

Koreksi Teks untuk Kesalahan Penulisan pada Artikel Berita menggunakan Model Bahasa

M. Dony Samdhila Yasin¹, Dr. Ade Romadhony, S.T., M. T.²

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹samdhila@students.telkomuniversity.ac.id,

²aderomadhony@telkomuniversity.ac.id,

Abstrak

Dalam era digital ini, kualitas penulisan artikel berita sangat penting untuk menjaga integritas dan kredibilitas informasi. Namun, sering terjadi kesalahan penulisan seperti ejaan dan tata bahasa yang dapat mengurangi kualitas artikel berita dan mempengaruhi pemahaman pembaca. Kesalahan ini mengindikasikan adanya kebutuhan akan metode yang efektif untuk mendeteksi dan memperbaiki kesalahan tersebut.

Penelitian ini bertujuan untuk meningkatkan kualitas artikel berita dengan mengimplementasikan teknologi koreksi teks menggunakan model bahasa IndoBERT. Tiga metode yang diterapkan adalah pretrained error detection, finetuned error detection, dan dictionary-based error detection. Dataset yang digunakan mencakup 200 artikel berita berbahasa Indonesia dengan total 55.356 kata, meliputi berbagai kategori berita.

Hasil penelitian menunjukkan bahwa metode dictionary-based error detection menghasilkan Overall Accuracy sebesar 89%, yang sangat efektif dalam mendeteksi dan memperbaiki kesalahan ejaan sederhana. Di sisi lain, model IndoBERT yang di-finetune dengan dataset berita menunjukkan peningkatan akurasi hingga 46% dibandingkan model pretrained, dan lebih unggul dalam menangani kesalahan kontekstual yang lebih kompleks.

Penelitian ini memberikan dampak signifikan dalam meningkatkan kualitas penulisan artikel berita dengan menggunakan teknologi koreksi teks berbasis model bahasa. Kesalahan penulisan dapat dikurangi, sehingga artikel menjadi lebih akurat dan mudah dipahami oleh pembaca. Hasil penelitian ini juga diharapkan dapat berkontribusi pada pengembangan teknologi koreksi teks otomatis di Indonesia.

Kata kunci: koreksi teks, kesalahan penulisan, artikel berita, model Bahasa
