

BAB I

PENDAHULUAN

Isu pengungsi Rohingya telah menjadi salah satu krisis kemanusiaan yang mendapat perhatian dunia internasional, khususnya di kawasan Asia Tenggara. Pergolakan politik dan kekerasan terhadap etnis Rohingya di Myanmar memaksa mereka mencari perlindungan di negara-negara tetangga termasuk Indonesia. Di tengah dinamika sosial yang terjadi, media sosial berperan besar sebagai platform untuk menyuarakan opini publik. Analisis terhadap sentimen yang terkandung dalam opini masyarakat menjadi penting untuk memahami persepsi dan pandangan mereka terkait isu ini. Penelitian ini bertujuan untuk melakukan perbandingan antara dua metode pelabelan sentimen yaitu InSet Lexicon dan VADER Lexicon serta klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) dan optimasi menggunakan *Particle Swarm Optimization* (PSO).

1.1. Latar Belakang

Krisis kemanusiaan yang melibatkan pengungsi Rohingya di Indonesia telah menarik perhatian dunia internasional pada tahun 2023. Situasi pengungsi ini diperparah oleh kekacauan politik dan ketegangan antara kelompok etnis Rohingya dan pemerintah Myanmar. Masyarakat di Aceh menolak kedatangan etnis Rohingya yang mencari suaka di Indonesia karena perang etnis di Myanmar. Etnis Rohingya, yang merupakan minoritas di Myanmar tidak dilindungi secara hukum dan kebijakan politik dan bahkan diusir oleh junta militer Myanmar. Meskipun menghadapi kesulitan, etnis Rohingya mencari suaka di negara-negara Asia Tenggara seperti Malaysia, Bangladesh, India, dan Indonesia. Di Aceh, diperkirakan sudah ada 2000 orang etnis Rohingya yang tiba. Penolakan masyarakat Aceh terhadap kedatangan etnis Rohingya disebabkan oleh pengalaman sebelumnya di mana mereka menerima jumlah kecil etnis Rohingya. Namun, etnis Rohingya kemudian melarikan diri dari tempat pengungsian, tidak menghormati norma agama dan norma sosial masyarakat setempat, serta melakukan tindakan kriminal yang mengganggu keamanan dan kondusivitas masyarakat Aceh [1]. Pada era digital saat ini, Media sosial merupakan platform

utama bagi warga Indonesia untuk berbagi pendapat, mengekspresikan pandangannya, dan melakukan diskusi secara publik terkait dengan pengungsi rohingya di Indonesia atau lebih tepatnya di Aceh pada tahun 2023. Media sosial merupakan media yang dapat digunakan pengguna dalam rangka mempresentasikan dirinya, berinteraksi, bekerja sama, berbagi informasi, maupun berinteraksi dengan pengguna lainnya [2].

X atau lebih dikenal dulunya sebagai Twitter merupakan sebuah platform media sosial yang memiliki dampak besar di Indonesia, menempati peringkat kelima dengan 30,2% pengguna berdasarkan survei Katadata Insight Center (KIC) pada tahun 2023 [3]. Berdasarkan informasi tersebut, X telah menjadi platform media sosial yang memiliki pengaruh yang cukup besar di Indonesia. Melalui X, pengguna dapat mengunggah konten sesuai keinginan. Konten tersebut berupa opini, sentimen, maupun *emoticon*, yang bisa menjadi data untuk menganalisis suatu trend atau topik tertentu [4]. Sebagai hasilnya, banyak pengguna X di Indonesia menggunakan platform tersebut untuk membagikan *tweet* yang mencerminkan pendapat atau perasaan mereka terhadap berbagai peristiwa, termasuk pengungsi rohingya di Indonesia tahun 2023. *Tweet* tersebut dapat memiliki sentimen positif maupun negatif. Penelitian ini bertujuan untuk melakukan analisis sentimen dalam *tweet* yang berkaitan dengan pengungsi rohingya di Indonesia pada tahun 2023 yang menggunakan berbagai metode analisis.

Analisis sentimen adalah sebuah studi yang menganalisis pendapat, sentimen, penilaian, sikap, evaluasi, dan emosi seseorang terhadap entitas seperti sebuah produk, layanan, organisasi, individual, masalah, topik atau peristiwa, dan sebagainya [5]. Analisis sentimen dapat memanfaatkan berbagai algoritma untuk melakukan klasifikasi teks seperti *Naïve Bayes*, *Support Vector Machine*, *Deep Neural Network*, dan algoritma *machine learning* lainnya. Algoritma pembelajaran Support Vector Machine (SVM) adalah model *Supervised Learning* yang digunakan untuk analisis data yang bertujuan untuk melakukan analisis regresi dan klasifikasi. Dalam prakteknya, algoritma SVM sering digunakan untuk berbagai permasalahan terkait dengan klasifikasi [6]. Algoritma SVM masih

memiliki kekurangan dalam menentukan parameter dan fitur yang sesuai untuk memaksimalkan performa yang dihasilkan. Dalam mengatasi kelemahan algoritma SVM, perlu dilakukan metode optimasi untuk memilih parameter yang tepat. Hal ini diharapkan dapat memperbaiki performa algoritma dan berpotensi meningkatkan akurasi hasil [7]. Penelitian ini mengambil data yang berupa sentimen dari *tweet* pengguna media X yang membahas mengenai pengungsi rohingya di Indonesia pada tahun 2023. Setelah itu, proses analisis dilakukan untuk mengelompokkan data menjadi 2 kategori sentimen yaitu positif dan negatif.

Salah satu pendekatan yang umum digunakan dalam analisis sentimen adalah metode *lexicon-based*, sebuah teks dianalisis berdasarkan kamus sentimen yang mengelompokkan kata-kata ke dalam kategori sentimen positif atau negatif. Kamus sentimen yang dapat digunakan pada *lexicon-based* antara lain yaitu InSet Lexicon, SentiWordNet, SentiStrength, VADER Lexicon, dll. Kamus InSet (Indonesia *Sentiment*) Lexicon berisi daftar kata yang mengandung sentimen positif maupun negatif serta sudah memiliki bobot nilai untuk kata nya. Kamus lexicon ini terdiri dari 3609 kata yang dianggap positif dan 6609 kata yang dianggap negatif. Setiap kata dalam kamus ini memiliki bobot dengan nilai yang berkisar dari -5 hingga +5 [8]. VADER Lexicon adalah kamus sentimen yang dirancang khusus untuk teks media sosial, dengan fokus pada kata-kata dan frasa yang sering muncul di platform media sosial seperti Twitter. VADER merupakan singkatan dari *Valence Aware Dictionary for Sentiment Reasoning*. VADER menggunakan kamus lexicon berbahasa inggris (en) [9].

Berdasarkan penelitian yang dilakukan [10] VADER Lexicon menghasilkan performa yang lebih baik jika dibandingkan dengan Textblob dalam melakukan klasifikasi sebuah *tweet* dalam bahasa inggris. Selain itu, pada penelitian yang dilakukan [11] memiliki hasil bahwa VADER menghasilkan akurasi yang lebih baik yaitu 77% jika dibandingkan dengan Textblob dan NLTK yang hanya mendapatkan persentase 74% dan 62%. Beberapa penelitian telah melakukan penelitian terkait analisis sentimen dengan menerapkan algoritma optimasi seleksi fitur. Penelitian [12] melakukan tinjauan pustaka terhadap 25 jurnal yang

memfokuskan pada penggunaan beberapa algoritma optimasi seleksi fitur seperti *Information Gain*, *Particle Swarm Optimization*, *Chi Square*, dan *N-Gram* untuk analisis sentimen menggunakan algoritma *Naive Bayes*. Hasilnya menunjukkan bahwa metode yang paling optimal adalah *Particle Swarm Optimalization* dengan rata-rata akurasi mencapai 89.08%. Pada penelitian yang dilakukan [13] memiliki hasil bahwa algoritma SVM tanpa optimasi menghasilkan akurasi sebesar 78.81%, sedangkan dengan optimasi akurasi meningkat menjadi 88.19%. Ini menunjukkan bahwa menggunakan optimasi *Particle Swarm Optimization* memberikan peningkatan yang signifikan pada hasil akurasi.

Berdasarkan semua uraian di atas dari penelitian sebelumnya, maka penelitian yang akan dilakukan yaitu tentang perbandingan antara 2 metode dalam pelabelan sentimen yaitu InSet Lexicon dan VADER Lexicon terkait sentimen masyarakat terhadap pengungsi rohingya di Indonesia tahun 2023 pada media sosial X (Twitter) serta menerapkan algoritma *Support Vector Machine* (SVM) dan optimasi menggunakan *Particle Swarm Optimization* (PSO) untuk menemukan parameter terbaik yang akan digunakan pada klasifikasi data sehingga dapat meningkatkan akurasi yang dihasilkan.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya maka rumusan masalah pada penelitian ini, yaitu :

1. Bagaimana perbandingan nilai akurasi yang dihasilkan dari evaluasi model antara data dengan pelabelan InSet Lexicon dan VADER Lexicon serta menerapkan algoritma SVM untuk klasifikasi data dan Optimasi menggunakan metode PSO?
2. Bagaimana perbandingan sentimen yang dihasilkan dari data dengan proses pelabelan menggunakan InSet Lexicon dan VADER Lexicon terhadap data *tweet* pengungsi rohingya di Indonesia?

1.3. Tujuan Penelitian

Berdasarkan hasil pemaparan dari rumusan masalah, maka tujuan dari penelitian ini, yaitu :

1. Mengetahui perbandingan nilai akurasi yang dihasilkan dari evaluasi model antara data dengan pelabelan InSet Lexicon dan VADER Lexicon serta menggunakan algoritma SVM untuk klasifikasi data dan Optimasi menggunakan metode PSO.
2. Mengetahui perbandingan sentimen yang dihasilkan dari data dengan proses pelabelan menggunakan InSet Lexicon dan VADER Lexicon terhadap data *tweet* pengungsi rohingya di Indonesia.

1.4. Manfaat Penelitian

Manfaat yang terdapat pada penelitian ini yaitu berkontribusi dalam pengembangan ilmu pengetahuan di bidang analisis sentimen, *machine learning*, dan optimasi algoritma. Perbandingan antara InSet Lexicon dan VADER Lexicon memberikan wawasan baru terkait efektivitas kamus sentimen dalam analisis sentimen. Hasil penelitian ini dapat menjadi acuan bagi penelitian selanjutnya. Penelitian ini juga membantu mengidentifikasi sentimen publik terhadap isu kemanusiaan seperti pengungsi Rohingya. Selain itu, manfaat teknologis juga terlihat dalam penelitian ini yaitu kombinasi metode SVM dan optimasi PSO dapat diaplikasikan lebih luas untuk menganalisis tren sosial atau isu publik lainnya. Dengan manfaat-manfaat tersebut, penelitian ini diharapkan mampu memberikan kontribusi signifikan baik dalam dunia akademik maupun untuk kepentingan praktis dalam menangani isu-isu sosial.

1.5. Batasan Masalah

Pada penelitian ini diperlukan batasan agar tidak terjadi pelebaran masalah. Adapun batasan masalah dalam penelitian ini sebagai berikut:

1. Penelitian ini membatasi pengumpulan data pada data *tweet* dengan bahasa Indonesia yang berkaitan tentang pengungsi rohingya di Indonesia yang diperoleh dari sosial media X (Twitter). Data *tweet* yang akan digunakan akan dibatasi pada periode tertentu yaitu data *tweet* yang telah diposting oleh pengguna pada tahun 2023.
2. Penelitian ini berfokus pada perbandingan antara dua metode pelabelan data pada analisis sentimen yaitu InSet Lexicon dan VADER Lexicon.

3. Penelitian ini menerapkan algoritma SVM sebagai metode klasifikasi dan optimasi parameter menggunakan PSO.
4. Penelitian ini menggunakan 2 kategori sentimen utama yaitu positif dan negatif.