

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Dengan perkembangan teknologi dan komunikasi membawa dampak perubahan yang signifikan terhadap interaksi sosial, khususnya di kalangan remaja. Media sosial, forum online, dan aplikasi pesan instan menjadi sarana komunikasi yang dominan. Namun, di balik manfaatnya, muncul masalah serius seperti *cyberbullying*. Menurut laporan dari *National Center for Education Statistic* 2022, sekitar 19,2% siswa melaporkan pernah mengalami *bullying* secara online (Nces, 2022). Menurut laporan dari Avast (2023), dalam beberapa tahun terakhir, jumlah kasus *cyberbullying* mengalami peningkatan, 37% remaja di seluruh dunia melaporkan pernah menjadi korban *cyberbullying*, dengan 59% diantaranya mengalami dampak negatif pada kesehatan mental. Selain itu, terdapat 70% remaja merasa bahwa *cyberbullying* adalah masalah serius, tetapi hanya 30% yang melaporkan insiden tersebut. Fenomena ini berdampak terhadap kesehatan mental dan sosial para korban.

Berdasarkan Studi global yang oleh *EClinicalMedicine* mengungkapkan bahwa 30,5% remaja di 83 negara melaporkan pernah menjadi korban *bullying*. Di Amerika Serikat, persentase siswa perempuan yang melaporkan *bullying* lebih tinggi 21,8% dibandingkan siswa laki-laki 16,7% (Biswas dkk., 2020). Bentuk *cyberbullying* yang paling umum meliputi komentar menyakitkan di media sosial (30,4%), pengucilan dari grup chat (28,9%), dan penyebaran rumor online (28,4%) (Patchin & Hinduja, 2024). Dampak dari *cyberbullying* sangat serius, baik secara psikologis maupun akademis. *National Center for Education Statistics* (2022) melaporkan bahwa korban *cyberbullying* seringkali mengalami penurunan harga diri, kesulitan dalam mengerjakan tugas sekolah dan gangguan dalam hubungan dengan keluarga dan teman.

Cyberbullying merupakan bentuk tindakan intimidasi, pelecehan, atau penindasan yang dilakukan melalui media digital. Meningkatnya kasus *cyberbullying* menjadi kekhawatiran yang serius tentang dampak pelecehan online

terhadap kesehatan mental dan keamanan individu. Salah satu pemicu utama adalah kurangnya pemahaman tentang etika berkomunikasi di platform ini, yang menciptakan lingkungan yang mendukung perilaku berbahaya dan merugikan. *Cyberbullying* dapat menyebabkan gangguan psikologis, kecemasan, depresi, hingga resiko bunuh diri pada korban, terutama dikalangan remaja dan dewasa (Purnayasa dkk., 2022). Oleh karena itu, diperlukan deteksi dini terhadap *cyberbullying* untuk menciptakan lingkungan digital yang lebih aman.

Untuk mengatasi permasalahan ini, deteksi manual tidak efektif karena volume data yang besar dan kompleksitas bahasa yang digunakan. Pendekatan berbasis Pemrosesan Bahasa Alami (NLP) sangat diperlukan untuk mendeteksi *cyberbullying*. Beberapa studi telah mengimplementasikan TF-IDF (*Term Frequency-Inverse Document Frequency*) sebagai ekstraksi fitur untuk mengidentifikasi teks yang mengandung ujaran *cyberbullying*. TF-IDF berfungsi untuk mengidentifikasi kata-kata yang memiliki bobot lebih tinggi dalam teks. Penerapan TF-IDF dalam *Analysis Sentiment Cyberbullying* menggunakan XGBoost menunjukkan bahwa metode ini efektif dalam mengidentifikasi pola teks yang relevan (Riza Kurniawanda et al., 2022).

Deteksi *cyberbullying* yang hanya menggunakan metode TF-IDF menghasilkan deteksi yang tidak akurat. Akibatnya, kata-kata yang sebenarnya tidak mengandung unsur *cyberbullying* bisa saja terdeteksi sebagai *cyberbullying*, atau sebaliknya. Untuk meningkatkan akurasi deteksi, digunakan *Machine Learning* dengan metode klasifikasi SVM (*Support Vector Machine*) sebagai classifier. Romindo melakukan pengujian TF-IDF dengan SVM dalam menganalisis komentar di media sosial TikTok dan menemukan bahwa kombinasi ini mampu meningkatkan performa klasifikasi dengan akurasi 88% (Romindo dkk., 2023). Selain itu, penelitian menunjukkan bahwa TF-IDF dengan SVM dapat memberikan hasil yang cukup akurat dalam mendeteksi ujaran kebencian di media sosial (Prabowo & Azizah, 2020).

Urgensi penelitian ini terletak pada kebutuhan sistem deteksi otomatis yang akurat, mengingat tingginya kasus *cyberbullying* dengan dampak psikologis yang serius, keterbatasan deteksi manual dalam menangani volume data yang besar,

kompleksitas bahasa alami yang membutuhkan pendekatan otomatis, serta potensi signifikan kombinasi TF-IDF dengan SVM yang telah terbukti efektif mengidentifikasi pemrosesan teks.

Meskipun demikian, terdapat beberapa tantangan dalam penerapan metode ini. Model klasifikasi berbasis statistik seringkali kesulitan dalam membedakan ujaran bercanda dan ujaran yang bersifat merendahkan (Alqahtani & Ilyas, 2024). Beberapa algoritma masih memiliki tingkat *false positive* yang tinggi, di mana teks yang sebenarnya tidak mengandung *cyberbullying* dapat salah diklasifikasikan (Julianto dkk., 2021). Dengan demikian, penelitian ini bertujuan untuk menerapkan metode TF-IDF dalam pendeteksian *cyberbullying* dengan memanfaatkan algoritma *Support Vector Machine* (SVM). Pendekatan ini diharapkan mampu meningkatkan akurasi dalam klasifikasi teks yang mengandung ujaran *cyberbullying*.

1.2 Perumusan Masalah

Deteksi *cyberbullying* dengan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) memiliki kekurangan terkait masih tingginya nilai *false detection* dalam konteks klasifikasi *multiclass*, serta masih terdapat ruang untuk peningkatan kinerja klasifikasi dengan mengintegrasikan metode TF-IDF dan algoritma SVM.

1.3 Pertanyaan Penelitian

Penelitian ini memiliki beberapa pertanyaan penelitian yaitu:

1. Bagaimana efektivitas metode TF-IDF dalam mengekstraksi fitur teks yang mengandung unsur *cyberbullying*?
2. Bagaimana kinerja metode SVM dalam meningkatkan akurasi deteksi *cyberbullying* dibandingkan dengan metode TF-IDF saja?
3. Bagaimana akurasi kombinasi TF-IDF dan SVM mampu mengurangi *false detection* dalam klasifikasi *cyberbullying*?

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan model deteksi *cyberbullying* menggunakan metode TF-IDF dalam *ekstraksi* fitur teks terkait *cyberbullying* dan *Support Vector Machine* (SVM) sebagai algoritma klasifikasi.

1.5 Manfaat Hasil Penelitian

Penelitian ini diharapkan mampu memberikan manfaat, di antaranya berupa:

1. Memberikan wawasan mengenai kinerja TF-IDF dalam mendeteksi *cyberbullying* menggunakan algoritma *Support Vector Machine* (SVM).
2. Menjadi referensi bagi penelitian selanjutnya yang berfokus pada deteksi *cyberbullying*, khususnya dalam penggunaan metode *ekstraksi* fitur dan algoritma *Support Vector Machine* (SVM).