

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Ketersediaan dokumen-dokumen yang tersebar di internet membutuhkan banyak riset di dalam ranah peringkasan teks otomatis [1]. Teks yang lebih pendek akan mengurangi waktu baca dan ukuran *file* [4]. Sehingga, ini memberikan benefit baik bagi manusia maupun komputer. *Automatic text summarization* (peringkasan teks otomatis) adalah sebuah pendekatan untuk menghasilkan sebuah ringkasan yang pendek dengan mempertahankan informasi kunci dan keseluruhan makna [1]. Terdapat dua jenis metode untuk peringkasan teks otomatis, yaitu ekstraktif dan abstraktif. Metode ekstraktif adalah metode untuk meringkas dokumen dengan cara memilih bagian dari kata atau kalimat di dalam dokumen [2], metode ini bergantung pada ekstraksi kalimat-kalimat dari teks aslinya [1]. Sedangkan, metode abstraktif adalah metode untuk meringkas dokumen dengan cara membuat kalimat-kalimat baru yang mempunyai informasi yang sama dengan dokumen aslinya [2]. Ringkasan abstraktif juga dinilai sebagai suatu perkiraan yang lebih dekat dengan ringkasan manusia karena bisa mengandung ekspresi yang tidak ada di dalam teks asli [1]. Dengan demikian, penelitian ini akan menerapkan metode abstraktif karena pendekatannya dalam membuat ringkasan mirip seperti halnya manusia.

Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), dan *gated recurrent neural networks* pada umumnya, telah menjadi pendekatan-pendekatan *state of the art* yang mapan dalam *sequence modeling* [3], contohnya pada peringkasan teks abstraktif. Sudah banyak penelitian yang mendorong batasan-batasan model *recurrent* dan arsitektur *encoder-decoder* [3]. Namun, model *recurrent* menghasilkan deretan *hidden state* [3], karena sifatnya yang bekerja secara berurutan, hal ini menjadi masalah dalam paralelisasi [4]. Rike Adelia dkk. [10] melakukan peringkasan teks abstraktif menggunakan Bidirectional Gated Recurrent Unit (BiGRU), salah satu arsitektur RNN, pada teks bahasa Indonesia. Dalam penelitiannya disebutkan bahwa model yang dibangun

berhasil mempelajari kata-kata tunggal dari sumber teks aslinya. Namun, ringkasan yang dihasilkan masih terdapat kekurangan dalam kohesi dan tata bahasa.

Pada tahun 2017, A. Vaswani dkk. [3] mengajukan arsitektur bernama Transformer yang memanfaatkan mekanisme *self-attention*. Model Transformer berhasil mencapai skor BLEU sebesar 28,4 pada *dataset* WMT 2014 English-to-German dan 41,8 skor BLEU pada *dataset* WMT 2014 English-to-French [3], menjadikan Transformer sebagai *state of the art* yang baru pada tahun tersebut. Berdasarkan eksperimen tersebut, Transformer menunjukkan hasil yang bagus secara kualitas dan dapat melakukan paralelisasi sehingga membutuhkan waktu yang lebih sedikit untuk dilatih [3], ini dikarenakan Transformer tidak menggunakan RNN. Selain translasi, Transformer juga bisa digunakan pada peringkasan. Anushka dkk. [4] mengimplementasikan Transformer untuk melakukan peringkasan dengan menggunakan *dataset* yang berisikan kumpulan berita dari BBC. Pada penelitiannya, satuan pengukuran yang digunakan untuk mengukur kualitas ringkasan adalah ROUGE, dan berhasil mencapai skor ROUGE-1 sebesar 0.47 dan ROUGE-2 sebesar 0.33.

Self-attention dapat menentukan apakah suatu kata berhubungan erat dengan kata yang lain dengan memberikan nilai *attention* yang lebih tinggi pada kata tersebut. Selain itu, terdapat *multi-head attention*, sebuah metode untuk melakukan *self-attention* lebih dari sekali. Terdapat indikasi bahwa *attention* menunjukkan perilaku yang berhubungan dengan struktur kalimat, baik secara sintaktis maupun semantik [3]. Dengan demikian, penulis tertarik untuk mengimplementasikan Transformer dalam peringkasan teks abstraktif dengan harapan model yang dibangun mampu menghasilkan ringkasan yang memiliki kohesi dan tata bahasa yang baik—dapat dipahami oleh pembaca.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijabarkan, penelitian ini akan mengevaluasi seberapa baik performa Transformer pada peringkasan teks abstraktif dalam bahasa Indonesia. Adapun evaluasi tersebut akan dilakukan

dengan cara mengukur ringkasan yang dihasilkan oleh model untuk meninjau kualitas dari ringkasan tersebut.

1.3 Maksud dan Tujuan Penelitian

Maksud dari penelitian ini adalah mengimplementasikan model Transformer pada peringkasan teks abstraktif. Adapun tujuan dari penelitian ini adalah mengukur performa model Transformer, yang dilatih menggunakan *dataset* yang sudah melalui tahap *preprocessing*, pada peringkasan teks abstraktif–pengukuran dilakukan menggunakan satuan ROUGE.

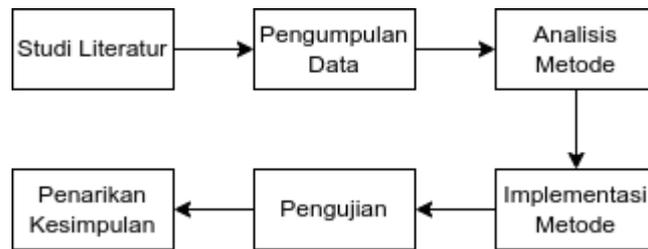
1.4 Batasan Masalah

Berikut ini batasan-batasan masalah yang menjadi ruang lingkup pekerjaan yang akan dilakukan.

1. *Dataset* yang digunakan adalah kumpulan dokumen bahasa Indonesia dari IndoSum yang berisikan 18774 berita.
2. Tahap *preprocessing* menggunakan:
 - a) *Lower casing* bertujuan untuk mengubah semua huruf alfabet pada teks menjadi *lower case* (huruf kecil).
 - b) *Filtering* atau penyaringan terhadap teks untuk menghilangkan karakter-karakter yang tidak akan digunakan.
 - c) *Tokenization* menggunakan SentencePiece untuk memecah teks menjadi *token*.
3. Menghasilkan ringkasan yang lebih pendek dari dokumen aslinya di mana panjang ringkasan adalah maksimal 50% dari panjang dokumen asli.
4. Menggunakan pengukuran ROUGE untuk evaluasi hasil ringkasan dari data uji.

1.5 Metode Penelitian

Metode penelitian yang diterapkan adalah metode deskriptif kuantitatif untuk mengamati dan menjelaskan fenomena yang dipelajari dan melakukan penarikan kesimpulan menggunakan angka. Penelitian ini mempelajari tentang peringkasan teks abstraktif menggunakan Transformer. Hasil penelitian yang diperoleh akan dievaluasi berdasarkan teknik pengukuran yang digunakan.



Gambar 1.1 Alur Penelitian

1. Studi Literatur

Penulis melakukan studi literatur dengan cara mempelajari sumber-sumber informasi dari jurnal ilmiah, buku, dan konten-konten yang tersedia di internet dalam bentuk teks dan video, terutama konten yang berkaitan dengan peringkasan teks abstraktif.

2. Pengumpulan Data

Dataset yang digunakan berisi berita-berita dalam bahasa Indonesia yang terdiri dari teks asli dan ringkasannya untuk menunjang kebutuhan pelatihan model Transformer pada peringkasan teks abstraktif.

3. Analisis Metode

Analisis metode meliputi pemilihan data masukan yang akan digunakan, tahapan-tahapan *preprocessing* yang akan digunakan, pengimplementasian model Transformer dengan *fine tuning*, hingga menentukan satuan pengukuran.

4. Implementasi Metode

Implementasi metode bertujuan untuk menerapkan metode-metode yang telah ditentukan berdasarkan tahap analisis metode, misalnya menerapkan metode-metode yang digunakan pada tahap *preprocessing* dan pendefinisian *hyper parameter* untuk *fine tuning*.

5. Pengujian

Tahap pengujian dilakukan setelah implementasi metode. Tahap ini bertujuan untuk mengevaluasi performa pada peringkasan teks abstraktif yang menggunakan model Transformer.

6. Penarikan Kesimpulan

Penulis akan melakukan penarikan kesimpulan berdasarkan hasil yang diperoleh dari implementasi dan evaluasi model Transformer pada peringkasan teks abstraktif yang telah dilakukan.

1.6 Sistematika Penulisan

Sistematika penulisan menggambarkan susunan bab yang akan dibahas pada penelitian beserta penjelasannya secara umum. Terdapat 5 (lima) bab pada penelitian ini yang terdiri dari:

BAB 1 PENDAHULUAN

Bab ini membahas tentang latar belakang masalah yang menjadi landasan penelitian ini dilakukan, rumusan masalah, maksud dan tujuan dari penelitian, lingkup batasan masalah, metode penelitian yang digunakan, dan sistematika penulisan untuk penelitian ini.

BAB 2 LANDASAN TEORI

Bab ini membahas tentang konsep-konsep yang digunakan pada penelitian. Mulai dari membahas tentang hal yang sangat mendasar seperti pengertian ringkasan, metode-metode yang digunakan untuk *preprocessing*, sampai dengan meninjau arsitektur Transformer secara detail.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini membahas tentang analisis data masukan dan metode-metode yang akan diterapkan untuk menangani bagaimana data tersebut akan diolah. Terdapat juga analisis mengenai perancangan sistem yang menggambarkan bagaimana metode-metode tersebut akan diimplementasikan.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini menjelaskan tentang implementasi dan pengujian dari sistem yang telah dibangun berdasarkan analisis dan perancangan yang telah dibuat pada bab 3. Tujuan dari implementasi dan pengujian sistem adalah untuk mengukur performa yang dihasilkan oleh model Transformer setelah menerapkan metode dan perancangan yang telah ditentukan.

BAB 5 KESIMPULAN DAN SARAN

Bab ini memuat kesimpulan dan saran berdasarkan penelitian yang telah dilakukan dalam membangun model Transformer untuk melakukan peringkasan teks abstraktif. Kesimpulan berisi penjelasan singkat mengenai hasil penelitian yang diperoleh, sedangkan saran berisi masukan-masukan yang bisa dijadikan pertimbangan untuk diterapkan pada penelitian selanjutnya.