

BAB I PENDAHULUAN

1.1 Latar Belakang

Pada era digital saat ini, media sosial sudah menjadi bagian tak terpisahkan dari kehidupan masyarakat. Salah satu platform yang mengalami pertumbuhan pesat adalah TikTok, yang memungkinkan penggunaannya berbagi video singkat sekaligus berinteraksi melalui kolom komentar. Berdasarkan laporan data dan statistik, Statista per Februari 2025 yang berjudul *"Countries with the largest TikTok audience as of February 2025"*, Indonesia menempati posisi kedua pengguna TikTok terbanyak di dunia dengan jumlah 107,7 juta pengguna setelah Amerika Serikat [1]. Popularitas TikTok tersebut turut menghadirkan gerak interaksi yang beragam, termasuk munculnya berbagai bentuk perilaku negatif di ruang digital, salah satunya berupa pelecehan seksual verbal di kolom komentar. Fenomena ini kerap ditemukan terutama pada konten publik figur Korea, seperti aktor, aktris, maupun idol K-Pop, yang memiliki basis penggemar besar di Indonesia seiring bertumbuhnya tren budaya Korea yang sedang populer [2]. Sayangnya meningkatnya popularitas budaya Korea (Korean Wave), tidak selalu menimbulkan interaksi yang bersifat positif. Keterlibatan emosional penggemar membuat beberapa pengguna mengekspresikan ketertarikannya melalui komentar berunsur seksual baik implisit maupun eksplisit yang tidak pantas

Pelecehan seksual verbal dalam kolom komentar dapat muncul dalam berbagai bentuk, mulai dari ungkapan yang menjurus pada tubuh, penampilan fisik, kata-kata vulgar, fantasi seksual, komentar bernuansa seksual, serta merendahkan individu secara seksual. Bentuk pelecehan seksual verbal ini tidak selalu eksplisit, dapat pula berupa candaan seksual maupun ungkapan sugestif (ajakan) yang disamarkan menggunakan bahasa slang, singkatan, metafora, atau istilah tertentu yang umum dalam komunikasi daring. Dampak dari pelecehan ini tidak bisa dianggap sepele. Dari sudut pandang korban baik perempuan maupun laki-laki, dapat menimbulkan rasa tidak nyaman dan tekanan psikologis hingga

mempengaruhi kualitas interaksi mereka dalam menggunakan media sosial. Selain itu, lingkungan digital yang semakin tidak sehat menjadi berisiko, terlebih karena pengguna TikTok tidak hanya orang dewasa, tetapi juga remaja yang rentan terpapar konten negatif [3]. Dari perspektif pengguna media sosial lainnya, yang membaca komentar tidak pantas juga dapat merasakan ketidaknyamanan, apalagi ketika komentar tersebut muncul berulang, sehingga menciptakan interaksi yang tidak sehat.

Dalam konteks tertentu, komentar bernuansa seksual dibungkus dalam bentuk candaan, pujian atau ekspresi kekaguman, sehingga dianggap wajar oleh sebagian pengguna. Fenomena ini berkaitan dengan perubahan norma sosial, yaitu menormalisasi perilaku yang sebelumnya dianggap tidak pantas. Komentar yang dinormalisasi ini dapat dengan mudah menyebar karena dianggap wajar, diperkuat minimnya konsekuensi di dunia maya, membuat pengguna merasa bebas mengekspresikan komentar tanpa mempertimbangkan dampak etis. Dampaknya, batas pujian maupun kekaguman semakin kabur, dan komentar berunsur objektifikasi seksual dianggap menjadi bagian lumrah dari interaksi di media sosial. Meskipun TikTok telah menerapkan sistem moderasi komentar otomatis, masih terdapat banyak keterbatasan dalam mendeteksi pelecehan seksual verbal, terutama jika disamarkan melalui bahasa informal yang umum ditemukan di media sosial, istilah tersirat, atau kode tertentu. Oleh karena itu, dibutuhkan pendekatan berbasis kecerdasan buatan yang mampu mendeteksi komentar bermuatan pelecehan seksual verbal secara lebih akurat, memahami konteks bahasa lebih mendalam serta adaptif terhadap karakteristik teks media sosial.

Dalam bidang analisis teks, berbagai algoritma *machine learning* telah digunakan untuk melakukan tugas klasifikasi, beberapa diantaranya seperti analisis sentimen, deteksi ujaran kebencian, maupun cyberbullying. Dua algoritma yang banyak diaplikasikan adalah Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN). SVM dikenal unggul dalam menangani data berdimensi tinggi serta menghasilkan klasifikasi yang akurat meskipun dengan jumlah data relatif terbatas [4]. SVM bekerja dengan fokus pada pencarian hyperplane optimal yang

dapat memisahkan kelas data dengan margin maksimum, sehingga model pemisahan kelas menjadi lebih presisi dan robust terhadap variasi data ambigu atau tumpang tindih. Dalam konteks media sosial TikTok yang bersifat singkat, informal, mengandung variasi bahasa, pola pemisahan antar kelas tidak dapat dilakukan secara linear. Oleh karena ini, pada penelitian ini digunakan svm dengan kernel Radial Basis Function (RBF), yang dipilih karena kemampuannya dapat memodelkan fitur dari ruang berdimensi rendah ke ruang berdimensi lebih tinggi secara non-linear. Dengan kemampuan dalam membuat batas non-linear yang dimiliki kernel RBF, SVM diharapkan dapat membangun model dengan batas margin yang lebih fleksibel dan adaptif terhadap variasi bahasa pada komentar TikTok.

Sementara itu, KNN merupakan algoritma berbasis jarak, bekerja dengan mengklasifikasikan data baru berdasarkan kedekatannya dengan sejumlah tetangga terdekat pada data latih. KNN algoritma yang sederhana, fleksibel, dan efektif dalam mendeteksi pola pada data, meskipun performanya dapat dipengaruhi oleh jumlah data dan pemilihan parameter k yang dipilih secara optimal untuk mendapatkan kinerja klasifikasi terbaik [5]. KNN dipilih karena kemampuannya dalam membandingkan kemiripan semantik antar dokumen atau kalimat, terutama jika representasi fitur yang digunakan mampu mencerminkan makna teks dengan baik. Kedua algoritma ini memiliki kelebihan dan kekurangan masing-masing. SVM cenderung lebih baik dalam menangani kasus klasifikasi kompleks, namun membutuhkan waktu komputasi lebih tinggi saat data besar. Di sisi lain, KNN lebih mudah diimplementasikan dan hasilnya bergantung pada distribusi data serta ukuran tetangga terdekat yang digunakan.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk memperoleh model terbaik dalam mendeteksi komentar bermuatan pelecehan seksual pada konten TikTok mengenai publik figur Korea melalui perbandingan kinerja dua algoritma klasifikasi, yaitu SVM dan KNN. Untuk meningkatkan kualitas hasil klasifikasi, penelitian ini menerapkan teknik pendekatan representasi teks berbasis LaBSE (*Language-agnostic BERT Sentence Embedding*) yang mampu menangani

karakteristik komentar media sosial yang singkat, multilingual, dan terdapat banyak variasi simbol. Kemampuan ini relevan dengan karakteristik komentar TikTok yang multilingual, menggabungkan bahasa Indonesia, Inggris, bahasa informal, maupun istilah asing lainnya. Penggunaan LaBSE diharapkan dapat menghasilkan representasi fitur yang tidak hanya melihat kemunculan kata tetapi juga makna semantik keseluruhan kalimat, sehingga kedua algoritma dapat bekerja dengan lebih optimal dalam mengenali komentar bermuatan pelecehan seksual. Pada SVM, representasi LaBSE membantu dalam membentuk batas keputusan yang lebih jelas antar kelas, sedangkan pada KNN, embedding meningkatkan kualitas pengukuran kemiripan antar komentar. Berdasarkan pertimbangan tersebut pendekatan ini diharapkan mampu memberikan gambaran komprehensif mengenai kinerja kedua algoritma dalam mendeteksi komentar bermuatan pelecehan seksual verbal pada konten TikTok mengenai publik figur Korea. Melalui evaluasi terhadap skenario pemodelan yang dilakukan, penelitian ini bertujuan menemukan algoritma yang memberikan performa klasifikasi paling optimal. Hasil penelitian ini diharapkan tidak hanya berkontribusi pada pengembangan kajian akademik, tetapi juga memberikan dasar teknis dalam pengembangan sistem deteksi otomatis yang lebih adaptif dan sensitif terhadap pola bahasa pengguna di platform digital, sehingga dapat mendukung terciptanya ruang digital yang lebih aman.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah pada penelitian ini yaitu :

1. Bagaimana kinerja algoritma *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) dalam mendeteksi komentar bermuatan pelecehan seksual pada konten TikTok mengenai publik figur Korea?
2. Algoritma manakah yang memiliki performa lebih efektif dalam mengidentifikasi komentar bermuatan pelecehan seksual pada platform TikTok?

1.3 Batasan Masalah

Adapun batasan masalah dari penyusunan penelitian skripsi ini adalah:

1. Penelitian ini menggunakan data teks komentar TikTok mengenai publik figur Korea Selatan
2. Analisis difokuskan pada algoritma *Support Vector Machine* dengan *kernel RBF* dan *K-Nearest Neighbor*
3. Metrik evaluasi yang digunakan adalah *confusion matrix*
4. Implementasi dan pengujian algoritma akan dilakukan menggunakan bahasa pemrograman python

1.4 Tujuan Penelitian

Adapun tujuan dari penyusunan penelitian skripsi ini adalah:

1. Menganalisis hasil kinerja algoritma *Support Vector Machine* dan *K-Nearest Neighbor* dalam mendeteksi komentar bermuatan seksual pada platform TikTok.
2. Membandingkan kinerja kedua algoritma untuk mengidentifikasi algoritma yang paling efektif dalam melakukan klasifikasi data komentar berunsur pelecehan seksual verbal.
3. Mengevaluasi hasil klasifikasi menggunakan matriks evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* guna melihat kualitas prediksi algoritma *Support Vector Machine* dan *K-Nearest Neighbor*

1.5 Manfaat Penelitian

Adapun manfaat dari penyusunan penelitian skripsi ini adalah:

1. Memberikan analisis mendalam mengenai kinerja algoritma *Support Vector Machine* dan *K-Nearest Neighbor* dalam mengklasifikasikan komentar bermuatan pelecehan seksual.
2. Berkontribusi pada pengembangan kajian akademik, memberikan dasar

teknis dalam pengembangan sistem deteksi otomatis yang lebih adaptif dan sensitif terhadap pola bahasa pengguna di platform digital.

3. Menjadi referensi untuk penelitian selanjutnya khususnya dalam pengembangan sistem deteksi teks berbasis *machine learning*.

1.6 Sistematika Penulisan

Penulisan penelitian ini mencakup 5 bab pokok pembahasan

BAB I PENDAHULUAN, berisi Latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

BAB II TINJAUAN PUSTAKA, berisi teori-teori dasar machine learning, algoritma *Support Vector Machine*, *K-Nearest Neighbor*, dan penelitian-penelitian terdahulu mengenai komparasi algoritma tersebut.

BAB III METODE PENELITIAN, bab ini menjelaskan mengenai metodologi yang digunakan dalam penelitian, termasuk pengumpulan data, pra-pemrosesan data, pelatihan, pengujian model, dan deployment.

BAB IV HASIL DAN PEMBAHASAN, bab ini berisi analisis kinerja kedua algoritma dalam klasifikasi data komentar.

BAB V PENUTUP, berisi kesimpulan dan saran untuk penelitian selanjutnya.