

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pariwisata adalah salah satu hal penting dalam pembangunan daerah karena mampu meningkatkan pendapatan masyarakat serta mendukung perekonomian lokal. Salah satu destinasi wisata alam unggul di Jawa timur Kabupaten Magetan adalah Telaga Sarangan. Telaga ini dikenal dengan panorama alamnya yang indah serta akses yang cukup dijangkau wisatawan dari berbagai daerah maupun Warga Negara Asing.

Kabupaten Magetan memiliki potensi besar dengan pariwisata telaga Sarangan sebagai ikon utamanya. Kontribusi telaga Sarangan terhadap Pendapatan Asli Daerah (PAD) meningkat setiap tahun selama 2016-2019, meskipun tahun 2020 mengalami penurunan akibat Virus Covid-19. Meski ada pembatasan aktivitas telaga Sarangan tetap mampu memberikan signifikan terhadap PAD. [1]

Namun Telaga Sarangan dapat meningkatkan kontribusi pada PAD di Tahun-Tahun yang akan datang. Tetapi terdapat permasalahan Ulasan wisatawan di berbagai platform menunjukkan beragam opini mulai dari keindahan alam, fasilitas, harga, hingga pelayanan. Analisis sentimen terhadap ulasan wisatawan tersebut penting dalam pengelola meningkatkan kualitas layanan. [1] Ketidakseimbangan kelas pada data sentimen menjadi persoalan penting karena model klasifikasi cenderung lebih mudah mengenali kelas mayoritas dibandingkan kelas minoritas. Pada dataset ulasan Telaga Sarangan, kelas positif memiliki jumlah paling dominan, sedangkan kelas netral dan negatif memiliki jumlah lebih sedikit. Kondisi ini dapat menyebabkan nilai akurasi terlihat tinggi, tetapi kemampuan model dalam mengenali kelas minoritas menjadi rendah. Oleh karena itu, penelitian ini menerapkan teknik resampling SMOTE dan ADASYN untuk meningkatkan representasi kelas minoritas pada data latih.

SMOTE dipilih karena mampu membentuk sampel sintetis baru berdasarkan kedekatan antar data minoritas menggunakan pendekatan k-nearest neighbors. Sementara itu, ADASYN digunakan karena memiliki mekanisme adaptif dalam membangkitkan sampel sintetis, yaitu lebih banyak menghasilkan data baru pada sampel minoritas yang sulit diklasifikasikan. Perbandingan kedua metode tersebut penting dilakukan untuk mengetahui teknik resampling yang lebih efektif dalam meningkatkan performa Random Forest pada klasifikasi sentimen ulasan wisata Telaga Sarangan.

1.2 Rumusan masalah

1.2.1 Permasalahan

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, beberapa permasalahan yang dapat diidentifikasi dalam penelitian ini adalah opini wisatawan terkait keindahan alam, fasilitas, harga dan pelayanan. Setelah itu penelitian ini juga ingin tahu tentang ketidak seimbangan jumlah ulasan positif, negatif ataupun netral untuk mempengaruhi akurasi Analisis sentimen bertujuan untuk menentukan teknik resampling yang efektif, Serta menilai kemampuan metode *Random Forest* dalam melakukan klasifikasi.

1.2.2 Metode Usulan

Penelitian ini mengusulkan penerapan metode *Random Forest* (RF) agar mendapatkan kinerja lebih baik dan memberikan hasil yang kuat dalam klasifikasi. Berdasarkan permasalahan yang telah diidentifikasi, penelitian ini berupaya menjawab beberapa pertanyaan sebagai berikut:

1. Bagaimana kinerja metode *Random Forest* tanpa teknik resampling dibandingkan dengan menerapkan teknik resampling data dalam klasifikasi sentimen wisata telaga sarangan ?
2. Bagaimana pengaruh penerapan proses hyperparameter tuning terhadap peningkatan akurasi model *Random Forest* dalam klasifikasi sentimen wisata Telaga Sarangan?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

1. Mengetahui bagaimana kinerja *random forest* dalam mengklasifikasikan data sentimen wisata telaga sarangan.
2. Menerapkan beberapa *teknik resampling* pada klasifikasi sentimen wisata telaga sarangan menggunakan metode *Random Forest*
3. Membandingkan hasil klasifikasi *random forest* setelah menggunakan teknik *resampling*.

1.3.2 Manfaat Penelitian

1. Mengetahui bagaimana kinerja *Random Forest* dalam mengklasifikasikan data sentimen wisata telaga sarangan
2. Menerapkan beberapa *teknik resampling* pada klasifikasi sentimen wisata telaga sarangan menggunakan metode *Random Forest*
3. Membandingkan hasil klasifikasi *Random Forest* setelah menggunakan teknik *resampling*.

1.4 Batasan Masalah

Dalam penelitian dilakukan mempunyai batasan-batasan masalah yang harus diperhatikan, yaitu:

1. Analisis sentimen di fokuskan pada ulasan positif, negatif, dan netral
2. Metode klasifikasi yang di gunakan *Random Forest* dengan menggunakan teknik *resampling* data
3. Membandingkan hasil klasifikasi *Random Forest* berdasarkan penggunaan beberapa teknik *resampling* data

1.5 Sistematika Laporan

Sistematika laporan adalah struktur dari penulisan laporan yang digunakan untuk pembuatan penelitian, seperti dibawah ini:

BAB I PENDAHULUAN

Pada penelitian bab ini berisi tentang latar belakang, perumusan masalah yang terdiri dari permasalahan, metode usulan, dan pertanyaan penelitian, tujuan dan manfaat penelitian, batasan masalah, dan penyusunan skripsi yang sistematis.

BAB II TINJAUAN PUSTAKA

Pada penelitian bab ini menjelaskan berbagai konsep dan teori yang menjadi landasan penelitian serta memberikan acuan dalam merumuskan solusi terhadap permasalahan yang diteliti.

BAB III METODE USULAN

Pada penelitian ini menggunakan data ulasan wisata Telaga Sarangan yang dikumpulkan dari google maps. Tahapan penelitian meliputi pengumpulan dan pembersihan data ulasan, preprocessing teks, penerapan teknik resampling, serta klasifikasi menggunakan metode *Random Forest*.

BAB IV METODE USULAN HASIL DAN PEMBAHASAN

Menampilkan hasil pengumpulan data, pelabelan, hasil *preprocessing*, TF-IDF, metode random forest, hasil evaluasi, serta perbandingan teknik resampling.

BAB V PENUTUP

Berisi kesimpulan yang menjawab tujuan dan pertanyaan penelitian, serta saran untuk penelitian selanjutnya