

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah membawa perubahan signifikan dalam dunia jurnalistik, khususnya dalam proses produksi dan publikasi berita[1]. Saat ini, banyak media yang telah beralih ke sistem digital berbasis web untuk mendukung efisiensi kerja, kolaborasi tim, serta kecepatan distribusi informasi[2]. Namun, di balik kemudahan tersebut, kualitas bahasa dalam penulisan berita masih menjadi tantangan yang cukup serius.

Salah satu permasalahan yang sering ditemukan dalam penulisan bahasa Indonesia adalah kesalahan penggunaan klitik, seperti “-ku”, “-mu”, dan “-nya”[3]. Kesalahan ini umumnya terjadi karena kurangnya pemahaman terhadap kaidah penulisan yang benar, terutama dalam konteks penulisan formal seperti berita[4]. Sebagai contoh, penulisan “rumah saya” yang seharusnya menjadi “rumah saya” (benar dalam konteks tertentu) seringkali tidak konsisten dengan aturan klitik dalam bahasa Indonesia, atau penulisan “buku ku” yang seharusnya ditulis “bukuku”.

Dalam praktik jurnalistik, kesalahan kecil seperti ini dapat mempengaruhi kredibilitas media dan kualitas informasi yang disampaikan kepada publik[5]. Oleh karena itu, diperlukan suatu sistem yang mampu membantu proses penapisan bahasa secara otomatis, khususnya dalam mendeteksi kesalahan penggunaan klitik.

Seiring dengan perkembangan teknologi kecerdasan buatan dan pemrosesan bahasa alami (Natural Language Processing), berbagai metode telah dikembangkan untuk mendeteksi kesalahan bahasa secara otomatis[6]. Salah satu pendekatan yang dapat digunakan adalah kombinasi antara metode berbasis aturan (rule-based) dan model machine learning untuk mendeteksi pola kesalahan dalam teks[7].

Pemilihan algoritma Random Forest pada penelitian ini didasarkan pada hasil beberapa penelitian terdahulu yang menunjukkan bahwa algoritma tersebut memiliki kemampuan klasifikasi yang baik pada data teks setelah dilakukan

ekstraksi fitur[8]. Selain mampu menangani data berdimensi tinggi, Random Forest juga memiliki ketahanan terhadap overfitting karena memanfaatkan kombinasi banyak pohon keputusan (decision tree), sehingga menghasilkan prediksi yang lebih stabil. Berdasarkan penelitian terdahulu yang dijadikan referensi dalam penelitian ini, Random Forest berhasil memberikan performa klasifikasi yang baik pada berbagai permasalahan analisis teks berbahasa Indonesia, sehingga algoritma ini dinilai sesuai untuk diterapkan dalam proses klasifikasi kesalahan penulisan klitik[9].

Berdasarkan permasalahan yang telah diidentifikasi, penelitian ini bertujuan untuk merancang dan membangun sistem berbasis web yang dapat mendeteksi kesalahan penulisan klitik dalam teks berita. Sistem ini tidak hanya menyediakan fitur deteksi kesalahan, tetapi juga mengimplementasikan alur kerja jurnalistik yang melibatkan peran reporter dan editor, sehingga menyerupai proses editorial yang nyata. Dengan demikian, diharapkan sistem ini dapat membantu meningkatkan kualitas penulisan berita sekaligus menjadi alat bantu dalam proses pembelajaran bahasa Indonesia.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut, rumusan masalah pada penelitian ini dapat dijabarkan sebagai berikut:

1. Bagaimana menerapkan metode Random Forest untuk mendeteksi kesalahan penulisan klitik pada teks berita bahasa Indonesia?
2. Bagaimana merancang sistem yang dapat melakukan koreksi kesalahan penulisan klitik pada teks berita secara otomatis?
3. Bagaimana mengimplementasikan sistem deteksi dan koreksi kesalahan klitik ke dalam aplikasi berbasis web untuk membantu proses penulisan dan penyuntingan berita?

1.3 Batasan Masalah

Untuk menjaga agar penelitian tetap terfokus dan sesuai dengan tujuan yang telah ditentukan, maka ditetapkan beberapa batasan masalah sebagai berikut:

1. Data yang digunakan berupa teks berita berbahasa Indonesia.

2. Sistem dibangun berbasis web menggunakan framework sederhana.
3. Algoritma yang digunakan merupakan kombinasi pendekatan rule-based dan model machine learning sederhana.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Penelitian ini bertujuan untuk mencapai beberapa hal sebagai berikut:

1. Mengimplementasikan metode Random Forest untuk mendeteksi kesalahan penulisan klitik pada teks berita bahasa Indonesia.
2. Merancang dan membangun sistem yang mampu melakukan koreksi kesalahan penulisan klitik secara otomatis pada teks berita.
3. Mengembangkan aplikasi berbasis web untuk mendukung proses deteksi dan koreksi kesalahan klitik pada penulisan dan penyuntingan berita.

1.4.2 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat sebagai berikut:

1. Manfaat Teoritis
 - a. Memberikan kontribusi dalam pengembangan metode deteksi kesalahan penulisan klitik pada teks berita bahasa Indonesia dengan memanfaatkan algoritma Random Forest.
 - b. Menambah referensi penelitian pada bidang Natural Language Processing (NLP), khususnya terkait penerapan machine learning dalam analisis kesalahan bahasa Indonesia.
 - c. Menjadi bahan referensi bagi penelitian selanjutnya yang berkaitan dengan pengembangan sistem koreksi atau penapisan teks bahasa Indonesia berbasis machine learning.
2. Manfaat Praktis
 - a. Membantu penulis dan editor berita dalam mendeteksi kesalahan penulisan klitik secara otomatis pada teks berita.
 - b. Mendukung proses penyuntingan berita agar menghasilkan tulisan yang lebih sesuai dengan kaidah bahasa Indonesia yang benar.

- c. Memberikan kemudahan melalui aplikasi berbasis web dalam proses deteksi dan koreksi kesalahan klitik pada teks berita.
- d. Menjadi media pembelajaran bagi pengguna dalam memahami penggunaan klitik yang tepat dalam bahasa Indonesia.

1.5 Sistematika Penulisan

Sistematika penulisan pada penelitian disusun menjadi beberapa bagian agar pembahasan lebih terstruktur dan cepat dipahami. Untuk itu sistematika penulisan penelitian ini adalah sebagai berikut:

1) BAB I PENDAHULUAN

Bab ini membahas latar belakang penelitian yang menjadi dasar pengembangan sistem deteksi kesalahan klitik pada teks berita bahasa Indonesia. Selain itu, dijelaskan pula rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan skripsi secara umum.

2) BAB II TINJAUAN PUSTAKA

Bab ini memuat teori-teori dan konsep yang berkaitan dengan penelitian, seperti klitik dalam bahasa Indonesia, (NLP, machine learning, sistem berbasis web, serta pendekatan rule-based. Selain itu, dibahas juga metode CRISP-DM, algoritma Random Forest, TF-IDF, Flask, dan penelitian terdahulu yang relevan sebagai landasan dalam pengembangan sistem.

3) BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode penelitian yang digunakan, yaitu CRISP-DM. Pada bab ini dibahas tahapan penelitian mulai dari identifikasi masalah, pengumpulan dataset, preprocessing data, ekstraksi fitur, pelatihan model, hingga implementasi sistem berbasis web. Selain itu, dijelaskan pula teknik pengumpulan data, variabel penelitian, serta teknik analisis data yang digunakan dalam penelitian.

4) BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi sistem deteksi kesalahan klitik berbasis web yang telah dikembangkan. Pembahasan meliputi proses preprocessing dataset,

pelatihan model machine learning, implementasi algoritma deteksi, integrasi sistem ke dalam website menggunakan Flask, serta hasil pengujian sistem. Selain itu, dilakukan pembahasan mengenai kemampuan sistem dalam mendeteksi kesalahan penggunaan klitik pada teks berita bahasa Indonesia.

5) BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan berdasarkan hasil penelitian yang telah dilakukan serta saran untuk pengembangan penelitian selanjutnya. Kesimpulan menjelaskan hasil pengembangan sistem dan performa algoritma yang digunakan, sedangkan saran ditujukan untuk pengembangan fitur, peningkatan akurasi sistem, serta pengembangan dataset yang lebih luas di masa mendatang.

