

1. Pendahuluan

1.1. Latar Belakang Masalah

Dalam beberapa tahun ini, terjadi pertumbuhan informasi berupa data teks yang sangat pesat, seperti pada website sehingga timbul kebutuhan untuk mengelola dan mengkategorikan data teks secara efektif. Kategorisasi teks merupakan salah satu solusi untuk masalah ini, yang mengklasifikasikan data teks ke dalam kategori yang ditetapkan. Pada umumnya data teks tersebut masih bersifat tidak terstruktur atau semi terstruktur sehingga perlu dilakukan konversi untuk mengubah data teks tersebut menjadi terstruktur. Proses konversi tersebut juga bisa dikatakan sebagai *text preprocessing*.

Text preprocessing berperan untuk mengubah dokumen menjadi terstruktur sehingga dapat diproses lebih lanjut oleh *classifier*. Salah satu tahapan penting dalam preprocessing adalah term weighting. Menurut studi yang dilakukan oleh Lan Man, pembobotan term merupakan langkah vital yang paling berpengaruh terhadap performansi klasifikasi teks[5], dan belum adanya penelitian yang membandingkan metode pembobotan secara komprehensif untuk korpus berbahasa Indonesia sehingga perlu dilakukan penelitian untuk membandingkan pembobotan term manakah yang paling baik performansinya khususnya untuk proses klasifikasi teks berbahasa Indonesia secara komprehensif untuk korpus-korpus berbahasa Indonesia.

Dalam Tugas Akhir ini dilakukan sebuah studi perbandingan yang komprehensif terhadap metode-metode pembobotan term untuk korpus-korpus bahasa Indonesia dan melakukan penelitian pengaruh pemilihan pembobotan term terhadap efektivitas klasifikasi. Metoda yang diteliti adalah seluruh metode yang dijumpai yaitu TF, IDF, ITF, TF-IDF, RF, TF-RF, OR, TF-OR, IG, TF-IG, NGL, GSS, QF, IQF, ICF, VRF, QF-IQF-ICF, CHI-SQUARE, TF-X2, BINARY, dan MI. Sedangkan korpus bahasa Indonesia yang akan dipakai yaitu korpus yang menggunakan penulisan formal relatif panjang seperti artikel berita SINDO, pertanyaan non formal yang pendek seperti Yahoo Answer, dan informasi dengan penulisan non formal seperti tweet.

1.2. Perumusan Masalah

Perumusan masalah yang diangkat dalam tugas akhir ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan metode pembobotan *term* untuk proses klasifikasi teks berbahasa Indonesia?
2. Metode pembobotan *term* apakah yang memiliki performansi paling baik untuk proses klasifikasi teks berbahasa Indonesia?
3. Bagaimana pengaruh metode pembobotan terhadap efektivitas klasifikasi dengan korpus berbahasa Indonesia ?

1.3. Batasan Masalah

Batasan masalah yang diangkat dalam tugas akhir ini adalah untuk pengujian efektivitas pembobotan menggunakan klasifikasi.

1.4. Tujuan

Tujuan dari Tugas Akhir ini adalah sebagai berikut:

1. Dapat mengimplementasikan beberapa metode pembobotan *term* pada proses klasifikasi teks berbahasa Indonesia.
2. Mengetahui metode pembobotan *term* berbahasa Indonesia manakah yang memiliki performansi paling baik (memiliki nilai *f-measure* paling baik) untuk proses klasifikasi teks berbahasa Indonesia.
3. Mengetahui pengaruh metode pembobotan *term* terhadap efektivitas klasifikasi dengan korpus berbahasa Indonesia.

1.5. Metodologi Penyelesaian Masalah

- a. Studi Literatur
Melakukan pencarian terhadap literatur yang sesuai dengan permasalahan meliputi perbandingan terhadap berbagai metode pembobotan *term*.
- b. Pengumpulan data
Melakukan pengumpulan data untuk mendukung pengimplementasian berbagai metode *term weighting*.
- c. Analisis dan Perancangan
Menganalisis kebutuhan sistem, dan proses-proses yang terjadi pada sistem yang dibangun
- d. Implementasi
Membangun perangkat lunak sesuai dengan hasil perancangan pada bahasa pemrograman java dan melakukan pengujian terhadap data yang didapatkan.
- e. Analisis

Menganalisis performansi metode pembobotan *term* terhadap pengkategorian teks.

- f. Kesimpulan
Melakukan pengambilan kesimpulan dari hasil testing yang didapatkan.
- g. Pembuatan Laporan Tugas Akhir
Mendokumentasikan kegiatan yang dilakukan pada pengerjaan tugas akhir ke dalam sebuah laporan.

1.6. Sistematika Penulisan

Tugas akhir ini akan dilaksanakannya dengan memiliki sistematika penulisan sebagai berikut:

BAB 1 Pendahuluan

Bab ini berisi mengenai penjelasan secara umum dalam tugas akhir, yang meliputi latar belakang masalah, perumusan masalah, tujuan, dan metodologi penyelesaian masalah.

BAB 2 Dasar Teori

Bab ini akan menjelaskan landasan teori yang berkaitan dengan pembangunan perangkat lunak.

BAB 3 Perancangan dan Implementasi

Bab ini berisi rancangan sistem sesuai dengan kasus yang dipilih, dan dilanjutkan dengan pengimplementasiannya.

BAB 4 Pengujian dan Analisis

Bab ini akan menjelaskan tentang pengujian terhadap perangkat lunak dan melakukan analisis terhadap *output* pengujian sebelumnya.

BAB 5 Kesimpulan dan Saran

Bab ini berisi kesimpulan dari tugas akhir mengenai perangkat lunak yang dibangun dan berisi juga saran bagi pengguna yang memakai atau yang ingin mengembangkan penelitian lebih lanjut.