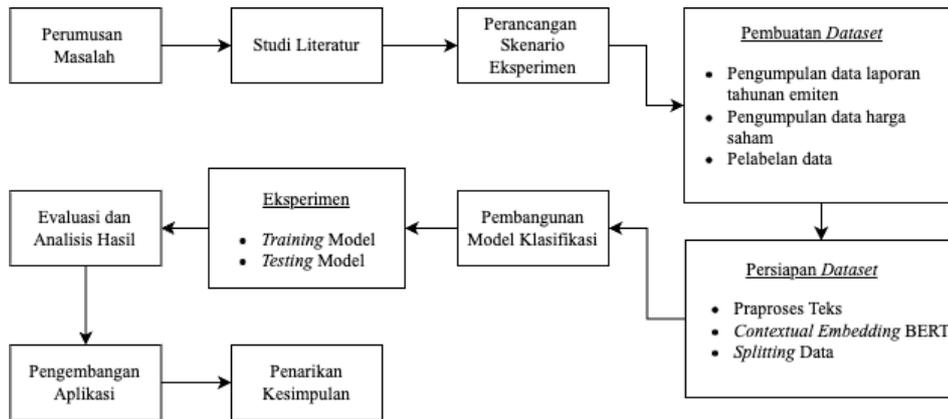


BAB III METODE PENELITIAN

3.1 Desain Penelitian

Desain penelitian merupakan kerangka kerja yang digunakan dalam melakukan penelitian. Penulis memaparkan kerangka kerja mulai dari tahap awal sampai dengan tahap akhir. Desain penelitian ditunjukkan pada Gambar 3. 1.



Gambar 3. 1 Desain Penelitian

Gambar 3.1 yang menunjukkan alur penelitian yang dilakukan dapat dijelaskan sebagai berikut:

1. Perumusan Masalah

Tahap awal dalam penelitian ini adalah merumuskan masalah. Tujuan perumusan masalah adalah untuk mengetahui latar belakang masalah dari tema penelitian. Selain itu, tahap ini juga berfokus untuk menemukan usulan solusi yang mungkin dilakukan untuk permasalahan tersebut.

Pada tahap ini, berbagai permasalahan terkait objek penelitian dikumpulkan sehingga ini akan menjadi fokus atau inti dari penelitian. Di samping itu, tahapan ini juga menentukan aspek-aspek penting yang akan menjadi pembahasan dalam laporan penelitian serta tujuan penelitian yang akan dilakukan.

2. Studi Literatur

Studi literatur merupakan tahapan mengumpulkan dan menganalisis teori-teori pendukung serta penelitian yang relevan melalui berbagai sumber seperti jurnal akademik, buku, artikel, dan sumber lainnya. Studi literatur ini bertujuan untuk menemukan kesenjangan penelitian (*research gap*), solusi yang sudah ada, dan menentukan metode yang tepat digunakan terhadap masalah penelitian.

Pada penelitian ini, literatur yang dipelajari meliputi topik instrumen investasi, saham, pasar saham, *word embedding*, *contextual embedding*, *deep learning*, *transformers*, BERT, *Long Short-Term Memory (LSTM)*, dan topik-topik pendukung lainnya. Studi literatur ini dapat dijadikan sebagai landasan teoritis untuk menjawab pertanyaan-pertanyaan penelitian dalam rumusan masalah. Selain itu, kajian terhadap literatur juga dapat dijadikan sebagai panduan dalam memahami metode yang akan diterapkan dalam penelitian.

3. Perancangan Skenario Eksperimen

Pada tahap ini, skenario eksperimen disusun berdasarkan penggunaan data, metode pemotongan teks (*truncation*), dan pengaturan *hyperparameter*. Secara rinci, skenario eksperimen terdiri dari penggunaan data dengan periode prediksi yang berbeda, penggunaan teks yang diterapkan metode pemotongan teks (*truncation*) berbeda, dan *hyperparameter*. Skenario-skenario ini akan menentukan berbagai eksperimen yang akan dilakukan pada tahap berikutnya.

4. Pembuatan *Dataset*

Tahapan ini terdiri dari langkah-langkah pembuatan *dataset*, langkah-langkah tersebut antara lain:

1) Pengumpulan data laporan tahunan

Data laporan tahunan merupakan kumpulan laporan tahunan emiten yang berisi informasi tekstual dan numerik tentang kondisi perusahaan selama satu tahun. Laporan tahunan diperoleh melalui proses pengunduhan pada masing-masing halaman *website* resmi emiten, yakni pada halaman

bagian keterbukaan informasi atau hubungan investor. Laporan tahunan yang diunduh memiliki periode awal pengumpulan data yang bervariasi, disesuaikan dengan ketersediaan laporan tahunan untuk masing-masing emiten dan periode akhir pengumpulan data laporan tahunan adalah tahun 2020.

2) Pengumpulan data harga saham

Selanjutnya, pengumpulan data harga saham yakni atribut harga penutupan saham yang disesuaikan atau *adjusted close price* (Adj Close). Harga penutupan saham yang disesuaikan adalah harga penutupan saham yang telah disesuaikan dengan perubahan seperti pembagian dividen dan pembagian saham untuk setiap harinya. Data harga saham ini diperoleh melalui proses *data-fetching* menggunakan *library yfinance* dengan awal periode pengambilan data mengikuti batas awal periode pengambilan data laporan tahunan. Sedangkan akhir periode pengambilan data harga saham sampai tahun 2020.

3) Pelabelan data

Label dalam penelitian ini merupakan atribut target untuk prediksi, yang ditentukan menggunakan metode *volatility based threshold*, yaitu pelabelan berdasarkan ambang batas yang dihitung dari volatilitas, atau fluktuasi harga aset dalam jangka waktu tertentu. Tahapan pelabelan meliputi *resampling* data harian menjadi tahunan dengan rata-rata harga penutupan saham yang disesuaikan, pembuatan atribut persentase perubahan harga penutupan dengan jeda waktu 1, 2, dan 3 tahun, serta proses *smoothing* atribut ini menggunakan *moving average* (MA) untuk mengurangi pengaruh nilai ekstrem. Nilai volatilitas kemudian dihitung sebagai standar deviasi dari *moving average*.

5. Persiapan *Dataset*

Dalam tahap ini, dilakukan serangkaian proses untuk mempersiapkan data sebelum diolah oleh model klasifikasi. Persiapan dataset terdiri dari beberapa proses, antara lain praproses teks, *contextual embedding* BERT, dan *splitting* data. Berikut ini adalah detail proses persiapan dataset:

1) Praproses teks

Langkah pertama dalam persiapan *dataset* adalah melakukan praproses teks. Serangkaian prosedur dilakukan pada teks laporan tahunan emiten dengan tujuan mengubah data teks yang tidak terstruktur menjadi terstruktur sehingga dapat diolah dengan mudah oleh model klasifikasi. Beberapa prosedur yang dilakukan dalam praproses teks, antara lain *case folding*, *special character remover*, *multispace remover*, dan *stopword remover*.

2) *Contextual embedding* BERT

Selanjutnya, *contextual embedding* BERT yakni tahap yang bertujuan untuk mengubah teks menjadi representasi vektor numerik. Seperti yang telah diketahui bahwa model klasifikasi hanya dapat menerima input numerik. Oleh karena itu, proses untuk mengubah teks menjadi representasi numerik penting dilakukan. Tahap ini terdiri dari dua proses utama, yaitu pemotongan teks (*truncation*) dan *sliding window embedding*. Dalam tahapan *sliding window embedding*, terdapat beberapa bagian yaitu tokenisasi (*tokenization*), pemecahan teks (*chunking*), *padding*, vektorisasi (*vectorization*), dan *mean pooling*.

3) *Splitting* data

Pada tahapan ini hasil vektor representasi numerik dilakukan pembagian (*splitting*) data untuk pelatihan (*training*) dan pengujian (*testing*). Data pelatihan digunakan dalam proses pelatihan, sedangkan data pengujian digunakan dalam proses evaluasi. Pembagian data ini penting dilakukan agar skenario eksperimen dan evaluasi dapat dilakukan.

6. Pembangunan Model Klasifikasi

Setelah data dipersiapkan pada tahapan sebelumnya, data kemudian dapat digunakan sebagai input model klasifikasi. Pada tahap ini, model klasifikasi dibangun dalam lingkungan pengembangan dengan bahasa pemrograman Python. Terdapat dua model klasifikasi yang dibangun yaitu model *baseline* dan model LSTM.

7. Eksperimen

Pada tahap ini, skenario yang telah disusun sebelumnya diterapkan pada model klasifikasi yang telah dibangun pada tahap sebelumnya. Berdasarkan eksperimen, maka akan diperoleh hasil berupa hasil latih dan prediksi untuk setiap skenario eksperimen. Lebih lanjut, tahapan eksperimen ini melibatkan dua proses yaitu pelatihan (*training*) dan pengujian (*testing*).

8. Evaluasi dan Analisis Hasil

Model yang dihasilkan pada proses eksperimen dilakukan evaluasi berdasarkan tingkat kinerjanya sehingga didapatkan model yang terbaik. Evaluasi model ini dilakukan menggunakan metrik evaluasi yang telah didefinisikan pada kajian teori.

9. Pengembangan Aplikasi

Setelah melakukan evaluasi dan analisis hasil, maka selanjutnya dilakukan pengembangan aplikasi untuk mengimplementasikan model prediksi yang telah dibuat. Aplikasi ini akan dikembangkan menggunakan *framework web* interaktif seperti streamlit. Terdapat beberapa fungsional utama yang pada aplikasi yang dibutuhkan dalam implementasi model prediksi.

10. Penarikan Kesimpulan

Berdasarkan hasil analisis pada tahap evaluasi, selanjutnya dilakukan proses penarikan kesimpulan yang bertujuan untuk menilai kinerja model yang telah dibangun sebagai bagian akhir dari penelitian ini.

3.2 Alat dan Bahan Penelitian

Alat dan bahan penelitian merupakan komponen penting dalam suatu penelitian. Alat mencakup seluruh instrumen, perangkat, atau peralatan yang diperlukan dalam penelitian. Bahan mencakup semua materi atau substansi yang digunakan dalam eksperimen.

3.2.1 Alat Penelitian

Dalam penelitian ini, alat yang digunakan yaitu perangkat keras (*hardware*), perangkat lunak (*software*), serta *library* dan *framework*.

1. Perangkat keras (*hardware*)
 - Apple Silicon M2 CPU 8-Core GPU 10-Core
 - Metal Performance Shaders (MPS) backend
 - SSD 512 GB
 - RAM 8 GB
2. Perangkat lunak (*software*)
 - Sistem Operasi MacOS Sonoma 14.5
 - Python
 - Anaconda Environment
 - Jupyter Lab
 - Visual Studio Code
3. *Library* dan *framework*
 - Keras
 - PyTorch
 - Transformers
 - LangChain
 - PyPDF
 - pandas, numpy, matplotlib, seaborn, joblib, datetime
 - nltk
 - yfinance
 - scikit-learn
 - gensim

3.2.2 Bahan Penelitian

Bahan penelitian yang dibutuhkan yaitu data laporan tahunan emiten dan data numerik saham. Data laporan tahunan diperoleh melalui *website* resmi setiap emiten pada bagian keterbukaan informasi. Sedangkan data numerik

saham diperoleh melalui proses *data-fetching* menggunakan *library yfinance* dan dijalankan dalam program dengan bahasa pemrograman Python. Daftar laporan tahunan emiten yang digunakan ditunjukkan pada Tabel 3. 1.

Tabel 3. 1 Daftar Laporan Tahunan Emiten

No	Emiten	Periode
1	ADHI	2014-2020
2	ADRO	2010-2020
3	ASII	2010-2020
4	ASRI	2010-2020
5	BABP	2013-2020
6	BBCA	2006-2020
7	BBNI	2005-2020
8	BBRI	2011-2020
9	BBYB	2016-2020
10	BMRI	2005-2020
11	BNII	2011-2020
12	BSDE	2016-2020
13	BUMI	2008-2020
14	CTRA	2016-2020
15	EMTK	2011-2020
16	FREN	2013-2020
17	ISAT	2006-2020
18	ITMG	2009-2020
19	KINO	2017-2020
20	KLBF	2010-2020
21	LPKR	2015-2020
22	MAYA	2015-2020
23	PLIN	2016-2020
24	PTBA	2013-2020
25	PTPP	2017-2020
26	PWON	2011-2020
27	SIDO	2015-2020
28	SMMA	2014-2020
29	SSIA	2016-2020

No	Emiten	Periode
30	TLKM	2016-2020
31	TOWR	2011-2020
32	UNVR	2005-2020
33	WIKA	2015-2020
34	WSKT	2014-2020