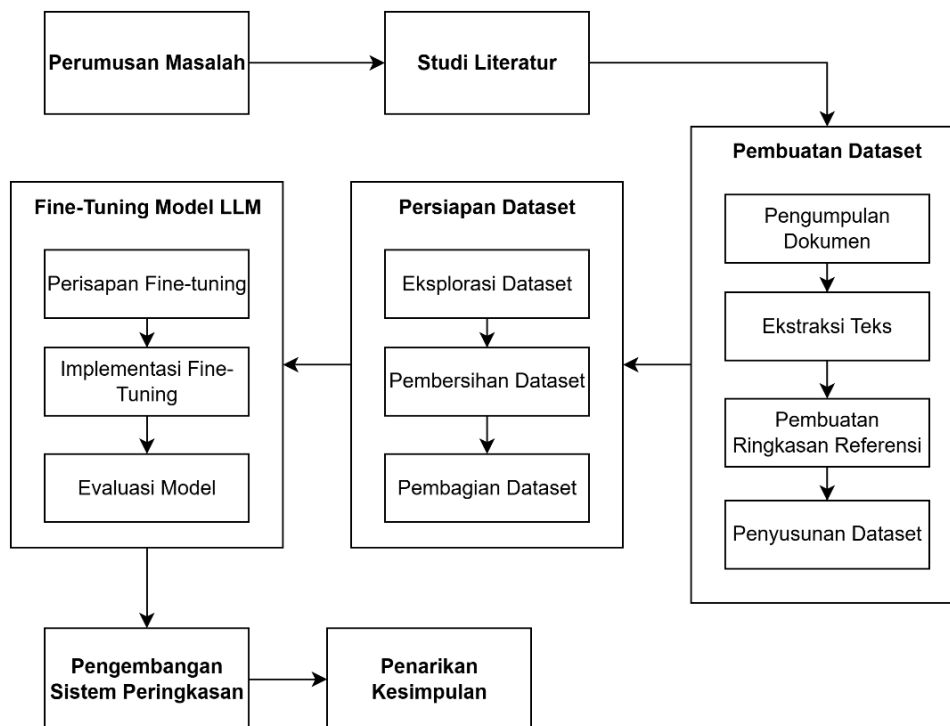


## BAB III

### 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. Perumusan Masalah

Pada tahap awal, peneliti melakukan identifikasi terhadap permasalahan yang ada terkait peringkasan teks dokumen laporan tahunan perusahaan. Permasalahan diformulasikan secara spesifik dengan menetapkan tujuan dan manfaat penelitian, serta menetapkan solusi berbasis teknologi terkini. Fokus penelitian diarahkan pada peringkasan teks dokumen laporan tahunan yang disusun dalam bahasa Inggris, dengan pertimbangan bahwa laporan tersebut

merupakan sumber informasi penunjang analisis saham. Lingkup penelitian dibatasi pada laporan tahunan perusahaan yang terdaftar di Bursa Efek Indonesia (BEI), sehingga sistem peringkasan yang dikembangkan diharapkan relevan dengan kebutuhan dalam proses analisis fundamental dalam konteks pasar saham Indonesia.

## 2. Studi Literatur

Studi literatur bertujuan untuk mengumpulkan, menelaah, dan menganalisis berbagai teori, temuan, serta metode dari penelitian-penelitian terdahulu yang relevan. Kegiatan ini dilakukan melalui sumber-sumber kredibel seperti jurnal ilmiah, buku akademik, prosiding konferensi, serta artikel ilmiah lainnya. Studi literatur membantu peneliti dalam memahami konteks permasalahan, menemukan kesenjangan penelitian (*research gap*), dan merumuskan pendekatan atau metode yang sesuai.

Pada penelitian ini, literatur yang dipelajari adalah terkait konsep laporan tahunan Perusahaan, arsitektur *Transformer*, *large language model*, dan sistem peringkasan teks otomatis. Pemahaman terkait hasil-hasil penelitian terdahulu juga dilakukan pada tahap ini. Studi literatur ini menjadi dasar teoritis dalam menjawab rumusan masalah dan sekaligus panduan dalam memahami metode yang digunakan dalam penelitian.

## 3. Pembuatan Dataset

Dataset dalam penelitian ini dibuat untuk menyediakan referensi pada tugas peringkasan dokumen tahunan perusahaan. Tahapan pembuatannya dilakukan secara sistematis sebagai berikut:

### a. Pengumpulan Dokumen

Laporan tahunan dikumpulkan dari situs resmi Bursa Efek Indonesia (BEI). Dokumen yang dipilih merupakan laporan dalam format PDF yang memuat versi Bahasa Inggris.

### b. Ekstraksi Teks

Proses ekstraksi dilakukan untuk mengambil konten teks dari *file* PDF laporan tahunan. Fokus utama pada bagian naratif berbahasa Inggris. Prioritas diberikan pada *Management Report* dan *Management Discussion*

*and Analysis (MD&A)* karena keduanya memuat informasi inti mengenai kinerja dan arah strategis perusahaan. Selain itu, bagian pendukung seperti *Corporate Governance*, *Risk Management*, *Corporate Social Responsibility (CSR)*, dan *Company Profile* juga disertakan berdasarkan kriteria relevan. Sebuah bagian dianggap relevan apabila mengandung pembahasan naratif mengenai aspek-aspek yang berdampak pada evaluasi perusahaan, seperti strategi, evaluasi kinerja, manajemen risiko, atau prospek masa depan. Elemen non-naratif seperti tabel, grafik, serta halaman kosong atau tidak relevan diabaikan dalam proses ini.

c. Pembuatan Ringkasan Referensi

Pembuatan ringkasan referensi (*ground truth*) dilakukan oleh peneliti yang bertindak sebagai anotator tunggal. Fokus utama peringkasan adalah menghasilkan ringkasan yang informatif dan relevan untuk kebutuhan analisis saham. Oleh karena itu, prioritas diberikan pada informasi-informasi krusial seperti: kinerja keuangan dan operasional perusahaan, strategi bisnis dan arah kebijakan, risiko utama beserta upaya mitigasinya yang dapat memengaruhi prospek perusahaan di masa mendatang. Setiap ringkasan disusun secara abstraktif dalam satu paragraf.

d. Penyusunan Dataset

Dataset akhir disusun dalam format *Comma-Separated Values (CSV)* dengan enam atribut utama. Setiap entri mewakili satu teks bagian laporan tahunan dari suatu perusahaan pada tahun tertentu, disertai ringkasan referensi. Struktur atribut yang digunakan ditampilkan pada Tabel 3.1.

Tabel 3. 1 Struktur Atribut Dataset

Atribut	Deskripsi
<i>id</i>	ID unik untuk setiap entri
<i>company</i>	Kode perusahaan
<i>year</i>	Tahun laporan tahunan

<i>section</i>	Nama bagian dalam laporan tahunan (misalnya: MD&A, <i>Management Report</i> , CSR, Governance, dll)
<i>text</i>	Teks sumber hasil ekstraksi dari laporan tahunan
<i>summary</i>	Ringkasan manual yang disusun oleh anotator

#### 4. Persiapan Dataset

Proses ini dilakukan untuk mempersiapkan dataset sebelum digunakan dalam proses *fine-tuning* LLM untuk tugas peringkasan teks laporan tahunan perusahaan. Proses ini terdiri dari tiga langkah utama yang sistematis sebagai berikut:

##### a. Eksplorasi Dataset

Proses ini dilakukan untuk memahami karakteristik data. Kegiatan ini meliputi identifikasi distribusi data, analisis statistik deskriptif, dan pencarian pola atau anomali. Tujuan utamanya adalah untuk mendapatkan wawasan yang akan menjadi dasar dalam menentukan strategi pembersihan dan transformasi data yang efektif.

##### b. Pembersihan Dataset

Berdasarkan hasil eksplorasi, dilakukan proses pembersihan untuk memperbaiki inkonsistensi dan meningkatkan kualitas data. Tindakan pada tahap ini mencakup penanganan data yang tidak sesuai, perbaikan kesalahan, dan standarisasi data.

##### c. Pembagian Dataset

Setelah data bersih dan terstandarisasi, langkah terakhir adalah membaginya menjadi tiga set data terpisah: data latih (*training set*), data validasi (*validation set*), dan data uji (*test set*). Pembagian ini dilakukan untuk memastikan proses *fine-tuning* berjalan efektif dan evaluasi performa model dapat dilakukan secara objektif pada data yang belum pernah dilihat model sebelumnya.

## 5. *Fine-Tuning* Model LLM

Tahapan ini bertujuan untuk mengadaptasi *Large Language Model* (LLM) agar memiliki kemampuan spesifik sesuai dengan tugas peringkasan teks laporan tahunan. Prosesnya dibagi menjadi tiga langkah: persiapan, implementasi, dan evaluasi. Penjelasan setiap langkah proses ini adalah sebagai berikut:

### a. Persiapan *Fine-tuning*

Proses ini dilakukan untuk menyiapkan semua komponen teknis sebelum proses pelatihan dijalankan. Persiapan diawali dengan penyiapan lingkungan komputasi yang memanfaatkan GPU dan instalasi *library* pendukung. Selanjutnya, dilakukan pemilihan model dasar (*base model*) untuk di-*fine-tune*. Pada tahap ini, konstruksi *prompt* juga dilakukan untuk memformat dataset sesuai dengan tugas yang akan dilatih.

### b. Implementasi *Fine-Tuning*

Tahap implementasi merupakan fase eksekusi inti dari proses *fine-tuning*. Pada tahap ini, model dilatih menggunakan teknik *Low-Rank Adaptation* (LoRA) pada data latih. Selama proses pelatihan, data validasi digunakan untuk memantau performa model guna mengukur kemampuannya dalam melakukan generalisasi. Hasil akhir dari keseluruhan proses ini adalah sebuah adapter LoRA yang dapat digunakan saat inferensi model.

### c. Evaluasi Model

Tahap evaluasi dilakukan untuk menilai peningkatan performa model setelah proses *fine-tuning*. Evaluasi dilakukan secara kuantitatif dengan menggunakan *test set* yang tidak pernah digunakan selama proses pelatihan. Dua jenis metrik digunakan pada tahap ini, yaitu ROUGE dan BERTScore.

Metrik ROUGE digunakan untuk mengukur kesamaan leksikal antara ringkasan yang dihasilkan model dan ringkasan referensi. Secara spesifik, pengukuran dilakukan pada ROUGE-1, ROUGE-2, dan ROUGE-L, yang masing-masing merepresentasikan tumpang tindih kata tunggal

(*unigram*), frasa dua kata (*bigram*), dan urutan kata terpanjang (*Longest Common Subsequence*) antara kedua ringkasan.

Sementara itu, BERTScore digunakan untuk mengukur kesamaan semantik dengan memanfaatkan representasi *embedding* dari model bahasa *pre-trained*. Metrik ini menilai sejauh mana makna ringkasan yang dihasilkan model sesuai dengan ringkasan referensi, meskipun penggunaan kata atau struktur kalimat berbeda.

Untuk memvalidasi efektivitas proses *fine-tuning*, skor ROUGE dan BERTScore dari model yang telah di-*fine-tune* dibandingkan secara dengan skor dari *baseline model* pada *test set* yang sama. Perbandingan ini memberikan gambaran kuantitatif mengenai dampak adaptasi yang dilakukan terhadap kualitas ringkasan yang dihasilkan model.

#### 6. Pengembangan sistem peringkasan

Penelitian ini mengembangkan aplikasi peringkasan otomatis berbasis website. Sistem dapat menerima masukan berupa dokumen PDF, kemudian memanfaatkan model LLM yang telah di-*fine-tuning* sebelumnya untuk meringkas bagian-bagian naratif penting yang difokuskan kepada aspek-aspek yang berguna dalam analisis saham.

#### 7. Penarikan Kesimpulan

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

### 3.2 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) i5-10110U CPU
  - b. RAM: 16 GB
  - c. SSD: 512GB
  - d. Monitor: resolusi 1920 x 1080
2. Perangkat lunak (*Software*)

- a. *Operating System*: Windows 11
- b. *Web Browser*: Google Chrome
- c. *Cloud Platform*: Google Colaboratory, Kaggle
- d. *Code Editor*: Visual Studio Code
- e. *UI Framework*: Gradio
- f. *Document Processing Service*: LlamaParse
- g. *Fine-tuning Library*: Unsloth

### 3.3 Bahan Penelitian

Bahan penelitian yang digunakan dalam penelitian ini terdiri dari dua komponen utama. Pertama, dokumen laporan tahunan perusahaan yang diperoleh dari situs resmi Bursa Efek Indonesia (BEI) digunakan sebagai sumber data untuk membangun dataset peringkasan. Kedua, model LLM LLaMA 3.2-3B *Instruct* yang telah dilatih sebelumnya (*pretrained*) digunakan sebagai fondasi model yang akan di-*finetune* untuk tugas peringkasan teks laporan tahunan.