

ABSTRACT

IMPLEMENTATION OF NAMED ENTITY RECOGNITION (NER) BASED ON DEEP LEARNING FOR INFORMATION EXTRACTION IN ONLINE NEWS ABOUT INFECTIOUS DISEASES

By

Anastasia Harum Mawadah

Infectious diseases remain a major public health challenge in Indonesia, with case trends continuing to rise. In the digital era, although information spreads rapidly through online media, the main challenge lies in extracting relevant data from the vast volume of available information. This study aims to apply the Named Entity Recognition (NER) method with the addition of Part-of-Speech (POS) Tagging and to develop BiLSTM, BiLSTM-CRF, and hybrid IndoBERT-BiLSTM-CRF models to extract information from online news related to infectious diseases. The results show that the NER method successfully identified and labeled entities such as locations, organizations, and individuals, making the information extraction process more efficient and systematic. The best-performing model was the hybrid combination of IndoBERT and BiLSTM-CRF, with an 80% training and 20% testing data split, yielding optimal accuracy, precision, recall, and F1-score values.

Keywords: Named Entity Recognition (NER), Part-of-Speech (POS) Tagging, BiLSTM, CRF, IndoBERT, Information Extraction, Online News, Infectious Diseases.

ABSTRAK

IMPLEMENTASI *NAMED ENTITY RECOGNITION* (NER) BERBASIS *DEEP LEARNING* UNTUK EKSTRAKSI INFORMASI PADA BERITA ONLINE MENGENAI PENYAKIT MENULAR

Oleh

Anastasia Harum Mawadah

Penyakit menular menjadi tantangan besar bagi kesehatan masyarakat di Indonesia dengan tren kasus yang terus meningkat. Di era digital, meskipun informasi tersebar cepat melalui media online, tantangan utama adalah mengekstraksi data yang relevan dari volume informasi yang sangat besar. Penelitian ini bertujuan menerapkan metode NER dengan penambahan POS Tagging dan membangun model BiLSTM, BiLSTM-CRF, serta IndoBERT-BiLSTM-CRF untuk mengekstraksi informasi dari berita online terkait penyakit menular. Hasil penelitian menunjukkan bahwa metode NER berhasil mengenali dan memberi label pada entitas seperti lokasi, organisasi, dan tokoh, sehingga proses ekstraksi informasi menjadi lebih efisien dan sistematis. Model terbaik diperoleh dari kombinasi IndoBERT dan BiLSTM-CRF dengan akurasi sebesar 95%. Pembagian data 80% untuk pelatihan dan 20% untuk pengujian, menghasilkan nilai akurasi, presisi, recall, dan f1-score yang optimal.

Kata-kata kunci: Penyakit Menular, Named Entity Recognition (NER), Part Of Speech (POS) Tagging, BiLSTM, CRF, IndoBERT, Berita Online, Ekstraksi Informasi.