

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Peringkasan sebuah teks adalah sebuah proses untuk menghasilkan ringkasan dari sebuah dokumen berbentuk berita, ulasan, artikel, atau makalah dari internet. Ringkasan adalah satu cara yang dapat membantu untuk mendapatkan sebuah informasi dari suatu dokumen yang panjang dalam waktu yang singkat. Peringkasan terdiri dari dua tipe yaitu ekstraktif dan abstraktif. Ekstraktif meringkas suatu dokumen dengan memilih sebagian dari kalimat yang ada di dalam dokumen asli. Abstraktif melakukan peringkasan dengan cara merubah teks asal melalui proses transformasi dokumen asli [1].

Penelitian peringkasan teks pertama kali dilakukan oleh Luhn pada tahun 1958, pada penelitian tersebut menghasilkan sebuah ringkasan yang menghitung nilai frekuensi kata dan kalimat untuk menghasilkan ringkasan yang baik [2]. Pada penelitian sebelumnya yang dilakukan oleh Kasyfi I dan Metty M pada tahun 2018 yang berjudul Implementasi Metode *Recurrent Neural Network* Pada Teks *Summarization* Dengan Teknik Abstraktif ini menggunakan salah satu jenis variansi *Recurrent Neural Network* (RNN) yang dinamakan *Long Short Term Memory* (LSTM), dan mendapatkan nilai rata-rata *recall non stemming* sebesar 41%, *precision* sebesar 81%, dan *F-measure* sebesar 54,27%. Sedangkan nilai rata-rata *recall* artikel berita dengan teknik *stemming* sebesar 44%, *Precision* sebesar 88% dan *F-measure* sebesar 58,20% [3].

Penelitian lainnya tentang *Recurrent Neural Network* mengenai Sistem Peringkasan teks otomatis dengan variansi *Long Short Term Memory* (LSTM) telah dilakukan oleh Kuncoro Yoko, Viny Christanti M, dan Janson Hendyli, dengan mendapatkan tingkat akurasi terbaik sebesar 99,8810%, dan *Q&A*

Evaluation menunjukkan akurasi sebesar 46,65% [4]. Besarnya tingkat akurasi dalam penelitian yang sudah ada telah membuktikan bahwa peringkasan teks otomatis menggunakan salah satu jenis varian RNN yang dinamakan *Long Short Term Memory* (LSTM) telah berhasil dilakukan.

Adapun penelitian yang dilakukan oleh Junyoung Chung, Caglar Gulcehre, Kyung Hyun Cho, dan Yoshua Bengio pada tahun 2014, menyimpulkan bahwa dengan menggunakan jumlah parameter tetap untuk semua model pada beberapa dataset GRU, dapat mengungguli unit LSTM baik dalam hal konvergensi dalam waktu CPU dan dalam hal parameter pembaruan dan generalisasi. Selain itu arsitektur GRU lebih sederhana dibandingkan dengan LSTM [5], karena GRU melibatkan perhitungan yang lebih sedikit, dengan begitu GRU dapat dilatih dengan cepat dibanding LSTM. Ada juga penelitian yang dilakukan oleh Rike Adelia, Suyanto Suyanto, dan Untari Novia W, tentang peringkasan teks abstractive menggunakan GRU dengan menggunakan model *word embedding word2vec* mendapatkan nilai *ROUGE-1* sebesar 0.11975 dan *ROUGE-2* sebesar 0.01199 [6].

Penelitian sebelumnya yang dilakukan oleh Arliyanti Nurdin, Bernadus Anggo Seno Aji, Anugrayani Bustamin dan Zaenal Abidin tentang perbandingan word embedding Word2Vec, Glove, dan FastText berdasarkan hasil evaluasi dari penelitian tersebut diantara ketiga word embedding tersebut, kinerja dari fasttext paling unggul dalam dataset yang digunakan [7]. FastText dapat memberikan representasi kata yang tidak ada dalam data latih, sehingga fasttext dapat mengatasi masalah *out of vocabulary*. Maka pada penelitian ini penulis akan menggunakan model *word embedding* berupa *fasttext* dan menggunakan salah satu metode jenis varian dari RNN yang disebut *Gated Recurrent Unit* (GRU), untuk melakukan peringkasan teks abstraktif, yang nantinya performa sistem diharapkan dapat bekerja lebih baik.

1.2 Identifikasi Masalah

Berdasarkan dari latar belakang yang telah diuraikan diatas, maka dapat diidentifikasi permasalahan yang dihadapi yaitu arsitektur LSTM yang digunakan rumit dan mengakibatkan banyaknya proses. Oleh karena itu penulis akan mengimplementasikan metode *Gated Recurrent Unit* dengan *word embedding* model berupa *FastText* untuk peringkasan teks berita bahasa indonesia.

1.3 Maksud dan Tujuan

Berdasarkan permasalahan yang ada, maksud dari penelitian ini adalah untuk mengimplementasikan Metode *Recurrent Neural Network* (RNN) dengan jenis varian yang disebut *Gated Recurrent Unit* (GRU) pada Aplikasi Peringkasan Teks Bahasa Indonesia. Adapun tujuan dari penelitian ini yaitu untuk mengukur performa yang didapatkan dari penggunaan metode GRU dalam peringkasan teks berbahasa Indonesia.

1.4 Batasan Masalah

Agar penelitian yang dilakukan lebih terarah sesuai dengan tujuan yang ingin dicapai, maka batasan masalah dalam penelitian ini adalah sebagai berikut:

- 1) Masukan :
 - a. Dokumen berbentuk teks.
 - b. Dokumen yang diambil merupakan paragraf dari berita berbahasa Indonesia.
 - c. Berita yang diambil merupakan berita dari website kompas.com dan detik.com.
 - d. Format data berupa CSV.
- 2) Proses :
 - a. Metode yang digunakan adalah salah satu jenis varian dari RNN yaitu *Gated Recurrent Unit* (GRU).

- b. Metode preprocessing yang digunakan yaitu *Case Folding*, *Tokenizing* kata, *Stopword*, dan vektorisasi dengan menggunakan metode *fasttext*.
- 3) Keluaran :
 1. Keluaran yang dihasilkan yaitu teks berupa hasil dari sistem peringkasan teks paragraf berita menggunakan metode *gated recurrent unit* dengan *stopword*, dan tanpa *stopword*.

1.5 Metode Penelitian

Dalam penelitian ini yang dilakukan adalah metode deskriptif. Metode ini digunakan karena pemecahan masalah berdasarkan fakta yang tampak atau apa adanya. Dimana dalam penelitian ini memiliki tahapan-tahapan seperti berikut.

1.5.1 Studi Literatur

Studi literatur dilakukan dengan mempelajari, meneliti, berbagai literatur dari berbagai macam sumber yang terdiri dari buku, jurnal ilmiah, situs internet yang berkaitan dengan metode dalam peringkasan teks otomatis, metode *Recurrent Neural Network*, dan metode *Gated Recurrent Unit*.

1.5.2 Metode Pengumpulan Data

Data yang akan dikumpulkan merupakan data yang berisi teks dari situs-situs berita seperti www.detik.com, dan www.kompas.com. Yang nantinya akan dijadikan sebagai data latih dan data uji.

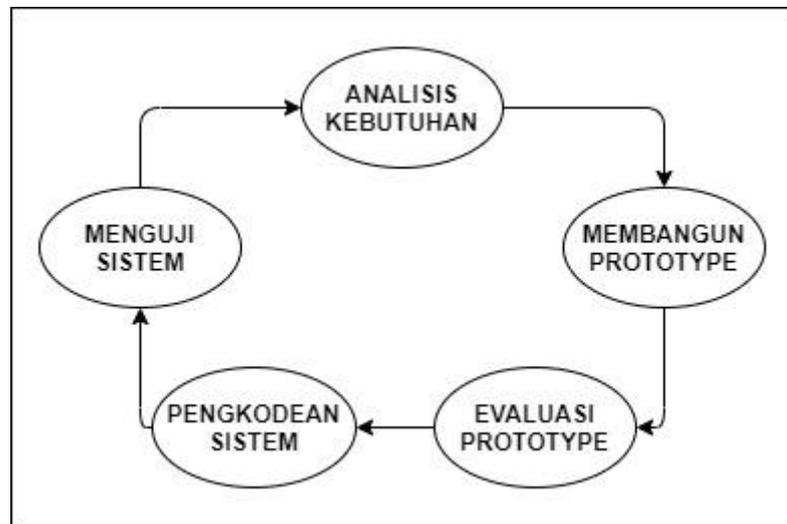
1.5.3 Analisis Metode

Tahap ini melakukan analisis pada metode yang akan digunakan, seperti metode yang akan digunakan pada tahap preprocessing, metode untuk vektorisasi hingga metode *Gated Recurrent Unit* (GRU).

1.5.4 Metode Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak yang akan digunakan pada penelitian ini adalah metode *prototype*. Karena setiap tahapannya dapat dievaluasi dan dirancang kembali jika tidak sesuai dengan yang diharapkan [8].

Tahapan dari metode *prototype* yang akan digunakan dapat dilihat pada Gambar 1.1.



Gambar 1. 1 Metode Prototype

Berikut penjelasan tahapan tahapan dalam metode prototype :

- 1) Analisis kebutuhan, merupakan menganalisis segala kebutuhan yang diperlukan dalam pembuatan sistem peringkasan teks, seperti dataset yang diperlukan, hingga analisis metode *Gated Recurrent Unit*.
- 2) Pembuatan prototype mengimplementasikan sistem menjadi bentuk prototype mulai dari tahapan preprocessing hingga pengujian, dengan menerapkan metode GRU pada program peringkasan teks.
- 3) Evaluasi Prototype, tahap ini untuk mengetahui apakah prototype sudah sesuai dengan yang akan dibentuk.
- 4) Pengkodean Sistem, pada tahap ini prototype yang sudah terbentuk diubah ke dalam bahasa pemrograman.
- 5) Evaluasi Sistem, perangkat yang telah diuji akan dievaluasi jika ada kekurangan pada program. Jika terdapat kekurangan pada hasil evaluasi, maka prototype akan dikaji ulang hingga sesuai dengan hasil yang diinginkan.

1.5.5 Pengujian Metode

Aplikasi yang telah selesai dibuat akan diuji secara menyeluruh, yang nantinya akan masuk kedalam pengujian metode untuk dapat mengetahui tingkat akurasi yang akan didapatkan dengan metode GRU untuk kasus peringkasan teks pada berita.

1.5.6 Penarikan Kesimpulan

Setelah setiap tahap selesai dilakukan, penulis akan melakukan penarikan kesimpulan dari hasil yang telah didapatkan pada penerapan metode GRU dalam kasus peringkasan teks otomatis yang dilakukan.

1.6 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran secara umum mengenai penelitian secara umum. Sistematika penulisan penelitian ini adalah sebagai berikut :

BAB 1 PENDAHULUAN

Bab ini akan membahas kerangka penelitian atau percobaan dalam penelitian, meliputi latar belakang permasalahan, identifikasi masalah, menentukan maksud dan tujuan penelitian, batasan masalah, metode penelitian serta sistematika penulisan dari penelitian.

BAB 2 TINJAUAN PUSTAKA

Bab ini memuat berbagai konsep dasar dan teori-teori yang berkaitan dengan penelitian yang dilakukan. Seperti *preprocessing* yang terdiri dari pemisahan kalimat, *case folding*, tokenisasi, *stopword removal*, vektorisasi serta metode *Gated Recurrent Unit*.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini berisi tentang tahapan analisis dan perancangan, Analisis yang akan dibahas berupa analisis masalah analisis data masukan, analisis proses, analisis data keluaran, analisis kebutuhan fungsional dan nonfungsional.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini berisi tentang implementasi dan pengujian dari perangkat lunak yang dibangun sesuai dengan analisis yang telah dibuat. Mulai dari implementasi hardware dan software yang digunakan, serta hasil pengujian. Hasil pengujian yang dilakukan yaitu mengukur performa dari *metode gated recurrent unit*.

BAB 5 KESIMPULAN DAN SARAN

Pada bab ini akan diuraikan kesimpulan dari penelitian yang telah dilakukan, serta akan membahas tentang saran yang dapat dijadikan masukan untuk penelitian selanjutnya.

