Abstract
Data imbalance is a condition between label classes that has data imbalance between classes. This condition causes the classification method to ignore classes that have a small number of samples, thus performing poorly. To deal with the problem of data imbalance, oversampling techniques are needed, one of which is SMOTE. SMOTE works by replicating new sample data on minority class data against majority class. In this study, the data that will be used is Twitter data for sentiment analysis of fuel price increases with 17,266 data. The classification methods used are CNN and SVM because they have accurate performance in classifying text data. In this study, K-fold cross validation was used to validate data and Confusion Matrix as a calculation evaluation to provide detailed model information. The results of this study are with SMOTE of Accuracy 78.54%, precision 79.09%, recall 78.80%, and F1-score 78.94% compared to no SMOTE Accuracy 77.65%, precision 74.14%, recall 80.97%, and F1-score 77.85% on SVM models. While the CNN model with SMOTE accuracy 94.03%, precision 93.14%, recall 93.63%, and F1-score 93.85% compared to no SMOTE accuracy of 93.88%, precision of 93.59%, recall of 94.08%, and F1-score of 93.84%. So it was concluded that SMOTE has a good influence on accuracy, precision, and F1-score. However, there is a decrease in recall results.

Keywords : Imbalance Data, Sentiment Analysis, SVM, CNN, SMOTE