

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pergeseran tren pariwisata ke arah digital menjadikan ulasan daring atau *Electronic Word of Mouth* (e-WOM) di platform Google Maps sebagai referensi utama wisatawan sebelum berkunjung. Fenomena ini sejalan dengan temuan literatur terdahulu yang membuktikan bahwa ulasan daring berpengaruh positif dan paling signifikan terhadap minat berkunjung kembali pada destinasi wisata dibandingkan faktor lainnya [1]. Bagi objek ekowisata yang mengandalkan keindahan alam dan kualitas layanan, reputasi digital menjadi aset vital yang perlu dipantau [2]. Sentimen negatif yang tidak tertangani dengan cepat berpotensi menurunkan citra destinasi dan berdampak pada stabilitas pendapatan objek wisata tersebut [3].

Penelitian ini mengambil studi kasus pada Ekowisata Sungai Mudal yang dikelola dengan model berbasis komunitas (*Community Based Tourism*). Berdasarkan data sekunder dari penelitian Ningrum dan Nugroho mengenai tata kelola kolaboratif, tercatat adanya fluktuasi kinerja kunjungan dan pendapatan yang dinamis pasca pandemi [4]. Ketersediaan data historis pendapatan dari literatur tersebut memungkinkan validasi korelasi dilakukan secara objektif tanpa memerlukan data primer internal yang bersifat rahasia [4]. Secara empiris, relevansi antara sentimen digital dan kinerja keuangan ini juga telah divalidasi pada skala makro oleh Islam et al. [5], yang membuktikan secara statistik bahwa sentimen di media sosial berperan signifikan dalam memoderasi dan meningkatkan kinerja finansial entitas bisnis. Namun hingga saat ini belum ada pengukuran kuantitatif yang mengadopsi logika korelasi tersebut pada skala ekonomi mikro pariwisata untuk menghubungkan tren sentimen ulasan daring dengan kinerja finansial secara spesifik di lokasi ini.

Permasalahan teknis utama dalam analisis ulasan adalah volume data yang besar dan tidak terstruktur, sehingga sulit bagi pengelola untuk memetakannya

secara manual [6]. Selain itu data ulasan pariwisata memiliki karakteristik alami yang tidak seimbang (*imbalanced data*), di mana jumlah ulasan positif cenderung jauh lebih banyak daripada ulasan negatif (keluhan) [7]. Jika diolah menggunakan algoritma klasifikasi standar, kondisi ini seringkali menyebabkan model mengalami bias mayoritas yaitu memiliki akurasi tinggi namun gagal mendeteksi keluhan pelanggan yang sebenarnya krusial [8].

Untuk mengatasi tantangan pengelolaan data teks yang kompleks tersebut, penelitian ini mengajukan penerapan analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM) yang dioptimasi secara hibrida [9]. Teknik *Adaptive Synthetic Sampling* (ADASYN) diterapkan untuk menyeimbangkan distribusi data latih dengan memfokuskan pembentukan data sintesis pada area perbatasan keputusan yang sulit dipelajari, sehingga sensitivitas model terhadap kelas minoritas dapat ditingkatkan [10]. Untuk memaksimalkan performa klasifikasi, integrasi *Particle Swarm Optimization* (PSO) digunakan untuk mencari parameter model terbaik secara otomatis [11]. Kombinasi metode ini diharapkan mampu menghasilkan prediksi sentimen yang akurat dan berimbang.

Penelitian ini tidak hanya bertujuan untuk mengukur performa teknis model, tetapi juga membuktikan validitas hubungan antara sentimen digital dan kinerja pendapatan riil melalui uji korelasi statistik. Guna memberikan manfaat praktis yang berkelanjutan, model klasifikasi yang telah dioptimasi selanjutnya diimplementasikan ke dalam sistem aplikasi berbasis web (*deployment*). Kehadiran sistem ini diharapkan dapat menjadi alat bantu strategis bagi pengelola Ekowisata dalam memantau kepuasan pengunjung serta merumuskan kebijakan berbasis data di masa depan.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana distribusi tren sentimen pengunjung Ekowisata Sungai Mudal berdasarkan data Google Maps periode 2019 hingga 2024?

2. Berapa tingkat akurasi model SVM yang dioptimasi menggunakan teknik ADASYN dan PSO dalam menangani ketimpangan data?
3. Seberapa kuat hubungan korelasi antara persentase sentimen positif ulasan daring terhadap rata-rata pendapatan bulanan objek wisata?

1.3 Batasan Masalah

Agar penelitian lebih terarah dan fokus, peneliti menetapkan batasan masalah sebagai berikut:

1. Objek penelitian berfokus pada ulasan Google Maps "Ekowisata Sungai Mudal" dan data sekunder pendapatan yang bersumber dari literatur penelitian terdahulu [4].
2. Rentang waktu data ulasan dan pendapatan yang disinkronisasi adalah dari Januari 2019 hingga September 2024.
3. Pelabelan kelas sentimen pada dataset ulasan dilakukan secara otomatis berdasarkan atribut rating bintang Google Maps, tanpa melalui proses anotasi manual oleh pakar linguistik (*human annotator*). Dengan demikian, penelitian ini membatasi diri dan mentoleransi adanya label *noise* yang mungkin timbul akibat ketidaksesuaian antara teks ulasan dengan rating bintang yang diberikan oleh pengunjung.
4. Pemodelan *Machine Learning* dalam penelitian ini difokuskan pada analisis sentimen secara retrospektif. Oleh karena itu, pembagian dataset menjadi data latih (*training set*) dan data uji (*testing set*) dilakukan secara acak bertingkat (*stratified random split*) berdasarkan proporsi kelas sentimen, dan tidak dirancang menggunakan skenario pembagian berbasis urutan waktu yang ketat (*time-aware split* atau *time-series forecasting*).
5. Pengembangan sistem dilakukan menggunakan bahasa pemrograman Python dengan pustaka Sastrawi untuk *preprocessing* bahasa Indonesia dan diimplementasikan (*deployment*) ke dalam bentuk aplikasi web berbasis *framework* Streamlit.
6. Ekstraksi fitur teks dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF).

7. Klasifikasi sentimen dilakukan menggunakan algoritma SVM *Kernel Linear* yang dikombinasikan dengan teknik ADASYN untuk penanganan data tidak seimbang (*imbalanced data*).
8. Optimasi parameter model (*Hyperparameter Tuning*) dilakukan menggunakan algoritma PSO.
9. Analisis hubungan variabel sentimen dan pendapatan dilakukan menggunakan metode statistik Korelasi *Pearson Product Moment*.

1.4 Tujuan Penelitian

Mengacu pada rumusan masalah yang telah ditetapkan, tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Menganalisis distribusi tren sentimen (positif dan non-positif) dari ulasan wisatawan dalam kurun waktu 5 tahun (2019–2024) untuk memetakan dinamika kepuasan pengunjung.
2. Mengukur performa dan efektivitas model SVM yang dioptimasi dengan teknik ADASYN dan PSO dalam menangani ketimpangan data ulasan berbahasa Indonesia.
3. Membuktikan secara statistik keberadaan hubungan linier antara reputasi digital (persentase sentimen positif) dengan kinerja pendapatan riil pariwisata.

1.5 Manfaat Penelitian

Manfaat Teoretis Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan ilmu informatika, khususnya implementasi *Natural Language Processing* (NLP) pada bidang pariwisata. Secara spesifik, penelitian ini memperkaya literatur mengenai efektivitas pendekatan hibrida yang menggabungkan algoritma klasifikasi, teknik resampling ADASYN, dan optimasi PSO dalam menangani masalah *imbalanced data* pada ulasan berbahasa Indonesia. Selain itu, penelitian ini menawarkan kerangka metodologi baru dalam memvalidasi hubungan statistik antara data teks tidak terstruktur dengan data kinerja keuangan riil.

Manfaat Praktis Hasil penelitian ini memberikan wawasan berbasis data

(*data-driven insights*) yang aplikatif bagi pengelola objek wisata (BUMDes) dan pemangku kepentingan terkait. Informasi mengenai tren sentimen dan korelasi nyata terhadap pendapatan dapat digunakan sebagai acuan pengambilan keputusan strategis untuk meningkatkan kualitas layanan dan efektivitas pemasaran digital. Lebih lanjut, ketersediaan sistem aplikasi dasbor yang dibangun dalam penelitian ini memungkinkan pengelola untuk memantau reputasi digital secara mandiri dan berkelanjutan tanpa memerlukan keahlian teknis pemrograman.

1.6 Sistematika Penulisan

Penyusunan skripsi ini dibagi menjadi lima bab dengan sistematika sebagai berikut:

BAB I PENDAHULUAN Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan yang memberikan gambaran umum mengenai arah penelitian

BAB II TINJAUAN PUSTAKA Bab ini memuat tinjauan pustaka dari penelitian terdahulu yang relevan serta landasan teori yang digunakan, meliputi *Text Mining*, Analisis Sentimen, SVM, ADASYN, PSO, dan teori korelasi statistik.

BAB III METODE PENELITIAN Bab ini menjelaskan tahapan penelitian secara rinci, mulai dari pengumpulan data sekunder, alur *preprocessing* data teks, skenario pengujian model, metode analisis korelasi, hingga perancangan antarmuka pengguna (*user interface*).

BAB IV HASIL DAN PEMBAHASAN Bab ini memaparkan hasil implementasi sistem, evaluasi performa model klasifikasi, visualisasi tren sentimen, analisis statistik mengenai hubungan antara sentimen ulasan dengan data pendapatan pariwisata, serta hasil *deployment* sistem ke dalam bentuk *dashboard* interaktif berbasis web.

BAB V PENUTUP Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian serta saran-saran yang konstruktif bagi pengembangan penelitian selanjutnya.