BAB 1
PENDAHULUAN

1.1. Latar Belakang Masalah

Peringkasan teks otomatis atau *automatic text summarization* merupakan bagian dari NLP (*natural language processing*) yang memungkinkan sebuah aplikasi dapat meringkas secara otomatis. Peringkasan teks otomatis dapat dilakukan dengan metode perangkingan terhadap teks dokumen. Perangkingan dilakukan sebagai proses penyeleksian kalimat pada teks dengan cara memberikan bobot nilai pada kalimat-kalimat tersebut sehingga kalimat yang mengandung informasi penting akan menempati urutan teratas dengan nilai bobot tertinggi dan nilai dipilih untuk dimasukkan sebagai kandidat kalimat ringkasan [1].

Pada penelitian [2] dilakukan peringkasan teks berbahasa Indonesia pada multi-dokumen dilakukan dengan cara mengklasifikasikan kalimat-kalimat dalam setiap artikel ke dalam kelas relevan (termasuk ke dalam ringkasan) dan kelas non-relevan (tidak termasuk ke dalam ringkasan) menggunakan SVM *classifier*. Selanjutnya dipilih *n* kalimat dari kelas positif berdasarkan nilai relevansinya. Nilai relevansi dapat berupa nilai probabilitas kemunculan kelas relevan atau nilai *marginal relevance* yang dihasilkan melalui MMR. Penelitian tersebut berhasil dilakukan dengan nilai parameter terbaik pada parameter C = 4.0 dan γ = 512.0, dan nilai *f-measure* yang diperoleh adalah sebesar 0.771 [2].

Selain menggunakan metode *maximum marginal relevance* untuk menghitung bobot akhir, ada salah satu metode yang juga bisa digunakan untuk pembobotan akhir salah satunya adalah metode *maximal marginal importance* (MMI). Penelitian [1] menggunakan metode *maximal marginal importance* (MMI) yang merupakan modifikasi dari metode *maximum marginal relevance* (MMR), dalam penelitian tersebut mengatakan bahwa metode *maximal marginal importance* dipilih karena melihat kekurangan pada *maximum marginal relevance* yang merupakan metode *query based* dimana dalam menentukan tingkat pentingnya suatu kalimat adalah dengan melihat kemiripan antara kalimat tersebut

dengan *query* yang diinputkan user. Sedangkan *maximal marginal importance* merupakan metode peringkasan yang general yang menentukan tingkat pentingnya suatu kalimat dengan melakukan penilaian fitur yang dikandung kalimat itu sendiri. Penelitian tersebut berhasil dilakukan dan menghasilkan nilai *threshold* sebesar 0.1.

Dari penelitian di atas menunjukkan bahwa metode *maximal marginal importance* dinilai cukup baik untuk melakukan pembobotan akhir pada proses peringkasan teks otomatis. Oleh karena itu, pada penelitian ini akan digunakan metode *support vector machine* dan *maximal marginal importance* untuk melakukan peringkasan teks otomatis pada artikel berita berbahasa Indonesia. Dengan adanya peringkas teks otomatis diharapkan dapat membantu pembaca dengan cepat dan mudah memahami makna dari beberapa artikel yang memiliki topik berita yang sama untuk mendapatkan intisari dari berita tersebut. Hasil ringkasan nantinya diharapkan dapat memberikan informasi baik.

1.2. Rumusan Masalah

Berdasarkan uraian pada latar belakang masalah pada penelitian ini terdapat permasalahan apakah penggunaan metode *support vector machine* (SVM) dan *maximal marginal importance* (MMI) pada peringkasan teks otomatis untuk multi dokumen berbahasa Indonesia bisa menghasilkan ringkasan yang baik?

1.3. Maksud dan Tujuan

1.3.1. Maksud

Adapun maksud dari penelitian ini adalah mengimplementasikan metode *support vector machine* (SVM) dan *maximal marginal importance* (MMI) untuk peringkasan teks otomatis multi-dokumen berbahasa Indonesia.

1.3.2. Tujuan

Sedangkan tujuan yang ingin dicapai dari penelitian ini adalah untuk mengetahui seberapa baik penggunaan metode *support vector machine* (SVM) dan *maximal marginal importance* (MMI) untuk menghasilkan ringkasan pada multi dokumen artikel berita berbahasa Indonesia.

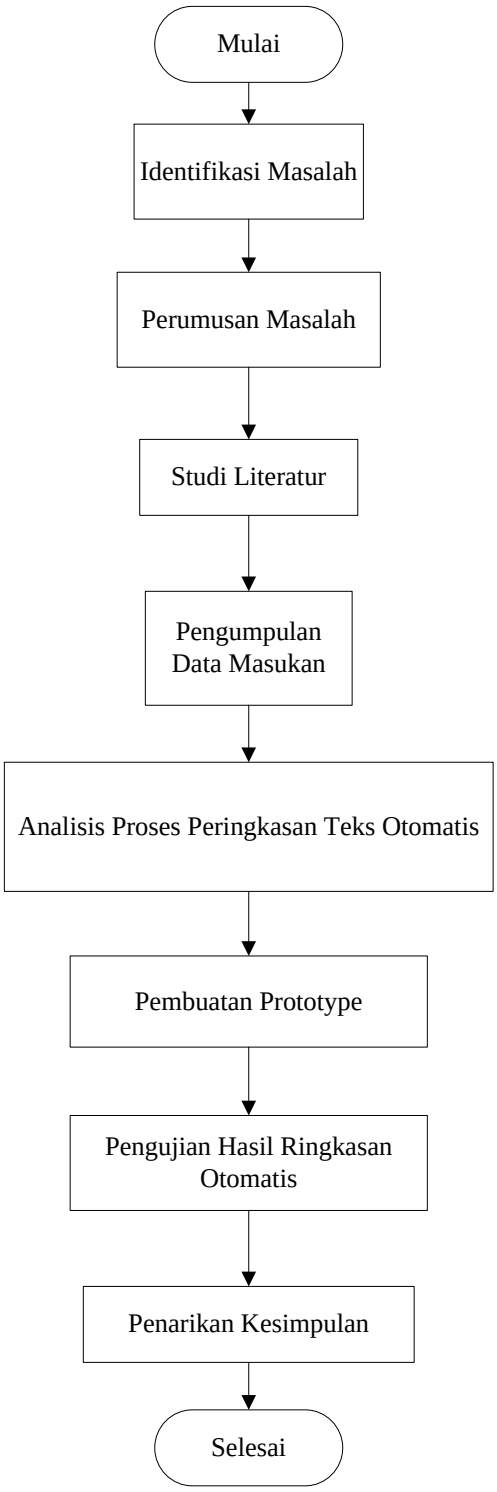
1.4. Batasan Masalah

Agar penelitian yang akan dilakukan lebih terarah dan sesuai dengan tujuan yang diinginkan, maka terdapat batasan masalah. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Sumber berita yang diambil berasal dari beberapa portal pemuat berita seperti detik, kompas, tempo, tribunnews, liputan6, dan bbc news.
2. Berita yang diambil adalah berita dengan beberapa topik, seperti olahraga, hukum, kriminal, dan teknologi.
3. Bentuk masukan teks ke sistem berupa teks dengan format file .txt.
4. Dokumen masukan ke dalam sistem yang dibuat adalah satu atau lebih dokumen (multi-dokumen) berita berbahasa Indonesia.
5. Peringkasan teks dilakukan secara *offline*.
6. Peringkas teks otomatis yang dibangun berbasis desktop dengan menggunakan bahasa pemrograman Java.

1.5. Metodologi Penelitian

Metodologi penelitian yang dilakukan dalam penelitian ini adalah metode kuantitif. Metode kuanttaif digunakan karena hasi akhir dari penelitin ini berupa persentase nilai performansi dari ringkasan yang dihasilkan sistem. Langkah-langkah penelitian dapat dilihat pada gambar 1.1 berikut:



Gambar 1. 1 Langkah-Langkah Penelitian

Berdasarkan alur penelitian pada gambar 1.1 dapat di jelaskan sebagai berikut:

1. Identifikasi Masalah

Tahap pertama dalam penelitian adalah mengidentifikasi masalah. Peneliti mengidentifikasi masalah ataupun kendala pada proses peringkasan teks otomatis.

2. Perumusan Masalah

Tahapan kedua adalah perumusan masalah. Peneliti merumuskan solusi dari identifikasi masalah yang telah diidentifkasi pada tahap sebelumnya.

3. Studi Literatur

Tahapan ini dilakukan untuk melakukan pengumpulan berbagai informasi mengenai penelitian dengan cara mengumpulkan literatur, jurnal, buku, *e-book*, paper dan bacaan yang berkaitan dengan topik penelitian yang akan dikerjakan.

4. Pengumpulan Data Masukan

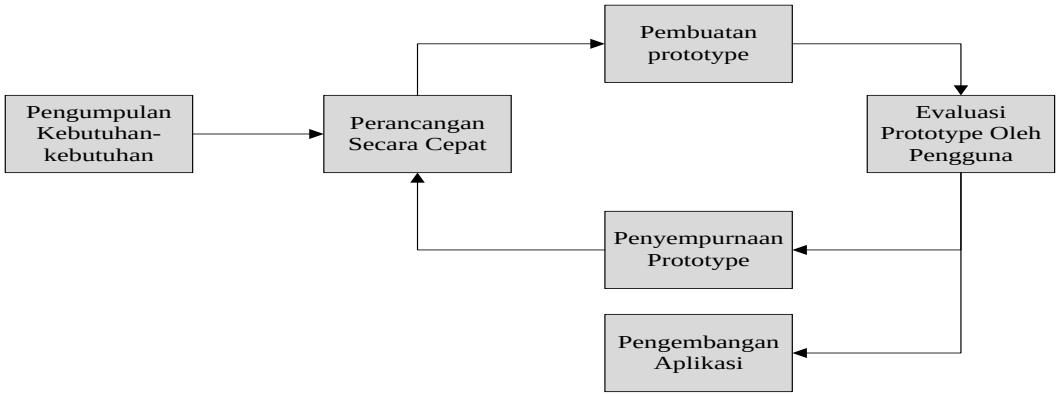
Pengumpulan data masukan dilakukan secara manual. Dimana data yang diambil merupakan berita dari portal pemuat berita seperti merdeka, republika, liputan6, kompas, viva, tribunnews dan tempo. Data akan dipilih secara acak sesuai dengan topik, disalin dan kemudian disimpan dalam format file .txt. Data masukan pada penelitian ini terdiri dari pengumpulan data masukan untuk proses *training* (pelatihan) dan pengumpulan data masukan untuk proses *testing* (pengujian).

5. Analisis Proses Peringkasan Teks otomatis

Pada tahap ini dilakukan analisis mengenai peringkasan teks otomatis. Analisis yang dilakukan meliputi analisis masalah, analisis data masukan, analisis proses preprocessing, analisis proses ekstraksi fitur, serta analisis metode yang dipakai pada penelitian ini, yaitu analisis klasifikasi kalimat menggunakan metode *support vector machine* (SVM) dan analisis pembentukan kalimat menggunakan metode *maximal marginal importance* (MMI).

6. Pembuatan Prototype

Dalam penelitian ini, metode pembangunan perangkat lunak yang digunakan adalah model *prototype* [3]. Model *prototype* yang digunakan dapat dilihat pada gambar 1.2. berikut:



Gambar 1. 2 Metode Protoype [3]

Berikut adalah langkah-langkah yang dilakukan dalam proses model *prototype*:

a. Pengumpulan Kebutuhan-Kebutuhan

Tahap ini merupakan tahap awal dalam Model Prototype. Kegiatan yang dilakukan yaitu menganalisis dan mengumpulkan kebutuhan-kebutuhan, meliputi analisis masalah, analisis data masukan, analisis *preprocessing*, analisis ekstraksi fitur dan analisis metode yang meliputi metode *support vector machine* dan *maximal marginal importance*. Pada tahap ini juga dilakukan analisis kebutuhan fungsional serta kebutuhan non fungsional.

b. Perancangan Secara Cepat

Pada tahap ini melakukan perancangan awal dari aplikasi yang akan dibangun, seperti perancangan antar muka dan basis data.

c. Pembuatan *Prototype*

Tahap ini termasuk ke dalam tahap implementasi dari desain yang sudah dibuat pada tahap perancangan cepat. Pada tahap ini implementasi dari proses perancangan cepat ke dalam bahasa pemrograman, yaitu Java.

d. Evaluasi *Prototype* Oleh Pengguna

Pada tahap ini melakukan pengujian terhadap aplikasi yang sudah dibuat. Melihat apakah metode yang diterapkan pada tahap *preprocessing*, tahap ekstraksi fitur, tahap klasifikasi kalimat dan pembentukan ringkasan sudah berjalan dengan baik serta apakah ada fungsionalitas program yang tidak berjalan dengan baik.

e. Penyempurnaan *Prototype*

Tahap ini merupakan tahap memperbaiki aplikasi dari segi fungsionalitas dan juga metode-metode yang digunakan belum berjalan dengan baik pada tahap evaluasi prototype yang sudah dilakukan.

f. Pengembangan Aplikasi

Tahap ini merupakan tahap akhir dimana fungsionalitas aplikasi dan metode-metode yang diterapkan sudah berjalan sesuai dengan tujuan yang diharapkan. Sampai tahap ini, program sudah bisa melakukan peringkasan teks otomatis.

7. Pengujian Hasil Ringkasan Otomatis

Pada tahap ini dilakukan pengujian terhadap hasil ringkasan yang dihasilkan tioleh sistem. Pengujian dilakukan dengan cara membandingkan hasil ringkasan yang dihasilkan oleh sistem dengan hasil ringkasan manual yang dibuat oleh *abstractor*.

8. Penarikan Kesimpulan

Tahapan akhir ini adalah menarik kesimpulan mengenai seberapa baik penggunaan metode yang telah diimplementasikan untuk melakukan peringkasan teks otomatis.

1.6. Sistematika Penulisan

Sistematika penulisan laporan penelitian skripsi ini disusun untuk memberikan gambaran umum mengenai penelitian yang dilakukan. Sistematika penulisan laporan penelitian ini adalah sebagai berikut :

BAB 1 PENDAHULUAN

Bab ini berisi uraian tentang latar belakang masalah, perumusan masalah, maksud dan tujuan penelitian, batasan masalah, metodologi penelitian yang meliputi studi literatur, metode pengumpulan data, tahap analisis dan kebutuhan sistem dan metode pembangunan perangkat lunak, serta sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini membahas berbagai konsep dasar dan teori yang berkaitan dengan topik penelitian seperti definisi peringkas teks otomatis, metode *support vector machine* (SVM), metode *maximal marginal importance* (MMI), tahap *preprocessing*, tahap ekstraksi fitur, serta perhitungan nilai performansi yaitu *recall*, *precision* dan *f-measure*. Selain itu dijelaskan juga mengenai teori perancangan aplikasi yang akan dibangun menggunakan pemodelan UML seperti *use case diagram, class diagram*, *activity diagram, sequence diagram* serta teori-teori tentang *tools* atau *software* pendukung yang digunakan dalam pembangunan peringkas teks otomatis ini.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini berisi tentang analisis masalah dari objek penelitian yang dilakukan untuk mengetahui masalah apa yang timbul, analisis sistem yang akan dibangun, analisis data masukan, analisis tahap *preprocessing*, analisis tahap ekstraksi fitur, analisis metode yang berupa analisis klasifikasi kalimat menggunakan SVM dan analisis pembentukan ringkasan dengan menggunakan metode MMI. Selain itu dilakukan analisis evaluasi ringkasan yang dihasilkan dengan nilai performansi (*recall*, *precission* dan *f-measure*). Pada bab ini dijelaskan juga mengenai perancangan tampilan sistem dan perancangan antarmuka dari aplikasi yang akan dibangun .

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang implementasi dari analisis dan perancangan yang telah dilakukan pada bab sebelumnya. Implementasi yang dilakukan terdiri dari implementasi sistem dan implementasi antar muka. Selain itu pada tahap ini dilakukan pengujian terhadap aplikasi yang dibangun serta melakukan evaluasi terhadap hasil rangkuman melalui nilai performansi *recall*, *precision* dan *f-measure*.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi 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 hasil pengujian mengenai seberapa baik ringkasan yang dihasilkan dari peringkas teks otomatis, serta menampilkan nilai performansi *recall*, *precision* dan *f-measure*.