah dilakukan handling imbalance data.

1.2. Rumusan Masalah

Rumusan masalah penelitian ini adalah sejauh mana performa akurasi, presisi, recall, dan F1-score dapat ditingkatkan dengan menggunakan pendekatan gabungan Long Short-Term Memory (LSTM) dan Recurrent Neural Network (RNN) dalam tugas klasifikasi dan deteksi teks tweet yang mengindikasikan depresi di platform Twitter.

1.3. Tujuan Penelitian

Tujuan penelitian ini yaitu mengetahui pengaruh metode dengan menggabungkan LSTM dan RNN, melakukan klasifikasi teks tweet depresi, dan bagaimana performa gabungan metode LSTM-RNN.

1.4. Batasan Masalah

Terdapat batasan masalah yang digunakan pada penelitian sehingga penelitian tetap memfokuskan melakukan eksperimen:

- a. Penelitian ini berfokus pada pengklasifikasi teks tweet depresi yang terdiri atas dua kelas, diantaranya: *Depresi* dan *Normal*.
- b. *Dataset yang digunakan dari penelitian sebelumnya* [10],[11],[12], dimana dataset diakses melalui website [github](#) yang berisi 4,403 data tweet.