

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Peringkasan pada teks dapat digunakan untuk mempermudah dalam penyampaian informasi yang dibutuhkan tanpa menghilangkan esensi dari informasi pada teks tersebut. Untuk membuat sebuah ringkasan pada suatu teks, dibutuhkan waktu dan tenaga apabila dilakukan secara manual. Sehingga diperlukan sebuah sistem yang dapat melakukan peringkasan teks dokumen secara otomatis agar pengerjaan dapat lebih efisien [1]. Peringkasan teks otomatis adalah proses pengurangan jumlah teks yang ada pada dokumen menggunakan *program* komputer untuk bisa menghasilkan sebuah ringkasan yang berisikan poin – poin penting sebagai intisari yang dimana hasil ringkasan tidak lebih dari dokumen yang asli [2].

Ada dua pendekatan dalam melakukan proses peringkasan teks, yaitu dengan menggunakan metode ekstraktif dan abstraktif. Metode ekstraktif adalah teknik penyusunan kalimat dengan mengambil kalimat – kalimat penting yang terdapat pada dokumen asli dan menggabungkannya menjadi dokumen yang lebih pendek. Sedangkan, metode abstraksi adalah teknik penyusunan kalimat dengan cara mengambil kalimat – kalimat penting pada dokumen asli lalu membuatnya dalam bentuk kalimat lain untuk dijadikan ringkasan. Metode ekstraktif memuat beberapa bagian seperti kalimat, frasa yang terdiri dari potongan – potongan teks yang akan membentuk sebuah ringkasan. Oleh karena itu, mengidentifikasi kalimat yang tepat untuk peringkasan merupakan faktor yang paling penting dalam metode ekstraktif [3].

Pada penelitian yang dilakukan oleh Achmad Ridok, dengan menggunakan metode Non-Negative Matrix Factorization (NMF) didapatkan nilai presisi dan recall dibawah 0,2. Namun, peringkasan dokumen secara otomatis menggunakan NMF, hasilnya kurang memuaskan. Hal ini ditandai dengan rendahnya nilai rata – rata presisi dan recall. Penyebab terjadinya hal ini adalah sumber data acuan

sebagai *benchmark* yaitu hasil ringkasan pakar mempunyai nilai rata – rata presisi 0.68667 dan recall 0.70642 [4].

Penelitian yang dilakukan oleh Bening Suryani Pratiwi, Shaufiah, Moch. Arif Bijaksana menyatakan bahwa hasil evaluasi didapat untuk hasil ringkasan artikel CNN Corpus bahasa Inggris mencapai hasil pada *compression rate* 10% - 20% dan nilai parameter  $\lambda = 0.3$  dengan nilai *recall* 0.383, *precision* 0.212, dan *f-measure* 0.257 untuk peringkasan tanpa menggunakan Word Sense Disambiguation. Penggunaan Maximal Marginal Relevance dan Word Sense Disambiguation sudah cukup baik untuk meminimalisirkan redundansi dan membantu mengurangi ambiguitas kata. Namun, hasil ringkasan tidak sebaik hasil ringkasan tanpa Word Sense Disambiguation. Ini disebabkan karena setiap term yang diganti menyesuaikan dengan frekuensi yang diambil belum tentu sesuai dengan hasil ringkasan yang dihasilkan [1].

Penelitian ini mengusulkan metode peringkasan ekstraktif dokumen yang dapat meningkatkan relevansi antar kalimat dengan menggunakan metode *non-negative matrix factorization* dan *word sense disambiguation*. Penggunaan *word sense disambiguation* digunakan untuk mengatasi kata – kata yang serupa namun sebenarnya memiliki makna yang berbeda, sehingga sedikit sulit untuk dipahami. Maka dari itu, dibutuhkan sebuah solusi untuk menentukan makna pada kalimat tersebut agar lebih dimengerti oleh pembaca.

Berdasarkan penelitian pada peringkasan teks sebelumnya, maka dapat disimpulkan bahwa peringkasan menggunakan kedua metode tersebut sudah cukup baik, tapi masih ada beberapa kekurangannya. Oleh karena itu, akan dilakukan pengkombinasian antara kedua metode tersebut dengan judul “Peringkasan Teks Otomatis Menggunakan Metode Non-Negative Matrix Factorization dan Word Sense Disambiguation Pada Berita Berbahasa Indonesia”.

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang masalah yang telah dijelaskan, maka perumusan masalah yang ada adalah bagaimana mengimplementasikan *Non-Negative Matrix Factorization* dan *Word Sense Disambiguation* pada peringkasan

teks otomatis serta mengukur akurasi dari hasil ringkasan yang dihasilkan pada artikel berita berbahasa Indonesia.

### **1.3 Maksud dan Tujuan**

Maksud dari penelitian ini adalah mengimplementasikan *non-negative matrix factorization* dan *word sense disambiguation* dalam peringkasan teks otomatis pada artikel berita berbahasa Indonesia. Tujuan yang ingin dicapai pada penelitian ini adalah mengukur akurasi dari hasil ringkasan teks otomatis untuk artikel berita berbahasa Indonesia dengan menggunakan *non-negative matrix factorization* dan *word sense disambiguation*.

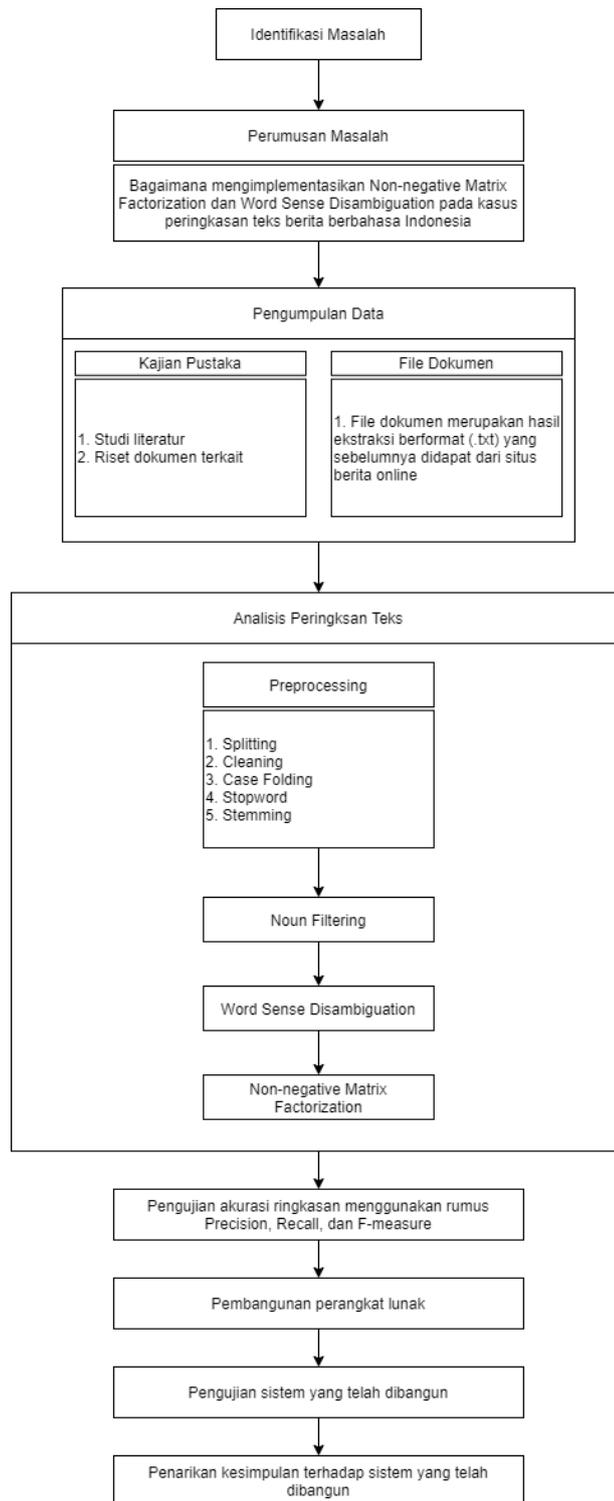
### **1.4 Batasan Masalah**

Batasan masalah dimaksudkan untuk menghindari penyimpangan dan kesalahan dari permasalahan utama serta tujuan yang dicapai. Batasan masalah yang terdapat dalam penelitian ini antara lain:

1. Peringkasan yang dilakukan adalah peringkasan terhadap artikel berita berbahasa Indonesia.
2. Metode peringkasan yang digunakan adalah metode peringkasan ekstraksi.
3. Dokumen sumber masukkan berformat .txt.
4. Dokumen yang digunakan adalah dokumen tunggal.
5. Bahasa yang digunakan adalah bahasa formal, bukan bahasa sehari – hari atau daerah.

### **1.5 Metode Penelitian**

Penelitian ini menggunakan metode kuantitatif. Metode kuantitatif merupakan metode yang bertujuan untuk melihat hasil akhir dari penelitian yang berupa persentase nilai performa dari ringkasan teks yang dihasilkan oleh sistem [5]. Tahapan metode penelitian ini dapat dilihat pada gambar 1.1 sebagai berikut:



**Gambar 1.1 Alur Penelitian**

Berdasarkan alur penelitian pada gambar 1.1 dapat dijelaskan sebagai berikut:

### **1. Identifikasi Masalah**

Tahapan pertama dalam penelitian adalah mengidentifikasi masalah. Penelitian mengidentifikasi masalah ataupun kendala pada proses peringkasan teks dan analisis kebutuhan algoritma.

### **2. Perumusan Masalah**

Tahapan kedua adalah perumusan masalah. Pada tahapan ini, peneliti merumuskan masalah bagaimana mengimplementasikan *Word Sense Disambiguation* dan *Non-negative matrix factorization* pada kasus peringkasan teks.

### **3. Pengumpulan Data**

Tahapan ketiga adalah pengumpulan data. Metode pengumpulan data yang dilakukan pada penelitian ini adalah sebagai berikut:

#### **a. Kajian Pustaka**

Kajian pustaka dilakukan dengan mempelajari jurnal, *paper*, buku dan situs *online* yang terkait dengan topik peringkasan teks, tahapan *preprocessing*, *Noun Filtering*, *Word Sense Disambiguation*, *Non-negative Matrix Factorization* dan pengujian akurasi menggunakan rumus *precision*, *recall* dan *f-measure*.

#### **b. File Dokumen**

File dokumen diambil dari beberapa situs berita *online* yang sudah terkenal, file tersebut kemudian diolah ke dalam bentuk format (.txt).

### **4. Analisis Peringkasan Teks**

Tahapan ini berupa analisis peringkasan teks. Adapapun analisis yang dilakukan dalam penelitian ini yaitu:

#### **a. Preprocessing**

*Preprocessing* merupakan tahapan untuk mempersiapkan teks menjadi data yang akan diolah pada tahapan berikutnya. *Preprocessing* pada penelitian ini terdiri dari beberapa tahapan yaitu proses *splitting*, *cleaning*, *case folding*, *stopword*, dan *stemming*.

**b. Noun Filtering**

Dalam penelitian ini, kata yang berambigu akan dibatasi hanya kata benda saja. Sebelum memasuki proses pendeteksian kata ambigu, terdapat proses penyaringan kata benda dari kumpulan data hasil *preprocessing* yang kemudian nantinya akan dianggap sebagai input dalam membentuk *non-negative matrix factorization*. Penyaringan tersebut dilakukan dengan cara melakukan pengecekan terhadap setiap kata dengan *database* kata benda sebagai acuan untuk mendapatkan kata benda.

**c. Word Sense Disambiguation**

Pada tahap ini, dilakukan identifikasi makna kata dengan menggunakan algoritma *lesk* yang bekerja dengan cara membandingkan definisi dari kata yang berambigu dengan definisi dari kata tetangganya berdasarkan definisi dari kamus kata.

**d. Non-negative Matrix Factorization**

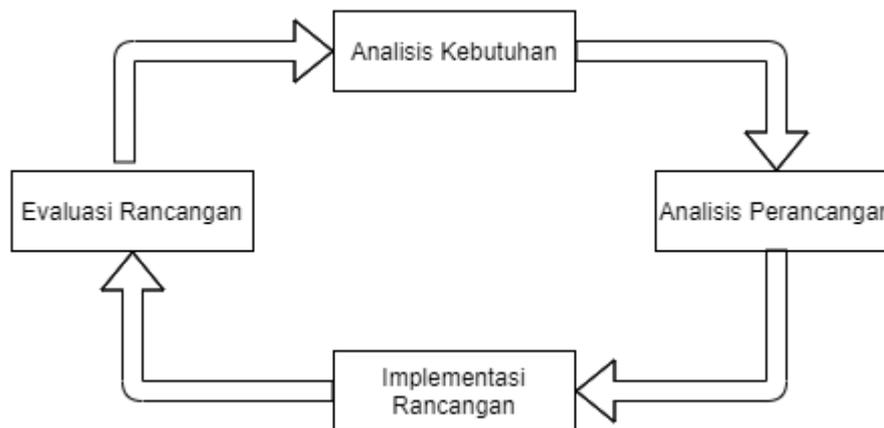
Merupakan tahap terakhir. Pada tahapan ini, hasil dari makna kata yang telah dibandingkan akan dipecah menjadi kalimat – kalimat dan menghitung frekuensi dari masing – masing term dalam kalimat yang direpresentasikan dengan bobot nilai.

**5. Pengujian Akurasi Ringkasan**

Tahapan ini merupakan pengujian akurasi hasil ringkasan menggunakan rumus *precision*, *recall* dan *f-measure*.

**6. Pembangunan Perangkat Lunak**

Tahapan ini, merupakan pembangunan perangkat lunak yang menggunakan metode *prototype*, karena metode ini merupakan tahap awal membangun suatu perangkat lunak yang digunakan untuk menunjukkan konsep, serta untuk mengetahui masalah dan solusi yang mungkin dapat untuk memperbaiki masalah tersebut dalam membangun perangkat lunak [6]. Adapun tahap proses dari *prototype* dapat dilihat pada gambar 1.2 berikut.



**Gambar 1.2 Model Prototype**

**a. Analisis Kebutuhan**

Tahapan ini, dilakukan analisis mengenai kebutuhan yang diperlukan dalam peringkasan teks. Analisis yang dilakukan berdasarkan hasil studi literatur yang telah dilakukan.

**b. Analisis Perancangan**

Tahapan ini, merupakan proses selanjutnya untuk menentukan rancangan pembangunan perangkat lunak yang akan dibangun dalam peringkasan teks.

**c. Implementasi Rancangan**

Tahapan ini, merupakan tahapan implementasi kedalam bahasa pemrograman sesuai dengan perancangan yang telah dibuat dalam analisis perancangan.

**d. Evaluasi Rancangan**

Tahapan ini, merupakan tahapan evaluasi pada *prototype* peringkasan teks yang dibangun sesuai atau masih ada kekurangan sesuai analisis kebutuhan. Sekaligus melakukan pengecekan perangkat lunak apabila dibutuhkan perbaikan karena kesalahan *error*, kembali ketahap awal untuk melakukan perbaikan sesuai dengan yang diharapkan dalam penelitian peringkasan teks otomatis. Pada tahapan ini juga, dilakukan pengujian hasil ringkasan teks yang dilakukan dengan menggunakan standar pengukuran akurasi yang biasa digunakan dalam penelitian peringkasan teks diantaranya *recall*, *precision*, dan *f-measure*.

**7. Pengujian Sistem**

Tahapan ini dilakukan pengujian sistem yang telah dibangun dengan mengecek semua fungsi yang ada apakah sudah berjalan dengan baik atau tidak.

## **8. Penarikan Kesimpulan**

Tahapan ini, merupakan penarikan kesimpulan dan saran mengenai seberapa baik penggunaan metode yang diterapkan untuk menghasilkan ringkasan teks otomatis.

### **1.6 Sistematika Penulisan**

Sistematika penulisan tugas akhir penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan tugas akhir ini adalah sebagai berikut:

#### **BAB 1 PENDAHULUAN**

Bab ini berisi penjelasan mengenai latar belakang masalah, rumusan masalah, maksud dan tujuan, batasan masalah, metode penelitian serta sistematika penulisan. Bertujuan agar dapat memberikan gambaran tentang urutan pemahaman dalam menyajikan laporan ini.

#### **BAB 2 LANDASAN TEORI**

Bab ini membahas mengenai landasan teori yang digunakan untuk menganalisis masalah dan teori yang dipakai dalam data penelitian yaitu teori mengenai *peringkasan teks otomatis*, algoritma *non-negative matrix factorization*, *word sense disambiguation*, bahasa pemrograman yang digunakan.

#### **BAB 3 ANALISIS DAN PERANCANGAN**

Bab ini menjabarkan penjelasan analisis dan perancangan sistem, perhitungan pada algoritma *non-negative matrix factorization* dan *word sense disambiguation* yang nantinya akan diaplikasikan pada peringkasan teks.

#### **BAB 4 IMPLEMENTASI DAN PENGUJIAN**

Bab ini berisi tentang pembangunan perangkat lunak, pengujian sistem dan pengujian akurasi algoritma *non-negative matrix factorization* dan *word sense disambiguation*.

#### **BAB 5 KESIMPULAN DAN SARAN**

Bab ini menjelaskan tentang kesimpulan yang diperoleh berdasarkan hasil perhitungan pada peringkasan dokumen menggunakan algoritma *word sense*

disambiguation serta saran – saran untuk pengembangan penelitian ini selanjutnya.