

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Teks adalah salah satu media komunikasi yang digunakan manusia dalam kehidupan sehari-hari. Hampir semua aktivitas yang kita lakukan tidak terlepas dari teks terutama bagi seorang mahasiswa. Teks memiliki peran penting dalam bidang pendidikan [1]. Selama menjalani masa perkuliahan mahasiswa dituntut untuk menulis berbagai macam dokumen seperti jurnal, makalah atau pun skripsi. Namun kenyataan di lapangan saat ini hal yang sering terjadi pada saat penulisan dokumen adalah kesalahan penulisan atau *typographical error*.

*Typographical error* tidak bisa lepas dari mahasiswa. Sejak dini mahasiswa telah dilatih kemampuan menulisnya dengan membuat laporan praktikum ataupun makalah untuk memenuhi tugas mata kuliah tertentu. Namun hal tersebut masih kurang karena pengetahuan dasar mahasiswa tentang cara penulisan Bahasa Indonesia dinilai masih kurang. Menurut Murtingsih, kesalahan penulisan kata yang tidak sesuai konteks sering dilakukan mahasiswa sebanyak 69,2% dari jumlah subjek penelitian yaitu 260 mahasiswa. Selain itu kesalahan penulisan imbuhan dan kata depan sebanyak 17,4% dari total 260 mahasiswa [2]. *Typographical error* cukup sering terjadi pada saat mahasiswa menulis jurnal, makalah atau pun skripsi. Mahasiswa sering kali mengalami revisi dikarenakan kesalahan penulisan yang tidak sesuai dengan kaidah Ejaan Yang Disempurnakan (EYD). Hal ini sesuai dengan yang ditulis oleh Javed, Juan dan Nazli bahwa kemampuan menulis lebih sulit dan paling akhir dikuasai [3].

Terdapat tiga faktor utama yang menyebabkan sering terjadinya *typographical error* pada dokumen yang dibuat oleh mahasiswa. Faktor pertama adalah kurangnya pemahaman mahasiswa dalam pembuatan atau penulisan dokumen Bahasa Indonesia yang baku dan sesuai dengan EYD. Hal ini sering sekali ditemukan pada tulisan-tulisan mahasiswa dilihat dari penempatan tanda

baca yang kurang tepat, penggunaan huruf kapital dan huruf kecil yang kurang tepat dan lain sebagainya.

Faktor kedua kurangnya kosakata baku yang dipahami oleh mahasiswa. Ini adalah masalah yang paling sering dihadapi mahasiswa. Kurangnya intensitas waktu membaca literatur membuat pengetahuan mahasiswa mengenai kosakata baku dinilai kurang.

Faktor ketiga yaitu kesulitan dalam menyatukan ide mahasiswa dan kutipan dari bacaan yang digunakan. Pada penulisan skripsi atau tugas akhir, mahasiswa dituntut untuk menulis sesuatu berdasarkan fakta yang diambil dari sumber-sumber terpercaya seperti jurnal, *paper*, skripsi atau tugas akhir yang sudah ada maupun buku yang berkaitan. Untuk menghindari plagiarisme mahasiswa harus menggunakan kutipan tersebut dengan merangkum atau menyatukan dengan idenya sendiri tanpa mengurangi makna dari kutipan tersebut.

Untuk membantu mahasiswa dalam mengatasi masalah kesalahan penulisan dokumen tersebut diperlukan suatu aplikasi yang dapat melakukan pengecekan kesalahan penulisan atau pengetikan teks. Saat ini sudah terdapat beberapa aplikasi yang serupa namun masih sangat kurang untuk pengecekan dokumen berbahasa Indonesia. Untuk itu dibuat suatu aplikasi *website* sederhana yang dapat melakukan pengecekan kesalahan penulisan atau pengetikan teks pada dokumen berbahasa Indonesia yang menerapkan *natural language processing* dengan menggunakan algoritma Jaro-Winkler *distance*. Dengan adanya aplikasi ini, diharapkan dapat membantu mahasiswa dalam mengatasi masalah kesalahan penulisan atau pengetikan teks.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang di atas, dapat dirumuskan masalah sebagai berikut:

1. Bagaimana membuat model NLP menggunakan algoritma Jaro-Winkler *distance* yang dapat melakukan identifikasi *typographical error* pada dokumen Berbahasa Indonesia?

2. Bagaimana membangun aplikasi yang dapat mengidentifikasi *typographical error* pada dokumen Berbahasa Indonesia?

### **1.3. Batasan Masalah**

Dalam penelitian permasalahan masih terbuka luas dan dapat melebar, maka dari itu dibuat batasan sebagai acuan untuk menjadi fokus penelitian ini. Terdapat beberapa batasan masalah yang digunakan yaitu:

1. Identifikasi *typographical error* hanya berlaku untuk kata Berbahasa Indonesia.
2. Identifikasi *typographical error* tidak berlaku untuk pengecekan kata entitas.
3. Penelitian hanya mencakup identifikasi *typographical error* atau kesalahan dalam susunan huruf pada kata. Tidak mencakup kesalahan secara semantik ataupun kontekstual.
4. Dataset yang digunakan adalah dataset KBBI yang diambil dari Github Jim Geovedi dan disempurnakan penulis.
5. Algoritma yang digunakan adalah Jaro-Winkler *distance*.
6. Aplikasi hanya dapat mengekstrak teks sederhana tanpa format teks nya.

### **1.4. Tujuan Penelitian**

Tujuan yang diharapkan dapat tercapai dari penelitian ini adalah sebagai berikut:

1. Membuat model NLP menggunakan algoritma Jaro-Winkler *distance* yang dapat melakukan identifikasi *typographical error* pada dokumen Berbahasa Indonesia.
2. Membangun aplikasi yang dapat mengidentifikasi kesalahan *typographical error* pada dokumen Berbahasa Indonesia.

## **1.5. Sistematika Penulisan**

Secara sistematis isi dari laporan ini disusun sebagai berikut :

### **BAB I Pendahuluan**

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, dan sistematika penulisan laporan.

### **BAB II Tinjauan Pustaka**

Bab ini berisi penelitian-penelitian yang sebelumnya telah dilakukan sebelumnya serta penjelasan mengenai teori yang digunakan sebagai dasar dan acuan untuk pemecahan masalah dalam penelitian.

### **BAB III Metodologi Penelitian**

Bab ini berisi penjelasan mengenai data yang akan digunakan, perangkat yang akan digunakan selama penelitian serta langkah-langkah dalam melakukan penelitian.

### **BAB IV Analisis dan Pembahasan**

Bab ini berisi penjelasan mengenai langkah-langkah *preprocessing* data, langkah-langkah pembuatan model identifikasi *typographical error* dan implementasi algoritma Jaro-Winkler *distance* pada model serta hasil evaluasi model.

### **BAB V Implementasi Model Pada Web Service dan Aplikasi Web**

Bab ini berisi penjelasan mengenai langkah-langkah pembuatan *web service* dan pengimplementasian model pada *web service*, langkah-langkah pembuatan sistem, implementasi *web service* pada sistem serta pengujian dari sistem aplikasi yang dibuat.

### **BAB VI Penutup**

Bab ini berisi kesimpulan dari hasil penelitian yang sudah dilakukan serta saran-saran yang dapat digunakan untuk pengembangan yang lebih lanjut.