

1. Pendahuluan

Pada teks berita terdapat banyak informasi yang penting seperti nama orang, nama organisasi atau nama tempat. Untuk memperoleh informasi dalam dokumen teks secara manual, manusia harus membaca isi seluruh teks berita. Apabila dokumen teks tersebut sangat panjang, maka diperlukan waktu yang lama bagi manusia untuk memperoleh informasi yang terdapat pada teks tersebut. Salah satu untuk mempermudah kasus tersebut menggunakan *NER*.

Named Entity Recognition atau *NER* adalah bagian dari *Text Mining* dan *Natural Language Processing* proses yang sangat berguna dalam proses penggalian informasi. *NER* pertama kali di usulkan pada tahun 1995 pada *MUC-6 Conference* [1]. Ekstraksi informasi adalah proses mencari informasi dari suatu dokumen atau bahasa alami sebagai masukannya dari hasil informasi yang bermanfaat berupa informasi yang terstruktur dengan format tertentu. Tugas utama *NER* adalah untuk mengidentifikasi dan mengklasifikasikan nama dalam teks ke dalam kelas yang telah ditentukan [2]. Ekstraksi informasi secara tegas untuk bidang *NER* digunakan untuk memudahkan mengolah data kata atau kalimat. Ini membantu mesin pencari dan juga membantu mengklasifikasikan tempat, waktu, dan organisasi. *NER* sangat penting dalam *NLP* dan merupakan tahap paling mendasar dalam mengekstraksi informasi. Dalam penerapannya terdapat tiga metode, *NER* berdasarkan aturan dan template, *NER based on machine learning*, *NER based on deep learning* [3].

Berdasarkan uraian tersebut maka pada penelitian ini akan dilakukan implementasi metode *Recurrent Neural Network (RNN)* pada kasus Ekstraksi Entitas Organisasi yang Terafiliasi Telkom University. Metode *RNN* dipilih dikarenakan pada penelitian sejenis menggunakan versi lain dari *RNN* seperti *LSTM (Long Short Term Memory)*, *BiLSTM (Bidirectional Long Short Term Memory)*, *CNN (Convolutional Neural Network)* dan lain-lain. Dengan itu pada penelitian ini hanya menggunakan metode *RNN* pada teks berbahasa Indonesia untuk melihat performa dan akurasi jika hanya menggunakan metode *RNN* pada entitas organisasi teks berbahasa Indonesia serta evaluasi performansi menggunakan *NER* dengan metode *RNN* pada portal berita Bahasa Indonesia *RNN* juga disebut jaringan umpan balik adalah jenis jaringan pada neural network dimana terdapat loop sebagai koneksi umpan balik dalam jaringan. Jaringan *RNN* adalah jaringan yang mengakomodasi output jaringan untuk menjadi input pada jaringan tersebut yang kemudian digunakan untuk menghasilkan output yang baru [4]. Baru-baru ini, *RNN* memiliki menunjukkan kesuksesan besar dalam beragam tugas *Natural Language Processing* seperti pengenalan suara, mesin terjemahan, dan model Bahasa [5]. Dampak dari penelitian ini adalah karena kurangnya penelitian lain yang menggunakan dataset bahasa Indonesia hanya dengan menggunakan metode *RNN* menjadikan hasil penelitian ini sebagai gambaran untuk penelitian selanjutnya.

Batasan masalah yang digunakan pada penelitian ini yaitu dataset yang digunakan adalah kumpulan berita berbahasa Indonesia yang membahas seputar Telkom University. Dataset yang digunakan berjumlah 20.061 kata, pada portal berita berbahasa Indonesia. Proses pelabelan pada kalimat Bahasa Indonesia dilakukan secara manual, Label aspek hanya terbagi menjadi 4 (empat) kategori yaitu LOC (lokasi), ORG (organisasi), PER (orang), TIM (waktu), O (other), dan entitas yang diambil berfokus pada label ORG (organisasi). Untuk mengukur performansi dari model *RNN* menggunakan *Confusion matrix* seperti *F1-Score*, *Recall*, *precision*, *accuracy*.

Adapun tujuan dari penelitian ini adalah membangun model *NER* dengan metode *RNN* pada Entitas Organisasi yang Terafiliasi Telkom University, evaluasi performansi dan akurasi *NER* menggunakan metode *RNN* pada Entitas Organisasi yang Terafiliasi Telkom University, dan juga mengukur pengaruh metode *RNN* pada *NER* pada Entitas Organisasi yang Terafiliasi Telkom University. Output yang diharapkan dari penelitian ini adalah model yang digunakan dapat mengkategorikan entitas kata pada dataset dengan akurasi yang maksimal menggunakan *confusion matrix* seperti *F1-Score*, *Recall*, *precision*, *accuracy*.

Bagian pertama ini adalah pendahuluan. Selanjutnya terdiri dari bagian Studi Terkait, Sistem yang Dibangun, Evaluasi, dan Kesimpulan. Pada bagian Studi Terkait berisikan studi atau riset yang telah dilakukan sebelumnya. Bagian Sistem yang Dibangun menjelaskan bagaimana pembangunan sistem secara umum dengan menggunakan skema dan penjelasan mengenai metode yang akan digunakan dalam membangun sistem. Evaluasi membuat hasil pengujian dan analisis dari sistem. Kesimpulan berisikan kesimpulan dan saran untuk penelitian selanjutnya.