

BAB III

METODOLOGI

3.1 Jenis Penelitian dan Sumber Data

Penelitian ini mengaplikasikan pendekatan kuantitatif dengan metode yang melibatkan pemrosesan angka dan algoritma matematis pada data historis saham PT Astra Internasional Tbk (ASII.JK). Peneliti memilih pendekatan ini untuk mengembangkan dan menerapkan berbagai model sistematis, teori, dan hipotesis yang terkait dengan fenomena yang sedang berlangsung.

Data yang dipakai dalam studi ini termasuk kategori data sekunder yang didapat dari *website* resmi *Yahoo Finance*, berupa data historis harga saham ASII.JK sejak tanggal 31 Oktober 2019 sampai dengan 31 Oktober 2024. Variabel yang dimanfaatkan dalam studi ini diantaranya adalah Harga Pembuka (*Open Price*), Harga Tertinggi (*High Price*), Harga Terendah (*Low Price*), Harga Penutup (*Close Price*), dan Volume [32].

3.2 Perangkat dan Sumber Daya Penelitian

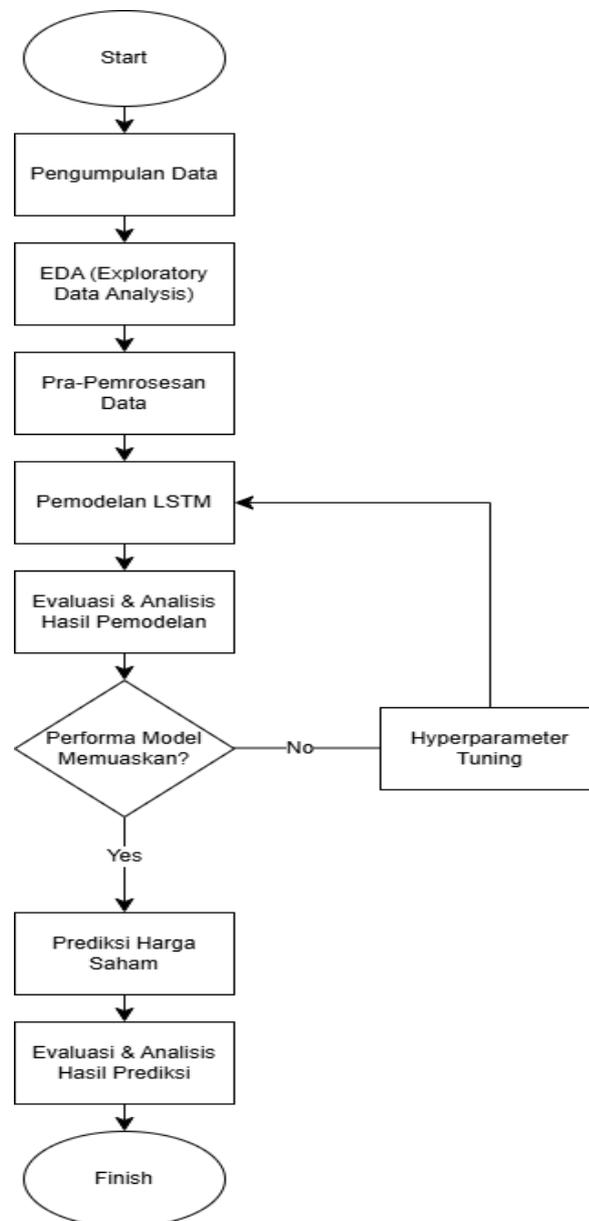
Dalam penelitian ini, berbagai perangkat dan sumber daya digunakan untuk mendukung proses penelitian mulai dari analisis data hingga implementasi model prediksi. Tabel 3 merangkum semua perangkat dan sumber daya yang dimanfaatkan dalam penelitian ini.

Tabel 3 Perangkat dan Sumber Daya

Jenis	Detail
Laptop	VivoBook ASUS X415MA_A416MA dengan prosesor Intel Celeron N4020 CPU @ 1.10 GHz, RAM 4 GB, sistem operasi 64-bit, x64-based processor.
Python	Versi 3.10.0
TensorFlow/Keras	Versi 2.10.0
Scikit-learn	Versi 1.4.2

3.3 Alur Penelitian

Penelitian ini mengawali prosesnya melalui pengumpulan data yang dilanjutkan dengan proses *Exploratory Data Analysis* (EDA), dan pra-pemrosesan data. Kemudian tahap pemodelan akan dilakukan dimana model LSTM akan dilatih sebelum akhirnya di evaluasi dan dianalisis. Setelah performa model dirasa sudah optimal, akan dilakukan prediksi harga saham untuk 7 hari ke depan sejak tanggal 31 Oktober 2024 untuk melihat sejauh mana model mampu memprediksi data baru yang belum pernah diproses sebelumnya. Gambar 12 adalah gambaran mengenai tahap-tahap yang akan dilalui pada penelitian ini.



Gambar 12 Alur Penelitian

3.4 Pengumpulan Data

Proses pengumpulan data dimulai dengan mengakses data historis harga saham ASII.JK di *Yahoo Finance* sejak tanggal 31 Oktober 2019 sampai dengan 31 Oktober 2024 menggunakan *library yfinance*. Penelitian ini berfokus pada penggunaan variabel Harga Pembuka (*Open Price*), Harga Tertinggi (*High Price*), Harga Terendah (*Low Price*), Harga Penutup (*Close Price*), dan Volume sebagai input [32]. Variabel ‘*Close*’ di sini akan berperan ganda, yaitu digunakan juga sebagai target untuk memprediksi harga penutupan saham di hari - hari berikutnya [33].

3.5 *Exploratory Data Analysis* (EDA)

Proses *Exploratory Data Analysis* (EDA) dilakukan untuk memahami karakteristik data harga saham sebelum model *machine learning* diterapkan. Dengan melakukan proses ini, kita dapat mengetahui apakah semua data yang digunakan memang diperlukan, dan apakah fitur yang kita pilih sudah cukup baik untuk dapat digunakan dalam model [34]. Pada penelitian ini, proses EDA yang dilakukan diantaranya adalah menghitung ringkasan statistik data harga saham ASII.JK untuk memahami karakteristik dasar data, memvisualisasikan distribusi data dari setiap fitur untuk mengidentifikasi potensi *outlier*, memvisualisasikan tren harga saham untuk memahami pergerakan harga saham, dan melihat korelasi antar fitur untuk mengidentifikasi hubungan linier antar variabel.

3.6 Pra-Pemrosesan Data

Pada tahap pra-pemrosesan data, data akan diolah agar memiliki format yang sesuai, menghilangkan potensi bias dalam skala fitur, serta membentuk struktur yang sesuai untuk model LSTM yang akan diterapkan. Beberapa langkah yang akan dilakukan dalam tahap pra-pemrosesan data pada penelitian ini adalah :

3.6.1 Normalisasi Data

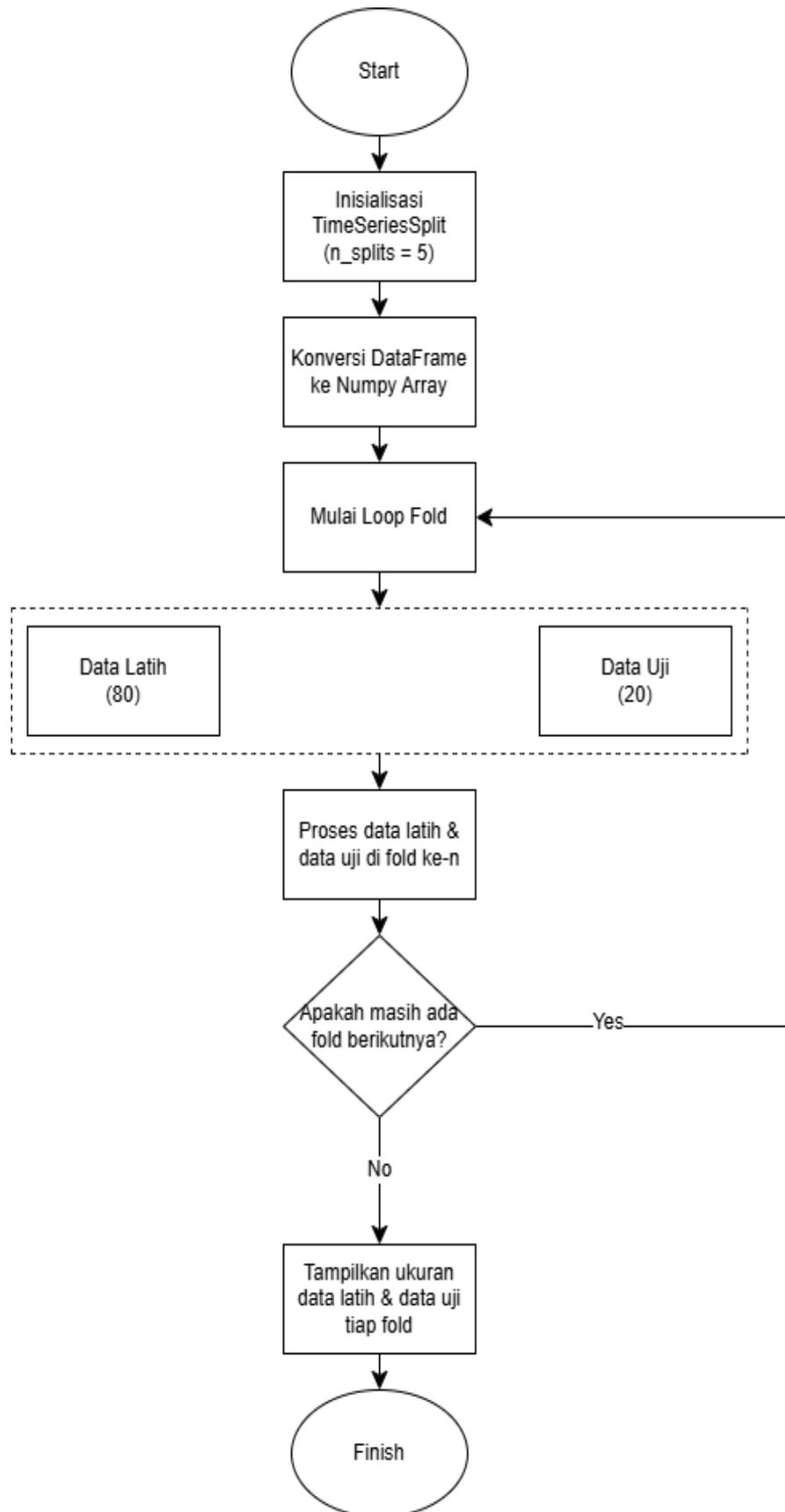
Normalisasi data akan dilakukan menggunakan teknik *MinMax Scaler* untuk menyetarakan skala semua fitur pada rentang 0 hingga 1. Transformasi ini membuat semua fitur memiliki skala yang sebanding tanpa mengubah distribusi relatif data antar fitur.



Gambar 13 Normalisasi Data

3.6.2 Pembagian Data

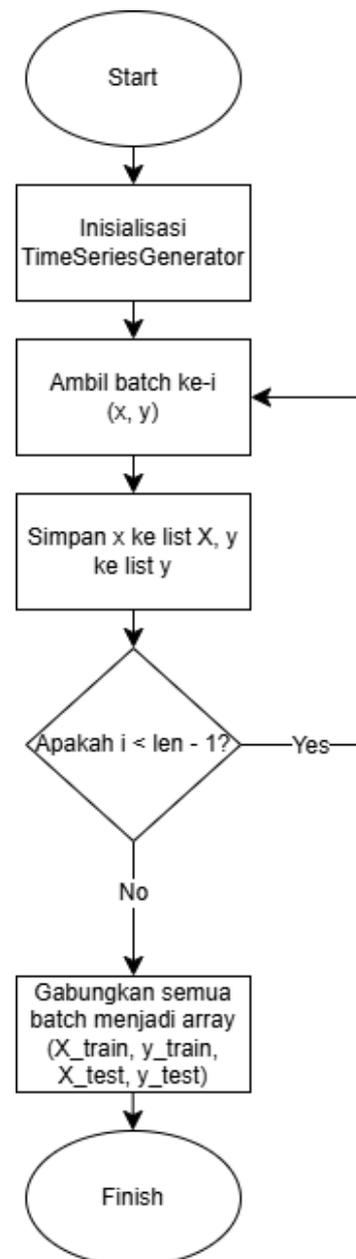
Pembagian data akan dilakukan untuk memisahkan data menjadi data latih (*training set*) dan data uji (*testing set*). Data akan dibagi dengan rasio pembagian sebesar 80 (data latih) : 20 (data uji) menggunakan metode *TimeSeriesSplit Cross Validation* sebanyak 5 *fold*. Data latih akan digunakan dalam tahap pelatihan model, sedangkan data uji akan digunakan dalam tahap pengujian untuk mengevaluasi performa model dalam memprediksi harga saham ASII.JK.



Gambar 14 Pembagian Data

3.6.3 Pembentukan *Sequence* dan Target

Data perlu diubah ke dalam bentuk *sequence* (urutan waktu) karena LSTM hanya bisa memproses data dalam *sequence*. Tujuannya adalah agar model dapat mempelajari pola dari data historis harga saham secara berurutan. Proses pembentukan *sequence* akan dilakukan menggunakan *TimeSeriesGenerator* yang akan secara otomatis membagi data menjadi *sequence* sepanjang n langkah waktu tertentu dan menentukan target pada langkah ke $(n+1)$.



Gambar 15 Pembentukan *Sequence* dan Target

3.7 Pemodelan LSTM

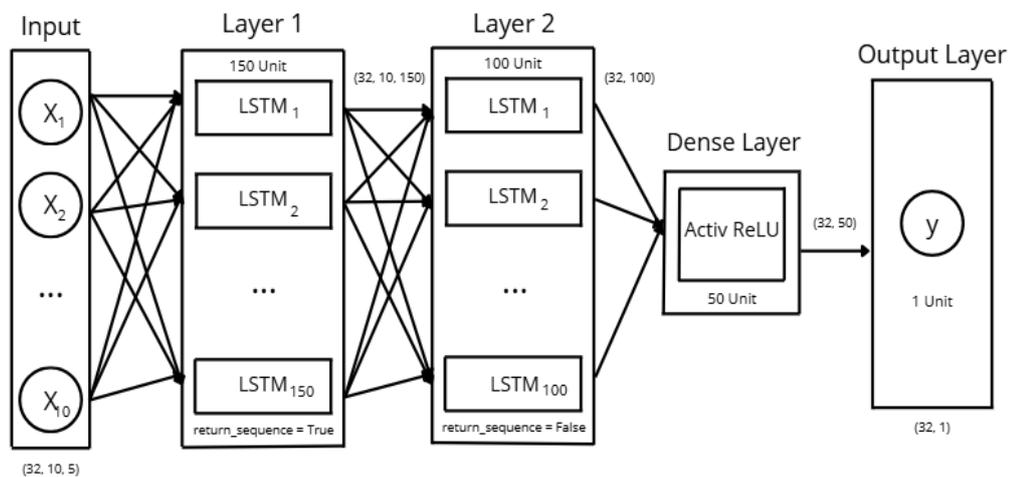
Model LSTM akan dibangun menggunakan *Sequential* API dari Keras. Data yang akan digunakan adalah data yang telah melewati tahap pra-pemrosesan data sebelumnya. Beberapa bagian dari proses pemodelan LSTM adalah :

3.7.1 Arsitektur Model

Struktur arsitektur yang akan digunakan meliputi 2 *Layer* LSTM dan 2 *Dense Layer* dengan rincian sebagai berikut :

- **LSTM Layer Pertama**, terdiri dari 150 unit neuron, *return_sequence=True*, dan *Dropout* sebesar 10%. Layer ini menggunakan fungsi aktivasi *tanh* secara *default*.
- **LSTM Layer Kedua**, terdiri dari 100 unit neuron, *return_sequence=False*, dan *Dropout* sebesar 10%.
- **Dense Layer Pertama**, terdiri dari 50 neuron dengan fungsi aktivasi ReLU.
- **Dense Layer Kedua (*Output Layer*)**, terdiri dari 1 neuron tanpa fungsi aktivasi.

Visualisasi arsitektur model LSTM yang digunakan, ditunjukkan pada Gambar 16.



Gambar 16 Arsitektur Model LSTM yang Digunakan

Gambar 16 menggambarkan alur data dari *input time series* hingga menghasilkan *output* berupa prediksi harga saham pada *output layer*.

3.7.2 Kompilasi dan Pelatihan

Model akan dikompilasi menggunakan fungsi *loss Mean Squared Error* (MSE) dan menggunakan Adam sebagai *optimizer* [5] dengan *learning rate* 0.001. Model akan dilatih menggunakan data latih selama 400 *epoch* dengan *batch size* sebesar 32. Selama proses pelatihan, model akan divalidasi menggunakan data uji untuk memantau performa model.

3.7.3 Visualisasi Loss

Visualisasi grafik *training loss* dan *validation loss* terhadap jumlah *epoch* juga akan ditampilkan untuk mengetahui performa pelatihan model. *Training loss* akan merepresentasikan kesalahan yang terjadi ketika model memprediksi nilai target dari data pelatihan, sedangkan *validation loss* akan menunjukkan kesalahan prediksi model terhadap data validasi yang tidak terlihat selama pelatihan. Dengan grafik ini, dapat diketahui apakah model mengalami *overfitting*, *underfitting*, atau telah mencapai kondisi pelatihan yang optimal.

3.8 Evaluasi & Analisis Hasil Pemodelan

Evaluasi & analisis dari hasil proses pemodelan akan dilakukan untuk menilai performa model dalam memprediksi harga saham. Hasil perhitungan metrik evaluasi utama berupa MAPE, RMSE, dan MAE dianalisis untuk menilai kinerja model. Jika hasil perhitungan metrik evaluasi menunjukkan nilai yang memuaskan, maka model dapat dinyatakan layak untuk digunakan dalam memprediksi harga saham ASII.JK dan dapat digunakan untuk analisis lebih lanjut. Namun jika hasil evaluasi menunjukkan bahwa performa model masih kurang optimal, maka *hyperparameter tuning* akan dilakukan untuk mengoptimalkan model dan meningkatkan hasil prediksi. Jumlah parameter akan diatur ulang dan model akan kembali di latih dan diuji hingga menghasilkan performa yang sesuai dengan yang diharapkan.

3.9 Prediksi Harga Saham

Untuk melihat sejauh mana model mampu menggeneralisasi dan memprediksi data yang benar-benar belum pernah dilihat sebelumnya, dilakukan prediksi harga

saham berdasarkan data historis terbaru yang tersedia. Proses ini dilakukan dengan mengambil n data terakhir dari data historis, mengubahnya ke dalam bentuk *sequence* untuk dijadikan input, dan memberikannya ke model LSTM untuk menghasilkan prediksi harga saham selama 7 hari ke depan sejak tanggal 31 Oktober 2024. Kemudian hasil prediksi akan diubah ke skala aslinya agar nilainya dapat dibandingkan langsung dengan harga yang sebenarnya melalui tabel perbandingan. Hasil perbandingan akan divisualisasikan melalui grafik perbandingan yang menyajikan harga aktual dan harga prediksi. Proses ini dilakukan untuk menilai kemampuan model dalam memprediksi harga saham ASII.JK dalam konteks dunia nyata.

3.10 Evaluasi & Analisis Hasil Prediksi

Hasil perbandingan antara harga aktual dan harga prediksi harga saham selama 7 hari setelah tanggal 31 Oktober 2024 akan dinilai performanya melalui perhitungan metrik evaluasi MAPE, RMSE, dan MAE. Hasil perhitungan akan dievaluasi dan dianalisis untuk menilai keandalan model dalam menghadapi data baru yang belum pernah dilihat sebelumnya. Jika hasil menunjukkan kesalahan prediksi yang rendah, maka model LSTM dapat dianggap efektif dan layak digunakan sebagai alat bantu dalam proses prediksi harga saham. Namun sebaliknya, jika hasil menunjukkan kesalahan yang tinggi, maka perlu dilakukan *hyperparameter tuning* dan pelatihan ulang pada model LSTM.