

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi internet berarti semakin banyak berita dapat dicari dan dipelajari tanpa banyak usaha. Teknologi internet juga menghubungkan banyak negara berbeda ke satu sama lain dan memperbolehkan pertukaran berita dengan cepat. Tentunya, dengan jumlah berita di internet terus bertambah, akan terdapat banyak berita yang sangat panjang dan padat sehingga akan memakan banyak waktu untuk membacanya. Selain itu, berita yang didapat di internet mungkin ditulis dalam bahasa yang berbeda dan tidak dapat dibaca orang Indonesia. Banyak berita yang ada di Internet ditulis dengan Bahasa Inggris dan ini dapat menyulitkan orang Indonesia yang tidak dapat menggunakan bahasa tersebut. Kedua halangan yang disebut tidak selalu muncul secara eksklusif dan berita dapat berupa panjang dan ditulis dalam bahasa lain.

*Deep Learning* adalah salah satu metode *Machine Learning* dimana arsitektur terbentuk dari sebuah *Neural Network* yang terdiri atas banyak layer. Teknologi *Deep Learning* dapat digunakan untuk memecahkan masalah yang dipaparkan sebelumnya melalui peringkasan dan penerjemahan teks berita tanpa bantuan dari manusia (Chaudhary & Patel, 2018). Untuk hal seperti peringkasan dan penerjemahan teks atau disebut dengan *Natural Language Processing* yang membutuhkan pengertian tinggi atas konteks dan posisi kata, *Deep Learning* cocok untuk mengatasi masalah ini. Arsitektur *Deep Learning* yang dapat digunakan untuk memroses teks dan memahami konteks serta posisi dari sebuah kata yakni *Recurrent Neural Network* atau RNN (Chaudhary & Patel, 2018).

Peringkasan teks atau *Text Summarization* merupakan proses dimana teks panjang diringkaskan sampai isi dari teks tersebut hanya inti topik yang dibicarakan teks tersebut (Ivanedra & Mustikasari, 2019). Untuk teks panjang seperti berita, menggunakan model *Deep Learning* yang dibuat

untuk peringkasan teks, kita dapat mendapatkan inti dari berita tersebut tanpa membaca banyak dari berita.

Penerjemahaan teks atau *Machine Translation* adalah proses dimana sebuah model *Deep Learning* melakukan penerjemahaan dari satu bahasa ke bahasa lain terhadap sebuah teks tanpa melakukan penerjemahan manual (Chaudhary & Patel, 2018). Jadi sebuah teks Bahasa Inggris dapat diubah menjadi teks Bahasa Indonesia tanpa bantuan dari manusia.

Bedasarkan pemaparan mengenai penerapan *Deep Learning* dan RNN, metode ini dapat digunakan untuk meringkas teks berita dan setelah itu menerjemahkan teks tersebut ke Bahasa Indonesia. Penelitian ini akan merumuskan sebuah sistem dimana jika teks berita yang berbahasa Inggris masuk akan menghasilkan sebuah teks berita yang telah dirinkas dan diterjemahkan. Proses ini akan dilakukan dengan pembuatan dua model RNN berbeda dimana dua model tersebut masing-masing dapat melakukan antara peringkasan atau penerjemahan teks. Keluaran dari model pertama, yaitu model peringkasan, akan menjadi masukan bagi model yang kedua. Model kedua, model penerjemahan, akan menghasilkan keluaran terakhir yaitu berita yang telah dirinkas dan diterjemahkan ke Bahasa Indonesia. Kedua model ini akan digabung menggunakan API yang dapat menerima teks lalu mengembalikan teks tersebut ke pengguna.

## **1.2 Rumusan Masalah**

Bedasarkan latar belakang yang sudah dipaparkan, rumusan masalah untuk penelitian ini adalah bagaimana membuat dan menerapkan dua model *Recurrent Neural Network* untuk peringkasan dan penerjemahaan teks dari Bahasa Inggris ke Bahasa Indonesia

## **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah untuk membuat sebuah Sistem Deep Learning menggunakan Recurrent Neural Network untuk melakukan peringkasan dan penerjemahan teks berita

## **1.4 Manfaat Penelitian**

Berikut adalah manfaat dari penelitian ini:

1. Sistem dapat membantu dalam meringkaskan dan menerjemahkan teks berita yang berbahasa Inggris agar lebih mudah dan cepat dibaca
2. Dapat digunakan untuk pengembangan sistem untuk ke bahasa lain tidak hanya Bahasa Inggris ke Bahasa Indonesia
3. Dapat berguna bagi pembaca yang tidak dapat menggunakan Bahasa Inggris untuk mencerna berita dan tidak kena misinformasi.
4. Perusahaan berita dapat menggunakan sistem untuk membuat judul yang sempurna untuk sebuah berita

### **1.5 Batasan Masalah**

Berikut merupakan beberapa batasan masalah bagi penelitian ini:

1. Model menerima teks berita berbahasa Inggris yang akan diringkas dan output model tersebut diterjemahkan oleh model lain sehingga mengeluarkan teks Bahasa Indonesia
2. Model menggunakan dataset Gigaword untuk peringkasan dan dataset CCMatrix untuk peringkasan dari Bahasa Inggris ke Bahasa Indonesia
3. Dua model yang dibuat akan dilatih dengan dataset ringkasan teks berita Bahasa Inggris dan dataset penerjemahan Bahasa Inggris ke Bahasa Indonesia

### **1.6 Luaran yang diharapkan**

Luaran penelitian ini adalah sebuah sistem *Recurrent Neural Network* yang dapat digunakan untuk melakukan peringkasan dan penerjemahan teks berita dari Bahasa Inggris ke Bahasa Indonesia. Sistem ini akan berupa API lokal yang dapat menerima input berupa teks dan akan mengembalikan teks yang sudah diringkas dan diterjemahkan

### **1.7 Sistematika Penulisan**

Sistematika penulisan penelitian ini sebagai berikut:

#### **BAB 1 PENDAHULUAN**

Pada bab ini, akan dijelaskan latar belakang penelitian serta perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, luaran

penelitian, dan sistematika penulisan. Bab ini menerangkan fondasi dari penelitian ini.

## BAB 2 TINJAUAN PUSTAKA

Bab ini menjelaskan ilmu dan teori yang akan digunakan untuk penelitian serta penelitian sebelumnya yang pernah dilakukan. Untuk penelitian ini, akan dijelaskan definisi dari ilmu fondasi penelitian seperti Deep Learning dan Natural Language Processing

## BAB 3 METODOLOGI PENELITIAN

Bab ini akan memaparkan rencana penelitian yang akan dilakukan dan menjelaskan tahap apa saja yang ada. Setiap tahap akan dijelaskan bagaimana tahap tersebut akan dijalankan dan apa saja yang dibutuhkan.