ABSTRAK

Penulisan yang baik dalam bahasa Indonesia sangat penting untuk memastikan bahwa pesan yang terkandung dalam teks dapat dipahami dengan baik dan benar oleh pembacanya. Namun, berbagai jenis tulisan, seperti karya ilmiah, konten media sosial, ataupun artikel berita, masih sering mengandung kesalahan tata bahasa, khususnya dalam aspek sintaksis dan morfologis. Kesalahan seperti tidak adanya subjek atau predikat, struktur kalimat yang tidak logis, serta penggunaan preposisi dan konjungsi yang tidak tepat dapat menyebabkan kesalahpahaman. Saat ini, alat bantu otomatis untuk mendeteksi kesalahan tata bahasa Indonesia masih terbatas. Penelitian ini telah berhasil mengembangkan model *deep learning* berbasis arsitektur Transformer yang hanya menggunakan lapisan *encoder* untuk mendeteksi kesalahan tata bahasa secara otomatis. *Dataset* yang digunakan berupa kumpulan kalimat berbahasa Indonesia yang berjumlah satu juta kalimat. Tahapan penelitian ini melibatkan *pre-processing* data yang mencakup normalisasi berupa *filtering* dan *cleaning*, generasi kalimat sintetis dengan tata bahasa yang salah, tokenisasi menggunakan SentencePiece dengan Unigram Language Model, serta pelabelan token. Setelah *pre-processing*, dilakukan pelatihan, validasi, dan pengujian model. Penelitian ini juga telah berhasil melakukan perbandingan model arsitektur Transformer antara 1, 2, 3, dan 6 *encoder layers*, serta Bi-LSTM yang menggunakan *self-attention*. Hasil terbaik berhasil dicapai oleh Transformer dengan 6 *encoder layers*, dengan akurasi 93,46% dan *F1-score* 75,08% pada tingkat token, serta akurasi 78,82% dan *F1-score* 78,38% pada tingkat kalimat. Hal ini menunjukkan bahwa kedalaman *encoder layer* pada Transformer berpengaruh positif terhadap efektivitas deteksi kesalahan tata bahasa dimana urutan token berperan sebagai faktor kunci terhadap akurasi kalimat.

Kata Kunci: kalimat, tata bahasa, sintaksis, morfologi, transformer, lapisan *encoder*