

ANALISIS SEMANTIK BERBASIS KONTEKS MENGGUNAKAN MODEL HYBRID INDOBERT–KNN UNTUK DETEKSI KOMENTAR CYBERBULLYING TIKTOK

Dita Nur Aulia Latuba*¹, Budiman², Winarti³

¹Program Studi Teknik Informatika, Universitas Darul Ulum, Jombang, Indonesia

*e-mail: ditalatuba03@gmail.com

Abstrak

Cyberbullying saat ini dipandang sebagai salah satu persoalan serius dalam lingkungan, khususnya TikTok, karena berpotensi menimbulkan dampak negatif terhadap kondisi psikologis dan kenyamanan pengguna. Sifat komentar yang pendek, tidak baku, serta bergantung konteks menyebabkan pendekatan klasifikasi yang hanya mengandalkan fitur leksikal sederhana belum mampu memahami makna secara utuh. Berdasarkan permasalahan tersebut, penelitian ini berfokus pada upaya merepresentasikan konteks semantik komentar berbahasa Indonesia secara lebih efektif untuk meningkatkan ketepatan deteksi cyberbullying. Tujuan penelitian adalah menganalisis efektivitas pendekatan representasi semantik berbasis konteks menggunakan model *hybrid* dalam mendeteksi komentar cyberbullying pada TikTok. Metode yang diusulkan menggabungkan *IndoBERT* sebagai *feature extractor* untuk menghasilkan vektor representasi semantik dan algoritma *K-Nearest Neighbor* (KNN) sebagai klasifier berdasarkan kedekatan fitur. Data yang digunakan berupa komentar TikTok berbahasa Indonesia yang telah dilabeli sebagai cyberbullying dan non-cyberbullying. Hasil pengujian menunjukkan bahwa representasi semantik berbasis *IndoBERT* mampu mendukung kinerja klasifikasi KNN secara efektif dan menghasilkan performa deteksi yang stabil. Model *hybrid* ini mencapai nilai akurasi sebesar 81% yang mengindikasikan kemampuan model dalam membedakan komentar bullying dan non-bullying secara efektif. Secara keseluruhan, penelitian ini membuktikan bahwa integrasi representasi semantik berbasis konteks dan klasifikasi berbasis kedekatan vektor berpotensi menjadi solusi yang relevan dan adaptif untuk deteksi cyberbullying berbasis teks pada platform TikTok.

Kata kunci: Cyberbullying; *IndoBERT*; *K-Nearest Neighbor*; TikTok

Abstract

Cyberbullying is currently considered a serious issue in the social environment, particularly on TikTok, due to its potential negative impact on users' psychological well-being and comfort. The short, informal, and context-dependent nature of comments means that classification approaches relying solely on simple lexical features are unable to fully capture meaning. Based on these issues, this research focuses on more effectively representing the semantic context of Indonesian-language comments to improve the accuracy of cyberbullying detection. The purpose of this study is to analyze the effectiveness of a context-based semantic representation approach using a hybrid model in detecting cyberbullying comments on TikTok. The proposed method combines *IndoBERT* as a feature extractor to generate semantic representation vectors and the *K-Nearest Neighbor* (KNN) algorithm as a classifier based on feature proximity. The data used consists of Indonesian-language TikTok comments that have been labeled as cyberbullying and non-cyberbullying. The test results show that *IndoBERT*-based semantic representation is able to effectively support KNN classification performance and produce stable detection performance. This hybrid model achieved an accuracy value of 81%, indicating the model's ability to effectively distinguish between bullying and non-bullying comments. Overall, this study proves that the integration of context-based semantic representation and Vector proximity-based classification has the potential to be a relevant and adaptive solution for text-based cyberbullying detection on the TikTok platform.

Keywords: Cyberbullying; *IndoBERT*; *K-Nearest Neighbor*; TikTok

1. PENDAHULUAN

Media sosial telah berevolusi menjadi ruang publik digital utama, membentuk pola komunikasi dan ekspresi masyarakat kontemporer [1]. Di antara berbagai platform, TikTok menonjol dengan pertumbuhan eksponensial dan tingkat partisipasi pengguna yang tinggi. Platform ini tidak hanya mendorong kreativitas dan pertukaran informasi, tetapi juga menjadi wadah bagi perilaku negatif seperti *cyberbullying* [2]. Cyberbullying didefinisikan sebagai tindakan agresi yang disengaja dan berulang melalui media digital dengan tujuan menyakiti atau merendahkan korban [3]. Dampaknya mencakup gangguan psikologis seperti stres, kecemasan, depresi, serta penarikan diri dari interaksi sosial [4].

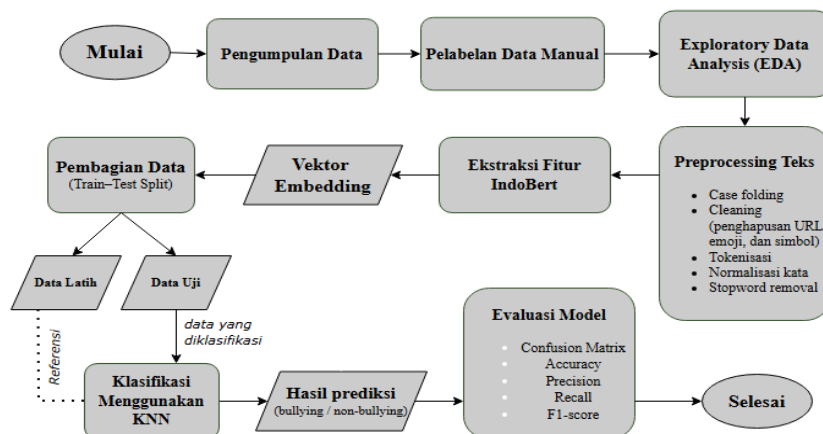
Sebagai upaya mengatasi permasalahan tersebut, pemanfaatan pendekatan otomatis berbasis Natural Language Processing (NLP) menjadi solusi yang relevan. Namun penerapan NLP pada teks media sosial

mewakili tantangan yang cukup kompleks, terutama akibat penggunaan bahasa yang tidak baku, seperti slang, singkatan, emotikon, serta praktik pencampuran bahasa. Keberagaman bentuk linguistik ini dapat memenuhi tahapan preprocessing dan berpotensi menurunkan kinerja model apabila tidak ditangani dengan metode yang sesuai [5]. Sejumlah penelitian sebelumnya menunjukkan bahwa algoritma pembelajaran mesin klasik, seperti K-Nearest Neighbor dan Naive Bayes [6][7] mampu mengidentifikasi pola linguistik sederhana, namun masih memiliki keterbatasan dalam memahami konteks semantik yang lebih kompleks, termasuk sarkasme dan ironi. Dengan berkembangnya deep learning, model berbasis transformer seperti BERT dan IndoBERT menunjukkan performa lebih stabil dan akurat dalam mendeteksi cyberbullying berbahasa Indonesia [8][9].

Meskipun demikian, sebagian besar studi masih menggunakan satu pendekatan dan berfokus pada platform seperti Twitter atau Instagram, sehingga pemanfaatan model hybrid yang menggabungkan kekuatan transformer dan algoritma klasik pada komentar TikTok masih terbatas [10]. Di sisi lain, penelitian forensik digital oleh [11] menyoroti bahwa pentingnya identifikasi bukti cyberbullying pada platform TikTok, namun pendekatan yang digunakan belum diarahkan pada pengembangan sistem klasifikasi otomatis berbasis NLP. Penelitian ini mengusulkan model hybrid IndoBERT-KNN sebagai alternatif, dimana dalam pendekatan ini, IndoBERT dimanfaatkan sebagai feature extractor untuk menghasilkan representasi vektor teks yang kaya secara semantik dan kontekstual, sehingga mampu menangkap makna komentar berdasarkan struktur kalimat dan konteks penggunaannya. Representasi semantik tersebut selanjutnya digunakan sebagai dasar pengukuran kedekatan antar komentar, yang kemudian dimanfaatkan oleh algoritma KNN dalam proses klasifikasi.

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif eksperimental yang bertujuan mengembangkan serta mengevaluasi sistem deteksi cyberbullying berbasis Natural Language Processing (NLP). Alur penelitian secara keseluruhan digambarkan dalam Gambar 1, yang dimulai dari pengumpulan data hingga evaluasi model.



Gambar 1. Alur penelitian.

2.1 Pengumpulan dan Persiapan Data

Fokus penelitian ini adalah pada komentar dari video publik bertema sosial, khususnya konten yang terkait dengan isu sosial populer dan berita selebriti yang viral. Video yang dianalisis dipilih berdasarkan tingkat interaksi pengguna yang tinggi serta indikasi kemunculan komentar dengan muatan perundungan. Proses pengumpulan data dilakukan secara otomatis melalui teknik *web scraping* menggunakan platform Apify, yang mendukung ekstraksi data dari media sosial TikTok [12]. Pendekatan ini digunakan untuk memperoleh data komentar beserta metrik interaksi seperti *like* dan *share* tanpa mengungkap identitas pribadi pengguna, sehingga tetap sesuai dengan prinsip etika dan anonimitas penelitian. Melalui proses tersebut, terkumpul sebanyak 1.079 komentar. Selanjutnya, dilakukan pelabelan manual oleh peneliti ke dalam dua kelas: cyberbullying (komentar mengandung penghinaan, pelecehan, intimidasi) dan non-cyberbullying (komentar netral, positif, atau diskusi sehat). Pelabelan mengacu pada pedoman Kementerian Komunikasi dan Informatika (Kominfo) Republik Indonesia terkait klasifikasi ujaran kebencian dan konten negatif di media sosial.

Setelah proses pelabelan manual selesai, Exploratory Data Analysis (EDA) dilakukan untuk memperoleh pemahaman mendalam tentang karakteristik dan struktur awal dataset [13]. EDA dilakukan sebagai tahap analisis awal untuk memahami dataset. Tahap ini meliputi pemeriksaan distribusi kelas guna mendeteksi ketidakseimbangan data, analisis panjang dan variasi komentar untuk menentukan strategi preprocessing, eksplorasi pola bahasa seperti slang dan kata kasar, serta identifikasi dan pembersihan noise seperti komentar duplikat atau kosong. Hasil EDA memberikan pemahaman awal yang mendukung tahapan *preprocessing* dan pemodelan selanjutnya.

2.2 Preprocessing

Dalam Natural Language Processing (NLP), makna komentar tidak hanya ditentukan oleh kata yang muncul, tetapi juga oleh struktur kalimat dan hubungan antar kata subjek, predikat, objek, keterangan (SPOK). Misalnya, kata yang sama bisa bersifat menghina dalam satu konteks, tetapi netral dalam konteks lain. Oleh karena itu, preprocessing dilakukan dengan hati-hati agar struktur SPOK dan konteks kalimat tetap terjaga.

Text Preprocessing merupakan tahapan kritis sebelum pemodelan machine learning untuk mempersiapkan dan membersihkan data teks dari *noise*, memerlukan pemahaman mendalam karakteristik dataset dan strategi yang tepat [14]. Proses ini meliputi langkah-langkah standar seperti case folding, cleaning, tokenisasi, normalisasi, dan stopword removal. Meskipun penelitian terkini menunjukkan efek preprocessing dapat bervariasi pada model transformer seperti BERT bergantung pada arsitektur model yang digunakan [15]. Dalam penelitian ini tahap preprocessing tetap dilaksanakan secara komprehensif untuk menjamin konsistensi data dan meminimalkan gangguan pada proses ekstraksi fitur oleh IndoBERT. Tahapan preprocessing data teks dalam penelitian ini meliputi beberapa Langkah seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Tahapan preprocessing.

Tahapan Preprocessing	Deskripsi	Contoh
Case Folding	Mengubah seluruh karakter menjadi huruf kecil (lowercase) untuk menyamakan bentuk kata dan menghindari perbedaan makna akibat variasi huruf besar–kecil	“DASAR Goblok!” menjadi “dasar goblok!”
Cleaning (URL, Emoji, Simbol)	Menghapus elemen non-teks yang tidak relevan, seperti URL, emoji, simbol, mention, hashtag, untuk mengurangi noise	“Dasar bodoh 🙄 🙄 !!!” menjadi “dasar bodoh”
Tokenisasi	Memecah teks menjadi unit kata (token) agar model bisa memahami hubungan antar kata	“kamu tidak berguna” menjadi [“kamu”, “tidak”, “berguna”]
Normalisasi Kata	Mengubah kata tidak baku atau bahasa informal menjadi bentuk baku sesuai kaidah Bahasa Indonesia	“gk” → “tidak”; “lu” → “kamu”; “bgt” → “banget”; “anjir” → “anjing”
Stopword Removal	Menghapus kata umum yang tidak memiliki pengaruh signifikan pada klasifikasi untuk fokus pada kata bermakna	“yang”, “dan”, “di”, “ke”, “dari”

2.3 Ekstraksi Fitur dan Embedding

Komentar hasil preprocessing kemudian ditransformasikan menjadi representasi vektor numerik menggunakan model IndoBERT sebagai *feature extractor*. IndoBERT, sebagai varian BERT yang telah dilatih (*pre-trained*) pada korpus bahasa Indonesia, menghasilkan vektor embedding berdimensi 768 untuk setiap komentar dengan mengambil representasi token [CLS] atau menghitung rata-rata hidden states lapisan terakhir. Vektor representasi tersebut memuat informasi semantik dan konteks teks yang berperan penting dalam mengenali bentuk cyberbullying yang kerap tersirat.

Tahap berikutnya adalah membagi embedding yang dihasilkan ke dalam data latih sebanyak 80% dan data uji sebesar 20%. Proses pembagian dilakukan pada level vektor untuk memastikan kesesuaian dengan

algoritma KNN yang beroperasi pada ruang fitur numerik. Data latih digunakan untuk membangun representasi kemiripan semantik sebagai dasar proses klasifikasi KNN, sedangkan data uji dimanfaatkan untuk menilai kemampuan model dalam menggeneralisasi dan mengklasifikasikan komentar yang belum pernah dikenali sebelumnya.

2.4 Klasifikasi KNN dan Hasil Prediksi

Vektor embedding hasil IndoBERT digunakan sebagai fitur masukan untuk algoritma K-Nearest Neighbor (KNN). KNN merupakan metode klasifikasi berbasis instans yang menentukan kelas data berdasarkan mayoritas kelas dari sejumlah K tetangga terdekat pada ruang fitur [16]. Pendekatan penggunaan representasi fitur berbasis IndoBERT yang dikombinasikan dengan KNN telah diterapkan pada penelitian klasifikasi sentimen berbahasa Indonesia dan menunjukkan kemampuan dalam merepresentasikan informasi kontekstual teks secara lebih baik [17]. Pada penelitian ini, nilai K ditentukan melalui eksperimen awal dan ditetapkan sebesar 5, sedangkan cosine similarity digunakan sebagai metrik jarak untuk mengukur kemiripan semantik antar vektor embedding. Proses klasifikasi dilakukan dengan menentukan kelas berdasarkan mayoritas dari K tetangga terdekat sebagaimana dirumuskan pada Persamaan (1).

$$Class(x_{\{test\}}) = \underset{c \in \{Bullying, Non-Bullying\}}{\operatorname{argmax}} \sum_{i=1}^{\{K\}} I(y_i = c) \quad (1)$$

Keterangan:

- $I(y_i = c) = 1$ jika tetangga ke-iii memiliki label C, 0 jika tidak.
- $\sum_{i=1}^{\{K\}} I(y_i = c)$ = jumlah tetangga dari K yang termasuk dalam kelas C
- $\operatorname{arg max}$ = memilih kelas yang memiliki jumlah tetangga terbanyak (mayoritas).

Dengan kata lain, KNN melihat 5 tetangga terdekat dari komentar baru dan memilih kelas yang paling sering muncul di antara mereka sebagai prediksi.

2.5 Evaluasi Model

Dataset embedding dibagi dengan rasio 80:20 untuk data latih dan data uji. Kinerja model hybrid dievaluasi menggunakan confusion matrix dan metrik turunannya: Accuracy, Precision, Recall, dan F1-Score untuk kelas bullying. Perhitungan metrik didasarkan pada nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN)[18].

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (3)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (4)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

3. HASIL PENELITIAN

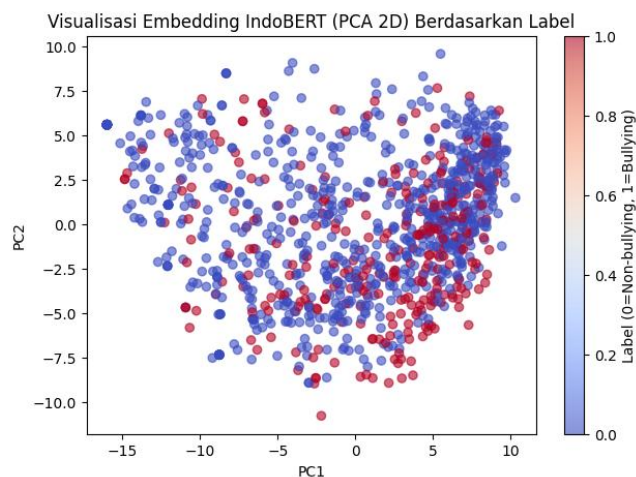
3.1 Karakteristik Dataset

Dataset akhir yang digunakan dalam penelitian ini terdiri dari 1079 komentar yang berhasil dikumpulkan dari platform TikTok. Hasil pelabelan manual menunjukkan distribusi kelas yang tidak seimbang (*imbalanced*), dengan 803 komentar (74.4%) tergolong dalam kategori non-cyberbullying dan 276 komentar (25.6%) tergolong sebagai cyberbullying. Karakteristik ini mencerminkan fenomena umum di media sosial, di mana interaksi negatif cenderung lebih sedikit namun memiliki dampak yang signifikan. Sebagian besar komentar memiliki panjang teks yang relatif pendek, sesuai dengan karakteristik komunikasi spontan dan cepat yang lazim ditemui pada platform TikTok. Distribusi ini menjadi pertimbangan penting

dalam evaluasi model, khususnya untuk metrik seperti *recall* dan F1-score yang sensitif terhadap ketidakseimbangan kelas.

3.2 Visualisasi Representasi Embedding

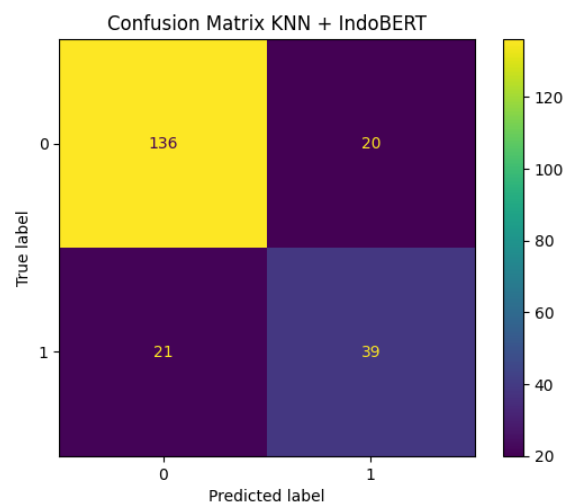
Untuk memahami sebaran data dalam ruang vektor, dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) dari 768 dimensi menjadi 2 dimensi. Hasil visualisasi pada Gambar 2 menunjukkan bahwa komentar dari kedua kelas membentuk pola pengelompokan (*clustering*) tertentu, meskipun masih terdapat area tumpang tindih (*overlap*). Hal ini mengindikasikan bahwa embedding IndoBERT mampu menangkap perbedaan semantik antara kedua kelas, namun kompleksitas dan ambiguitas bahasa di media sosial menyebabkan batasan yang tidak linier.



Gambar 2. Visualisasi Embedding IndoBERT Menggunakan PCA 2D Berdasarkan Label.

3.3 Hasil Klasifikasi dan Evaluasi Model

Model hybrid IndoBERT-KNN diuji pada 216 data uji (20% dari total dataset). *Confusion matrix* yang dihasilkan disajikan pada Gambar 3.



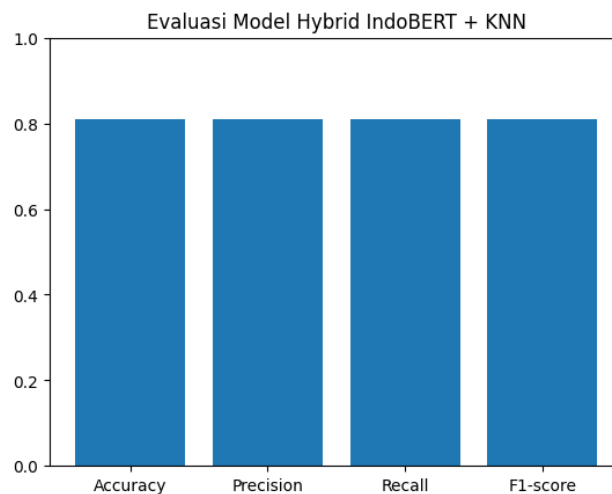
Gambar 3. Confusion Matrix hasil klasifikasi hybrid IndoBERT-KNN.

Berdasarkan Tabel 1, nilai *True Positive* (TP) adalah 39 dan *False Negative* (FN) adalah 21, yang berarti model berhasil mendeteksi 39 komentar bullying dengan benar, namun masih melewatkan 21 komentar bullying (salah klasifikasi sebagai non-bullying). Di sisi lain, model sangat baik dalam mengidentifikasi komentar non-bullying dengan *True Negative* (TN) sebanyak 136. Perhitungan metrik evaluasi menghasilkan nilai seperti yang dirangkum pada Tabel 2. Model mencapai akurasi keseluruhan sebesar 81.02%. Untuk kelas bullying (kelas minoritas), diperoleh presisi 0.66, *recall* 0.65, dan F1-score 0.66.

Tabel 2. Hasil Evaluasi Model Hybrid IndoBERT–KNN.

Metrik	Nilai
Accuracy	0.81
Precision	0.66
Recall	0.65
F1-score	0.66

Perbandingan visual dari keempat metrik utama ditampilkan pada Gambar 4. Grafik menunjukkan bahwa meskipun performa pada kelas bullying lebih rendah dibandingkan dengan akurasi keseluruhan, nilai presisi, *recall*, dan F1-score relatif seimbang, mengindikasikan bahwa model tidak terlalu bias terhadap satu metrik tertentu



Gambar 4. Perbandingan metrik evaluasi model Hybrid IndoBERT–KNN

4. DISKUSI

Hasil penelitian menunjukkan bahwa model hybrid IndoBERT-KNN dapat mendeteksi komentar cyberbullying pada TikTok dengan kinerja yang memadai (akurasi 81%). Kemampuan ini bersumber dari sinergi antara kekuatan kedua komponen: IndoBERT berhasil menghasilkan representasi vektor yang merefleksikan konteks dan makna semantik komentar, bahkan untuk teks informal dan singkat khas TikTok. Selanjutnya, KNN memanfaatkan representasi ini untuk mengelompokkan komentar baru berdasarkan kemiripannya dengan sampel yang telah dipelajari di ruang embedding. Nilai presisi 66% pada kelas bullying mengindikasikan bahwa dari semua komentar yang diprediksi sebagai bullying, sekitar dua pertiganya benar-benar bullying. Sementara *recall* 65% menunjukkan model mampu mengidentifikasi sekitar 65% dari total komentar bullying yang ada dalam dataset. Ketidakeimbangan distribusi kelas (hampir 3:1) diduga menjadi faktor yang mempengaruhi kinerja model pada kelas minoritas (bullying), di mana model cenderung lebih akurat dalam mendeteksi kelas mayoritas (non-bullying). Fenomena ini umum terjadi pada masalah klasifikasi dengan data *imbalanced*.

Jika dibandingkan dengan penelitian terdahulu, hasil akurasi ini sejalan dengan temuan [10] yang menggunakan hybrid CNN-LSTM dengan IndoBERT pada data Twitter (82.05%). Penelitian ini mengonfirmasi keefektifan IndoBERT untuk pemrosesan bahasa Indonesia informal. Namun, penelitian oleh [8] yang menggunakan IndoBERT secara *fine-tuned* melaporkan akurasi lebih tinggi (96.7%). Perbedaan ini dapat disebabkan oleh beberapa faktor: 1) Perbedaan karakteristik dataset (Twitter vs. TikTok), 2) Skala dan kualitas dataset, 3) Pendekatan *fine-tuning* vs. *feature extraction* yang digunakan dalam penelitian ini. Keunggulan pendekatan hybrid yang diusulkan terletak pada efisiensi komputasi, karena tidak memerlukan *fine-tuning* IndoBERT yang mahal sumber dayanya, serta kemudahan interpretasi klasifikasi berbasis kemiripan.

Tantangan utama yang teridentifikasi adalah deteksi terhadap komentar bullying yang implisit, bersifat sarkastik, atau menggunakan bahasa figuratif. *False Negative* yang terjadi (21 kasus) banyak berasal dari komentar semacam ini, di mana makna negatifnya sangat bergantung pada konteks dan pengetahuan dunia nyata yang mungkin belum sepenuhnya tertangkap oleh model. Selain itu, *False Positive* (20 kasus) dapat terjadi pada komentar non-bullying yang menggunakan kata-kata keras namun dalam konteks candaan atau ekspresi yang tidak ditujukan untuk merendahkan.

5. KESIMPULAN

Berdasarkan hasil penelitian, model hybrid yang mengombinasikan IndoBERT sebagai feature extractor dan KNN sebagai classifier berhasil diterapkan untuk deteksi cyberbullying pada komentar TikTok berbahasa Indonesia. Melalui tahapan preprocessing, ekstraksi fitur kontekstual, dan klasifikasi berbasis kemiripan, yang terbukti efektif dalam mengelompokkan komentar bullying dan non-bullying dengan performa yang relatif baik. Hasil evaluasi menunjukkan akurasi sebesar 81%, dengan nilai presisi, recall, dan F1-score pada kelas bullying masing-masing sebesar 0,66, 0,65, dan 0,66. Visualisasi PCA juga menunjukkan pola pemisahan data yang cukup jelas meskipun masih terdapat ambiguitas pada beberapa komentar. Temuan ini menunjukkan bahwa pendekatan hybrid IndoBERT–KNN efektif untuk klasifikasi teks pada domain cyberbullying berbahasa Indonesia dan berpotensi menjadi pendekatan yang relevan untuk pengembangan sistem moderasi komentar di platform TikTok.

DAFTAR PUSTAKA

- [1] E. Sylvain and M. Talpade, "Exploring The Characteristics of Cyberbullied TikTokers Based on Their Ethnicity," *Int. J. Arts, Humanit. Soc. Sci.*, vol. 05, no. 06, pp. 49–53, 2024, doi: 10.56734/ijahss.v5n6a8.
- [2] F. G. Aser, S. Paramita, and Sudarto, "Fenomena Cyberbullying di Media Sosial TikTok," *Kiwari*, vol. 1, no. 3, pp. 449–453, 2022, doi: 10.24912/ki.v1i3.15763.
- [3] D. Cyberbullying, T. Mental, and D. A. N. Perilaku, "Triwikrama : Jurnal Ilmu Sosial," vol. 6, no. 1, pp. 42–56, 2024.
- [4] Sadaruddin, A. F. K. Nur, Kasmawati, and K. N. Fitrah, "Cyberbullying: Ancaman Mental Siswa di Era Digital," *J. Pendidik. dan Teknol.*, vol. 1, no. April, pp. 16–25, 2025, [Online]. Available: <http://jurnal-fkip-uim.ac.id/index.php/teknos/article/view/394>
- [5] H. Saleous, M. Gergely, and K. Shuaib, "Review Article Exploring NLP-Based Solutions to Social Media Moderation Challenges," vol. 2025, 2025, doi: 10.1155/hbe2/9436490.
- [6] N. F. Hasan and V. Wati, "Deteksi Cyberbullying pada Facebook Menggunakan Algoritma K-Nearest Neighbor-Detect Cyberbullying on Facebook Using K-Nearest Neighbor Algorithm," vol. 1, no. 1, pp. 35–44, 2021.
- [7] I. S. Arfan, S. Fauziah, and I. Nawangsih, "Sentiment Analyst of Cyber Bullying in X Using Naïve Bayes Algorithm Analisa Sentimen Terhadap Cyber Bullying di X Menggunakan Algoritma Naïve Bayes," vol. 4, no. October, pp. 1411–1419, 2024.
- [8] Y. D. Novandian *et al.*, "IndoBERT-based Indonesian Cyberbullying Detection with Multi-stage Labeling," *Proc. - 2024 Int. Semin. Appl. Technol. Inf. Commun. Smart Emerg. Technol. a Better Life, iSemantic 2024*, no. December, pp. 515–521, 2024, doi: 10.1109/iSemantic63362.2024.10762553.
- [9] Kurniawanda, R., Tobing, and A. A., "Cyberbullying Detection on Instagram Using XGBoost," *Indones. J. Data Sci.*, vol. 3, no. 2, pp. 85–96, 2023.
- [10] A. A. Hafiza and E. B. Setiawan, "Peningkatan Deteksi Cyberbullying pada Platform ' X ' Menggunakan IndoBERT dan Model Hybrid CNN-LSTM," vol. 6, no. 2, pp. 655–672, 2025.
- [11] P. Rona and I. Riadi, "Cyberbullying Detection on TikTok using Association of Chief Police Officers," *Int. J. Comput. Appl.*, vol. 186, no. 3, pp. 6–13, 2024, doi: 10.5120/ijca2024923359.
- [12] I. Jufri, Arianto, and A. F. Sonni, "Analisis Respon Publik dan Sentimen Video Berbagi Willy Salim yang Viral di TikTok," *J. Ris. Jurnalistik dan Media Digit.*, pp. 77–88, 2025, doi: 10.29313/jrjmd.v5i1.7017.
- [13] A. Pramudyantoro, E. Utami, and D. Ariatmanto, "Penggabungan K-Nearest Neighbors Dan Lightgbm Untuk Prediksi Diabetes Pada Dataset Pima Indians: Menggunakan Pendekatan

- Exploratory Data Analysis,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 3, pp. 1133–1144, 2024, doi: 10.29100/jipi.v9i3.4966.
- [14] S. García-méndez, F. De Arriba-pérez, and E. Costa-montenegro, “Special Issue on Advancements in Natural Language Processing , Semantic Networks , and Sentiment Analysis,” 2025.
- [15] A. Kurniasih and L. P. Manik, “On the Role of Text Preprocessing in BERT Embedding-based DNNs for Classifying Informal Texts,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, pp. 927–934, 2022, doi: 10.14569/IJACSA.2022.01306109.
- [16] E. F. Fasttext, T. D. A. N. Indobert, P. Metode, and K. N. Untuk, “Eksplorasi fitur fasttext, tf-idf dan indobert pada metode k-nearest neighbor untuk klasifikasi sentimen 1,2,3,4,” vol. 7, no. 1, pp. 49–60.
- [17] M. H. Fiqri, R. Rudiman, and N. A. Verdikha, “Analisis Model Klasifikasi Sentimen Publik Terhadap Kebijakan Keberlanjutan IKN Menggunakan BERT Sebagai Feature Extractor dan K-Nearest Neighbor (KNN),” vol. 7, no. 2, pp. 1332–1342, 2025, doi: 10.47065/bits.v7i2.8168.
- [18] K. Hadi and E. Utami, “Analysis of K-NN with the Integration of Bag of Words , TF-IDF , and N-Grams for Hate Speech Classification on Twitter,” vol. 12, no. 2, pp. 289–298, 2024.