

1 Pendahuluan

1.1. Latar Belakang

Makalah ilmiah sebagai salah satu sumber ilmu pengetahuan sangat dibutuhkan oleh para pelajar dan peneliti, terutama untuk kebutuhan studi literatur. Namun ketersediaannya yang sangat banyak telah menjadi suatu permasalahan tersendiri, dimana sang penuntut ilmu harus memeriksa setiap makalah yang dianggap memiliki keterkaitan terhadap topik tertentu untuk memperoleh literatur yang paling cocok untuknya [1].

Salah satu cara untuk mendapatkan sumber literatur yang cocok adalah dengan meringkas sendiri makalah yang berkaitan atau membaca ringkasan lain tentang makalah tersebut. Menggunakan “peringkasan otomatis” merupakan salah satu terobosan yang dapat membantu dalam memperoleh ringkasan dari makalah ilmiah tertentu dengan cepat.

Beberapa teknik peringkasan yang sudah ada memberikan hasil ringkasan dalam bentuk kumpulan kalimat singkat yang menggambarkan isi dari suatu makalah ilmiah. Namun sajian berupa uraian kalimat saja tidak dapat menggambarkan hubungan kerja, kekuatan atau kelebihan, dan batasan dari makalah yang diringkaskan secara langsung. Dalam penelitian ini ditawarkan solusi terhadap masalah ini, yaitu dengan menyajikan hasil ringkasan yang terstruktur sehingga pembaca dapat mengidentifikasi isi ringkasan secara langsung. Hasil terstruktur tersebut didapat dengan mengelompokkan hasil berdasarkan poin-poin yang biasa digunakan dalam *review*. Klasifikasi yang merupakan salah satu metode dalam *text mining* sangat cocok untuk permasalahan peringkasan seperti ini. Sebelumnya sudah banyak penelitian yang membahas klasifikasi untuk teks dengan pendekatan *machine learning*, seperti Bayesian [6, 16] dan Support Vector Machine (SVM) [8, 16].

Pada penelitian [1] telah dilakukan peringkasan makalah ilmiah dengan mengklasifikasikan kalimat sitasi kedalam beberapa kelas. Dalam proses klasifikasinya, diperoleh performa terbaik dengan *classifier* Naive Bayes dan kombinasi fitur *adjective*, *verb*, dan *bigram*.

Dalam penelitian ini diusulkan pengembangan sistem yang mengacu [1], dengan menggunakan *classifier* Naive Bayes dan Support Vector Machine (SVM) yang tidak dilakukan pada penelitian tersebut. Fitur yang digunakan disini merupakan kombinasi fitur yang juga digunakan pada [1] dengan menambah tambahan fitur *unigram* dan *noun*. Dalam penelitian ini juga dipaparkan fitur dan *classifier* terbaik bagi peringkasan makalah ilmiah terstruktur.

1.2. Perumusan Masalah

Perumusan masalah yang akan dibahas adalah:

1. Bagaimana kombinasi fitur yang paling baik untuk proses klasifikasi kalimat sitasi.
2. *Classifier* mana yang paling baik untuk proses ini

Ruang lingkup yang harus diperhatikan dalam pengerjaan tugas akhir ini adalah:

1. Data input yang digunakan berupa kumpulan kalimat sitasi dari beberapa makalah ilmiah.
2. Klasifikasi hasil ringkasan diasumsikan berupa *single-label classification*, artinya setiap data (kalimat sitasi) hanya akan memiliki satu label dari kelas-kelas yang ada. Karena diterapkannya asumsi tersebut, maka tidak ada data yang termasuk ke dalam dua atau lebih kelas.
3. *Preprocessing* yang dilakukan adalah *feature extraction*.
4. Pengukuran evaluasi yang dipakai adalah *precision*, *recall*, dan *F-measure* [2].
5. Pengerjaan tugas akhir ini berfokus pada proses klasifikasi untuk menentukan kelompok fitur dan *classifier* terbaik dalam kasus peringkasan makalah ilmiah terstruktur, sehingga tidak melakukan keseluruhan proses peringkasan makalah ilmiah.

1.3. Tujuan

Tujuan dari Tugas Akhir ini adalah:

1. Menemukan kelompok fitur terbaik yang digunakan dalam peringkasan makalah ilmiah.
2. Menentukan *classifier* terbaik yang digunakan dalam peringkasan makalah ilmiah.

1.4. Metodologi Penyelesaian Masalah

Metodologi yang digunakan untuk menyelesaikan masalah dalam Tugas Akhir ini meliputi:

1. Studi Literatur.
Mencari referensi pendukung seperti buku, karya ilmiah, dan jurnal penelitian untuk memperoleh pemahaman dan pengetahuan terkait masalah dalam *Text Mining* seperti Fitur, Ekstraksi Fitur, Klasifikasi, dan Evaluasi.
2. Pengumpulan Data.
Mengumpulkan data berupa kalimat sitasi dari beberapa makalah ilmiah. Pengumpulan data dilakukan secara manual dengan membaca dan memahami isi dari sampel makalah dan mengambil kalimat sitasi di dalamnya. Kemudian setiap kalimat sitasi yang didapat diberi label atau kelas.

3. Perancangan Sistem.
Merancang sistem untuk proses klasifikasi peringkasan makalah ilmiah dengan menerapkan teori-teori yang sudah dipelajari dan alat-alat pendukung yang biasa digunakan dalam klasifikasi teks. Dalam perancangan sistem ini, diterapkan konsep yang berorientasikan objek dengan bahasa pemrograman Java.
4. Implementasi dan Pengujian.
Mengimplementasikan sistem yang dirancang terhadap masalah dalam beberapa skenario pengujian. Pada pengujian sistem digunakan data set berdasarkan skenario yang dibuat untuk mengukur performansi sistem.
5. Analisis Hasil dan Kesimpulan.
Hasil dari pengujian sistem dianalisis dengan menerapkan skema uji yang telah dibuat. Kemudian merumuskan kesimpulan dari hasil analisis yang didapat.

1.5. Sistematika Penulisan

Penulisan Tugas Akhir ini dibagi menjadi beberapa bab. Bab 1 menjelaskan tentang pendahuluan yang mencakup latar belakang, rumusan masalah, tujuan, dan metodologi. Bab 2 menjelaskan tentang tinjauan pustaka yang berisikan tentang *text mining*, *preprocessing*, klasifikasi, evaluasi, dan lain-lain. Bab 3 menjelaskan tentang perancangan sistem yang berisikan tentang pembangunan sistem dalam pengerjaan Tugas Akhir ini. Bab 4 menjelaskan pengujian, hasil dari proses klasifikasi yang dijalankan, serta analisis untuk performnsinya. Pada bab 5 dijelaskan kesimpulan dari pengerjaan tugas akhir ini dan saran untuk penelitian yang lebih lanjut.