

# 1. Pendahuluan

## Latar Belakang

Beberapa tahun terakhir, *parsing* menjadi sangat populer dalam ruang lingkup NLP (*Natural Language Processing*) dengan adanya *Dependency Parser* [9]. *Parsing* merupakan langkah penting dalam beberapa aplikasi yang melibatkan analisis dokumen, seperti *knowledge extraction*, *question answering*, *summarization*, atau *filtering* [1]. Secara umum, NLP menggunakan hasil yang diperoleh dari *parsing* menjadi dasar untuk pemrosesan lebih lanjut, dan ketepatan dalam *parsing* akan menghasilkan peningkatan dalam akurasi sistem [10]. *Dependency parsing* merupakan pendekatan yang digunakan untuk melakukan analisis sintaksis otomatis bahasa alami yang terinspirasi dari *dependency grammar*. Tujuan utama dari *dependency parsing* adalah untuk melakukan analisis struktur dependensi secara otomatis dari kalimat input yang diberikan [9].

Banyak sekali penelitian sebelumnya yang telah menghasilkan parser, seperti *PCFG Parser* [7], *Malt Parser* [13], *Arabic Parser* [4], *Chinese Parser* [14], dan lainnya. Namun, hampir semua *Dependency Parser* yang ada melakukan klasifikasi berdasarkan jutaan fitur indikator jarang (*sparse*). Fitur ini tidak hanya buruk dalam mengambil kesimpulan, tetapi juga membatasi kecepatan *parsing* secara signifikan sehingga *parsing* yang dihasilkan tidak maksimal. Untuk mengatasi permasalahan tersebut, dilakukan pergantian penggunaan fitur jarang (*sparse*) menjadi fitur padat (*dense*). Dengan penggunaan fitur padat dapat secara efektif mengurangi *sparsity* antar kata, dan memperoleh titik awal yang baik untuk membangun fitur kata dan interaksinya [2].

Pada penelitian ini, penulis menggunakan metode klasifikasi Jaringan Saraf Tiruan dalam *Transition-Based Dependency Parser* untuk membuat keputusan *parsing*. Metode klasifikasi Jaringan Saraf Tiruan hanya mempelajari dan menggunakan sejumlah kecil fitur padat *dense*, sehingga metode ini dapat bekerja dengan cepat dan ringkas. Metode ini juga dapat mempelajari POS (*Part of Speech*) *Tagging*, dan *Dependency Relation*. Dalam melakukan *parsing*, penulis menggunakan *Dependency Parser* yang menerapkan metode Jaringan Saraf Tiruan berdasarkan penelitian sebelumnya yaitu *A Fast and Accurate Dependency Parser using Neural Networks* [2]. Tujuan yang ingin dicapai dari penelitian ini adalah untuk menghasilkan klasifikasi yang cepat dan ringkas, baik dari segi akurasi LAS (*Labelled Attachment Score*) dan UAS (*Unlabelled Attachment Score*) maupun kecepatan *parsing* dengan menggunakan *Transition-Based Dependency Parser*.

## Topik dan Batasannya

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, topik-topik yang terkait dalam tugas akhir ini sebagai berikut:

### 1. *Dependency Parser*

Hasil dari tugas akhir ini adalah nilai LAS (*Labelled Attachment Score*) dan UAS (*Unlabelled Attachment Score*) serta kecepatan *parsing* yang diperoleh dari hasil *parsing* oleh *Transition-Based Dependency Parser*, sebelum melakukan *parsing* setiap kata pada dataset akan diubah menjadi vektor dengan *word-embeddings*, kemudian dilakukan *parsing* menggunakan metode klasifikasi Jaringan Saraf Tiruan, dan hasil dari *parsing* tersebut akan menghasilkan nilai akurasi LAS dan UAS serta kecepatan *parsing*.

### 2. *Input dan Output*

*Input* dari sistem adalah dataset Bahasa Arab, China, Inggris dan Indonesia yang telah dipisah menjadi 3 bagian berdasarkan aturan pemisah standar PTB3 yaitu *train*, *dev*, dan *test*. *Output* dari sistem adalah nilai LAS dan UAS serta kecepatan *parsing* yang diperoleh dari hasil *parsing* dengan *Dependency Parser*.

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka batasan dari permasalahan yang ada pada tugas akhir ini adalah:

- Dataset yang digunakan adalah Bahasa Arab, China, Inggris, dan Indonesia.
- Pengujian sistem menggunakan 2 *hyperparameter*.

## Tujuan

Tujuan penelitian ini adalah untuk melakukan *parsing* yang cepat dan ringkas menggunakan 2 *hyperparameter* pada dataset Bahasa Arab, China, Inggris, dan Indonesia, kemudian menghitung nilai akurasi LAS dan UAS serta kecepatan *parsing* yang dihasilkan oleh *Transition-Based Dependency Parser*.

## Organisasi Tulisan

Selanjutnya pada bab 2 akan dijelaskan mengenai *Dependency Parsing*, *Transition-Based Dependency Parsing*, Jaringan Saraf Tiruan, dan *Evaluation Metrics*, kemudian pada bab 3 akan dijelaskan sistem yang akan dibangun, selanjutnya pada bab 4 mengenai evaluasi dan analisis hasil pengujian sistem, diakhiri dengan bab 5 yaitu kesimpulan dan saran.