

ABSTRAK

Silabifikasi adalah sebuah proses dari kata menjadi serangkaian suku kata. Silabifikasi dapat digunakan dalam Natural Language Processing (NLP) seperti pengenalan suara, *text-to-speech*, deteksi rima, dan banyak lagi. Silabifikasi kata bahasa Indonesia akan mengacu pada Pedoman Umum Ejaan Bahasa Indonesia (PUEBI). Korpus untuk penelitian ini berisi kata utama dan rangkaian suku kata tersebut. Korpus penelitian ini menggunakan "*50k KBBI 5-k fold*" (Korpus kata formal) dan digabungkan dengan "*103k Named Entity 5-k fold*" (Korpus kata entitas bernama). Evaluasi model ini menggunakan *Word Error Rate* (WER). WER model *deep learning* sebelumnya untuk suku kata masih tinggi dengan 3.75% WER. Tujuan dari penelitian ini adalah untuk menurunkan WER untuk metode *deep learning* menggunakan Transformer dan Syllable Tag karena dapat menerima *contextual dependency* yang panjang. Hasil evaluasi dari model ini adalah 3.68% WER dan dapat digunakan secara universal untuk kata-kata bahasa Indonesia karena margin antara Kata Formal dan Kata Entitas Bernama mendekati hasil rata-rata. Dengan demikian, model ini saat ini merupakan model yang lebih baik untuk model *deep learning* silabifikasi bahasa Indonesia berdasarkan rata-rata WER yang lebih rendah dibandingkan dengan model *deep learning* lainnya.

Kata Kunci: silabifikasi, bahasa Indonesia, *deep learning*, transformer