

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Penelitian ini bertujuan untuk mengembangkan model prediksi saham yang lebih akurat dan adaptif dengan memanfaatkan metode Empirical Mode Decomposition (EMD) sebagai tahap *preprocessing* utama dalam pengolahan data. EMD digunakan untuk memecah sinyal harga saham yang bersifat non-linear dan non-stationary menjadi beberapa komponen sederhana (Intrinsic Mode Functions atau IMF), sehingga pola tren dan fluktuasi harga dapat dianalisis secara lebih detail. Hasil dekomposisi tersebut kemudian dijadikan masukan bagi model *Long Short-Term Memory* (LSTM) yang dikenal mampu menangkap pola ketergantungan jangka panjang pada data deret waktu (*time series*).

Selanjutnya, algoritma Manta Ray Foraging Optimization (MRFO) diterapkan untuk melakukan optimasi terhadap parameter-parameter penting (hyperparameter tuning) pada model LSTM, dengan tujuan meningkatkan performa prediksi dan mengurangi kesalahan peramalan. Dengan kombinasi ketiga metode tersebut, diharapkan model yang dihasilkan mampu memberikan prediksi harga saham yang lebih stabil, akurat, dan efisien.

Tabel 3. 1. Tabel Data Saham BBCA

Date	Close	High	Low	Open	Volume
2020-10-12	5182.4663 09	5200.16901 2	5160.33793 89281	5169.1 89281	8291850 0
2020-10-13	5182.4663 09	5186.89198 4	5107.22982 1	5169.1 89281	4825050 0
2020-10-14	5222.2983 4	5222.29834	5142.63616 2	5169.1 90221	8311750 0
2020-10-15	5120.5073 24	5204.59517 1	5098.37894 3	5200.1 69495	8206700 0

Date	Close	High	Low	Open	Volume
2020-10-16	5098.3793 95	5124.93345 4	5071.82533 5	5120.5 07777	8905150 0
2020-10-19	5222.2983 4	5222.29834	5054.12263 1	5089.5 28043	5687150 0
Date	Close	High	Low	Open	Volume

Tabel 3.1 menyajikan data historis harga saham Bank Central Asia (BBCA) yang diambil dalam rentang waktu tahun 2020 hingga 2025. Data tersebut mencakup berbagai informasi penting seperti harga pembukaan, harga penutupan, harga tertinggi, harga terendah, serta volume transaksi harian. Rentang waktu lima tahun ini dipilih agar model dapat menangkap dinamika pasar yang terjadi pada berbagai kondisi ekonomi, termasuk masa pandemi, periode pemulihan ekonomi, hingga fluktuasi pasar global.

Berdasarkan data pada Tabel 3.1 tersebut, penelitian ini bertujuan untuk melakukan proses prediksi harga saham BBCA dengan menggunakan dua pendekatan berbeda, yaitu *univariate*. Pendekatan *univariate* hanya menggunakan satu variabel, yaitu harga saham itu sendiri (biasanya harga penutupan), untuk memprediksi nilai di masa mendatang.

3.2 Metode Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan explanatory–prediktif, serta menggunakan desain studi kasus pada tiga saham perbankan LQ45, yaitu BBCA, BBRI, dan BMRI. Pendekatan kuantitatif dipilih karena penelitian ini berfokus pada pengolahan data numerik secara sistematis untuk mengukur kinerja model prediksi berdasarkan nilai akurasi yang diperoleh.

Dalam penelitian ini, data historis harga saham digunakan sebagai dasar untuk membangun dan melatih model prediksi berbasis EMD–MRFO–LSTM. Proses analisis dilakukan dengan tujuan menjelaskan bagaimana kombinasi metode tersebut dapat meningkatkan kemampuan model dalam mengenali pola

pergerakan harga saham yang bersifat non-linier dan non-stasioner, sekaligus menghasilkan prediksi yang lebih akurat. Fokus utama penelitian kuantitatif ini ditunjukkan melalui evaluasi performa model menggunakan berbagai metrik kesalahan serta pengujian statistik untuk memastikan ketepatan dan signifikansi hasil prediksi.

3.1.1 Metode Data Mining

Dalam kerangka kerja data mining, terdapat berbagai framework seperti SEMMA (Sample, Explore, Modify, Model, and Assess), CRISP DM (Cross-Industry Standard Process for Data Mining), KDD (Knowledge Discovery in Database), dan lainnya. Penjelasan serta perbandingan framework-framework tersebut disajikan pada Tabel 3.2 di bawah ini [43-47]

Tabel 3. 2. Data Mining Framework

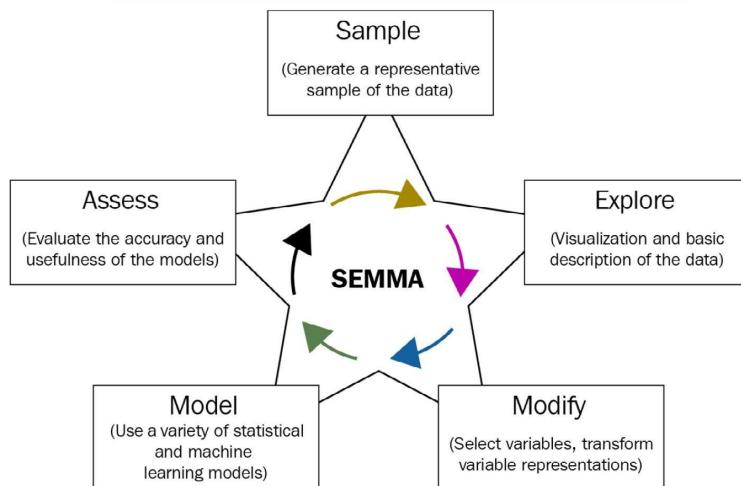
Kriteria	CRISP-DM	KDD	SEMMA
Tahapan	Business understanding, Data understanding, Data preparation, Modeling, Evaluation, Deployment	Data selection, Data transformation, Data cleaning, Modeling, Evaluation, Interpretationn	Sample, Explore, Modify, Model, Assess
Tujuan	Fokus utamanya Adalah pemahaman bisnis yang jelas dan menghasilkan solusi yang relevan dengan tujuan bisnis	Proses KDD menekankan pentingnya pemilihan data yang relevan, <i>preprocessing</i> , dan evaluasi	Memfokuskan pada eksplorasi data dan pemodelan
Kelebihan	Fleksibel, dapat diterapkan di berbagai industri	Pendekatan luas dan mencakup lebih banyak tahapan	Fokus pada eksplorasi data yang mendalam
Kekurangan	Tidak mendalam dalam aspek teknis	Kurang fleksibel dan prosedurnya bisa kaku	Kurang fokus pada tujuan bisnis

Pada hasil dari tabel 3.2, dapat dilihat bahwa SEMMA (Sample, Explore, Modify, Model, Assess) menjadi sebuah kerangka yang kuat dan efektif untuk

digunakan dalam proses analisis data. Hal ini terjadi karena SEMMA memfokuskan pada eksplorasi data secara mendalam pada tahap awal, yang memungkinkan untuk memperoleh wawasan yang lebih baik mengenai karakteristik data yang akan dianalisis. Dengan melakukan eksplorasi data secara intensif, kita dapat mengidentifikasi pola-pola tersembunyi, anomali, dan hubungan antar variabel yang mungkin tidak terlihat pada pandangan pertama. Pemahaman yang lebih baik terhadap

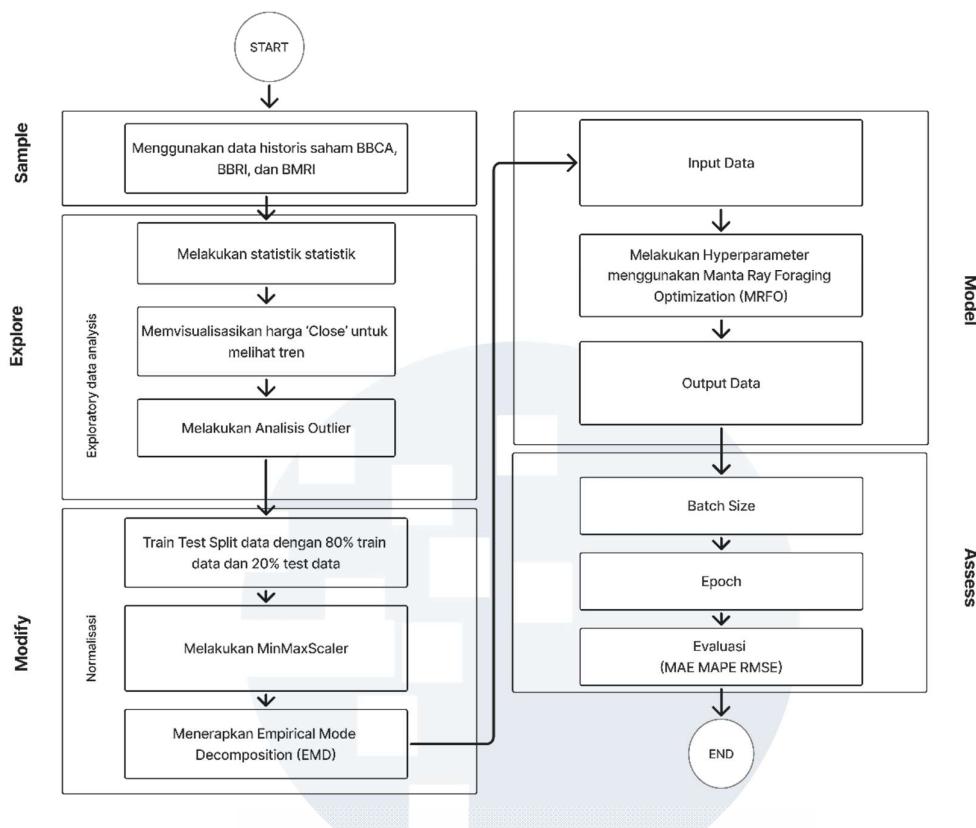
3.1.2 Alur Penelitian

Alur penelitian ini mengadopsi dan mengimplementasikan kerangka kerja SEMMA (Sample, Explore, Modify, Model, Assess) sebagai dasar untuk proses analisis data. Kerangka kerja SEMMA ini dipilih karena kemampuannya dalam memberikan langkah-langkah terstruktur yang dapat memandu peneliti dalam mengelola dan menganalisis data secara sistematis. Dalam konteks penelitian ini, kerangka kerja SEMMA telah disesuaikan dengan kebutuhan dan tujuan spesifik penelitian, agar lebih relevan dengan topik yang sedang dianalisis. Alur penelitian yang diusulkan dapat digambarkan secara jelas melalui diagram yang disajikan pada gambar 3.1 di bawah ini.



Gambar 3. 1 Alur SEMMA [42]

Namun, alur penelitian ini mengalami beberapa penyesuaian sehingga dirumuskan seperti pada gambar 3.2 di bawah ini.



Gambar 3. 2 Alur Penelitian

Penelitian ini tidak melakukan pengujian terhadap berbagai jenis model prediksi lain, seperti *Autoregressive Integrated Moving Average* (ARIMA), *Support Vector Regression* (SVR), atau model *deep learning* lainnya. Pembatasan ini dilakukan untuk menjaga fokus penelitian pada evaluasi efektivitas model hibrida EMD–MRFO–LSTM dibandingkan dengan model *Long Short-Term Memory* (LSTM) tunggal sebagai *baseline*.

Pemilihan *Long Short-Term Memory* (LSTM) sebagai model pembanding didasarkan pada karakteristik data harga saham yang bersifat *non-linear* dan *non-stasioner*, di mana LSTM telah terbukti memiliki kemampuan yang baik dalam memodelkan ketergantungan jangka panjang pada data deret waktu (*time series*). Dengan membatasi perbandingan hanya pada LSTM standar dan model hibrida yang diusulkan, analisis peningkatan performa dapat dilakukan secara lebih terfokus dan objektif, tanpa dipengaruhi oleh perbedaan karakteristik dasar antar model yang tidak sejenis.

Selain itu, pembatasan jumlah model juga bertujuan untuk menghindari kompleksitas eksperimen yang berlebihan serta memastikan kedalaman analisis pada model yang diusulkan, sehingga hasil penelitian dapat lebih terarah dan sesuai dengan tujuan penelitian.

3.1.2.1. Sample

Tahap ini merupakan langkah awal dalam proses penelitian, yaitu pengambilan atau pencuplikan data dari sumber yang relevan. Pada tahap ini, peneliti memperoleh data historis harga saham yang mencakup *Open*, *High*, *Low*, *Close*, dan *Volume* (OHLCV) dari perusahaan X sebagai penyedia data resmi. Data tersebut memuat informasi harga saham untuk tiga emiten perbankan besar, yaitu Bank Central Asia (BBCA), Bank Rakyat Indonesia (BBRI), dan Bank Mandiri (BMRI).

3.1.2.2. Explore

Tahap ini bertujuan untuk memahami karakteristik dan struktur dasar dari data yang telah dikumpulkan dari perusahaan X. Pada tahap ini, peneliti melakukan analisis statistik terhadap data historis saham untuk memperoleh gambaran umum mengenai sebaran dan kecenderungan data. Analisis tersebut meliputi perhitungan nilai rata-rata, median, maksimum, minimum, dan standar deviasi dari setiap variabel harga saham seperti *open*, *high*, *low*, *close*, dan *volume*.

Selain itu, peneliti juga melakukan pemeriksaan terhadap nilai ekstrem dan deteksi *outlier* guna memastikan bahwa data yang akan digunakan dalam tahap pemodelan bersih serta bebas dari nilai-nilai yang dapat mengganggu akurasi model. Tahap eksplorasi ini menjadi langkah penting untuk memahami pola dasar data dan memastikan kualitas dataset sebelum dilanjutkan ke tahap modifikasi dan pemodelan.

3.1.2.3. Modify

Tahap modifikasi merupakan langkah krusial dalam penelitian ini karena berfokus pada pembersihan, transformasi, dan penyiapan data agar dapat digunakan secara optimal pada proses pemodelan. Pada tahap ini, data terlebih dahulu dibagi menjadi *training set* dan *testing set* dengan proporsi

80% untuk pelatihan dan 20% untuk pengujian. Pembagian dilakukan sebelum proses normalisasi untuk mencegah terjadinya *data leakage*, sehingga skala data pada *testing set* tidak mempengaruhi parameter normalisasi yang dihitung dari *training set*. Selanjutnya, data dinormalisasi menggunakan metode *Min–Max Scaling* berdasarkan nilai minimum dan maksimum dari *training set*.

Setelah proses normalisasi, dibentuklah *fixed sliding window* dengan panjang 30 hari historis ($t-30$ hingga $t-1$) sebagai *input* untuk memprediksi harga pada hari ke-31 (t). Pendekatan ini tidak memperbarui *window* menggunakan hasil prediksi model, sehingga konsisten dengan praktik umum pada *baseline Long Short-Term Memory* (LSTM) dan memungkinkan perbandingan yang adil dengan model *Hybrid*.

Tahap modifikasi juga mencakup penerapan *Empirical Mode Decomposition* (EMD), yaitu teknik dekomposisi sinyal deret waktu *non-linear* dan *non-stationary* menjadi beberapa komponen *Intrinsic Mode Functions* (IMFs). Setiap IMF merepresentasikan tingkat osilasi yang berbeda pada data, sehingga mampu memisahkan *noise* dan menonjolkan pola utama yang lebih stabil. Dengan demikian, tahap modifikasi menghasilkan *dataset* yang lebih bersih, terstruktur, dan siap digunakan dalam pembangunan model prediksi harga saham.

3.1.2.4. Model

Pada tahap ini, peneliti membangun dan melatih model prediksi menggunakan *Long Short-Term Memory* (LSTM), sebuah varian dari *Recurrent Neural Network* (RNN) yang dirancang untuk menangkap pola ketergantungan jangka panjang pada data deret waktu yang bersifat *non-linear*. Arsitektur model yang digunakan terdiri dari tiga lapisan LSTM (3-layer LSTM), yang disusun berdasarkan referensi dan temuan pada penelitian terdahulu.

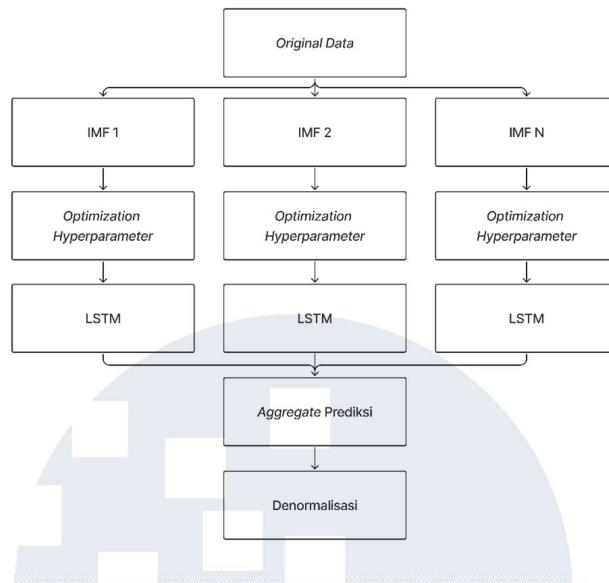
Untuk meningkatkan performa model dan mengurangi ketergantungan pada pemilihan *hyperparameter* secara manual, penelitian ini menerapkan

algoritma *Manta Ray Foraging Optimization* (MRFO) sebagai metode optimasi hyperparameter. MRFO digunakan untuk mencari kombinasi parameter yang menghasilkan nilai validation loss terendah. Terdapat enam hyperparameter utama yang dioptimasi, yaitu jumlah neuron pada tiga lapisan LSTM, nilai dropout, learning rate, serta batch size. Rentang pencarian hyperparameter disajikan pada Tabel 3.3 berikut:

Tabel 3.3 Parameter *Hyperparameter* MRFO

Hyperparameter	Rentang Pencarian
Neuron Layer 1	32 – 128
Neuron Layer 2	32 – 128
Neuron Layer 3	16 – 64
Dropout	0.1 – 0.5
Learning Rate	0.0001 – 0.001
Batch Size	16 – 128

Penggunaan MRFO dalam optimasi memungkinkan pemilihan konfigurasi model yang lebih adaptif terhadap karakteristik setiap komponen IMF hasil dekomposisi EMD. Sehingga, kombinasi antara EMD, MRFO, dan LSTM sebagai model *Hybrid* diharapkan mampu menghasilkan prediksi harga saham yang lebih stabil, sensitif terhadap dinamika pasar, serta memiliki tingkat akurasi yang lebih tinggi dibandingkan model LSTM standar. Berikut ini disajikan *workflow* pemodelan *Hybrid* yang diusulkan dalam penelitian ini.



Gambar 3. 3 *Workflow Model Hybrid*

3.1.2.5. Assess

Tahap terakhir dalam proses SEMMA adalah evaluasi (assess), yang bertujuan untuk menilai sejauh mana model yang telah dibangun mampu memberikan hasil prediksi yang akurat terhadap data harga saham. Pada tahap ini, model *Hybrid* EMD–MRFO–LSTM yang telah dilatih dievaluasi menggunakan data uji (testing data) yang tidak disertakan dalam proses pelatihan, dengan tujuan untuk mengukur kemampuan generalisasi model terhadap data baru.

Hasil prediksi kemudian dibandingkan dengan data aktual untuk menilai tingkat kesalahan atau deviasi antara nilai prediksi dan nilai sebenarnya. Evaluasi dilakukan menggunakan beberapa metrik statistik yaitu Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan Mean Absolute Percentage Error (MAPE). Ketiga metrik ini digunakan untuk mengukur performa model dari sisi ketepatan, keakuratan, dan stabilitas hasil prediksi. Semakin kecil nilai kesalahan yang dihasilkan dari ketiga metrik tersebut, maka semakin baik kinerja model dalam memprediksi harga saham BBCA, BBRI, dan BMRI selama periode penelitian.

3.3 Teknik Pengumpulan Data

Teknik pengumpulan data pada penelitian ini menggunakan metode dokumentasi, yaitu dengan mengumpulkan data sekunder yang diperoleh dari perusahaan X sebagai penyedia informasi keuangan yang terpercaya. Data yang dikumpulkan berupa catatan historis harga saham dari tiga emiten perbankan besar, yaitu BBCA, BBRI, dan BMRI, dalam rentang waktu tahun 2020 hingga 2025. Data tersebut meliputi beberapa variabel penting seperti harga pembukaan, harga penutupan, harga tertinggi, harga terendah, dan volume perdagangan. Pengumpulan data dilakukan secara daring melalui portal resmi perusahaan X dan kemudian disesuaikan dengan kebutuhan penelitian untuk tahap analisis dan pemodelan prediksi saham.

3.4 Teknik Analisis Data

Ada berbagai tools yang berguna untuk memilih teknik analisis data, dengan Jupyter Notebook yang paling sering digunakan, namun penelitian ini menggunakan *tools* dari visual studio code untuk Teknik analisis data. Perbandingan antara kedua *tools* ini dapat dilihat pada tabel 3.4 berikut

Tabel 3.4 Perbandingan *Tools*

Kriteria	Jupyter Notebook	Visual Studio Code
Interface	Berbasis web dengan notebook interaktif	Berbasis aplikasi desktop dengan tampilan yang dapat disesuaikan melalui ekstensi.
Kemudahan Penggunaan	Mudah digunakan untuk pemula dalam eksplorasi data dan analisis sederhana	Membutuhkan konfigurasi awal tetapi lebih fleksibel untuk pengembang berpengalaman
Kolaborasi	Ideal untuk berbagi dan dokumentasi dengan format yang mendukung literasi pemrograman	Lebih kuat untuk kolaborasi berbasis kode melalui Git, integrasi langsung dengan sistem

Kriteria	Jupyter Notebook	Visual Studio Code
		kontrol versi, dan kemampuan debugging tingkat lanjut.
Integrasi Ekosistem	Mendukung pustaka Python seperti Pandas, Matplotlib, dan Scikit-learn secara langsung.	Mendukung berbagai bahasa (Python, C++, JavaScript, dll.), alat debugging, terminal terintegrasi, dan ekstensi pihak ketiga untuk hampir semua kebutuhan pengembangan.
Performance	Ideal untuk dataset kecil hingga sedang.	Lebih optimal untuk proyek besar dengan kemampuan manajemen sumber daya yang lebih baik, termasuk multi-threading dan debugging real-time.
Visualisasi	Sangat baik untuk analisis visual dengan dukungan alat visualisasi bawaan.	Membutuhkan ekstensi tambahan atau pihak ketiga.

Visual Studio Code adalah alat yang sangat cocok untuk melakukan prediksi saham. Hal ini disebabkan oleh berbagai kemudahan yang ditawarkannya, seperti kemampuan untuk berkolaborasi, integrasi yang baik dengan berbagai *library* untuk analisis data visual, dan fitur-fitur lainnya. Selain itu, Visual Studio Code juga mendukung integrasi yang seamless dengan Jupyter Notebook, sehingga mempermudah proses pemodelan, analisis, dan kegiatan lainnya secara efisien.