

BAB I PENDAHULUAN

1.1 Latar Belakang

Informasi politik semakin penting bagi kehidupan masyarakat di era komputer dan internet yang terus berkembang. Berita politik, terutama artikel berita politik di Indonesia, berfungsi sebagai dasar bagi masyarakat untuk memahami isu-isu politik yang sedang berlangsung[1]. Namun, di tengah arus informasi yang begitu besar untuk mencari dan menganalisis data tentang entitas organisasi politik secara manual menjadi sulit dan memakan waktu[2]. Solusinya, diperlukan prosedur otomatis yang efisien untuk mengekstraksi entitas organisasi dari artikel berita politik di Indonesia.

Dalam domain artikel berita politik, identifikasi entitas melibatkan pengenalan dan pemahaman terhadap berbagai entitas penting seperti nama orang atau pelaku politik, nama partai politik, serta organisasi-organisasi yang memiliki keterkaitan dalam dunia politik. Identifikasi entitas ini bukan hanya mencakup organisasi-organisasi politik yang terlibat, tetapi juga mencakup entitas-entitas yang berkaitan dalam ranah politik[3]. Tujuannya adalah untuk membuat calon pembaca mudah menemukan dan mengakses informasi tertentu melalui internet. Misalnya, mengidentifikasi partai politik atau organisasi politik tertentu juga dapat membantu pembaca memahami latar belakang, platform, dan pandangan politik yang dilaksanakan oleh organisasi atau partai politik tersebut.

Named Entity Recognition (NER) merupakan komponen krusial dalam pemrosesan bahasa alami (NLP) yang bertujuan mengidentifikasi entitas seperti nama orang, organisasi, waktu, dan lokasi dalam teks. Fungsi NER tidak hanya terbatas pada pengenalan entitas, melainkan juga berperan dalam tugas NLP lainnya seperti ekstraksi data dan pembuatan kueri otomatis. Dengan kemampuannya memahami dan mengekstraksi entitas dalam teks, NER menjadi pondasi penting dalam pengembangan aplikasi NLP yang lebih canggih[4].

Metode pembelajaran mesin berbasis aturan mencakup berbagai pendekatan seperti *Hidden Markov Model* (HMM), *Maximum Entropy*, *Decision Tree*, *Support Vector Machines* (SVM), *Conditional Random Fields* (CRF), dan *Hybrid approach*[5]. Selain itu, teknik-teknik lain seperti *Recurrent Neural Network* (RNN) dan variasinya, seperti *Long Short-Term Memory* (LSTM), juga

telah sukses diterapkan dalam sejumlah masalah prediksi urutan, termasuk *Named Entity Recognition* (NER), pemodelan bahasa, dan pengenalan suara[6].

Long Short-Term Memory (LSTM) dan *Bidirectional Long Short-Term Memory* (BiLSTM) merupakan metode tradisional yang bisa digunakan dalam mengekstraksi entitas. Namun metode tersebut memiliki permasalahan yang untuk mengekstraksi entitas pada berita, memiliki biaya komputasi yang tinggi dan kesulitan dalam interpretasi model dalam mengekstraksi entitas merupakan tantangan utama dalam menggunakan metode LSTM dan BiLSTM[7].

Berdasarkan penelitian sebelumnya *Classification of Turkish News Content by Deep Learning Based LSTM Using Fasttext*. Meskipun model berhasil mencapai akurasi sebesar 96.18% menggunakan model *Fasttext* pada dataset yang berisi 2.520.000 berita dalam berbagai kategori, terdapat beberapa tantangan yang perlu diatasi. Salah satu kendala yang dihadapi adalah waktu eksekusi yang cukup lama selama pelatihan model. Selain itu, perbaikan dalam interpretasi model dan optimasi proses komputasi masih diperlukan untuk meningkatkan efisiensi klasifikasi teks berita secara keseluruhan[8].

Berdasarkan penelitian sebelumnya tentang *Named-Entity Recognition for Indonesian Language using Bidirectional LSTM-CNNs* walaupun berhasil mencapai akurasi sebesar 79.43% dengan menggunakan dataset berupa 700.000 kata Bahasa Indonesia yang diambil dari Wikipedia, diperlukan usaha tambahan untuk mengatasi kendala biaya komputasi yang tinggi, kerumitan waktu eksekusi, serta perbaikan interpretasi model dalam mengekstraksi entitas pada berita berbahasa Indonesia[9].

Berdasarkan penelitian sebelumnya tentang *News text named entity Recognition based on BI-LSTM-CRF model*, hasil penelitian tersebut menghasilkan akurasi sebesar 94.53% dengan menggunakan dataset bahasa mandarin tentang berita diskriminasi. Kombinasi *BI-LSTM* dengan *CRF* berperan penting dalam meningkatkan akurasi, karena *CRF* mampu mempertimbangkan hubungan antar label secara lebih efektif, sehingga metode ini sangat baik dan akurat dalam mengekstraksi entitas-entitas penting pada berita[10].

Melihat kondisi tersebut dalam penelitian ini, bermaksud untuk mengimplementasikan metode *Bidirectional LSTM* dan *CRF* (LSTM-CRF) pada artikel berita politik di Indonesia dengan bertujuan untuk meningkatkan efisiensi dan akurasi dalam ekstraksi entitas organisasi partai politik, organisasi terkait, dan untuk memfasilitasi pemahaman yang lebih mendalam terhadap dinamika politik yang sedang berlangsung di Indonesia.

1.2 Rumusan masalah

1.2.1 Permasalahan

Pada penelitian sebelumnya dengan menggunakan LSTM dan BiLSTM yang sudah dipaparkan dalam latar belakang, memiliki biaya komputasi yang tinggi dan kesulitan dalam interpretasi model dalam mengekstraksi entitas organisasi dalam artikel berita politik.

1.2.2 Solusi Permasalahan

Solusi permasalahan yang digunakan adalah dengan menggunakan metode *Bidirectional LSTM* dan *CRF*. Metode ini digunakan untuk mengekstraksi entitas organisasi politik secara sistematis dalam data berita politik.

1.2.3 Pertanyaan Penelitian

1. Berapa lama waktu komputasi yang dibutuhkan metode *Bidirectional LSTM* dan *CRF* dalam melatih model untuk mengekstraksi entitas organisasi dalam artikel berita politik?
2. Bagaimana kinerja metode *Bidirectional LSTM* dan *CRF* dalam mengekstraksi entitas organisasi dalam artikel berita politik yang diukur dengan nilai *Precision*, *Recall*, *F1-Score* dan *Accuracy*?

1.3 Tujuan Dan Manfaat

1.3.1 Tujuan

Penelitian ini bertujuan mengetahui berapa lama komputasi dan kinerja metode *Bidirectional LSTM* dan *CRF* dalam mengekstraksi entitas organisasi dalam artikel berita politik yang diukur dengan nilai *Precision*, *Recall*, *F1-Score* dan *Accuracy*.

1.3.2 Manfaat

Adapun manfaat dari penelitian ini adalah :

- a. Mengetahui berapa lama komputasi yang dibutuhkan dan kinerja metode *Bidirectional LSTM* dan *CRF* dalam mengekstraksi entitas organisasi pada data atikel berita politik yang dikur dengan nilai *Precision*, *Recall*, *F1-Score* dan *Accuracy*.
- b. Dengan melakukan ekstraksi entitas dari suatu artikel berita politik pembaca lebih cepat memperoleh informasi dari artikel berita politik tersebut.

1.4 Batasan Masalah

Adapun batasan-batasan yang akan diberikan agar permasalahan tidak terlalu luas sehingga sesuai dengan tujuan penulis sebagai berikut:

1. Data yang digunakan hanya artikel berita politik yang ada di Indonesia.
2. Dataset diambil dari sumber detik.com dan CNN Indonesia mulai tanggal 01 desember 2023 sampai 29 february 2024
3. Penelitian ini berfokus pada ekstraksi entitas organisasi atau yang berkaitan dengan politik seperti nama orang, lokasi, waktu.

1.5 Sistematika Laporan

Sistematika penulisan pada laporan skripsi disusun untuk memberikan pembahasan singkat atau gambaran umum, dan telah disesuaikan dengan format penulisan skripsi jurusan Teknik Informatika Universitas Trunojoyo Madura sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini membahas tentang latar belakang, perumusan masalah, manfaat, tujuan, batasan masalah dan rancangan sistematika penulisan yang diajukan.

BAB II KAJIAN PUSTAKA

Bab ini membahas landasan teori terkait materi dan metode yang akan digunakan dalam penelitian, termasuk konsep *Bidirectional LSTM*, *CRF*, dan *Named Entity Recognition*.

BAB III METODE USULAN

Bab ini membahas metode usulan, perancangan, dan arsitektur sistem yang akan digunakan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini dijelaskan tentang pembahasan penelitian dengan menjelaskan penelitian berlangsung dan mendapatkan hasil pada penelitian.

BAB V PENUTUP

Bab ini dijelaskan tentang kesimpulan dari penelitian ini dan saran yang dapat dilakukan dan dikembangkan oleh penelitian kali ini.