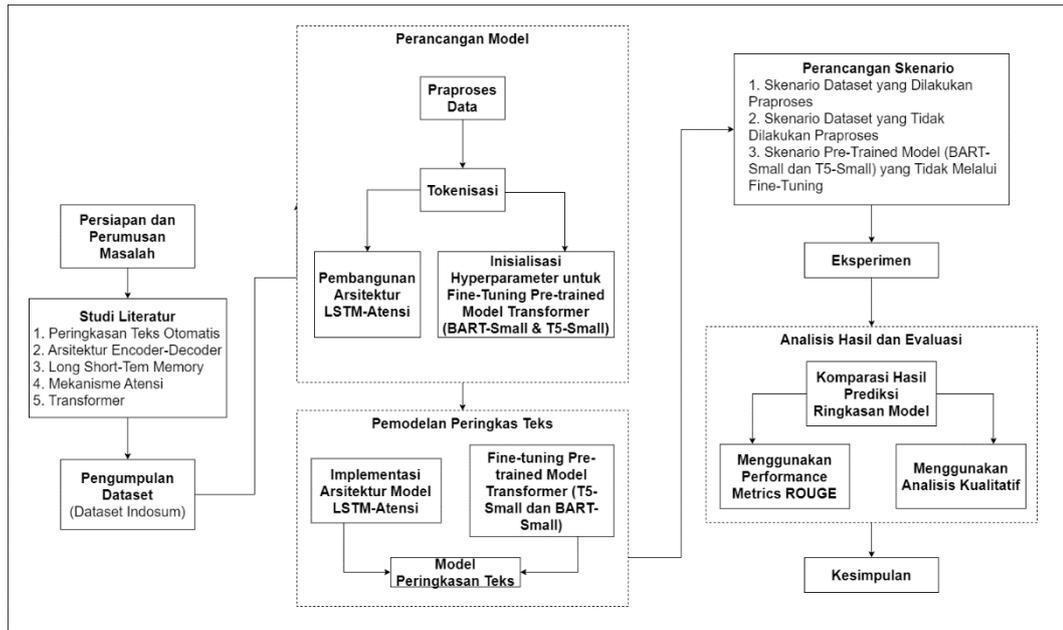


BAB III METODOLOGI PENELITIAN

3.1. Desain Penelitian

Desain penelitian merupakan kerangka kerja yang digunakan oleh penulis selama melakukan penelitian untuk memberikan gambaran terkait kegiatan yang dilakukan selama penelitian. Desain penelitian ditunjukkan pada Gambar 3.1.



Gambar 3.1 Desain Penelitian

Kerangka kerja penelitian yang ditunjukkan pada Gambar 3.1 memiliki langkah-langkah sebagai berikut.

1. Persiapan dan Perumusan Masalah

Pada tahap ini dilakukan proses identifikasi masalah yang selanjutnya akan dilanjutkan dengan perumusan masalah dan pencarian algoritma atau metode untuk menyelesaikan masalah tersebut. Tahap ini meliputi identifikasi latar belakang terkait rumusan masalah serta tujuan dan manfaat yang diperoleh dari penelitian dengan rumusan masalah tersebut. Masalah yang ditemukan dan diangkat menjadi topik penelitian ini adalah terkait peringkasan teks berita berbahasa Indonesia secara otomatis menggunakan *Natural Language Processing*.

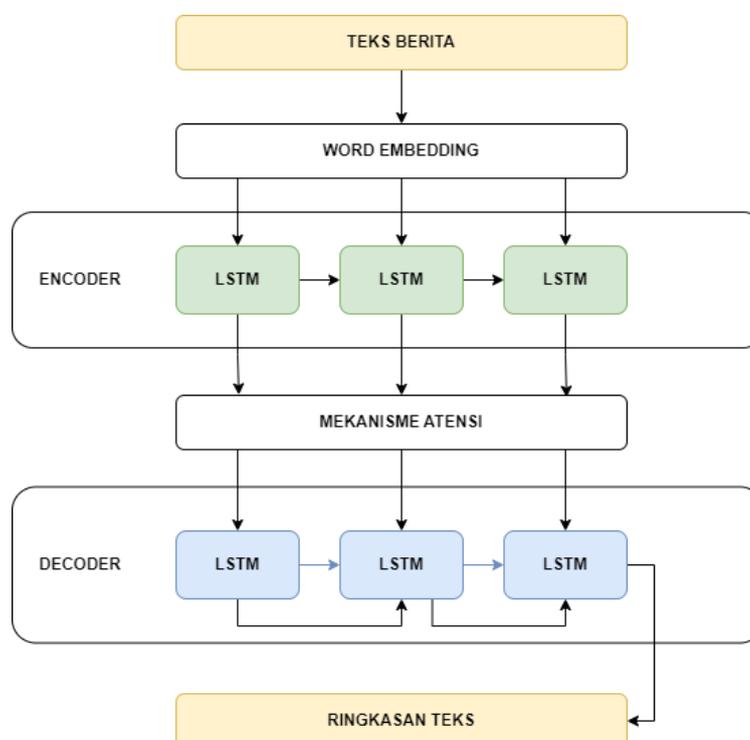
2. Studi Literatur

Pada tahap studi literatur dilakukan dengan mempelajari secara lebih mendalam terkait rumusan masalah dan metode yang akan digunakan selama penelitian. Pada penelitian terkait peringkasan teks berita berbahasa Indonesia secara otomatis menggunakan *Deep Learning* ini, literatur yang dipelajari adalah terkait konsep peringkasan teks, setelah itu mempelajari tentang peringkasan teks secara otomatis menggunakan *Natural Language Processing*, dan juga salah satu algoritma *Deep Learning* yaitu *Long Short Term Memory (LSTM)*. Teori terkait mekanisme atensi, arsitektur *transformer* dan *fine-tuning* juga dipelajari untuk mendukung penelitian. Pemahaman dataset dan hasil-hasil penelitian terdahulu juga dilakukan pada tahap ini.

3. Pengumpulan Data

Data yang digunakan dalam penelitian peringkasan teks berita berbahasa Indonesia secara otomatis ini didapat dari dataset Indosum (Kurniawan & Louvan, 2018) yang berisi sekitar 19.000 berita berbahasa Indonesia. Data utama yang akan digunakan untuk peringkasan teks adalah data artikel berita dan data ringkasan asli dari berita tersebut.

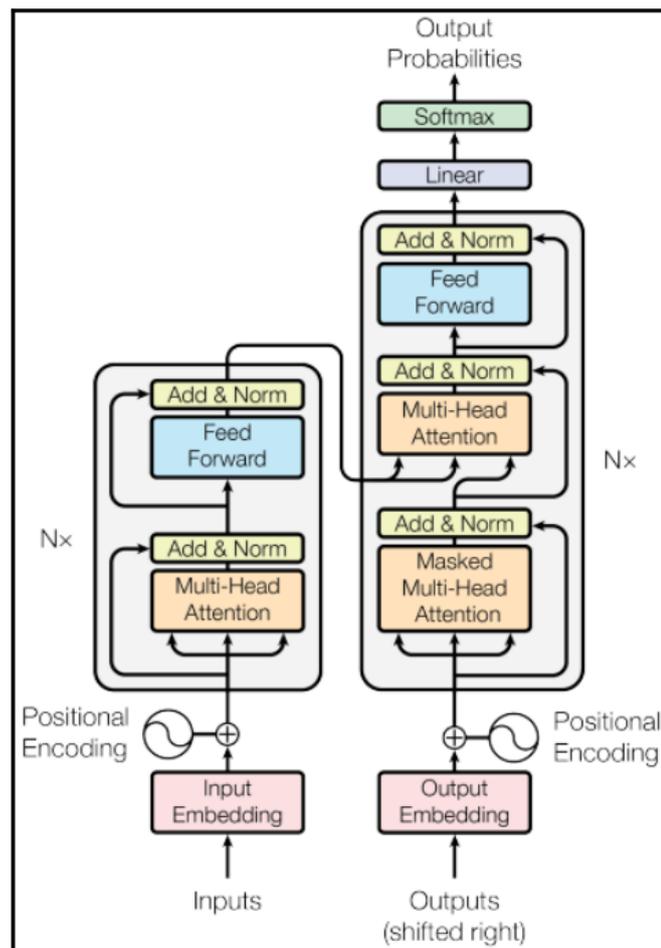
4. Perancangan Model



Gambar 3.2 Arsitektur Model LSTM-Atensi

Pada tahap ini dilakukan perancangan arsitektur model seperti penentuan banyaknya *layer* pada model, lalu penentuan jumlah *unit* setiap *layer*, penggunaan *activation function*, penggunaan banyaknya *epoch* dalam eksperimen, dan sebagainya. Arsitektur model LSTM yang diaplikasikan mekanisme atensi menggunakan *encoder-decoder*, model ini dinamakan model LSTM-Atensi. Pada bagian *encoder*, terdapat dua *layer* LSTM dan di bagian *decoder* terdapat satu *layer* LSTM. Mekanisme atensi akan diimplementasikan dengan menggunakan *output hidden state* dari *encoder* dan *decoder*. Ilustrasi arsitektur model LSTM-Atensi ditunjukkan oleh Gambar 3.2.

Selain itu, untuk *fine-tuning model Transformer* (T5-Small dan BART-Small) juga diperlukan inisialisasi *hyperparameter*. Arsitektur *pre-trained model transformer* pada dasarnya sama seperti arsitektur *transformer* milik (Vaswani dkk., 2017) yang ditunjukkan oleh Gambar 3.3.



Gambar 3.3 Arsitektur *Encoder-Decoder Transformer* (Vaswani dkk., 2017)

5. Pembangunan Model

Setelah perancangan model, akan dilakukan implementasi model LSTM-Atensi dan *fine-tuning pre-trained model* T5-Small dan BART-Small. Tahap ini akan menghasilkan *output* berupa model peringkasan teks sebelum diimplementasikan skenario eksperimen.

6. Perancangan Skenario Eksperimen

Terdapat tiga skenario eksperimen yang akan dilakukan pada penelitian ini, yaitu skenario pelatihan model menggunakan dataset yang dilakukan praproses, skenario pelatihan model menggunakan dataset yang tidak dilakukan praproses, dan skenario penggunaan *pre-trained model* tanpa melalui proses *fine-tuning*. Skenario yang berkaitan dengan dataset akan diimplementasikan pada model LSTM-Atensi dan *pre-trained model* T5-Small dan BART-Small. Sedangkan skenario tanpa *fine-tuning* hanya dilakukan pada *pre-trained model* saja.

7. Eksperimen

Pada tahap eksperimen, semua skenario eksperimen yang telah dibuat sebelumnya akan diimplementasikan pada model yang telah dibuat. Semua hasil dari eksperimen akan dicatat dan kemudian dianalisis pada tahap berikutnya. Pengukuran hasil dari eksperimen ini menggunakan metode *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE).

8. Analisis dan Evaluasi Hasil

Setelah semua skenario eksperimen dilakukan, hasil-hasil dari eksperimen tersebut akan dianalisis dan dievaluasi. Evaluasi dilakukan dengan melakukan komparasi antara hasil prediksi ringkasan yang dihasilkan model LSTM-Atensi dan *Transformer* (BART-Small dan T5-Small) dengan melihat nilai ROUGE dari hasil prediksi ringkasan. Selain itu juga dilakukan analisis kualitatif untuk melihat kualitas kalimat pada hasil prediksi ringkasan.

9. Kesimpulan

Setelah semua proses selesai, dilakukan penarikan kesimpulan dan saran untuk digunakan dalam penelitian selanjutnya.

3.1.1 Alat Penelitian

Alat yang digunakan dalam penelitian ini melibatkan beberapa perangkat keras dan beberapa perangkat lunak yang dispesifikasikan sebagai berikut.

1. Perangkat keras (*Hardware*) menggunakan sebuah laptop dengan
 - a. Processor : Intel(R) Core(TM) i3-10110U CPU
 - b. RAM : 4 GB
 - c. SSD : 512GB
 - d. Monitor : resolusi 1366 x 768
2. Perangkat lunak (*Software*)
 - a. Sistem Operasi Microsoft Windows 11
 - b. Google Chrome Web Browser
 - c. Google Colab
 - d. Kaggle
 - e. Visual Studio Code

3.1.2 Bahan Penelitian

Bahan penelitian yang digunakan dalam penelitian ini adalah sebuah dataset bernama Indosum yang didapatkan dari sebuah artikel (Kurniawan & Louvan, 2018). Dataset ini berisi sekitar 20.000 data berita berbahasa Indonesia yang berasal dari beberapa halaman berita seperti CNN Indonesia, Kumparan, dan lain-lain. Dataset ini berisi 6 kategori yaitu *entertainment*, olahraga, teknologi, *showbiz*, *inspiration*, dan *headline*. Atribut yang terdapat pada dataset ini adalah sebagai berikut.

1. Judul
2. Kategori
3. Isi berita
4. Link URL sumber
5. Ringkasan abstraktif yang dibuat secara manual oleh penutur asli bahasa Indonesia

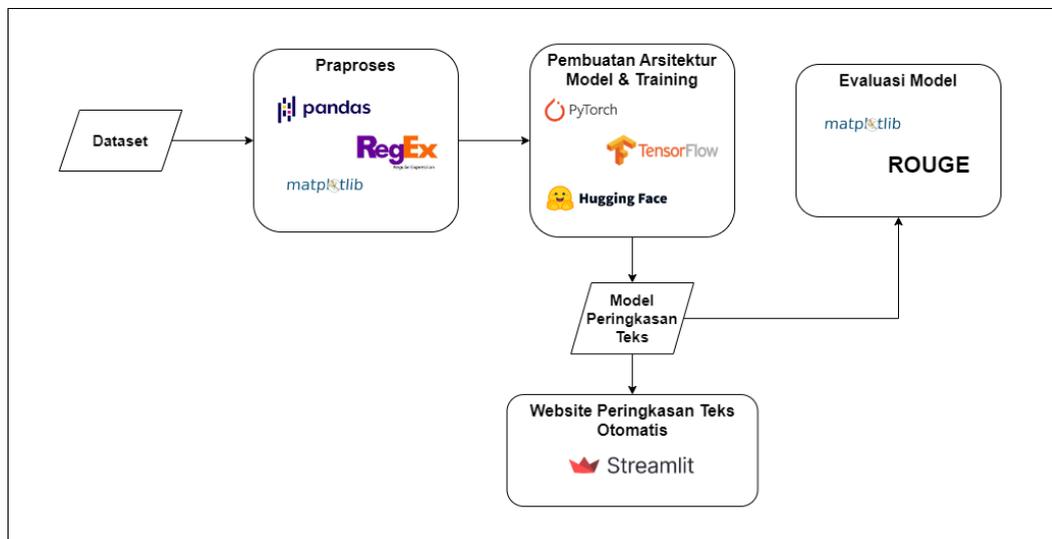
3.2. Desain Sistem

Penelitian ini menggunakan beberapa *library* Python baik untuk tahap praproses sampai tahap pelatihan model. Tahap praproses menggunakan *library* Pandas untuk mengolah dataset, Regex untuk membersihkan data, dan Matplotlib

untuk menampilkan Exploratory Data Analysis (EDA). Sedangkan untuk pembangunan dan pelatihan model, penelitian ini memerlukan *framework Tensorflow* untuk membangun model LSTM-Atensi dan *framework Pytorch* serta *library transformers* milik *Huggingface* yang terdapat *pre-trained model* T5-Small dan BART-Small.

Evaluasi hasil ringkasan dan pelatihan model dilakukan dengan menggunakan Matplotlib untuk menampilkan grafik *loss* dan *library* ROUGE untuk mengukur kemiripan hasil ringkasan yang diprediksi dengan ringkasan asli. Sebuah website juga dibangun untuk mensimulasikan kinerja model dalam melakukan peringkasan teks. Website tersebut dibangun dengan menggunakan *library* Streamlit. Desain sistem ini diilustrasikan oleh Gambar 3.4.

Desain rancangan antarmuka dari halaman website peringkasan teks ditunjukkan oleh Gambar 3.5. Pada halaman website ini akan terdapat bagian *input* teks dan *dropdown* untuk pemilihan model peringkasan. Lalu, hasil prediksi ringkasan yang dihasilkan model akan ditampilkan pada bagian kanan halaman website.



Gambar 3.4 Desain Sistem



Gambar 3.5 Desain Antarmuka Website Peringkasan Teks