

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Dewasa ini ketersediaan informasi yang semakin berkembang pesat menandakan bahwa kebutuhan untuk mendapatkan informasi semakin besar. Informasi yang dibutuhkan mengalami perkembangan mulai dari informasi yang bersifat umum hingga yang bersifat khusus. Banyaknya informasi dokumen yang tersedia mendorong manusia untuk mencari cara untuk mendapatkan informasi dan dokumen yang tepat dalam waktu yang singkat.

Permasalahan yang timbul apabila jumlah simpanan data dokumen menjadi sangat besar dan tidak terorganisir mengakibatkan proses pencarian dokumen tertentu tidak efektif [1]. Oleh karena itu, diperlukan suatu strategi pengelompokan otomatis dokumen-dokumen tersebut. Klasifikasi merupakan salah satu metode yang bertujuan untuk mendefinisikan kelas dari sebuah objek kelasnya ke dalam satu atau lebih kelompok yang telah dikenal sebelumnya secara otomatis berdasarkan isi dokumen.

*Extreme Learning Machines* (ELM) merupakan salah satu metode pembelajaran yang digunakan untuk klasifikasi otomatis. Alasan utama keberhasilan dari ELM adalah kemampuannya dalam memperoleh fungsi model yang memberikan tingkat akurasi dan kecepatan yang baik serta mampu menangani jumlah data yang besar [2].

Huang [3] telah mengusulkan algoritma pembelajaran baru *Single Hidden Layer Feedforward Neural Network* (SLFN) disebut *Extreme Learning Machine* (ELM). Di algoritma ini bobot input dan bias layer tersembunyi dipilih secara acak. Formulasi ELM mengarah ke pemecahan sistem persamaan linear dalam hal bobot yang tidak diketahui menghubungkan lapisan tersembunyi ke lapisan keluaran. Solusi dari sistem persamaan linear umum ini diperoleh menggunakan *Moore-Penrose pseudo inverse*.

ELM digunakan untuk klasifikasi data berupa text dengan tingkat akurasi dan kecepatan yang baik. Tetapi untuk proses klasifikasi dokumen seringkali ditemukan hasil yang kurang baik dikarenakan jumlah kata setiap dokumen yang

besar dan bervariasi sehingga harus dikelompokkan terlebih dahulu agar tingkat akurasi model yang diusulkan menjadi lebih baik [4]. Selain dilakukan pengelompokan kata juga dilakukan seleksi ciri berupa kumpulan perwakilan kata yang akan digunakan untuk data masukan.

Oleh karena itu, pada penelitian ini akan mengimplementasikan metode *Extreme Learning Machines* (ELM) untuk menyelesaikan permasalahan klasifikasi dokumen.

### **1.2 Identifikasi Masalah**

Berdasarkan uraian latar belakang di atas, maka masalah dari penelitian ini adalah bagaimana mengimplementasikan metode *Extreme Learning Machines* (ELM) dengan seleksi fitur pada klasifikasi dokumen untuk menghasilkan dokumen yang tepat.

### **1.3 Maksud dan Tujuan**

Berdasarkan permasalahan maka maksud dari penelitian ini adalah mengimplementasikan metode *Extreme Learning Machines* (ELM) pada klasifikasi dokumen dengan menggunakan seleksi fitur. Sedangkan tujuan yang diharapkan akan dicapai dalam penelitian ini adalah untuk mengukur akurasi dari hasil klasifikasi metode *Extreme Learning Machines* (ELM) dengan seleksi fitur pada klasifikasi dokumen.

### **1.4 Batasan Masalah**

Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Kategori yang ditentukan berdasarkan kelompok keilmuan Teknik Informatika, yaitu:
  - a. Kelompok A (Sistem Informasi)
  - b. Kelompok B (Rekayasa Pengetahuan Dan Perangkat Lunak)
  - c. Kelompok C (Multimedia Interaktif, Jaringan komputer dan internet, Keamanan Sistem, dan Teknologi IT)
  - d. Kelompok D (Teknologi Web dan Mobile)
  - e. Kelompok E (Ilmu Komputer)
2. Dokumen yang digunakan adalah dokumen Abstrak(judul dan isi) Skripsi Teknik Informatika UNIKOM.

3. Jumlah dokumen yang digunakan untuk klasifikasi adalah 150 buah. 100 untuk data latih dan 50 untuk data uji.
4. Tahapan *preprocessing* yang dilakukan adalah *filtering*, *case folding*, *tokenizing*, *stopword removal*, pembobotan *tf-idf* dan normalisasi.
5. *Stoplist* yang digunakan adalah Tala.
6. Metode yang digunakan adalah ELM dan K-Means.
7. Kamus sinonim bahasa Indonesia didapatkan dari alumni UNIKOM Teknik Informatika dengan penelitian “Pembangunan Sumber Daya Kamus Sinonim Bahasa Indonesia”.
8. Bahasa program yang akan digunakan adalah C#.
9. Perangkat lunak yang akan dikembangkan berbasis desktop.

### 1.5 Metodologi Penelitian

Metodologi penelitian adalah sekumpulan kegiatan dan prosedur yang digunakan dalam penelitian. Metodologi juga merupakan analisis teoritis mengenai suatu cara atau metode yang akan digunakan dalam penelitian. Dalam penelitian yang dilakukan metodologi yang akan dilakukan terdiri dari beberapa tahap seperti pada Gambar 1.1 berikut:



**Gambar 1. 1 Metode Penelitian**

#### 1) Studi Literatur

Studi literatur dilakukan dengan mempelajari teori- teori melalui buku, artikel, jurnal dan bahan lain yang berkaitan dengan metode tahap yang digunakan untuk mengklasifikasi dokumen dan algoritma yang digunakan.

#### 2) Analisis Kebutuhan Algoritma

Menganalisis kebutuhan algoritma mulai dari inputan (dokumen yang mengandung kata-kata yang dibutuhkan untuk penelitian), proses (tahapan praproses inputan yang diubah menjadi hasil sesuai format yang dipakai dalam

penelitian), sampai output yang dihasilkan untuk dapat diimplementasikan pada program.

### 3) Implementasi

Pada tahap ini dilakukan pembangunan dari klasifikasi dokumen untuk dokumen berbahasa Indonesia. Dimulai dari tahapan *preprocessing*, seleksi menggunakan algoritma K-Means dan implementasi ELM untuk menghasilkan klasifikasi, analisis kebutuhan non fungsional, analisis kebutuhan fungsional dan perancangan sistem (struktur tabel, perancangan antarmuka dan jaringan semantik).

### 4) Pengujian

Pengujian dilakukan dengan membandingkan hasil klasifikasi oleh sistem dengan klasifikasi awal (label) sehingga mendapatkan hasil perbandingan. Hasil perbandingan tersebut akan menjadi evaluasi apakah sistem berjalan cukup baik atau tidak.

### 5) Penarikan Kesimpulan

Pada tahap ini dilakukan penarikan kesimpulan yang didapat dari hasil klasifikasi yang dihasilkan oleh sistem untuk dokumen berbahasa Indonesia.

## **1.6 Sistematika Penulisan**

Sistematika penulisan ini disusun untuk memberikan gambaran umum tentang penelitian yang dilakukan. Sistematika penulisan secara umum adalah sebagai berikut:

### **BAB 1 PENDAHULUAN**

Bab ini terdiri dari latar belakang masalah, perumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, dan sistematika penulisan.

### **BAB 2 LANDASAN TEORI**

Bab ini Berisi pembahasan umum rangkaian penjelasan yang berhubungan dengan teori-teori yang akan digunakan dalam penelitian.

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

Bab ini menjelaskan analisis dan perancangan mulai dari analisis masalah, analisis sistem, analisis kebutuhan non-fungsional dan fungsional serta perancangan sistem.

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

Bab ini menjelaskan implementasi dari analisis dan perancangan sistem hasil implementasi selanjutnya akan dilakukan pengujian. Pengujian yang dilakukan adalah menguji akurasi hasil klasifikasi dokumen dengan hasil klasifikasi manual.

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

Bab ini berisi kesimpulan dari penelitian serta saran yang diuraikan dengan tujuan agar adanya pengembangan penelitian yang serupa dengan mengimplementasikan metode yang lebih baik lagi untuk hasil akurasinya.



