

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Tugas akhir merupakan salah satu syarat kelulusan yang harus dipenuhi oleh mahasiswa untuk menyelesaikan pendidikan di perguruan tinggi. Ilmu yang didapat selama perkuliahan dituangkan ke dalam suatu penelitian yang akan menghasilkan luaran berupa dokumen tugas akhir (Yusra dkk, 2016).

Semakin bertambahnya mahasiswa tiap tahunnya, maka semakin banyak pula koleksi dokumen tugas akhir. Semakin banyaknya dokumen tugas akhir menyebabkan sulitnya mengkategorikan dokumen tugas akhir jika harus dilakukan secara manual. Mengelompokan tugas akhir berdasarkan judul saja kurang efektif karena ada judul tugas akhir yang dapat di kategorikan ke dalam lebih dari satu program studi khususnya di lingkungan fakultas teknik. Salah satu cara yang berhasil dalam mengkategorikan dokumen tugas akhir dalam jumlah banyak yaitu dengan klasifikasi dokumen tugas akhir.

Penelitian klasifikasi tugas akhir sendiri sudah pernah dilakukan antara lain yaitu Yusra dkk (2016) melakukan penelitian berjudul Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor*, menghasilkan kesimpulan bahwa metode *Bayes Classifier* dan *K-Nearest Neighbor* mampu mengklasifikasikan tugas akhir mahasiswa jurusan Teknik Informatika. Lestari dkk (2019) melakukan penelitian berjudul Klasifikasi Teks Berbasis Ontologi Untuk Dokumen Tugas Akhir Berbahasa Indonesia, menghasilkan kesimpulan bahwa pengujian yang dilakukan kepada sistem yang telah dibuat menghasilkan nilai akurasi 87%.

Universitas Muhammadiyah Jember (UM JEMBER) merupakan salah satu perguruan tinggi swasta yang berada di Kota Jember. salah satu fakultas yang terdapat di UM JEMBER adalah Fakultas Teknik yang di dalamnya terdapat lima jurusan yaitu Teknik Informatika, Teknik Sipil, Teknik Mesin, Teknik Elektro dan Manajemen Informatika. Sebagai syarat kelulusan, mahasiswa diwajibkan untuk

menyusun tugas akhir. Namun tugas akhir yang ada di Fakultas Teknik Universitas Muhammadiyah Jember belum ada yang mengklasifikasikan berdasarkan program studi.

Dalam *data mining* terdapat beberapa metode pengklasifikasian salah satu diantaranya yaitu KNN (*K-Nearest Neighbor*) & *Rocchio*. Penelitian – penelitian menggunakan metode KNN & *Rocchio* telah banyak diterapkan dan menyebutkan hasil dari dua algoritma tersebut memiliki akurasi yang cukup bagus dalam metode klasifikasi, berdasarkan penelitian yang dilakukan oleh Moldagulova dan Sulaiman (2017), yang membahas tentang penggunaan algoritma KNN untuk klasifikasi dokumen tekstual. Dimana dalam penelitiannya algoritma KNN menghasilkan akurasi sebesar 97,14 % untuk $K = 5$. Begitu juga dengan metode *Rocchio*, berdasarkan penelitian yang dilakukan oleh Afriza dan Adisantoso (2018), yang membahas tentang metode klasifikasi *Rocchio* untuk analisis *hoax* dengan algoritma *multinomial naive bayes* sebagai pembanding dimana algoritma *Rocchio* menghasilkan akurasi yang lebih besar yaitu 83,501 %.

Berdasarkan uraian di atas, maka dalam penelitian ini akan dilakukan perbandingan antara metode KNN dan *Rocchio* untuk mengklasifikasikan tugas akhir berdasarkan abstrak dan bab 1 di Fakultas Teknik Universitas Muhammadiyah Jember kedalam 5 program studi yaitu Teknik Sipil, Teknik Informatika, Teknik Mesin, Teknik Elektro, dan Manajemen Informatika dengan judul “PERBANDINGAN ALGORITMA *K-NEAREST NEIGHBOR* (KNN) DAN *ROCCHIO* DALAM KLASIFIKASI TUGAS AKHIR UNIVERSITAS MUHAMMADIYAH JEMBER STUDY KASUS FAKULTAS TEKNIK”

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan di atas, maka permasalahan yang akan dibahas dalam penelitian ini adalah sebagai berikut: Berapakah tingkat akurasi, presisi, dan *recall* perbandingan algoritma *K-Nearest Neighbor* (KNN) dan *Rocchio* dalam klasifikasi tugas akhir Fakultas Teknik Universitas Muhammadiyah Jember berdasarkan abstrak dan bab 1?

1.3 Batasan Masalah

Agar permasalahan tidak menyimpang pada tujuan penelitian, maka berikut beberapa batasan yang perlu dibuat, yaitu:

1. Jumlah dataset yang digunakan dalam penelitian ini sebesar 150 *record* dengan masing – masing 30 *record* untuk tiap kelas.
2. Dokumen yang digunakan berupa abstrak dan bab 1 pada tugas akhir Fakultas Teknik yang ada di *Repository* Universitas Muhammadiyah Jember yang diambil secara acak .
3. Bahasa pemrograman yang digunakan dalam penelitian ini adalah *Python*.
4. *Tools* yang digunakan dalam penelitian ini adalah *Jupyter Notebook*.
5. Apabila terdapat *record* data yang memiliki lebih dari satu kelas pada KNN maka diasumsikan benar jika salah satu kelas prediksi sama dengan kelas sebenarnya.
6. Objek penelitian ini berupa kelas SIPIL, MI, TI, MESIN, & ELEKTRO.
7. Merupakan klasifikasi *single class*.
8. Menghitung nilai akurasi, presisi, dan *recall*.
9. Menggunakan modul *stemming* Sastrawi.
10. Menggunakan modul *NLTK* untuk *Stopword*.
11. Menggunakan modul *sklearn* untuk algoritma *KNN* dan *Rocchio*

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

Membandingkan tingkat akurasi, presisi, dan *recall* dari algoritma *K-Nearest Neighbor* (KNN) dan *Rocchio* dalam klasifikasi tugas akhir Fakultas Teknik Universitas Muhammadiyah Jember.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini yang dilakukan sebagai berikut:

1. Bagi penulis:

Hasil dari penelitian ini diharapkan dapat menambah wawasan dalam bidang *text mining*, serta sebagai syarat kelulusan penulis.

2. Bagi instansi:

Diharapkan dapat memberikan kontribusi secara keilmuan berupa hasil perbandingan algoritma *K-Nearest Neighbor* dan *Rocchio* dalam klasifikasi tugas akhir mahasiswa Fakultas Teknik Universitas Muhammadiyah Jember.

3. Bagi peneliti lain:

Hasil penelitian ini dapat digunakan sebagai referensi untuk penelitian-penelitian berikutnya di bidang klasifikasi dokumen dengan metode *text mining*.

