

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sektor Pariwisata merupakan salah satu dari banyak industri yang telah diuntungkan dari berkembangnya teknologi informasi dengan memberikan kemudahan akses informasi, kemudahan pelayanan bagi wisatawan, dan fitur-fitur lainnya yang dapat menguntungkan wisatawan [1]. Dalam sektor pariwisata, aplikasi *google maps* dengan fitur *google review* menjadi salah satu fitur penting dalam era big data untuk mengumpulkan jejak wisatawan yang berkunjung ke suatu tempat. Dengan memanfaatkan fitur lokasi ponsel yang terus menyala, *google review* akan mengirimkan *notification* otomatis pada *handphone* sehingga wisatawan dapat memberikan sebuah penilaian terhadap tempat yang baru saja dikunjungi pada aplikasi *google maps* [2].

Sumenep adalah salah satu kabupaten di Jawa Timur yang terletak di ujung timur pulau Madura. Sumenep mempunyai banyak destinasi wisata salah satunya wisata Taman Adipura Sumenep yang merupakan ikon dari Kota Sumenep. Tempatnya yang *instagramable* dan berdekatan dengan Masjid Jamik sebagai masjid tertua dikabupaten Sumenep dengan arsitektur uniknya, menjadikan daya tarik tersendiri bagi wisatawan terhadap Taman Adipura Sumenep. [3]. Banyaknya pengunjung yang datang, tidak sedikit yang memberikan ulasan atau *opini* salah satunya pada aplikasi *google Maps*. Ulasan ini menjadi hal yang sangat penting untuk diperhatikan, pasalnya pengunjung akan mencari informasi terlebih dahulu dari internet baik pada *media social* atau *google review* sebelum memutuskan untuk berkunjung ke sebuah tempat wisata [2]. Dengan ulasan tersebut dapat dilakukan sebuah analisis sentimen untuk melihat sentimen masyarakat serta menjadikan evaluasi sederhana terhadap wisata Taman Adipura Sumenep.

Analisis sentimen merupakan proses mengestrak dan mengelolah data tekstual secara otomatis dari teks bahasa yang di proses menggunakan metode komputasi untuk mendapatkan informasi dengan tujuan melihat kecenderungan penilaian opini terhadap sebuah objek, sentimen yang diberikan bisa cenderung

sentimen positif, negatif atau netral [4]. Salah satu metode yang dapat digunakan untuk proses pengklasifikasian analisis sentimen adalah *Random Forest*.

Pada penelitian terdahulu tentang analisis sentimen yang berjudul *Sentiment Analysis of Social Media Network Using Random Forest Algorithm*. Penelitian ini menggunakan metode *Random Forest* dan kemudian dibandingkan dengan *Support Vector Machine* untuk memastikan *Random Forest* memberikan hasil terbaik. Hasil dari penelitian ini metode *Random Forest* berhasil memberikan akurasi terbaik sebesar 97% dibandingkan metode *Support Vector Machine* dengan akurasi 92% [5]. Penelitian selanjutnya tentang Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma *Naive Bayes*, *Random Forest* Dan *Support Vector Machine*. Penelitian ini menghasilkan metode *Random Forest* memiliki nilai akurasi tertinggi dari kedua algoritma lainnya yaitu sebesar 97,16% serta menunjukkan adanya peningkatan akurasi dari penelitian sebelumnya sebesar 7,16% [6]. Kelebihan metode *Random Forest* adalah dapat mengelola data dengan jumlah yang besar, mampu mengklasifikasikan data dengan atribut yang tidak lengkap, tidak rentan terhadap *noise* dan *outlier*, serta lebih kecil untuk terjadinya *overfitting* [7]. Namun, metode *Random Forest* memiliki kelemahan ketika bekerja dengan jumlah fitur yang banyak [8][9].

Dalam proses klasifikasi banyaknya fitur yang digunakan dapat menyebabkan penurunan performa dan kompleksitas waktu pada proses *training*, sehingga dibutuhkan proses seleksi fitur untuk mereduksi atau mengurangi jumlah fitur yang besar dan diambil fitur yang tidak relevan atau bermanfaat. Beberapa metode seleksi fitur yang dapat diterapkan diantaranya yaitu *Gain Ratio*, *Information Gain*, *Chi Square* dan masih banyak yang lainnya. Terdapat beberapa penelitian tentang seleksi fitur seperti pada penelitian yang berjudul “Analisis Sentimen Maskapai Penerbangan Menggunakan Metode *Naïve Bayes* dan Seleksi Fitur *Information Gain*”. Metode *Information Gain* melakukan seleksi fitur dengan cara menghitung berapa banyak kemunculan dan ketidakmunculan informasi dari sebuah *term* yang berperan untuk membuat keputusan klasifikasi yang benar dalam kelas apapun. Setelah itu bobot *Information Gain* dari setiap *term* atau kata akan diranking berdasarkan bobot tertinggi sehingga menghasilkan fitur terbaik. Penelitian ini menghasilkan nilai rata-rata akurasi, *precision*, *recall*

yang lebih baik ketika menggunakan *Information Gain* sebesar 86% dibandingkan tanpa *Information Gain* sebesar 81% [10].

Berdasarkan dari latar belakang dan permasalahan tersebut, maka penulis mengajukan penelitian dengan judul “Analisis Sentimen Pada Ulasan Wisata Taman Adipura Sumenep Dengan Penerapan Metode *Random Forest* Dan Seleksi Fitur *Information Gain*”.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Random Forest mengalami penurunan performa saat bekerja dengan jumlah fitur yang banyak, sehingga dibutuhkan sebuah seleksi fitur untuk mengurangi fitur tersebut.

1.2.2 Metode Usulan

Penambahan seleksi fitur *Information Gain* pada metode *Random Forest* untuk klasifikasi ulasan objek wisata Taman Adipura Sumenep pada *Google maps*.

1.2.3 Pertanyaan Penelitian

1. Berapa nilai akurasi yang didapatkan metode *Random Forest* tanpa seleksi fitur *Information Gain* dengan distribusi data yang berbeda?
2. Berapa waktu dan akurasi terbaik yang didapatkan metode *Random Forest* dengan seleksi fitur *Information Gain* dan menggunakan *threshold* berapa yang paling optimal ?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Menerapkan seleksi fitur *Information Gain* untuk mengurangi jumlah fitur agar dapat meningkatkan performa *Random Forest* dalam mengklasifikasikan sentimen masyarakat terhadap wisata Taman Adipura Sumenep.

1.3.2 Manfaat Penelitian

Manfaat yang diharapkan dari klasifikasi sentimen pada penelitian ini adalah sebagai berikut:

1. Mengetahui pengaruh seleksi fitur *Information Gain* terhadap metode *Random Forest* dalam mengklasifikasikan sentimen masyarakat terhadap Taman Adipura Sumenep.
2. Partisipasi dalam penyediaan informasi ilmiah tentang *Random Forest* dan seleksi fitur *Information Gain* dalam klasifikasi teks.

1.4 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Teks yang digunakan berbahasa Indonesia.
2. Data yang digunakan pada penelitian ini diambil dari ulasan *google maps*.
3. Klasifikasi yang dilakukan dibagi menjadi 2 kategori yaitu: positif dan negatif.
4. Data yang digunakan berjumlah 1972 ulasan.
5. Data diambil dari tanggal 12 November 2022.
6. Bahasa pemrograman yang dipakai adalah Python.
7. Metode yang digunakan *Random Forest* dengan seleksi fitur *Information Gain*.

1.5 Sistematika Penulisan

Sistematika penulisan proposal skripsi ini terdiri dari beberapa bab, yaitu sebagai berikut:

1. BAB I PENDAHULUAN

Bab ini berisi latar belakang, perumusan masalah yang meliputi permasalahan, metode usulan, dan pertanyaan penelitian, kemudian selanjutnya terdapat tujuan dan manfaat penelitian, batasan-batasan penelitian dan sistematika penulisan.

2. BAB II KAJIAN PUSTAKA

Bab ini berisi penjelasan mengenai teori-teori pendukung dari berbagai sumber yang berhubungan dengan penelitian ini.

3. BAB III METODE USULAN

Bab ini berisi penjelasan tentang data penelitian, arsitektur sistem, dan perkiraan jadwal pengerjaan.