
Abstrak

Pemeriksa ejaan adalah alat untuk mendeteksi dan memperbaiki berbagai kesalahan ejaan. Meskipun mungkin sepele bagi manusia, pendeteksian dan koreksi ejaan bisa sangat berguna untuk mesin, karena mesin tidak dapat mendeteksi kesalahan ejaan dan memperbaikinya secara otomatis. Dalam Natural Language Processing (NLP), mendeteksi dan memperbaiki kesalahan ejaan adalah tugas yang telah banyak dilakukan untuk menormalkan data, karena sebagian besar teks mentah berisik dan memiliki banyak kesalahan ejaan. Dalam beberapa tahun terakhir, Long Short-term Memory (LSTM) telah terbukti memberikan hasil yang luar biasa dalam memecahkan masalah berurutan, termasuk koreksi ejaan. Dalam makalah ini, kami mengusulkan model LSTM yang mengkodekan kata input pada tingkat karakter, yang juga menggunakan konteks kata dan tag POS sebagai fitur. Kami melakukan percobaan pada set data buatan berdasarkan artikel Wikipedia bahasa Indonesia yang kami buat dengan mensimulasikan beberapa kesalahan ejaan artifisial di tingkat karakter dan mengujinya pada dataset nyata, sebagian besar adalah artikel berita online Indonesia. Evaluasi pada dataset uji memberikan akurasi 83,76%.

Kata kunci: Pemeriksa Ejaan Teks Bahasa Indonesia, LSTM, deep learning