

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Cyberbullying mencakup tindakan sengaja menyebarkan konten yang menyakitkan atau jahat melalui berbagai platform online, seperti pada situs media sosial, email, ruang obrolan, dan pesan instan. Selain itu, bentuk pelecehan ini dapat muncul melalui perangkat komunikasi digital seperti smartphone [1].

Platform media sosial memungkinkan individu untuk terhubung dan berbagi kontennya, seringkali memunculkan berbagai pendapat [2]. Pada Januari 2023, menurut laporan dari We Are Social, Indonesia memiliki total populasi sebanyak 276,4 juta, dengan 212,9 juta orang menggunakan internet dan 167,0 juta menggunakan media sosial. Statistik ini menunjukkan bahwa lebih dari setengah populasi Indonesia menggunakan media sosial.

Baik Twitter maupun Instagram adalah platform media sosial di mana cyberbullying dapat terjadi [2]. Contoh cyberbullying dalam komentar Instagram adalah "Makin jelek aja anaknya, padahal ibu ayahnya cakep", "Muka nya muka kolot wkwk bukan muka bayi lucu gt", dan "ANAK TOLOL INI MULAI AKTIF YA BUND".

Natural Language Processing (NLP) adalah teknik yang banyak digunakan untuk analisis sentimen dalam klasifikasi teks. Popularitas pertumbuhannya yang cepat telah mendorong para peneliti untuk fokus pada pengumpulan data, analisis web, klasifikasi teks, dan temuan informasi [3]. Bidang NLP mengintegrasikan linguistik, komputasi, dan statistik untuk memungkinkan komputer memahami bahasa manusia [4].

Supervised learning adalah pendekatan yang memerlukan data yang diberi label secara manual untuk melatih klasifikator. Pendekatan pembelajaran terawasi telah menarik minat para peneliti untuk deteksi sentimen, mengimplementasikan teknik seperti SVM, LSTM, ANN, Naive Bayes, Decision Tree, dan banyak lagi dengan berbagai bahasa dan dataset [5].

Salah satu teknik dasar dalam temuan informasi dan penambangan teks adalah skema pembobotan Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini memberikan pentingnya kata-kata berdasarkan frekuensinya dalam suatu dokumen relatif terhadap distribusinya di seluruh dokumen. Rasio ini didasarkan pada kata-kata yang muncul secara sering dalam suatu dokumen tetapi jarang muncul di seluruh dokumen, dianggap lebih informatif dan relevan dengan konten spesifik dokumen tersebut. Spesifikasinya ditentukan oleh distribusi, atau "penggunaan kata," bukan oleh makna inheren mereka. Oleh karena itu, kata-kata yang kurang sering muncul di seluruh koleksi dokumen dan lebih spesifik diberi bobot lebih tinggi, karena keberadaan mereka dalam suatu dokumen dianggap lebih mengindikasikan topik atau tema dokumen tersebut [6].

Word2Vec memperkenalkan arsitektur model inovatif untuk menghasilkan representasi vektor yang berkelanjutan dengan kata-kata dari dataset yang sangat besar. Representasi ini telah menunjukkan peningkatan yang luar biasa dalam tugas kesamaan kata dibandingkan dengan metode berbasis jaringan saraf sebelumnya. Secara khusus, Word2Vec dapat menghasilkan vektor kata berkualitas tinggi dari dataset sebesar 1,6 miliar kata dalam waktu kurang dari satu hari, mencapai hasil terbaik dalam mengevaluasi kesamaan kata baik secara sintaksis maupun semantik [7].

FastText adalah pendekatan baru berdasarkan model skipgram yang diusulkan, di mana kata-kata direpresentasikan sebagai kumpulan n-gram karakter. Pendekatan ini tidak hanya memungkinkan representasi kata-kata yang mungkin tidak muncul dalam data latih tetapi juga menawarkan waktu pemrosesan yang cepat. Evaluasi pada tugas-tugas kesamaan kata dan analogi melibatkan sembilan bahasa, menunjukkan bahwa representasi ini melampaui representasi kata morfologis yang baru-baru ini diusulkan [8].

Beberapa penelitian terkait klasifikasi teks menggunakan Bahasa Indonesia telah dilakukan sebelumnya. Iswanto dan Poerwoto [9]. menggunakan klasifikasi Naïve Bayes, Maximum Entropy, dan Support Vector Machines dalam pra-pemrosesan dataset Twitter. Metode ini mencapai recall dan precision hingga

85,50%. Kurniasari dan Setyanto [10]. dalam studinya mengusulkan penggunaan Recurrent Neural Network (RNN)-Long Short Memory Term (LSTM) untuk mengklasifikasikan polaritas sentimen kalimat Bahasa Indonesia. Dalam evaluasinya, mereka menggunakan dataset dari situs perjalanan, terdiri dari 25.000 ulasan yang didistribusikan secara merata antara sentimen positif dan negatif. Model LSTM mereka menunjukkan hasil impresif, mencapai accuracy 95,0%.

Dalam beberapa tahun terakhir, banyak penelitian dilakukan tentang deteksi cyberbullying di media sosial, dengan banyak menggunakan pembelajaran mesin untuk klasifikasi. Secara khusus, Kumiawanda dan Tobing, menggunakan XGBoost dan TF-IDF pada tahun 2022 [11]. Laxmi et al. mengimplementasikan Doc2vec dan CNN pada tahun 2021 [12]. Pada tahun 2018, Luqyana et al. menggabungkan SVM dan TF-IDF, mencapai accuracy tertinggi di antara studi-studi yang disebutkan, terutama ketika fokus pada dataset berbahasa Indonesia [13]. Pada tahun yang sama, Nurrahmi dan Nurjanah, menerapkan SVM dan KNN [14]. Dan Andriansyah et al. menggunakan DTM dan SVM pada tahun 2017 [15].

Dalam penelitian analisis cyberbullying ini, peneliti bertujuan untuk melakukan eksperimen dengan menggunakan Long Short-Term Memory (LSTM) untuk klasifikasi. LSTM, sebagai varian dari jaringan saraf rekuren, ditandai dengan keadaan memori dan struktur selular bertingkat [16]. Peneliti juga bermaksud untuk menggunakan berbagai teknik teks representasi, termasuk TF-IDF, Word2vec, dan FastText.

## 1.2 Rumusan Masalah

Dari latar belakang pada bagian 1.1, rumusan masalah yang didapatkan yaitu:

1. Bagaimana pengaruh representasi teks TF-IDF, Word2vec, dan FastText terhadap accuracy pada saat digunakan dengan klasifikasi cyberbullying menggunakan teknik Long Short-Term Memory (LSTM) ?
2. Dari ketiga word embedding seberapa signifikan perbedaan performa pada klasifikasi cyberbullying dengan menggunakan

teknik Long Short-Term Memory (LSTM) ?

3. Bagaimana pengaruh masing masing word-embedding terhadap overfitting dengan menggunakan LSTM pada jumlah unit yang berbeda beda ?

### 1.3 Batasan Masalah

Adapun batasan masalah yang ada pada penelitian ini antara lain, yaitu:

1. Dataset yang digunakan merupakan kombinasi dari kaggle.com dengan nama dataset Cyberbullying Bahasa Indonesia dan penelitian sebelumnya oleh Luqyana et al [13], dan Alfina et al [17].
2. Kombinasi menunjukkan dataset dari Kaggle yang bernama Cyberbullying Bahasa Indonesia<sup>1</sup> dan dataset pada penelitian sebelumnya oleh Luqyana et al [13] dan Alfina et al [17] sebanyak 1763 data.
3. Dataset menggunakan Bahasa Indonesia.
4. Penelitian ini menggunakan klasifikasi biner. Atau klasifikasi ini hanya memiliki 2 kelas yaitu *cyberbullying* dan *tidak cyberbullying*.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini yaitu untuk mengetahui pengaruh word embedding TF-IDF, Word2vec, dan FastText terhadap performa klasifier LSTM ketika digunakan pada kasus cyberbullying. Hasil dari penelitian ini ditunjukkan dengan menggunakan metode evaluasi yaitu confusion matrix dan cross validation yang hasilnya merupakan persentase dari accuracy, precision, recall, dan f1-score.

### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini untuk akademik, penelitian ini bisa menjadi referensi bagi mahasiswa yang selanjutnya akan meneliti di ruang lingkup yang

<sup>1</sup> Dataset Kaggle: "Cyberbullying Bahasa Indonesia." Diakses pada 28 Juli 2023, pukul 14:43 WIB. [Tautan Dataset: <https://www.kaggle.com/datasets/citrhnn/cyberbullying-bahasa-indonesia/data>]

sama yaitu teks klasifikasi pada kasus cyberbully. Bagi peneliti sendiri diharapkan penelitian ini dapat diimplementasikan untuk mengurangi dampak dari cyberbullying.

## **1.6 Sistematika Penulisan**

Sistematis penulisan penelitian ini dirinci dalam urutan sebagai berikut:

### **BAB I PENDAHULUAN**

Bagian ini mencakup esensi permasalahan dan gambaran umum penelitian, meliputi latar belakang masalah, perumusan masalah, batasan masalah, tujuan, maksud, manfaat penelitian, metode penelitian, dan susunan penulisan skripsi.

### **BAB II KAJIAN PUSTAKA**

Bagian ini membahas literatur yang relevan, mencakup dasar-dasar teori yang terkait dengan topik penelitian dan perangkat lunak yang digunakan.

### **BAB III METODE PENELITIAN**

Bagian ini memuat tahapan penelitian, algoritma, serta perangkat dan data yang akan diterapkan dalam penelitian.

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

Dalam bagian ini, dilakukan implementasi dan perincian desain yang telah disusun. Disajikan hasil-hasil dari setiap tahap penelitian, termasuk analisis, desain, implementasi, dan pengujian.

### **BAB V PENUTUP**

Bagian ini berisi simpulan dari hasil penelitian dan saran untuk meningkatkan aspek-aspek yang perlu diperbaiki pada penelitian tersebut.