

Evaluasi Kinerja Algoritma *Graph Neural Network* (GNN) dan *Convolutional Neural Network* (CNN) untuk Deteksi *Cyberbullying* pada Komentar *Twitter*

Muhammad Rizki Nurfiqri¹, Fitriyani²

^{2,2,3}Fakultas Informatika, Universitas Telkom, Bandung

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹fiqrimrn@students.telkomuniversity.ac.id, ²fitriyani@telkomuniversity.ac.id,

Abstrak

Cyberbullying telah meningkat akibat perluasan jaringan media sosial dan kemajuan teknologi internet, yang menghadirkan tantangan besar bagi komunitas daring. Penelitian sebelumnya menggunakan teknik *Support Vector Machine* (SVM) dan mencapai tingkat akurasi sebesar 71,25%. Namun, mengingat sifat dinamis dari perilaku perundungan siber dan kebutuhan akan metodologi deteksi yang lebih kuat, topik ini tetap menjadi tantangan. Penelitian ini menyelidiki penerapan teknik *Convolutional Neural Network* (CNN) dan *Graph Neural Network* (GNN) dalam mendeteksi perundungan siber di *Twitter*. CNN dan GNN dipilih karena kemampuan jaringan saraf untuk menangkap pola yang kompleks dalam data teks dan jaringan. Hasil eksperimen menunjukkan bahwa metode GNN secara konsisten mengungguli CNN dalam hal *f1-score*, akurasi, presisi, dan *recall*. Metode GNN mencapai akurasi sebesar 80,25%, melebihi CNN yang mencapai 68,43%, dengan menggunakan 20 epoch. Selanjutnya, optimasi GNN dengan menerapkan berbagai jumlah epoch mencapai akurasi tinggi sebesar 92,78% ketika menggunakan 200 epoch. Hal ini memvalidasi efektivitas GNN dalam mendeteksi perundungan siber di *Twitter*.

Kata kunci: *Cyberbullying*, *Convolutional Neural Network* (CNN), *Graph Neural Network* (GNN), *Twitter*, Perbandingan performansi

Abstract

*Cyberbullying incidents have surged due to the expansion of social media networks and advancements in internet technology, presenting a substantial challenge in online communities. Previous research utilized Support Vector Machine (SVM) techniques and obtained an accuracy rate of 71.25%. However, given the dynamic nature of Cyberbullying behaviors and the necessity for more robust detection methodologies, the topic remains challenging, this study investigates the application of Convolutional Neural Network (CNN) and Graph Neural Network (GNN) techniques in detecting Cyberbullying on Twitter. We chose CNN and GNN due to the capacity of neural networks to capture intricate patterns in textual and network data. The results of the experiment show that the GNN method consistently outperforms CNN in terms of *f1-score*, accuracy, precision, and recall. The GNN method achieves an accuracy of 80.25%, surpassing CNN 68.43%, by employing 20 epochs. Then the optimization of GNN by implementing various numbers of epochs reaches a high accuracy of 92.78 % when using 200 epochs. This validates the effectiveness of GNN in detecting Cyberbullying on Twitter.*

Keywords: *Cyberbullying*, *Convolutional Neural Network* (CNN), *Graph Neural Network* (GNN), *Twitter*, *Performance Comparison*

1. PENDAHULUAN

1. Latar Belakang

Di Indonesia, internet telah menyebar luas ke seluruh negeri, menandai kemajuan signifikan dalam teknologi informasi. Kini, banyak platform media sosial telah menjadi bagian integral dari kehidupan sehari-hari masyarakat, tidak hanya media massa konvensional. Data terbaru dari We Are Social menunjukkan bahwa jumlah pengguna internet di Indonesia mencapai 212 juta pada Januari 2023, atau sekitar 77% dari populasi, yang

menunjukkan inklusi digital yang semakin meluas [1]. Transformasi digital ini juga membawa tantangan, seperti masalah keamanan siber dan kebutuhan akan literasi digital. Untuk memastikan kemajuan ini terus berlanjut dan memberikan manfaat bagi semua lapisan masyarakat, tantangan-tantangan ini harus diatasi [2].

Di sisi negatif perkembangan media sosial adalah munculnya *cyberbullying*. *Cyberbullying* dapat berupa intimidasi, pelecehan, atau ancaman yang dilakukan melalui media digital seperti internet, media sosial, atau pesan teks [3]. Pada tahun 2023, sebuah studi oleh Hendry bertujuan untuk mendapatkan pemahaman yang luas tentang strategi yang dapat digunakan untuk mencegah dan menangani *cyberbullying*. Upaya ini mengakui pentingnya melibatkan berbagai perspektif dalam menghasilkan respons yang efektif [4].

Sebuah studi yang diterbitkan di Emerald Insight oleh Bharti, Yadav, Kumar, dan Yadav (2022) menemukan bahwa pelecehan siber meningkat di media sosial, khususnya terhadap remaja, dan berdampak negatif pada kesehatan mental mereka. Mereka meneliti metode pembelajaran mesin dan pembelajaran mendalam untuk mengidentifikasi *cyberbullyin*[5]. Blog HubSpot (2021) juga membahas masalah pelecehan yang meningkat di *Twitter*. Salah satu contohnya adalah bagaimana *Twitter* menggunakan algoritma pembelajaran mesin untuk memfilter hasil pencarian, sehingga konten dari akun yang dilaporkan atau dianggap ofensif tidak ditampilkan secara mencolok. Ini menunjukkan bahwa konten ofensif masih bisa terlihat di platform tersebut, tetapi upaya dilakukan untuk menguranginya [6].

Penelitian sebelumnya telah menyajikan berbagai metode untuk mendeteksi *Cyberbullying* di media sosial. Pendekatan yang umum digunakan termasuk Pemrosesan Bahasa Alami (NLP) dan teknik Pembelajaran Mesin (ML) untuk mengklasifikasikan teks berdasarkan karakteristik seperti agresivitas, penggunaan kata-kata negatif, dan pola perilaku pengguna [7]. Dalam penerapan pembelajaran mesin, metode tradisional seperti *Support Vector Machines* (SVM), *K-Nearest Neighbors* (KNN), dan *Decision Trees* telah banyak diterapkan karena kemampuannya memproses data yang beragam, termasuk data dari media sosial seperti *Twitter* [8]. Metodemetode ini tidak hanya mengidentifikasi tanda-tanda *Cyberbullying* tetapi juga mengklasifikasikan tingkat keparahan *Cyberbullying* berdasarkan fitur-fitur yang dihasilkan dari konten media sosial [9].

Penerapan pendekatan *Neural Network* seperti *Graph Neural Network* (GNN) dan *Convolutional Neural Network* (CNN) telah menjadi populer untuk mendeteksi *cyberbullying*. GNN sangat berguna untuk menganalisis hubungan antar pengguna di platform sosial karena dapat menganalisis data terstruktur seperti grafik [10]. Sebaliknya, CNN dapat mengidentifikasi tanda-tanda dalam teks yang menunjukkan *Cyberbullying* karena kemampuannya memproses gambar dan teks [11]. Kedua pendekatan ini bekerja sama untuk mendeteksi *cyberbullying*. Studi ini mempertimbangkan efektivitas algoritma GNN dan CNN dalam mendeteksi *Cyberbullying* di *Twitter*. Pemilihan algoritma GNN dan CNN didasarkan pada kemampuan mereka menangani pola data yang kompleks, seperti yang ditunjukkan oleh penelitian sebelumnya [11] dan [10].

1.2. Topik

Penelitian ini berfokus pada penerapan teknik-teknik pembelajaran mesin, khususnya *Convolutional Neural Network* (CNN) dan *Graph Neural Network* (GNN), dalam mendeteksi insiden perundungan siber di platform media sosial *Twitter*. Penelitian ini menitikberatkan pada kemampuan jaringan saraf untuk menangkap pola-pola kompleks dalam data teks dan jaringan, serta mengevaluasi kinerja dari kedua teknik tersebut dalam konteks deteksi perundungan siber.

1.3. Batasan Masalah

Lingkup Data: Penelitian ini terbatas pada data yang diperoleh dari platform media sosial *Twitter*. Data yang digunakan terdiri dari teks yang diposting oleh pengguna *Twitter*, dan penelitian ini tidak mencakup data dari platform media sosial lainnya. **Metodologi:** Penelitian ini membatasi teknik pembelajaran mesin yang digunakan pada *Convolutional Neural Network* (CNN) dan *Graph Neural Network* (GNN). Metode lain seperti *Support Vector Machine* (SVM), yang juga digunakan dalam penelitian terdahulu, hanya akan digunakan sebagai pembandingan dalam analisis kinerja. **Parameter Evaluasi:** Kinerja deteksi diukur menggunakan metrik *f1-score*, akurasi, presisi, dan *recall*. Penelitian ini tidak mempertimbangkan metrik lainnya seperti *Area Under Curve* (AUC) atau *Mean Squared Error* (MSE). **Optimasi Epoch:** Optimasi kinerja GNN dilakukan dengan menguji berbagai jumlah epoch, dengan fokus utama pada capaian kinerja terbaik pada 20 epoch dan 200 epoch. Penelitian ini tidak mengeksplorasi teknik optimasi lain seperti tuning hiperparameter lebih lanjut atau penggunaan arsitektur jaringan yang berbeda. Batasan-batasan ini ditetapkan untuk memastikan fokus yang jelas dan terarah dalam penelitian, serta memungkinkan analisis yang mendalam terhadap metode yang dipilih.

1.4. Tujuan

Penelitian ini juga bertujuan untuk melakukan perbandingan efektivitas antara kedua metode tersebut dalam mengidentifikasi pola kompleks pada data teks dan jaringan. Selain itu, penelitian ini berfokus pada pengoptimalan kinerja GNN melalui penggunaan berbagai jumlah epoch untuk mencapai tingkat akurasi deteksi yang lebih tinggi.

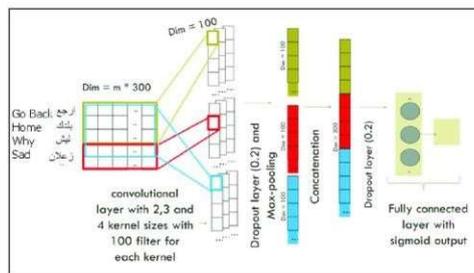
2. KAJIAN PUSTAKA

Pada tahun 2020, sebuah studi di Sardar Patel Institute of Technology oleh Rahul Ramesh Dalvi, Sudhanshu Baliram Chavan, dan Aparna Halbe menggunakan SVM dan *Naïve Bayes* untuk mendeteksi *Cyberbullying* di *Twitter*. Mereka menerapkan model mereka secara real-time menggunakan *Twitter* API setelah mengumpulkan data dari berbagai sumber seperti Kaggle dan GitHub. Studi tersebut menunjukkan bahwa SVM memiliki kinerja lebih baik dengan akurasi 71,25%, namun kedua metode memiliki kelemahan dalam menafsirkan konteks dan sentimen tweet, serta bias dalam *Dataset*, yang dapat mempengaruhi generalisasi hasil. Untuk mendeteksi *Cyberbullying*, sampel data yang representatif dan interpretasi yang akurat sangat penting, menurut studi ini dan penelitian sebelumnya [12].

2.1. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) adalah alat penting untuk pembelajaran mendalam karena dapat menyelesaikan masalah sulit yang tidak dapat diselesaikan oleh metode pembelajaran mesin tradisional. Dengan menggunakan struktur berlapis-lapis, ini menemukan pola dan karakteristik penting dari data masuk seperti gambar atau teks [13].

CNN melakukan operasi konvolusi dengan perkalian elemen demi elemen antara input array fitur dan input array angka. Ini dicapai dengan menerapkan kernel pada input gambar atau teks. Tergantung pada arsitektur dan fungsinya, kernel dapat berukuran 3x3 atau 5x5, misalnya [14]. Untuk CNN, *Flowchart* terdiri dari input data, diikuti oleh beberapa lapisan konvolusi dan pooling, dan kemudian, untuk klasifikasi akhir, data diolah melalui lapisan yang penuh dengan koneksi. Setiap tahap melakukan tugas tertentu dalam memproses dan mengkonsolidasi informasi [15]. Dibawah ini adalah *Visualisasi arsitektur* sebagai berikut:



Gambar 1 Arsitektur CNN

Gambar 1 menggambarkan arsitektur CNN dari berbagai jenis lapisan yang dapat diterapkan. Berisi lima lapisan yaitu lapisan input(embedding layer), lapisan konvolusi, yang pada dasarnya terdiri dari 3 lapisan konvolusi paralel dengan ukuran kernel yang berbeda seperti a poolin layer, a hidden dense layer, output layer. [16]

2.2. Graph Neural Network (GNN)

Metode pembelajaran mesin *Graph Neural Networks* (GNN) digunakan untuk memproses data dalam struktur grafik. GNN telah menjadi lebih efisien dan efektif sejak diperkenalkan. Menurut Wu et al. (2021), GNN telah menjadi penting untuk berbagai tujuan, seperti memprediksi interaksi protein dan membuat sistem rekomendasi [17]. GNN berhasil menangkap hubungan antara node di grafik, memungkinkan pembelajaran fitur yang lebih mendalam. Menurut Zhou et al. (2018), GNN memperbarui informasi node berdasarkan tetangganya dengan teknik agregasi untuk memperbarui node berdasarkan tetangganya [18]. GNN dapat diimplementasikan untuk mendeteksi *Cyberbullying* dengan memanfaatkan struktur jaringan sosial dalam data. Ruan et al. (2019) menunjukkan bagaimana GNN dapat digunakan untuk mengklasifikasi komentar pada media sosial dengan menggunakan struktur grafik untuk memperkaya fitur yang diekstraksi [19]. Berikut dibawah ini adalah visualisasi untuk *Flowchart* penelitian ini: