Abstract

Social media is online media where users can participate, share, and create content such as forums, blogs, social networks, wikis and virtual worlds. One of the social media with a widely used communication platform is Twitter, with various "tweets" with different contents being thrown every day, such as topics related to politics, social, government, services, etc., with opinions in terms of positive, neutral, and negative. The "tweet" sentence in Twitter has a limit of 280 characters, therefore causing problems such as vocabulary mismatches. Therefore, this research applies Word2Vec to overcome these problems and the Support Vector Machine (SVM) classification method. This research also uses TF-IDF feature extraction to weight the vocabulary. Synthetic Minority Oversampling Technique or SMOTE is used to overcome imbalanced data, and with the comparison of SMOTE Tomek Links, and SMOTE ENN extensions. The results of this study obtained the highest F1-score performance results using Support Vector Machine classification and SMOTE + ENN data imbalance techniques with TF-IDF feature extraction, feature selection, and Word2Vec feature expansion on twitter + news corpus in the signal aspect with an increase of 28.69% with F1-score 98.58%, and in the service aspect with an increase of 27.79% with F1-score 95.44%.

Keywords: Aspect-Based Sentiment Analysis, Support Vector Machine, Word2vec, Twitter, Telkomsel