

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi informasi sangat bergantung pada penggunaan internet, karena dengan adanya internet dapat membantu kegiatan komunikasi menjadi cepat dan efisien, serta mencari informasi yang diperlukan juga terasa sangat mudah. Karakteristik Internet, dapat diakses kapan saja, di mana saja dan oleh siapa saja, memungkinkan pengguna untuk mencari dan mengakses informasi sesuai dengan kebutuhannya.

Media sosial adalah media online di mana pengguna dapat dengan mudah berpartisipasi, berbagi, dan membuat konten seperti blog, jejaring sosial, wiki, forum, dan dunia virtual. Blog, jejaring sosial, dan wiki adalah bentuk media sosial yang paling umum digunakan oleh orang-orang di seluruh dunia [1]. Dengan adanya media sosial sangat memudahkan pengguna untuk mendapatkan informasi dan membangun jejaring sosial. Melalui media sosial, jejaring sosial dengan model individu dan kelompok telah menjadi barang sosial yang tak terelakkan. Kemudahan komunikasi juga telah menjadi realitas sosial. Orang dapat menerima dan menggunakan semua informasi yang dibutuhkan dalam konteks hubungan sosial tersebut [2]. Salah satu media sosial tempat orang bebas mengutarakan pendapatnya adalah Twitter. Dalam beberapa tahun terakhir, Twitter telah memberikan pengaruh banyak pengaruh dalam menciptakan sumber informasi [3]. Di media sosial Twitter pengguna dapat membuat sebuah tweet dengan maksimal karakter per tweet baru-baru ini ditingkatkan dari 140 menjadi 280, sehingga mendorong lebih banyak fleksibilitas dalam interaksi [4].

Twitter adalah salah satu platform media sosial terbesar dan paling populer di dunia, di mana pengguna dapat membagikan pemikiran, ide, dan pandangan mereka secara publik. Namun, hal ini seringkali berdampak negatif, karena munculnya kalimat ofensif atau kasar di Twitter [5]. Berdasarkan data

yang terdapat dalam situs cyberbullying.org, sebanyak 15% murid sekolah di Amerika Serikat terkena kalimat ofensif di internet yang awalnya hanya 9,3% selama 30 hari [6].

Masalah kalimat ofensif di Twitter sebagian besar disebabkan oleh karakteristik platform media sosial tersebut. Twitter memungkinkan pengguna untuk mengungkapkan opini mereka secara anonim, tanpa ada batasan pada bahasa atau isi pesan yang mereka bagikan. Hal ini menyebabkan adanya penyalahgunaan kebebasan berekspresi, dan seringkali muncul kalimat ofensif, penghinaan, dan pelecehan [4].

Masalah kalimat ofensif di Twitter perlu mendapat perhatian serius, dan diperlukan solusi yang efektif untuk mengatasi hal tersebut. Salah satu solusinya adalah dengan menggunakan teknologi NLP, seperti klasifikasi kalimat ofensif, untuk mendeteksi dan memfilter konten yang bersifat ofensif atau kasar di Twitter.

Metode *Machine Learning* atau pembelajaran mesin yang sering digunakan dalam penerapan analisis sentimen, seperti *Decision Tree*, *Naive Bayes Classifier*, *Random Forest*, *Support Vector Machine (SVM)* dan lainnya [7]. Dari beberapa metode *machine learning* yang disebutkan, SVM merupakan metode terbaik yang memiliki keunggulan berupa kemampuannya dalam mengolah data berdimensi tinggi [8], tetapi SVM masih memiliki kelemahan saat menghitung data dalam jumlah besar [9]. Sementara masalah umum yang sering ditemui dalam teknik analisis sentimen adalah banyaknya atribut yang digunakan, sehingga volume data yang digunakan juga besar, hal ini dapat menurunkan efisiensi *classifier* dan menyebabkan akurasi yang kurang optimal [10]. Untuk mengatasi hal tersebut, diperlukan sebuah algoritma *feature selection* yang dapat meningkatkan kinerja SVM dalam mengklasifikasikan analisis sentimen.

Particle Swarm Optimization (PSO) adalah teknik optimisasi yang digunakan untuk mencari solusi optimal dalam masalah yang kompleks dengan mencari partikel terbaik dalam ruang pencarian. Ketika SVM digabungkan dengan PSO, maka akan menghasilkan SVM berbasis *Particle Swarm*

Optimization (PSO-SVM). Kombinasi ini membantu meningkatkan kinerja SVM dengan memperbaiki hyperparameter dan parameter yang digunakan dalam SVM. Dengan menggunakan PSO untuk mengoptimalkan parameter SVM, kita dapat memperoleh model yang lebih baik dan efisien [11].

Berdasarkan permasalahan diatas maka penulis melakukan penelitian tugas akhir menggunakan metode *Support Vector Machine* (SVM) berbasis *Particle Swarm Optimization* (PSO) untuk mengklasifikasi kalimat ofensif pada media sosial *Twitter*. Adapun tujuan dari penelitian ini yaitu untuk meningkatkan hasil akurasi hasil klasifikasi model SVM terhadap sentimen kalimat ofensif di media sosial twitter dengan menggunakan *feature selection* PSO.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sebelumnya sudah dipaparkan, maka dapat dirumuskan menjadi beberapa rumusan masalah sebagai berikut :

1. Membandingkan performa kinerja model klasifikasi teks tanpa seleksi fitur dan dengan seleksi fitur?
2. Berapa akurasi yang diperoleh dari algoritma *Support Vector Machine* (SVM) berbasis *optimasi Particle Swarm Optimization* (PSO) terhadap *Tweet* yang mengandung kalimat ofensif?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini yaitu untuk :

1. Mengetahui perbandingan performa dari kinerja model klasifikasi teks tanpa seleksi fitur dan dengan seleksi fitur.
2. Mengetahui nilai akurasi yang diperoleh dari algoritma *Support Vector Machine* (SVM) berbasis *optimasi Particle Swarm Optimization* (PSO) terhadap *Tweet* yang bersifat ofensif di *Twitter*.

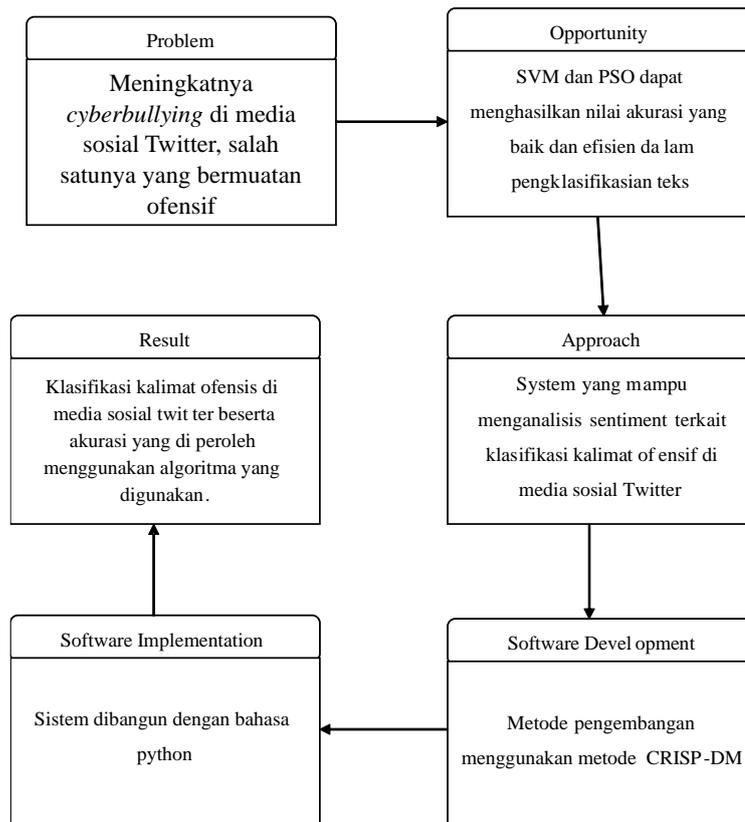
1.4 Batasan Masalah

Agar penelitian ini tidak keluar dari pokok permasalahan yang dirumuskan, maka ruang lingkup penelitian dibatasi pada :

1. Data yang digunakan diambil dari media sosial Twitter dengan *keyword* “bodoh”, “iblis”, “bobrok”, “hancur”, “lemah”, “jelek”[12], “politik”. Kata-kata tersebut merupakan contoh kata yang bersifat ofensif.
2. Cakupan penelitian hanya sebatas menghitung akurasi algoritma.
3. Data komentar yang dianalisis menggunakan bahasa Indonesia, termasuk bahasa Indonesia yang tidak baku.
4. *Emoticon* merupakan karakter yang dihapus dari dataset.
5. Data komentar diambil pada tanggal 01 Maret 2023.
6. Hasil klasifikasi terbagi menjadi 2 kelas, yaitu ofensif dan tidak ofensif.

1.5 Kerangka Pemikiran

Kerangka pemikiran ini diawali dengan permasalahan yaitu terkait meningkatnya tweet yang bersifat ofensif atau menyerang individu dan golongan di media sosial Twitter. Dengan adanya teknologi deep learning, untuk mengklasifikasikan teks bisa jauh lebih baik dengan hasil akurasi yang tinggi. Implementasi dari sistem ini adalah menggunakan bahasa pemrograman Python dan menggunakan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) untuk metode pengembangannya. Hasil dari penelitian ini adalah analisis sentimen terkait klasifikasi kalimat ofensif di media sosial *Twitter*. Pengembangan Kerangka Pemikiran bisa dilihat pada gambar berikut :



Gambar 1. 1 Kerangka pemikiran

1.6 Sistematika Penulisan

Sistematika dalam penulisan tugas akhir ini dibagi menjadi 5 bab dengan uraian sebagai berikut:

BAB I PENDAHULUAN

Bab pendahuluan memuat Latar Belakang Penelitian, Perumusan Masalah Penelitian, Tujuan Penelitian, Batasan Masalah Penelitian, Metodologi Penelitian, Kerangka Pemikiran Penelitian, dan Sistematika Penulisan.

BAB II KAJIAN LITERATUR

Bab kajian literatur berisi landasan teori-teori yang relevan dengan sebagai sebuah dasar dan rujukan terhadap penelitian yang dilakukan.

BAB III METODOLOGI PENELITIAN

Bab ini berisi mengenai metodologi penelitian berupa pemahaman bisnis, pemahaman data, persiapan data, pemodelan, dan evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Bab hasil dan pembahasan memuat tentang penjelasan dari temuan dan hasil dari penelitian yang telah dilakukan.

BAB V SIMPULAN DAN SARAN

Bab penutup berisi mengenai kesimpulan terhadap penelitian yang telah dilakukan serta membahas mengenai saran-saran untuk penelitian yang lebih baik.