

# **BAB 1**

## **PENDAHULUAN**

### **1.1. Latar Belakang Masalah**

Ekstraksi Informasi adalah proses ekstraksi yang mengacu pada ekstraksi otomatis informasi terstruktur seperti entitas, hubungan antara entitas, dan atribut yang menggambarkan entitas dari sumber yang tidak terstruktur [1]. Karya tulis ilmiah menjadi salah satu contoh dokumen yang dapat dilakukannya ekstraksi dari informasi yang tersedia. Spesifikasi dasar dari tugas ekstraksi mencakup hanya jenis struktur yang akan diekstraksi [1]. Untuk mendeteksi jenis struktur yang akan diekstraksi pada dokumen karya tulis ilmiah sulit dilakukan untuk dokumen yang memiliki format berbeda, dikarenakan jenis struktur yang akan diekstraksi pada dokumen yang memiliki format berbeda dapat memberikan hasil yang berbeda. Oleh karena itu dibutuhkan metode untuk mengekstraksi informasi dari dokumen karya tulis ilmiah dengan kondisi format yang berbeda

Berdasarkan penelitian sebelumnya terhadap pengenalan dokumen karya tulis ilmiah untuk dokumen skripsi berbahasa Indonesia yang melakukan ekstraksi informasi menggunakan sistem berbasis aturan untuk mendapatkan identitas pada dokumen skripsi, diantaranya cover, dan abstrak, dan abstract [2]. Berdasarkan pengujian 3 dokumen yang memiliki format tahun 2017, diperoleh akurasi sebesar 100%. Sedangkan apabila pengujian dilakukan terhadap 50 dokumen yang disertakan dengan format tahun 2013 diperoleh rata-rata akurasi sebesar 57% [3]. Salah satu solusi untuk mengangani permasalahan tersebut adalah menggunakan machine learning.

Beberapa penelitian telah dilakukan terhadap penggunaan machine learning di bidang ekstraksi informasi, penelitian terhadap ekstraksi informasi pada karya tulis ilmiah menggunakan metode learning vector quantization (LVQ) dan memperoleh akurasi sebesar 78% [3]. Kemudian ekstraksi informasi pada makalah ilmiah dengan pendekatan supervised learning [4], dengan perolehan rata-rata metode SVM sebesar 98.81%. Begitupun pada bidang lainnya seperti penelitian

terhadap analisis sentiment review film [5], yang melakukan perbandingan terhadap metode ANN, SVM dan NB dengan hasil prolehan akurasi ANN sebesar 51,80%, SVM sebesar 81,10% dan NB sebesar 74%. Kemudian penelitian terhadap berita *Online Menggunakan Metode Support Vector Machine dan K- Nearest Neighbor* [6], menyatakan bahwa metode yang lebih unggul adalah metode *Support Vector Machine* dengan perolehan akurasi sebesar 93,2%. Oleh karena itu, algoritma *Support Vector Machine (SVM)* memiliki kemungkinan untuk memperoleh akurasi yang lebih baik apabila dilakukan pada bidang ekstraksi informasi

Dalam metode klasifikasi pembobotan fitur menjadi faktor yang penting untuk menentukan bobot nilai setiap kelas kategori. Pada penelitian yang dilakukan oleh Firdamdani Sasmita [3], hasil akurasi didapatkan dipengaruhi oleh 2 faktor, salah satunya adalah pembobotan fitur yang kurang signifikan. Penggunaan fitur yang tepat dapat meningkatkan akurasi [4], oleh karena itu perlu dilakukannya analisis terhadap pengaruh pembobotan fitur yang digunakan dan meminimalisir fitur yang kurang berpengaruh. Dalam mengetahui pengaruh setiap fitur yang digunakan *information gain* menjadi salah satu solusi dalam seleksi fitur yang kurang berpengaruh. Dengan *information gain* fitur-fitur yang digunakan dapat diketahui nilai pengaruh terhadap data training yang dilakukan. Berdasarkan penelitian yang dilakukan oleh Kiki Prima Wijaya dan Much Aziz Muslim [7], dalam mengoptimalkan fitur yang digunakan, dilakukannya seleksi fitur dengan pendekatan *information gain* berhasil menunjukkan peningkatan sebesar 0,75% pada akurasi dari 97,75% menjadi 98,5% dalam proses diagnose.

Berdasarkan uraian diatas, maka pada penelitian ini akan dibangunnya sistem ekstraksi informasi dengan metode *Support Vector Machine (SVM)* yang menggunakan pendekatan *information gain* dalam peningkatan akurasi pada dokumen karya tulis ilmiah, dengan batasan dokumen bahasa Indonesia dan dokumen merupakan skripsi dari Program Studi Teknik Informatika Universitas Komputer Indonesia (UNIKOM).

## 1.2. Identifikasi Masalah

Berdasarkan uraian pada latar belakang, identifikasi masalah yang didapat pada penelitian ini sebagai berikut.

1. Ekstraksi informasi menggunakan *Rule Based* mengalami penurunan akurasi terhadap dokumen karya tulis ilmiah yang memiliki format berbeda.
2. Terdapatnya fitur kurang signifikan dalam pembobotan yang mempengaruhi akurasi hasil klasifikasi.

## 1.3. Maksud Dan Tujuan

Maksud dari penelitian ini adalah untuk menganalisis dan mengimplementasikan metode Support Vector Machine (SVM) dengan pendekatan information gain untuk seleksi fitur pada kasus ekstraksi informasi karya tulis ilmiah. Adapun tujuan dari penelitian ini adalah untuk mengukur seberapa besar akurasi yang dihasilkan dari penerapan Support Vector Machine (SVM) dengan dan tanpa pendekatan information gain pada ekstraksi informasi karya tulis ilmiah.

## 1.4. Batasan Masalah

Adapun batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut.

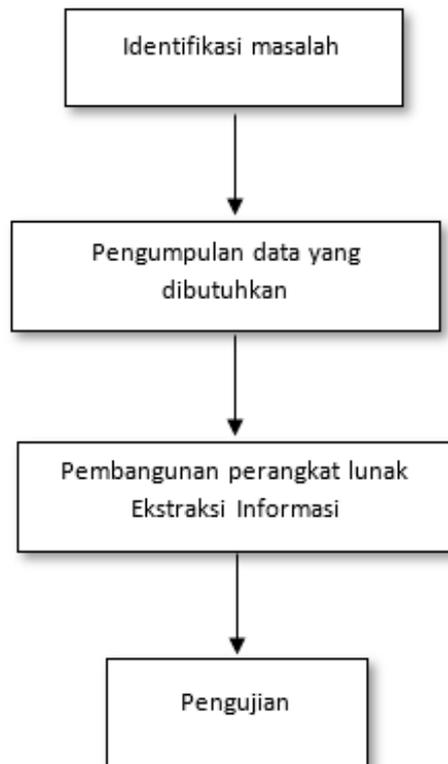
1. Data set yang digunakan untuk pelatihan dan pengujian merupakan dokumen karya tulis ilmiah skripsi Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Komputer Indonesia
2. Dokumen yang diuji adalah dokumen yang berbahasa Indonesia.

## 1.5. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah Deskriptif. Metode deskriptif merupakan metode yang menggambarkan fakta-fakta dan informasi dalam situasi atau kejadian dimana sekarang secara sistematis, faktual

dan akurat. Metode penelitian ini memiliki empat tahapan, yaitu tahap identifikasi masalah, pengumpulan data, pembangunan perangkat lunak dan pengujian.

Langkah-langkah yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 1.1



**Gambar 1.1 Alur Pengerjaan**

Adapun penjelasan langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut.

#### **1.5.1. Identifikasi Masalah**

Pada tahap ini dilakukan identifikasi masalah terhadap kebutuhan pada penelitian ekstraksi informasi karya tulis ilmiah dan analisis kebutuhan algoritma yang digunakan.

### **1.5.2. Pengumpulan Data**

Metode pengumpulan data yang digunakan dalam membantu penelitian ini menggunakan dua cara, yaitu dengan cara studi literatur dan dokumen.

#### **1. Studi literatur**

Melakukan pencarian terhadap berbagai sumber tertulis, baik berupa buku-buku, arsip, majalah, artikel, dan jurnal, atau dokumen-dokumen yang relevan dengan permasalahan yang dikaji. Sehingga informasi yang didapat dari studi kepustakaan ini dijadikan rujukan untuk memperkuat argumentasi-argumentasi yang ada.

#### **2. Pengumpulan Dokumen**

Sumber data diambil dari skripsi Program Studi Teknik Informatika Universitas Komputer Indonesia.

### **1.5.3. Pembangunan Perangkat Lunak**

Metode pembangunan perangkat lunak yang digunakan dalam penelitian ini adalah model waterfall. Adapun tahapan dari model waterfall adalah Requirements analysis and definition, System and software design, Implementation and unit testing, Integration and system testing dan Operation and maintenance.

### **1.5.4. Pengujian**

Menjelaskan tentang pengujian untuk menghitung keakuratan dari metode SVM yang diimplementasikan untuk ekstraksi informasi dokumen katya tulis ilmiah skripsi Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Komputer Indonesia dengan metode *Confusion Matrix*.

## **1.6. Sistematika Penulisan**

Laporan tugas akhir ini disusun dengan sistematika dengan tujuan agar hasil dari laporan lebih terarah dan tidak menyimpang dari permasalahan yang telah dibuat. Dengan pembahasan sebagai berikut.

## **BAB 1 PENDAHULUAN**

Dalam bab ini membahas tentang latar belakang, rumusan masalah, maksud dan tujuan, manfaat, batasan masalah serta metodologi penelitian yang digunakan dalam implementasi klasifikasi program acara televisi.

## **BAB 2 TINJAUAN PUSTAKA**

Dalam bab ini berisi tentang landasan teori penunjang yang dapat mendukung pokok pembahasan dari ekstraksi informasi menggunakan Support Vector Machine (SVM) dengan pendekatan information gain.

## **BAB 3 ANALISIS DAN PERANCANGAN**

Dalam bab ini akan memaparkan analisis serta perancangan. Analisis yang akan dibahas berupa analisis sistem, analisis algoritma, analisis kebutuhan fungsional, analisis kebutuhan non-fungsional. Pada perancangan yang akan dibahas adalah perancangan antarmuka.

## **BAB 4 IMPLEMENTASI DAN PENGUJIAN**

Dalam bab ini menjelaskan tentang implementasi dan pengujian dari perangkat lunak yang dibangun sesuai dengan analisis dan perancangan yang telah dibuat. Implementasi yang akan dibahas adalah implementasi antarmuka. Pada bagian pengujian yang akan dibahas adalah pengujian sistem yang telah dibuat.

## **BAB 5 KESIMPULAN DAN SARAN**

Dalam bab ini akan menyimpulkan hasil dan nilai akurasi yang didapat serta memberikan saran untuk penelitian dalam pengembangan selanjutnya.