

ABSTRAK

Pariwisata merupakan sektor strategis yang berkontribusi signifikan terhadap pertumbuhan ekonomi Sumenep, Madura. Namun, pengelolaan ulasan pariwisata secara manual menghadapi tantangan dalam menangani volume data yang besar dan beragam. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pariwisata menggunakan algoritma Random Forest dengan membandingkan dua metode seleksi fitur, yaitu Mutual Information (MI) dan Chi-Square (Cs). Selain itu, penelitian ini juga menerapkan Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan data. Fokus penelitian mencakup enam destinasi wisata unggulan di Sumenep, yaitu Pantai Sembilan, Museum Kraton, Bukit Tinggi, Pantai Lombang, Goa Soekarno, dan Gili Labak. Algoritma Random Forest dipilih karena kemampuannya dalam menangani data besar dan mengurangi risiko overfitting. Seleksi fitur dilakukan untuk mengidentifikasi variabel yang paling relevan, sementara SMOTE diterapkan untuk meningkatkan akurasi model pada data yang tidak seimbang. Hasil penelitian menunjukkan bahwa kombinasi Random Forest dengan Mutual Information menghasilkan akurasi 84,39%, precision 80,00%, recall 84,00%, dan F1-Score 81,00%, yang merupakan performa terbaik dibandingkan dengan metode lainnya. Sebaliknya, penerapan SMOTE cenderung menurunkan akurasi model (74,13%), meskipun precision meningkat menjadi 81,00%. Hasil ini memberikan wawasan penting bagi pemangku kepentingan pariwisata di Sumenep dalam memahami sentimen pengunjung. Dengan analisis sentimen berbasis machine learning ini, diharapkan dapat mendukung pengambilan keputusan yang lebih efektif dalam pengelolaan dan pengembangan sektor pariwisata di Sumenep.

Kata Kunci : Pariwisata Sumenep, Analisis Sentimen, *Big Data*, *Random Forest*, *Mutual Information*, *Mutual Information*, *Oversampling SMOTE*