

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *Machine Learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dari pengetahuan yang terkait dari berbagai *database* besar. Pada *data mining* ada beberapa algoritma, diantaranya adalah klasifikasi. Klasifikasi adalah algoritma pada *data mining* yang menggolongkan atau menentukan suatu kriteria dari suatu data yang didasarkan pada data sebelumnya yang sudah dipelajari[1].

Dalam penggunaan klasifikasi, tidak hanya diterapkan pada data numerik, akan tetapi juga bisa diterapkan pada data teks. Penambangan teks atau *text mining* memiliki arti secara luas sebagai proses penggalian informasi di mana pengguna berinteraksi dengan sekumpulan dokumen menggunakan alat analisis yang merupakan komponen dalam penambangan data[2]. Penambangan teks atau *text mining* adalah teknik yang digunakan untuk menyelesaikan suatu kasus seperti *information retrieval classification*, dan juga *information extraction*. Konsep *text mining* yang dipakai dalam klasifikasi adalah dokumen teks dengan tujuan untuk mengklasifikasikan dokumen sesuai dengan pembahasan topik[3].

Salah satu konsep yang sering digunakan dalam *text mining* adalah analisis sentimen. Analisis Sentimen, juga dikenal sebagai *opinion mining*, adalah cabang dari penambangan data yang tujuan dasarnya yaitu untuk menganalisis, memahami, memproses, dan mengekstrak data tekstual dalam bentuk pendapat atau ulasan tentang entitas seperti produk, layanan, organisasi, individu, dan topik tertentu. Metode klasifikasi seperti *Support Vector Machine* (SVM), *Naive Bayes*, dan *Decision Tree* sering digunakan untuk membantu dalam proses analisis sentimen[4].

Metode SVM memiliki tingkat akurasi yang cukup baik dalam pengklasifikasian data teks. Pada penelitian sebelumnya tentang “Perbandingan

Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter” mendapatkan hasil akurasi SVM 89,70% dan KNN 88,76 %[5]. Selanjutnya pada penelitian tentang “Perbandingan Metode *Decision Tree* dan *Support Vector Machine* untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI” mendapatkan hasil akurasi SVM 94,76%, *Decision Tree* 87,45%[6]. Selanjutnya penelitian tentang “*Sentiment Analysis For Extracting Student Opinion Data On Higher Education Services Using The Naive Bayes Classifier and Support Vector Machine Methods*” mendapatkan hasil *Naive Bayes* 96,56% dan SVM 99,09%[7]. Dari penelitian tersebut metode SVM memiliki kinerja yang cukup baik dibandingkan dengan metode klasifikasi lainnya dalam pengklasifikasian data teks.

Pada penelitian ini akan dilakukan sentimen analisis terhadap ulasan masyarakat mengenai tempat wisata Api Tak Kunjung Padam. Data ulasan disini diambil dari ulasan *Google Maps Review* dimana tujuannya untuk mengetahui penerapan metode klasifikasi *Support Vector Machine* dalam pengklasifikasikan sentimen positif ataupun sentimen negatif pada data ulasan pengunjung objek wisata tersebut dengan harapan metode tersebut bekerja cukup baik dan menghasilkan akurasi diatas 80% dan juga hasil analisis ulasan tersebut dapat memberikan informasi yang berguna bagi pemerintah khususnya Dinas Kebudayaan dan Pariwisata Kota Pamekasan, pengelola objek wisata Api Tak Kunjung Padam maupun pihak lain yang membutuhkan dalam mengelola tempat wisata menjadi lebih terarah dan baik.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Pariwisata merupakan salah satu sektor dari sekian banyak yang mempengaruhi tingkat ekonomi suatu daerah. Tidak hanya itu wisata juga termasuk sumber pendapatan daerah. Untuk itu pemerintah memerlukan informasi mengenai kualitas objek wisatanya dimana opini masyarakat adalah sumber informasi bagi pemerintah yang dapat dijadikan sebagai evaluasi objek wisata. Informasi mengenai objek wisata Api Tak Kunjung Padam

adalah ulasan pada *Google Maps*. Namun, *Google Maps* tidak mempunyai kemampuan untuk menganalisa informasi yang ada untuk ditarik sebuah kesimpulan. Maka dari itu suatu pendekatan diperlukan, untuk mendapatkan informasi mengenai kualitas suatu objek wisata berdasarkan kesimpulan dari sentimen atau opini masyarakat.

1.2.2 Metode Penelitian

Untuk menyelesaikan permasalahan yang ada dalam penelitian ini akan digunakan metode *Support Vector Machine*.

1.2.3 Pertanyaan Penelitian

Adapun pertanyaan penelitian yang didapat dari rumusan masalah diatas yaitu;

1. Berapa akurasi yang dihasilkan metode *Support Vector Machine* berdasarkan parameter *Hyperplane C* dan *gamma* untuk klasifikasi sentimen positif dan negatif pada data ulasan *Google Maps* mengenai tempat wisata Api Tak Kunjung Padam?
2. Bagaimana memvisualisasikan kata yang sering muncul pada ulasan wisata Api Tak Kunjung Padam?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui akurasi yang dihasilkan metode *Support Vector Machine* untuk mengklasifikasikan sentimen positif dan negatif pada data ulasan *Google Maps* mengenai tempat wisata Api Tak Kunjung Padam.
2. Untuk mendapatkan informasi kata apa saja yang sering muncul pada ulasan Api Tak Kunjung Padam.

1.3.2 Manfaat Penelitian

Adapun manfaat dilakukannya penelitian ini adalah:

1. Untuk mengetahui kinerja dari metode *Support Vector Machine* dalam melakukan klasifikasi data teks.
2. Dapat mejadi rujukan dan refrensi tambahan tentang klasifikasi data teks dengan menggunakan metode *Support Vector Machine*.

1.4 Batasan Masalah

Berikut adalah batasan-batasan penelitian yang dilakukan dalam penelitian ini:

1. Data dalam penelitian ini adalah data ulasan *Google Maps* dengan jumlah 861 ulasan.
2. Metode yang digunakan pada penelitian ini yaitu metode *Support Vector Machine*.
3. Bahasa pemrograman menggunakan bahasa pemrograman *Python*.

1.5 Sistematika Penulisan

Adapun sistematika proposal yang digunakan terdiri dari beberapa bagian, yaitu

BAB I PENDAHULUAN

Bagian pendahuluan berisi latar belakang , rumusan masalah, tujuan dan manfaat, batasan masalah dan sistematika proposal.

BAB II LANDASAN TEORI

Bagian ini membahas mengenai beberapa teori untuk menunjang penelitian, diantaranya yaitu definisi *Text mining*, *text preprocessing*, teori tentang klasifikasi dan juga metode *Support Vector Machine*.

BAB III ANALISA DAN PERANCANGAN SISTEM

Secara garis besar dibagian ini membahas tentang penerapan konsep dasar yang sebelumnya telah dijelaskan pada Bab II untuk menganalisis dan merancang sistem sesuai tahap-tahap untuk menyelesaikan masalah tersebut dengan menggunakan metode *Support Vector Machine*.

BAB IV HASIL DAN PEMBAHASAN

Pada Bab ini dijelaskan mengenai implementasi sistem yang dikerjakan sesuai dengan rancangan yang dibuat pada Bab sebelumnya. Dimana terdapat hasil dari tahap *crawling* data pada situs *Google Maps*, hasil dari tahap pelabelan manual, hasil dari tahap *preprocessing* data (*case folding, cleaning, tokenizing, konversi slangword, stopword removal, stemming*), hasil dari tahap pembobotan dengan TF-IDF, hasil dari tahap klasifikasi dengan metode *Support Vector Machine*, hasil dari tahap evaluasi model menggunakan *confusion matrix*, terakhir terdapat hasil dari uji skenario.

BAB V PENUTUP

Pada bab ini dijelaskan tentang kesimpulan dari hasil penelitian yang meliputi visualisasi kata, hasil *accuracy* serta hasil nilai *precision, recall* dan *f-1 score* yang didapatkan. Selain itu juga terdapat saran bagi penelitian berikutnya yang akan mengangkat topik yang sama yaitu analisis sentimen