

BAB I PENDAHULUAN

1.1 Latar Belakang

Dalam kehidupan sehari-hari seringkali terjadi perbedaan pendapat terhadap sebuah isu. Perbedaan pendapat dapat dilihat secara luas pada media sosial, antara lain *Facebook, Instagram, Twitter, Youtube*[1]. Bentuk perbedaan pendapat dapat diamati dari kolom komentar dari sebuah postingan, dari komentar sebuah postingan tersebut dapat dikelompokkan menjadi dataset sentimen yang berbentuk pendapat positif dan negatif. Dataset analisis sentimen seringkali cenderung tidak seimbang dimana jumlah sampel positif dan negatif sangat berbeda sehingga dapat berdampak negatif pada klasifikasi dimana model cenderung menjadi bias terhadap kelas minoritas[2].

Imbalance data merupakan situasi kelas atau label pada dataset tidak seimbang, artinya sampel untuk setiap kelas tidak merata. Dalam konteks klasifikasi, ini berarti bahwa satu atau beberapa kelas memiliki frekuensi yang signifikan lebih tinggi atau lebih rendah dengan kelas-kelas lainnya. Hal tersebut dapat mengakibatkan pada kinerja klasifikasi tiap algoritma[3].

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi berbasis *supervised learning* yang digunakan untuk memprediksi kelas data berdasarkan pola yang dipelajari selama proses pelatihan. Dalam proses klasifikasi, SVM bekerja dengan membentuk sebuah *hyperplane* yang berfungsi sebagai batas pemisah antara kelas positif dan kelas negatif. *Hyperplane* yang optimal yang memiliki jarak paling jauh terhadap titik data terdekat dari masing-masing kelas, yang dikenal sebagai *support vectors*. Jarak tersebut disebut sebagai *margin*, dan *margin* yang semakin besar cenderung mampu meningkatkan kemampuan generalisasi model serta meminimalkan kesalahan prediksi pada data baru. Proses pencarian pemisahan terbaik dilakukan dengan memaksimalkan nilai *margin* dari *hyperplane* tersebut, sehingga diperoleh batas keputusan yang paling optimal. Berbeda dengan SVM yang mencari garis pemisah terbaik dengan jarak paling lebar antar kelas, algoritma *Naive Bayes* menentukan kelas berdasarkan perhitungan

peluang dan menganggap setiap fitur tidak saling berkaitan. Karena cara kerjanya berbeda, keduanya juga memiliki kelebihan dan kekurangan masing-masing dalam melakukan klasifikasi data[4]. Namun penelitian ini menerapkan metode *Support Vector Machine* (SVM) yang dikenal efektif dalam menangani masalah klasifikasi berbasis teks, dengan dukungan *kernel Radial Basis Function (RBF)* untuk meningkatkan kemampuan pemisahan kategori data[5].

1.2 Rumusan Masalah

Rumusan masalah yang digunakan pada penelitian ini adalah bagaimana perbandingan model SVM menggunakan SMOTE dan tanpa menggunakan SMOTE pada data sentimen kemudian akan dihasilkan akurasi yang bertujuan untuk menentukan kinerja terbaik dari penggunaan SMOTE

1.3 Batasan Masalah

1. Menggunakan 5 Dataset yang diperoleh dari data publik
2. Hanya memiliki dua kelas yaitu kelas positif dan kelas negatif
3. Metode yang digunakan dalam analisis sentimen adalah *Support Vector Machine*

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengetahui perbandingan model SVM menggunakan SMOTE dan tanpa menggunakan SMOTE pada data sentimen dengan tujuan menentukan apakah penggunaan SMOTE meningkatkan akurasi klasifikasi sentimen, serta untuk menentukan model SVM mana yang memberikan kinerja terbaik dalam mengklasifikasikan sentimen.

1.5 Manfaat Penelitian

a. Manfaat Teoritis

Penelitian ini diharapkan memberikan kontribusi baru terhadap pengetahuan dalam bidang analisis sentimen, khususnya dalam konteks penggunaan *Support Vector Machine* (SVM) dengan dan tanpa *Synthetic Minority Oversampling Technique* (SMOTE) untuk menangani

ketidakseimbangan dataset

b. Manfaat Praktis

Menyediakan petunjuk bagi praktisi dan peneliti dalam memilih pendekatan yang paling efektif ketika menggunakan SVM untuk analisis sentimen dengan ketidakseimbangan dataset. Ini dapat membantu pengguna memilih strategi yang sesuai dengan kebutuhan spesifik mereka

1.6 Sistematika Penulisan

Pada bagian ini diuraikan sistematika penulisan dari penyusun skripsi ini yang disajikan secara sistematis sebagai berikut:

BAB I PENDAHULUAN, berisi Latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, sistematika penelitian.

BAB II TINJAUAN PUSTAKA, berisi tinjauan pustaka, dasar-dasar teori yang digunakan,

BAB III METODE PENELITIAN, didalamnya terdapat tinjauan umum tentang objek penelitian, analisis masalah, solusi yang ditawarkan, rancangan,

BAB IV HASIL DAN PEMBAHASAN, bab ini merupakan tahapan yang penulis lakukan dalam mengembangkan aplikasi, testing hingga penerapan aplikasi di objek penelitian,

BAB V PENUTUP, berisi kesimpulan dan saran yang dapat peneliti rangkum selama proses penelitian,