

Abstrak (Indonesia)

Penelitian ini membandingkan dua teknik resampling, yaitu SMOTE dan ADASYN, untuk mengatasi masalah data sentimen yang tidak seimbang pada ulasan wisata Telaga Sarangan. Data ulasan diambil dari Google Maps dan dibagi menjadi tiga kategori sentimen, yaitu positif, negatif, dan netral. Masalah utama dalam penelitian ini adalah jumlah ulasan positif jauh lebih banyak dibandingkan ulasan netral dan negatif. Jika kondisi ini dibiarkan, model cenderung lebih mudah mengenali kelas positif, tetapi kurang baik dalam mengenali kelas netral dan negatif. Tahapan penelitian dilakukan mulai dari pengumpulan data, pelabelan sentimen, preprocessing teks, pembagian data latih dan data uji, ekstraksi fitur menggunakan TF-IDF, penerapan SMOTE dan ADASYN pada data latih, hingga klasifikasi menggunakan algoritma Random Forest. Hasil pengujian menunjukkan bahwa metode SMOTE memperoleh akurasi sebesar 80%, sedangkan ADASYN memperoleh akurasi sebesar 78%. Artinya, jika dilihat dari akurasi saja, SMOTE sedikit lebih unggul dibandingkan ADASYN. Namun, hasil evaluasi tidak cukup hanya dilihat dari akurasi. Pada penelitian ini, ADASYN menghasilkan nilai macro F1-score sebesar 0,62, sedangkan SMOTE memperoleh macro F1-score sebesar 0,58. Nilai ini menunjukkan bahwa ADASYN lebih baik dalam mengenali semua kelas sentimen secara lebih seimbang, terutama kelas minoritas seperti netral dan negatif. Dengan demikian, SMOTE lebih unggul dari sisi akurasi, tetapi ADASYN lebih baik jika tujuan penelitian adalah menghasilkan model yang lebih adil dalam mengenali setiap kelas sentimen.

Kata Kunci: Analisis Sentimen, ADASYN, Random Forest, SMOTE, TF-IDF, Telaga Sarangan.

Abstrak (English)

This study compares two resampling techniques, namely SMOTE and ADASYN, to address the problem of imbalanced sentiment data in tourist reviews of Telaga Sarangan. The review data were collected from Google Maps and divided into three sentiment categories, namely positive, negative, and neutral. The main problem in this study is that the number of positive reviews is much higher than the number of neutral and negative reviews. If this condition is left untreated, the model tends to recognize the positive class more easily but performs less well in recognizing the neutral and negative classes. The research stages were carried out starting from data collection, sentiment labeling, text preprocessing, splitting the data into training data and testing data, feature extraction using TF-IDF, applying SMOTE and ADASYN to the training data, and classification using the Random Forest algorithm. The test results show that the SMOTE method achieved an accuracy of 80%, while ADASYN achieved an accuracy of 78%. This means that, when viewed only from accuracy, SMOTE is slightly superior to ADASYN. However, the evaluation results are not sufficient if viewed only from accuracy. In this study, ADASYN produced a macro F1-score of 0.62, while SMOTE obtained a macro F1-score of 0.58. This value indicates that ADASYN is better at recognizing all sentiment classes in a more balanced manner, especially minority classes such as neutral and negative. Therefore, SMOTE is superior in terms of accuracy, but ADASYN is better if the purpose of the study is to produce a model that is fairer in recognizing each sentiment class.

Keywords: *Sentiment Analysis, ADASYN, Random Forest, SMOTE, TF-IDF, Telaga Sarangan.*