

## BAB I PENDAHULUAN

### 1.1. Latar Belakang Masalah

Penggunaan media sosial untuk berkomunikasi sesama warga pengguna internet (*netizen*) sudah sangat umum dan menyebar ke seluruh lapisan masyarakat Indonesia. Setiap netizen bebas mengungkapkan ide atau hal yang dirasakannya melalui postingan di media sosial yang kadang tidak terkendali dari segi etika penggunaan bahasa yang dipakai. Ada hal yang perlu dikonfirmasi pada dunia maya, karena pada sistem online masyarakat bisa sangat mengenal satu sama lain tidak hanya secara lokal, tetapi juga internasional. Pada saat terjadi penyebaran berita fitnah maupun ujaran kebencian, biasanya *Internet Service Provider (ISP)* menjadi pihak yang dianggap bertanggung jawab dan memiliki kewenangan atas penyebaran pihak ketiga tersebut. [1]

Penyebaran ujaran kebencian yang semakin marak akan dapat menimbulkan efek negatif baik bagi di kalangan pengguna media sosial itu sendiri, maupun bagi pihak lain yang berkaitan dengan dari ujaran kebencian tersebut. Salah satu cara lama yang dapat dilakukan oleh pengguna media sosial untuk mereduksi tindakan ujaran kebencian adalah dengan membuat pesan tertulis dan melaporkan akun yang menyebarkan ujaran kebencian tersebut. Hal yang akan menjadi masalah kedepannya adalah kemungkinan bagi pelaku media sosial untuk secara otomatis menghapus ujaran kebencian yang tersebar di *platform* media sosial yang mereka gunakan. Oleh karena itu, ada beberapa tahapan yang dapat dilakukan berkaitan dengan ujaran kebencian dimana penggunaan alat

elektronik dengan bahasa mesin dapat memainkan peran untuk membedakan antara kalimat biasa dengan ujaran kebencian. [2]

Ujaran kebencian menjadi topik yang mendapat perhatian untuk ditelaah bagaimana bahasa mesin dengan algoritma *Long Short-Term Memory (LSTM)* dapat melakukan suatu analisis sehingga dapat diamati muatan kebenciannya ada atau tidak. Didalam algoritma *Long Short-Term Memory* yang merupakan salah satu jenis arsitektur dari *Recurrent Neural Network (RNN)*, para programmer biasanya dapat mendeteksi kalimat ujaran tersebut secara pembelajaran mendalam (*deep learning*). Model ramalan bersifat statis untuk prediksi nilai indikator polusi udara dalam deret waktu merupakan penerapan dari *Long Short Term Memory* berbasis *Reccurent Neural Network*. Hal ini disebabkan bahwa terdapat penyesuaian model dinamis terhadap model prediksi statis berdasarkan nilai observasi dari polutan tersebut. Untuk menganalisa kinerja model, penelitian lain memperbaikinya dengan menggunakan kombinasi dari *mean square error* dari model LSTM-Kalman dan model statis LSTM. Selain itu, penelitian ini juga menggunakan nilai *R-squared* (koefisien determinasi) dari model prediksi yang dibangun.

Setelah penelitian Kalman, penelitian sberikutnya berkaitan dengan LSTM adalah pada salah satu isu penting pada sistem transportasi cerdas yaitu permasalahan untuk meramalkan kondisi lalu lintas jangka pendek. Pada penelitian yang disebutkan diatas, terdapat suatu model pengajuan berdasarkan algoritma LSTM yang mempertimbangkan data traffic yang besar pada struktur hirearkinya untuk terintegrasi pada korelasi *temporal-spatial* pada LSTM untuk

menghasilkan peramalan kondisi lalu lintas yang tepat. Model yang dihasilkan juga berdasarkan pada *Internet Of Vehicles* (IOVs) serta analisis korelasi. [3]

Model prediksi trajektori lainnya untuk melihat kondisi lalu lintas padat jangka panjang dengan memanfaatkan LSTM adalah pada imodifikasi penambahan *shortcut connections* dengan tujuan akhir membangun model prediksi trajektori *spatio temporal*. Percobaan *trajectory* ini dilakukan dengan menggunakan *dataset* NGSIM I-80 dan US-101. Secara implisit model ini dapat mengukur interaksi dari kendaraan yang memiliki tenaga dan dapat menjangkau adanya kelemahan pada model LSTM klasik dimana LSTM klasik tidak dapat secara simultan menjelaskan interaksi spasial antar kendaraan serta hubungan waktu antara trajektori deret waktu. [4]

Implementasi LSTM lain adalah pada bidang keamanan jaringan. Pengelola jaringan perlu mewaspadaikan akan adanya kemungkinan serangan dari luar atau dalam jaringan, yang berniat mencari kelemahan jaringan, atau mengakses data yang sifatnya rahasia. Pada umumnya *firewall* digunakan untuk mengamankan jaringan. Namun itu saja tidak cukup, untuk lebih mengamankan jaringan dibangunlah *Intrusion Detection System*. Pada penelitian, model LSTM diimplementasikan pada masalah *Intrusion Detection System* (IDS) khususnya menggunakan salah satu *benchmark dataset* yaitu KDD Cup '99. Data KDD Cup'99 yang digunakan adalah versi 10% data dimana terdapat 4 jenis serangan yaitu Probe, Denial of Service (DoS), User to Root (U2R), dan Remote 2 Local (R2L). Pada penelitian ini dipilih metode Random Forest sebagai metode pemilihan fitur. Data KDD Cup'99 memiliki lebih dari 40 fitur dan secara keseluruhan terdiri dari lebih dari 400.000 *records* yang akan berakibat pada

mahalnya biaya komputasi apabila tidak dilakukan tahapan pemilihan fitur terlebih dahulu. [5]

Hasil klasifikasi dengan menggunakan LSTM mencapai keakuratan 98.85%. Penelitian membahas tentang desain, implementasi, dan evaluasi dari suatu aplikasi *mobile* yang dinamakan *Third-Eye*, yang dapat digunakan pada ponsel untuk memonitor *PM2.5*, yaitu partikel dengan diameter 2.5 mikron atau kurang dari itu. Hal ini umum digunakan pada pengukuran kualitas udara di perkotaan. LSTM dan *Convolutional Neural Network (CNN)* digunakan untuk melatih model inferensia *PM2.5*. Hasil ujicoba menunjukkan *Third-Eye* memiliki akurasi klasifikasi mencapai 81.55%. Hasil penelitian menunjukkan bahwa model *hybrid* yang diklaim sederhana secara komputasi, lebih sederhana dari model fisis, cepat dan efisien yang menggabungkan model prediksi yang bersifat *hybrid data-driven Ensemble Empirical Mode Composition (EEMD)* serta LSTM yang dapat memprediksi suhu harian permukaan tanah (*Daily Land Surface Temperatur/Daily LST*). Model ini dinamakan *EEMD Long Short-Term Memory*. Data yang digunakan adalah data dari Dongting Lake, Cina pada 1 January 2014 sampai dengan 31 Desember 2016. [6]

Kinerja dari model prediksi diukur dengan parameter *Mean Square Error (MSE)*, *Mean Absolute Error (MAE)*, *Mean Absolute Percentage Error (MAPE)*, *Root Mean Square Error (RMSE)*, Koefisien Korelasi Pearson dan Koefisien Efisiensi Nash-Sutcliffe. Hasil penelitian menyimpulkan bahwa model yang diteliti dapat bekerja lebih baik dari 5 model lain dalam memprediksi *Daily LST*, yaitu model yang menggunakan LSTM, BERT, *Empirical Mode Decomposition (EMD)-RNN* dan *EEMD-RNN*. [7]

## 1.2. Identifikasi Masalah

Berdasarkan latar belakang masalah di atas, ada beberapa hal yang dapat diidentifikasi, yaitu :

- a. Adanya ujaran kebencian yang menggunakan kalimat bahasa Indonesia yang *abusive*
- b. Penangkalan deteksi kalimat yang *abusive* terlalu sulit untuk dilakukan
- c. Algoritma LSTM berdasarkan penelitian terdahulu dapat mendeteksi kalimat yang *abusive* di media sosial
- d. Sosialisasi penggunaan internet yang ramah dalam pemakaian bahasa

## 1.3. Rumusan Masalah

Berdasarkan identifikasi dari permasalahan di atas, maka penelitian ini merumuskan masalah bagaimana mendeteksi ujaran kebencian dalam media sosial dan merepresentasikannya dengan metode algoritma *Long Short Term Memory* dan *Bidirectional Encoder Representations dari Transformers*?

## 1.4. Ruang Lingkup Pembahasan

Batasan masalah dari penelitian ini adalah sebagai berikut :

- a. Metode yang digunakan pada penelitian kali ini adalah pendekatan *Long Short Term Memory networks* dan *Bidirectional Encoder Representations dari Transformers*.
- b. Data yang digunakan hanya data pada tahun 2023
- c. Media sosial yang digunakan adalah Twitter

## 1.5. Tujuan Penelitian

Adapun tujuan penelitian dalam tesis ini adalah untuk melihat kinerja metode *Long Short Term Memory networks* dan *Bidirectional Encoder Representations*

dari *Transformers* dalam menganalisis kalimat bahasa yang mengandung kebencian pada media sosial dengan pendekatan.

## **1.6. Manfaat Penelitian**

### **1.6.1. Bagi Pengguna Media Sosial**

Manfaat dari penelitian ini bagi pengguna media sosial adalah memberikan wawasan pengetahuan tentang kategori ujaran yang termasuk kebencian atau tidak sehingga pengguna media sosial dapat mawas diri dan mampu bersikap serta beropini yang bertanggung jawab pada suatu group media sosial.

### **1.6.2. Bagi Perguruan Tinggi**

Manfaat penelitian dalam tesis ini untuk perguruan tinggi adalah :

- a. Dapat melakukan perencanaan tentang peluang kerjasama penelitian dengan para penyedia internet (*Internet Service Provider*) sebagai bentuk pengabdian kepada masyarakat.
- b. Menambah koleksi pustaka ilmiah pada *digital library* kampus.

### **1.6.3. Bagi Mahasiswa**

Manfaat penelitian dalam tesis ini adalah :

- a. Memberikan wawasan pengetahuan di bidang *deep learning* dengan pendekatan *Long Short Term Memory networks* dan *Bidirectional Encoder Representations dari Transformers* yang dapat diterapkan kepada perusahaan pengelola media sosial.
- b. Dapat memenuhi persyaratan untuk menyandang predikat sebagai sarjana Magister Teknik Informatika (MTI) pada Institut Bisnis dan Informatika Darmajaya Bandar Lampung.

## **1.7. Sistematika Penulisan**

Untuk lebih memahami lebih jelas proposal penelitian ini, maka materimateri yang ada dalam proposal ini dikelompokkan menjadi beberapa sub bab dengan sistematika penulisan sebagai berikut :

### **BAB I PENDAHULUAN**

Berisi tentang latar belakang, identifikasi masalah, rumusan masalah, tujuan dan manfaat penelitian serta sistematika penulisan.

### **BAB II TINJAUAN PUSTAKA**

Bab ini berisi tentang teori-teori yang menjadi landasan penulisan proposal penelitian ini. Teori-teori berupa definisi yang diambil dari kutipan buku yang berkaitan dengan kinerja sistem informasi, dengan pendekatan *Long Short Term Memory networks* dan *Bidirectional Encoder Representations* dari *Transformers* yang berhubungan dengan penulisan tesis ini.

### **BAB III METODE PENELITIAN**

Bab ini berisi tentang metodologi penelitian yang digunakan yaitupendekatan *Long Short Term Memory networks* dan *Bidirectional Encoder Representations* dari *Transformers* beserta langkah-langkahnya.

### **BAB IV HASIL PENELITIAN DAN PEMBAHASAN**

Bab ini membahas tentang hasil penelitian yang sudah dilakukan pada bab sebelumnya disertai dengan ulasan dan perhitungan yang sistematis pada dua pendekatan yang digunakan.

### **BAB V PENUTUP**

Bab ini berisi kesimpulan dan saran yang berkaitan dengan hasil penelitian.