

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis sentimen adalah proses perhitungan dan analisis terhadap pendapat masyarakat dalam menanggapi atau mengomentari suatu isu atau topik yang disampaikan secara tertulis dalam berbentuk teks[1]. Pada umumnya analisis sentimen sering dimanfaatkan oleh perusahaan yang menerbitkan produk yang menyediakan layanan untuk menganalisis pendapat dari konsumen terhadap produk tersebut[2]. Namun, tidak hanya dalam dunia bisnis saja, analisis sentimen dapat digunakan untuk mengkategorikan opini positif dan negatif dari masyarakat terhadap suatu objek wisata baik wisata didalam negeri ataupun diluar negeri yang melibatkan banyak individu sehingga objek wisata tersebut dapat ditinjau kembali untuk mendapatkan respon positif dan menarik perhatian para wisatawan. Apabila mendapatkan sentimen negatif, maka pihak yang bertanggung jawab terhadap objek wisata tersebut dapat dengan cepat mengambil tindakan untuk menanggulangnya.

Berbagai ulasan terhadap sebuah objek wisata salah satunya KBS yang telah berdiri dari tahun 1916 membuat para wisatawan seringkali menimbang ketika ingin mengunjunginya. Ulasan tersebut berisi tanggapan positif dan negatif dari berbagai kalangan masyarakat di Tripadvisor dan *Google maps*. Penggunaan analisis sentimen dapat diterapkan dalam pengklasifikasikan ulasan di Tripadvisor dan *Google Maps* terhadap Wisata KBS menggunakan algoritma klasifikasi yang ada, salah satunya SVM.

Algoritma perhitungan SVM digunakan dalam penelitian ini untuk menemukan *hyperplane* yang optimal dengan memaksimalkan jarak antar kelas. *Hyperplane* adalah fitur yang dapat digunakan untuk memisahkan *class* [1]. Metode SVM akan menjadi bias jika data tidak seimbang. Dalam analisis sentimen pada dataset yang tidak seimbang akan menjadi sulit karena pengklasifikasian cenderung mendeteksi kelas mayoritas daripada kelas minoritas. Proses yang diperlukan yaitu melakukan penyeimbangan data. Penyeimbangan data dapat dilakukan oleh metode SMOTE[19]. SMOTE akan melakukan *oversampling* terhadap kelas minoritas agar jumlah kelas positif dan negatif seimbang.

Beberapa model algoritma klasifikasi sentimen yang umum digunakan adalah *Random Forest*, *SVM*, dan *Naives Bayes*. Penelitian sebelumnya yang telah dilakukan oleh Rahayu Prihatini Saputri telah menganalisis sentimen terkait wisatawan candi borobudur dengan perbandingan 2 metode yaitu *K-Nearest Neighbors*(k-NN) dan *SVM* dan ekstraksi fitur N-Gram. Hasil dari penelitian tersebut *SVM* lebih unggul dari k-NN dengan akurasi 87% untuk *SVM* dan 41% untuk K-NN[2]. Penelitian terkait dilakukan oleh Frizka Fitriana yang telah menganalisis sentimen terkait kasus vaksin Covid-19 di jejaring sosial *Twitter* menggunakan metode *Naïve Bayes* dan *SVM*. Hasil tertinggi yang diperoleh *SVM* dengan nilai akurasi 90,47% sedangkan *Naïve Bayes* adalah 88,64%[3]. Penelitian terkait dilakukan oleh Elza Fitriana Saraswita yang telah menganalisis sentimen terkait *E-wallet* di media sosial *Tripadvisor* dan *Google Maps*. Pada penelitian tersebut memanfaatkan seleksi fitur *RFE* untuk menambah akurasi. Hasil pengujian dari penelitian tersebut terbukti meningkatkan akurasi hingga 81%[1]. Penelitian lain dilakukan oleh Nur Syafiqah, Penelitian tersebut melakukan analisis terhadap ulasan film di situs *IMDb*, Penelitian ini menggunakan algoritma *SVM-RFE*. Pada penelitian ini hasil akurasi terbaik terdapat pada jumlah fitur berada di 700 *top features* dari 1000 *features* yaitu sebesar 80.05%.

Berdasarkan penjabaran dari penelitian terdahulu menunjukkan bahwa klasifikasi dengan menggunakan metode *SVM* dan memanfaatkan seleksi fitur *RFE* ini sangat baik dalam klasifikasi data dilihat dari hasil akurasi, maka penulis mengajukan penelitian dengan judul “Analisis sentimen Wisata Kebun Binatang Surabaya

dengan Metode *Support Vector Machine* dan *Recursive Feature Elimination*”.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Dengan adanya *Tripadvisor* dan *Google Maps* memudahkan seseorang dalam mencari ulasan-ulasan yang mengenai wisata KBS. Pengklasifikasian sentimen tanggapan berupa tanggapan positif dan negatif. dapat dilakukan dengan metode *SVM*. Namun Algoritma *SVM* tidak dapat bekerja dengan baik ketika dataset memiliki banyak *noise*, data tidak seimbang dan jika jumlah fitur melampaui data training[6]. Sehingga dalam penelitian kali ini akan mencoba mengeliminasi fitur terendah dengan seleksi fitur *RFE*.

1.2.2 Metode Usulan

Penelitian ini mengusulkan metode SVM yang berfungsi sebagai algoritma pengklasifikasian sentimen dengan data ulasan dari Tripadvisor dan *Google Maps*. Data yang diperoleh melalui Tripadvisor dan *Google Maps* akan dilabelkan oleh dua orang mahasiswa dan akan divalidasi oleh seorang guru bahasa Indonesia. Data akan di bagi menjadi dua untuk data latih dan data uji. Setelah itu, data akan melakukan *text preprocessing*. setelah data diolah *text preprocessing* maka data akan dibobotkan. Pembobotan kata dilakukan dalam proses ekstraksi fitur dengan metode TF-IDF, setelah melalui ekstraksi fitur maka ada empat proses yaitu, proses pengklasifikasian SVM pada data tidak seimbang, proses pengklasifikasian SVM pada data seimbang, proses pengklasifikasian SVM dengan RFE pada data tidak seimbang, dan proses pengklasifikasian SVM dengan RFE pada data seimbang.

1.2.3 Pertanyaan Penelitian

Bagaimana pengaruh penyeimbangan data dan seleksi fitur RFE terhadap hasil kinerja algoritma SVM berupa perhitungan waktu dan hasil akurasi terhadap analisis sentimen masyarakat terkait Wisata KBS?

1.3 Tujuan & Manfaat

1.3.1 Tujuan Penelitian

Setelah pertanyaan penelitian ada tujuan yang akan dihasilkan dari penelitian ini ialah seberapa besar pengaruh penerapan dari penyeimbangan data dan seleksi fitur RFE terhadap hasil kinerja algoritma SVM berupa perhitungan waktu dan hasil akurasi terhadap analisis sentimen terkait Wisata KBS.

1.3.2 Manfaat

Setelah tujuan penelitian, manfaat yang diharapkan yang akan dihasilkan dari penelitian ini ialah mengetahui pengaruh penerapan penyeimbang data dan seleksi fitur RFE untuk memaksimalkan kinerja algoritma SVM terhadap analisis sentimen masyarakat terkait Wisata KBS sehingga dapat diolah untuk penelitian yang akan datang dan sebagai media pendukung terhadap Wisata KBS agar melakukan tinjauan kembali untuk mendapatkan respon positif dan menarik perhatian para wisatawan. Apabila mendapatkan sentimen negatif, maka dapat dengan cepat mengambil tindakan untuk menanggulangnya.

1.4 Batasan-batasan

1. Dalam penelitian diperlukan adanya penentuan batasan masalah, supaya penelitian tetap terstruktur dan tidak menyimpang atau pelebaran pokok permasalahan yang akan diteliti nantinya.
2. Dalam pelabelan manual menggunakan bantuan dari 2 orang mahasiswa dan akan divalidasi oleh seorang guru Bahasa Indonesia Shafira Ega Alya Namira, S.Pd.
3. Penelitian terfokus pada analisis opini berbahasa Indonesia dari masyarakat dengan kata kunci "Kebun Binatang Surabaya" melalui Tripadvisor dan *Google Maps*, data akan dilabelkan kelas positif dan negatif.

1.5 Sistematika Laporan

Sistematika penulisan ditunjukkan untuk memberikan pembahasan secara ringkas atau uraian secara garis besar, sistematika yang telah disesuaikan dengan format penulisan skripsi teknik informatika Universitas Trunojoyo Madura sebagai berikut:

BAB I: PENDAHULUAN

Bab ini memberikan latar belakang, rumusan masalah, keterbatasan, tujuan, kekuatan, dan sistem penulisan yang diusulkan.

BAB II : TINJAUAN PUSTAKA

Pada bab ini berkaitan dengan penelitian terdahulu sebagai tinjauan pustaka yang akan membantu untuk mendasari penelitian berupa teori yang terkait dengan Analisis sentimen, SMOTE, Tripadvisor dan *Google Maps*, SVM, TF-IDF, RFE, dan kerangka teori lainnya yang terkait dengan tema.

BAB III : METODE USULAN

Pada bab ini menerangkan alur dari penelitian, bagaimana data itu diproses melalui *text preprocessing* dilanjutkan dengan pembobotan kata memakai metode TF-IDF setelah pembobotan kata data akan diseimbangkan, setelah data seimbang akan dilakukan proses seleksi fitur RFE. Sehabis seleksi fitur langsung ketahap klasifikasi dengan metode SVM.

BAB IV : TINJAUAN PUSTAKA

Pada bab ini menerangkan hasil klasifikasi sentimen dengan skenario yang telah di tunjukkan pada metode usulan dan akan dilakukan pengukuran akurasi dengan menggunakan *confusion matrix* sehingga mendapatkan *accuracy*, *precision*, *recall*, dan *f1-score*.

BAB V : PENUTUP

Pada bab ini menjelaskan kesimpulan dari hasil dan pembahasan serta saran untuk pengembangan sistem selanjutnya.