

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Kehadiran teknologi internet telah mempermudah hidup manusia dalam menggali informasi dan komunikasi. Dampak dari internet tanpa disadari sangat melekat pada manusia, sehingga setiap hari dalam hidup manusia telah bergantung pada teknologi tersebut. Banyak dampak dapat dimanfaatkan oleh beberapa orang yang berupaya untuk mengembangkan suatu media komunikasi maupun untuk menguntungkan diri sendiri dan suatu kelompok. Media sosial adalah salah satu contoh dari kemajuan teknologi. Munculnya media sosial berdampak pada pengguna media sosial yang terus bertambah disetiap harinya. Tak hanya kalangan orang dewasa, anak-anak yang masih dibangku sekolah dasar-pun telah mengenal media sosial. Namun, tanpa disadari dengan adanya media sosial yang tidak mengenal usia penggunanya. Pengguna media sosial yang berasal dari seluruh kalangan usia dapat memberikan berbagai macam permasalahan.

Instagram merupakan media sosial yang pada dasarnya berfungsi untuk berbagi foto maupun video pada sesama pengguna. Sejak September 2017 pengguna aktif instagram telah tercatat kurang lebih 800 juta orang (Yusuf, 2017). Pengguna Instagram tentunya tak hanya orang dewasa namun anak - anak dan remaja ikut menggunakan media sosial ini. Pemanfaatan yang beragam dari pengguna Instagram dibuat sebagai akun pribadi, baik untuk orang biasa maupun artis hingga menjadi sarana bisnis perseorangan. Namun dari manfaat penggunaan Instagram masih banyak pengguna yang belum memahami etika - etika dalam bersosialisasi pada dunia maya. Permasalahan tersebut terjadi dengan adanya tindakan *cyberbullying*. *Cyberbullying* yang merupakan tindakan bullying yang sering terjadi di dunia maya. Banyak pengguna yang masih belum menyadari bahwa ulasan atau komentar yang dilontarkan merupakan tindakan *cyberbullying*. Hal ini tentunya menjadi salah satu peringatan bagi pengguna Instagram agar dapat meminimalisir perbuatan merugikan tersebut. *Cyberbullying* yang telah terjadi di Indonesia dialami oleh seorang remaja asal Indonesia yang berusia 13 tahun di bulan April 2017. Hal tersebut terjadi ketika remaja itu berniat untuk

membagikan tiket secara gratis kepada khalayak umum dengan beberapa syarat. Namun hal ini memberikan dampak negatif untuk dirinya. Berbagai kabar tidak baik tersebar secara cepat di dunia maya dan pada akhirnya muncul komentar negatif yang bersifat menyakiti hati anak. Sebagai dampaknya remaja yang berkeseharian riang, menjadi murung dan selalu menangis (Muttya, 2017). Permasalahan *cyberbullying* pada kolom komentar Instagram menjadi hal yang penting untuk dikaji sebagai pemrosesan teks. Hasil yang diharapkan dari analisis ini adalah untuk mengidentifikasi dan mengungkap pola *cyberbullying* dalam komentar di Instagram, meningkatkan kesadaran pengguna, serta mendorong pengembangan alat deteksi otomatis dan rekomendasi kebijakan yang efektif dalam menangani *cyberbullying*.

Analisis sentimen merupakan proses yang bertujuan untuk menentukan isi dari dataset yang berbentuk teks (dokumen, kalimat, sentimen, dll) bersifat positif dan negative (Sari, dkk. 2022). Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi yang terkandung dalam suatu kalimat opini. Dari berbagai referensi penelitian yang telah dilakukan oleh Veny, dkk (2019), metode Naive Bayes memiliki keunggulan pada akurasi model dibandingkan Decision Tree sama Random Forest dalam analisis sentimen di media sosial Twitter dalam kasus kampanye anti-LGBT di Indonesia dengan nilai akurasi 83,43%. Dari penelitian sebelumnya pada metode klasifikasi support vector machine dalam analisis sentimen *cyberbullying* pada komentar Instagram yang telah dilakukan oleh Wanda, dkk (2018) menunjukkan bahwa yang didapatkan hasil akurasi tertinggi sebesar 90% dengan komposisi data latih 50% dan data uji 50%. Untuk penelitian yang dilakukan oleh Garg, dkk (2024), metode Decision Tree memiliki keunggulan pada akurasi model dibandingkan Random Forest, Naive Bayes, dan K-Nearest Neighbor (K-NN) dalam analisis sentimen dan indikator pasar utama untuk prediksi tren pencatatan IPO dengan nilai akurasi 82%. Sedangkan penelitian yang dilakukan oleh Auliya, dkk (2021) tentang implementasi algoritma K-Nearest Neighbor (K-NN) untuk analisis sentimen publik pada pembelajaran daring menunjukkan bahwa yang didapatkan hasil akurasi tertinggi sebesar 84,93% dari hasil melakukan

pengujian tetangga terdekat dengan  $K=10$ . Hal ini menjadi topik peneliti untuk memberikan solusi terhadap *cyberbullying* yang terjadi di media sosial Instagram. Untuk pengumpulan komentar pengguna aplikasi sosial media Instagram pada penelitian ini dilakukan dengan *web scraping*. Setelah data komentar terkumpul, dilakukan praproses teks untuk membersihkan data teks agar siap untuk dianalisis lebih lanjut. Pemilihan algoritma Naive Bayes, Decision Tree, dan K-Nearest Neighbor (K-NN) dalam menganalisis sentimen komentar *cyberbullying* di Instagram didasarkan pada keunggulan masing-masing algoritma dalam memberikan pendekatan yang komprehensif dan efektif. Naive Bayes unggul dalam menangani data teks besar secara cepat dan efisien serta memberikan akurasi tinggi pada data yang tidak seimbang (Husain dkk., 2020). Decision Tree menawarkan kemudahan interpretasi hasil dan akurasi kompetitif yang mendukung pemahaman sentimen dan pengambilan keputusan (Rashid dkk., 2021). K-Nearest Neighbor (KNN) fleksibel dalam menangani pola data kompleks dan non-linear serta memberikan akurasi tinggi dengan parameter optimal (Ali dkk., 2021). Ketiga algoritma ini secara keseluruhan memberikan pendekatan yang komprehensif dan efektif untuk analisis sentimen.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang diatas, permasalahan yang akan dikaji dalam penelitian ini adalah sebagai berikut :

1. Bagaimana membandingkan algoritma Naive Bayes, Decision Tree, dan K-Nearest Neighbor (K-NN) untuk mendapatkan yang terbaik ?
2. Bagaimana menentukan algoritma yang paling terbaik untuk klasifikasi sentimen pada komentar Instagram?
3. Bagaimana mengidentifikasi komentar yang mengandung elemen *cyberbullying* dalam data komentar Instagram?

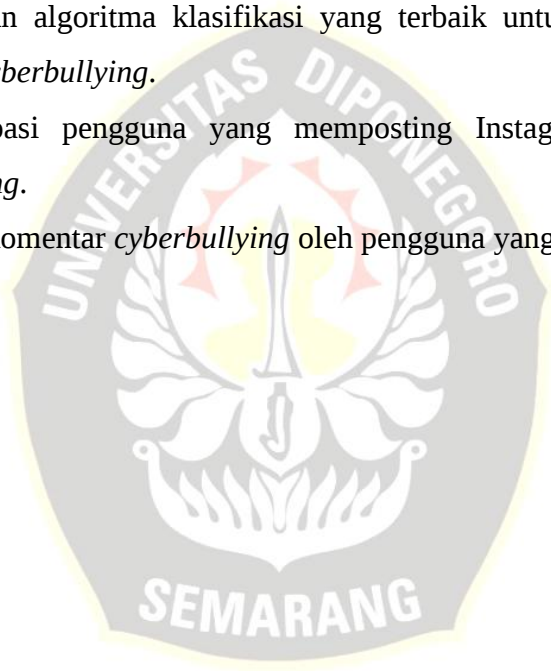
### 1.3 Tujuan

Dari latar belakang masalah yang telah diberikan, penelitian ini dilakukan dengan tujuan untuk mengidentifikasi komentar yang mengandung cyberbullying menggunakan metode sentimen analisis yang terbaik dengan membandingkan algoritma klasifikasi Naive Bayes, Decision Tree, dan K-Nearest Neighbor (K-NN).

### 1.4 Manfaat

Manfaat dari hasil penelitian ini adalah :

1. Mendapatkan algoritma klasifikasi yang terbaik untuk analisis sentimen komentar *cyberbullying*.
2. Mengantisipasi pengguna yang memposting Instagram dari komentar *cyberbullying*.
3. Mencegah komentar *cyberbullying* oleh pengguna yang berkomentar.



SEKOLAH PASCASARJANA