BAB 1 PENDAHULUAN

1.1 Latar Belakang

Tugas akhir merupakan suatu karya akademik yang ditulis sebagai syarat untuk menyelesaikan suatu program studi di suatu universitas. Tugas akhir disusun untuk memenuhi syarat memperoleh gelar berdasarkan jenjangnya. Tugas akhir ini sering juga disebut dengan skripsi, thesis, atau disertasi, tergantung pada jenjang pendidikan dan program studi yang diikuti. Dengan demikian, tugas akhir merupakan suatu proyek ilmiah yang disiapkan oleh mahasiswa untuk menyelesaikan studinya melalui proses berpikir ilmiah, kreativitas, integrasi, dan sesuai dengan keahlian keilmuannya, untuk memenuhi persyaratan penyelesaian gelar dan program (Septiana, dkk, 2016).

Universitas Muhammadiyah Jember merupakan salah satu perguruan tinggi swasta di wilayah Jawa Timur. Dengan total 9 Fakultas, salah satunya yaitu Fakultas Teknik, yang terdiri dari 7 program studi yaitu Teknik Informatika, Teknik Sipil, Teknik Mesin, Teknik Elektro, Teknik Lingkungan, Sistem Informasi dan Menejemen Informatika. Dengan 7 program study yang dimiliki, Fakultas Teknik menghasilkan keluaran berupa dokumen Tugas Akhir mahasiswa yang telah menyelesaikan program pendidikannya. Banyaknya lulusan dari fakultas teknik tentunya berbanding lurus dengan banyaknya tugas akhir yang bertambah setiap tahunnya. Perkembangan teknologi dan informasi membawa dampak positif dan negatif, seiring dengan perkembangan teknologi dan informasi tersebut jumblah data dan informasi yang tersedia terus meningkat dengan sangat pesat. Pesatnya pertumbuhan jumlah data dan informasi merupakan permasalahan umum yang perlu ditangani. Seperti halnya pada pengklasifikasian dokumen untuk data abstrak Tugas Akhir di Fakultas Teknik Universitas Muhammadiyah Jember.

Abstrak secara ringkas merupakan representasi dari isi dalam dokumen yang dapat bersifat akurat. Abstrak menggunakan berbagai frasa dalam dokumen yang memuat potongan teks yang dibuat oleh pembuat ringkasan, dan bukan merupakan kutipan langsung dari penulis. Abstak tersebut mendefinisikan istilah-istilah secara lebih rinci dan jelas dalam setiap bidang penelitian tugas akhir. Klasifikasi dengan menggunakan abstrak masih sangat sulit jika dilakukan secara

manual. Banyaknya kata, kemiripan kata antara satu abstrak dengan abstrak lainya dan keanekaragaman topik dalam abstrak menjadi alasan sulitnya klasifikasi secara manual. Pengidentifikasian pola dan pencarian informasi kontelektual diperlukan untuk proses klasifikasi. Sehingga yang dapat dilakukan untuk mencapai hasil tersebut adalah dengan menggunakan *N-gram* yang merupakan fitur penting dalam berbagai proses *text maining.* Namun pada penelitian sebelumnya yang dilakukan oleh Feni Shofiya, (2020) dengan judul "Perbandingan Algoritma *Support Vector Machine (SVM)* Dan *Multinomial Naive Bayes (MNB)* Dalam Klasifikasi Abstrak Tugas Akhir (Studi Kasus : Fakultas Teknik Universitas Muhammadiyah Jember)", penelitian tersebut hanya dilakukan dengan mengklasifikasian tugas akhir dan membandingkan 2 algoritma. Sehingga, untuk penelitian ini apabila diteliti lebih lanjut didapatkan hasil klasifikasi abstrak Tugas Akhir dengan menggunakan *N-gram* sebagai fitur yang akan digunakan dalam proses pengklasifikasian.

N-gram merupakan sejumlah substring atau penggalan kata yang diperoleh dari suatu kalimat, metode *N-gram* digunakan untuk menghasilkan kata atau karakter. Metode *N-gram* dapat juga digunakan untuk mengekstrak fragmen karakter dari kata yang kemudian dapat dibentuk berdasarkan kata sebelum dan sesudahnya. *N-gram* dapat dibagi menjadi beberapa jenis berdasarkan jumlah segmen kata atau substring yang dihasilkan. Jenis *N-gram* seperti *Unigram, Bigram, Trigram,* dan seterusnya. tergantung pada nomor n dalam *N-Gram*. (Nurhidayat & Dewi, 2023).

Seleksi dan ekstraksi fitur merupakan bagian penting dari pengolahan suatu data. seleksi fitur yaitu sebuah proses pemilihan fitur subset dari fitur-fitur yang ada dalam data. Pemilihan fitur dilakukan untuk mengurangi sebagian besar *noise* yang tidak relevan, yang akan mengakibatkan banyak terjadi kesalahan dalam proses pengklasifikasi. Dengan dilakukannya pemilihan fitur, dapat mengurangi dimensi ruang fitur dan meningkatkan perform dalam klasifikasi teks. Sedangkan ekstraksi fitur yaitu salah satu teknik penting dalam reduksi data untuk menemukan fitur yang paling penting agar dapat dengan mudah untuk digunakan lebih lanjut dalam menganalisis dan melatih suatu model (Mulyani dkk., 2021).

Penelitian tentang data mining memiliki banyak metode salah satunya Algoritma *K-Nearest Neighbors. K-Nearest Neighbors* (KNN) yang termasuk

dalam model *supervised learning* pada algoritma *machine learning* klasifikasi, yaitu Algoritma dalam *machine learning* di mana model belajar dari data yang sudah memiliki label atau kelas yang ditentukan sebelumnya. *K-Nearest Neighbors* adalah metode klasifikasi yang memanfaatkan jarak ketetanggan terdekat atau kemiripan antara suatu data dengan data lainnya yang akan digunakan dalam proses klasifikasi nantinya (Nurhidayat & Dewi, 2023).

Penggunaan metode *K-Nearest Neighbors* pada penelitian ini, didasari pada penelitian sebelumnya oleh Louis Madaerdo Sotarjua & Dian Budhi Santosa, yang berfokus pada klasifikasi dan perbandingan kinerja Algoritma *K-Nearest Neighbors (KNN), Decision Tree* dan *Random Forest.* Pada hasil klasifikasi disimpulkan bahwa tingkat akurasi terbesar untuk algoritma *K-Nearest Neighbors (KNN)* sebesar 86.57 %, untuk algoritma *Decision Tree* sebesar 85.29%, dan algoritma *Random Forest* sebesar 86.37%. (Sotarjua & Santosa, 2022). Dengan mempertimbangkan aspek dari hasil penelitian tersebut juga beberapa penelitian lainnya, penelitian ini dianggap akan lebih relevan dengan menggunaan algoritma *K-Nearest Neighbors (KNN)* sebagai proses klasifikasi otomatis.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka dapat disimpulkan bahwa rumusan masalah yang dapat diambil adalah :

1. Bagaimana penggunaan *N-gram* dapat mempengaruh hasil akurasi pada klasifikasi abstrak Tugas Akhir di Fakultas Teknik Universitas Muhammadiyah Jember dengan menggunakan metode *K-Nearest Nightbors* (KNN)?
2. Apakah menggunaan kombinasi *ekstraksi fitur* dan *seleksi fitur* dapat meningkatkan nilai akurasi dibandingkan dengan tidak menggunakan kombinasi tersebut?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Mengetahui sejauh mana pengaruh *N-gram* dalam mempengaruhi hasil klasifikasi dengan menggunakan metode *K-Nearest Nightbors* (KNN).

2. Mengetahui apakah nilai hasil akurai, presisi dan *Recall* klasifikasi dapat mengalami peningkatan jika menggunaan kombinasi ekstraksi dan seleksi fitur.

1.4 Manfaat Penelitian

Manfaat yang dapat diambil dari penelitian ini adalah :

1. Memberi pemahaman lebih terkait *N-gram* dalam analisis *data mining* untuk bentuk *text*.
2. Memberikan pemahaman mengenai pebandingan sebuah klasifikasi dengan menggunakan *N-gram* atau tidak.
3. Memberikan kontribusi untuk peneliti selanjutnya dalam pengembangan metode analisis teks dan klasifikasi.

1.5 Batasan penelitian

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

1. Penelitian ini berfokus pada penggunaan abstrak Tugas Akhir di Fakultas Teknik Universitas Muhammadiyah Jember.
2. Jumlah dataset yang digunakan untuk penelitian ini sebanyak 100 *record* dengan 20 *record* untuk setiap masing masing kelas.
3. Sumber data yang digunakan terbatas pada tugas akhir bagian abstrak yang ada pada *repositori* Universitas Muhammadiyah Jember, penggunaan sumber data lainya seperti literatur dan observasi tidak termasuk.
4. Penelitian ini mengklasifikasian menjadi 5 kelas , yaitu Teknik Informatika, Teknik Mesin, Teknik Elektro, Teknik Sipil, dan Menejemen Informatika
5. Penelitian ini hanya berfokus pada penggunaan klasifikasi metode *K-Nearest Nighboars* (KNN) sebagai metode metode utama, sehingga algoritma klasifikasi lainnya tidak dapat dipertimbangkan.
6. Penelitian ini menggunakan bahasa pemrograman *python* dan *tools jupyter nootbook.*

7. Penelitian ini menggunakna *library Natural Language Toolkit* (NLTK)
8. Menggunakan *library stemming* sastrawi
9. Menggunakan *library sklearn* untuk pengimplementasi algoritma *K-Nearest Neighbors* (KNN).