LAPORAN TUGAS AKHIR

Analisis Sentimen Menggunakan Metode Naive Bayes, Maximum Entropy, dan Support Vector Machine pada Dokumen Berbahasa Inggris dan Dokumen Berbahasa Indonesia Hasil Penerjemahan Otomatis

Disusun oleh:

Franky

1204000343

Fakultas Ilmu Komputer

Universitas Indonesia

Depok, 2008

HALAMAN PERSETUJUAN

Judul Tugas Akhir:

Analisis Sentimen Menggunakan Metode Naive Bayes, Maximum Entropy, dan Support Vector Machine pada Dokumen Berbahasa Inggris dan Dokumen Berbahasa Indonesia Hasil Penerjemahan Otomatis

Nama: Franky

NPM: 1204000343

Laporan tugas akhir ini telah diperiksa dan disetujui.

Depok, Juli 2008

Pembimbing Tugas Akhir

Dr. Hisar Maruli Manurung

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 dibandingkan 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.

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas penyertaannya untuk kehidupan yang dilalui dalam pengerjaan tugas akhir ini. Terima kasih yang sebesar-besarnya untuk Pak Ruli sebagai pembimbing tugas akhir yang memperhatikan dan membimbing dalam tiap tahapan pengerjaan tugas akhir serta memberikan ide-ide dalam menyelesaikan masalah yang ditemui. Terima kasih juga untuk sumber daya yang diberikan terutama untuk komputer yang memadai.

Terima kasih pada keluarga di rumah yang telah memberikan dukungan berupa penghidupan dan sumber daya untuk tugas akhir ini. Terima kasih juga untuk Mimi yang selalu ada terutama saat makan siang dan sebagai teman bicara. Selain itu, penulis juga ingin mengucapkan terima kasih kepada Desmond dan Eliza sebagai teman yang membantu dan memberikan pendapat yang membangun dalam penyusunan laporan tugas akhir. Ucapan terima kasih juga diberikan untuk teman-teman di laboratorium *information retrieval* (IR) yang cukup ramai dan membantu secara langsung atau tidak langsung lewat kata-kata, tindakan atau pengaruh yang dipancarkan. Terima kasih juga untuk Ibu Mirna yang selalu membawa sesuatu yang luar biasa untuk dikonsumsi saat pertemuan anggota lab pada hari Rabu. Terima kasih untuk penguji, Ibu Mirna dan Pak Stef, yang memberikan saran dan kritik yang positif untuk tugas akhir ini.

Akhir kata, terima kasih kepada teman-teman angkatan 2004 dan kepada seluruh elemen Fakultas Ilmu Komputer Universitas Indonesia yang turut serta mendukung dan mewarnai kehidupan selama berada di Fasilkom ini. Kritik dan saran yang membangun akan diterima dengan tulus. Semoga penelitian yang dilakukan ini membantu dalam pengembangan ilmu komputer di Indonesia.

Jakarta, 27 Juni 2008

Franky

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL