

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tempat wisata merupakan salah satu tempat untuk hiburan di masa libur bersama keluarga maupun kerabat terdekat. Wisata pantai mempunyai beberapa kategori, yang mana kegiatan berwisata yang sering dikenal salah satunya yaitu wisata rekreasi pantai. Rekreasi merupakan kegiatan favorit masyarakat yang dilakukan untuk memulihkan kondisi fisik yang letih juga kondisi mental setelah beberapa hari mengalami kelelahan dalam bekerja [1]. Seringkali masyarakat menghabiskan waktu saat liburan untuk pergi ke pantai karena dapat dilakukan berbagai kegiatan seperti bermain ombak, membuat istana pasir, atau sekedar berjemur di pantai. Itulah alasan mengapa Pantai selalu menjadi tempat rekreasi favorit masyarakat.

Kabupaten Sumenep merupakan salah satu kabupaten yang berada di Provinsi Jawa Timur, secara geografis terletak diantara 113° 32' - 116° 16' BT dan 4° 55' - 7° 24' LS [2]. Kabupaten Sumenep memiliki banyak tempat wisata diantaranya wisata pantai, bangunan sejarah, sebagainya. Diantaranya adalah wisata pantai Sembilan ini. Pantai yang cukup terkenal di Kabupaten Sumenep. Namun walaupun cukup terkenal dengan keindahan pantainya, masih terdapat beberapa kekurangan dari berbagai aspek seperti kebersihan, keamanan dan keindahan.

Hal ini dapat dilihat dari review ulasan masyarakat yang berkunjung ke tempat wisata. Pantai Sembilan ini memiliki banyak ulasan dibandingkan pantai lain yang ada di Kabupaten Sumenep, hal ini menjadikan alasan bahwa masyarakat khususnya warga Kabupaten Sumenep sendiri sering berlibur di pantai ini. Ulasan masyarakat mengenai wisata Pantai Sembilan ini dapat dilihat melalui Fasilitas yang disediakan *Google* yang bernama *Google Maps*. Merupakan salah satu fasilitas yang dapat menulis ulasan untuk tempat yang pernah dikunjungi [3].

Melalui Fasilitas *Google Maps* didapatkan data ulasan opini masyarakat mengenai *review* suatu tempat wisata misalnya seperti tempat wisata Pantai Sembilan. Pada fitur komentar dapat dilakukan untuk memfilterisasi komentar dengan rating tertinggi dan rating terendah, maupun komentar terbaru [3]. Ulasan dari masyarakat dapat dikategorikan ke dalam ulasan positif, negatif, maupun netral. Untuk me-review ulasan opini masyarakat pada *Google Maps* tersebut dapat dilakukan dengan Analisis Sentimen. Analisis Sentimen antara lain adalah merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan suatu data informasi dari ulasan masyarakat atau opini masyarakat [4].

Pada penelitian ini penulis menggunakan pendekatan strategi klasifikasi Algoritma *Random Forest* karena algoritma metode yang sederhana dan kinerjanya menyerupai pendekatan lain. Algoritma *Random Forest* merupakan suatu algoritma pengklasifikasian teks yang terdiri dari beberapa *Decision Tree* dengan setiap cabangnya dibangun menggunakan vektor yang tidak beraturan [5]. Algoritma ini bersifat *ensemble* yang berarti meningkatkan akurasi klasifikasi dengan cara menggabungkan beberapa metode klasifikasi [5]. Berdasarkan penelitian terdahulu, *Random Forest* memiliki kinerja yang cukup akurat dan optimal, karena dapat mengatasi noising dan missing value serta akurat untuk mengklasifikasikan data dalam jumlah yang cukup besar [6]. Penulis juga menggunakan Metode TF-IDF, merupakan metode yang dapat dimanfaatkan untuk ekstraksi fitur dan pembobotan kata. TF-IDF merupakan metode seleksi fitur dan pembobotan kata yang digunakan untuk mendapatkan nilai dari kata dalam dokumen atau kalimat [7]. Rumus umum TF-IDF berasal dari penggabungan formula perhitungan *raw TF* dan formula IDF dengan cara mengalikan nilai *Term Frequency* (TF) dan nilai *Inverse Document Frequency* (IDF) [8]. Fitur yang dihasilkan dari metode TF-IDF akan sangat banyak karena data yang digunakan berupa data teks. Oleh karena itu, diperlukan proses seleksi fitur untuk memilih fitur atau atribut yang dianggap penting.

Sebelumnya telah dilakukan penelitian analisis sentimen dengan berbagai metode, diantaranya adalah penelitian yang melakukan analisis sentimen terhadap objek wisata di Sulawesi Selatan dengan menggunakan Metode *Random Forest* memiliki tingkat akurasi cukup tinggi sebesar 82%[6]. Selanjutnya penelitian yang melakukan analisis sentimen aplikasi Ruangguru dengan menggunakan *Naive Bayes Classifier*, *Support Vector Machine*, dan *Random Forest*, menunjukkan performa pengujian dengan menggunakan model algoritme *random forest* memiliki akurasi yang sangat tinggi sebesar 97,16% dibanding dengan beberapa model klasifikasi seperti *naïve bayes* sebesar 94,16% dan *support vector machine* (SVM) sebesar 96,01% [9]. Dari hasil analisa tersebut disebutkan pada penelitian-penelitian sebelumnya bahwa algoritma *Random Forest* memiliki tingkat akurasi yang cukup baik.

Dari latar belakang penelitian diatas, maka didapatkan permasalahan mengenai “Analisis Sentimen Persepsi Masyarakat Tentang Wisata Pantai Sembilan Menggunakan *Random Forest*”. Penelitian ini bertujuan untuk mengelompokkan persepsi ulasan masyarakat menjadi tiga kelas yakni positif, negatif, dan netral menggunakan metode *Random Forest* sehingga dapat mengetahui seberapa akurasi opini masyarakat mengenai Pantai Sembilan, Kabupaten Sumenep.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Untuk mengetahui persepsi masyarakat mengenai Pariwisata Pantai Sembilan, Kabupaten Sumenep, diperlukan informasi yang diberikan melalui opini masyarakat tersebut. *Google Maps* merupakan salah satu *platform* yang digunakan untuk menentukan arah tujuan atau tempat yang ingin didatangi, serta terdapat fitur ulasan yang dapat digunakan untuk mengetahui opini masyarakat mengenai tempat tersebut. Tetapi, *Google Maps* tidak mempunyai fitur untuk mengagregasi opini menjadi sebuah kesimpulan. Dari pernyataan diatas tersebut, diperlukan analisis sentimen untuk memetakan opini masyarakat kedalam opini positif, negatif dan netral. Pengklasifikasian dari data tersebut menggunakan metode *Random Forest Classifier*. Proses klasifikasi meliputi text preprocessing, pembobotan kata menggunakan metode TF-IDF dan klasifikasi menggunakan algoritme *Random Forest*.

1.2.2 Metode Usulan

Dalam penelitian ini yaitu melakukan sentimen analisis pada Wisata Pantai Sembilan dengan mengusulkan metode *Random Forest Classifier* sebagai algoritma pengklasifikasian data serta menambahkan TF-IDF sebagai pembobotan kata untuk menunjang hasil akhir dari penelitian ini.

1.2.3 Pertanyaan Penelitian

- 1) Bagaimana pengaruh jumlah dan kedalaman *tree* terhadap hasil penelitian?
- 2) Bagaimana performa *accuracy*, *recall*, *precision*, dan *F1-Score* yang didapatkan menggunakan metode *Random Forest Classifier*?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

- 1) Mengetahui pengaruh jumlah dan kedalaman *tree* terhadap hasil penelitian
- 2) Mendapatkan hasil dari performa *accuracy*, *recall*, *precision*, dan *F1-Score* dengan menggunakan Metode *Random Forest Classifier*

1.3.2 Manfaat Penelitian

- 1) Bagi Sektor Pariwisata
 - a. Mengetahui informasi mengenai opini pengunjung terhadap objek wisata Pantai Sembilan.
 - b. Sebagai sarana evaluasi bagi pemerintahan setempat pada sektor pariwisata untuk dapat meningkatkan sarana dan prasarana yang dinilai masih perlu untuk ditingkatkan.

- 2) Bagi Pengunjung
 - a. Mengetahui opini dari pengunjung lain yang pernah mengunjungi objek wisata Pantai Sembilan sehingga dapat dijadikan acuan untuk layak atau tidaknya objek wisata tersebut dikunjungi.
- 3) Bagi Penulis
 - a. Mengimplementasikan ilmu pengetahuan yang didapatkan masa perkuliahan dalam bidang data mining.

1.4 Batasan Masalah

Adapun batasan masalah dari penelitian adalah sebagai berikut :

- 1) Data yang diambil adalah dari Ulasan Masyarakat pada *Google Maps*.
- 2) Data yang diambil menggunakan Teknik *Scrapping* dengan memasukkan url dari *website* dengan kata kunci “Pantai Sembilan Gmaps”.
- 3) Data yang berhasil diambil adalah sebanyak 1.981 data.
- 4) Data yang diambil adalah ulasan dalam bentuk Bahasa Indonesia.
- 5) Analisis Sentiment dilakukan menggunakan beberapa tingkatan aspek. Aspek yang dimiliki pariwisata Pantai Sembilan, Kabupaten Sumenep antara lain meliputi Aspek Kebersihan, Umum, Keramaian, Akses Jalan, dan Kondisi Ombak.
- 6) Klasifikasi terdiri dari 3 kelas yaitu positif, negatif, dan netral.
- 7) Perhitungan TF-IDF, dan *Random Forest Classifier* dilakukan dengan bantuan library python.

1.5 Metodologi Penulisan Skripsi

Berikut merupakan sistematika penulisan yang akan digunakan pada penyusunan laporan yang disesuaikan dengan format penulisan Skripsi Program Studi Sistem Informasi Fakultas Teknik Universitas Trunojoyo Madura :

BAB I : PENDAHULUAN

Pada bab ini terdapat latar belakang penelitian, tujuan dan manfaat penelitian, rumusan masalah, serta batasan masalah dan sistematika penulisan.

BAB II: KAJIAN PUSTAKA

Pada tinjauan pustaka membahas hal-hal dasar yang berisi teori-teori yang berkaitan dengan analisis sentimen dengan menggunakan metode *Random Forest Classifier* dan hal-hal yang mendukung dalam hal melakukan analisis sentimen ini. Juga penelitian terdahulu.

BAB III: METODE USULAN

Pada bab ini membahas mengenai rancangan dari sistem terkait alur untuk melakukan pengumpulan data yakni scrapping data dari *Google Maps*, alur untuk melakukan *preprocessing* data, alur untuk melakukan pembobotan menggunakan TF-IDF, alur untuk melakukan klasifikasi dengan metode *Random Forest Classifier*.

BAB IV : HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan mengenai implementasi sistem yang dikerjakan sesuai dengan rancangan yang dibuat pada bab sebelumnya. Dimana terdapat skenario pengujian, hasil *scrapping* data menggunakan *WebHarvy*, hasil dari tahap pelabelan manual, hasil dari tahap *preprocessing* data (*cleansing, case folding, tokenizing, stopword removal, konversi slangword, dan stemming*) hasil dari tahap pembobotan dengan TF-IDF, hasil dari tahap *Random Forest Classifier*, hasil dari tahap evaluasi model menggunakan *confusion matrix* dan *k-fold cross validation*, terakhir terdapat hasil dari uji skenario.

BAB V : PENUTUP

Pada bab ini dijelaskan mengenai kesimpulan hasil penelitian yang meliputi hasil *accuracy* yang didapatkan, terdapat juga saran bagi penelitian berikutnya yang akan mengangkat topik yang sama yaitu analisis sentimen.