

1. Pendahuluan

Latar Belakang

Dengan adanya teknologi internet membuat banyak peranan yang belum pernah terjadi sebelumnya dalam meningkatkan jumlah pengguna media sosial. Selain itu, pengguna juga terbiasa dengan ide untuk mengekspresikan perasaan dan emosi mereka dengan orang lain dengan menggunakan platform dengan menggunakan teks [1, 2]. Fenomena ini telah menghasilkan produksi dan pembuatan berbagai macam data, yang dapat dianalisis untuk menilai sentimen sehingga bermanfaat bagi individu dan organisasi, terutama mengingat produksi data yang sangat besar ini [3]. Namun, pada penelitian [4] identifikasi dan penyaringan informasi yang ada di aplikasi media sosial untuk menganalisis sentimen merupakan tantangan. Beberapa faktor tersebut dikarenakan adanya data yang tidak terstruktur, perbedaan bahasa, keragaman situs web dan platform media sosial, dan data yang heterogen tentang pendapat individu. Salah satunya Twitter, Twitter merupakan salah satu media sosial yang banyak digunakan dalam sehari-hari di mana Indonesia menempati peringkat sembilan dunia menurut Social and Hootsuite Survei pada tahun 2019 [5]. *Tweet* pada Twitter memiliki batas 280 karakter dengan isi yang beragam serta isu dengan topik yang berbeda-beda.

Analisis sentimen merupakan metode untuk mengekstrak data teks untuk mendapatkan informasi tentang pandangan dari suatu individu atau kelompok untuk memberikan suatu penilaian atau opini pribadi terhadap isu atau kejadian yang ada. Analisis sentimen juga bisa digunakan sebagai metode untuk menganalisis data opini dan mengklasifikasikan teks yang didapatkan berdasarkan tipenya menjadi beberapa kelas yang berbeda [1, 2]. Dang dkk. [6] mengidentifikasi tiga pendekatan utama dalam analisis sentimen: berbasis pembelajaran mesin, berbasis leksikon, dan hibrida. Teknik berbasis leksikon dikategorikan menjadi dua pendekatan: berbasis korpus dan kamus. Dalam pendekatan berbasis kamus, klasifikasi sentimen dilakukan dengan memanfaatkan kamus istilah seperti yang terdapat di WordNet dan SentiWordNet. Sebaliknya, pendekatan analisis yang memiliki basis korpus bergantung pada analisis statistik konten dengan memanfaatkan metode yang terkait dengan *K-nearest Neighbors*, *Hidden Markov Models* (HMM), dan *Conditional Random Field* (CRF).

Deep learning merupakan salah satu sub bidang dalam *machine learning* yang memiliki arsitektur yang lebih mendalam dibandingkan dengan sub bidang *machine learning* yang lainnya dalam menyelesaikan suatu masalah prediksi maupun masalah klasifikasi [7]. Arsitektur umum *deep learning* adalah Deep Believe Network, Deep Neural Network, Deep Convolutional Neural Network, dan Deep Recurrent Neural Network (DRNN). *Deep learning* telah banyak digunakan untuk memecahkan berbagai permasalahan analisis sentimen, di antaranya oleh Mariel dkk. [8] dalam penelitian perbandingan algoritma *Deep Learning Neural Network* dengan *SVM* dan *Naive Bayes* untuk teks bahasa Indonesia. Dalam penelitiannya dapat dilihat hasil pengujian dengan model tersebut memperlihatkan bahwa metode terbaik menggunakan fitur ekstraksi *Bag of Word* adalah dengan menggunakan metode *Deep Learning Neural Network* yang memperoleh keakuratan sebesar 91%. Hasil yang diperoleh tersebut lebih baik dibandingkan dengan metode *Naive Bayes* yang dengan keakuratan sebesar 86% dan metode *Support Vector Machine* dengan keakuratan sebesar 89%.

Ekspansi fitur salah satu metode dalam meningkatkan akurasi, di mana pada penelitian [9, 10], ekspansi fitur dengan berbagai jenis *word embedding* membuat hasil performansi pada model klasifikasi menjadi lebih baik.

Pada tugas akhir ini, penulis menggunakan metode *Long Short-Term Memory* (LSTM) serta penambahan *word embedding* untuk ekspansi fitur pada analisis sentimen, metode ini dipilih karena LSTM merupakan salah satu jenis dari *Recurrent Neural Network* (RNN) di mana dilakukan modifikasi pada RNN dengan menambahkan *memory cell* yang bisa menyimpan informasi untuk jangka waktu yang lama. LSTM diusulkan sebagai cara untuk mengatasi terjadinya *vanishing gradient* pada RNN saat memproses data sequential yang panjang. sehingga menambah akurasi dari suatu prediksi [11].

Topik dan Batasannya

Berikut merupakan Batasan masalah dalam penelitian ini adalah sebagai berikut:

- 1.2.1 Metode *word embedding* yang digunakan yaitu *fastText* dan *Word2vec* yang digunakan sebagai ekspansi fitur menggunakan korpus *tweet* dan wikipedia.
- 1.2.2 Data yang digunakan adalah data *tweet* dengan kata kunci yang sudah ditentukan sebanyak 37.523 *tweet* menggunakan Bahasa Indonesia. data *tweet* diambil dari 4 Agustus 2022 sampai 10 Oktober 2022.

Tujuan

Tujuan dari penelitian ini adalah mendapatkan hasil kinerja yang lebih baik menggunakan metode *deep learning* LSTM serta penambahan ekspansi fitur menggunakan *Word embedding* dalam analisis sentimen.