

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Kesalahan berbahasa merupakan persoalan umum yang sering ditemui baik dalam komunikasi informal maupun dalam penulisan formal seperti karya ilmiah, dokumen hukum, dan produk jurnalistik [1, 2]. Salah satu jenis kesalahan yang cukup krusial adalah kesalahan dalam penggunaan tanda baca, karena dapat mengaburkan makna dan menyebabkan kesalahpahaman terhadap informasi yang disampaikan. Dalam konteks jurnalistik, ketepatan berbahasa sangat penting karena media massa memiliki tanggung jawab untuk menyampaikan informasi secara jelas dan akurat kepada masyarakat. Oleh karena itu, pengecekan terhadap tata bahasa, termasuk tanda baca, menjadi aspek yang tidak dapat diabaikan, meskipun proses ini seringkali terkendala oleh tekanan tenggat waktu.

Pada penelitian sebelumnya dilakukan dengan menganalisis karangan siswa sekolah dasar secara manual, di mana masih banyak terjadi 4 kekeliruan dalam penggunaan titik, koma, dan tanda tanya, bahkan pada tingkat pendidikan dasar [3]. Hasil ini menunjukkan bahwa kesalahan kebahasaan dalam teks jurnalistik masih cukup tinggi. Pada penelitian berikutnya, dari 18 berita daring yang terbit di Karanganyarnews pada periode 24 Oktober hingga 15 November 2024, dilakukan pengecekan tanda baca secara manual dan ditemukan 15 kesalahan ejaan dan 6 kesalahan tanda baca [4].

Pengecekan tanda baca yang sesuai dengan kaidah Ejaan yang Disempurnakan (EYD) V [5] membutuhkan ketelitian dan waktu yang tidak sedikit. Di sinilah peran teknologi berbasis *Artificial Intelligence* (AI) dan *Machine Learning* (ML) menjadi signifikan, karena mampu menghadirkan solusi otomatis dan efisien dalam mendeteksi serta memperbaiki kesalahan berbahasa, termasuk kesalahan tanda baca.

Sebagai bagian dari pengembangan teknologi pengecekan tata bahasa Indonesia, Universitas Multimedia Nusantara telah mengembangkan aplikasi U-Tapis, sebuah sistem berbasis Natural Language Processing (NLP) yang bertujuan untuk membantu mendeteksi dan memperbaiki kesalahan tata bahasa secara otomatis sesuai dengan EYD V [6, 7]. U-Tapis telah memiliki berbagai modul yang berfokus pada beragam aspek bahasa, seperti U-Tapis Kata Baku [8], U-

Tapis Kata Depan [9], U-Tapis Kata Majemuk [10], U-Tapis Kata Terikat [11], U-Tapis Saltik [12], U-Tapis Konjungsi, U-Tapis Peluluhan Kata [13], U-Tapis Kata Keterangan dan U-Tapis Kalimat [14].

Penelitian ini merupakan lanjutan dari pengembangan tersebut, dengan fokus pada mengoreksi kesalahan penggunaan tanda baca yang telah terdeteksi pada penelitian sebelumnya. Pada penelitian sebelumnya, pendeteksian kesalahan penggunaan tanda baca menghasilkan akurasi sebesar 69% menggunakan model *Recurrent Neural Network*. Melanjutkan dari deteksi, model yang digunakan dalam penelitian ini adalah *Long Short-Term Memory* (LSTM), sebuah varian dari *Recurrent Neural Network* (RNN) yang memiliki keunggulan dalam menangani data berurutan dan dapat mengingat memori dalam jangka panjang dan jangka pendek [15]. LSTM dipilih karena kemampuannya dalam memahami konteks kalimat secara lebih dalam, yang sangat relevan dalam menentukan keakuratan penggunaan tanda baca.

Penggunaan model *Recurrent Neural Network* (RNN) dan *Long Short-Term Memory* (LSTM) pada penelitian sebelumnya menghasilkan akurasi pengujian sebesar 65% dan peningkatan akurasi penerapan menjadi 79,46% [16]. Meskipun menunjukkan performa yang cukup baik, model RNN memiliki keterbatasan dalam menangani dependensi jangka panjang antar kata dalam kalimat, terutama ketika struktur kalimat menjadi lebih kompleks. Oleh karena itu, dalam penelitian ini digunakan model LSTM, yang merupakan pengembangan dari RNN dengan kemampuan tambahan berupa memori jangka panjang. LSTM terbukti lebih stabil dan efektif dalam memahami konteks sekuensial, sehingga diharapkan mampu menghasilkan koreksi tanda baca yang lebih akurat dibandingkan model RNN konvensional.

Melalui penelitian ini, diharapkan U-Tapis dapat dilengkapi dengan modul tambahan yang secara khusus menangani kesalahan tanda baca. Hasilnya tidak hanya dapat mempercepat proses penyuntingan teks jurnalistik, tetapi juga berkontribusi dalam upaya pelestarian bahasa Indonesia yang baik dan benar, serta mendorong pemanfaatan AI dalam pengolahan bahasa alami.

1.2 Rumusan Masalah

1. Bagaimana cara U-Tapis dapat mengoreksi kesalahan penggunaan tanda baca menggunakan *Long Short-Term Memory*?
2. Bagaimana bagaimana mengimplementasikan LSTM untuk mengoreksi

kesalahan penggunaan tanda baca pada aplikasi U-Tapis?

3. Bagaimana performa dari pengkoreksi kesalahan pada penggunaan tanda baca menggunakan *Long Short-Term Memory*?

1.3 Batasan Permasalahan

1. Penelitian hanya berfokus pada mengkoreksi penggunaan tanda baca titik, koma, dan seru.
2. Pengkoreksian tanda baca hanya dikhususkan untuk penggunaan tanda baca yang relevan dengan kaidah jurnalistik.
3. Relevansi penggunaan huruf kapital setelah kata yang tidak terdapat tanda baca adalah kata yang tidak termasuk pada Kamus Besar Bahasa Indonesia.
4. Hasil output berupa deskripsi kesalahan dari penggunaan tanda baca dan hasil koreksi dari teks.

1.4 Tujuan Penelitian

1. Mengembangkan modul U-Tapis yang mampu melakukan koreksi otomatis terhadap kesalahan penggunaan tanda baca dalam teks Bahasa Indonesia.
2. Menerapkan arsitektur *Long Short-Term Memory* (LSTM) untuk mengimplementasikan proses koreksi tanda baca secara efektif dalam modul U-Tapis.
3. Mengevaluasi performa model LSTM dalam mengoreksi kesalahan tanda baca, yang diukur menggunakan metrik akurasi, *F1-Score*, serta tingkat *False Positive* untuk menilai ketepatan dan keandalan sistem.

1.5 Manfaat Penelitian

1. Meningkatkan penerapan *Natural Language Processing* (NLP) dalam proses penyuntingan teks secara otomatis.
2. Memberikan solusi melalui aplikasi U-Tapis yang mampu mendeteksi kesalahan tanda baca menggunakan pendekatan *Recurrent Neural Network* (RNN).

3. Memberikan solusi melalui aplikasi U-Tapis yang mampu mendeteksi kesalahan tanda baca menggunakan pendekatan *Long Short-Term Memory*

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Bab ini membahas latar belakang permasalahan, perumusan masalah, batasan penelitian, tujuan yang ingin dicapai, manfaat dari penelitian, serta sistematika penulisan laporan secara keseluruhan.
- Bab 2 LANDASAN TEORI
Bab ini memuat kajian teori dan konsep-konsep yang relevan sebagai dasar dari penelitian. Topik yang dibahas mencakup teori tanda baca, *Deep Learning*, *Natural Language Processing (NLP)*, arsitektur *Recurrent Neural Network (RNN)*, *Long Short-Term Memory (LSTM)*, *dot product*, *Random Search*, *Rectified Linear Unit (ReLU)*, *Sigmoid*, serta *Classification Report* yang digunakan untuk mengukur performa model.
- Bab 3 METODOLOGI PENELITIAN
Bab ini menjelaskan tahapan metodologi yang digunakan dalam penelitian, mulai dari studi literatur, pengumpulan dan praproses data, penerapan aturan linguistik untuk validasi tanda baca, perancangan arsitektur model RNN untuk deteksi kesalahan tanda baca, pengembangan model LSTM untuk koreksi otomatis, hingga evaluasi performa dari kedua model yang diimplementasikan.
- Bab 4 HASIL DAN DISKUSI
Bab ini berisi uraian mengenai implementasi metode yang dijelaskan pada Bab 3, serta analisis hasil dari model RNN dalam mendeteksi kesalahan dan LSTM dalam memberikan saran koreksi tanda baca pada aplikasi U-Tapis. Hasil evaluasi dan diskusi performa model turut disajikan dalam bab ini.
- Bab 5 SIMPULAN DAN SARAN
Bab ini menyajikan kesimpulan dari hasil penelitian yang telah dilakukan, serta memberikan saran untuk pengembangan lebih lanjut di masa depan, khususnya terkait peningkatan akurasi dan efisiensi dalam sistem koreksi tanda baca berbasis deep learning.