

Abstrak

Twitter adalah situs media sosial berbasis mikroblog yang diluncurkan pada 13 Juli 2006. Pada Maret 2020, sebanyak 476.696 tweet tentang kebijakan pemerintah dalam penyebaran COVID-19 di Twitter ditangkap oleh Institute for Development of Economics and Finance (Indef). Kebijakan pemerintah mempunyai arti baku, yaitu suatu keputusan yang dibuat secara sistematis oleh pemerintah dengan maksud dan tujuan tertentu yang berkaitan dengan kepentingan umum, baik yang dilakukan secara langsung maupun tidak langsung. Analisis sentimen menganalisis pendapat, sentimen, evaluasi, sikap, dan emosi orang dari bahasa tertulis. Dalam dekade ini, Analisis Sentimen telah menjadi bidang penelitian yang trendi. Tujuan dari makalah ini adalah untuk memfokuskan bagaimana mengimplementasikan word2vec menggunakan kata kesamaan sebagai perluasan fitur untuk meminimalkan ketidakcocokan kosakata dalam Analisis Sentimen Twitter menggunakan "word embeddings". Penelitian ini berisi 11.395 tweet untuk sebuah dataset, dimana dataset tersebut akan digunakan dalam dua klasifikasi yaitu Algoritma Support Vector Machine dan Algoritma Jaringan Syaraf Tiruan. Keluaran dari Word2Vec akan digunakan untuk perluasan fitur dalam penelitian ini, dimana algoritma perluasan akan memeriksa setiap baris pada korpus yang memiliki vektor kesamaan dengan kata tersebut dan akan mengganti kata dengan kesamaan kata tersebut jika nilainya 0. Dataset pada Feature Expansion menggunakan 142.545 artikel dari media Indonesia. Hasil dari penelitian ini adalah JST lebih baik dari SVM, dimana JST tanpa perluasan fitur mendapatkan 68,89 % dan menggunakan perluasan fitur mendapatkan 72,58%. Untuk SVM, akurasi akhir tanpa perluasan fitur adalah 63,95 %, dan menggunakan perluasan fitur mendapatkan 68,56 %. Penelitian ini membuktikan bahwa perluasan fitur dapat meningkatkan akurasi akhir.

Kata Kunci: Sentiment Analysis, SVM, ANN, Word2Vec, TF-IDF