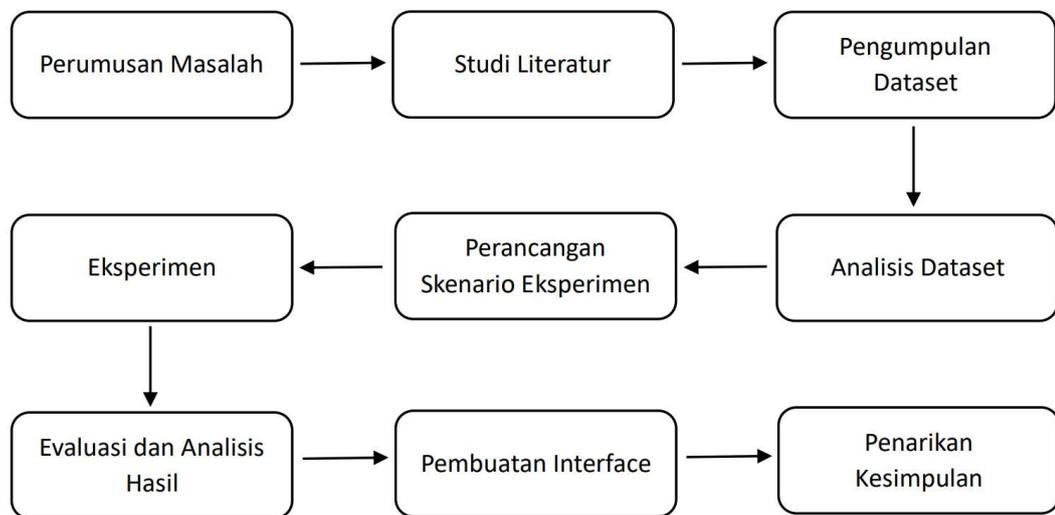


BAB III

METODOLOGI PENELITIAN

3.1 Desain Penelitian

Desain penelitian merupakan serangkaian alur kerja yang digunakan dalam melakukan penelitian. Bagian ini berisi gambaran mengenai kerangka kerja yang akan dilakukan dari awal hingga berakhirnya penelitian. Kerangka kerja tersebut diilustrasikan dalam Gambar 3.1.



Gambar 3.1 Desain Penelitian

Berikut penjelasan dari Gambar 3.1 yang akan menjadi alur penelitian yang dilakukan oleh penulis:

1. Perumusan Masalah

Perumusan masalah merupakan tahap awal yang dilakukan untuk memahami latar belakang masalah, menentukan solusi yang dapat menyelesaikan permasalahan dari tema penelitian yang akan dilakukan serta menentukan tujuan penelitian yang akan dilakukan. Selain itu, pada tahap perumusan masalah juga terdapat poin-poin permasalahan yang akan menjadi pembahasan dalam penelitian.

2. Studi Literatur

Studi literatur adalah tahapan untuk mengumpulkan data-data yang menunjang penelitian berupa teori pendukung dari penelitian yang telah dilakukan sebelumnya dalam bentuk jurnal, buku, situs, *e-book* dan sumber lainnya.

Adapun literatur yang dikumpulkan berkaitan dengan tema penelitian seperti berita, peringkasan teks, peringkasan teks, peringkasan berita, *deep learning*, LSTM, *encoder* dan *decoder*, mekanisme atensi, *Transformer*, dan topik pendukung lainnya. Literatur tersebut dapat digunakan sebagai bahan rujukan dalam menjawab rumusan masalah dan digunakan untuk memahami algoritma metode untuk menyelesaikan permasalahan yang ada pada penelitian.

3. Pengumpulan Dataset Berita

Dataset berita merupakan kumpulan artikel berita yang diterbitkan pada situs portal berita *online*. Data berita yang digunakan pada penelitian ini didapat dari liputan6.com dan bbc.com yang tersedia secara *online*. Tahapan ini diperlukan untuk membantu mengumpulkan jumlah kata dalam artikel berita, sehingga artikel dapat diringkas.

4. Analisis Dataset

Setelah dataset terkumpul, selanjutnya dilakukan proses analisis dataset dan praproses untuk mempersiapkan data pada saat proses *training* data berlangsung. Analisis data mencakup penggabungan dataset, pemeriksaan data *redundant* dan data *null* serta proses pembersihan data berupa penghapusan nama awal platform dan nama kota pada awal artikel berita.

5. Perancangan Skenario Eksperimen

Perancangan skenario eksperimen dilakukan untuk menentukan eksperimen apa saja yang akan dilakukan saat penelitian. Perancangan ini mencakup identifikasi variabel-variabel yang akan diuji, pemilihan dataset, serta penentuan metode dan teknik yang akan digunakan. Selain itu, pada tahap ini juga dilakukan penyusunan prosedur langkah demi langkah yang akan diikuti selama eksperimen, termasuk penentuan metrik evaluasi yang akan digunakan untuk mengukur keberhasilan setiap eksperimen.

6. Eksperimen

Eksperimen akan melakukan implementasi dari tahapan yang telah direncanakan pada skenario eksperimen yang telah dibuat sebelumnya. Setiap eksperimen dilakukan dengan tujuan untuk menguji hipotesis atau mendapatkan data yang diperlukan untuk menjawab pertanyaan penelitian. Hasil dari setiap eksperimen dicatat secara detail untuk di analisis lebih lanjut. Dalam konteks penelitian ini, eksperimen dapat mencakup tanpa proses *fine-tuning* dan *fine-tuning* model *Transformer* seperti T5 dan Pegasus, baik pada dataset Liputan6 dan XL-Sum yang telah dipraproses maupun yang belum, dengan tujuan untuk memperoleh hasil peringkasan berita yang optimal.

7. Evaluasi dan Analisis Hasil

Setelah semua eksperimen selesai dilaksanakan, tahap berikutnya adalah melakukan evaluasi dan analisis terhadap hasil yang diperoleh. Pada tahap ini, setiap hasil eksperimen dianalisis secara mendalam untuk menilai kinerja model yang telah di ujicobakan. Evaluasi dilakukan menggunakan metrik ROUGE, yang berfungsi untuk mengukur kualitas hasil ringkasan yang dihasilkan oleh model. Hasil evaluasi kemudian dianalisis untuk mengidentifikasi pola, tren, dan *insight* yang bisa diperoleh, serta untuk menentukan model mana yang memberikan hasil terbaik. Analisis ini juga melibatkan perbandingan antara berbagai skenario eksperimen, termasuk pengaruh dari proses praproses terhadap performa model.

8. Pembuatan *Interface*

Tahap ini berfokus pada pengembangan sebuah aplikasi web sederhana yang bertujuan untuk memvisualisasikan dan mempresentasikan hasil dari penelitian ini. Aplikasi tersebut dibuat menggunakan *framework* Streamlit yang memungkinkan pembuatan aplikasi web secara cepat dan interaktif. Sehingga pengguna dapat dengan mudah memasukkan teks berita dan mendapatkan hasil ringkasan yang dihasilkan oleh model.

9. Penarikan Kesimpulan

Tahap penarikan kesimpulan merupakan tahap akhir dalam proses penelitian ini. Pada tahap ini, peneliti menyusun kesimpulan berdasarkan hasil analisis dan evaluasi yang telah dilakukan sebelumnya. Kesimpulan yang ditarik harus mampu menjawab tujuan penelitian dan memberikan gambaran

yang jelas tentang kinerja sistem yang telah dikembangkan. Kesimpulan yang baik harus disertai dengan rekomendasi untuk penelitian lebih lanjut, termasuk saran untuk perbaikan dan pengembangan sistem di masa depan.

3.2 Alat dan Bahan Penelitian

Bagian ini menjelaskan secara detail alat dan bahan yang digunakan penulis dalam melakukan penelitian.

3.2.1 Alat Penelitian

Alat yang digunakan oleh penulis untuk penelitian berupa:

1. Perangkat keras, yang terdiri dari:
 - Processor AMD Ryzen 5 5500U with Radeon Graphics
 - AMD Radeon(TM) Graphics
 - SSD 512 GB
 - RAM 16 GB
2. Perangkat lunak, yang terdiri dari:
 - Sistem Operasi Windows 11 Home Single Language
 - Google Colab,
 - Kaggle
 - Paperspace
 - Python
 - Visual Studio Code
 - Google Chrome Web Browser

3.2.2 Bahan Penelitian

Bahan yang diperlukan untuk melakukan penelitian yaitu dataset yang berisi artikel berita bahasa Indonesia Selain dataset, bahan yang digunakan dalam penelitian ini yaitu jurnal, buku serta dokumentasi *library* dari Hugging Face. Berikut penjelasan dari setiap bahan yang digunakan dalam penelitian.

1. Dataset yang digunakan dalam penelitian ini adalah dataset Liputan6 yang berasal dari portal berita liputan6.com (Koto et al., 2020) dan dataset XL-Sum yang berasal dari portal berita bbc.com yang terdiri dari berbagai bahasa termasuk bahasa Indonesia (Hasan et al., 2021). Atribut yang terdapat pada

dataset Liputan6 adalah *id*, *url*, *clean_artcile*, *clean_summary* dan *extractive_summary*. Sedangkan pada dataset XL-Sum atributnya adalah *id*, *url*, *title*, *text* dan *summary*. Penggunaan kedua dataset ini memungkinkan penelitian untuk mengevaluasi kinerja model pada dua jenis dataset yang berbeda sehingga memberikan wawasan yang lebih komprehensif tentang kemampuan model dalam peringkasan berita otomatis.

2. Jurnal dan buku digunakan sebagai sumber bacaan yang mendukung kajian teori dan metode penelitian untuk peringkasan berita otomatis menggunakan arsitektur *Transformer* model T5 dan Pegasus. Jurnal dan buku yang relevan menyediakan pemahaman tentang perkembangan terbaru dalam bidang peringkasan teks otomatis, teknik *fine-tuning pre-trained models*, serta aplikasi model *Transformer* dalam tugas-tugas pemrosesan bahasa alami (NLP).
3. Dokumentasi *library* digunakan untuk membangun model peringkasan berita otomatis, serta untuk memfasilitasi proses analisis dan evaluasi. *Library-library* seperti *Transformer* dari Hugging Face digunakan untuk mengakses model-model T5 dan Pegasus yang telah di *pre-trained*, serta untuk menerapkan *fine-tuning* pada dataset yang digunakan. Selain itu, *library* PyTorch digunakan untuk menangani proses komputasi dan pelatihan model. Dokumentasi dari *library-library* ini memainkan peran penting dalam memastikan bahwa implementasi dilakukan dengan benar, serta untuk mengatasi masalah teknis yang mungkin muncul selama proses eksperimen. Dokumentasi ini juga berguna untuk mereplikasi hasil penelitian, memastikan bahwa semua langkah yang diambil dapat dipahami dan diikuti oleh peneliti lain di masa depan.