

# **BAB I**

## **PENDAHULUAN**

### **I.1. Latar Belakang**

Dalam dunia plagiarisme (penjiplakan) yang semakin marak terjadi di era modern saat ini, dimana dunia plagiarisme mudah dilakukan karena semakin majunya teknologi digital seperti dalam pembuatan laporan tulisan ilmiah di bidang akademik baik tingkat sekolah maupun perguruan tinggi. Terjadinya plagiarisme disebabkan karena kebiasaan peneliti yang malas dan ingin serba cepat dalam menyelesaikan tugas-tugasnya, sehingga kemungkinan peneliti akan mencari referensi yang terkait baik di internet maupun perpustakaan (Andry Hery Purba, 2017).

Untuk menyelesaikan deteksi dokumen teks, kita dapat menggunakan beberapa macam Algoritma seperti rabin karp, K-means, serta K-gram untuk memecahkannya. Meskipun Algoritma penyelesaiannya banyak, namun tidak semua Algoritma tersebut bagus dalam deteksi dokumen teks.

Algoritma winnowing merupakan algoritma yang digunakan dalam deteksi kesamaan menggunakan fungsi hashing. Winnowing adalah algoritma yang digunakan untuk melakukan proses document fingerprinting. Algoritma winnowing melakukan penghitungan nilai-nilai hash dari setiap k-gram, untuk mencari nilai hash selanjutnya digunakan fungsi rolling hash. Kemudian dibentuk window dari nilai-nilai hash tersebut. Dalam setiap window dipilih nilai hash minimum. Jika ada lebih dari satu hash dengan nilai minimum, dipilih nilai hash yang paling kanan. Kemudian semua nilai hash terpilih disimpan untuk dijadikan

fingerprint dari suatu dokumen. Fingerprint ini yang akan dijadikan dasar pembandingan kesamaan antara teks yang telah dimasukkan.

Beberapa penelitian terkait dengan bahasan tulisan ini telah dilakukan oleh (Andysah Putera Utama Siahaan, Sugianto, (2017). telah melakukan penelitian mengenai Implementasi Algoritma K-gram pada Deteksi kemiripan dokumen. Tujuan dari penelitian tersebut adalah mengimplementasikan algoritma K-gram pada deteksi dokumen teks agar dapat membuktikan bahwa Algoritma K-gram dapat digunakan dalam deteksi kemiripan dokumen teks.

Beberapa penelitian terkait dengan bahasan tulisan ini telah dilakukan oleh Upper Saddle River (2017) dengan pembahasan yang sama dengan membandingkan Algoritma Winnowing dengan Algoritma K-means. Pada Algoritma K-means terdapat Kelemahan beberapa kasus yang tidak dapat terselesaikan pada deteksi kemiripan dokumen bahwa Jika hanya terdapat beberapa titik sampel data, maka cukup mudah untuk menghitung dan mencari titik terdekat dengan k titik yang diinisialisasi secara random. Namun jika terdapat banyak sekali titik data (misalnya satu milyar buah data), maka perhitungan dan pencarian titik terdekat akan membutuhkan waktu yang lama. Proses tersebut dapat dipercepat, namun dibutuhkan struktur data yang lebih rumit seperti kD-Tree atau hashing dibandingkan dengan Algoritma winnowing merupakan algoritma yang digunakan dalam deteksi kesamaan menggunakan fungsi hashing. Winnowing adalah algoritma yang digunakan untuk melakukan proses document fingerprinting. Algoritma winnowing melakukan penghitungan nilai-nilai hash dari setiap k-gram, untuk mencari nilai hash selanjutnya digunakan fungsi rolling hash. Kemudian dibentuk window dari nilai-nilai hash tersebut. Dalam setia

window dipilih nilai hash minimum. Jika ada lebih dari satu hash dengan nilai minimum, dipilih nilai hash yang paling kanan. Kemudian semua nilai hash terpilih disimpan untuk dijadikan fingerprint dari suatu dokumen. Fingerprint ini yang akan dijadikan dasar pembandingan kesamaan antara teks yang telah dimasukkan.

Dalam proses penyelesaiannya banyak terdapat Algoritma-Algoritma pencarian yang dapat diterapkan. Dalam hal ini Penulis memilih Algoritma Wnnowing. Berdasarkan uraian diatas penulis mengambil judul **"Penerapan Aplikasi Deteksi Kemiripan Dokumen Teks Menggunakan Metode Wnnowing"**

## **I.2. Ruang Lingkup Permasalahan**

Adapun ruang lingkup permasalahan yang ada dalam penulisan skripsi ini adalah sebagai berikut :

### **I.2.1 Identifikasi Masalah**

Adapun identifikasi masalah yang ada dalam penulisan skripsi ini adalah sebagai berikut :

1. Masih banyak yang belum mengetahui cara deteksi kemiripan dokumen teks dengan menerapkan Algoritma Wnnowing.
2. Belum diketahuinya akurasi yang dihasilkan dalam deteksi dokumen teks dengan menggunakan Algoritma Wnnowing.

### **I.2.2. Rumusan Masalah**

Berdasarkan latar belakang masalah yang telah diuraikan, maka dapat diselesaikan masalah yang dibahas dalam penulisan ini adalah :

1. Bagaimana menciptakan sebuah aplikasi deteksi pada sebuah dokumen?
2. Bagaimana mengetahui tingkat akurasi deteksi kemiripan dua buah dokumen teks dengan Algoritma Winnowing.

### **I.2.3. Batasan Masalah**

Adanya batasan masalah ini dapat memberi batasan pada sebuah penelitian agar tetap dalam lingkup pembahasan masalah yang ada. Adapun batasan masalahnya sebagai berikut :

1. Dalam penelitian ini hanya membahas tentang Penggunaan Algoritma winnowing Pada mendeteksi kemiripan dokumen teks.
2. Bagaimana penggunaan algoritma Winnowing untuk mencari solusi masalah deteksi tingkat kesamaan dokumen teks dengan baik.
3. Bahasa pemrograman yang digunakan adalah Java NetBeans
4. *Output* yang dihasilkan berupa simulasi aplikasi deteksi kemiripan dokumen teks yang telah dijalankan pada pemrograman Java.

## **I.3 Tujuan dan Manfaat**

### **I.3.1. Tujuan**

Dengan mengetahui rumusan masalah di atas maka tujuan dari penelitian ini adalah sebagai berikut :

1. Membangun sebuah sistem aplikasi yang mampu mendeteksi dokumen teks.

2. Menerapkan Algoritma Winnowing pada suatu aplikasi deteksi dokumen teks.

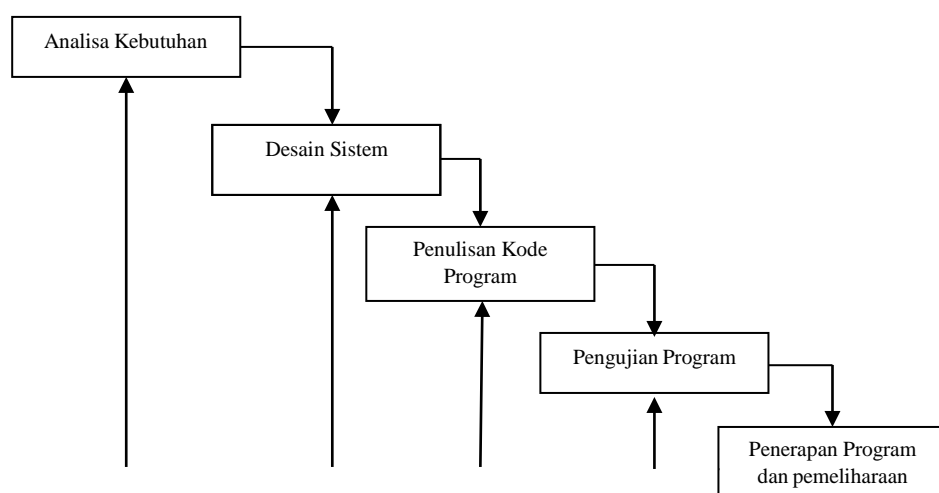
### I.3.2 Manfaat

Berdasarkan latar belakang di atas, maka dapat di deskripsikan manfaat dari penelitian ini adalah sebagai berikut :

1. Aplikasi yang dirancang dapat digunakan untuk sebagai mencari dua objek dokumen yang asli dan palsu..
2. Menguji Algoritma Winnowing dalam penggunaan sistem deteksi cek kemiripan dokumen teks.

### I.4. Metodologi Penelitian

Penelitian ini akan melalui beberapa tahapan. Tahapan dalam penelitian ini dapat di modelkan pada diagram *waterfall* yang ditunjukkan pada gambar I.1. sebagai berikut :



**Gambar I.1. Diagram Waterfall**

Adapun penjelasan dari prosedur perancangan yang dilakukan penulis dalam penelitian ini, adalah sebagai berikut:

## **1. Analisis Kebutuhan**

Pada tahapan ini merupakan analisa terhadap kebutuhan yang diperlukan untuk mencapai tujuan penelitian yang akan dilakukan dengan mempelajari teori dasar yang mendukung penelitian, pencarian dan pengumpulan data-data yang dibutuhkan. Untuk mengumpulkan data yang dibutuhkan sebagai berikut :

### *a. Pengamatan (Observation)*

Data dikumpulkan langsung dengan cara melakukan pengamatan ini penulis diberi kesempatan untuk melakukan pengamatan Metode dengan pengamatan dan penerapan secara sistematis mengenai objek yang di teliti yaitu dokumen yang sudah di uji dan sudah ada sebelumnya dan menjadikannya sebagai contoh.

### *b. Studi Pustaka*

Penulis juga melakukan studi kepustakaan, yaitu dengan mempelajari buku-buku, jurnal, e-jurnal dan website yang relevan guna memberi dan menambah pemahaman dan pengetahuan penulis, serta sebagai bahan referensi untuk menulis tugas akhir ini.

## **2. Desain Sistem**

pada tahap ini dilakukan desain perangkat lunak menggunakan pemodelan UML yaitu *use case diagram*, *activity diagram*, *sequence diagram* dan *class diagram*.

### *a. Usecase diagram*

*Use case diagram* yaitu salah satu jenis diagram pada UML yang menggambarkan interaksi antara sistem dan aktor, *use case diagram* juga dapat men-deskripsikan tipe interaksi antara sipemakai system dengan sistemnya.

b. *Class diagram*

*Class diagram* yaitu penjelasan detail tiap-tiap kelas di dalam model desain dari system yang memperlihatkan aturan-aturan dan tanggung jawab entitas yang menentukan perilaku sistem.

c. *Activity diagram*

*Activity diagram* atau diagram aktivitas yaitu salah satu jenis diagram pada UML yang dapat memodelkan proses-proses apa saja yang terjadi pada sistem.

d. *Sequence diagram*

*Sequence diagram* yaitu salah satu jenis diagram pada UML yang menjelaskan interaksi objek yang berdasarkan urutan waktu, *sequence diagram* juga dapat menggambarkan urutan atau tahapan yang harus dilakukan untuk dapat menghasilkan sesuatu seperti pada *use case diagram*.

### **3. Implementasi Sistem**

Untuk dapat dimengerti oleh komputer atau PC, maka desain tersebut harus diubah bentuknya menjadi bentuk yang dapat dimengerti oleh komputer atau PC, yaitu melalui proses coding yang merupakan bentuk bahasa pemrograman. Tahap implementasi merupakan penerapan dari tahap desain sistem..

### **4. Pengujian Sistem**

Setelah sistem yang dirancang selesai diimplementasikan menjadi sebuah aplikasi, tahap yang selanjutnya yaitu tahap pengujian, dari spesifikasi, desain dan

pengkodean. Dalam penelitian ini pengujian sistem dilakukan melalui pengujian black – box terhadap seluruh fungsi dalam aplikasi. Pengujian black – box merupakan salah satu pengujian aplikasi atau perangkat lunak yang berfokus pada persyaratan fungsional perangkat lunak..

## **5. Perawatan Sistem**

Ketika dijalankan mungkin saja masih ada errors yang mungkin tidak ditemukan sebelumnya, atau ada penambahan fitur baru yang belum ada pada sistem tersebut. Pengembangan diperlukan ketika adanya perubahan sistem yang lebih mudah di pahami masyarakat, atau ketika ada kendala saat aplikasi yang di gunakan bermasalah.

### **I.5. Kontribusi Penelitian**

Kontribusi keilmuan bertujuan untuk menjabarkan perbedaan diantara penelitian terdahulu dengan penelitian saat ini.

Penelitian yang dilakukan oleh Zakarias Situmorang (2017) “Analisis Perbandingan Algoritma Rabin-Karp Dan Levenshtein Distance Dalam Menghitung Kemiripan Teks”. Pada penelitian ini implementasinya tidak menggunakan struktur hash seperti kebanyakan solusi deteksi dokumen teks lain. Penulis menggunakan matriks untuk struktur data dari deteksi kesamaan itu sendiri dan array of matrix

Penelitian yang dilakukan oleh Nur Alamsyah (2017) “PERBANDINGAN ALGORITMA WINNOWING DENGAN ALGORITMA RABIN KARP UNTUK MENDETEKSI PLAGIARISME PADA KEMIRIPAN TEKS JUDUL SKRIPSI” Terdapat beberapa metode untuk membuat deteksi deteksi

plagiarisme pada kemiripan teks judul skripsi. Bisa menggunakan algoritma oberself, dll. Penulis menggunakan metode winnowing, salah satu sistemnya position independence, yaitu pencocokan teks file seharusnya tidak bergantung pada posisi kata-kata sehingga kata dengan urutan posisi berbeda masih dapat dikenali jika terjadi kesamaan.

berdasarkan penjabaran dari beberapa penelitian diatas perbedaan penelitian yang dilakukan penulis menggunakan metode Winnowing dimana algoritma Winnowing ini mendeteksi dokumen teks tercepat. Dalam hal ini penulis memilih menggunakan algoritma winnowing karena algoritma winnowing mampu menyelesaikan contoh kasus deteksi kemiripan dokumen teks dengan baik.

## **I.6. Sistematika Penulisan**

Adapun sistematika penulisan yang diajukan dalam skripsi ini adalah sebagai berikut :

### **BAB I : PENDAHULUAN**

Pada bab ini menerangkan tentang latar belakang, ruang lingkup permasalahan, tujuan dan manfaat, metode penelitian dan sistematika penulisan.

### **BAB II : TINJAUAN PUSTAKA**

Bab ini berisi teori-teori yang berkaitan dengan Penggunaan algoritma Winnowing pada deteksi kemiripan dokumen teks.

### **BAB III : ANALISIS DAN PERANCANGAN**

Bab ini berisi tentang proses-proses perancangan aplikasi dengan teori yang berkaitan, perancangan alur kerja aplikasi, antar muka pemakai, dan perancangan penelitian.

**BAB IV : HASIL DAN UJI COBA**

Bab ini berisi tentang hasil implementasi aplikasi, serta analisa hasil pengujian.

**BAB V : KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan yang diambil dari analisa, desain, dan implementasi serta uji coba yang telah dilakukan. Selain itu bab ini juga berisi saran yang bermanfaat dalam pengembangan aplikasi di waktu yang akan datang.