

# BAB I

## LATAR BELAKANG

### 1.1 Latar Belakang Masalah

Tugas Akhir merupakan salah satu syarat wajib untuk mendapat gelar sarjana. Tugas Akhir di beberapa program studi memiliki beberapa kategori, salah satunya pada prodi Teknik Informatika S1 di Universitas Muhammadiyah Jember. Semakin lama dokumen Tugas Akhir semakin bertambah banyak. Banyaknya dokumen yang tersedia mendorong manusia untuk mencari cara untuk mendapatkan informasi dan dokumen yang tepat dalam waktu singkat. Apabila jumlah dokumen yang tersedia sangat besar, proses pencarian secara manual akan menghabiskan waktu dan tenaga. Sehingga akan lebih mudah jika dokumen tersebut sudah diketahui sesuai dengan kategorinya masing-masing. Penyusunan dokumen sesuai dengan kategori yang ada sangat diperlukan untuk menyesuaikan dengan kategori yang terkandung pada dokumen tersebut secara otomatis.

Salah satu proses untuk menyelesaikan masalah tersebut adalah dengan proses *Text Mining*. *Text Mining* merupakan proses mengelompokkan suatu teks ke dalam suatu kategori tertentu. Algoritma kategorisasi teks saat ini telah banyak berkembang, antara lain: *Support Vector Machines (SVM)*, *Naive Bayessian*, *Decision Tree*, dan *K-Nearest Neighbor (K-NN)* (G. Toker and Ö. Kirmemi ). Algoritma K-NN termasuk suatu algoritma yang sederhana, namun cukup efektif dalam melakukan kategorisasi teks. Bukan hanya mudah dan efisien, sifat dari algoritma *K-Nearest Neighbor* sendiri bersifat *self-learning*, dimana algoritma ini dapat mempelajari struktur data yang ada dan mengkategorikan dirinya sendiri (Zhang , 2009).

Penelitian ini akan difokuskan pada kategorisasi Tugas Akhir melalui informasi dalam dokumen abstraksi tugas akhir. Penelitian dilakukan dengan mengklasifikasi data dokumen teks abstraksi ke kategori bidang minat yang ada. Sehingga diharapkan mampu membuat sistem kategorisasi dokumen tugas akhir menggunakan metode *k-nearest neighbor*.

## 1.2 Rumusan Masalah

Seberapa tinggi kinerja sistem kategorisasi yang didapat dari algoritma *K-Nearest Neighbor* pada kasus pengkategorisasian dokumen tugas akhir yang diwakili oleh abstraksi tugas akhir.

## 1.3 Batasan Masalah

Batasan masalah penelitian ini meliputi :

1. Kategorisasi dilakukan pada dokumen Abstraksi Tugas Akhir Teknik Informatika Universitas Muhammadiyah Jember.
2. Dokumen berasal dari penelitian tugas akhir yang dilakukan oleh Muhammad Isa Hidayatullah yang berjudul Klasifikasi Dokumen Tugas Akhir Menggunakan Metode Support Vector Machine dengan Pemilihan Fitur Chi-Square.
3. File yang akan di kategorisasi berisi file teks bahasa Indonesia
4. Penelitian difokuskan untuk mengetahui kategori dokumen menggunakan metode klasifikasi *k-nearest neighbor*.
5. Studi kasus yang diambil pada Bidang Minat Fakultas Teknik Prodi Informatika Universitas Muhammadiyah Jember.
6. Kategori dokumen ditentukan dari Bidang Minat Teknik Informatika yaitu Jaringan, Rekayasa Perangkat Lunak(RPL) dan Bisnis Cerdas.
7. Dokumen yang digunakan berjumlah 100 dokumen yang akan dibagi menjadi 85% (85 dokumen) data latih dan 15% (15 dokumen) data uji.

### **1.3 Tujuan Penelitian**

Mengaplikasikan algoritma *K-Nearest Neighbor* untuk mengukur akurasi dalam mengkategorisasikan dokumen tugas akhir kedalam Bidang Minat dengan mengambil studi kasus untuk program studi Teknik Informatika.

### **1.4 Manfaat Penelitian**

Penelitian ini diharapkan dapat menambah metode klasifikasi dokumen dan membantu dalam mengorganisasikan dokumen secara cepat dan memiliki kinerja yang baik.