

## ABSTRACT

NER as one of the processing parts of NLP has the objective to identify entities that exist in a text or sentence. Identification of these entities is based on existing general classifications such as person, location, time, facilities, and geopolitical entities. The entity classification has been used by NER in various fields such as Information Extraction (IE), question answer system, search engine, machine translation, text summarization, and Information Retrieval (IR). The wide field of use of NER causes research to continue to grow. NER research especially in Indonesian has a problem, that is NER has a high dependence on vocabulary. These dependencies have implications for the resulting performance. This research was conducted to solve these problems by developing deep learning models based on the architecture of BiLSTM. The development is carried out through several stages, that is 1) combining two forms of embedding, namely ELMo and C2W embedding, 2) modification of BiLSTM by adding self-attention mechanism, and 3) The use of CRF linear chains as a function of connection at the output layer. the results of the performance evaluation using 5 fold cross-validation of deep learning models reached precision 82,9%, recall 82,0% and F1-score 82,5%.

**Keywords:** NER, *deep learning model*, *architecture*, ELMo, C2W, BiLSTM, *self-attention*, CRF

## INTISARI

NER sebagai salah satu bagian pemrosesan pada NLP memiliki tujuan untuk mengidentifikasi entitas yang ada pada sebuah teks atau kalimat. Identifikasi entitas tersebut dilakukan berdasarkan klasifikasi umum yang ada seperti person/orang, lokasi, waktu, fasilitas, dan entitas geo-politik. klasifikasi entitas tersebut telah digunakan NER dalam berbagai bidang seperti *Information Extraction* (IE), *question answer system*, mesin pencari, mesin translasi, summarisasi teks, dan *Information Retrieval* (IR). Luasnya bidang penggunaan NER menyebabkan penelitian terhadapnya terus berkembang. Penelitian NER khususnya dalam Bahasa Indonesia memiliki permasalahan yaitu ketergantungan NER terhadap vocabulary. Ketergantungan tersebut berimplikasi terhadap performa yang dihasilkan. Penelitian ini dilakukan untuk menyelesaikan permasalahan tersebut dengan mengembangkan *deep learning model* berdasarkan arsitektur dari BiLSTM. Pengembangan tersebut dilakukan melalui beberapa tahapan yaitu 1) mengkombinasikan dua bentuk *embedding* yaitu ELMo dan C2W *embedding*, 2) modifikasi BiLSTM dengan menambahkan mekanisme *self-attention*, dan 3) Penggunaan rantai linier CRF sebagai fungsi koneksi pada lapisan *output*. Hasil evaluasi *5 fold cross-validation* diperoleh performa *deep learning model* mencapai nilai *precision* 82,9%, *recall* 82,0% dan *F1-score* 82,5%..

**Kata kunci:** NER, *deep learning model*, arsitektur, ELMo, C2W, BiLSTM, *self-attention*, CRF