

1 Pendahuluan

1.1 Latar Belakang

Dalam mengkategorisasikan suatu teks diperlukan data yang terstruktur, sedangkan data yang didapatkan adalah data yang belum terstruktur. Maka dari itu diperlukan tahapan *text preprocessing*. *Text preprocessing* adalah tahapan untuk mengubah suatu data yang tidak terstruktur menjadi data yang terstruktur. Tahap *text preprocessing* memiliki beberapa proses, yaitu *case folding*, *stopwords removing*, *stemming*, dan *term weighting*.

Proses yang penting dalam kategorisasi teks adalah proses *term weighting*. *Term Weighting* adalah proses pemberian bobot pada *term* atau kata yang ada di dalam suatu dokumen. Dalam Tugas Akhir ini berfokus pada proses *term weighting* dengan membandingkan dua metode yaitu TF.CHI² dan TF.RF. TF.CHI² adalah salah satu metode pembobotan kata hasil dari penggabungan antara metode TF dan metode CHI². TF.RF adalah metode pembobotan kata hasil dari penggabungan dua metode yaitu metode TF dan RF.

Metode TF.CHI² dan TF.RF dipilih untuk dianalisis perbandingannya karena pada penelitian sebelumnya menganalisis perbandingan metode TF.IDF dan TF.RF yang menghasilkan kesimpulan bahwa metode TF.RF lebih baik dibandingkan TF.IDF karena memperhitungkan kategori dimana *term* muncul pada suatu dokumen. Pada penelitian yang lain [10] menyebutkan bahwa TF.CHI² lebih baik dari TF.IDF pada SVM - *based*, maka TF.CHI² dipilih untuk dianalisis perbandingannya dengan metode TF.RF agar dapat mengetahui metode yang paling efektif dan memiliki performansi kesesuaian yang paling baik dengan SVM dalam proses *term weighting*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas, permasalahan yang dapat diambil adalah sebagai berikut :

1. Bagaimana menganalisis performansi metode *term weighting* TF.CHI² dan TF.RF dalam kategorisasi dokumen teks.
2. Dari metode *term weighting* TF.CHI² dan TF.RF yang akan dibandingkan, metode mana yang memiliki performansi kesesuaian paling baik dengan SVM

1.3 Tujuan

Tujuan yang akan dicapai pada Tugas Akhir ini adalah sebagai berikut :

1. Menganalisis *term weighting* TF.CHI² dan TF.RF untuk menyelesaikan permasalahan kategorisasi dokumen teks dengan melihat akurasi dari performansinya.
2. Menentukan metode *term weighting* yang paling baik kesesuaiannya dengan SVM diantara TF.CHI² dan TF.RF

1.4 Batasan Masalah

Batasan masalah yang ada pada Tugas Akhir ini adalah sebagai berikut :

1. Pengkategorisasian dokumen hanya dilakukan pada teks berbahasa Indonesia.
2. Tidak mengimplementasikan metode klasifikasi *Support Vector Machine* (SVM)

1.5 Metodologi / Desain

Metodologi yang digunakan dalam pembuatan aplikasi ini adalah sebagai berikut :

1. Pengumpulan data dan Studi Literatur
Melakukan pengumpulan data dokumen abstrak dari Tugas Akhir yang ada di Fakultas Informatika Universitas Telkom dan dokumen artikel dari berbagai situs di *internet* dan mempelajari berbagai literature yang berhubungan dengan topik yang diambil.
2. Perancangan
Melakukan analisa kebutuhan system, fungsionalitas, dan proses – proses yang terjadi didalam sistem tersebut.
3. Implementasi
Membangun sistem sesuai dengan perancangan yang ada.
4. Testing
Melakukan pengecekan terhadap sistem, apakah sistem tersebut mengandung *bug* atau *error*.
5. Analisis
Menganalisa dan melakukan pengukuran performansi sistem khususnya pada metode *term weighting* yang paling baik terhadap metode klasifikasi SVM.
6. Pengambilan Kesimpulan
Mengambil kesimpulan dari hasil analisis testing yang telah dilakukan .
7. Penyusunan laporan
Mendokumentasikan berbagai tahapan kegiatan pembangunan sistem ini.