

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi seiring waktu semakin cepat, khususnya *smartphone* yang berbasis android. Adanya *smartphone* yang berbasis android ini berdampak pula semakin banyaknya media sosial yang bermunculan. Salah satu media sosial yang masih sering di gunakan masyarakat ialah *Twitter*. Peringkat media sosial berdasarkan jumlah pengguna di Indonesia *Twitter* berada di posisi ke-lima setelah *Youtube*, *Whatsapp*, *Facebook* dan *Instagram* [1]. Banyak aktivitas yang dapat dilakukan melalui *Twitter*, seperti membuat *tweet*, *me-re-tweet* milik akun orang lain, menyukai *tweet* serta mengirim pesan pribadi antar akun atau disebut *direct message* [2].

Setelah mengetahui banyaknya aktivitas yang dapat dilakukan, tidak jarang banyak pihak yang memanfaatkan *Twitter* untuk melakukan kejahatan yang dapat merugikan pihak lain yang biasa disebut dengan *Cybercrime*. *Cybercrime* adalah suatu kejadian atau kasus yang berhubungan dengan teknologi komputer (dunia maya) dimana seorang korban menderita berupa kerugian dan seorang pelaku dengan sengaja memperoleh keuntungan [3]. Jenis-jenis kejahatan kategori *cybercrime* yaitu *cyber terorisme*, *cyber pornography* dan *cyber harassment* [4]. *Cyber harassment* bisa disebut juga dengan pelecehan online dapat berupa mengancam atau melecehkan melalui email, pesan instan, atau memposting informasi secara online. Dalam kasus ini menargetkan terhadap orang tertentu, baik dengan secara langsung menghubungi mereka atau dengan menyebarkan informasi pribadi mereka, menyebabkan mereka tertekan, takut, atau marah [5].

Pada tahun 2016, menurut lembaga Country Business Head Twitter Indonesia, menyatakan bahwa kicauan oleh pengguna *twitter* Indonesia dari bulan Januari hingga Desember 2016 yakni mencapai 4,1 miliar *tweet* [6]. Sebanyak 4,1 milyar *tweet* tersebut akan dikatakan percuma apabila disalah gunakan ke sesuatu yang negatif, tentu hal ini bertentangan dengan etika berinternet yang diatur oleh pemerintah yang terdapat di Undang-Undang Nomor 11 tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE) [7].

Undang-Undang Nomor 11 tahun 2008 tentang Informasi dan Transaksi Elektronik disahkan pada tanggal 21 April 2008 sebagaimana telah diubah

menjadi Undang-Undang Nomor 19 Tahun 2016 (UU ITE) dan menjadi *cyber law* pertama di Indonesia [8]. Menurut Komnas Perempuan pada tahun 2017 mencatat sebanyak 91 laporan pelecehan online. Selain itu pelecehan seksual dalam transportasi online juga sering terjadi, yakni dengan mencuri data milik konsumennya. Di tahun 2017 kasus siber yang dilaporkan atau tindakan aktif kepolisian mencapai 5.061 kasus yaitu dengan penyelesaian 3.325 kasus untuk kasus kebencian [9]. Hal ini menandakan penanganan kasus *harassment* di Indonesia masih tergolong lambat, sehingga korban merasa ragu untuk melaporkan kepihak berwenang seperti polisi atau orang tua mereka.

Pada penelitian ini menggunakan proses *text mining*, yaitu suatu teknik yang digunakan untuk menangani masalah klasifikasi, pengelompokan, ekstraksi informasi, dan pencarian informasi [10] serta memiliki sifat hampir sama dengan data mining namun hanya fokus pada teks daripada bentuk data yang lebih terstruktur [11]. Melakukan secara khusus melibatkan penggunaan teknologi *Natural Language Processing* (NLP), yang menerapkan prinsip-prinsip linguistik komputasional untuk menguraikan dan menginterpretasikan set data. *Natural Language Processing* berkaitan dengan komputer dapat digunakan untuk memahami dan memanipulasi teks bahasa alami (*natural language*) untuk mendapatkan informasi tertentu [12].

Pada penelitian ini menggunakan TF-IDF (*Term Frequency-Invers Document Frequency*) sebagai seleksi fitur. Metode TF-IDF ini terkenal efisien, mudah, dan memiliki hasil yang akurat [13]. Secara sederhana, metode TF-IDF digunakan untuk mengetahui berapa sering kata yang muncul dalam suatu dokumen. Pada penelitian yang dilakukan oleh Rizky Tri Wahyuni, Dhidik Prastiyanto dan Eko Suprpto yang berjudul Penerapan Algoritma *Cosine Similarity* dan pembobotan TF-IDF pada sistem Klasifikasi Dokumen Skripsi yang menunjukkan hasil penelitian dengan presentase tingkat ketepatan klasifikasi sistem dengan sebesar 98% [14].

Ada beberapa penelitian yang sudah dilakukan seperti oleh Tri Mueri Sandes yang berjudul Klasifikasi *Tweet* Pelecehan Online Pada *Twitter* Dengan Menggunakan Metode *Naïve Bayes* pada tahun 2018 [15]. Pada penelitian tersebut menggunakan metode *naïve bayes* yang memiliki akurasi 85% dan kategori nya yang terdapat lima kelas yaitu *Racist*, *Embarrass*, *Physically threatned*, *Sexual Harassed* dan *Neutral* berbeda dengan penelitian ini hanya terdapat dua kelas. Sedangkan penelitian berikutnya yaitu dilakukan oleh Amer Saeed Ali Al-Katheri & Maheyzah Md Siraj yang berjudul "*Classification of Sexual Harassment on Facebook Using Term Weigthing Schemes*" pada tahun 2018 [16]. Penelitian ini tentang pelecehan seksual yang terjadi pada media sosial *Facebook*. Metode yang digunakan *Support Vector*

Machine dengan tiga pembobotan yaitu *Entropy*, *TFIDF*, dan *Modified TFIDF*. Penelitian ini menggunakan metode SVM untuk mengklasifikasi dua dataset berdasarkan akurasi dan presisi serta hasil percobaan menunjukkan bahwa *Modified TFIDF* lebih tinggi dibandingkan *Entropy* dan *TFIDF*. Hasil akurasi dengan metode SVM dan *Modified TFIDF* yaitu 30 fitur 78%, pada 60 dan 90 fitur 96.50%.

Setelah mempelajari penelitian terdahulu, penelitian ini akan membandingkan kinerja dua metode yaitu *Naïve Bayes Classifier* dan *Support Vector Machine*. Pemilihan metode *Naïve Bayes Classifier* karena merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki kemampuan dan akurasi tinggi [15]. Metode *Support Vector Machine* (SVM) sebagai pembanding dengan *Naïve Bayes Classifier* karena merupakan suatu teknik untuk melakukan prediksi, baik prediksi dalam kasus regresi maupun klasifikasi [17].

Metode *Naïve Bayes Classifier* dan *Support Vector Machine* termasuk klasifikasi *supervised learning* [18]. Berdasarkan penelitian terdahulu algoritma *Naïve Bayes Classifier* dan *Support Vector Machine* merupakan algoritma klasifikasi yang banyak digunakan oleh para peneliti dibidang *text mining* [17]. *Naïve Bayes Classifier* menggunakan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi dengan melihat frekuensi pada data training [19]. Sedangkan *Support Vector Machine* ialah algoritma yang bekerja menggunakan pemetaan nonlinear untuk mengubah data pelatihan asli menjadi dimensi yang lebih tinggi yaitu dengan menggunakan *hyperplane* sebagai pemisah dua kelas sebuah data [17].

Dapat dilihat dari penjelasan di atas, penelitian ini akan dituangkan ke dalam Laporan Tugas Akhir yang berjudul “Perbandingan Metode *Naïve Bayes Classifier* dan *Support Vector Machine* Untuk Klasifikasi *Cyber Harassment* Pada *Twitter*”. Alasan pengambilan judul ini berharap dari penelitian yang dilakukan terhadap data yang ada di *Twitter* untuk mengklasifikasi jenis kasus *cyber harassment* secara cepat dan akurat dengan perbandingan dua metode tersebut serta kelebihanannya belum ada yang melakukan penelitian ini.

1.2. Perumusan Masalah

Metode *Naïve Bayes Classifier* dan *Support Vector Machine* termasuk klasifikasi *supervised learning* namun memiliki cara kerja yang berbeda maka metode apakah yang terbaik dari dua metode tersebut untuk klasifikasi *cyber harassment*.

1.3. Pertanyaan Penelitian

Berdasarkan uraian di atas, maka penulis dapat merumuskan beberapa pertanyaan pada penelitian ini yaitu berapa nilai akurasi masing-masing metode yaitu *Naïve Bayes Classifier* dan *Support Vector Machine* untuk klasifikasi *cyber harassment*?

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang ada, maka dapat diketahui tujuan dari penelitian ini yaitu:

1. Mengukur nilai akurasi masing-masing metode yaitu *Naïve Bayes Classifier* dan *Support Vector Machine* untuk klasifikasi *cyber harassment*.
2. Menjelaskan perbandingan kinerja metode *Naïve Bayes Classifier* dan *Support Vector Machine* (SVM) untuk klasifikasi *cyber harassment*.

1.5. Batasan Masalah

Berdasarkan rumusan masalah dan tujuan penelitian, maka diperoleh batasan-batasan masalah penelitian sebagai berikut:

1. Hanya untuk klasifikasi data *tweet* atau teks tidak untuk foto atau gambar.
2. *Tweet* berisi kata kunci dari berbagai jenis-jenis *cyber harassment* yaitu *Sexual Harassment*, *Racial Harassment*, *Appearance-related Harassment*, *Intellectual Harassment*, dan *Political Harassment* yang memiliki potensi terjadinya *harassment*.
3. Kategori terdiri dari dua kelas yaitu mengandung *cyber harassment* dan bukan *cyber harassment*.
4. *Tweet* menggunakan Bahasa Indonesia.
5. *Tweet* dengan konten promosi produk diabaikan.
6. *Emoticons* atau simbol diabaikan pada saat proses analisis.
7. Kolom *tweet* tidak diaktifkan fitur "Lindungi Tweet Anda" atau *hidden* oleh pengguna *Twitter*.

1.6. Manfaat Penelitian

Berdasarkan penelitian ini, manfaat yang dapat diambil yaitu:

1. Dapat menjadi pencegahan dini agar korban dapat terhindar dari *cyber harassment*.
2. Menambah ilmu pengetahuan untuk bidang IT tentang metode *Naïve Bayes Classifier* dan *Support Vector Machine*.