

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Dunia telah berubah menjadi era digital. Perubahan yang besar dapat dilihat pada segala bidang, terutama bidang teknologi informasi. Informasi didapatkan dalam berbagai bentuk seperti dokumen, gambar, video, suara, dll. Dokumen-dokumen buku yang menjadi sumber informasi memiliki identitas yang berbeda. Agar mempermudah pengguna untuk men-sitasi suatu dokumen buku, perlu dilakukan ekstraksi dokumen buku.

Penelitian mengenai *Information Extraction* (IR) pada dokumen sudah banyak dilakukan oleh para peneliti dengan berbagai metode dan memperlihatkan hasil yang bervariasi.

Penelitian yang dilakukan oleh Romario Yudo Herlambang, Rekyan Regasari Mardi Putri dan Randy Cahya Wihandika dalam “IMPLEMENTASI METODE K-NEAREST NEIGHBOUR DENGAN PEMBOBOTAN TF.IDF.ICF UNTUK KATEGORISASI IDE KREATIF PADA PERUSAHAAN”, menunjukkan bahwa hasil penelitian dengan jumlah  $k=1$  memiliki hasil akurasi yang cukup baik yaitu 90%. Penelitian yang dilakukan [ [HYPERLINK \l "Rom" 1](#) ]. Adam Sulaiman mengenai “EKSTRAKSI INFORMASI PADA DOKUMEN SURAT MASUK DENGAN ALGORITMA FUZZY K-NEAREST NEIGHBOUR (FUZZY K-NN)” menunjukkan bahwa dari hasil pengujian, metode *K-Nearest Neighbours* memiliki akurasi yang sangat baik yaitu rata-rata diatas 85%2}}.

Berdasarkan penjelasan diatas, dalam penelitian ini akan dilakukan implementasi metode *K-Nearest Neighbour* mengekstraksi identitas dari suatu dokumen buku.

## 1.2 Rumusan Masalah

Rumusan masalah menurut latar belakang diatas adalah :

1. Apakah ekstraksi dokumen buku menggunakan *K-Nearest Neighbourss* perlu dilakukan
2. Berapa besarnya akurasi yang dihasilkan dari penggunaan metode *K-Nearest Neighbours*

## 1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah mengimplementasi metode *K-Nearest Neighbour* untuk ekstraksi dokumen buku. Adapun, tujuan dari penelitian ini adalah untuk mengevaluasi keefektifan *K-Nearest Neighbour* sebagai metode Ekstraksi Dokumen buku.

## 1.4 Batasan Masalah

Di dalam penelitian ini terdapat batasan-batasan yang menjadi ruang lingkup penelitian, batasan-batasan tersebut adalah :

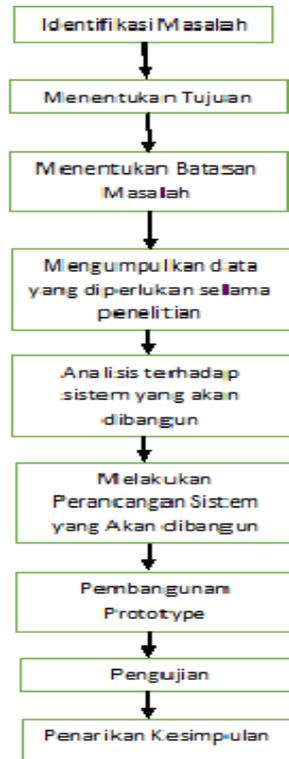
1. Data *latih* dan data *uji* diperoleh dari sampul buku.
2. Jenis dokumen teks adalah dokumen yang berformat .jpg, .png, .jpeg
3. Pengelompokan dokumen buku akan disesuaikan dengan kategori-kategori yang telah ditentukan.

## 1.5 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini yaitu metode penelitian *deskriptif* yang merupakan suatu metode dalam meneliti status sekelompok manusia, suatu objek, suatu set kondisi, suatu sistem pemikiran, ataupun suatu kelas peristiwa pada masa sekarang. Tujuan dari penelitian deskriptif ini adalah untuk membuat deskripsi, gambaran atau lukisan secara sistematis, faktual dan akurat mengenai fakta-fakta, sifat-sifat serta hubungan antar fenomena yang diselidiki. Serta memiliki ciri-ciri menerangkan hubungan, menguji hipotesis-hipotesis, membuat prediksi serta mendapatkan makna(kesimpulan) dan implikasi dari suatu masalah yang ingin dipecahkan. alur

penelitian yang akan dilakukan pada penelitian ini dapat dilihat pada gambar 1.0 sebagai berikut :

:



**Gambar 1.1 Alur Penelitian**

### **1.5.1 Identifikasi Masalah**

Melakukan identifikasi terhadap masalah yang ada dari penelitian yang akan dilakukan. Dengan membaca jurnal untuk melihat kekurangan dari penelitian sebelumnya.

### **1.5.2 Penentuan Tujuan Penelitian**

Setelah melakukan identifikasi, kemudian akan dilakukan penentuan tujuan yang akan dicapai.

### **1.5.3 Menentukan Batasan**

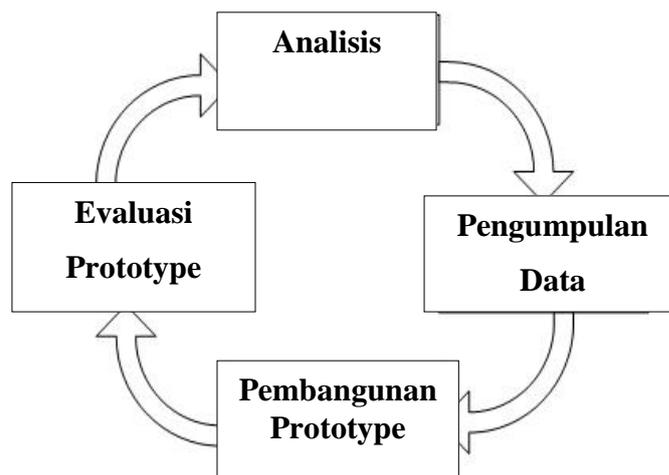
Dalam proses ini, melakukan penentuan batasan-batasan yang ada dalam pembuatan laporan.

#### 1.5.4 Metode Pengumpulan Data

Pada penelitian ini metode yang digunakan adalah studi. Studi literatur yang diperoleh dari sumber-sumber tertulis, baik yang tercetak maupun elektronik, seperti buku, *e-book*, jurnal, *paper*, dan sumber-sumber yang berhubungan dengan data yang diperlukan.

#### 1.5.5 Metode Pembangunan Prototype

Metode pembangunan sistem yang digunakan adalah *prototype* karena pada pembangunan sistem pengklasifikasian didasarkan pada ide pengembangan awal, diuji keberhasilannya, dan menyempurnakannya dengan kembali ke pembangunan. Tahapan dari metode *prototype* dapat dilihat pada Gambar 1.1.



**Gambar 1.2 Model Prototype** [ [HYPERLINK \|"Roz17" 3](#) ]

1. Analisis

Analisis yang dilakukan yaitu mulai penerimaan input teks dokumen hingga melakukan proses pengklasifikasian, analisis penentuan keefektifan metode *K-Nearest Neighbours* untuk Ekstraksi Informasi.

2. Pengumpulan data

Pada tahap ini peneliti akan mengumpulkan berupa file data berita untuk data masukan sistem.

3. Pembangunan *prototype*

Merupakan tahap pengimplementasian dari proses analisis dan kebutuhan sistem yang sudah didapatkan. Pada tahap ini pengembang mencoba mengimplementasikan *K-Nearest Neighbours* pengklasifikasian dokumen.

#### 4. Evaluasi *prototype*

Program akan diuji, dimana uji coba dilakukan untuk mengetahui kekurangan pada program. Jika masih ada kekurangan, maka *prototype* direvisi dengan tahapan-tahapan yang sebelumnya telah dilakukan. Proses ini akan terus berulang sampai *prototype* yang dihasilkan mendekati harapan yang diinginkan.

#### 5. Pengujian

fase ini adalah tahap pengujian untuk menghitung keakuratan dari metode *K-Nearest Neighbours* pada Klasifikasi Dokumen.

### **1.6 Sistematika Penulisan**

Sistematika penulisan menjelaskan tentang gambaran umum penelitian yang dijalankan. Adapun 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, dan sistematika penulisan yang dimaksudkan agar dapat memberikan gambaran tentang urutan pemahaman dalam penelitian pengklasifikasian informasi dokumen menggunakan metode *K-Nearest Neighbours*.

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

Bab ini membahas mengenai landasan teori yang digunakan untuk menganalisis masalah yaitu, teori mengenai penjelasan klasifikasi dokumen, *text preprocessing*, pembobotan text, metode *K-Nearest Neighbours*.

#### **BAB 3 ANALISIS KEBUTUHAN IMPLEMENTASI ALGORITMA**

Bab ini menguraikan penjelasan analisis perhitungan pada metode *K-Nearest Neighbours* untuk diaplikasikan pada pengklasifikasian informasi dokumen.

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

Pada bab ini membahas mengenai pengujian akurasi pengklasifikasian dokumen buku menggunakan metode *K-Nearest Neighbours*. Implementasi yang dilakukan terdiri dari implementasi perangkat keras, implementasi perangkat lunak dan implementasi antarmuka.

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

Pada bab ini menjelaskan tentang kesimpulan yang diperoleh dari hasil penelitian serta saran untuk pengembangan pengklasifikasian dokumen.