

BAB I

PENDAHULUAN

1.1. Latar Belakang

Keindahan Alam pada suatu negara adalah salah satu hal yang dapat menghasilkan kontribusi positif dengan memanfaatkan dan mengoptimalkan potensi yang dimiliki. Indonesia terkenal dengan sejumlah wisata dengan keindahan alam dan keanekaragaman yang dimiliki yang membuat para wisatawan yang berkunjung terus mengalami peningkatan, sehingga Industri pariwisata mengalami perkembangan yang sangat pesat dari tahun ke tahun [1]. Malang termasuk salah satu kabupaten yang ada di Jawa Timur yang terkenal di dalamnya memiliki segudang wisata yang layak dan banyak dikunjungi oleh wisatawan dengan keindahan alam yang dimiliki. Saat ini pengembangan industri dalam bidang pariwisata lebih banyak membantu dan mendorong dalam meningkatkan pendapatan negara khususnya dan juga masyarakat [2]. Kondisi tersebut membuat pemerintah pada tingkat daerah melakukan perbaikan di daerahnya masing-masing demi meningkatkan kunjungan dalam sektor pariwisata dengan mengoptimalkan citra pariwisata yang lebih baik. Maka dari itu, dibutuhkan informasi untuk mengetahui opini para wisatawan mengenai wisata yang ada di kabupaten Malang. Sehingga dapat membantu pemerintah kabupaten Malang untuk bahan evaluasi dalam pengembangan wisata kabupaten Malang dengan tujuan menarik wisatawan dan memajukan sektor pariwisata dikarenakan wisata dan budaya merupakan sektor yang sangat berpengaruh untuk berlangsungnya suatu daerah melalui media sosial.

Menurut laporan terbaru agensi marketing *We Are Social* dan platform manajemen media sosial Hootsuite pada tahun 2022, bahwa terdapat sekitar 204,7 juta pengguna internet di Indonesia dari total populasi 277,7 juta penduduk jiwa yang ada di Indonesia. 191,4 juta diantaranya merupakan pengguna aktif media sosial. Media sosial yang banyak digunakan oleh pengguna diantaranya yaitu Twitter, Facebook, dan lain sebagainya. Twitter adalah salah satu media sosial yang sangat banyak digunakan di kalangan pengguna internet, hal ini dikarenakan penggunaannya yang sederhana dan mudah, pengguna juga dapat dengan bebas berpendapat dan beropini [3]. Hal tersebut juga dimanfaatkan oleh para wisatawan yang ingin mengutarakan berbagai ulasan atau komentar terkait perjalanan wisata yang telah dilalui menggunakan sosial media. Banyaknya ulasan dan komentar yang diberikan sehingga dibutuhkan sebuah cara untuk menganalisis yaitu dengan menerapkan analisis sentimen.

Analisis sentimen adalah sebuah pemodelan dalam pengelolaan informasi yang didapat mengenai sikap seseorang dengan melakukan pengelompokan polaritas dalam menanggapi suatu kejadian ke dalam beberapa kelas dalam sebuah teks [4]. Pengelompokan tersebut dapat dilakukan

untuk mengetahui teks tersebut masuk kedalam kelas positif, negatif, atau netral. Banyaknya penelitian yang mengenai analisis sentiment seperti penelitian analisis sentimen yang dilakukan oleh Kartika Fithriasari pada tahun 2022 terkait analisis sentimen wisatawan candi borobudur menggunakan metode Support Vector Machine dan K-Nearest Neighbor. Berdasarkan hasil pengujian yang telah dilakukan, dihasilkan akurasi yang terbaik dengan menggunakan metode Support Vector Machine (SVM) dibanding metode K-Nearest Neighbor [5].

Metode Support Vector Machine (SVM) adalah salah satu metode klasifikasi yang diciptakan oleh Vladimir Vapnik dengan menggunakan metode *machine learning* yang memprediksi kelas berdasarkan pola dari hasil proses training[6]. SVM memiliki kelebihan yaitu mampu mengklasifikasikan teks dalam data yang banyak serta lengkap [7]. Support Vector Machine (SVM) adalah metode terbaik dari beberapa metode lain karena mampu melakukan komputasi data sehingga tingkat akurasi yang dihasilkan tinggi [8]. Meskipun hasil akurasi yang didapatkan oleh metode SVM memiliki nilai yang tinggi, metode SVM memiliki kelemahan ketika data yang digunakan memiliki dimensi yang tinggi. Menurut Yu dan Liu, data yang memiliki dimensi tinggi adalah data yang memiliki jumlah fitur atau atribut yang banyak sehingga terdapat informasi yang kurang relevan dikarenakan memakan waktu yang cukup lama sehingga mengakibatkan menurunnya kinerja dari SVM [9]. Berdasarkan permasalahan tersebut maka dibutuhkan seleksi fitur untuk mengurangi fitur agar memperlancar kerja komputasi metode SVM sehingga mempercepat dalam proses klasifikasi. Menurut penelitian yang dilakukan oleh Purnamasari, et al, menggunakan seleksi fitur *information gain* dengan metode SVM mendapatkan hasil dengan akurasi yang tinggi dibandingkan dengan menggunakan seleksi fitur yang ada [10]. Agar proses klasifikasi berjalan dengan sempurna maka diperlukan proses pembobotan kata menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) agar dapat menghitung dan menentukan bobot dari setiap kata sehingga dapat melakukan penyeleksian kata yang kurang berbobot.

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan suatu metode yang berguna dalam melakukan pembobotan kata pada dokumen. Metode ini merupakan gabungan antara *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). *Term frequency* (TF) berguna untuk perhitungan bobot kata pada sebuah kalimat, semakin banyak jumlah kata muncul, maka semakin tinggi nilai bobotnya. Begitupun sebaliknya, apabila semakin sedikit jumlah kata yang muncul, maka nilai bobot akan semakin kecil. Sedangkan *Inverse Document Frequency* (IDF) dipakai sebagai metode pengurangan nilai bobot kata (Term) yang lebih banyak pada dokumen [11].

Seperti pada penelitian analisis sentimen yang telah dilakukan oleh Darwis, Dedi Pratiwi, Eka Shintya Pasaribu, dan A Fericon Octaviansyah. Pada penelitian ini menggunakan metode *Support Vector Machine* dan *TF-IDF*. Data yang didapatkan dari hasil *crawling* memperoleh 2000 data dan menghasilkan 1890 data yang diklasifikasikan menjadi kedalam 3 kelas yaitu positif, negatif dan netrall. Hasil akurasi yang didapatkan adalah 82%. Sementara itu peneliti yang dilakukan oleh Catur Supriyanto mengenai analisis sentimen pariwisata kota semarang menggunakan algoritma *support vector machine* pada tahun 2019. Berdasarkan hasil dari proses pengujian yang telah dilakukan terhadap tweets sentimen pariwisata kota semarang dari media sosial twitter sebanyak 1300 sentimen positif dan 700 sentimen negatif menggunakan algoritma *Support Vector Machine* dengan nilai akurasi 91.95% [12]. Berdasarkan hasil penelitian yang telah dilakukan sebelumnya, dapat diperoleh kesimpulan bahwa metode SVM dapat menyelesaikan permasalahan analisis sentimen terhadap komentar *twitter* dengan memperoleh akurasi yang terbaik.

Oleh karena itu, berdasarkan penulisan latar belakang tersebut, maka penulis ingin melakukan penelitian kembali menggunakan metode *support Vector Machine (SVM)* dengan mengajukan penelitian dengan judul Analisis Sentiment komentar para wisatawan mengenai wisata di Kabupaten malang dengan menggunakan TF-IDF dan algoritma *Support Vector Machine*. Penelitian ini menggunakan TF-IDF sebagai pembobotan kata serta menggunakan seleksi fitur *Information Gain* untuk mengurangi dimensi fitur dalam meningkatkan akurasi dari metode *support vector machine (SVM)* pada ulasan wisata Kabupaten Malang. Dengan dilakukan penelitian ini, diharapkan sistem yang dihasilkan dapat digunakan untuk mengukur sentimen ulasan masyarakat mengenai tempat wisata di Kabupaten Malang.

1.2. Perumusan Masalah

1.2.1. Permasalahan

Permasalahan dalam penelitian ini adalah pengembangan industri dalam bidang pariwisata lebih banyak membantu dan mendorong dalam meningkatkan pendapatan negara khususnya dan juga masyarakat. Kondisi tersebut membuat pemerintah pada tingkat daerah melakukan perbaikan sehingga dibutuhkan informasi untuk mengetahui opini para wisatawan mengenai wisata yang ada di kabupaten Malang sebagai bahan evaluasi dalam pengembangan wisata kabupaten malang. Penelitian ini menggunakan metode *Support Vector Machine (SVM)* untuk analisis sentimen komentar paran wisatawan di kabupaten malang yang memiliki kelemahan ketika data yang digunakan memiliki dimensi yang tinggi sehingga mengakibatkan menurunnya kinerja dari SVM.

1.2.2. Solusi permasalahan

Berdasarkan permasalahan yang ada, penelitian ini mengusulkan metode *Support Vector Machine (SVM)* yang digunakan untuk klasifikasi komentar para wisatawan di kabupaten Malang dengan membutuhkan penggunaan *information gain* untuk mengurangi dimensi fitur serta memerlukan perubahan kernel yang digunakan untuk menghasilkan klasifikasi yang akurat dari SVM.

1.2.3. Pertanyaan Penelitian

1. Bagaimana pengaruh perubahan kernel terhadap akurasi hasil analisis sentimen menggunakan algoritma *Support Vector Machine (SVM)* dengan menggunakan TF-IDF terhadap komentar pariwisata yang ada pada Kabupaten Malang?
2. Bagaimana hasil klasifikasi metode *Support Vector Machine (SVM)* dengan menggunakan TF-IDF terhadap analisis sentimen mengenai komentar wisata di Kabupaten Malang?

1.3. Tujuan

1. Mengetahui uji akurasi terhadap perubahan kernel hasil analisis sentimen mengenai komentar para wisatawan pada media sosial Twitter terkait wisata yang ada pada kabupaten Malang dengan menggunakan metode *support vector machine (SVM)*.
2. Mengetahui hasil klasifikasi metode *Support Vector Machine (SVM)* dengan menggunakan TF-IDF sebagai pembobotan kata untuk menghitung banyaknya kemunculan kata dalam sebuah dokumen yang bersumber pada media sosial Twitter mengenai komentar wisata di Kabupaten Malang untuk pengelolaan dan pengembangan tempat wisata.

1.4. Manfaat

Dalam penelitian ini penulis berharap manfaat yang dicapai dalam penelitian adalah sebagai berikut:

1. Dengan menggunakan Algoritma *Support Vector Machine (SVM)* dapat melakukan klasifikasi sentimen terhadap data ulasan pariwisata di kabupaten Malang.
2. Dengan adanya penelitian ini dapat bermanfaat sebagai sumber pembelajaran untuk penelitian selanjutnya serta untuk menambah wawasan.

1.5. Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Algoritma yang digunakan adalah algoritma Support Vachine Machine dan tidak menggunakan algoritma lain untuk perbandingan.
2. Data yang digunakan merupakan dokumen tweets yang bersumber dari media sosial Twitter berbahasa indonesia dengan menggunakan kata kunci “wisata kabupaten malang”.
3. Data memiliki tiga kelas yaitu kelas positif, kelas negatif dan kelas netral.

1.6. Sistematika Proposal

Penyusunan proposal ini berdasarkan sistematika sebagai berikut:

Bab I Pendahuluan

Bab ini merupakan bagian awal yang berisi mengenai penjelasan latar belakang masalah, Perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika proposal.

Bab II Landasan Teori

Bab ini berisi mengenai pembahasan teori-teori yang digunakan dalam penelitian untuk menyelesaikan permasalahan mengenai analisis sentimen pada dokumen Twitter dengan metode SVM dan TF-IDF serta dasar teori yang digunakan.

Bab III Metodologi Penelitian

Bab ini berisi mengenai rancangan sistem terkait bagaimana alur dalam mengumpulkan data dari media sosial twitter, alur *preprocessing* data, alur dalam melakukan pembobotan menggunakan TF-IDF, dan alur untuk melakukan klasifikasi dengan metode *Support Vector Machine (SVM)*.

Bab IV Hasil pengujian dan Implementasi

Bab ini memuat mengenai implementasi sistem sesuai dengan rancangan yang telah dibuat pada bab sebelumnya yang berisi tahap *preprocessing* data (*tokenizing, cleansing, case folding*, konversi *slangword, stemming, stopword*), hasil data tahap pembobotan dengan TF-IDF, Hasil dari tahap *Support Vector Machine*, hasil dari tahap evaluasi model menggunakan *confusion matrix*, dan juga terdapat hasil dari uji skenario.

Bab V Penutup

Bab ini memuat terkait kesimpulan penelitian yang sudah dilakukan meliputi akurasi yang dihasilkan, serta saran untuk pengembangan lebih lanjut.