

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Permasalahan dimensi data yang tinggi berpengaruh terhadap peran atribut pada sistem akurasi analisis sentimen dengan hasil yang kurang memuaskan. Analisis sentimen merupakan proses mengumpulkan persepsi atau komentar masyarakat terkait topik tertentu pada media sosial [1]. Analisis sentimen bertujuan untuk mengetahui kecenderungan komentar masyarakat positif atau negatif terhadap suatu wacana untuk selanjutnya dilakukan pemodelan [2]. Pengurangan atribut merupakan salah satu cara efektif yang digunakan untuk meningkatkan kinerja dengan mengurangi beberapa atribut atau fitur dalam aplikasi *machine learning*. *feature selection* dilakukan sebelum klasifikasi.

Feature selection yang sering digunakan pada penelitian sebelumnya yakni *particle swarm optimization* (PSO) dan *information gain* dengan tujuan untuk mengoptimalkan atau meningkatkan nilai akurasi suatu algoritma seperti penelitian yang dilakukan oleh Ika dan Hilman pada tahun 2018 dengan judul *Hybrid Method of Information Gain and Particle Swarm Optimization for Selection of Features of SVM-Based Sentiment Analysis*. Penelitian tersebut bertujuan untuk memperoleh nilai akurasi terbaik dari penerapan algoritma *support vector machine* (SVM) dengan dua *feature selection* dalam satu sistem. *Feature selection* yang digunakan yakni PSO dan *Information gain*. Pengujian dilakukan sebanyak 4 kali percobaan yakni percobaan pertama akurasi yang diperoleh dengan menggunakan algoritma

SVM saja, percobaan kedua menggunakan algoritma SVM dan satu metode *feature selection* yakni PSO, percobaan ketiga menggunakan algoritma SVM dan satu metode *feature selection* yakni *information gain*, selanjutnya percobaan keempat menggunakan algoritma SVM dan dua *feature selection* yakni PSO dan *information gain*. Hasil penelitian menunjukkan bahwa penerapan algoritma SVM dengan menggunakan dua *feature selection* yakni PSO dan *information gain* memperoleh nilai akurasi terbaik yakni 94.80 daripada penerapan algoritma SVM dengan PSO dan *information gain* yang diterapkan masing-masing [3].

Berdasarkan tinjauan literatur yang dilakukan pada penelitian sebelumnya, dalam penelitian ini menggunakan salah satu metode *supervised learning*. Metode atau tipe *Supervised learning* mencakup beberapa metode diantaranya Regresi Logistic, K-Nearest Neighbor (K-NN), *Support Vector Machine*, Naïve Bayes, Decision Tree dan Random Forest. Pada penelitian ini akan menggunakan metode *Support Vector Machine* (SVM) yang digunakan untuk mengklasifikasikan kelas sentimen positif dan negatif pada media sosial twitter dengan topik yang diambil terkait wacana pemekaran Papua. *Support Vector Machine* (SVM) memiliki kemampuan mengidentifikasi hyperlane yang berbeda, sehingga meningkatkan margin dua kelas yang berbeda, namun *Support Vector Machine* (SVM) memiliki kelemahan yakni sulitnya pemilihan fitur yang sesuai [4].

Kesulitan pemilihan fitur yang tepat pada SVM bisa mengakibatkan nilai akurasi yang diperoleh tidak stabil, untuk mengatasi hal tersebut dapat menggunakan *feature selection* pada pemilihan fitur diantaranya ada *Genetic Algoritma* (GA), *Ant Colony Optimization* (ACO), *Artificial Bae Colony*

Optimization (ABCO), Cat Swarm Optimization (CSO), Cockroach Swarm Optimization Algorithm (CSOA), Glowworm Swarm Optimization (GSO), Particle Swarm Optimization (PSO). Dari berbagai Teknik optimasi tersebut pada penelitian ini akan diterapkan *Particle Swarm Optimization (PSO)* dan Genetic Algorithm (GA) untuk menentukan parameter optimal untuk mendapatkan nilai akurasi terbaik pada *Support Vector Machine (SVM)*.

Kelebihan yang dimiliki PSO yakni kecepatan dalam menyelesaikan suatu permasalahan optimasi lebih cepat serta dapat digunakan untuk memecahkan masalah optimasi pada berbagai aplikasi di berbagai bidang [5]. Sedangkan kelebihan GA yakni mampu melakukan optimasi masalah dengan masalah yang kompleks dan ruang pencarian yang sangat luas [6].

Penerapan *Particle Swarm Optimization (PSO)* dan *Genetic Algorithm (GA)* pada algoritma *Support Vector Machine (SVM)* berdasarkan peluang penelitian pada literatur yang ada dapat diterapkan pada analisis sentimen pemekaran Papua untuk mendapatkan hasil akurasi terbaik.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan pada latar belakang maka fokus rumusan masalah pada penelitian ini sebagai berikut :

1. Bagaimana cara meningkatkan performansi dari penerapan algoritma *Support Vector Machine* pada klasifikasi analisis sentimen menggunakan *feature selection Particle Swarm Optimization (PSO)* dan *Genetic Algorithm (GA)*?

2. Bagaimana hasil akurasi terbaik berdasarkan perbandingan dari penerapan algoritma *Support Vector Machine* (SVM) menggunakan *feature selection Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA) ?

1.3 Tujuan Penelitian

Tujuan dilakukannya penelitian ini yaitu untuk :

1. Meningkatkan performansi penerapan algoritma *Support Vector Machine* pada klasifikasi analisis sentimen menggunakan *feature selection Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA).
2. Mendapatkan hasil akurasi terbaik berdasarkan perbandingan dari penerapan algoritma *Support Vector Machine* (SVM) menggunakan *feature selection Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA).

1.4 Manfaat Penelitian

1. Mampu meningkatkan performansi dari algoritma *Support Vector Machine* (SVM) dengan penerapan *feature selection Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA).
2. Memberikan informasi dari analisis sentimen yang dilakukan pada *tweet* pemekaran Papua dari media sosial Twitter untuk melihat kecenderungan publik.
3. Penelitian ini bisa dijadikan referensi untuk penelitian selanjutnya khususnya dengan tema terkait.

1.5 Batasan Masalah

Berdasarkan permasalahan yang didapatkan, supaya penelitian tidak keluar dari topik maka ditentukan batasan masalah. Adapun batasan masalah pada penelitian ini diantaranya :

1. *Tweet* yang digunakan untuk analisis adalah *tweet* berbahasa Indonesia.
2. Metode yang digunakan dalam penelitian ini adalah CRISP-DM.
3. Algoritma yang diterapkan adalah *Support Vector Machine* (SVM).
4. Menerapkan *Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA).
5. Analisis sentimen dilakukan pada *tweet* dari media sosial Twitter dengan topik pemekaran Papua.
6. Data yang digunakan merupakan data *tweet* pada aplikasi Twitter terkait pemekaran Papua dengan rentang waktu dari 1 Juli 2022 sampai dengan 25 Juli 2022.
7. Media yang digunakan untuk pengumpulan data yaitu Twitter API.
8. *Tweet* akan diklasifikasikan menjadi 2 kelas yakni sentimen positif dan negatif.

1.6 Metodologi Penelitian

Metodologi penelitian meliputi pembahasan terkait waktu dan tempat penelitian, tahapan penelitian, pendekatan penelitian, jenis penelitian, variabel penelitian, serta objek penelitian. Tahapan metode penelitian dalam penelitian ini adalah sebagai berikut :

1. Literatur *review*

Pada tahap literatur *review* dilakukan pencarian terkait penelitian sebelumnya mengenai analisis sentimen dan algoritma yang digunakan yang mampu mendukung penelitian yang akan dilakukan secara teoritis. Sumber literatur yang digunakan diantaranya dari buku, skripsi terdahulu, jurnal, dan sumber Pustaka dari internet.

2. Pengumpulan dan Analisis Data

Data yang diperlukan dalam penelitian merupakan data tweet pada sosial media Twitter terkait dengan Pemekaran Papua untuk dijadikan sampel dalam membandingkan kinerja algoritma *Support Vector Machine* (SVM) dengan menerapkan *feature selection Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA) untuk memperoleh hasil akurasi terbaik.

3. Pengolahan Data

Pengolahan data dilakukan untuk memperoleh data yang valid dan berkualitas. Tahapan pengolahan data yang digunakan yakni metode CRISP-DM yang terdiri dari: *Business Understanding, Data Understanding, Data Preparation.*

4. *Modelling*

Pada tahap ini akan dilakukan pemodelan menggunakan algoritma *Support Vector Machin* (SVM) dengan menerapkan *feature selection Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA) untuk membandingkan hasil akurasi sehingga diperoleh hasil akurasi terbaik.

5. *Evaluation and Deployment*

Proses evaluasi dilakukan dengan tujuan untuk menguji performa model dari algoritma *Support Vector Machine* (SVM) menggunakan *Confusion Matrix* untuk mendapatkan nilai *accuracy*, *precision*, *recall*, dan *f-score* dengan hasil data dalam bentuk kurva ROC untuk mengukur nilai AUC. Pada tahap *deployment* merupakan tahap untuk penyajian hasil..

6. Penarikan Kesimpulan

Penarikan kesimpulan merupakan tahapan terakhir pada penelitian berupa analisa perbandingan algoritma dengan *feature selection* yang berbeda dengan mengimplementasikan model *machine learning*.

1.7 Sistematika Penulisan

Adapun aturan dan sistematika penulisan yang digunakan dalam penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab pendahuluan berisi latar belakang umum seperti gambaran secara garis besar tentang isi laporan, rumusan masalah, batasan permasalahan pada penelitian, tujuan penelitian, manfaat penelitian yang diperoleh, metode penelitian serta sistematika penulisan laporan tugas akhir.

BAB II LANDASAN TEORI

Bab landasan teori meliputi teori - teori yang saling berhubungan dengan penelitian seperti konsep serta metode dan algoritma yang terkait dengan penelitian

ini. Pada bab ini juga berisi penjelasan dari penelitian sebelumnya yang relevan dengan penelitian ini.

BAB III METODOLOGI

Bab metodologi berisi uraian metode yang digunakan dalam melakukan penelitian, mulai dari waktu dan tempat penelitian, objek penelitian, variabel penelitian, matriks penelitian serta tahapan-tahapan yang dilakukan pada penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab hasil dan pembahasan berisi pemaparan hasil serta pembahasan terhadap perancangan pada bab sebelumnya, yaitu bagaimana alur sebuah data apabila diolah dengan berdasarkan metode *Support Vector Machin* (SVM) dengan menerapkan *Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA) serta bagaimana suatu sistem dikembangkan berdasarkan algoritma yang digunakan dengan mengimplementasikan model *Machine Learning* dengan tahapan yang dilalui berdasarkan metode CRISP-DM.

BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bab terakhir berisi kesimpulan dan saran dari hasil penelitian, serta merupakan garis besar dari metode penelitian yang telah dilakukan. Kesimpulan adalah hasil akhir dari penelitian yang dilakukan, sedangkan Saran berisi tentang rekomendasi sesuai dengan keterbatasan yang ada pada sistem.