

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pantai Sembilan merupakan destinasi yang banyak dikunjungi oleh masyarakat Madura. Pantai tersebut terletak di Pulau Giligenting, Desa Bringsang, Kabupaten Sumenep. Oleh karena itu, masyarakat Madura, khususnya di Kabupaten Sumenep, memiliki potensi yang signifikan untuk memberikan dampak positif terhadap kehidupan masyarakat lokal. Dalam hal ini, sektor ekonomi dan pariwisata merupakan salah satu industri yang paling penting di Indonesia untuk dikembangkan[1]. Karena masyarakat memanfaatkan hal tersebut untuk memperoleh lapangan pekerjaan guna meningkatkan perekonomian, maka industri ini perlu diciptakan[2]. Oleh karena itu, masyarakat menciptakan lapangan pekerjaan dengan mendirikan berbagai usaha, seperti banana boat, kano, dan snorkeling. Dalam rangka mengunjungi Pantai Sembilan, beberapa individu memanfaatkan *Google maps* sebagai sarana untuk menentukan destinasi liburan yang ideal. Banyak program yang tersedia untuk pencarian destinasi online. Program berbasis internet *Google maps* adalah salah satu program yang dapat digunakan oleh bisnis secara gratis[3]. Mengingat bahwa sejumlah besar masyarakat Indonesia masih memanfaatkan *Google maps* untuk mencari lokasi tempat wisata, serta mayoritas pengguna memberikan ulasan pada platform tersebut. Dalam proses pengambilan data ulasan, diperlukan pula penerapan pembelajaran mesin (*machine learning*), yang dapat mempermudah dalam pengumpulan dan pengolahan data tersebut.

Machine learning adalah salah satu cabang dari kecerdasan buatan (*Artificial Intelligence*) dan ilmu komputer yang berfokus pada penerapan data serta algoritma untuk menirukan cara manusia dalam belajar, sehingga secara bertahap dapat meningkatkan tingkat akurasi[4]. Dalam *machine learning*, diperlukan pula data mining. Penggunaan data mining akan memfasilitasi proses ekstraksi informasi dari berita, sehingga kita dapat mengidentifikasi kata kunci yang berhubungan dengan istilah tertentu di berbagai kategori. Dalam tahap pengumpulan berita yang akan melalui proses data mining, diperlukan penerapan teknik web crawling[5].

Dengan banyaknya ulasan yang diberikan oleh para pengguna, sejumlah peneliti melakukan data mining atau crawling data.

Metode crawling data merupakan suatu teknik yang digunakan untuk memperoleh data dari suatu situs web secara otomatis dengan memanfaatkan program komputer. Teknik ini memungkinkan pengambilan data dari berbagai sumber secara efisien dan terstruktur, sehingga dapat dimanfaatkan untuk berbagai keperluan, seperti penelitian, analisis, dan pengembangan aplikasi[6]. Pengumpulan data melalui teknik crawling bertujuan untuk memperoleh informasi berbentuk teks dari ulasan yang ditulis oleh sejumlah wisatawan di *Google maps*[7]. Penggalan informasi dari internet, atau yang dikenal sebagai web mining, dapat menjadi salah satu solusi untuk mengumpulkan informasi mengenai wisata kuliner dengan memanfaatkan internet sebagai sumber data[8]. Setelah pengumpulan data yang dilakukan oleh peneliti, selanjutnya dilakukan analisis sentimen.

Analisis sentimen merupakan suatu teknik yang digunakan untuk mendeteksi opini terkait suatu subjek, seperti individu, organisasi, atau produk, dalam suatu kumpulan data[9]. Dalam analisis sentimen ini terdapat klasifikasi yang bertujuan untuk menetapkan kategori yang relevan. Sentimen dikategorikan ke dalam tiga kategori, yaitu positif, netral, dan negatif, sehingga pengguna dapat dengan mudah memilih dan membaca opini sesuai dengan preferensi yang diinginkan[10]. Tujuan dari analisis ini adalah untuk mengetahui pendapat yang diungkapkan dalam kalimat atau dokumen tersebut, apakah bersifat positif atau negatif[11].

Berdasarkan penelitian yang telah dilakukan sebelumnya, analisis sentimen ChatGPT terhadap Twitter yang diperoleh dari eksperimen menunjukkan bahwa ketika nilai K Tetangga = 5 pada KNN, ukuran penilaian yang Dihasilkan adalah tingkat akurasi sebesar 80,4% dan tingkat presisi. nilai presisi mencapai 80,7%, sedangkan nilai recall adalah 80,4%. Sebaliknya Naïve Bayes menghasilkan tingkat akurasi sebesar Skor akurasi mencapai 81,7%, dengan tingkat presisi sebesar 81,9%, serta nilai recall sebesar 81,7%[12]. Analisis sentimen pada data Twitter terkait dengan Piala Dunia menunjukkan hasil percobaan dengan memanfaatkan metode klasifikasi Naive Bayes. Mendapatkan akurasi 82%, presisi 83%, recall 90%, dan f1-score sebesar 90%. Dengan menggunakan metode support vector machine, diperoleh akurasi 85%, presisi 86%, serta f1-score sebesar 91%[13].

Dalam pelaksanaan analisis sentimen menggunakan *Naive Bayes Classifier*, diperlukan keseimbangan data. Salah satu metode yang dilakukan untuk mencapai keseimbangan tersebut adalah dengan menggunakan (Synthetic Minority Over-sampling Technique) *SMOTE*.

Dalam analisis sentimen, seringkali dijumpai isu ketidakseimbangan data. Oleh karena itu, dilakukan upaya penyeimbangan data untuk memastikan agar data tersebut berada dalam keadaan seimbang. *SMOTE* mampu melakukan duplikasi data secara sintesis, sehingga masalah perbedaan distribusi data dapat diatasi[14]. Teknik ini melakukan sintesis sampel baru dari kelas minoritas dengan tujuan untuk menyeimbangkan dataset melalui proses pengambilan sampel ulang dari sampel kelas minoritas[15]. Dalam proses penyeimbangan data, tidak hanya metode *SMOTE* yang digunakan, melainkan juga ADASYN sebagai metode perbandingan. ADASYN (Adaptive Synthetic) merupakan suatu metode pendekatan pengambilan sampel yang digunakan dalam pembelajaran dengan data yang tidak seimbang[16]. Pada kesempatan ini, saya menggunakan metode *SMOTE* karena telah banyak digunakan dalam berbagai penelitian sebelumnya.

Berdasarkan studi-studi sebelumnya yang telah menunjukkan keberhasilan metode Naive Bayes Classifier dalam menganalisis sentimen, serta pentingnya teknik *SMOTE* dalam menangani ketidakseimbangan data, penulis melakukan analisis sentimen terkait ulasan publik tentang wisata Pantai Sembilan di *Google maps* dalam penelitian ini. Dalam penelitian ini, diterapkan metode Naive Bayes Classifier dengan bantuan teknik *SMOTE* untuk menghasilkan model klasifikasi yang lebih seimbang dan tepat. Diharapkan hasil dari penelitian ini dapat memberikan wawasan yang lebih objektif mengenai pandangan wisatawan terhadap Pantai Sembilan, serta berfungsi sebagai acuan dalam mengembangkan strategi pengelolaan pariwisata di Kabupaten Sumenep.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian yaitu sebagai berikut:

1.2.1 Permasalahan

Permasalahan yang dihadapi dalam penelitian ini adalah volume data yang sangat besar, yang mengakibatkan kebutuhan waktu yang signifikan serta usaha yang melelahkan dalam proses pelabelan data. Selain itu, terdapat ketidaksesuaian

antara ulasan dan rating yang mengakibatkan pelabelan yang kurang akurat. Selain itu, terdapat ketidakseimbangan pada data. Selain itu, algoritma fitur yang digunakan dalam metode tersebut kurang relevan, sehingga mengakibatkan tingkat optimalitas yang rendah dalam pengklasifikasian sentimen. Dengan menerapkan Naive Bayes Classifier (NBC) dan SMOTE, analisis sentimen ini dapat dioptimalkan di objek wisata Pantai Sembilan.

1.2.2 Metode Usulan

Oleh karena itu, peneliti mengusulkan untuk melaksanakan pelabelan secara otomatis dengan memanfaatkan metode Python TextBlob, yang mampu melabelkan kalimat-kalimat ulasan. Selain itu, penelitian ini juga akan melakukan klasifikasi menggunakan metode *Naive Bayes Classifier*, yang telah terbukti memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan metode yang digunakan dalam penelitian sebelumnya.

1.2.3 Pertanyaan Penelitian

1. Bagaimana performa metode *Naive Bayes Classifier* terhadap analisis sentimen ulasan Pantai Sembilan?
2. Bagaimana pengaruh pengguna *SMOTE* dan tidak terhadap performa *Naive Bayes Classifier*?

1.3 Tujuan Dan Manfaat

1.3.1 Tujuan Penelitian

Setelah merumuskan pertanyaan penelitian, tujuan dari penelitian ini adalah untuk mengetahui nilai performa dari metode *Naive Bayes Classifier* (NBC). Selain itu, penelitian ini juga bertujuan untuk mengevaluasi apakah metode *Naive Bayes Classifier* (NBC) dapat mencapai nilai akurasi yang tinggi dengan penerapan teknik *SMOTE*.

1.3.2 Manfaat penelitian

Menyusun nilai performa dari metode *Naive Bayes Classifier* (NBC) dapat dilakukan melalui analisis sentimen pada ulasan *Google Maps*. Dengan memanfaatkan metode *Naive Bayes Classifier* (NBC) dan *SMOTE* dalam analisis ini, dapat dievaluasi performa dari metode tersebut dalam konteks analisis sentimen. Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan terhadap pengembangan pengetahuan lebih lanjut di bidang ini.

1.4 Batasan Masalah

1. Data yang digunakan untuk penelitian ini diperoleh dari *Google maps* yang merupakan data ulasan komentar terbaru dari tempat wisata Pantai Sembilan. yang berjumlah 1101 dokumen ulasan
2. Pengambilan data dilakukan dengan *Scrapping* data pada *Google maps* sebanyak 1101 ulasan mulai dari 21 Februari 2019 – 8 Maret 2025
<https://www.google.com/maps> (<https://drive.google.com/file/d/1AXy05G79JFDMu0uuTD4cb11HlcGhJAp7/view?usp=sharing>)
3. Data teks yang digunakan berupa data yang menggunakan bahasa Indonesia
4. Pembagian kategori atau label dibagi menjadi dua yaitu kategori positif dan negatif
5. Memanfaatkan Split dataset dengan rasio perbandingan 80:20.

1.5 Sistematika Penelitian

Sistematika penulisan pada penelitian ini bisa dilihat dibawah ini:

BAB I PENDAHULUAN

Bab ini berisi penjelasan mengenai tinjauan pustaka dan penelitian terkait yang digunakan sebagai acuan dalam penelitian ini.

BAB II KAJIAN PUSTAKA

Dalam bab ini, dilakukan analisis terhadap literatur dan teori yang berkaitan dengan penelitian ini, yang diperoleh dari berbagai sumber. Beberapa teori yang diterapkan dalam penelitian ini mencakup penjelasan mengenai Pantai Sembilan, *Preprocessing*, *SMOTE*, Klasifikasi Naive Bayes, *Confison Matrix*, serta studi yang berkaitan dengan Klasifikasi Naive Bayes (NBC) dalam analisis sentimen ulasan *Google maps*.

BAB III METODE PENELITIAN

Dalam bab ini dirujuk langkah-langkah penelitian agar penelitian yang dilakukan mencapai tujuan yang maksimal. Tahapan penelitian meliputi tahap penelitian, perancangan sistem, pengumpulan data, tahap *Preprocessing*, tahap klasifikasi, tahap evaluasi dan tahap skenario pengujian.

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini berisi penjelasan dari penerapan dan hasil analisis sentimen menggunakan metode *Naive Bayes* dan *SMOTE*.

BAB V PENUTUP

Dalam bab ini berisi rangkuman kesimpulan yang diperoleh dari hasil penelitian. Selain itu, pada bab ini terdapat saran dari penulis untuk penelitian selanjutnya.