

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi dan internet telah mengubah cara manusia berkomunikasi dan berinteraksi. Salah satu wujud dari perkembangan ini adalah munculnya media sosial, yakni platform digital yang memungkinkan penggunanya saling berinteraksi, berkomunikasi, dan berbagi beragam konten seperti tulisan, gambar, suara, hingga video secara online (Ju dkk., 2023). Kehadiran media sosial membuat komunikasi menjadi lebih sederhana dan efektif serta memungkinkan penyebaran informasi secara luas dan cepat. Selain sebagai sarana komunikasi, media sosial kini memegang peran penting dalam berbagai bidang seperti pendidikan, bisnis, hiburan, dan kampanye sosial. Platform yang paling banyak digunakan saat ini antara lain WhatsApp, TikTok, Facebook, YouTube, Twitter, dan Instagram (Santoso dkk., 2023).

Platform yang saat ini diminati adalah TikTok, dengan jumlah pengguna aktif bulanan yang telah melampaui 1 miliar. Popularitas TikTok yang terus meningkat menjadikannya sebagai salah satu platform utama dalam interaksi digital, khususnya di kalangan generasi muda dan remaja. TikTok merupakan media sosial dengan beragam jenis konten yang memungkinkan penggunanya untuk berkreasi, dan menikmati berbagai video singkat dengan berbagai tema dan kreativitas. Di sisi lain, popularitas platform ini juga meningkatkan risiko penyebaran komentar berunsur *cyberbullying*. Sifat kontennya yang interaktif serta adanya fitur anonimitas membuat beberapa pengguna merasa bebas melakukan tindakan perundungan tanpa takut terkena dampak hukum secara langsung (Halimah dkk., 2024). Bentuk-bentuk komentar yang tergolong sebagai *cyberbullying* di TikTok sangat beragam dan umumnya bersifat langsung menyerang, seperti penghinaan fisik (*body shaming*), ejekan, hinaan verbal, serta komentar yang merendahkan harga diri, kemampuan, atau status sosial seseorang. Contohnya seperti komentar: “WKwkw kek udh tinggi bat yaa tu orang”, atau “maklum orkay” merupakan bentuk ejekan yang menyiratkan kritik atau

merendahkan seseorang secara tidak langsung. Meskipun tidak mengandung hinaan kasar secara eksplisit, komentar semacam ini mencerminkan bahwa bentuk *cyberbullying* tidak selalu harus bersifat *vulgar*, tetapi juga bisa berupa sindiran, ejekan halus, atau komentar pasif-agresif yang tetap berdampak negatif secara psikologis (Budiasa dkk., 2021).

Cyberbullying merupakan bentuk perundungan yang terjadi di dunia digital melalui berbagai platform online. Fenomena ini telah berkembang menjadi isu krusial yang berdampak pada berbagai aspek kehidupan, terutama di kalangan anak muda, termasuk pelajar dan mahasiswa (Nugraha & Astuti, 2023). Dalam konteks *cybersecurity*, *cyberbullying* dikategorikan sebagai kejahatan siber berbasis konten (*content-related cybercrime*), yaitu kejahatan yang terjadi melalui media sosial dalam bentuk penghinaan, pelecehan, ancaman, serta penyebaran informasi palsu yang dapat merugikan individu atau kelompok tertentu (Halimah dkk., 2024). Akibatnya, korban tidak hanya mengalami tekanan mental, seperti stres, kecemasan, dan depresi, tetapi juga menghadapi gangguan dalam kehidupan sosial maupun akademik mereka (Nugraha & Astuti, 2023). Interaksi antara pelaku dan korban dalam *cyberbullying* bersifat reaktif, di mana semakin besar intensitas serangan pelaku, semakin kuat pula respon emosional korban, sehingga meningkatkan efek negatif dalam lingkungan digital. Oleh karena itu, diperlukan upaya pencegahan melalui peningkatan kesadaran, penerapan regulasi yang lebih ketat untuk mengurangi kejadian *cyberbullying* di dunia maya (Awalia Z Abidin dkk., 2023).

Semakin maraknya kasus *cyberbullying* beserta dampak buruk yang ditimbulkannya menunjukkan pentingnya sistem yang mampu mendeteksi, mengidentifikasi, dan mengklasifikasikan komentar-komentar yang mengandung unsur perundungan di media sosial. Komentar yang bersifat menyerang tidak selalu disampaikan secara *eksplisit* dan terang-terangan, melainkan bisa hadir dalam berbagai bentuk ujaran negatif, seperti hinaan, ejekan, atau sindiran. Tantangan utama dalam deteksi *cyberbullying* terletak pada keberagaman ekspresi bahasa yang digunakan oleh pelaku, sehingga diperlukan pendekatan klasifikasi teks yang

efektif untuk membedakan antara komentar yang bersifat *bullying* dan komentar yang tidak mengandung unsur perundungan (Awalia Z Abidin dkk., 2023).

Salah satu metode yang umum digunakan untuk mendeteksi *cyberbullying* adalah *machine learning*, di mana algoritma seperti *Naïve Bayes*, *Support Vector Machine* (SVM), *Random Forest*, dan *K-Nearest Neighbor* (KNN) termasuk yang paling diterapkan. Masing-masing algoritma memiliki pendekatan dan karakteristik tersendiri dalam membedakan komentar yang termasuk kategori *cyberbullying* dan yang tidak, tergantung pada parameter, fitur, serta dataset yang digunakan dalam pelatihan model (Setiawan dkk., 2022).

Beberapa penelitian sebelumnya telah menerapkan algoritma *K-Nearest Neighbor* (KNN) dalam mendeteksi *cyberbullying*. Salah satu penerapannya, adalah dalam analisis sentimen komentar di TikTok, yang menunjukkan hasil yang cukup baik sebesar 91% (Wijaya & Suwandhi, 2024). Sementara itu, penelitian lain pada Instagram menggunakan algoritma *K-Nearest Neighbor* (KNN) untuk membedakan komentar *bullying* dan *non-bullying*, dengan hasil akurasi rata-rata 58,83%; akurasi tertinggi 77% ($k=13$, data 90:10, fold ke-6); akurasi terendah 35% ($k=11$, data 90:10, fold ke-1) (Candra & Nanda Rozana, 2020). Meskipun *K-Nearest Neighbor* (KNN) menunjukkan performa yang cukup baik, algoritma ini masih memiliki keterbatasan, terutama dalam menangani dataset yang besar serta menentukan nilai k yang optimal agar dapat meningkatkan akurasi dalam mendeteksi *cyberbullying*.

Aktivitas komentar pada satu konten atau akun TikTok dapat mencapai lebih dari 1.000 komentar. Jumlah ini tergolong sebagai dataset skala sedang yang masih dapat diolah secara efektif menggunakan algoritma *K-Nearest Neighbor* (KNN). Selain itu, data yang dihasilkan dari media sosial umumnya berbentuk teks tidak terstruktur, sehingga memerlukan tahapan *pre-processing* sebelum dapat diklasifikasikan menggunakan algoritma *machine learning* (Setiawan dkk., 2022). Dalam konteks ini, algoritma KNN tetap relevan untuk digunakan dalam mendeteksi komentar yang mengandung unsur *cyberbullying* di TikTok.

Dalam penelitian berbasis klasifikasi teks, istilah *ground truth* mengacu pada label data yang dijadikan acuan kebenaran dalam proses pelatihan dan evaluasi

model. Keberadaan *ground truth* sangat penting untuk menjamin validitas hasil klasifikasi. Ketidaktepatan atau ketidakkonsistenan label dapat menurunkan akurasi model serta menimbulkan bias dalam prediksi (Stutz dkk., 2025).

Berdasarkan permasalahan tersebut, algoritma *K-Nearest Neighbor* (KNN) telah terbukti memiliki performa yang cukup baik dalam mendeteksi *cyberbullying*. Namun, masih terdapat beberapa tantangan, seperti penurunan performa saat diterapkan pada dataset berukuran besar. Dengan itu, penelitian ini dilakukan untuk menilai kembali performa algoritma *K-Nearest Neighbor* (KNN) pada dataset baru yang belum memiliki *ground truth*.

1.2 Perumusan Masalah

Meskipun algoritma *K-Nearest Neighbor* (KNN) telah menunjukkan performa yang baik dalam mendeteksi komentar *cyberbullying* pada dataset standar dalam pembelajaran mesin, performanya belum dievaluasi secara menyeluruh pada dataset baru yang belum memiliki *ground truth* yang divalidasi.

1.3 Pertanyaan Penelitian

Adapun pertanyaan yang menjadi fokus dalam penelitian ini meliputi:

1. Bagaimana implementasi algoritma *K-Nearest Neighbor* (KNN) untuk mendeteksi *cyberbullying* di TikTok?
2. Bagaimana performa algoritma *K-Nearest Neighbor* (KNN) dalam mendeteksi *cyberbullying* di TikTok berdasarkan akurasi?

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mengkaji efektivitas *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan komentar bermuatan *cyberbullying* pada dataset yang berasal dari sumber data aktual dan belum tervalidasi sebelumnya.

1.5 Manfaat Hasil Penelitian

Penelitian ini diharapkan dapat memberikan manfaat berupa:

1. Memberikan wawasan mengenai performa algoritma *K-Nearest Neighbor* (KNN) dalam mendeteksi komentar dari dataset yang belum memiliki *ground truth*.

2. Menjadi referensi dalam deteksi perundungan digital di media sosial melalui pemanfaatan algoritma *machine learning* yang sesuai dan cukup baik, khususnya dalam menangani dataset yang belum memiliki *ground truth* sebelumnya.

PERPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANI
YOGYAKARTA