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

**Proposal**



**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 dengan bertujuan untuk memungkinkan komputer mengenali dan mengekspresikan emosi (Picard, n.d.). Mendeteksi 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 (Fanesya et al., 2019). 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). Serta 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), dimana 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, yang akan dibagi menjadi 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, 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 2012 sampai 2018 (Saputri et al., 2019). Maka dari itu *twitter* merupakan media sosial yang cocok untuk dijadikan data penelitian ini.

Penelitian yang berkaitan dengan deteksi emosi pernah dilakukan dengan *Naïve Bayes* dipadukan kombinasi 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 mengkategorikan menjadi 8 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). Penelitian lain pernah menggunakan *Vector Space Model* untuk mendeteksi emosi dari kuisioner menjadi 3 emosi(senang, sedih, kecewa), hasil akurasi deteksi emosi dengan frasa adalah 92,59% (Sitorus & Dachlan, 2017).

Ada tantangan yang harus dihadapi dalam deteksi emosi yaitu susahnya pengklasifikasian data uji, karena nilai-nilai antar kelas hampir sama sehingga mempengaruhi tingkat akurasi, dalam hal ini penurunan akurasi, masalah ini ditemukan oleh Fera Fanesya(2019) dalam penelitiannya dalam mendeteksi emosi pada twitter menggunakan metode *Naïve Bayes*. Dalam penelitian tersebut data yang digunakan sebanyak 4396 *tweet* dengan berbandingan data latih 80% dan data uji 20%, mendapatkan hasil akurasi yang kurang maksimal yaitu 55,54% saja. Untuk meminimalisir hal tersebut, dapat dilakukan dengan menggunakan metode *deep learning*.

*Deep learning* merupakan teknik dalam *machine learning* dimana pemprosesan informasi nonlinear memanfaatkan banyak lapisan untuk klasifikasi, ekstraksi fitur ,dan pengenalan pola (Nurrohmat & SN, 2019). Metode *deep learning* yang dapat digunakan untuk mendeteksi emosi adalah *Recurrent Neural Network* (RNN). Kelebihan RNN adalah mampu menangkap informasi tentang urutan panjang menggunakan sel memori (Patel & Tiwari, 2019). Namun RNN memiliki kekurangan yaitu *short-term memory*. Pada RNN memori lama akan tertimpa atau tergantikan dengan memori baru sehingga tidak dapat membawa informasi dari *time step* ke *time step* berikutnya (Bengio et al., 1994). 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, Untuk mengubah *string* atau teks menjadi vector angka dapat menggunakan *word embedding*. Selain itu *word embedding* ini juga mempengaruhi akurasi yang hasilkan oleh *deep learning*, seperti yang dilakukan oleh Boy Utomo 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 yaitu *out of vocabulary*, masalah ini muncul karena ketidakmampuan merepresentasikan vector dari kata yang tidak ada dalam korpus. 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 (Arliyanti Nurdin, Bernadus Anggo Seno Aji, Anugrayani Bustamin, 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 sistem cerdas yang dapat mengklasifikasikan emosi menjadi bahagia, marah, takut, jijik, sedih, dan terkejut secara efisien dengan tingkat akurasi yang tinggi, sehingga hasil deteksi emosi tersebut dapat bermanfaat.

* 1. **Perumusan Masalah**

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

1. Bagaimana tingkat akurasi algoritma *Long Short Term Memory*(LSTM) dalam deteksi emosi media sosial *twitter*?
   1. **Batasan Masalah**

Pada penelitian ini, permasalahan dibatasi sebagai berikut:

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

Tujuan pada penelitian ini adalah sebagai berikut:

1. Dapat mengetahui 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 untuk berbagai aspek bidang.

* 1. **Metodologi Penelitian**
  2. **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,analisis sentimen, jaringan syaraf tiruan*,* backpropagation*.*

**Bab III Analisis dan Perancangan Sistem**

Pada bagian ini akan membahas mengenai analisa dan perancangan sistem hingga implementasi sistem dalam analisis sentiment pada komentar YouTube.

**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 kesimpulan dari hasil penelitian dan saran yang diajukan oleh penulis untuk dikembangan penelitian selanjutnya.