

Word2Vec pada Analisis Sentimen dengan *Synthetic Minority Oversampling Technique* dan Algoritma *Boosting*

Rayhan Rahmanda¹, Erwin Budi Setiawan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹rayhanr@students.telkomuniversity.ac.id, ²erwinbudisetiawan@telkomuniversity.ac.id

Abstrak

Opini pelanggan menjadi aspek penting dalam menentukan kesuksesan perusahaan atau penyedia jasa. Dengan menentukan sentimen dari opini yang ada perusahaan dapat menggunakannya sebagai bahan evaluasi untuk memperbaiki kualitas dari layanan atau produk yang diberikan. Analisis sentimen dapat digunakan sebagai pengukur sentimen opini dengan data masukan berupa korpus yang akan diklasifikasikan ke dalam kelas positif, netral, ataupun negatif untuk memperoleh tingkat kepuasan pelanggan terhadap suatu produk atau jasa. Dengan menggunakan salah satu pendekatan analisis sentimen yaitu berbasis aspek maka perusahaan dapat menganalisis dengan lebih spesifik dan mengetahui aspek apa yang perlu ditingkatkan. Sejauh ini diketahui bahwa belum ada penelitian analisis sentimen berbasis aspek yang mencoba menggunakan ekspansi fitur. Sehingga pada penelitian ini dilakukan analisis sentimen berbasis aspek terhadap pengguna Telkomsel di Twitter. Data yang digunakan berupa 16.992 tweet yang membahas mengenai beberapa aspek seperti pelayanan dan sinyal Telkomsel. Ekspansi fitur word embedding ditambahkan pada penelitian ini dengan menggunakan Word2Vec untuk meminimalisir ketidakcocokan kosakata yang diakibatkan oleh keterbatasan kata pada tweet yang ada. Dari pengujian yang dilakukan, penggunaan Logistic Regression yang dikombinasikan dengan Word2Vec, Synthetic Minority Oversampling Technique (SMOTE), dan Algoritma Boosting mendapat akurasi tertinggi sebesar 95,10% untuk aspek sinyal dan dengan menggunakan hyperparameter aspek pelayanan mendapatkan akurasi 93,34%.

Kata kunci: analisis sentimen, logistic regression, word2vec, SMOTE, algoritma boosting, twitter, telkomsel
