

ABSTRAK

Pemberlakuan kebijakan pembelajaran jarak jauh (PJJ) pada masa pandemi covid-19 memperoleh banyak tanggapan, kritik dan saran yang berisi cuitan positif dan negatif dari masyarakat melalui *twitter*. Cuitan tersebut sulit dipilah untuk mendapatkan kata dari sentimen positif dan sentimen negatif karena kata yang digunakan beragam. Mengatasi hal itu analisis sentimen dilakukan untuk mengekstrak informasi terkait kebijakan PJJ. Untuk mendapatkan hasil evaluasi terbaik menggunakan perbandingan dengan 4 skenario. Tahap penelitian dimulai dengan pengambilan data *twitter*, *cleaning data*, pelabelan positif dan negatif, *preprocessing* data diantaranya *tokenize*, *casefolding*, *stemming* dan *stopword removal*, TF-IDF untuk pembobotan kata dan *SMOTE* menangani ketidakseimbangan data, klasifikasi menggunakan algoritma *Naïve Bayes* dan *Particle Swarm Optimization* (PSO) sebagai seleksi fitur. Pengujian *Naïve Bayes* yang dikombinasikan dan PSO dengan 4 skenario menghasilkan akurasi yang cukup baik, dengan nilai rata-rata akurasi 97.29%, *recall* 94.72%, dan *precision* 99.84%. Hasil akurasi tertinggi pada rasio 60:40 dengan algoritma *Naïve Bayes* dan PSO sebesar 97.88%. Pengaruh PSO terhadap performa model cukup baik terlihat pada akurasi yang meningkat.

Kata Kunci: Analisis Sentimen, *CRISP-DM*, *Naïve Bayes*, *Particle Swarm Optimization*, PJJ, *SMOTE*.

ABSTRACT

The implementation of distance learning (PJJ) policies during the Covid-19 pandemic received many responses, criticisms and suggestions containing positive and negative tweets from the public via Twitter. These tweets are difficult to sort out to get terms from positive sentiment and negative sentiment because the words used vary. Overcoming this, sentiment analysis is carried out to extract information related to PJJ policies. To get the best evaluation results using a comparison with 4 scenarios. The research phase began with Twitter data collection, data cleaning, positive and negative labeling, data preprocessing including tokenize, casefolding, stemming and stopword removal, TF-IDF for word weighting and SMOTE handling data imbalance, classification using the Naïve Bayes algorithm and Particle Swarm Optimization (PSO) as feature selection. Combined Naïve Bayes testing and PSO with 4 scenarios produces a fairly good accuracy, with an average value of 97.29% accuracy, 94.72% recall, and 99.84% precision. The highest accuracy results at a ratio of 60:40 with the Naïve Bayes algorithm and PSO of 97.88%. The effect of PSO on model performance is quite good as can be seen in the increased accuracy.

Keywords: CRISP-DM, *Naïve Bayes*, *Particle Swarm Optimization*, PJJ, *Sentiment Analysis*, SMOTE