

# BAB I

## Pendahuluan

### 1.1. Latar Belakang

Plagiarisme menjadi satu permasalahan yang cukup serius untuk dunia pendidikan Indonesia sekarang, karena plagiarisme sudah termasuk kedalam tindakan kriminal dengan mengambil karya orang. Semakin berkembangnya teknologi dan zaman, semakin banyak pula cara yang dapat digunakan untuk melakukan plagiarisme dan dapat berdampak kepada generasi muda Indonesia yang bermental plagiarisme[8].

Penelitian ini menggunakan *Text Alignment* dengan metode *N-Gram* dan algoritma *Winnowing*. Metode ini dipilih karena algoritma *Winnowing* sendiri memiliki keunggulan dibandingkan algoritma yang menggunakan *Fingerprint* lainnya seperti algoritma *Manber* dan algoritma *Rabin-Karp*. Kemudian akan mengukur akurasi, *Precision*, *Recall* untuk memperjelas akurasi yang di dapat dalam metode ini.

Proses pada penelitian ini yaitu melakukan *preprocessing* yang dimana menggunakan *case folding*, *filtering* dan *white insensitivity*. Kemudian pembentukan rangkaian *n-gram* untuk selanjutnya dapat menghitung nilai *hash*. Dari nilai *hash* yang didapat, bisa dilanjutkan untuk membentuk *window* dari rangkaian *hash* yang sudah didapat dan dilanjutkan dengan pemilihan *fingerprint*. Nilai *fingerprint* yang dipilih akan dicocokkan menggunakan *jaccard coefficient* untuk mengukur kemiripan dari teks. Agar penelitian ini lebih detail untuk diklasifikasikan maka menggunakan *confusion matrix* dalam menentukan pengukuran dalam bentuk *accuracy, recall, precision* dan *f1-score*[9].

## 1.2. Rumusan Masalah

Berdasarkan latar belakang, perumusan permasalahan yang didapat yaitu pengimplementasian dari karakter *N-gram* yang dirangkai sehingga dapat diolah dengan algoritma *Winnowing* dalam mendeteksi plagiarisme.

Performansi algoritma *Winnowing* dalam melakukan deteksi plagiarisme, dimana algoritma *winnowing* sendiri adalah yang berperan dalam sistem dan pengaruh dari *preprocessing* terhadap performansi sistem.

## 1.3. Batasan Masalah

Batasan masalah pada penelitian kali ini antara lain:

- a. Menggunakan korpus dari PAN 2013 dan PAN 2014 yang berformat .txt.
- b. Dokumen yang digunakan hanya dokumen berbahasa inggris.
- c. Setelah mengolah dokumen, olahan tersebut ada dua yaitu HasilSistem (Olahan program) dan Training (Olahan Manual).
- d. *N-gram* yang digunakan masih  $N=4$ , Karena  $N=4$  mendapatkan hasil yang lebih beragam sedangkan jika menggunakan yang lain akurasi yang didapat jauh lebih rendah.

## 1.4. Tujuan

Penelitian ini memiliki tujuan untuk mengimplementasikan dan menganalisis sistem deteksi plagiarisme dengan menggunakan *Text Alignment* dengan karakter *N-Gram* dan algoritma *Winnowing*, Dimana yang paling banyak berperan dalam penelitian ini yaitu *Window* dan *fingerprint* pada algoritma *Winnowing* tersebut dalam mendeteksi plagiarsme.

Mengetahui pengaruh *preprocessing* dan *window* pada *winnowing* terhadap performansi sistem yang dibangun. Untuk mengetahui performansi pada penelitian ini menggunakan *Confusion Matrix* dalam mengukur *Recall*, *Precision* dan *F1-score*.

## 1.5. Organisasi Tulisan

Adapun organisasi penulisan laporan tugas akhir ini disusun sesuai dengan rencana sebagai berikut:

- BAB I Pendahuluan

Di bab pertama ini menjelaskan mengenai gambaran umum dari tugas akhir. Yang dijelaskan pada bab ini diantaranya latar belakang, perumusan masalah, batasan masalah, dan tujuan penelitian.

- **BAB II Studi Literatur**

Bagian ini berisi teori/studi/literatur yang mendukung dalam penyelesaian tugas akhir ini.

- **BAB III**

Bagian ini menjelaskan bagaimana proses dari perhitungan *accuracy*, *Precision*, *Recall*, dan *F1-Score* dalam pendeteksi plagiarisme.

- **BAB IV Evaluasi**

Bagian ini berisi dua sub-bagian, yaitu Hasil Pengujian dan Analisis Hasil Pengujian. Pengujian dan analisis yang dilakukan selaras dengan tujuan TA sebagaimana yang dijelaskan pada pendahuluan.

- **BAB V Kesimpulan**

Pada bab terakhir ini menyajikan kesimpulan dan saran dari hasil penelitian tugas akhir yang sudah dikerjakan.