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

Sebagai syarat untuk memperoleh gelar sarjana S-1 di Program Studi Informatika, Jurusan Teknik Informatika, Fakultas Teknik Industri, Universitas Pembangunan Nasional “Veteran” Yogyakarta



**DISUSUN OLEH :**

**M. ALFA RIZA**

**123170027**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK INDUSTRI**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”**

**YOGYAKARTA**

**2021**

**BAB I PENDAHULUAN**

* 1. **Latar Belakang**

Deteksi emosi penting diberbagai bidang seperti pendidikan, bisnis, rekrukmen karyawan. 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 (Ekman, 1999), label-label ini bersifat universal pada budaya yang berbeda.

Saat ini media sosial menjadikan penggunanya cenderung berekspresi emosi 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% (Fanesya et al., 2019). 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% (Rohman, Utami, et al., 2019). Rohman (2019) melakukan penelitian menggunakan Term Frequency-Inverse Document Frequency(TF-IDF) hanya menghasilkan akurasi 59% dari 6 emosi, namun jika 1 emosi saja dan menghasilkan akurasi cukup baik sebesar 87,23% (Rohman et al., 2019). Penelitian yang dilakukan oleh Ardiada(2019) menggunakan *Support Vector Machine*(SVM) dan *K-Nearest Neighbour*. SVM menghasilkan *precision* 45,64%, *recall* 50,20%, dan akurasi cukup bagus sebesar yaitu 81,04%, sedangkan untuk KNN menghasilkan *precision*, *recall*, dan akurasi yang kurang maksimal yaitu *precision* 34,21%, *recall* 45,95%, akurasi 79,70% (Ardiada et al., 2019). Penelitian Haryadi (2019) membandingkan antara metode SVM, Long Short Term Memory (LSTM) dan Nested Long Short Term Memory (NLSTM). SVM menghasilkan akurasi, *precision*, *recall* terendah yaitu akurasi 98,679%, *precision* 98,53%, *recall* 98,22%, sedangkan akurasi tertinggi didapatkan oleh NLSTM dengan *word embedding word2vec* menghasilkan akurasi 99,167%, *precision* 99,21%, *recall* 98,83%, namun *precision* dan *recall* tertinggi didapatkan oleh LSTM dengan *word embedding word2vec* menghasilkan *precision* 99,22%, *recall* 98,86% sedangkan akurasi tidak jauh berbeda dengan NLSTM yaitu 99,154% (Haryadi dan Kusuma, 2019). 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* (Dwi et al., 2020). Penelitian sebelumnya menunjukkan NLSTM memiliki akurasi yang paling baik, namun LSTM menunjukkan *precision* dan *recall* yang paling baik, hal ini berarti LSTM memiliki performa yang lebih baik. LSTM efektif untuk mengolah data sekuensial seperti teks, karena memiliki *memory cell* (Miedema, 2018).

LSTM dengan *word embedding word2vec* banyak digunakan karena *word2vec* mampu mengenali kata-kata yang mirip, *word embedding* Glo-Ve juga banyak digunakan karena mampu memperoleh hubungan semantik antarkata dengan *co-occurrence* (Nurdin et al., 2020), namun *word2vec* dan Glo-Ve tidak dapat menangani kesalahan penulisan sehingga menurunkan akurasi (Lim et al., 2020). Hal ini terjadi karena *word2vec* dan Glo-Ve tidak dapat merepresentasikan vektor dari kata yang belum pernah ditemui, masalah ini dinamakan *out of vocabulary*(OOV), *word embedding fasttext* dapat menangani masalah OOV dengan memperhatikan informasi *subword* menggunakan *n-gram* (Bojanowski et al., 2017). Sehingga penelitian ini menggunakan metode *Long Short Term Memory* (LSTM) untuk mendapatkan *precision*, *recall* yang baik dan menggunakan *word embedding fasttext* agar mendapatkan akurasi yang lebih baik.

Penelitian ini akan menerapkan *Long Short Term Memory* (LSTM) dengan *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 seperti pendidikan, bisnis, rekrukmen karyawan.

* 1. **Perumusan Masalah**

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

Evaluasi 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. Data penelitian diambil dari akun beberapa *influencer* di Indonesia sebagai data utama dan data trending selama 5 April 2021 sampai dengan 11 April 2020.
   1. **Tujuan Penelitian**

Tujuan penelitian ini adalah:

Menghitung tingkat akurasi, *precision*, *recall* 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.

* 1. **Tahapan Penelitian**

Tahapan penelitian yang dipakai pada penelitian ini sebagai berikut :

**1.6.1 Rencana dan Tahapan Penelitian**

Rencana dan tahapan penelitian 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*, data diambil menggunakan teknik *web scrapping* dan data tersebut diberi label secara manual.

1. Analisis Sistem

Analisis sistem ini merupakan penganalisaan terhadap kebutuhan dalam pembuatan sistem. Sehingga dapat diketahui kebutuhan-kebutuhan 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 selanjutnya sehingga mendapatkan hasil yang lebih baik.

**1.6.2 Metode Pengembangan Sistem**

Metode pengembangan sistem yang digunakan pada penelitian ini adalah *prototyping*. *Prototyping* digunakan pada penelitian ini untuk mengembangkan model menjadi final, sehingga sistem yang dikembangkan lebih cepat dan lebih hemat dibandingkan metode pengembangan sistem lainnya. *Prototyping* memiliki beberapa tahapan, yaitu (Pressman, 2015) :

1. *Communication*

Tahap awal pengembangan sistem melakukan komunikasi dan kolaborasi dengan pelanggan atau pemangku kepentingan untuk memahami tujuan dan kebutuhan sistem sehingga membantu menentukan fitur dan fungsi sistem.

1. *Planning*

Tahap *planning* berguna untuk membantu pengembang saat pembuatan sistem. Tahap *planning* mendeskripsikan tugas teknis, resiko yang mungkin terjadi, sumber daya yang dibutuhkan, hasil produk, dan jadwal pengerjaan sistem.

1. *Model*

Tahap ini pengembang membuat *model* dari sistem yang akan dibuat sehingga pengembang dapat memahami kebutuhan sistem dan design yang sesuai untuk menunjang kebutuhan tersebut.

1. *Construction*

Tahap *construction*, pengembang memulai pembuatan sistem tahap sebelumnya, selain itu tahap ini juga melakukan pengujian atau *testing* untuk menemukan kesalahan pada pembuatan sistem.

1. *Deployment*

Tahap ini sistem yang telah dibuat dikirimkan kepada pelanggan baik semua fitur selesai maupun sebagian untuk mendapatkan evaluasi produk dan memberikan *feed back* berdasarkan evaluasi.

**1.6.3 Pengujian Sistem**

Pengujian sistem yang digunakan pada penelitian ini adalah metode *black box testing*, yaitu pengujian yang menekankan fungsionalitas tanpa mengetahui *coding* dari sistem tersebut. *Black box testing* bertujuan untuk mengukur kinerja dari sistem yang telah dibangun.

**1.6.4 Pengujian Penelitian**

Pengujian penelitian yang digunakan pada penelitian ini adalah *confusion matrix*. *Confusion matrix* digunakan untuk mengukur tingkat performa model klasifikasi dengan menghitung nilai *precision*, *recall*, *accuracy*. Terdapat empat istilah pada *confusion matrix* yang merepresentasikan hasil, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN).

* 1. **Sistematika Penulisan**

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

**Bab I Pendahuluan**

Pada bagian ini membahas masalah kehidupan, *research gap*, masalah metode, dan solusi masalah metode. Bab ini terdiri dari latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, tahapan penelitian, dan sistematika penulisan.

**Bab II Tinjauan Pustaka**

Tinjauan pustaka memuat tentang uraian dan pembahasan tentang teori, konsep, model, metode, sistem, atau analisis pustaka ilmiah yang berkaitan dan bersifat fundamental terhadap masalah yang akan diselesaikan. 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 Metodologi Penelitian dan Pengembangan Sistem**

Pada bagian ini akan membahas mengenai jenis penelitian, strategi dan rancangan penelitian, analisis kebutuhan sistem, metode penelitian, akuisisi data, pengolahan data, perancangan sistem, simulasi komputasi, pengembangan sistem, pengujian sistem.

**Bab IV Hasil dan Pembahasan Implementasi**

Pada bab ini akan merealisasikan metodologi penelitian menjadi produk akhir, selain itu bab ini berisi hasil pelaksanaan dan pengujian metode penelitian yang disajikan secara sistematis dan logis. Pembahasan implementasi juga membahas mengenai hasil penelitian yang diperoleh untuk menjawab masalah penelitian. Selain itu berisi pemahaman baru yang didapatkan dari hasil penelitian.

**Bab V Penutup**

Pada bagian ini berisi hasil penelitian untuk membuktikan hasil mampu menjawab rumusan masalah sehingga dapat ditarik kesimpulan berdasarkan tujuan. Saran juga dibahas pada bagian ini sehingga dapat dikembangkan pada penelitian lebih lanjut.