**Deteksi Emosi pada Twitter Menggunakan Metode Long Short Term Memory (LSTM) dan Fasttext**

TUGAS AKHIR

Tugas Akhir ini sebagai salah satu syarat untuk memperoleh gelar sarjana Informatika Universitas Pembangunan Nasional “Veteran” Yogyakarta



**Disusun Oleh :**

1. **ALFA RIZA**

**NIM : 123170027**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK INDUSTRI**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”**

**YOGYAKARTA**

**2020**

**BAB 1**

**PENDAHULUAN**

* 1. **Latar Belakang**

Emosi adalah kunci perasaan dan pikiran orang (Jabreel & Moreno, 2019). Perasaan intens yang diarahkan pada sesuatu atau seseorang sebagai respon terhadap peristiwa internal maupun eksternal bagi individu digambarkan sebagai emosi (Gaind et al., 2019). Emosi bersifat umum dan penting dalam semua aspek kehidupan manusia (Fanesya et al., 2019). Deteksi emosi adalah bagian dari area yang lebih luas dari komputasi afektif bertujuan untuk memungkinkan komputer mengenali dan mengekspresikan emosi (Picard, 1997). Penelitian deteksi emosi memiliki banyak kegunaan dalam mengambil keputusan seperti di bidang pendidikan, bisnis, politik, psikologi, sehingga menunjukkan pentingnya deteksi emosi (Fanesya et al., 2019). Emosi dapat dideteksi melalui suara, ekspresi wajah, gerakan tangan, gerakan tubuh, detak jantung, tekanan darah, teks (Consoli, 2009). Teks relatif lebih mudah digunakan untuk deteksi emosi karena emosi dipicu oleh situasi tertentu serta menggambarkan emosi dari situasi tanpa kata-kata (Bata et al., 2015). Teks tidak hanya memuat informasi namun juga emosi (Hirat & Mittal, 2015). Model emosi yang sangat populer adalah model emosi Ekman. Model Ekman membagi emosi menjadi 6 label emosi yaitu bahagia, marah, takut, jijik, sedih, dan terkejut (Kowalska & Wróbel, 2017), label-label ini bersifat universal pada budaya yang berbeda. Sehingga deteksi emosi dengan data teks memiliki banyak manfaat dan akan menjadi fokus pada penelitian ini, emosi pada penelitian ini diberi 6 label yaitu bahagia, marah, takut, jijik, sedih, dan terkejut.

Saat ini teknologi berkembang pesat, media sosial menjadikan orang cenderung berekspresi emosi mereka melalui postingan teks (Saputri et al., 2019). Pada media sosial, setiap acara, berita atau aktivitas di seluruh dunia dapat dibagikan, didiskusikan, diposting dan dikomentari oleh jutaan orang (Gaind et al., 2019). Salah satu media sosial yang memiliki laju pertumbuhan pengguna tertinggi di indonesia adalah *twitter*, pengguna aktif *twitter* di Indonesia menempati posisi ketiga di Asia Pasifik dari tahun 2012 sampai tahun 2018 (Saputri et al., 2019). Maka dari itu *twitter* merupakan media sosial yang dipilih untuk dijadikan data penelitian ini.

Penelitian yang berkaitan dengan deteksi emosi pernah dilakukan dengan *Naïve Bayes* dengan fitur N-gram untuk medeteksi emosi media sosial *twitter*, hasil penelitian ini didapatkan akurasi tertinggi 55,54% (Fanesya et al., 2019). Deteksi emosi juga pernah diimplementasikan pada *facebook* dengan membandingkan leksikon dan NLP, penelitian tersebut memiliki kategori 6 emosi dan menghasilkan akurasi 55,54% dari 34.872 kata untuk lexicon dengan emolex, sedangkan NLP menghasilkan akurasi 61,53% (Rohman et al., 2019). *Support Vector Machine*(SVM) dan *K-Nearest Neighbour* juga pernah digunakan untuk mendeteksi emosi pada media sosial *twitter*, penelitian tersebut menghasilkan *precision* 45,64%, *recall* 50,20%, dan akurasi 81,04% untuk SVM, sedangkan untuk KNN menghasilkan *precision* 34,21%, *recall* 45,95%, serta akurasi 79,70% (Ardiada et al., 2019). TF-IDF juga pernah digunakan untuk mendeteksi emosi dari media sosial *facebook* dengan 365 dataset yang terdiri dari 265 data latih dan 100 data uji, penelitian tersebut menghasilkan akurasi 59% dari 6 emosi, hal ini karena jumlah data antar emosi tidak sama, sehingga peneliti menguji dengan 1 emosi saja dan menghasilkan 87,23% (Rohman et al., 2020).

Ada tantangan yang harus dihadapi dalam deteksi emosi yaitu sulitnya mengklasifikasikan data uji, karena nilai-nilai antar kelas hampir sama sehingga mempengaruhi tingkat akurasi menjadi turun, masalah ini ditemukan oleh Fanesya (2019) dalam penelitiannya untuk mendeteksi emosi pada twitter menggunakan metode *Naïve Bayes*. Data yang digunakan pada penelitian tersebut sebanyak 4396 *tweet* dengan perbandingaan data latih 80% dan data uji 20%, penelitian tersebut menghasilkan akurasi yang kurang maksimal yaitu 55,54% saja. Permasalahan tersebut dapat diatasi dengan menggunakan metode *deep learning*.

Metode *deep learning* yang dapat digunakan untuk mendeteksi emosi adalah *Recurrent Neural Network* (RNN). RNN mampu menangkap informasi tentang urutan panjang data menggunakan sel memori (Patel & Tiwari, 2019). Namun RNN memiliki kekurangan yang dinamakan *vanishing gradient*. Pada RNN, *vanishing gradient* terjadi karena nilai *gradient* bernilai 0 ataumendekati 0 seiring bertambahnya data yang dilatih RNN, karena nilai *gradient* digunakan untuk memperbaharui bobot sehingga tidak dapat memberi banyak kontribusi dan berhenti melakukan proses *learning* (Bengio et al., 1994). *Long short term memory* (LSTM) dapat mengatasi masalah tersebut dengan *memory cell* dan *gate units* (*input gate, forget gate, output gate*) sehingga LSTM dapat membaca, menyimpan, dan memperbaharui informasi (Rao & Spasojevic, 2016).

Data yang diproses oleh *deep learning* harus terlebih dahulu dikonversikan ke dalam bentuk angka, karena *deep learning* tidak dapat memproses data *string* atau teks, *string* atau teks dapat diubah menjadi vector angka menggunakan *word embedding*. Selain itu, *word embedding* ini juga mempengaruhi akurasi yang hasilkan oleh *deep learning*, seperti yang dilakukan oleh Manalu (2020) mendapatkan algoritma LSTM memiliki akurasi yang lebih baik jika menggunakan *word embedding* sebesar 86,76%, sedangkan tanpa *word embedding* hanya mendapatkan akurasi sebesar 84,14% (Utomo, 2020) .Beberapa metode *word embedding* adalah *word2vec* dan GloVe. Namun *word2vec* dan GloVe memiliki kelemahan yang dinamakan *out of vocabulary*, masalah ini muncul karena ketidakmampuan merepresentasikan vector dari kata yang belum pernah ditemui. Metode *fasttext* dapat menangani masalah tersebut, ketiga metode tersebut pernah dibandingkan untuk klasifikasi text dan menghasilkan fastext merupakan metode terbaik dengan nilai *f-measure* 0.979, sedangkan *word2vec* mendapatkan nilai *f-measure* 0,925 dan GloVe mendapatkan nilai f-measure 0,958 (Nurdin et al., 2020). Sehingga, pada penelitian ini, metode *Long Short Term Memory* (LSTM) akan menggunakan *word embedding fasttext* agar mendapatkan hasil akurasi yang maksimal.

Oleh karena itu, penelitian ini akan menerapkan *Long Short Term Memory* (LSTM) dan *word embedding* *fasttext* untuk mendeteksi emosi dengan 6 label emosi yaitu, bahagia, marah, takut, jijik, sedih, dan terkejut. Penelitian ini akan menggunakan data yang bersumber dari *twitter*. Penelitian ini diharapkan dapat menghasilkan sebuah aplikasi yang dapat mengklasifikasikan emosi menjadi bahagia, marah, takut, jijik, sedih, dan terkejut dengan lebih akurat dari beberapa penelitian sebelumnya, sehingga hasil deteksi emosi tersebut dapat berguna dalam mengambil keputusan di berbagai bidang.

* 1. **Perumusan Masalah**

Berdasarkan latar belakang yang telah dijelaskan, maka didapatkan rumusan masalah sebagai berikut:

Bagaimana tingkat akurasi algoritma *Long Short Term Memory* (LSTM) untuk mendeteksi emosi pada media sosial *twitter*?

* 1. **Batasan Masalah**

Pada penelitian ini, permasalahan dibatasi sebagai berikut:

1. Data penelitian didapatkan dari media sosial *twitter* dari tahun 2018 sampai tahun 2021.
2. Data penelitian hanya menggunakan Bahasa Indonesia.
3. Pada penelitian ini emosi yang diklasifikasi menjadi 6 label sesuai model Ekman yaitu bahagia, marah, takut, jijik, sedih, dan terkejut.
   1. **Tujuan Penelitian**

Tujuan penelitian ini adalah:

Menghitung tingkat akurasi algoritma *Long Short Term Memory*(LSTM) untuk deteksi emosi pada media sosial *twitter*.

* 1. **Manfaat Penelitian**

Manfaat yang didapatkan dari penelitian ini adalah dapat mendeteksi emosi pada media sosial *twitter* sehingga dapat digunakan mengambil keputusan di berbagai bidang.

* 1. **Metodologi Penelitian**

Alur metodologi penelitian yang dilakukan sebagai berikut:

1. Studi Literatur

Studi literatur dilakukan dengan cara mencari dan menghimpun data ataupun sumber-sumber pustaka yang dapat mendukung penelitian serta memberikan informasi untuk menyelesaikan permasalahan pada penelitian ini yang bersumber dari jurnal, artikel, skripsi, tesis dan paper yang dapat dipertanggungjawabkan.

1. Pengumpulan Data

Data yang akan digunakan bersumber dari media sosial *twitter* dan data tersebut sudah diberi label.

1. Analisis Sistem

Analisis sistem ini merupakan penganalisaan terhadap kebutuhan dalam pembuatan sistem. Sehingga dapat diketahui kebutuhan-kebutuha apa saja yang diperlukan dalam pembuatan sistem pada penelitian ini.

1. Perancangan Sistem

Pada bagian perancangan ini akan melakukan pemodelan terhadap sistem yang akan dibuat berdasarkan hasil analisis yang telah dilakukan sebelumnya.

1. Implementasi Perangkat Lunak

Tahap ini akan mengimplementasikan sistem yang sudah dirancang sebelumnya.

1. Pengujian dan Analisis

Pada tahap ini akan dilakukan pengujian terhadap implementasi yang telah dikerjakan. Kemudian akan dilakukan analisis berdasarkan hasil dari pengujian.

1. Kesimpulan dan Saran

Pada penelitian ini akan diberikan kesimpulan dari hasil penelitian yang telah dilakukan dan akan menyertakan saran yang dapat dikembangkan untuk penelitian di masa depan sehingga mendapatkan hasil yang lebih baik.

* 1. **Sistematika Penulisan**

Sistematika penulisan yang digunakan dalam menyusun laporan penelitian ini adalah sebagai berikut :

**Bab I Pendahuluan**

Pada bagian ini membahas tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

**Bab II Tinjauan Pustaka**

Tinjauan pustaka memuat tentang dasar teori yang digunakan untuk analisis dan perancangan sistem serta implementasi pada penelitian ini. Selain itu juga sebagai bahan referensi dan pondasi untuk memperkuat argumentasi dalam penelitian ini. Teori-teori yang sesuai dengan penelitian ini antara lain media sosial, deteksi emosi, jaringan syaraf tiruan, *deep learning*, *Recurrent Neural Network*(RNN) , *Long Short Term Memory*(LSTM), *word embedding*, *fasttext.*

**Bab III Analisis dan Perancangan Sistem**

Pada bagian ini akan membahas mengenai analisa dan perancangan sistem hingga implementasi sistem dalam deteksi emosi.

**Bab IV Hasil, Pengujian dan Pembahasan**

Pada bab ini akan menyajikan hasil penelitian yang berisi hasil penerapan atau implementasi dari perancangan yang telah dibuat pada bab sebelumnya. Selain itu berisi pengujian terhadap hasil penelitian beserta pembahasannya.

**Bab V Kesimpulan dan Saran**

Pada bagian ini berisi hasil akurasi metode *Long Short Term Memory* (LSTM) dan *fasttext* sehingga dapat ditarik kesimpulan. Saran juga dibahas pada bagian ini sehingga dapat dikembangkan pada penelitian berikutnya.