BAB 1
PENDAHULUAN

1.1.Latar Belakang Masalah

Perkembangan teknologi semakin pesat menyebabkan kebutuhan akan informasi yang sangat besar dan tidak terbatas, terutama informasi dalam bentuk jurnal, artikel, dan berita. Artikel yang ada pada situs portal berita umumnya bersifat *real time* dan *up to date* yang menyebabkan artikel berita memiliki jumlah kalimat/paragraf yang sangat banyak[4]. Dibutuhkan waktu yang lama untuk membaca secara keseluruhan untuk dapat memehami dan mengetahui isi dari berita tersebut. Sehingga diperlukan sebuah sistem peringkasan teks otomatis. Dengan adanya sistem peringkasan teks otomatis diharapkan pembaca akan lebih mudah untuk memahami dan mengetahui isi dari berita tersebut.

Banyak metode penelitian yang membahas tentang peringkasan teks otomatis salah satunya, yaitu Rendra Pratama pada tahun 2014. Rendra Pratama menggunakan *K-means clustering* untuk meringkas dokumen tunggal berita artikel dalam bahasa Indonesia. Pada penelitiannya jumlah kluster yang sudah ditetapkan atau konstan, sedangkan jumlah kluster sangat berpengaruh terhadap ringkasan[13]. Pada penelitian tersebut pemilihan kalimat penting berdasarkan bobot kalimat paling besar pada tiap cluster dengan mengambil *compression* sebesar 25% dan 40%. Pengujian hasil ringkasan sistem pada *compression* 40% menghasilkan rata-rata recall 60%, precision 62%, dan f-measure 61%. Sedangkan untuk *compression* 25% menghasilkan rata-rata recall 37%, precision 44%, dan f-measure 40%. Penilaian hasil ringkasan sistem oleh expert judgement pada *compression* 40% menghasilkan nilai rata-rata 3,15 atau pada range diterima, sedangkan pada *compression* 25% menghasilkan nilai rata-rata 1,83 atau berada pada range cukup diterima[18]. Dalam hal ini seberapa besar tingkat akurasi apabila tingka tidak menggunakan tingkat *compression* dan penentuan jumlah *cluster* ditentukan berdasarkan jumlah kalimat[13].

Berdasarkan pada hal tersebut, maka penulis dalam penelitian ini akan membuat *Text Summarization* dengan metode *k-means* pada artikel berita berbahasa Indonesia. Hasil ringkasan nantinya diharapkan dapat memberikan informasi yang lebih akurat.

1.2. Identifikasi Masalah

Berdasarkan uraian pada latar belakang masalah diatas, identifikasi masalah dalam penelitian ini adalah seberapa besar tingkat akurasi ringkasan yang dihasilkan sistem peringkasan teks pada artikel berita berbahasa Indonesia jika tingkat *compression* dan penentuan jumlah *cluster* disesuaikan berdasarkan pada jumlah kalimat.

1.3. Maksud dan Tujuan

Penelitian ini bermaksud membuat sistem peringkas teks pada artikel berita berbahasa Indonesia. Sedangkan tujuan yang ingin dicapai dari penelitian ini adalah mengetahui tingkat akurasi pada sistem peringkasan teks otomatis dengan metode *K-Means* pada artikel berita berbahasa Indonesia.

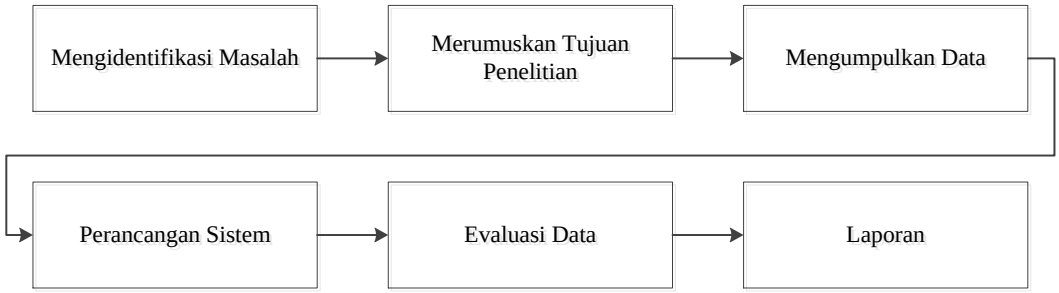
1.4. Batasan Masalah

Agar penelitian ini tidak menyimpang dari latar belakang dan tujuan yang akan dicapai, maka terdapat batasan masalah sebagai berikut.

1. Dokumen yang digunakan sebagai inputan berupa dokumen berita berbahasa Indonesia dengan ekstensi .txt dan dokumen tunggal berita berbahasa Indonesia.
2. Proses preprocessing yang dilakukan dalam penelitian ini terdiri dari *Token* Kalimat, *casefolding, Filtering, Tokenizing* dan *Stopword Removal*.
3. Output yang dihasilkan dari sistem peringkasan teks berupa ringkasan teks.

1.5. Metodologi Penelitian

Metode penelitian implementasi *text suummarization* dengan metode *K-Means* ini menggunakan metode deskriptif. Metode deskriptif adalah metode yang digunakan untuk menggambarkan atau menganalisis suatu hasil penelitian tetapi tidak digunakan untuk membuat kesimpulan yang lebih [15]. Metode penelitian meliputi metode pengumpulan data dan metode pembangunan perangkat lunak. Adapun skema dalam penelitian ini, untuk lebih jelasnya lihat gambar 1.1 berikut.



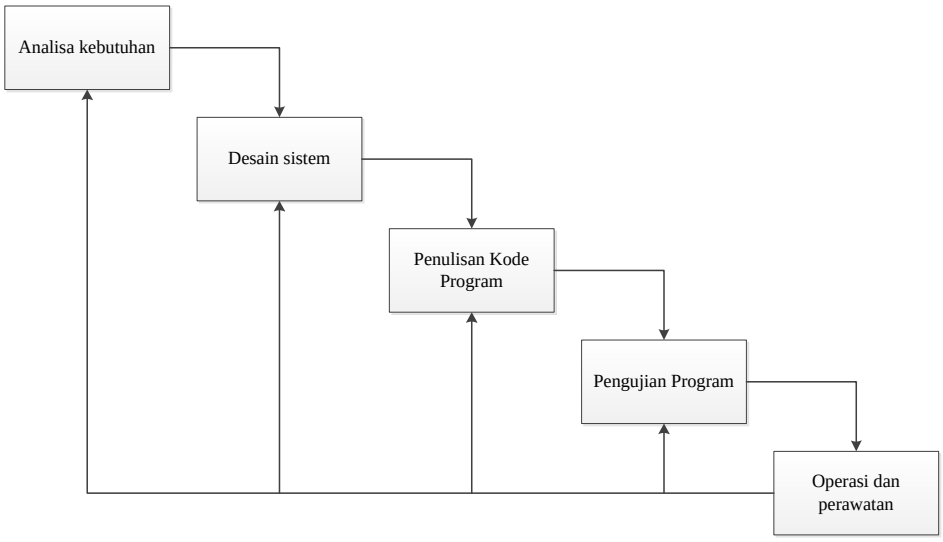
Gambar 1. 1 Skema Penelitian

1.5.1. Metode Pengumpulan Data

Dalam penelitian ini, penulis melakukan beberapa metode untuk memperoleh data atau informasi dalam menyelesaikan permasalahan. Metode yang dilakukan tersebut adalah Studi literatur. Studi literatur dilakukan dengan mempelajari teori-teori melalui buku, artikel, jurnal dan bahan lain yang berkaitan dengan metode tahap yang digunakan untuk meringkas teks dan algoritma *K-Means.*

1.5.2. Metode Pembangunan Perangkat Lunak

Metode pembangunan perngkat lunak yang digunakan untuk implementasi text summarization artikel berita berbahasa indonesia menggunakan model *waterfall sommervile*[3] yang melakukan pendekatan secara sistematis dan berurutan. Berikut alur proses model *waterfall sommerfille.*



Gambar 1. 2 Diagram *Waterfall Sommerfille[3]*

1. Analisa kebutuhan

   Pada tahap ini dilakukan pengumpulan kebutuhan penelitian secara lengkap mengenai peringkasan teks menggunakan motode k-means. Data yang digunakan dalam format .txt.

2. Desain sistem

   Setelah kebutuhan data telah dikumpulkan, tahap selanjutnya adalah perancangan sistem secara keseluruhan. Untuk kebutuhan perangkat lunak dalam pengolahan data diperlukan software yang berjalan pada operasi sistem windows dalam hal ini peneliti menggunakan bahasa pemrograman PHP. Yang akan berjalan disebuah halaman website.

3. Penulisan kode program

   Setelah perancangan sistem secara keseluruhan, maka tahap selanjutnya adalah mengimplementasikan ke dalam bahasa pemrograman. Tahap preprocessing(pemisahan kalimat, *casefolding, filtering, tokenisasi* dan *stopword*) yang kemudian dilakukan pembobotan tiap kata menggunakan TF-

IDF. Hasil dari TF-IDF ini kemudian akan dilakukan *clustering* dengan metode *K-Means*.

4. Pengujian program

   Tahapan ini digunakan untuk mengintegrasikan program, apabila program yang dibuat berjalansesuai dengan harapan maka text summarization siap digunakan oleh pengguna sistem.

5. Operasi dan perawatan

   Pada tahapan ini sistem sudah siap untuk digunakan, lalu dilakukan pengembangan sistem seperti penambahan fitur dan fungsi baru yang mungkin akan muncul untuk kebutuhan pengguna.

1.6. Sistematika Penulisan

Sistematika penulisan pada penelitian ini disusun untuk memberikan gambaran umum tentang kasus yang akan dilakukan. Sistematika penulisan tugas akhir ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini menerangkan secara umum mengenai latar belakang, rumusan masalah, menentukan maksud dan tujuan, batasan masalah, metodologi penelitian, metode pembangunan perangkat lunak, serta sistematika penulisan penelitian *text summarization* artikel berita berbahasa indonesia.

BAB 2 LANDASAN TEORI

Pada bab ini membahas berbagai konsep dasar dan teori-teori yang berkaitan dengan topik penelitian tugas akhir. Pembahasan dimulai dari penjelasan berita, jenis-jenis berita. Dilanjutkan dengan peringkasan teks dan metode yang digunakan.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini berisi mengenai perancangan untuk melakukan *text summarization* artikel berita berbahasa indonesia. Perancangan *text summarization* meliputi persiapan dokumen artikel, tahap *preprocessing* (pemisahan kalimat,*casefolding*,

filtering, *tokenizing*, *stopword removal*), pembobotan kata dengan TF-IDF dan *clustering* dengan metode *k-means.*

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Pada bab ini berisi tentang tahapan yang dilakukan dalam penelitian secara garis besar dimulai dari tahapan persiapan sampai pada penarikan kesimpulan, pada tahap ini *text summarization* akan diuji dengan menggunakan data testing.

BAB 5 KESIMPULAN DAN SARAN

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