

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Pada era digital ini, masyarakat Indonesia menjadikan media sosial sebagai bagian penting dalam kehidupan sehari-hari untuk berinteraksi secara daring [1]. Berdasarkan data statistik yang dilansir oleh Data Reportal, jumlah pengguna media sosial di Indonesia diperkirakan meningkat hingga 80,23% dari total populasi pada tahun 2025, dengan sebagian besar diantaranya berasal dari kalangan remaja [2]. Menurut riset Reuters Institute, media sosial yang paling sering digunakan oleh kalangan usia tersebut meliputi Twitter atau yang kini dikenal dengan nama X, Facebook, Instagram, TikTok, dan YouTube [3]. Tingginya tingkat partisipasi remaja dalam media sosial memberikan peluang besar bagi mereka untuk memperluas jejaring sosial dan mengeksplorasi kreativitas [4].

Akan tetapi, salah satu permasalahan yang dihadapi oleh kalangan remaja dalam menggunakan media sosial adalah rentannya mereka terhadap tindakan Kekerasan Berbasis Gender *Online* (KBGO) yang kian meningkat setiap tahunnya [5]. Berdasarkan data dari SAFEnet Indonesia, jumlah kasus KBGO meningkat drastis dari 118 kasus pada triwulan pertama 2023 menjadi 480 kasus pada triwulan pertama 2024, di mana 57% di antaranya berasal dari kelompok usia 18–25 tahun, sementara 26% merupakan anak-anak yang berusia di bawah 18 tahun [6]. Salah satu bentuk KBGO yang marak terjadi adalah pelecehan seksual terhadap perempuan dan anak, yang banyak ditemukan di platform media sosial X. Kerawanan platform ini diperkuat oleh studi Tim Peneliti Monash University pada rentang waktu 1 September 2023 hingga Januari 2024. Dari berbagai kategori ujaran kebencian yang dianalisis, studi tersebut secara spesifik mengidentifikasi sebanyak 3.528 unggahan dan komentar yang mengandung muatan seksual atau vulgar. Temuan ini menjadikan X sebagai platform yang rawan digunakan untuk menyebarkan pelecehan seksual berbasis *online* [7]. Dampak dari tindakan ini tidak hanya merugikan korban secara sosial tetapi juga secara psikologis, yang meliputi depresi, kecemasan, hingga kemunculan rasa ingin bunuh diri [8].

Dengan semakin meningkatnya kasus KBGO di platform media sosial X, maka sangat penting untuk dilakukan pencegahan terhadap komentar pelecehan seksual pada remaja. Salah satu upaya pencegahan adalah dengan melakukan

deteksi dini terhadap komentar yang berpotensi mengandung pelecehan seksual di media sosial tersebut [9][10]. Dalam hal ini, berbagai penelitian telah dilakukan, seperti yang dilakukan oleh Hasyim dkk. Penelitian tersebut mengembangkan sistem deteksi ujaran seksisme pada platform media sosial X menggunakan algoritma *machine learning* tradisional. Algoritma yang digunakan adalah *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB) yang diterapkan membangun model klasifikasi, dimana akurasi yang dihasilkan mencapai 80,36% untuk SVM dan 68,62% untuk NB [9]. Namun, penelitian tersebut memiliki cakupan yang masih terlalu luas karena fokusnya hanya pada ujaran seksisme secara umum, sehingga belum secara khusus mengkaji dan mengoptimalkan deteksi pelecehan seksual secara lebih mendalam [11].

Meskipun komentar pelecehan seksual merupakan komentar negatif berbasis gender, karakteristiknya seringkali mirip dengan komentar negatif berbasis non gender seperti ujaran kebencian dan perundungan daring (*cyberbullying*) [12]. Keduanya sama-sama bersifat ofensif yang dapat menimbulkan dampak psikologis buruk bagi remaja [8][13]. Oleh karena itu, pendekatan atau algoritma yang efektif dalam mendeteksi komentar negatif berbasis non-gender dapat dijadikan acuan awal dalam mengembangkan sistem deteksi pelecehan seksual secara lebih spesifik.

Dalam hal pendeteksian jenis komentar negatif berbasis non-gender, berbagai penelitian telah dilakukan [14][15][16]. Salah satunya adalah penelitian yang dilakukan oleh Heri Santoso, dkk., yang menjadi acuan penting bagi studi ini. Dalam penelitian tersebut, komentar *cyberbullying* di Instagram dideteksi menggunakan algoritma *Random Forest* yang berhasil memperoleh kinerja dasar (*baseline*) dengan akurasi sebesar 84%. Hasil tersebut menunjukkan efektivitas dari algoritma *Random Forest* dalam mengidentifikasi pola bahasa negatif berbasis non-gender [14][15].

Berdasarkan paparan tersebut, algoritma *Random Forest* menunjukkan relevansi tinggi untuk dikaji lebih lanjut dalam konteks deteksi komentar pelecehan seksual verbal di media sosial X yang memiliki pola bahasa negatif berbasis gender. Namun, penelitian Heri Santoso dkk., masih terbatas pada penyajian performa *baseline*, yang dicapai hanya dengan menggunakan parameter *default* [14]. Kinerja model seperti ini seringkali belum mencerminkan potensi maksimalnya [17]. Sehingga, untuk benar-benar mengukur dan memperoleh performa terbaik dari algoritma *Random Forest* pada kasus ini, diperlukan pendekatan yang lebih mendalam dengan mengatur kombinasi parameter terbaik melalui proses *hyperparameter tuning* [17].

Oleh karena itu, penting untuk dilakukan studi yang mendalam guna meninjau bagaimana algoritma *Random Forest* dapat diimplementasikan secara optimal, serta mengevaluasi sejauh mana tingkat akurasi dan performanya dalam mengidentifikasi komentar yang mengandung unsur pelecehan seksual di media sosial X.

1.2 Rumusan Masalah

Berdasarkan paparan mengenai urgensi penanganan pelecehan seksual verbal dan celah penelitian yang telah diidentifikasi, maka penelitian ini difokuskan untuk menjawab beberapa rumusan masalah utama sebagai berikut:

1. Bagaimana implementasi algoritma *Random Forest* pada model deteksi komentar pelecehan seksual verbal Berbahasa Indonesia di platform media sosial X?
2. Bagaimana performa algoritma *Random Forest* dalam mendeteksi komentar pelecehan seksual verbal berbahasa Indonesia di platform media sosial X berdasarkan metrik evaluasi yang relevan?

1.3 Batasan Permasalahan

Berdasarkan rumusan masalah yang telah dipaparkan, penelitian ini memiliki beberapa batasan yang bertujuan untuk mempertajam fokus dalam penelitian, yaitu:

1. Sumber dan Jenis Data yang Dideteksi

Penelitian ini terbatas pada deteksi komentar pelecehan seksual verbal yang bersifat eksplisit dalam Bahasa Indonesia dari media sosial X. Fokus deteksi adalah pada komentar yang mengandung kata kunci (*keywords*) spesifik yang berkonotasi merendahkan *gender* perempuan, seperti “jalang” [18], “lonte” [19], “pelacur” [20], “jablay” [21], “tobrut” [19], serta istilah kekerasan seksual seperti “perkosa” [21] dan “digilir” [20]. Dengan demikian, penelitian ini tidak mencakup deteksi pelecehan yang bersifat implisit atau sarkastik.

2. Jumlah dan Rentang Waktu Pengambilan Data

Data untuk penelitian ini diambil melalui teknik *scrapping* dari platform media sosial X dengan rentang waktu 1 September 2023 hingga Januari 2024.

Pemilihan periode ini didasarkan pada temuan relevan dari Tim Peneliti Monash University, yang dalam studinya pada rentang waktu yang sama secara spesifik mengidentifikasi 3.528 unggahan dan komentar bermuatan seksual atau vulgar [7]. Angka ini menunjukkan tingginya volume kasus yang relevan pada periode tersebut dan memperkuat justifikasi pemilihan rentang waktu penelitian.

3. Penyeimbangan dan Jumlah Akhir Dataset

Dataset awal hasil *crawling* memiliki komposisi kelas yang tidak seimbang, terdiri dari 625 data untuk kelas *Sexual Harassment* (mayoritas) dan 375 data untuk kelas *Non Sexual Harassment* (minoritas). Untuk menangani ketidakseimbangan ini, diterapkan teknik *random undersampling* dengan mengurangi sampel pada kelas mayoritas menjadi 382 data [22]. Dengan demikian, batasan penting dalam penelitian ini adalah penggunaan total 757 data (382 kelas *Sexual Harassment* dan 382 kelas *Non Sexual Harassment*) untuk keseluruhan proses.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disampaikan, penelitian ini memiliki tujuan sebagai berikut:

1. Untuk mengimplementasikan algoritma *Random Forest* pada model deteksi komentar pelecehan seksual verbal Berbahasa Indonesia di platform media sosial X.
2. Untuk menganalisis performa algoritma *Random Forest* dalam mendeteksi komentar pelecehan seksual verbal berbahasa Indonesia di platform media sosial X berdasarkan metrik evaluasi yang relevan.

1.5 Manfaat Penelitian

Berdasarkan rumusan masalah dan tujuan penelitian yang akan dilakukan, diharapkan penelitian ini dapat memberikan manfaat sebagai berikut.

1. Manfaat bagi Peneliti

- (a) Mendapatkan pemahaman yang lebih mendalam tentang bagaimana cara penerapan algoritma *Random Forest* dalam mendeteksi komentar

pelecehan seksual verbal berbahasa Indonesia di platform media sosial X.

- (b) Berhasil mengimplementasikan algoritma *Random Forest* dalam mendeteksi komentar pelecehan seksual verbal berbahasa Indonesia di platform media sosial X.
- (c) Mendapatkan wawasan baru terkait dengan hasil performa model *machine learning* yang telah terbukti efektif pada kasus yang berbeda.

2. Manfaat bagi Peneliti Lain

- (a) Memberikan kontribusi baru dengan mengaplikasikan metode yang sudah terbukti efektif pada kasus yang berbeda, khususnya pada deteksi komentar pelecehan seksual di media sosial.
- (b) Memberikan wawasan baru terkait dengan performa model *machine learning* yang telah terbukti efektif pada jenis kasus yang berbeda, sehingga membuka peluang untuk penerapan model yang sama pada kasus lain yang serupa.
- (c) Menjadi acuan bagi penelitian lebih lanjut yang mengeksplorasi potensi penggunaan algoritma *Random Forest* untuk mendeteksi berbagai jenis komentar kebencian dan kekerasan lainnya seperti komentar diskriminasi rasial, komentar kebencian berbasis agama, dan sebagainya.
- (d) Menjadi referensi bagi peneliti lain untuk mengembangkan sistem aplikasi *parental control* yang dapat mengintegrasikan model *machine learning* yang telah dikembangkan dalam penelitian ini.

3. Manfaat bagi Orang Tua dan Remaja

- (a) Memberikan kontribusi dan langkah awal agar orang tua dapat dengan mudah memonitor dan mendeteksi komentar pelecehan seksual pada aktivitas media sosial remaja sehingga mereka terhindar dari dampak buruknya.

1.6 Sistematika Penulisan

Berdasarkan ruang lingkup penelitian yang telah dipaparkan, penulisan laporan ini disusun secara sistematis agar memudahkan pembahasan dan

pemahaman penelitian yang dilakukan. Adapun sistematika penulisan laporan ini adalah sebagai berikut:

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Memaparkan latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
- Bab 2 LANDASAN TEORI
Memaparkan kajian teori yang terkait dengan topik penelitian, diantaranya adalah teori dasar dari media sosial dan remaja, Kekerasan Gender Berbasis *Online*, pelecehan seksual berbasis *online*, *data collection*, *data labelling*, *text preprocessing*, *feature extraction*, *data splitting*, algoritma *Random Forest*, *ensemble learning*, *hyperparameter tuning*, evaluasi model, serta jabaran penelitian terdahulu.
- Bab 3 METODOLOGI PENELITIAN
Memaparkan gambaran umum alur penelitian dan tahapan metode penelitian yang digunakan terdiri dari studi literatur, *dataset collection*, *data labelling*, *text preprocessing*, *feature extraction*, *data splitting*, *Random Forest*, *hyperparameter tuning* dan evaluasi model.
- Bab 4 HASIL DAN DISKUSI
Menyajikan hasil analisis yang mencakup *data collection*, *text preprocessing*, pelatihan model, evaluasi model menggunakan algoritma *Random Forest* tanpa dan menggunakan *hyperparameter tuning*.
- Bab 5 KESIMPULAN DAN SARAN
Memberikan kesimpulan dari hasil penelitian dan merekomendasikan saran untuk penelitian lanjutan