

Analisis sentimen menggunakan metode Naive Bayes, maximum entropy, dan support vector machine pada dokumen berbahasa Inggris dan dokumen berbahasa Indonesia hasil penerjemahan otomatis

Franky, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=123194&lokasi=lokal>

Abstrak

Sentimen merupakan opini atau penilaian penulis dokumen mengenai topik yang dibahas dalam dokumen tersebut. Analisis sentimen merupakan suatu tugas yang melakukan polarisasi dokumen berupa pengklasifikasian dokumen ke dalam sentimen positif dan negatif. Penggunaan metode Naive Bayes, Maximum Entropy, dan Support Vector Machine telah ditunjukkan mampu untuk menangkap informasi sentimen dari dokumen review film pada domain bahasa Inggris (Pang, Lee, & Vaithyanathan, 2002). Laporan tugas akhir ini menjelaskan percobaan yang mengaplikasikan kembali metode Naive Bayes, Maximum Entropy, dan Support Vector Machine untuk analisis sentimen pada dokumen berbahasa Indonesia hasil penerjemahan otomatis menggunakan kamus bilingual dan program penerjemah, pada dokumen review film.

Hasil analisis sentimen yang didapat dibanding kan dengan hasil analisis sentimen pada dokumen berbahasa Inggris. Percobaan analisis sentimen dilakukan dengan memvariasikan metode penerjemahan dan pengolahan data, fitur yang digunakan, dan informasi nilai fitur berupa nilai kemunculan fitur (presence), frekuensi, normalisasi nilai frekuensi, dan pembobotan menggunakan tf-idf. Baseline untuk analisis sentimen pada bahasa Indonesia dibuat dengan metode klasifikasi yang sederhana.

Hasil yang didapat menunjukkan bahwa analisis sentimen menggunakan machine learning untuk dokumen berbahasa Indonesia hasil penerjemahan otomatis dapat dilakukan, dengan akurasi tertinggi sebesar 78.82%. Hasil ini lebih baik dari akurasi yang didapat dari baseline sebesar 52.43% tetapi tidak melebihi akurasi tertinggi pada dokumen berbahasa Inggris sebesar 80.09%, namun cukup dekat. Penggunaan fitur yang diambil dari 25% bagian terakhir dokumen memberikan hasil yang lebih baik dari penggunaan fitur yang diambil dari keseluruhan dokumen. Sementara, metode Support Vector Machine secara umum memberikan hasil analisis sentimen dengan akurasi yang lebih baik dari metode machine learning lain yang digunakan.