

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Peringkasan teks merupakan proses pengambilan informasi penting yang berasal dari satu atau beberapa sumber yang kemudian menghasilkan sebuah ringkasan dengan panjang teks kurang dari 50% dari teks aslinya. Peringkasan teks dapat dilakukan dengan menggunakan teknik abstraktif dan teknik ekstraktif [1]. Teknik abstraktif adalah teknik peringkasan dengan membuat kalimat baru yang lebih singkat, dengan tidak menghilangkan maksud dari teks aslinya. Sedangkan teknik ekstraktif adalah teknik peringkasan dengan memilih sebagian kalimat tertentu dari kalimat yang ada dalam teks aslinya. Pemilihan kalimat dalam teknik ekstraktif, berdasarkan pada pemberian nilai bobot pada kalimat. Pemberian nilai bobot kalimat dilakukan dengan menggunakan fitur tertentu [2].

Penelitian terhadap peringkasan teks berbahasa Indonesia [3] menggunakan metode *Support Vector Machine* dengan 6 fitur kalimat yang digunakan adalah posisi kalimat, data numerik, tanda koma terbalik, panjang kalimat, kata kunci dan keseluruhan kalimat. Hasilnya diperoleh *recall* sebesar 42.8%, nilai *precision* sebesar 60.4%, dan nilai *f-measure* sebesar 49.6%. Penelitian lainnya juga menggunakan metode *Support Vector Machine* dalam peringkasan teks [4] dengan menggunakan 8 fitur kalimat yang diantaranya adalah panjang kalimat, posisi kalimat, data numerik, kata tematik dalam kalimat, kalimat yang menyerupai judul, kemiripan kalimat dengan kalimat lain, ikatan leksikal dengan kalimat sebelumnya dan ikatan leksikal dengan kalimat sesudahnya. Hasilnya didapat rata-rata dari nilai *recall* 55.2%, nilai *precision* 56%, dan nilai *f-measure* sebesar 53.65%. Berdasarkan penelitian tersebut bahwa metode *Support Vector Machine* terjadi penurunan *precision* yang disebabkan oleh perbedaan penggunaan fitur kalimat. Oleh karena itu dibutuhkan suatu cara penyeleksian fitur untuk mengoptimalkan penggunaan fitur kalimat secara tepat.

Seleksi fitur adalah proses pemilihan fitur yang memiliki posisi penting dalam suatu dimensi fitur. Proses pemilihan fitur, dilakukan dengan menghilangkan fitur yang dianggap tidak relevan, sehingga mengoptimalkan kinerja dari mesin pembelajaran [5]. Penelitian tentang seleksi fitur [6] dalam klasifikasi berita, menghasilkan bahwa metode *Particle Swarm Optimization* mendapatkan hasil terbaik yaitu rata-rata nilai *precision* 77.4619 dibandingkan dengan metode seleksi fitur lain yaitu *Information Gain*, *Chi-Square*, dan *Genetic Algorithm* dengan nilai rata-rata *precision* berturut-turut 70.3791, 72.204 dan 76.6511. Penelitian seleksi fitur *Particle Swarm Optimization* [7] juga menghasilkan peningkatan akurasi menjadi 95.42% lebih baik 29.61% dibandingkan dengan tanpa menggunakan seleksi fitur *Particle Swarm Optimization* dengan akurasi 65.81%. Berdasarkan dari penelitian tersebut bahwa seleksi fitur *Particle Swarm Optimization* memberikan peningkatan akurasi yang baik, maka penelitian ini akan menggunakan seleksi fitur *Particle Swarm Optimization* pada metode *Support Vector Machine* untuk mendapatkan kombinasi fitur kalimat yang tepat dalam kasus peringkasan teks berbahasa Indonesia.

## 1.2 Rumusan Masalah

Berdasarkan penjelasan yang ada pada latar belakang diatas, maka dapat dirumuskan permasalahannya yaitu fitur kalimat apa sajakah yang dapat mengoptimalkan kinerja *Support Vector Machine* dalam melakukan peringkasan teks otomatis.

## 1.3 Maksud dan Tujuan Penelitian

Maksud dari penelitian ini adalah mengimplementasikan *Particle Swarm Optimization* untuk pemilihan fitur dalam peringkasan teks dengan menggunakan metode *Support Vector Machine*. Sedangkan tujuan yang ingin dicapai dari penelitian ini adalah mengetahui fitur yang tepat untuk peringkasan teks menggunakan metode *Support Vector Machine*.

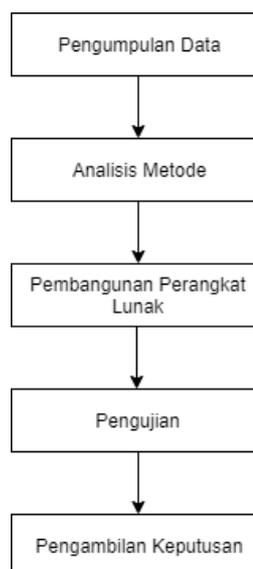
#### 1.4 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Dokumen yang digunakan dalam peringkasan adalah berita berbahasa Indonesia.
2. Dokumen berita diambil bersumber dari portal berita **www.detik.com** dan **www.cnnindonesia.com**.
3. Tema dari dokumen berita masukan yaitu berita tentang olahraga Sepakbola.
4. Format data masukan berupa .csv.
5. Jumlah data sebanyak 60 berita.
6. Total peringkasan yang dilakukan sebanyak 25% [8].
7. Fitur ekstraksi teks yang digunakan berjumlah 12 fitur.

#### 1.5 Metodologi Penelitian

Pada penelitian ini, metodologi penelitian yang digunakan yaitu metodologi deskriptif [9]. Metodologi deskriptif dipilih karena penelitian dilakukan dengan mengumpulkan informasi status gejala atau keadaan menurut apa adanya kemudian dipaparkan dalam bentuk laporan. Tahapan yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 1.1 berikut.



**Gambar 1.1 Alur Penelitian**

### 1.5.1 Pengumpulan Data

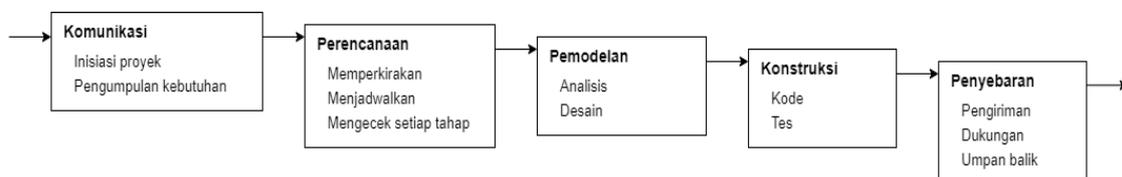
Metode pengumpulan data dilakukan dengan metode Studi Literatur yaitu dengan mengumpulkan dan mempelajari berbagai hasil penelitian, jurnal, paper, buku referensi, dan bacaan-bacaan yang ada kaitannya dengan peringkasan teks dokumen, pemrosesan teks, *Particle Swarm Optimization*, *Support Vector Machine* dan penggunaannya.

### 1.5.2 Analisis Metode

Dalam tahap analisis metode dilakukan proses analisa metode-metode apa saja yang akan digunakan pada penelitian ini seperti analisis data masukan, analisis tahap *preprocessing*, analisis ekstraksi fitur, analisis metode *Particel Swarm Optimization*, dan analisis *Support Vector Machine*.

### 1.5.3 Metode Pembangunan Perangkat Lunak

Dalam tahap ini metode yang digunakan oleh peneliti adalah Metode *waterfall* [10]. Metode *waterfall* dipilih karena setiap tahap dilakukan secara sistematis atau secara berurutan sehingga kebutuhan sistem bersifat stabil. Tahapan dalam model *waterfall* menurut Pressman dapat dilihat pada Gambar 1.2 berikut.



**Gambar 1.2 Metode Waterfall**

#### 1) Komunikasi

Tahapan komunikasi atau *communication* adalah tahap awal dalam membangun perangkat lunak yaitu dengan mengumpulkan kebutuhan perangkat lunak yang akan dibangun seperti data teks berita, mendefinisikan fungsi dan fitur, serta notasi-notasi yang digunakan dalam perhitungan PSO dan SVM.

#### 2) Perencanaan

Tahap perencanaan atau *planning*, dilakukan dengan menguraikan estimasi dari tugas-tugas teknis yang akan dikerjakan, kemudian dilakukan penjadwalan kerja

yang akan dilakukan, dan dilakukan proses *tracking* terhadap pengerjaan pembangunan perangkat lunak.

### 3) Pemodelan

Tahap pemodelan atau *modeling* adalah tahap dilakukannya penerjemahan kebutuhan perangkat lunak kedalam perancangan *software* dan dilakukan pemodelan arsitektur sistem. Tahap ini akan berfokus pada perencanaan seperti Diagram Konteks, DFD, dll, agar dapat dilihat gambaran besar dari yang akan dilakukan.

### 4) Konstruksi

Tahap Konstruksi atau *construction* merupakan tahap dilakukannya penerjemahan bentuk desain menjadi kode atau *coding*, dengan kata lain dilakukan implementasi hasil dari pemodelan *software* ke dalam bentuk atau bahasa yang dapat dibaca oleh komputer. Kemudian setelah selesai, dilakukan pengujian agar dapat dilihat kesalahan-kesalahan yang ada dan kemudian dapat diperbaiki.

### 5) Penyebaran

Tahap Penyebaran atau *deployment* adalah tahapan akhir dalam pengerjaan sebuah sistem. Tahap ini dilakukan perbaikan dan evaluasi agar sistem dapat sesuai dengan fungsinya yaitu melakukan pemilihan fitur kalimat untuk peringkasan teks otomatis menggunakan PSO-SVM.

## 1.5.4 Pengujian

Pada tahap ini dilakukan pengujian, agar diketahui fitur yang tepat untuk peringkasan teks otomatis menggunakan metode *Support Vector Machine*. Dalam pengujian akurasi digunakan metode *Confussion Matrix*. Sedangkan pengujian dari fungsionalitas menggunakan metode *black box*.

## 1.5.5 Pengambilan Keputusan

Pada tahap ini dilakukan kesimpulan dan saran mengenai penelitian ini, seberapa baik penggunaan metode pemilihan fitur *Particle Swarm Optimization* dan metode klasifikasi *Support Vector Machine* dalam melakukan peringkasan teks berita, dan juga saran untuk penelitian selanjutnya.

## **1.6 Sistematika Penulisan**

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

### **BAB 1 PENDAHULUAN**

Bab ini berisi uraian tentang latar belakang masalah yang mendasari pentingnya penelitian, perumusan masalah, batasan masalah, maksud dan tujuan penelitian dalam melakukan pemilihan fitur kalimat dalam peringkasan teks otomatis menggunakan metode PSO-SVM dan sistematika penulisan.

### **BAB 2 TINJAUAN PUSTAKA**

Pada bab ini membahas tentang berbagai tinjauan dan teori-teori dari literatur pendukung yang akan digunakan dalam penelitian tugas akhir. Adapun teori-teori yang akan ditulis pada bab ini adalah teori peringkasan teks, *preprocessing* (*case folding, split sentence, filtering, tokenization, stopword removal*), ekstraksi fitur (panjang kalimat, kalimat yang mengandung angka, posisi kalimat, kalimat yang menyerupai judul, kemiripan antar kalimat, koneksi antar kalimat, penjumlahan bobot koneksi antar kalimat, ikatan leksikal dengan kalimat sebelumnya, ikatan leksikal dengan kalimat setelahnya, *positive keyword, negative keyword, Term Frequency-Invers Document Frequency*), metode pemilihan fitur *Particle Swarm Optimization*, metode klasifikasi *Support Vector Machine*, bahasa pemrograman yang digunakan dan metode pengujian.

### **BAB 3 ANALISIS**

Bab ini menguraikan penjelasan analisis masalah, analisis data masukan, analisis *preprocessing*, analisis perhitungan ekstraksi fitur, analisis metode pemilihan fitur *Particle Swarm Optimization*, analisis metode klasifikasi *Support Vector Machine* dan diuraikan analisis perancangan aplikasi.

### **BAB 4 PENGUJIAN DAN PEMBAHASAN**

Bab ini berisi tentang hasil implementasi dari analisis metode dan perancangan yang dilakukan pada bab sebelumnya. Selain itu disertai dengan skenario dan hasil dari sistem pemilihan ekstraksi fitur.

## **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.

