

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Seiring perkembangan zaman, teknologi *artificial intelligence* dan *machine learning* juga berkembang secara pesat [1]. Pengaplikasian *artificial intelligence* bermacam-macam, mulai dari bidang pengobatan [2], *finance* [3], *Internet of Things (IOT)* [4], hingga pemrosesan bahasa alami. Pemrosesan bahasa alami atau yang sering disebut sebagai *Natural Language Processing (NLP)* merupakan cabang dari kecerdasan buatan yang berfokus pada interaksi antara manusia dan komputer melalui bahasa manusia [5]. Teknologi ini berfungsi untuk membuat komputer dapat mengerti bahasa yang digunakan oleh manusia, mulai dari per kata hingga kalimat. Contoh dari fungsi dari NLP adalah pembuatan sistem sentimen analisis [6], mesin pencarian [7], penerjemah [8], chatbot [9,10], hingga sistem rekomendasi [11].

Pengembangan dari NLP lainnya adalah pada bidang jurnalistik [12]. Dalam bidang jurnalistik, kecepatan dalam menyajikan informasi telah menjadi tuntutan primer yang harus dihadapi oleh para praktisi media. Dalam upaya memenuhi kebutuhan akan berita yang aktual dan relevan, jurnalis sering kali terjebak dalam tekanan untuk menghasilkan konten dengan cepat, tanpa memperhatikan secara teliti penggunaan bahasa yang tepat [13]. Namun, sebagaimana disadari, bahasa bukan sekadar alat untuk menyampaikan informasi, melainkan juga sebuah instrumen penting dalam pembentukan dan pengembangan budaya intelektual suatu bangsa [14]. Bahasa yang digunakan oleh seorang jurnalis tidak hanya mencerminkan kualitas pemberitaannya maupun kewibawaan suatu media, tetapi juga akan menjadi panutan bagi masyarakat. Implementasi dari kecerdasan buatan, khususnya NLP dapat menjawab dari masalah tersebut.

Salah satu kesalahan bahasa yang sering dibuat oleh jurnalistik adalah kesalahan dalam peluluhan kata. Peluluhan kata adalah kata yang merupakan bagian dari morfofonemik, yaitu perubahan fonem yang terjadi sebagai akibat pertemuan antara morfem (kata atau suku kata) yang satu dan morfem lain [15]. Perubahan fonem awal pada kata dasar terjadi saat prefiks "me-" ditambahkan. Prefiks "me-" dalam pembentukan kata sering kali menghasilkan variasi morfem awal, termasuk alomorf seperti "mem-", "men-", "meng-", dan "meny-" yang

merupakan hasil dari kemunculan prefiks nasal "meN-" [16]. Contoh umumnya adalah penggunaan kata "mempunyai" sering kali ditulis sebagai memunyai, atau kata "menyakiti" sering kali ditulis "mensakiti".

Salah satu pengembangan aplikasi yang dapat membantu memperbaiki kesalahan kata adalah U-Tapis [17, 18]. Aplikasi U-Tapis melibatkan analisis statistik, bahasa komputasi, dan kecerdasan buatan dalam memperbaiki kesalahan penggunaan bahasa Indonesia [19]. Namun, U-Tapis masih belum memiliki fitur untuk memperbaiki kata luluh. Oleh karena itu, fitur peluluhan kata dimutakhirkan dengan menggunakan algoritma *Damerau-Levenshtein Distance* dan BERT. *Damerau-Levenshtein Distance*, turunan dari *Levenshtein Distance*, banyak digunakan oleh peneliti untuk mengkalkulasi kesamaan kata yang dapat berfungsi sebagai pendeteksi saltik dari *editor* [20] [21]. Model BERT juga memiliki manfaat untuk memprediksi kata selanjutnya menggunakan konteks kalimat yang sedang digunakan [22].

Berdasarkan latar belakang yang telah diuraikan sebelumnya, dirancanglah sebuah model *machine learning* yang mampu mendeteksi kesalahan peluluhan kata. Model ini diharapkan mampu untuk mendeteksi kesalahan dengan tingkat akurasi yang tinggi, disertai dengan rekomendasi kata yang benar sehingga dapat membantu jurnalis dalam proses penyuntingan berita di portal berita.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang telah dijabarkan sebelumnya, terdapat rumusan masalah yang dapat dibentuk sebagai berikut.

- (1) Bagaimana cara penerapan algoritma *Damerau-Levenshtein Distance* dan BERT untuk mendeteksi kesalahan peluluhan kata?
- (2) Bagaimana tingkat akurasi yang dihasilkan untuk penerapan algoritma *Damerau-Levenshtein Distance* dan BERT untuk mendeteksi kesalahan peluluhan kata?

1.3 Batasan Permasalahan

Berikut adalah batasan-batasan masalah pada penelitian ini.

- (1) Berita yang digunakan bersumber dari Tribunnews sebanyak 1000 berita.

- (2) Penelitian ini berfokus pada kesalahan penggunaan peluluhan kata yang memiliki kata dasar 's', 'p', 't', 'k', tanpa kluster dan satu suku kata.
- (3) Implementasi menggunakan bahasa pemrograman Python untuk mendeteksi kesalahan peluluhan kata yang menggunakan algoritma *Damerau-Levenshtein Distance* dan BERT.
- (4) Deteksi kesalahan kata luluh bersifat *uncased*, yaitu tidak dapat membedakan nama entitas.

1.4 Tujuan Penelitian

Berikut adalah tujuan dari penelitian ini.

- (1) Mengimplementasikan algoritma *Damerau-Levenshtein Distance* dan BERT untuk mendeteksi kesalahan peluluhan kata dan memberikan rekomendasi yang sesuai.
- (2) Mengukur tingkat akurasi yang dihasilkan menggunakan *Confussion Matrix* untuk implementasi algoritma *Damerau-Levenshtein Distance* dan BERT dalam mendeteksi kesalahan peluluhan kata dan memberikan rekomendasi yang sesuai.

1.5 Manfaat Penelitian

Berikut adalah manfaat dari penelitian ini.

- (1) Membantu untuk mendeteksi kesalahan peluluhan kata dan memberikan rekomendasi kata yang sesuai.
- (2) Menjadi sarana yang bermanfaat dalam mengimplementasikan pengetahuan mengenai perkembangan teknologi melalui pengimplementasian algoritma *Damerau-Levenshtein Distance* dan BERT untuk mendeteksi kesalahan peluluhan kata.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut.

(1) Bab 1 PENDAHULUAN

Bab ini berisi penjelasan tentang latar belakang, rumusan masalah, batasan masalah, tujuan, dan manfaat dari penelitian ini.

(2) Bab 2 LANDASAN TEORI

Bab ini berisi teori-teori yang digunakan dalam penelitian ini. Teori-teori tersebut didapatkan melalui penelitian-penelitian terdahulu yang berkaitan dengan penelitian ini.

(3) Bab 3 METODOLOGI PENELITIAN

Bab ini menjelaskan tentang alur pengerjaan penelitian ini yang dimulai dari pengumpulan data, *text preprocessing*, implementasi, pengujian, dan evaluasi.

(4) Bab 4 HASIL DAN DISKUSI

Bab ini berisi hasil penelitian yang dilakukan. Hasil penelitian didapatkan melalui penerapan algoritma *Damerau-Levenshtein Distance* dan BERT berupa hasil pendeteksian kesalahan peluluhan kata dan pemberian solusi kata yang benar untuk setiap kesalahan.

(5) Bab 5 SIMPULAN DAN SARAN

Bab ini berisi simpulan mengenai hasil dari penelitian yang sudah dilakukan dan saran yang dapat digunakan lagi untuk penelitian selanjutnya.

