

## 1. Pendahuluan

Pada artikel berita terdapat informasi penting berisi berbagai informasi seperti nama pribadi, nama tempat, nama institusi dan lain-lain. Penelitian ini menggunakan sumber data berita teks berbahasa Indonesia untuk dapat memperoleh entitas organisasi yang terafiliasi Telkom University, contoh teks berita terafiliasi Telkom University adalah “Universitas Telkom membuka beasiswa bagi siswa SMA/SMK/MA yang memiliki Kartu Indonesia Pintar (KIP) dengan nama Beasiswa KIP Kuliah Merdeka. Seleksi ini dilakukan dengan menggunakan nilai rapor. Proses pendaftaran tidak memungut biaya sepeserpun alias gratis. Adapun, jadwal pendaftaran dibuka sejak 10 Mei hingga 25 Juni 2021 mendatang”. Sehingga dalam kasus ini yang dapat mempermudah untuk mendapatkan informasi dalam teks berita adalah dengan menggunakan *Named Entity Recognition*.

*Named Entity Recognition (NER)* merupakan salah satu bagian dari *Natural Language Processing (NLP)*. *NER* banyak digunakan dalam aplikasi *NLP* dan kecerdasan buatan seperti ekstraksi informasi, terjemahan mesin, pencarian informasi, penjawab pertanyaan dan lain-lain[1]. Tujuan utama *NER* adalah mengidentifikasi entitas dengan makna khusus dalam teks, terutama termasuk nama pribadi, nama tempat, nama institusi, kata benda, dan lain-lain[2]. Berdasarkan uraian tersebut tentang *NER* maka penelitian ini mengimplementasikan metode *Bidirectional LSTM (Long Short Term Memory)-Conditional Random Field (CRF)* untuk kasus ekstraksi entitas organisasi pada berita yang terafiliasi Telkom University. Metode *bidirectional LSTM-CRF* menjadi pilihan dalam penelitian ini dikarenakan *bidirectional LSTM-CRF* tersebut memiliki hasil *f1-score* cukup baik diatas 80% untuk berbagai bahasa yang telah diuji. Terbukti pada beberapa penelitian seperti penelitian oleh *Guillaume Lampe, dkk*[3], membuktikan bahwa model *bidirectional LSTM-CRF* mendapatkan *f1-score* 90,33%, penelitian serupa lainnya oleh *L. T. Anh, dkk*[4], mendapatkan hasil *precision*, *recall* dan *f1-score* dari mengkombinasikan model *bidirectional LSTM* dengan *CRF* yaitu 96,49%, 97,19% dan 96,84%. Hasil dari beberapa penelitian serupa dapat disimpulkan bahwa model *bidirectional LSTM-CRF* merupakan model yang terbaik untuk sekarang. Oleh sebab itu, metode *bidirectional LSTM-CRF* terpilih untuk menyelesaikan kasus ekstraksi entitas organisasi yang terafiliasi Telkom University.

Manfaat Teoritis dari tugas akhir ini adalah memberikan manfaat bagi setiap institusi yang ingin menggunakan tugas akhir ini untuk mencari informasi penting pada kumpulan artikel berita. Manfaat Praktis dari tugas akhir ini adalah bermanfaat bagi penulis, secara personal penulis akan mendapatkan manfaat berupa pengetahuan dan wawasan baru terkait implementasi metode *Bidirectional LSTM-CRF* pada artikel berita terkait Telkom University. Batasan masalah pada penelitian ini adalah dataset yang digunakan berupa kumpulan artikel berita berbahasa Indonesia yang membahas seputar Telkom University dengan jumlah dataset 20.061 kata, label informasi yang digunakan terbagi menjadi empat label, yaitu *Person (PER)*, *Location (LOC)*, *Organization (ORG)*, *Time (TIM)* dan fokus pada penelitian ini hanya berfokus pada label *organization (ORG)*. Pengukuran performansi dari model *bidirectional LSTM-CRF* menggunakan *confusion matrix* seperti *accuracy*, *precision*, *recall* dan *f1-score*. Adapun tujuan dari penelitian ini adalah membangun model *NER* dengan metode *bidirectional LSTM-CRF* pada ekstraksi entitas organisasi yang terafiliasi Telkom University, evaluasi performansi dan *f1-score* menggunakan metode *bidirectional LSTM-CRF* pada entitas organisasi yang terafiliasi Telkom University, dan juga mengukur pengaruh metode *bidirectional LSTM-CRF* 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 *accuracy*, *precision*, *recall* dan *f1-Score*. Berdasarkan tujuan tersebut pada penelitian ini berhasil membangun model *NER* dengan metode *bidirectional LSTM-CRF* untuk ekstraksi entitas organisasi yang terafiliasi Telkom University dengan mengolah dataset teks berita berbahasa Indonesia menghasilkan nilai *f1-score* untuk entitas organisasi diatas 80%. Pada penelitian ini telah dilakukan ketiga skenario pengujian yaitu dengan mengatur jumlah *embedding dimension*, jumlah *units* dan jumlah *batch size* pada model. Dapat disimpulkan bahwa dengan jumlah *embedding dimension* 100, jumlah *units* 50 dan jumlah *batch size* 16 merupakan hasil yang terbaik dengan hasil *precision* untuk B-ORG dan I-ORG mencapai 83% dan 89%, *recall* untuk B-ORG dan I-ORG mencapai 90% dan 74%, *f1-score* untuk B-ORG dan I-ORG mencapai 86% dan 81%. Dataset yang cukup memadai, sehingga dapat disimpulkan model *bidirectional LSTM-CRF* mampu menyelesaikan permasalahan *NER* untuk dataset teks berbahasa Indonesia dengan mendapatkan hasil yang cukup baik diatas 80%.