**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 bersifat umum dan penting dalam semua aspek kehidupan manusia (Fanesya et al., 2019). Penelitian deteksi emosi memiliki banyak kegunaan dalam mengambil keputusan seperti di bidang pendidikan untuk memprediksi keadaan emosi siswa dengan platform *e-learning* sehingga dapat digunakan untuk meningkatkan kualitas belajar (Daouas dan Lejmi, 2018), bidang bisnis digunakan untuk review kualitas (Nurlaila, Saptono, & Wiranto, 2017), dibidang politik digunakan untuk melacak dukungan poitik, dibidang psikologi digunakan untuk mengidentifikasi depresi dan kecemasan seseorang (Yuliana dan Faizah, 2018), selain itu deteksi emosi juga bermanfaat untuk rekrutmen karyawan, koseling , dan profil psikologis karena merupakan bagian dari prediksi kepribadian (Dandannavar et al., 2018), oleh karena itu deteksi emosi penting diberbagai bidang. 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 mengsgambarkan emosi dari situasi tanpa kata-kata (Bata et al., 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 dan Wróbel, 2017), label-label ini bersifat universal pada budaya yang berbeda.

Saat ini media sosial menjadikan penggunanya cenderung berekspresi emosi mereka melalui postingan teks. 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 oleh Fanesya(2019) menggunakan metode *Naïve Bayes* dengan kombinasi fitur N-gram, namun penelitian ini hanya menghasilkan akurasi tertinggi 55,54%. Penelitian yang dilakukan oleh Rohman (2019) menghasilkan akurasi kurang maksimal sebesar 55,54% dari 34.872 kata untuk lexicon dengan emolex, sedangkan NLP menghasilkan hanya akurasi 61,53%. Penelitian yang dilakukan oleh Ardiada(2019) menggunakan *Support Vector Machine*(SVM) dan *K-Nearest Neighbour* menghasilkan *precision* 45,64%, *recall* 50,20%, dan akurasi cukup bagus sebesar 81,04% untuk SVM, sedangkan untuk KNN menghasilkan *precision*, *recall*, dan akurasi yang kurang maksimal yaitu *precision* 34,21%, *recall* 45,95%, akurasi 79,70. Rohman(2020) melakukan penelitian menggunakan Term Frequency-Inverse Document Frequency(TF-IDF) menghasilkan akurasi hanya 59% dari 6 emosi, namun jika 1 emosi saja dan menghasilkan akurasi cukup baik sebesar 87,23%. Penelitian Dwi (2020) menggunakan *Long Short Term Memory* (LSTM) dengan *word embedding* Glo-Ve menghasilkan akurasi hanya 50% saja, *precision* hanya33%, *recall* hanya 38%, karena dalam proses *training* terjadi *underfitting*. Penelitian Haryadi (2019) membandingkan antara metode SVM, LSTM dan nested LSTM menghasilkan akurasi terendah 98,679%, *precision* 98,53%, *recall* 98,22% untuk SVM, sedangkan akurasi tertinggi didapatkan oleh nested LSTM dan *word embedding word2vec* dengan akurasi 99,167%, *precision* 99,21%, *recall* 98,83%, namun *precision* dan *recall* tertinggi didapatkan oleh LSTM dan *word embedding word2vec* dengan *precision* 99,22%, *recall* 98,86% sedangkan akurasi tidak jauh berbeda dengan nested LSTM yaitu 99,154%,. Penelitian sebelumnya menunjukkan Nested LSTM memiliki akurasi yang paling baik, namun LSTM menunjukkan *precision* dan *recall* yang paling baik, kedua metode tersebut merupakan *deep learning*.

*Word embedding* 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* yang sering digunakanadalah *word2vec* dan GloVe, karena mampu mengenali kata-kata yang mirip, 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) untuk mendapatkan *precision*, *recall* yang baik dan menggunakan *word embedding fasttext* agar mendapatkan akurasi yang lebih baik.

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 yang memiliki akurasiyang lebih baik, 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, *precision*, *recall* algoritma *Long Short Term Memory* (LSTM) dengan *word embedding fasttext* 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.
4. Algoritma *word embedding* yang digunakan pada peneitian ini adalah fasttext
   1. **Tujuan Penelitian**

Tujuan penelitian ini adalah:

Menghitung tingkat akurasi algoritma *Long Short Term Memory*(LSTM) dengan *word embedding fasttext* 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, seperti pendidikan, bisnis, politik, psikologi, rekrukmen karyawan.

* 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, *deep learning*, *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.