

## BAB 1

### PENDAHULUAN

#### 1.1 Latar Belakang Masalah

Bahasa Indonesia, yang merupakan bahasa resmi di Indonesia, digunakan oleh sekitar 200 juta penutur di seluruh dunia [5]. Bahasa ini dikenal memiliki aturan ejaan yang cukup kompleks, sehingga penerapan pemeriksa ejaan menjadi penting dalam meningkatkan kualitas tulisan berbahasa Indonesia [6]. Perkembangan teknologi membuat pemeriksa ejaan otomatis semakin banyak dimanfaatkan, terutama pada aplikasi pengolah kata dan bidang *Natural Language Processing* (NLP) [7].

Kemajuan teknologi merupakan fenomena yang wajar dijumpai dalam berbagai aspek kehidupan manusia, mulai dari komunikasi, pendidikan, bisnis, hingga media visualisasi [8]. Pemanfaatan kecerdasan buatan khususnya memberikan dampak signifikan di berbagai bidang, termasuk dunia pendidikan maupun jurnalisme [9] [10]. Pada praktik jurnalistik, penulisan berita harus mengikuti kaidah dan prosedur tertentu agar memenuhi standar penulisan yang berlaku [11].

Pada konteks penelitian jurnalisme, salah satu aplikasi pengolahan kata dan bidang NLP untuk kalimat berita supaya mengikuti prosedur penulisan adalah U-Tapis. U-Tapis menggarisbawahi pentingnya pemanfaatan AI dan NLP, yang terbukti membawa dampak positif terhadap perkembangan dunia jurnalistik [12]. Sejumlah penelitian terkait U-Tapis telah dilakukan [13]-[19], namun deteksi kesalahan penulisan klitik belum pernah dibahas. Penelitian ini berfokus pada deteksi kesalahan penggunaan tulisan klitik sebagai upaya lanjutan dalam pengembangan sistem pemeriksa ejaan yang lebih komprehensif.

Bahasa jurnalistik memiliki peran penting dalam menyampaikan informasi secara tepat dan akurat kepada masyarakat melalui media massa [20]. Kesalahan penulisan atau ketikan masih sulit dihindari karena faktor seperti kurangnya konsentrasi maupun kebiasaan mengetik dengan tergesa-gesa. Penggunaan ejaan yang benar diperlukan untuk menjaga kejelasan makna, mencegah kesalahpahaman, serta meningkatkan kualitas tulisan. Hal ini menjadi semakin relevan dalam penulisan karya ilmiah, laporan, makalah, hingga berita [21].

Berbagai penelitian telah mengkaji metode pemeriksa ejaan dalam bahasa

Indonesia. Studi Yanfi [21], misalnya, meninjau beragam pendekatan dan mencatat bahwa metode campuran (*hybrid method*), seperti kombinasi *rule-based* dengan algoritma pembelajaran mesin, sering digunakan untuk meningkatkan akurasi deteksi kesalahan ejaan. Pendekatan ini dinilai mampu memanfaatkan keunggulan aturan linguistik eksplisit sekaligus kemampuan generalisasi model berbasis data.

Sejalan dengan temuan tersebut, penelitian Zhu [22] menunjukkan bahwa algoritma *Decision Tree* memiliki tingkat akurasi yang tinggi dalam mendeteksi kesalahan penulisan. Pendekatan ini kemudian dikembangkan menjadi *Random Forest*, yang mampu mengurangi korelasi antar pohon keputusan serta meningkatkan akurasi prediksi melalui mekanisme pemilihan fitur dan data secara acak [23]. *Random Forest* dikenal memiliki ketahanan terhadap *noise* dan *outlier*, serta efisien untuk diterapkan pada permasalahan klasifikasi teks [24].

Berdasarkan pertimbangan tersebut, penelitian ini mengusulkan pendekatan hibrida dengan menggabungkan model *Random Forest* dan metode *rule-based* untuk mendeteksi kesalahan penulisan klitik. Pendekatan serupa telah menunjukkan hasil yang menjanjikan pada berbagai penelitian sebelumnya mencapai akurasi 97,16% [25], serta deteksi kesalahan Sal-Tik yang dilaporkan mencapai akurasi 100% [17]. Hasil-hasil tersebut mengindikasikan bahwa Random Forest efektif dalam menangani pola kesalahan teks yang kompleks.

Kombinasi antara pendekatan berbasis aturan, *Random Forest*, dan pemrosesan fitur teks memungkinkan proses analisis berita dilakukan secara otomatis dengan tingkat ketepatan yang tinggi, tanpa mengorbankan kecepatan pemrosesan yang dibutuhkan di era digital [26] [27]. Penelitian ini diharapkan dapat mendukung jurnalis Indonesia dalam proses pengetikan dan penulisan berita yang sesuai kaidah bahasa, sekaligus menjadi rujukan bagi penelitian selanjutnya dalam bidang jurnalisme digital dan *Natural Language Processing* (NLP).

## 1.2 Rumusan Masalah

Rumusan masalah yang dirumuskan berdasarkan latar belakang di atas adalah sebagai berikut:

1. Bagaimana mendeteksi penyalahgunaan tulisan klitik menggunakan *Random Forest* dan *Rule-based*?
2. Bagaimana hasil evaluasi kinerja sistem pemeriksa ejaan pada model *Random Forest* dalam mendeteksi kesalahan penggunaan klitik?

### **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan model pemeriksa ejaan menggunakan algoritma *Random Forest* dan *Rule-Based* untuk mendeteksi kesalahan penulisan klitik sesuai kaidah EYD.
2. Mengevaluasi kinerja sistem pemeriksa ejaan pada model *Random Forest* dalam mendeteksi kesalahan penggunaan klitik berdasarkan akurasi, presisi, *recall*, dan *F1-score* pada dataset Bahasa Indonesia.

### **1.4 Urgensi Penelitian**

Bahasa Indonesia memiliki aturan ejaan yang kompleks, sehingga pemeriksa ejaan sangat dibutuhkan untuk menjaga ketepatan dan kejelasan makna tulisan. Kesalahan penulisan klitik masih sering ditemukan dalam naskah berita maupun karya ilmiah, sementara penelitian mengenai deteksi kesalahan klitik dengan pendekatan algoritma masih terbatas. Kemajuan teknologi *Natural Language Processing* (NLP) dan kecerdasan buatan memungkinkan proses pemeriksaan ejaan dilakukan secara otomatis dan cepat, yang sangat relevan dengan kebutuhan dunia jurnalistik dan publikasi digital. Kombinasi pendekatan *Rule-Based* dan algoritma *Random Forest* pada penelitian ini digunakan untuk mengembangkan sistem pemeriksa ejaan klitik untuk membantu kinerja jurnalis serta penulis lainnya.

### **1.5 Luaran Penelitian**

Luaran yang dihasilkan dari penelitian ini meliputi:

1. Sebuah model *Random Forest* yang berfungsi untuk mendeteksi kesalahan penggunaan klitik sesuai kaidah EYD, sehingga dapat diintegrasikan dengan sistem pemeriksa ejaan maupun aplikasi jurnalistik lainnya.
2. Artikel ilmiah yang memuat hasil penelitian ini dan direncanakan untuk dipublikasikan pada jurnal nasional terakreditasi SINTA.

### **1.6 Manfaat Penelitian**

Penelitian ini diharapkan memberikan beberapa manfaat yang relevan dengan pengembangan sistem pemeriksa ejaan Bahasa Indonesia. Pertama, penelitian ini

dapat meningkatkan akurasi sistem pemeriksa ejaan, khususnya dalam mendeteksi kesalahan penulisan klitik yang sebelumnya belum terakomodasi dalam sistem U-Tapis. Kedua, hasil penelitian ini berpotensi mendukung peningkatan kualitas penulisan jurnalistik dengan membantu jurnalis menjaga ketepatan penggunaan ejaan sesuai kaidah bahasa Indonesia. Selain itu, penelitian ini turut memperluas penerapan *Natural Language Processing* (NLP) dan kecerdasan buatan dalam bidang kebahasaan, terutama dalam konteks bahasa Indonesia yang memiliki aturan ejaan kompleks. Lebih jauh, penelitian ini juga dapat menjadi dasar bagi pengembangan penelitian lanjutan yang berfokus pada sistem pemeriksa ejaan otomatis maupun analisis sintaksis berbasis pembelajaran mesin. Dengan demikian, penelitian ini berkontribusi dalam meningkatkan efisiensi proses penyuntingan teks sekaligus memperkuat pemanfaatan teknologi AI dalam dunia jurnalistik dan pendidikan.

