

Abstrak

Imbalance Data merupakan kondisi diantara kelas label memiliki ketidakseimbangan data antar kelas. Kondisi ini menyebabkan metode klasifikasi mengabaikan kelas yang memiliki jumlah sampel sedikit sehingga memberi performa kurang baik. Untuk menangani masalah *imbalance Data* maka diperlukan teknik *oversampling*, salah satu nya SMOTE. SMOTE bekerja dengan cara mereplikasi data sampel baru pada data kelas minoritas terhadap kelas mayoritas. Pada penelitian ini, data yang akan digunakan berupa data *Twitter* terhadap analisis sentimen kenaikan harga BBM dengan 17.266 data. Metode klasifikasi yang digunakan adalah CNN dan SVM karena memiliki performa yang akurat dalam mengklasifikasi data teks. Dalam penelitian ini digunakan *K-fold cross validation* untuk mem validasi data dan *Confusion Matrix* sebagai evaluasi perhitungan untuk memebrikan informasi model secara detail. Hasil Penelitian ini adalah dengan SMOTE sebesar *Accuracy* 78.54% , *precision* 79.09%, *recall* 78.80%, dan *F1-score* 78.94%. dibandingkan tanpa SMOTE *Accuracy* 77.65% , *precision* 74.14%, *recall* 80.97%, dan *F1-score* 77.85% pada model SVM. Sedangkan model CNN dengan SMOTE *accuracy* 94.03% , *precision* 93.14%, *recall* 93.63%, dan *F1-score* 93.85%. dibandingkan tanpa SMOTE *accuracy* 93.88% , *precision* 93.59%, *recall* 94.08%, dan *F1-score* 93.84%.Maka disimpulkan SMOTE memberi pengaruh baik pada nilai *accuracy*, *precision*, dan *F1-score*. Namun mengalami penurunan pada hasil *recall*.

Kata kunci : Imbalance Data, Analisis Sentimen, SVM, CNN, SMOTE