

ABSTRAK

Isu korupsi Pertamina yang ramai diperbincangkan di media sosial menunjukkan pentingnya dilakukan analisis sentimen guna memahami persepsi publik. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat ke dalam kategori positif dan negatif menggunakan metode *Support Vector Machine* (SVM). Data diperoleh melalui proses *crawling* dari media sosial X dan dilabeli secara manual. Ditemukan adanya ketidakseimbangan kelas, di mana ulasan negatif lebih dominan dibandingkan ulasan positif. Untuk mengatasi hal tersebut, dilakukan perbandingan dua teknik penyeimbangan data, yaitu *text augmentation* dan SMOTE. Proses pra-pemrosesan mencakup *cleaning*, *case folding*, *tokenization*, *normalization*, *stopword removal*, dan *stemming*. Fitur teks diekstraksi menggunakan metode TF-IDF. Pengujian dilakukan dalam dua skenario. Skenario pertama menggunakan *text augmentation* dengan optimasi parameter *grid search*, menghasilkan akurasi tertinggi 99,44% menggunakan kernel Linear dan RBF. Skenario kedua menggunakan SMOTE menghasilkan akurasi 97,54% dengan kernel Linear. Evaluasi metrik pada skenario pertama menunjukkan nilai presisi, *recall*, dan F1-Score masing-masing 99%. Pada skenario kedua, diperoleh nilai presisi 98%, *recall* 97%, dan F1-Score 97%. Hasil penelitian menunjukkan bahwa *text augmentation* memberikan performa klasifikasi yang lebih optimal dibandingkan dengan SMOTE.

Kata Kunci : Analisis Sentimen, *Grid Search*, SMOTE, *Support Vector Machine*, *Text Augmentation*, TF-IDF.

ABSTRACT

The Pertamina corruption issue that is widely discussed on social media shows the importance of sentiment analysis to understand public perception. This study aims to classify public sentiment into positive and negative categories using the Support Vector Machine (SVM) method. Data was obtained through a crawling process from social media X and labeled manually. Class imbalance was found, where negative reviews were more dominant than positive reviews. To overcome this, a comparison of two data balancing techniques was conducted, namely text augmentation and SMOTE. The pre-processing process included cleaning, case folding, tokenization, normalization, stopword removal, and stemming. Text features were extracted using the TF-IDF method. Testing was carried out in two scenarios. The first scenario used text augmentation with grid search parameter optimization, resulting in the highest accuracy of 99.44% using Linear and RBF kernels. The second scenario used SMOTE resulting in an accuracy of 97.54% with a Linear kernel. Evaluation metrics in the first scenario showed precision, recall, and F1-Score values of 99% each. In the second scenario, precision values of 98%, recall of 97%, and F1-Score of 97% were obtained. The results showed that text augmentation provided more optimal classification performance compared to SMOTE.

Keywords: *Sentiment Analysis, Grid Search, SMOTE, Support Vector Machine, Text Augmentation, TF-IDF.*