# BAB I
# PENDAHULUAN

## 1.1 Latar Belakang

Keanekaragaman serta jumlah dokumen abstrak tugas akhir pada saat ini sudah berkembang sangat pesat. Hal tersebut menyebabkan pencarian informasi pada dokumen abstrak tugas akhir semakin sulit. Bentuk dokumen abstrak tugas akhir saat ini tidak hanya berupa kertas, melainkan juga berupa digital. Dokumen digital tersebut memiliki beragam jenis format seperti *.txt, .pdf* dan sebagainya. Dokumen dalam bentuk digital akan mudah untuk di duplikasi. Dokumen digital juga mudah untuk penyimpanannya, tidak seperti dokumen kertas yang memakan banyak tempat.

Jumlah dokumen abstrak tugas akhir yang terus meningkat menyebabkan dokumen tersebar tidak terorganisir dengan baik karena terjadinya penumpukan. Hal ini akan menyulitkan seseorang dalam mendapatkan informasi yang diinginkan. Solusi untuk mengatasi permasalahan ini adalah dengan menerapkan metode yang dapat mengklasifikasikan dokumen abstrak tugas akhir berdasarkan kesamaan isi dokumen.

Klasifikasi ialah sebuah teknik untuk memproses pengelompokan sejumlah data ke dalam kelas tertentu yang telah dilabeli berdasarkan kemiripan isi, sifat dan pola dengan membandingkannya dengan kelas yang telah ada. Penelitian ini merupakan penelitian lanjutan dari (Radili, 2016). Pada penelitian tersebut tidak menggunakan *filtering* pada tahap *text preprocessing*.

Penelitian mengenai *winnowing* sebelumnya telah dilakukan oleh (Kurniawati & Wicaksana, 2008) yang menyimpulkan bahwa algoritma *winnowing* lebih baik daripada algoritma *manber* dalam mencari kemiripan isi dokumen, karena algoritma *winnowing* menghasilkan informasi terkait posisi *fingerprint* pada dokumen serta memberikan jaminan diperolehnya dokumen yang mirip. Penelitian mengenai pengelompokan dokumen sebelumnya diteliti oleh (Sanjaya & Absar, 2015) yang berkesimpulan bahwa algoritma *winnowing*

digunakan untuk menghasilkan *fingerprint*, sehingga apabila karakter kata yang muncul sebagai *fingerprint* antar dokumen tidak sama, maka proses klasifikasi tidak relevan, tingkat akurasi yang didapat adalah 80%. Penelitian selanjutnya dilakukan oleh (Radili, 2016) yang menerapkan algoritma *winnowing fingerprint* sebagai seleksi fitur dan *naive bayes* sebagai pengelompokan dokumen pada abstrak skripsi. Penelitian ini berhasil mendapatkan tingkat akurasi terbaik, yaitu 84,76% dan juga menghasilkan nilai *k-gram* terbaik, yaitu 8. Penelitian yang dilakukan oleh (Ridho, 2013) menyimpulkan bahwa nilai bilangan prima terbaik adalah 2 dan untuk nilai *window* terbaik adalah 8. Penilitian yang di lakukan oleh (Abghari, 2013) menyimpulkan penggunaan pembobotan *TF-IDF* dapat meningkatkan kinerja *precision, recall,* serta *F1-measure* pada *multinomial logistic regression*

Sedangkan penelitian mengenai *logistic regression* sebelumnya diteliti oleh (Al-tahrawi, 2015) yang menyimpulkan bahwa *logistic regression* memiliki kinerja klasifikasi yang sangat akurat dengan *precision* 96,46, *recall* 91,67, *F1-measure* 94,0171. Hasil penelitian yang dilakukan oleh (Rajagukguk, 2015) menyimpulkan bahwa metode *logistic regression* lebih baik daripada *naive bayes* dengan persentase akurasi 83,33% dan 81,75%. (Rianto & Wahono, 2015) menyimpulkan bahwa tingkat akurasi *logistic regression* menunjukkan hasil yang paling baik ketimbang algoritma *naive bayes*. (Jurafsky & Martin, 2015) menyatakan *Logistic Regression* termasuk ke dalam *family of classifiers*. *Logistic Regression* dikenal sebagai pengklasifikasi ekponensial atau *log-linier*. Secara teknis, *Logistic Regression* mengacu pada klasifikasi yang mengklasifikasikan observasi menjadi salah satu dari dua kelas. *Multinomial Logistic Regression* digunakan saat mengelompokkan ke dalam lebih dari dua kelas. Berdasarkan penelitian yang telah disebutkan diatas, maka pada penelitian ini akan melakukan "Penerapan Metode *Winnowing Fingerprint* dan *Multinomial Logistic Regression* pada pengelompokan dokumen abstrak tugas akhir".

Penelitian ini ditujukan untuk mengelompokan dokumen sesuai dengan kelasnya. Pada penelitian ini penulis menggunakan *winnowing fingerprint* untuk seleksi fitur dan *TF-IDF* untuk proses pembobotan, serta *multinomial logistic*

*regression* sebagai pengklasifikasi. Pemilihan metode *multinomial logistic regression* dikarenakan menurut (Jurafsky & Martin, 2015) yang menyatakan bahwa berdasarkan penelitian yang telah mereka lakukan berpendapat bahwa pengklasifikasi diskriminatif seperti *multinomial logistic regression* atau yang biasa dikenal sebagai *the maximum entropy* (*MaxEnt classifier*) kebanyakan menghasilkan akurasi yang lebih akurat.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka didapat sebuah rumusan masalah, "Bagaimana menerapkan metode *winnowing fingerprint* dan multinomial *logistic regression* untuk pengelompokan dokumen abstrak tugas akhir?".

## 1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam pembuatan tugas akhir ini adalah sebagai berikut:

1. Menerapkan metode *multinomial logistic regression* untuk pengelompokan dokumen abstrak tugas akhir.
2. Menguji tingkat akurasi dari penerapan metode multinomial *logistic regression*.

## 1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan adalah data sekunder dari penelitian Radili tahun 2016 yaitu berasal dari situs ITS www.digilib.its.ac.id/repository/undergraduate (Institut Teknologi Sepuluh Nopember) *Digital Repository*, Fakultas Teknologi Informasi, Jurusan Teknik Informatika.
2. Bagian dalam abstrak yang digunakan adalah judul, konten abstrak dan kata kunci (*keyword*).

## 1.5 Sistematika Penulisan

**BAB I PENDAHULUAN**

Pembahasan berisi mengenai hal umum dari penelitian atau Tugas Akhir ini yang terdiri dari latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, dan sistematika penulisan.

**BAB II LANDASAN TEORI**

Berisi tentang pengetahuan dasar dari penelitian atau Tugas Akhir yang penulis lakukan. Baik itu berupa pengertian *Text Mining*, *Winnowing Fingerprint*, *TF-IDF* dan *Multinomial Logistic Regression*.

**BAB III METODOLOGI PENELITIAN**

Berisi penjelasan tahap-tahap penelitian atau Tugas Akhir yang penulis lakukan. Mulai dari identifikasi, perumusan masalah, pengumpulan data, analisa dan perancangan, implementasi dan pengujian, serta kesimpulan dan saran.

**BAB IV ANALISA DAN PERANCANGAN**

Berisi tentang analisa dari aplikasi yang akan dibangun dan metode *Winnowing Fingerprint*, *TF-IDF* dan *Multinomial Logistic Regression* yang dilakukan dalam penelitian ini.

**BAB V IMPLEMENTASI DAN PENGUJIAN**

Berisi implementasi dari hasil analisa dan perancangan aplikasi yang dibangun dan pengujian dari metode yang digunakan dalam pembangunan aplikasi tersebut.

**BAB VI PENUTUP**

Berisi kesimpulan dari hasil penelitian yang dilakukan dan saran yang diberikan atas hasil penelitian untuk peneliti selanjutnya.