

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Semakin dikembangkannya teknologi kecerdasan buatan (*artificial intelligence*) dengan bermunculannya implementasi seperti sistem pakar, pengenalan suara, pengenalan citra dan lain sebagainya memberi manfaat yang besar bagi kehidupan manusia. Teknologi tersebut saat ini telah disematkan untuk berbagai keperluan seperti sistem tes psikologi, pengenalan suara untuk melakukan pencarian pada *Google*, ataupun efek *beauty* pada kamera ponsel.

Implementasi lainnya mampu digunakan untuk mengolah teks seperti halnya dalam kasus klasifikasi data teks otomatis. Salah satu kegiatan mengklasifikasikan data teks biasa dilakukan di perpustakaan untuk menentukan kelas/kategori suatu dokumen. Suatu dokumen pada perpustakaan biasanya diklasifikasikan dalam standar tertentu seperti *DDC (Dewey Decimal Classification)*. Pada perpustakaan di dalam perguruan tinggi, dokumen skripsi pun juga diklasifikasikan.

Namun apabila proses klasifikasi tersebut dilakukan secara manual akan berdampak pada waktu dan banyaknya sumber daya manusia yang terlibat untuk menangani hal tersebut. Selain itu sangat mungkin jika terdapat kekeliruan karena sifat manusia, seperti lelah ataupun lupa. Kekeliruan yang timbul adalah penentuan kelas dokumen skripsi yang tidak sesuai. Sehingga dari permasalahan tersebut perlu adanya solusi permasalahan klasifikasi pada dokumen skripsi dari segi waktu,

tenaga, serta pikiran.

Dokumen skripsi memungkinkan untuk dilakukan klasifikasi hanya dengan menggunakan judul dari setiap dokumen tersebut. Karena, judul skripsi berisikan kata-kata yang jelas menunjukkan topik dari isi dokumen tersebut. Dengan kata lain, judul skripsi menunjukkan kata kunci terhadap isi dokumen.

Untuk melakukan dokumen tersebut secara terkomputerisasi, tidak serta merta langsung dapat dilakukan dengan perhitungan menggunakan metode klasifikasinya. Hal itu disebabkan karena tipe data dari judul skripsi adalah teks atau tidak dapat diolah langsung pada perhitungan angka dalam metode klasifikasi. Sehingga diperlukan adanya beberapa tahapan. Adapun yang dikenal dengan istilah *text mining*. *Text mining* adalah satu langkah dari analisis teks yang dilakukan secara otomatis oleh komputer untuk menggali informasi yang berkualitas dari suatu rangkaian teks yang terangkum dalam sebuah dokumen [1].

Di dalam teks mining terdapat beberapa metode klasifikasi yang dapat digunakan seperti *Support Vector Machine* (SVM), *Naïve Bayes Classifier* (NBC), *K-Means Classifier*, dan *K-Nearest Neighbor* (KNN) [2]. Dari 4 metode tersebut *Naïve Bayes Classifier* merupakan salah satu pilihan yang menarik untuk menangani permasalahan klasifikasi judul skripsi. Hal tersebut didasarkan dengan kelebihan yang terdapat pada metode *Naïve Bayes*. Kelebihan dari metode ini adalah memiliki akurasi yang tinggi, mudah diimplementasikan, dan dapat menangani permasalahan banyak kelas [3] [4]. Dengan berdasar pertimbangan, metode ini cocok untuk menangani kasus klasifikasi teks dengan proses yang mudah dan memiliki akurasi yang tinggi. Serta pada penelitian ini juga akan

menguji akurasi *Naïve Bayes Classifier* dengan berbagai percobaan pada bagian *preprocessing* untuk memberikan hasil pembobotan yang baik.

## 1.2 Rumusan Masalah

Sebagaimana latar belakang masalah yang telah dijabarkan di atas, maka masalah dapat dirumuskan sebagai berikut:

1. Bagaimana mengimplemenasikan *Naïve Bayes Classifier* untuk mengklasifikasikan skripsi berdasarkan judul?
2. Bagaimana kinerja *Naïve Bayes Classifier* pada penelitian ini?

## 1.3 Batasan Masalah

Agar penelitian berjalan dengan efektif, maka dibuat batasan masalah sebagai berikut:

1. Data yang digunakan berasal dari portal Sistem Informasi Perpustakaan Universitas Islam Indonesia.
2. Algoritma yang digunakan pada pengklasifikasian ini adalah *Naïve Bayes Classifier*.
3. Data dibagi menjadi 60% data *training* dan 40% data *testing*.
4. Pengujian model dengan metode pemecahan kata *unigram*, *bigram*, dan *trigram*, beserta kombinasinya.
5. Digunakan *stopword list* Tala untuk proses *filtering* pada tahap *preprocessing*.
6. Digunakan *Sastrawi Streammer* sebagai *stemming* kata berbahasa Indonesia.

7. Digunakan metode pembobotan term/kata *Tf-Idf (Term Frequency Inverse Document Frequency)*.
8. Menggunakan server *database MySQL*.
9. Menggunakan bahasa pemrograman *Python*.

#### **1.4 Maksud dan Tujuan Penelitian**

Adapun maksud dan tujuan penelitian ini adalah sebagai berikut:

1. Mengimplementasikan *Naïve Bayes Classifier* pada klasifikasi skripsi berdasarkan judul.
2. Mengukur kinerja yang diperoleh dari sistem *Naïve Bayes Classifier* yang dibuat pada penelitian ini.

#### **1.5 Manfaat Penelitian**

Manfaat penelitian yang diharapkan pada penelitian ini adalah:

1. Memberi wawasan bagaimana cara mengimplementasikan *Naïve Bayes Classifier* pada permasalahan klasifikasi data teks berbahasa Indonesia.
2. Mengetahui kinerja sistem *Naïve Bayes Classifier* yang dibuat pada penelitian ini.
3. Sistem yang dibuat dapat digunakan oleh pengelola perpustakaan dalam mengklasifikasikan dokumen skripsi dari data judul skripsi yang telah diklasifikasikan sebelumnya.
4. Sebagai rujukan pada penelitian lainnya dalam kasus klasifikasi pada data teks.

## 1.6 Metodologi Penelitian

Metode penelitian merupakan cara utama yang digunakan peneliti untuk mencapai tujuan dan menentukan jawaban atas masalah yang diajukan. Adapun metode penelitian yang akan digunakan adalah sebagai berikut:

### 1.6.1 Metode Pengumpulan Data

Metode pengumpulan data diperlukan untuk mendapatkan berbagai informasi yang dibutuhkan dalam mencapai tujuan penelitian. Adapun metode yang digunakan adalah studi pustaka dan studi dokumen. Studi pustaka merupakan teknik pengumpulan data dengan berbagai buku, literatur, catatan, serta berbagai laporan yang berkaitan dengan masalah yang ingin dipecahkan. Studi dokumen merupakan teknik pengumpulan dokumen tertulis maupun elektronik dari instansi/ lembaga yang diperlukan untuk mendukung penelitian.

### 1.6.2 Metode Analisis Data

Analisis data merupakan tahap lanjut dari pada pengumpulan data. Pada tahap ini data yang terkumpul akan diidentifikasi, sehingga didapat data-data yang diperlukan dalam tahap perancangan sistem. Metode analisis data yang digunakan adalah dengan analisis kebutuhan data dan analisis kebutuhan fungsional serta non-fungsional. Adapun data *CSV* yang sudah dipersiapkan pada tahap sebelumnya, data tersebut akan diolah pada tahap *text mining*. *Text mining* penelitian ini terdiri dari *preprocessing* data teks, tahap pembobotan term dengan metode *tf-idf*, tahap *Naïve Bayes Classifier* dan evaluasi menggunakan *Confusion Matrix*.

### 1.6.3 Metode Perancangan Sistem

Untuk merancang sistem dibutuhkan model dari perancangan sistem, sehingga gambaran dari sistem yang akan dibangun menjadi jelas. Pada penelitian ini digunakan model perancangan *ERD (Entity Relationship Diagram)*, rancangan struktur tabel, *DFD (Data Flow Diagram)*, dan *flowchart*.

### 1.6.4 Metode Pengembangan Sistem

Pada penelitian ini digunakan metode pengembangan sistem yaitu *waterfall*. Metode pengembangan *waterfall* adalah model pengembangan sistem yang bersifat sistematis, serta berurutan dalam mengembangkan *software*. Adapun tahapan dalam model ini yaitu, analisis kebutuhan sistem, perancangan sistem, implementasi sistem, pengujian sistem, serta pemeliharaan.

### 1.6.5 Metode Pengujian

Metode pengujian merupakan tahap uji coba untuk menentukan kelayakan aplikasi yang telah dibangun. Dengan adanya tahap ini sistem dapat dipastikan berjalan sesuai dengan harapan. Adapun *testing* yang akan dilakukan pada sistem ini yaitu dengan metode *black box testing*. *Black box testing* merupakan pengecekan yang didasarkan pada kebutuhan fungsionalitas sistem yang dibangun.

## 1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun ke dalam beberapa bab sebagai berikut:

### 1. BAB I : PENDAHULUAN

Bab ini dibahas mengenai latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan

penelitian, manfaat penelitian, metode penelitian, serta sistematika penulisan.

## 2. BAB II : LANDASAN TEORI

Pada bab landasan teori berisi tinjauan pustaka yang merupakan suatu penelitian terdahulu yang dijadikan rujukan dalam penelitian ini serta terdapat teori-teori yang dijadikan sebagai landasan atau dasar dari penelitian ini.

## 3. BAB III : ANALISIS DAN PERANCANGAN

Bab analisis dan perancangan membahas mengenai analisis penelitian beserta tahapan perancangan dari sistem yang digunakan dalam pengujian kinerja model.

## 4. BAB IV : IMPLEMENTASI DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai implementasi dari hasil analisis yang terdapat pada bab sebelumnya beserta hasil dari pengujian kinerja model *Naïve Bayes Classifier* dalam menangani permasalahan klasifikasi judul skripsi.

## 5. BAB V : PENUTUP

Merupakan bab terakhir yang berisikan tentang kesimpulan dari hasil penelitian dan saran yang diharapkan berguna sebagai rujukan pada penelitian selanjutnya.