

BAB 5

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan penelitian yang dilakukan, penggunaan *Random Forest* yang dibantu dengan pendekatan *rule-based* terbukti mampu mendeteksi kesalahan penggunaan klitik. Kombinasi antara pembelajaran mesin dan aturan linguistik menghasilkan akurasi sebesar 0,99. Penggunaan *rule-based* membantu mengurangi kesalahan yang mungkin muncul akibat kekurangan data pelatihan, sehingga model dapat mendeteksi klitik yang salah. Hasil penelitian ini menunjukkan bahwa integrasi antara metode statistik dan pengetahuan linguistik dapat melakukan koreksi kata, terutama dalam konteks bahasa Indonesia yang memiliki banyak variasi penggunaan klitik.

5.2 Saran

Berdasarkan hasil penelitian ini, terdapat beberapa saran untuk penelitian selanjutnya:

1. Pemanfaatan Indo-T5 untuk koreksi kesalahan klitik pada teks berita. Penelitian selanjutnya disarankan menggunakan model T5 agar sistem tidak hanya mampu mendeteksi kesalahan klitik, tetapi juga memberikan koreksi yang lebih akurat dan kontekstual sesuai pola bahasa dalam teks berita.
2. Penerapan model *Conditional Random Fields* (CRF) untuk memetakan pola klitik pada level token, sehingga sistem mampu mengenali posisi kesalahan klitik secara terstruktur.

**UNIVERSITAS
MULTIMEDIA
NUSANTARA**