BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Informasi yang ada pada saat ini tidak hanya bersumber dari satu dokumen saja, melainkan dari beberapa dokumen. Selain itu beberapa informasi dapat mempunyai pandangan yang berbeda tentang suatu topik yang sama. Salah satu informasi yang terdapat dalam dokumen adalah berita. Dokumen berita bisa bersumber dari beberapa media. Saat ini banyak sekali media yang memuat berita serta dipublikasikan memiliki pembahasan yang sama dalam satu topik sejenis, tetapi hanya ditulis dengan tata bahasa dan cara penyampaiannya yang berbeda. Untuk dapat menyerap informasi penting dari kumpulan dokumen berita dengan satu topik yang sama, setidaknya pembaca harus membaca satu-persatu dokumen berita. Ketersediaan informasi berita yang semakin banyak tersebut menjadikan ringkasan sebagai kebutuhan untuk mendapatkan isi berita secara ringkas[1]. Ringkasan merupakan teks singkat dan padat yang dapat dianggap pengganti dari keseluruhan dokumen, karena tetap mempertahankan kandungan informasi penting yang dimiliki dari sumber dokumennya. Ringkasan dapat membantu menghasilkan rangkuman singkat yang diambil dari sebuah dokumen berita, tetapi bagaimana cara membangun suatu sistem yang dapat menghasilkan sebuah rangkuman dari kumpulan dokumen berita dengan satu topik sejenis.

Berdasarkan penelitian[1] sebelumnya, ringkasan dapat dihasilkan dari ekstraksi kalimat dengan menggunakan metode *Maximum Marginal Relevance* (MMR). Metode MMR merupakan salah satu metode ekstraksi ringkasan yang dapat digunakan untuk meringkas *single*-dokumen maupun *multi*-dokumen. Hasil dari penelitian[1] menyimpulkan bahwa, metode MMR dapat digunakan untuk meringkas *single*-dokumen secara otomatis dangan rata-rata *recall* 60%, *precision* 77% dan *f-measure* 66% dalam kasus peringkasan teks otomatis berita berbahasa Indonesia menggunakan metode MMR. Untuk menghasilkan ringkasan yang relevan, sebelumnya dapat dilakukan pemilihan kandidat kalimat ringkasan[2]. Pemilihan kandidat kalimat ringkasan dapat dilakukan dengan mengklasifikasikan

kalimat kedalam kelas positif dan negatif, sehingga kalimat positif yang dihasilkan dijadikan kandidat dalam ringkasan berupa rangkuman singkat. Salah satu metode klasifikasi yang akan digunakan untuk mengklasifikasikan kalimat pada penelitian ini adalah metode *Support Vector Machine* (SVM).

Metode SVM merupakan salah satu metode *machine learning* yang memaksimumkan akurasi prediksi dengan mencari bidang pembatas terbaik dari dua kelas dalam ruang fitur. Metode SVM bekerja dengan mengambil himpunan data masukan, kemudian memprediksi untuk setiap masukan yang diberikan ke dalam salah satu kelas dari dua kelas yang ada yaitu kelas positif dan negatif[3]. Berdasarkan penelitian[4] metode SVM memberikan unjuk kerja yang lebih baik dari pada metode *Naïve Bayes Classifier* (NBC) dalam kasus mengklasifikasikan opini berbahasa Inggris dan opini berbahasa Indonesia dengan rata-rata akurasi hasil penelitian metode NBC menghasilkan 78.38% dan 74.39% sedangkan metode SVM menghasilkan 88.51 % dan 75.06%. Penjelasan mengenai metode SVM dan MMR di atas, menimbulkan permasalahan baru yaitu bagaimana cara mengimplementasikan metode SVM dan MMR untuk menghasilkan sebuah rangkuman yang relevan. Permasalahan tersebut dapat diselesaikan dengan mengetahui nilai performansi *recall*, *precision* dan *f-measure* sejauh mana rangkuman yang dihasilkan benar-benar relevan

Berdasarkan uraian di atas, maka diimplementasikanlah metode SVM dan MMR untuk menghasilkan sebuah rangkuman yang relevan dari kumpulan dokumen berita dengan satu topik sejenis.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah disampaikan, dapat dirumuskan masalah yang timbul diantaranya:

1. Bagaimana cara membangun suatu sistem yang dapat menghasilkan sebuah rangkuman dari kumpulan dokumen berita dengan satu topik sejenis.
2. Bagaimana cara mengimplementasikan metode SVM dan MMR untuk menghasilkan sebuah rangkuman yang relevan.

1.3 Maksud dan Tujuan

Maksud dari penelitian yang dilakukan adalah mengimplementasikan metode *Support Vector Machine* dan *Maximum Marginal Relevance* dari kumpulan dokumen berita dengan topik sejenis.

Sedangkan tujuan yang akan dicapai diantaranya:

1. Menghasilkan suatu sistem yang dapat menampilkan sebuah rangkuman dari kumpulan dokumen berita dengan satu topik sejenis.
2. Mengetahui nilai performansi *recall*, *precision* dan *f-measure* sejauh mana rangkuman yang dihasilkan benar-benar relevan.

1.4 Batasan Masalah

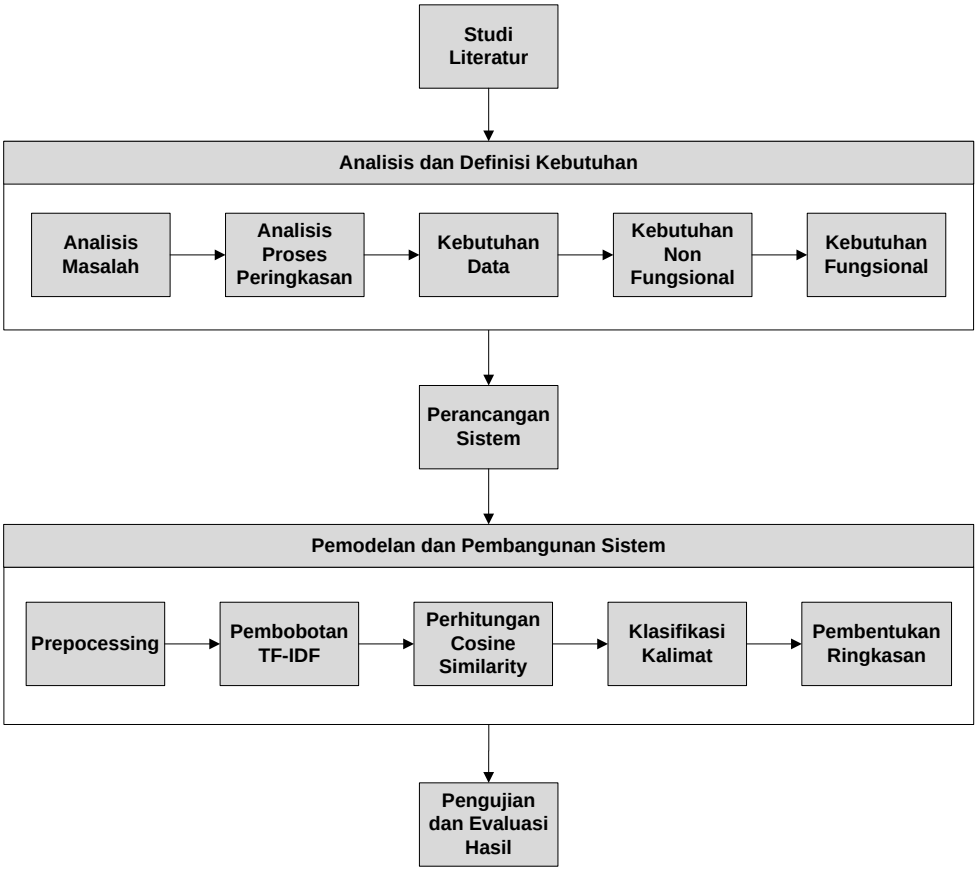
Pembatasan masalah dimaksudkan agar dapat memberikan pemahaman yang terarah dan pembahasan tidak menyimpang dari pokok perumusan masalah yang ada, sehingga mencapai sasaran yang ditentukan sesuai dengan apa yang diharapkan.

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Peringkasan yang dilakukan untuk menghasilkan rangkuman adalah peringkasan terhadap kumpulan dokumen berita bahasa Indonesia.
2. Data dokumen berita yang digunakan diambil dari media *online* pemuat berita bahasa Indonesia dan dipilih berdasarkan satu topik sejenis. Kemudian disimpan dengan format *file* .doc dengan jumlah sampel 4 dokumen berita.
3. Proses *stemming* menggunakan algoritma Porter Stemmer bahasa Indonesia.
4. Kalimat kelas positif menjadi kandidat kalimat dalam rangkuman.
5. Model ringkasan yang digunakan berbasis ekstraktif.
6. Evaluasi ringkasan yang dilakukan berbasis evaluasi intrinstik.
7. *Output* yang dihasilkan adalah sebuah rangkuman dan nilai performansi *recall*, *precision* serta *f-measure.*
8. Pendekatan pembangunan perangkat lunak yang digunakan adalah pendekatan berorientasi objek.

1.5 Metodologi Penelitian

Metodologi penelitian penyelesaian masalah yang dilakukan dalam penelitian ini dapat dilihat dengan langkah-langkah penelitian pada Gambar 1.1 berikut:



Gambar 1.1 Langkah-Langkah Penelitian

Berikut penjelasan lengkap dari Gambar 1.1:

1. Studi Literatur

   Merupakan teknik pengumpulan data dengan cara membaca dan mempelajari serta mengumpulkan literatur. Sumber-sumber yang didapatkan berupa jurnal, buku, paper dan beberapa informasi yang relevan dengan penelitian.

2. Analisis dan Definisi Kebutuhan meliputi:

   a. Analisis Masalah

      Analisis masalah dilakukan untuk memahami masalah yang timbul dan mencari solusi untuk memecahkan masalah dalam menghasilkan sebuah rangkuman dari kumpulan dokumen berita bahasa Indonesia dengan satu topik sejenis.

b. Analisis Proses Peringkasan

Analisis proses peringkasan meliputi 5 proses tahap utama tahap *Preprocessing*, tahap Pembobotan TF-IDF, tahap Perhitungan *Cosine Similarity* (*query relevance* dan *similarity* kalimat), tahap klasifikasi kalimat dengan metode *Support Vector Machine* dan tahap pembentukan ringkasan dengan metode *Maximum Marginal Relevance*.

c. Kebutuhan Data

Data yang dibutuhkan berupa dokumen berita berbahasa Indonesia yang diambil dari media *online* pemuat berita dan dipilih berdasarkan satu topik sejenis. Kemudian disimpan dengan format *file* .doc.

d. Kebutuhan Non Fungsional

Kebutuhan non fungsional menjelaskan kebutuhan *hardware*, *software* dan *brainware* dalam pembangunan aplikasi ini.

e. Kebutuhan Fungsional

Kebutuhan fungsional dengan pendekatan berorientasi objek menggunakan pemodelan UML terdiri dari *use case diagram, use case scenario, activity diagram, class diagram* dan *sequence diagram*.

3. Perancangan Sistem

Perancangan sistem dilakukan untuk mengetahui rancangan dari basis data dan antarmuka dari pembangunan aplikasi ini.

4. Pemodelan dan Pembangunan Sistem

Pemodelan sistem disini merupakan model sistem proses peringkasan dari 5 proses tahap utama yang telah dilakukan analisis sebelumnya untuk menghasilkan rangkuman, dan mengimplementasikannya ke dalam bahasa pemrograman untuk mulai dalam pembangunan aplikasi.

5. Pengujian dan Evaluasi Hasil

Pengujian dilakukan pada aplikasi yang dibangun, serta mengevaluasi hasil rangkuman yang dihasilkan dengan melihat nilai performansi *recall*, *precision* dan *f-measure* untuk mengetahui sejauh mana rangkuman yang dihasilkan benar-benar relevan.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum mengenai penelitian yang dilakukan sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini berisi tentang gambaran umum mengenai isi laporan skripsi yang berisi penjelasan meliputi latar belakang masalah, rumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian dan sistematika penulisan dalam penelitian tentang implementasi metode *Support Vector Machine* (SVM) dan *Maximum Marginal Relevance* (MMR) untuk menghasilkan rangkuman dari kumpulan dokumen berita dengan topik sejenis.

BAB 2 TINJAUAN PUSTAKA

Bab ini menjelaskan tentang berbagai konsep dan teori-teori para ahli dari literatur pendukung untuk menghasilkan rangkuman yang relevan dari kumpulan dokumen berita dengan topik sejenis dengan mengimplementasikan metode SVM dan MMR. Adapun teori-teori yang digunakan seperti definisi dan bentuk ringkasan, tahap *Preprocessing*, pembobotan TF-IDF, perhitungan *cosine similarity*, metode SVM, metode MMR, perhitungan nilai performansi (*recall*, *precision*, *f-measure*). Teori perancangan aplikasi yang akan dibangun menggunakan pemodelan UML seperti *use case diagram, use case scenario, activity diagram, class diagram, sequence diagram* serta teori-teori tentang *tools* atau *software* pendukung yang digunakan dalam pembangunan aplikasi ini.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini berisi penjelasan mengenai analisis masalah dari objek penelitian yang dilakukan untuk mengetahui masalah apa yang timbul dan mencoba memecahkan masalah tersebut, analisis sistem, analisis kebutuhan data masukan, analisis tahap *Preprocessing*, analisis pembobotan TF-IDF, analisis perhitungan *cosine similarity*, analisi klasifikasi kalimat dengan metode SVM, analisis pembentukan ringkasan menggunakan metode MMR dan evaluasi ringkasan yang dihasilkan dengan nilai performansi (*recall, precision* dan *f-measure*). Selain itu dijelaskan juga mengenai analisis kebutuhan perancangan sistem dan perancangan antarmuka dari aplikasi yang akan dibangun.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini berisi tentang implementasi dari analisis dan perancangan yang telah dilakukan pada bab sebelumnya. Implementasi yang dilakukan terdiri dari implementasi sistem berupa implementasi perangkat keras, implementasi perangkat lunak, implementasi basis data dan implementasi antarmuka. Selain itu dilakukan pengujian sistem terhadap aplikasi yang dibangun melalui rencana pengujian, skenario pengujian, serta mengevaluasi hasil rangkuman yang dihasilkan melalui nilai performansi (*recall*, *precision* dan *f-measure*). Untuk menghasilkan kesimpulan dari hasil pengujian sistem. Informasi yang ditampilkan dari implementasi dan pengujian sistem yang dilakukan berupa sebuah rangkuman beserta dan nilai performansi.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi penjelasan mengenai kesimpulan yang didapatkan dari pembahasan pada bab-bab sebelumnya, serta saran yang dapat diberikan oleh peneliti untuk penelitian atau pengembangan selanjutnya. Pada bagian kesimpulan berisi tentang apakah dengan penerapan metode SVM dan MMR rangkuman yang dihasilkan benar-benar relevan dilihat dari nilai performansi (*recall*, *precision* dan *f-measure*). Sedangkan pada bagian saran berisi tentang hal-hal yang belum bisa terselesaikan pada penelitian yang telah dilakukan, serta hal-hal yang bisa dilakukan untuk pengembangan selanjutnya.