

1. Pendahuluan

Pengguna internet di Indonesia, menurut studi Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) dari tahun 2024, mencapai 221 juta orang [1]. Media sosial adalah salah satu platform yang sering digunakan oleh pengguna internet di era modern untuk berinteraksi di dunia maya. Jumlah pengguna media sosial yang semakin meningkat di dunia maya juga disertai dengan kebebasan berekspresi. Mereka tampaknya menemukan ruang untuk mengekspresikan apa pun yang mereka pikirkan dan rasakan. Namun, ini telah menciptakan masalah baru, salah satunya adalah *cyberbullying*.

Cyberbullying, atau yang sering disebut sebagai perundungan daring, merujuk pada tindakan menggunakan internet untuk menyakiti, mengintimidasi, atau menakut-nakuti orang lain [2]. Ini umumnya dilakukan melalui pengiriman pesan, komentar, atau konten digital yang negatif, merendahkan, atau penuh kebencian. Perilaku ini dapat dilakukan oleh siapa saja, tanpa memandang usia, baik anak-anak, remaja, maupun orang dewasa. Fitur-fitur seperti komentar anonim, kemudahan menyebarkan informasi, dan kurangnya regulasi ketat di beberapa platform sering kali dimanfaatkan untuk menyebarkan kebencian atau menyerang individu atau kelompok.

Platform media sosial dengan jangkauan luas, seperti X (sebelumnya Twitter), sering menjadi ruang utama untuk *cyberbullying*. Di platform ini, pengguna dengan mudah mengekspresikan pendapat atau kritik yang kadang-kadang meningkat menjadi serangan balasan, mencari-cari kesalahan, ancaman, intimidasi, dan merendahkan satu sama lain, yang mengandung unsur *cyberbullying*. Tindakan *cyberbullying* sering terjadi melalui pengiriman tweet atau pesan singkat oleh pengguna media sosial, seperti komentar keras atau pernyataan yang merendahkan individu atau kelompok tertentu. Oleh karena itu, sangat penting untuk membuat sistem deteksi *cyberbullying* di platform X yang menggunakan *feature expansion* berbasis *word embedding*.

Pengiriman tweet dalam tindakan *cyberbullying* sering kali mengandung kata-kata yang sangat pendek, tata bahasa yang salah, dan variasi kata yang menyebabkan kesulitan dalam pemahaman serta kesalahan dalam penggunaan kosakata. Oleh karena itu, *word embedding* dapat digunakan untuk *feature expansion* dan mengurangi kesalahan kosakata. *Word embedding* adalah metode untuk menentukan vektor dari sebuah kata dan konteksnya dalam sebuah korpus, yang kemudian dapat dicocokkan dengan kriteria tertentu. Beberapa contoh teknik *feature expansion word embedding* termasuk Word2Vec, FastText, dan GloVe[3].

Dalam studi [4], teknik *feature expansion* menggunakan FastText yang terintegrasi dengan *Hybrid Deep Learning* CNN-BiLSTM mencapai nilai akurasi tinggi sebesar 80,55%. Studi lain yang telah dilakukan [5] menggunakan *feature expansion* Word2Vec dengan model *deep learning* CNN-LSTM, mencapai nilai akurasi sebesar 79,26%. Selain metode-metode ini, *Global Vectors* (GloVe) dapat digunakan untuk memperluas fitur representasi kata. GloVe dianggap sebagai metode yang efisien dan efektif untuk mengajarkan representasi vektor kata. GloVe adalah model *unsupervised learning* untuk *word representation* yang menggunakan *log-bilinear regression* berbasis *global information* dari corpus teks. Dalam penelitian [6], hasil *feature expansion* GloVe yang diklasifikasikan menggunakan model *Artificial Neural Network* (ANN), *Random Forest* (RF), dan *Logistic Regression* (LR) menunjukkan bahwa *feature expansion* GloVe cukup efektif dalam meningkatkan kinerja dengan tingkat akurasi sebesar 88,59%.

Berdasarkan studi [7], sebuah model hybrid untuk analisis sentimen dikembangkan dengan menggabungkan Bidirectional Long Short-Term Memory (Bi-LSTM) dan Convolutional Neural Networks (CNN). Model ini memanfaatkan kemampuan Bi-LSTM dalam mengenali *long-term dependencies* dalam urutan data teks serta kemampuan CNN dalam mengekstraksi fitur lokal melalui teknik *convolutional* dan *pooling*. Model ini menggunakan pre-trained *word embeddings* dengan GloVe serta pendekatan term frequency-inverse document frequency (TF-IDF) untuk representasi teks. GloVe membantu dalam membentuk representasi vektor yang memberikan makna kontekstual yang lebih dalam, sedangkan TF-IDF digunakan untuk mengukur signifikansi suatu kata dalam sebuah dokumen.

Penerapan *hybrid deep learning* yang dikombinasikan dengan *word embedding* lainnya telah dilakukan oleh [8] untuk mendeteksi teks *cyberbullying* dalam bentuk tweet dengan memanfaatkan bidang *Natural Language Processing* dan *hybrid deep learning* model Res-CNN-BiLSTM, di mana GloVe digunakan sebagai *feature extraction*. Penelitian ini mencapai akurasi sebesar 92%. Namun, penelitian ini memiliki keterbatasan terkait dataset dari tweet di Twitter. Tweet yang ada sering kali disingkat karena batasan karakter, sehingga sulit dipahami, dan terdapat variasi yang luas dalam tweets yang digunakan, yang menyebabkan ketidakkonsistenan kosakata. Oleh karena itu, pembuatan sistem deteksi *cyberbullying* di X dengan kemampuan *word embedding* yang ditingkatkan diperlukan [3].

Berdasarkan penelitian [4] mengenai algoritma klasifikasi, dapat disimpulkan bahwa mengintegrasikan *Convolutional Neural Networks* (CNN) dan *Bidirectional Long Short-Term Memory* (BiLSTM) memiliki keunggulan dari kedua teknik tersebut dalam menangkap *local patterns* dan *long-range dependencies* [9]. Keunggulan dari metode GloVe *word embeddings* juga telah terbukti dalam penelitian dibandingkan dengan metode lainnya. Selain itu, GloVe memberikan hasil yang lebih baik dan lebih cepat, serta menghasilkan hasil terbaik terlepas dari kecepatan prosesnya [10].

Kontribusi utama dari penelitian ini adalah *hybrid deep learning* model dan *feature expansion* seperti GloVe yang digunakan untuk mengidentifikasi *cyberbullying* dalam bahasa Indonesia di X. Studi mengenai deteksi *cyberbullying* menggunakan dataset berbahasa Indonesia masih terbatas pada *supervised learning* atau *deep learning* dan belum memanfaatkan *feature expansion*, sebagaimana dijelaskan dalam penelitian terkait sebelumnya. Seperti yang telah disebutkan, dataset Twitter berbahasa Indonesia memiliki tantangan tersendiri, sehingga pemilihan GloVe

sebagai *feature expansion* akan membantu mengatasi ambiguitas bahasa. Oleh karena itu, beberapa skenario akan diuji, seperti memilih rasio data yang berbeda dan menggunakan n-grams terbaik untuk *feature extraction*. Penggunaan GloVe sebagai *feature expansion* dan ketiadaan GloVe akan dibandingkan untuk mengkaji dampaknya sebagai *feature expansion* pada model CNN dan BiLSTM. Kombinasi Hybrid Deep Learning models, CNN-BiLSTM dan BiLSTM-CNN, juga akan diuji.

Bagian 2 akan menjelaskan metode untuk mendeteksi *cyberbullying* dengan memanfaatkan Hybrid Deep Learning dan GloVe sebagai *feature expansion*, dan Bagian 3 akan membahas hasil dan diskusi. Akhirnya, Bagian 4 akan menyajikan kesimpulan dari penelitian yang telah dilakukan

