

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Manusia merupakan makhluk sosial, dan memiliki kebutuhan akan berkomunikasi. Media sosial adalah salah satu platform yang sering digunakan di era modern ini untuk berkomunikasi dengan cara mengungkapkan pikiran dan perasaan. Banyaknya data dan jejak digital pada sosial media memberikan sejumlah besar pengetahuan implisit dan perspektif baru bagi para peneliti untuk memahami pemikiran pengguna sosial media. Media sosial Twitter adalah situs media sosial populer. Kontennya sebagian besar bersifat publik dan sangat ringkas (tidak lebih dari 140 karakter per tweet). Twitter menyediakan API gratis yang dapat digunakan untuk mengalirkan data[1]. Karenanya, dirasa cukup efektif dan efisien memilih Twitter untuk menganalisis stres psikologikal dari konten pengguna.

Penggunaan media sosial, sebuah fenomena yang telah menjadi bentuk komunikasi utama bagi kaum muda di berbagai belahan dunia. Tidak diragukan lagi, penggunaan media sosial memiliki memiliki manfaat yang positif, namun, sejumlah bukti menunjukkan penggunaan sosial media juga memiliki keterkaitan dengan masalah kesehatan mental bagi remaja sebagai pengguna[2]. Jumlah waktu yang dihabiskan oleh remaja untuk menggunakan media digital online meningkat, disamping itu waktu untuk tidur dan interaksi sosial secara langsung menurun, seiring juga dengan menurunnya tingkat kebahagiaan[3]. Untuk dapat menganalisis konten sosial media maka harus dilakukannya pengklasifikasian teks. Salah satu hal

yang dilakukan dalam mengklasifikasikan teks pendek di Twitter adalah analisis sentimen[4]. Pada penelitian ini analisis sentimen digunakan untuk mengetahui stres psikologikal dari sebuah tweet. Sehingga dapat dengan mudah mengetahui psikologikal stres dari pengguna Twitter yang melakukan tweet.

Terdapat beberapa metode di masing-masing bidang ilmu tersebut, salah satunya yang akan di gunakan di penelitian ini adalah Support Vector Machine (SVM) dari ruang lingkup Machine Learning dan Long Short-Term Memory (LSTM) di lingkup Deep Learning. Pada penelitian ini akan menggunakan pembelajaran supervised learning, dimana pada supervised learning data akan di labelkan sebelum di training [5]. SVM merupakan metode klasifikasi populer dan terbukti efektif pada sebagian besar masalah klasifikasi [6]. Dalam sebuah penelitian yang sama dalam hal menganalisis stres psikologikal, LSTM mampu memberikan kinerja yang terbaik [7].

Adanya perkembangan teknologi dengan berbagai macam platform sangat membantu proses deteksi stres pada remaja [8]. Oleh karena itu, dilakukanlah penelitian perbandingan antara dua algoritma Support Vector Machine dan Long Short-Term Memory yang digunakan untuk text classification dari segi metrik akurasi dalam menganalisis stres psikologikal pada tweet. Perbandingan ini bertujuan untuk mengetahui algoritma mana yang memiliki metrik akurasi terbaik.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan di atas, maka didapatkan rumusan masalah penelitian ini adalah :

1. Bagaimana implementasi algoritma SVM dan LSTM dalam mendeteksi stres psikologikal pada tweet.
2. Bagaimana hasil perbandingan antara algoritma SVM dan LSTM dalam mengklasifikasikan teks pada konten sosial media Twitter.

## 1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Objek yang menjadi penelitian ini adalah konten berupa pernyataan (*tweet*) pada sosial media Twitter dengan kata kunci yang mengandung stres psikologikal. Jumlah *tweet* yang digunakan sebagai data pada penelitian ini kurang lebih berjumlah 4400 *tweets* berbahasa inggris yang di ambil pada tanggal 1 Januari 2019 sampai dengan 20 April 2019.
2. Fokus dari penelitian ini hanya membandingkan dua model algoritma yaitu Support Vector Machine (SVM) dan Long Short-Term Memory (LSTM).
3. Penelitian ini mengukur nilai *accuracy*, *precision*, *recall* dan *f1-score* model algoritma SVM dan LSTM melalui tahapan evaluasi dan validasi dengan menggunakan metode *confusion matrix* dan *k-fold cross validation*.

#### **1.4 Tujuan Penelitian**

Tujuan penelitian dalam skripsi ini adalah :

1. Mengimplementasi dan membandingkan akurasi antara algoritma SVM dan LSTM dalam melakukan klasifikasi teks.
2. Mengetahui apakah kedua algoritma tersebut dapat diandalkan dalam mengklasifikasikan teks terhadap stres psikologikal.

#### **1.5 Manfaat Penelitian**

Hasil penelitian ini diharapkan dapat bermanfaat untuk :

1. Penelitian ini atau yang sejenis berguna untuk mengetahui stress psikologikal yang ada pada tweet dari pengguna twitter. Hasil yang didapatkan bisa dijadikan referensi atau alat bantu bagi perkembangan profesionalisme di bidang psikologi dan teknologi di masa mendatang.
2. Hasil penelitian ini dapat dijadikan bahan atau referensi bagi para peneliti berikutnya yang akan membahas klasifikasi teks khususnya analisis sentimen pada sosial media Twitter.

#### **1.6. Metode Penelitian**

Metode yang digunakan dalam penelitian kuantitatif ini adalah eksperimental. Pada tahap ini berisi prosedur atau langkah-langkah yang harus ditempuh, waktu penelitian, sumber data, serta dengan cara apa data tersebut diperoleh dan diolah atau dianalisis.

### 1.6.1. Metode Pengumpulan Data

Pengumpulan data merupakan suatu metode atau prosedur yang digunakan untuk mendapatkan suatu informasi tentang apa saja yang harus dikerjakan pada saat melakukan pengelolaan data. Pada tahap pengumpulan data terdapat beberapa hal yang harus dilakukan yaitu :

#### a. Studi Literatur

Studi literatur dilakukan untuk mempelajari penelitian-penelitian dengan tema sejenis yang pernah dilakukan sebelumnya. Penelitian-penelitian ini dijadikan referensi pada penelitian ini. Sumber yang dipakai dapat berupa buku, jurnal, *paper*, *website*, dan lain-lain. Pada penelitian ini, referensi yang dicari adalah pendeteksian stres pada sosial media dan metode-metode klasifikasi teks yang di pakai untuk analisis pada konten sosial media.

#### b. Pengumpulan Data melalui Twitter

Data input yang akan dipakai pada penelitian ini berupa konten pada sosial media Twitter (*tweet*) dengan kata kunci tertentu. Untuk mengumpulkan data tweet menggunakan alat *collection* Twitter bernama *GetOldTweets3*. Data yang sudah dikumpulkan kemudian akan dibagi menjadi data latih dan data uji. Data latih akan diberikan porsi sebesar 80% dan data uji sebesar 20%. Data latih akan diberi label secara manual, sedangkan data uji tidak diberi label. Dalam penelitian ini, penulis secara manual memilih hashtag yang menunjukkan keadaan perasaan penulis stres atau tidak, dan menggunakannya sebagai label positif (*stress*) dan label negatif (*unstress*).

### 1.6.2. Metode Analisis

Penggunaan metode analisis bertujuan untuk mengenali penyebab masalah yang dihadapi serta mengambil keputusan yang tepat dari hasil analisis tersebut. Semua data yang terkumpul dari objek penelitian dan studi pustaka akan dianalisis guna mendapatkan hasil untuk pemecahan masalah yang timbul.

### 1.6.3. Metode Perancangan

Pada tahap ini akan dilakukan perancangan model yang akan dibuat. Tahapan-tahapan dalam proses perancangan model ini antara lain *pre-processing* data, memfilter kata-kata yang dianggap penting, mengubah kata ke bentuk kata dasar, pembobotan pada kata dengan metode TF-IDF, perancangan model dari data latih dengan menggunakan metode Support Vector Machine dan Long Short-Term Memory.

### 1.6.4. Metode Pengujian

Model yang telah dibuat kemudian diuji dan dievaluasi. Pengujian dilakukan untuk mengetahui nilai akurasi dari model yang telah dibangun. Selain menguji model, kedua model yang telah dibangun akan dibandingkan, untuk mengetahui model yang memiliki akurasi yang lebih baik. Pengujian menggunakan metode *confusion matrix* dengan mencari *accuracy*, *precision*, *recall* dan *f1 score*. Kemudian model akan divalidasi menggunakan metode Cross Validation untuk memeriksa apakah model melakukan klasifikasi secara *overfitting* atau *underfitting*.

### 1.7. Sistematika Penulisan

Rincian sistematika penulisan yang digunakan dalam penelitian ini adalah sebagai berikut:

**BAB I PENDAHULUAN**, bab ini memuat paparan latar belakang penelitian, rumusan masalah, batasan masalah dalam penelitian, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

**BAB II LANDASAN TEORI**, bab ini berisi pembahasan tinjauan pustaka tentang penelitian sebelumnya yang dijadikan referensi pada penelitian ini, serta dasar - dasar teori yang digunakan dalam penelitian ini.

**BAB III METODOLOGI PENELITIAN**, bab ini berisi analisis dari penelitian serta perancangan model yang akan dibuat.

**BAB IV HASIL DAN PEMBAHASAN**, bab ini berisi hasil dan pembahasan dari eksperimen yang telah dilakukan serta hasil pengujian model algoritma yang digunakan.

**BAB V KESIMPULAN DAN SARAN**, bab ini berisi kesimpulan dari penelitian serta saran untuk penelitian selanjutnya.