

ABSTRAK

Penelitian terdahulu dalam klasifikasi sentimen terbatas pada skala dataset yang kecil dan kurang menyertakan perbandingan komprehensif antara berbagai metode *oversampling*. Ketidakseimbangan data (*Imbalanced data*) merupakan tantangan dalam analisis sentimen, dominasi kelas mayoritas menyebabkan model klasifikasi cenderung bias dan mengabaikan kelas minoritas. Penelitian ini bertujuan untuk meningkatkan algoritma klasifikasi dengan menerapkan tiga metode *oversampling*, yaitu *Synthetic Minority Oversampling Technique* (SMOTE), *Borderline-SMOTE*, dan *Random Oversampling*. Dataset yang digunakan terdiri dari 10.000 ulasan pengguna Tokopedia yang dikumpulkan melalui web scraping. Tiga algoritma klasifikasi *Random Forest*, *Naïve Bayes*, dan *Support Vector Machine* (SVM) digunakan dalam proses pelatihan model. Evaluasi kinerja dilakukan menggunakan metrik *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *Area Under the Curve* (AUC). Penelitian ini merupakan studi pertama yang melakukan komparasi tiga metode *oversampling* pada dataset besar (10.000 ulasan) dengan tiga algoritma berbeda dalam konteks *e-commerce* Indonesia. Hasil penelitian menunjukkan bahwa kombinasi SMOTE dan algoritma *Naïve Bayes* memberikan hasil yang stabil, dengan *Accuracy* sebesar 86,71%, *Precision* 80,43%, *Recall* 70,38%, *F1-Score* 75,07%, dan AUC 90,62%. Temuan ini membuktikan bahwa metode SMOTE secara efektif mampu meningkatkan kinerja klasifikasi.

Kata kunci: *Analisis Sentimen*, *Imbalanced Data*, *Oversampling*, *Random Forest*, *Naïve Bayes*, *Support Vector Machine*.