

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pada saat ini, perkembangan teknologi informasi sangat cepat, salah satunya adalah penggunaan internet (Wahyuni, Prastiyanto, & Supraptono, 2017). Berkembangnya internet berdampak terhadap bertambahnya jumlah informasi yang mengakibatkan sangat sulit untuk mendapatkan informasi secara efisien (Indrianto, Fauzi, & Muflikhah, 2017). Salah satu informasi yang sangat banyak dicari adalah berita (Wangsa, Utomo, & Nugroho, 2014).

Berita merupakan kumpulan informasi tentang banyak peristiwa penting terjadi dan terbaru secara berkala (Wahyuni et al., 2017). Pada tahun 2013, media online bernama kapanlagi yang merupakan salah satu penggerak pertama di industri media online Indonesia menyatakan bahwa perusahaan tersebut membuat sekitar 500 artikel perhari untuk lima anak media mereka. Jika dikalkulasikan dalam setahun satu perusahaan media online saja dapat menghasilkan 182.500 artikel.

Dengan banyaknya artikel berita yang dibuat setiap harinya dan perkembangan internet yang mendukung menyebabkan informasi berita online sangatlah mudah untuk didapatkan akan tetapi yang menjadi permasalahan di era informasi ini adalah mendapatkan informasi yang diinginkan dengan waktu yang pendek atau singkat (Indrianto et al., 2017).

Oleh karena itu diperlukan sebuah sistem yang dapat menyajikan informasi secara singkat namun mengandung informasi yang penting dari teks aslinya, hal itu disebut sebagai ringkasan (Indrianto et al., 2017). Dengan adanya ringkasan, pembaca dapat dengan cepat dan mudah memahami isi sebuah berita tanpa harus membaca keseluruhan isi berita (Budiyono & Solihin, 2014).

Peringkasan teks diklasifikasikan menjadi dua yaitu peringkasan ekstraktif dan peringkasan abstraktif. Peringkasan ekstraktif adalah peringkasan teks yang menghasilkan ringkasan dengan memilih beberapa kalimat penting dari teks aslinya. Sedangkan peringkasan abstraktif adalah peringkasan teks yang

menghasilkan ringkasan dengan memahami teks asli dan menghasilkan kalimat baru dari teks yang diringkaskan (Mandar & Gunawan, 2018).

Salah satu metode yang sering digunakan untuk peringkasan abstraktif adalah *Recurrent Neural Network* (RNN). RNN atau jaringan syaraf berulang adalah bagian dari jaringan syaraf tiruan yang dapat memproses data sekuensial (Yoko, Christanti M, & Hendryli, 2018).

RNN telah digunakan pada banyak penelitian peringkasan abstraktif seperti (Nallapati, Zhou, & Santos, 2016) yang menambahkan *attention mechanism*, *switching generator pointer* dan *hierarchical attention*. Kemudian pada penelitian (Hasselqvist, Helmertz, & Kågebäck, 2017) dengan menambahkan *query-based* pada pembentukan ringkasannya. Setelah itu pada penelitian (Paulus, Xiong, & Socher, 2017) dengan menambahkan *attention mechanism* dan pembelajaran baru yang menggabungkan antara *maximumlike-lihood* dan *reinforcement learning*. Kemudian penelitian terdekat yaitu (Y. Chen & Bansal, 2018) dengan menambahkan *reinforce-selected sentence rewriting*.

Keempat penelitian tersebut menggunakan *Rouge* untuk mengevaluasi modelnya. Perbandingan nilai akurasi terhadap CNN/Daily Mail dataset dapat dilihat pada Table 1.1.

**Tabel 1. 1 Perbandingan nilai akurasi penelitian-penelitian peringkasan abstraktif**

Penelitian	Rouge-1	Rouge-2	Rouge-L
(Nallapati et al., 2016)	36.40	17.77	33.71
(Hasselqvist et al., 2017)	33.81	18.19	29.22
(Paulus et al., 2017)	41.16	15.75	39.08
(Y. Chen & Bansal, 2018)	40.88	17.80	38.54

Dari Table 1.1 diatas dapat dilihat bahwa model RNN pada penelitian (Paulus et al., 2017) menghasilkan nilai *Rouge-1* terbaik dibandingkan dengan penelitian lainnya. Selain pada CNN/Daily Mail *dataset*, model yang dinamai *A Deep Reinforced Model* ini juga menghasilkan nilai *Rouge* yang baik pada New York Times *dataset* dengan nilai *Rouge-1*, *Rouge-2*, dan *Rouge-L* masing-masing

sebesar 47.22, 30.51, dan 43.27. Sebelum banyaknya penelitian yang menggunakan metode LSTM, terdapat penelitian peringkasan teks dengan menggunakan metode *Rules-Based*. *Rules-Based* digunakan untuk memecahkan masalah dengan aturan yang dibuat berdasarkan pengetahuan (S. H. Chen, Jakeman, & Norton, 2008). Pada penelitian (El-Kilany & Saleh, 2012) peringkasan teks dihasilkan dengan memanfaatkan *dependency parser* dan menghasilkan nilai *Rouge* yang baik, yaitu *Rouge-1* sebesar 48.8, dan *Rouge-L* sebesar 44 pada DUC 02 *dataset*. Oleh karena itu penulis melakukan penelitian untuk mengetahui nilai *Rouge* dari *Deep Reinforced Model* dan *Rules-Based* dengan menggunakan *dependency parser* untuk peringkasan kalimat Bahasa Indonesia.

Berdasarkan latar belakang diatas, maka dilakukan penelitian dengan judul “*Deep Reinforced Model* dan *Rules-Based* untuk Peringkasan Kalimat Bahasa Indonesia”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah dari penelitian ini adalah :

- a. Bagaimana mengimplementasikan *Deep Reinforced Model* dan *Rules-Based* untuk peringkasan kalimat Bahasa Indonesia?
- b. Bagaimana nilai akurasi dari *Deep Reinforced Model* dan *Rules-Based* yang sudah diimplementasikan?

## 1.3 Batasan Masalah

Batasan masalah yang ditetapkan pada penelitian ini adalah sebagai berikut:

- a. Bahasa dari kalimat yang digunakan dalam penelitian ini adalah Bahasa Indonesia.
- b. Penelitian dilakukan pada dataset yang berupa teks.
- c. Data kalimat diambil dari berita dengan topik Lalu Lintas
- d. Data berita diambil dari situs berita online yaitu [cnnindonesia.com](http://cnnindonesia.com).

## 1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini berdasarkan rumusan masalah yang telah disebutkan adalah:

- a. Mengimplementasi *Deep Reinforced Model* dan *Rules-Based* untuk peringkasan kalimat Bahasa Indonesia.
- b. Melakukan pengujian untuk mengetahui nilai *Rouge* dari *Deep Reinforced Model* dan *Rules-Based*.

### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

- a. Memberikan kalimat yang lebih singkat tanpa menghilangkan inti informasi dari kalimat.
- b. Memberikan hasil analisis kinerja dari sistem peringkasan kalimat Bahasa Indonesia yang menggunakan *Deep Reinforced Model* dan *Rules-Based*.
- c. Menghasilkan *dataset* yang dapat digunakan untuk penelitian peringkasan kalimat selanjutnya.

### 1.6 Sistematika Penulisan

Dalam mempermudah penulisan skripsi, penulis membuat sistematika penulisan yang terdiri dari:

## BAB I PENDAHULUAN

Bab I menjelaskan mengenai latar belakang masalah, rumusan masalah, batasan masalah, serta tujuan dan manfaat dari penelitian. Latar belakang menjelaskan alasan dilakukannya penelitian dan pengambilan judul penelitian ini yang meliputi pentingnya peringkasan kalimat dalam perkembangan teknologi yang semakin pesat dan metode yang akan digunakan untuk membangun sistem peringkasan kalimat.

## BAB II KAJIAN PUSTAKA

Bab II terdiri dari kajian pustaka singkat mengenai teori-teori yang dipergunakan pada penelitian ini, misalnya *Recurrent Neural Network*, *Long-short Term Memory*, *Rules-Based*, *Dependency Parser* dan *Rouge*. *Long-short Term Memory* adalah salah satu jenis arsitektur RNN yang digunakan untuk mengimplementasikan *Deep Reinforced Model* dan *Dependency Parser* digunakan untuk membuat aturan penghapusan kata untuk peringkasan kalimat berdasarkan

*Rules-Based*. Sedangkan *Rouge* digunakan untuk mengevaluasi hasil implementasi dari *Deep Reinforced Model* dan *Rules-Based*.

### **BAB III METODOLOGI PENELITIAN**

Bab III menjelaskan mengenai langkah-langkah yang akan dilakukan pada penelitian ini. Langkah-langkah tersebut meliputi desain penelitian serta alat dan bahan penelitian. Desain penelitian menjelaskan tahap penelitian mulai dari pengumpulan data sampai dokumentasi. Sedangkan alat dan bahan penelitian menjelaskan alat apa saja yang digunakan dalam melakukan penelitian, seperti perangkat keras, perangkat lunak, dan library yang digunakan serta bentuk data yang digunakan.

### **BAB IV HASIL PENELITIAN DAN PEMBAHASAN**

Bab IV terdiri dari hasil penelitian serta pembahasan dari setiap tahapan penelitian yang telah disusun pada metodologi penelitian. Dalam pembahasan akan dijelaskan bagaimana data penelitian dibuat, tahapan-tahapan yang dilakukan sebelum memproses data pada *Deep Reinforced Model* dan *Rules-Based*, implementasi *Deep Reinforced Model* dan *Rules-Based* yang dilakukan, skenario pengujian yang akan dilakukan untuk mendapatkan nilai *Rouge* yang terbaik dan hasil pengujian dari skenario tersebut.

### **BAB V KESIMPULAN DAN SARAN**

Bab V terdiri dari kesimpulan dari penelitian yang telah dilakukan dan saran untuk pengembangan selanjutnya. Kesimpulan berisi simpulan dari keseluruhan penelitian yang sudah dilakukan sedangkan saran berisi kelemahan-kelemahan dari penelitian yang sudah dilakukan dan bagaimana cara untuk memperbaiki kelemahan tersebut.