

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Dengan kemajuan pesat dalam bidang kecerdasan buatan dan pembelajaran mesin, teknologi semakin memainkan peran krusial dalam berbagai aspek kehidupan, termasuk penyebaran informasi yang cepat di era digital saat ini. Namun, tantangan masih ada dalam menjaga kualitas penggunaan Bahasa Indonesia yang baik dan benar, terutama dalam berita daring. Meskipun Bahasa Indonesia telah ditetapkan sebagai bahasa resmi negara dan harus dipatuhi sesuai dengan EYD, beberapa media cetak masih menghadapi kendala dalam mematuhi norma bahasa [1][2][3]. Dalam konteks ini, jurnalisme perlu memprioritaskan kepatuhan terhadap norma kebahasaan, bukan hanya efektivitas komunikasi semata, khususnya dalam penulisan berita daring yang semakin populer [4]. Diperlukan langkah-langkah konkret untuk memastikan bahwa teknologi yang terus berkembang ini tidak hanya menjadi alat untuk menyebarkan informasi, tetapi juga membantu menjaga kualitas bahasa dalam konteks jurnalisme digital.

Dengan bantuan teknologi, wartawan Indonesia dapat mengatasi masalah penggunaan bahasa yang tidak sesuai standar di dunia jurnalisme. Ketidaksesuaian standar Bahasa Indonesia memiliki beberapa faktor penyebab, diantaranya, banyaknya kata serapan dari Bahasa Inggris, bahasa gaul yang semakin populer, dan kurang penguasaan kosa kata Bahasa Indonesia. Dengan kemajuan media digital dan teknologi, proyek yang ditujukan untuk menyelesaikan masalah kesalahan penulisan kata baku pada berita daring menjadi semakin penting dan relevan. Tujuan utama proyek-proyek ini adalah untuk meningkatkan kualitas tulisan dalam media digital dengan menggunakan berbagai algoritma dan sistem cerdas yang tersedia untuk mendeteksi dan memperbaiki kesalahan bahasa. Oleh karena itu, teknologi tidak hanya berfungsi sebagai alat untuk menyebarluaskan informasi, tetapi juga dapat membantu meningkatkan kualitas bahasa dalam jurnalisme digital.

Salah satu upaya untuk mengatasi permasalahan wartawan mengenai kesalahan penulisan dalam berita adalah proyek U-Tapis. U-Tapis adalah proyek yang bertujuan untuk mengatasi kesalahan bahasa Indonesia dalam berita dan jurnalistik. U-Tapis adalah aplikasi yang dirancang untuk menghilangkan kesalahan bahasa Indonesia dalam pekerjaan jurnalistik[5][6][7][8][9][10][11]

Mengembangkan modul untuk mendeteksi dan memperbaiki kesalahan penulisan kata baku dalam aplikasi U-Tapis dapat membantu mengatasi masalah penggunaan bahasa yang tidak sesuai dalam berita, terutama yang berkaitan dengan kesalahan penulisan kata baku. Untuk memeriksa kesalahan ejaan bahasa, algoritma Damerau-Levenshtein, metode n-gram, metode penggabungan trigram, distribusi bobot otomatis, model Bayesian, dan algoritma Levenshtein Distance [5][12][13][14][15][16].

Ene mencatat bahwa algoritma Levenshtein Distance dapat diterapkan dalam pembelajaran leksikon dengan seleksi kata-kata yang memiliki kesamaan tertentu dengan kata yang benar [17]. Beberapa contoh proyek yang berhasil dikembangkan dengan algoritma Levenshtein Distance adalah pengukuran fonetik antara dialek-dialek Belanda dan penyejajaran urutan serta pencarian kesamaan dalam basis data biologis [18][19]. Penelitian yang dilakukan oleh Susilawati mendukung gagasan bahwa Levenshtein Distance merupakan algoritma yang cocok digunakan untuk mengukur *similarity* pada dua *string*, Levenshtein Distance memiliki akurasi yang cukup baik dalam mendeteksi kesalahan kata dibandingkan dengan algoritma Cosine similarity. Hasil pengujian penggunaan algoritma Levenshtein Distance untuk mengoreksi kesalahan kata pada 12 dokumen teks Microsoft Word dengan variasi jumlah kesalahan ejaan menunjukkan tingkat akurasi sebesar 93.5%. Sementara itu, algoritma cosine similarity menghasilkan akurasi sebesar 77.9% dalam pengujian yang sama [20]. Melalui pendekatan ini, penelitian bertujuan untuk mengidentifikasi kata-kata tidak baku yang sering muncul dalam teks berita daring dan memberikan alternatif kata baku yang sesuai dengan norma bahasa Indonesia. Penelitian sebelumnya oleh Okuda juga menyoroti bahwa Algoritma Levenshtein Distance memiliki potensi dalam memperbaiki kesalahan ejaan, seperti substansi, penyisipan, dan penghapusan, dengan tingkat akurasi yang lebih tinggi dibandingkan metode lain [21]. Oleh karena itu, diharapkan bahwa penerapan algoritma ini dapat mengurangi insiden kesalahan penulisan kata tidak baku dalam konteks berita daring.

## 1.2 Rumusan Masalah

1. Berapa tingkat akurasi deteksi kesalahan penulisan kata baku dalam bahasa Indonesia yang dapat dicapai oleh algoritma Levenshtein Distance?
2. Mengapa menggunakan algoritma Levenshtein Distance dalam mendeteksi kesalahan penulisan kata baku dalam bahasa Indonesia?

3. Bagaimana pengaruh penggunaan algoritma Levenshtein Distance terhadap peningkatan akurasi deteksi kesalahan penulisan kata baku dalam konteks bahasa Indonesia, serta sejauh mana metode ini dapat memperbaiki dan memberikan saran yang tepat untuk kesalahan tersebut?

### **1.3 Batasan Masalah**

Penelitian terfokus secara eksklusif pada pengembangan sebuah modul yang bertujuan untuk mendeteksi dan memberikan solusi terhadap kesalahan penulisan kata tak baku dalam teks bahasa Indonesia. Fokus penelitian tidak akan meluas ke aspek lain dari kesalahan penulisan, seperti kesalahan tata bahasa atau struktur kalimat. Selain itu, batasan penelitian ini mencakup pembatasan penggunaan modul hanya pada bahasa Indonesia, sehingga tidak akan mempertimbangkan atau mengakomodasi teks dalam bahasa lain. Modul yang dikembangkan akan memanfaatkan algoritma Levenshtein Distance untuk mendeteksi kesalahan penulisan dan menyajikan usulan kata baku yang sesuai. Namun, penelitian ini tidak akan membahas implementasi atau aplikasi langsung dari kata-kata baku tersebut ke dalam teks. Dengan demikian, batasan masalah ini mengarahkan penelitian pada pengembangan modul yang spesifik dan terfokus, dengan tujuan meningkatkan kualitas penulisan dalam bahasa Indonesia melalui deteksi dan perbaikan kesalahan penulisan kata tak baku.

### **1.4 Tujuan Penelitian**

1. Menghitung Akurasi Algoritma Levenshtein Distance dalam Model Deteksi dan Perbaikan Kesalahan Penulisan Kata Baku dalam Berita Daring. Penelitian bertujuan untuk mengevaluasi seberapa baik algoritma Levenshtein Distance dapat mengenali kesalahan penulisan kata baku dalam konteks berita daring.
2. Menghitung Nilai F1 Score dari Hasil Uji Coba Algoritma Levenshtein Distance dalam Model Deteksi dan Perbaikan Kesalahan Penulisan Kata Baku dalam Berita Daring. Penelitian bertujuan untuk mengidentifikasi dan menganalisis seberapa baik algoritma Levenshtein Distance dalam mengukur performa model deteksi dan perbaikan kesalahan penulisan kata baku dalam berita daring. F1 Score digunakan sebagai metrik evaluasi utama untuk menilai keseimbangan antara

*precision* dan *recall* dalam model, memberikan gambaran komprehensif tentang keakuratan dan kelengkapan proses deteksi serta perbaikan kesalahan penulisan kata baku.

## **1.5 Manfaat Penelitian**

1. Meningkatkan Kualitas Penulisan Berita Daring.  
Dengan mengukur akurasi algoritma Levenshtein Distance dalam mendeteksi dan memperbaiki kesalahan penulisan kata baku, penelitian ini akan memberikan kontribusi dalam meningkatkan kualitas berita daring. Dengan lebih akuratnya deteksi dan perbaikan kesalahan penulisan, berita daring akan lebih konsisten dan sesuai dengan standar bahasa Indonesia yang baku.
2. Mengetahui Efek dari Model Terhadap Efisiensi Waktu dan Tenaga Jurnalis.  
Dengan menghitung F1 score dari hasil uji coba algoritma Levenshtein Distance, penelitian ini akan memberikan wawasan tentang efisiensi waktu dan tenaga dalam proses deteksi dan perbaikan kesalahan penulisan. Informasi ini dapat membantu meningkatkan efisiensi dan produktivitas dalam pekerjaan jurnalistik, sehingga waktu dan tenaga dapat dialokasikan dengan lebih baik.

## **1.6 Sistematika Penulisan**

Dengan tujuan memastikan bahwa pembahasan karya ilmiah ini terfokus pada pokok masalah dan tidak melebar ke masalah lain, penulis mengikuti beberapa aturan penulisan ilmiah sebagai berikut.

### **1.6.1 BAB 1 PENDAHULUAN**

Menjelaskan Latar Belakang Masalah, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian dan Sistematika Penulisan.

### **1.6.2 BAB 2 LANDASAN TEORI**

Menjelaskan definisi dan teori terkait Bahasa Indonesia, Kata Baku, *Natural Language Processing*, *Text Preprocessing*, Algoritma Levenshtein Distance, dan Pustaka FuzzyWuzzy.

### **1.6.3 BAB 3 METODOLOGI PENELITIAN**

Menjelaskan metodologi penelitian yang dilakukan pada penelitian yang mencakup penjabaran Diagram Metode Penelitian, studi literatur, pengumpulan data, Preprocessing teks berita daring, penjelasan algoritma, implementasi algoritma, pengembangan situs web, dan uji coba serta evaluasi.

### **1.6.4 BAB 4 HASIL DAN DISKUSI**

Menjelaskan hasil penelitian yang mencakup spesifikasi sistem, *dataset*, proses preprocessing, permodelan, fitur tambahan, penjabaran aplikasi web, dan evaluasi program.

### **1.6.5 BAB 5 SIMPULAN DAN SARAN**

Menjelaskan simpulan dari keseluruhan penelitian serta saran-saran guna menunjang penelitian selanjutnya.

