# bab i

# pendahuluan

## Latar Belakang

Plagiarisme merupakan tindakan kejahatan yang sudah tidak asing lagi dalam dunia akademik, khususnya di Indonesia. Plagiarisme merupakan perbuatan yang tidak baik disebabkan mengambil karya seperti ide, gagasan, dan pikiran orang lain tanpa izin, tanpa menyatakan karya yang diambilnya merupakan karya orang lain atau tidak menyantumkan sumbernya, dan mengakui bahwa karya itu miliknya. Tindakan tersebut sama seperti tindakan mencuri (Sunardi *et al.*, 2018). Tindakan plagiarisme ini semakin mudah dengan semakin pesatnya perkembangan dunia teknologi, serta semakin mudahnya mendapatkan suatu informasi. Hal ini menyebabkan banyak orang-orang yang mengambil karya orang lain dari internet, tanpa mencantumkan sumbernya.

Dalam dunia pendidikan, khususnya tingkat perguruan tinggi, kesamaan dokumen atau plagiarisme bukanlah suatu hal yang baru. Namun sudah terjadi semenjak teknologi belum secanggih ini. Pelajar maupun mahasiswa ketika mengerjakan tugas-tugas sekolah maupun kuliah sangat jarang mencantumkan sumbernya atau referensi yang digunakan dalam pengerjaan tugas, yang menyebabkan banyaknya kesamaan tugas siswa maupun mahasiswa satu sama lain. Pelajar maupun mahasiswa banyak melakukan manipulasi pembuatan tugas seperti karya ilmiah, makalah, kuliah kerja praktek, maupun skripsi yang menyebabkan kreatifitas seseorang tidak tersalurkan dengan baik.

Semakin maraknya tindakan plagiarisme, salah satunya pembuatan skripsi di kampus Budi Luhur, sehingga untuk mengatasi tindakan plagiarisme ini perlu dilakukan pencegahan awal untuk mengurangi praktik plagiarisme. Pendeteksian dapat dilakukan secara manual, namun tidak efektif sebab dibutuhkan waktu yang cukup lama, karena banyaknya dokumen skripsi yang perlu diperiksa nilai kemiripannya dengan dokumen skripsi yang sudah dipublikasi ke perpustakaan Budi Luhur. Cara menangani hal tersebut yaitu dengan membuat sebuah sistem yang dapat mendeteksi nilai plagiarisme dari dokumen skripsi mahasiswa Budi Luhur. Sistem ini di fokuskan untuk pendeteksian awal, sehingga dokumen yang diperiksa dari sistem ini adalah dokumen abstrak, di mana data dokumen abstrak dari skripsi mahasiswa Budi Luhur yang sudah dipublikasi ke perpustakaan dijadikan sebagai set data dari penelitian ini.

Beberapa penelitian sebelumnya telah membuat pendeteksian plagiarisme menggunakan beberapa algoritme yang mempunyai fungsi sama sebagai dokumen *fingerprint*, tetapi permasalahan yang sering terjadi pada hasil atau tingkat keakurasian (Sunardi *et al.*, 2018). Penggunaan model Bayesian juga telah digunakan, namun tidak sepenuhnya dapat mendeteksi kemiripan. Pada penelitian yang dilakukan oleh (Rahmatulloh *et al.*, 2019) dalam pendeteksian plagiarisme dokumen berbahasa indonesia, penelitian ini menggunakan kinerja algoritme *fingerprint* dan *winnowing*, hasil tertinggi pada analisis tes akurasi algoritme *fingerprint* sebesar 92,8% dengan nilai *threshold* 0,1 dan nilai *N-gram* berada di 3, algoritme *winnowing* sebesar 91,8%. hasil yang berbeda ditunjukkan pada tingkat relevansi akurasi dengan topik, hasil akurasi dari algoritme *winnowing* mendapatkan korelasi yang lebih kuat 37,1% dibandingkan *fingerprint* 33,6%.

Pada penelitian ini sistem yang akan dibuat menggunakan metode *N-gram*, *Jaccard Similarity,* dan algoritme *Winnowing*. *N-gram* merupakan algoritme yang akan digunakan untuk mengambil potongan huruf sejumlah n dan mempunyai pengaruh yang tinggi terhadap hasil *similarity*. Algoritme yang akan digunakan untuk mengetahui tingkat kemiripan teks atau dokumen yaitu *Jaccard Similarity* atau sering disebut *Jaccard Coefficient*. *Jaccard Similarity* merupakan algoritme yang berfugsi untuk membandingkan antar dua dokumen dengan menghitung kemeripan atau perbedaan dari dokumen tersebut (Sunardi *et al.*, 2018). Untuk menjalankan *n-gram* dan *Jaccard Similarity* diperlukan algoritme yang berfungsi sebagai dokumen *fingerprint* dan algoritme yang akan digunakan untuk mendukung tersebut yaitu algoritme *winnowing*. Algoritme *winnowing* merupakan algoritme yang dapat digunakan untuk membantu mendeteksi kemiripan kata atau dokumen yang akan digunakan sebagai suatu cara untuk mendeteksi adanya suatu tindak plagiariasme. Diharapkan dengan adanya alat pendeteksi plagirisme menggunakan metode *n-gram* dan *Jaccard similarity* terhadap algoritme *winnowing* dapat memberikan keakurasian yang tepat dan mengatasi tindakan suatu plagiarisme yang terjadi di dunia akademik, tepatnya di kampus Budi Luhur.

## Perumusan Masalah

Berdasarkan latar belakang yang telah disebutkan di atas maka didapatkan rumusan masalah yaitu :

1. Bagaimana menerapkan metode *N-gram*, *Jaccard Similarity,* dan algoritme *Winnowing* untuk pemeriksaan kesamaan abstrak pada dokumen skripsi mahasiswa Budi Luhur?
2. Bagaimana mengukur tingkat persentase kesamaan dalam penerapan metode *N-gram*, *Jaccard Similarity,* dan algoritme *Winnowing*?

## Batasan Masalah

Adapun batasan atau ruang lingkup masalah antara lain sebagai berikut :

1. Aplikasi menggunakan bahasa pemrograman PHP.
2. *Platform* yang digunakan hanya berbasis *web.*
3. Deteksi plagiarisme hanya dilakukan pada dokumen abstrak.
4. Fitur *import* hanya dapat mengenali file masukan berupa *.pdf.*
5. Hanya dapat mengenali tulisan di dalam file masukan.
6. Waktu pemprosesan meningkat seiring dengan jumlah data yang diproses.
7. Pemeriksaan dokumen hanya dilakukan untuk mendapatkan persentase *similarity* dari dokumen yang diuji, tidak untuk mendapatkan *output* berupa keterangan plagiat atau tidak plagiat.

## Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut :

1. Menerapkan metode *n-gram* untuk mengambil potongan karakter huruf pada dokumen yang akan dicocokan.
2. Menerapkan metode *Jaccard similarity* untuk menghitung berapa persen kesamaan pada sebuah dokumen yang diproses menggunakan algoritme *winnowing.*
3. Mengimplementasikan rancangan model yang dibuat ke dalam aplikasi berbasis web.

## Manfaat

Adapun manfaat dari penelitian ini adalah sebagai acuan untuk pendeteksi awal plagiarisme pada dokumen skripsi mahasiswa Budi Luhur yang bersumber dari dokumen abstrak. Sehingga diharapkan tidak terjadi lagi plagiarisme dalam penulisan dokumen skripsi di Universitas Budi Luhur.

## Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan tugas akhir ini adalah sebagai berikut:

**BAB I : PENDAHULUAN**

Bagian ini berisi tentang latar belakang, rumusan masalah, batasan masalah, manfaat dan tujuan penelitian, dan juga membahas mengenai sistematika penulisan.

**BAB II : LANDASAN TEORI**

Bagian ini berisi tentang algoritme dan metode yang akan dibahas, serta teori-teori yang berkaitan dengan penelitian ini, yaitu pengertian dan pemahaman plagiarisme, *text mining*, data *mining, preprocessing, n-gram,* *Jaccard Similarity,* Algoritme *Winnowing*, serta studi literatur.

**BAB III : METODOLOGI PENELITIAN**

Bagian ini berisi tentang sumber data penelitian, penerapan atau tahapan metode yang digunakan. Bab ini juga berisi tentang rancangan pengujian dari ekstraksi informasi yang didapat.

**BAB IV : HASIL DAN PEMBAHASAN**

Bagian ini berisi mengenai lingkungan percobaan sistem yang dibuat, implementasi metode, *flowchart* tahapan metode, dan uraian algoritme pada proses, serta analisa pengujian sistem yang telah dibangun apakah data hasil pengelompokan yang didapat sudah sesuai dan relevan.

**BAB V : PENUTUP**

Bagian ini berisi tentang kesimpulan yang dapat ditarik dari penelitian dan saran untuk pengembangan lebih lanjut mengenai topik terkait dalam penelitian berikutnya.