BAB 1
PENDAHULUAN

1.1. Latar Belakang Masalah

Seiring dengan perkembangan teknologi, untuk mendapatkan informasi tentunya semakin mudah, salah satu caranya adalah dengan mengakses suatu media online. Salah satu media online yang memuat informasi adalah website berita, diantaranya adalah www.liputan6.com dan www.kompas.com. Banyaknya situs yang memuat berita menimbulkan pembahasan berita dengan topik yang sama selalu berulang walaupun dengan judul yang berbeda. Oleh karena itu, dibutuhkan suatu peringkasan berita untuk mempermudah pembaca dalam mencari informasi yang lengkap tentang suatu topik berita tanpa harus membaca semua berita yang ada. Salah satu cara peringkasan adalah peringkasan manual yaitu peringkasan yang dilakukan secara manual oleh seorang individu, tetapi dengan cara peringkasan manual akan memakan banyak waktu dan sumberdaya. Supaya mempermudah dalam peringkasan dibutuhkan suatu alat/sistem yang bisa otomatis meringkas isi dari beberapa dokumen [1].

Ada berbagai cara untuk melakukan peringkasan teks otomatis multi dokumen, salah satunya pada penelitian yang dilakukan Deni Fitriaman, Masayu Leylia Khodra, bambang Rianto Trilaksono. Dalam penelitian ini dilakukan peringkasan teks multi dokumen dengan menggunakan metode SVM (*Support Vektor Machine*). Adapun metode SVM pada penelitian tersebut digunakan untuk mengklasifikasikan kalimat yang masuk pada peringkasan. Selanjutnya kalimat hasil peringkasan disusun menggunakan metode MMR (*Maximum Marginal Relevance*). Dari penelitian ini didapatkan nilai rata-rata *recall* 0.773, *precission* 0.771, *f-measure* 0.771 [1].

Pengembangan metode SVM (*Support Vector Machine)* yaitu metode RVM (*Relevance Vector Machine*) [2]. Pada penelitian yang dilakukan oleh Christopher M. Bishop, Michael E. Tipping didapatkan kesimpulan tingkat error dari SVM lebih tinggi dari RVM yaitu SVM 10,6% dan RVM 9,3% [3, 2].

Penelitian tentang pembobotan ekstraksi fitur telah dilakukan oleh Zulkifi, Agung Toto Wibowo, Gia Septiana [4]. Pada penelitian tersebut didapatkan bahwa ekstraksi fitur yang baik adalah posisi kalimat, kemiripan antar kalimat, kalimat yang menyerupai judul dokumen, kalimat yang mengandung kata entiti, kalimat yang mengandung data numerik, panjang kalimat, koneksi antar kalimat dan penjumlahan bobot koneksi antar kalimat dengan akurasi 53%.

Berdasarkan hal tersebut, maka dalam penelitian ini membuat peringkasan teks otomatis multi dokumen menggunakan metode RVM. Hasil dari penelitian ini diharapkan bisa menghasilkan peringkasan teks yang lebih akurat.

1.2. Rumusan Masalah

Berdasarkan uraian pada latar belakang masalah diatas, rumusan masalah dalam penelitian ini adalah apakah RVM dapat digunakan untuk peringkasan multi dokumen.

1.3. Maksud dan Tujuan

Penelitian ini bermaksud membuat sistem peringkasan teks otomatis pada multi dokumen. Sedangkan tujuan yang ingin dicapai dari penelitian ini adalah menguji /mengukur seberapa baik ringkasan yang dihasilkan dengan menggunakan metode RVM pada peringkasan teks multi dokumen dengan mengetahui nilai performasi (*recall, precision dan f-measure)* dan akurasi peringkasan pada multi dokumen dengan menggunakan metode RVM.

1.4. Batasan Masalah

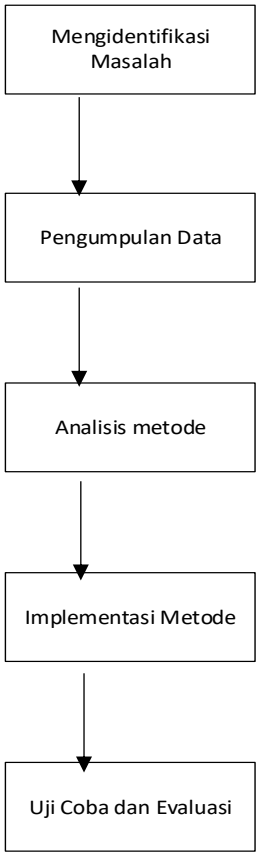
Adapun beberapa batasan-batasan dalam penelitian ini yaitu :

1. Dokumen yang diringkas adalah berita berbahasa Indonesia.
2. Dokumen diambil dari **www.liputan6.com, www.kompas.com.**
3. Input teks berupa berita dengan tema yang sama untuk pelatihan dan dokumen dengan tema yang sama untuk pengujian.
4. Tema dari data masukan yaitu olahraga sepak bola.
5. Daftar *stoplist* untuk proses *stopword removal* yang digunakan adalah hasil penelitian dari Tala, Fadillah Z yang dilampirkan pada halaman D-1.
6. Jenis peringkasan yang digunakan adalah *ekstraksi*.

7. Metode pembobotan yang dipakai menggunakan fitur ekstraksi diantaranya posisi kalimat, kemiripan antar kalimat, kalimat yang menyerupai judul dokumen, kalimat yang mengandung kata entiti, kalimat yang mengandung data numerik, panjang kalimat, koneksi antar kalimat dan penjumlahan bobot koneksi antar kalimat.
8. Output yang dihasilkan dari sistem peringkasan teks berupa ringkasan teks dan jumlah ringkasan yang dihasilkan tidak dibatasi.

1.5. Metodologi Penelitian

Metode pada penelitian yang akan digunakan dalam penelitian ini dapat dilihat pada Gambar 1.1.



Gambar 1.1 Metodologi Penelitian

1.5.1. Identifikasi Masalah

Pada penelitian ini dilakukan identifikasi masalah dengan cara menganalisis penelitian sebelumnya yang sudah dilakukan dimana salah satu hasil analisis adalah belum ada penelitian peringkasan teks multi dokumen menggunakan RVM, sehingga belum diketahui seberapa baik penggunaan RVM dalam peringkasan teks multi dokumen.

1.5.2. Pengumpulan data

Pengumpulan data dilakukan dengan cara mengumpulkan berbagai literatur dan referensi yang berkaitan dengan penelitian yang dilakukan, mulai dari buku, jurnal, artikel. Selain itu dilakukan juga pengumpulan dokumen-dokumen berbahasa Indonesia yang akan dijadikan bahan pelatihan dan pengujian dari penelitian yang akan dilakukan.

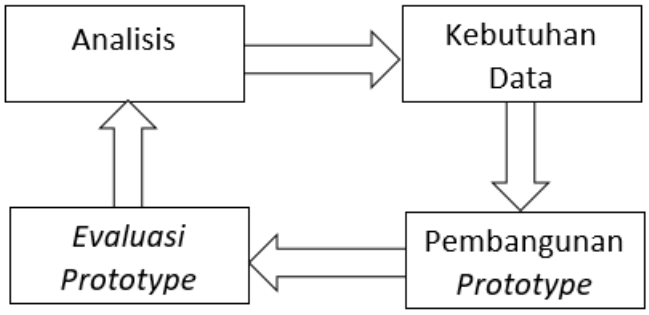
1.5.3. Analisis Metode

Dalam penelitian ini analisis dibagi menjadi beberapa bagian, yaitu analisis masalah yang diangkat, analisis data masukan, analisis tahap *preprocessing*, analisis peringkasan dengan melakukan penghitungan manual dari metode yang digunakan, serta analisis kebutuhan fungsional dan non fungsional dari perangkat lunak yang akan dibangun. Setelah analisis dilakukan, dibuat suatu perancangan model perangkat lunak yang sebelumnya telah dianalisis kebutuhan fungsionalnya, dimana perangkat lunak yang dibangun untuk mensimulasikan metode yang diterapkan.

1.5.4. Implementasi Metode

Setelah melakukan tahapan analisis metode *Relevance Vector Machine* yang akan digunakan untuk peringkasan teks otomatis, langkah selanjutnya adalah implementasi metode tersebut dengan cara membangun perangkat lunak berbasis *web* menggunakan bahasa pemrograman *php*. Implementasi dilakukan berdasarkan analisis dan perancangan yang telah dilakukan sebelumnya dari mulai dari tahap *preprocessing*, tahap pelatihan, hingga ke tahap pengujian hasil setelah proses pelatihan.

Implementasi pembangunan perangkat lunak peringkasan teks multi dokumen menggunakan metode *prototype*. Karena pada pembangunan sistem peringkasan teks terus menerus diperbaiki [6].



Gambar 1.2 Metode *Prototype*

1. Analisis

   Pada tahap ini dilakukan analisis dari studi litelatur penelitian sebelumnya yang telah dikerjakan, analisis data yang akan menjadi masukan, analisis pembobotan fitur, analisis metode *Relevance Vector machine*.

2. Kebutuhan data

   Pada tahapan ini dilakukan mengumpulkan data yang dibutuhkan sistem, dimana data dari sistem ini didapat dari situs liputan6.com dan kompas.com.

3. Pembangunan *prototype*

   Merupakan tahap implementasi dari proses analisis metode dan perancangan sistem yang sudah didapatkan. Pada tahapan ini mengimplementasikan pembobotan fitur ekstraksi dan metode *Relevance Vector Machine*.

4. Evaluasi Prototype

   Program akan diuji, dimana pengujian dilakukan untuk mengetahui kekurangan pada program. Jika masih ada kekurangan, maka akan diperbaiki kembali melalui tahapan-tahapan yang ada pada metode *prototype.* Proses ini akan diulang sampai *prototype* yang dihasilkan mendekati yang diinginkan [6].

1.5.5. Uji Coba dan Evaluasi

Melakukan uji coba dengan menggunakan perangkat lunak yang telah dibuat. Uji coba yang dilakukan adalah dengan melakukan proses pengujian dengan menggunakan dokumen-dokumen uji yang telah disiapkan, melihat akurasi dari setiap percobaan, serta membuat kesimpulan terhadap hasil percobaan tersebut apakah metode *Relevance Vector Machine* memiliki akurasi yang baik atau tidak.

1.6. Sistematika Penulisan

Sistematika penulisan skripsi peringkasan teks disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan penelitian ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini menerangkan tentang secara umum mengenai latar belakang, rumusan masalah, menentukan maksud dan tujuan, batasan masalah, metodologi penelitian, metode pembangunan perangkat lunak, serta sistematika penulisan yang dimaksudkan agar dapat memberikan gambaran tentang urutan dalam menyajikan skripsi peringkasan teks.

BAB 2 LANDASAN TEORI

Bab ini membahas mengenai peringkasan teks multi dokumen, preprocessing, metode pembobotan fitur ekstraksi dan metode *Relevance Vector Machine* (RVM), *Cosine Similarity* dengan pembobotan *term* menggunakan *tf-idf* hingga model dari analisis dan perancangan yang akan dilakukan.

BAB 3 ANALISIS KEBUTUHAN ALGORITMA

Bab ini menguraikan penjelasan analisis dan perancangan sistem, perhitungan pada fitur ekstraksi, dan metode *Relevance Vector Machine* (RVM) untuk diaplikasikan pada peringkasan teks multi dokumen.

BAB 4 PENGUJIAN DAN PEMBAHASAN

Bab ini berisi tentang pembahasan dari analisis metode dan perancangan sistem yang dilakukan pada bab sebelumnya. Selain itu dilakukan pengujian peringkasan teks multi dokumen menggunakan

metode *Relevance Vector Machine* (RVM), dengan skenario dan hasil pengujian.

BAB 5 KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan dari hasil penelitian yang telah dilakukan dan saran yang diharapkan dapat dikembangkan untuk penelitian selanjutnya.