

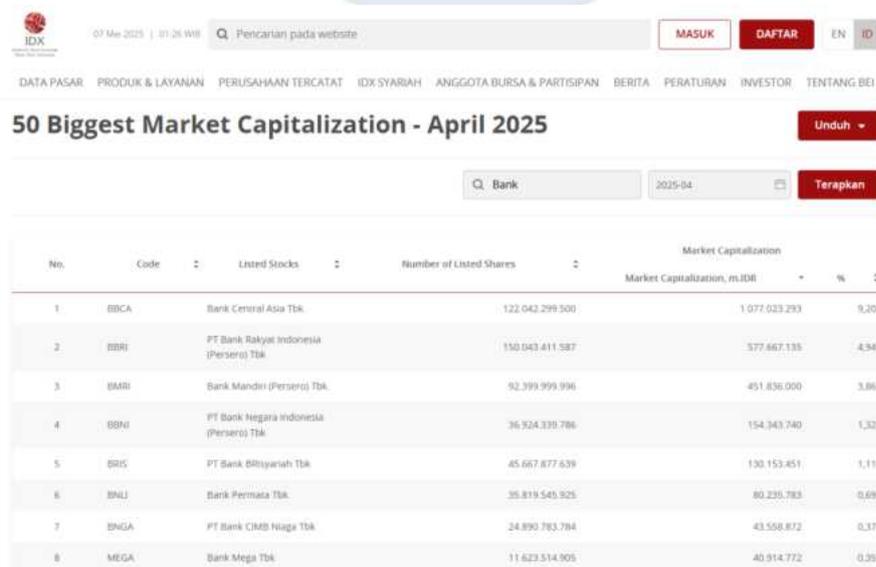
BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Penelitian ini dilakukan untuk melakukan analisis prediksi harga saham pada sektor perbankan. Data yang digunakan dalam penelitian ini merupakan data historis harga saham dari tiga perusahaan perbankan dengan kapitalisasi pasar terbesar di Indonesia. Data historis diambil melalui situs Yahoo Finance dengan rentang waktu 20 tahun terakhir, yaitu dari 1 Januari 2005 hingga 1 Maret 2025. Adapun ketiga perusahaan sektor perbankan yang menjadi objek penelitian ini adalah PT Bank Central Asia Tbk. (BBCA.JK), PT Bank Rakyat Indonesia (Persero) Tbk. (BBRI.JK), dan PT Bank Mandiri (Persero) Tbk. (BMRI.JK).

Ketiga perusahaan ini dipilih berdasarkan data dari website resmi Bursa Efek Indonesia, yang mencatat mereka sebagai tiga bank dengan kapitalisasi pasar terbesar di papan utama BEI. Data ini diperoleh dan didokumentasikan pada Gambar 3.1.



No.	Code	Listed Stocks	Number of Listed Shares	Market Capitalization	
				Market Capitalization, m.IDB	%
1	BBCA	Bank Central Asia Tbk.	122.042.299.500	1.077.023.293	9,20
2	BBRI	PT Bank Rakyat Indonesia (Persero) Tbk.	150.043.411.587	577.867.133	4,94
3	BMRI	Bank Mandiri (Persero) Tbk.	92.399.999.996	451.836.000	3,86
4	BBNI	PT Bank Negara Indonesia (Persero) Tbk.	36.924.339.786	154.343.740	1,32
5	BRIS	PT Bank BSI Syariah Tbk.	45.667.877.639	130.153.451	1,11
6	BNLI	Bank Permata Tbk.	35.819.545.925	80.235.783	0,69
7	BNIGA	PT Bank CIMB Niaga Tbk.	24.890.783.784	43.558.872	0,37
8	MEGA	Bank Mega Tbk.	11.623.514.905	40.914.772	0,35

Gambar 3. 1 Daftar Saham Perbankan dengan Kapitalisasi Pasar Tertinggi (April 2025)
Sumber: [9]

Ketiga saham tersebut dipilih karena memiliki tingkat likuiditas yang tinggi dan stabilitas yang kuat dalam pergerakan harga. Selain itu, saham-saham tersebut

juga masuk dalam kategori blue chip, serta menjadi komponen utama dalam indeks-indeks unggulan seperti LQ45 dan IDX80.

3.2 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan algoritma *deep learning*, yaitu LSTM, GRU, dan RNN untuk memprediksi harga saham pada sektor perbankan. Proses penelitian dilakukan secara bertahap dan sistematis mengikuti metode CRISP-DM (*Cross Industry Standard Process for Data Mining*). Metode ini dipilih karena fleksibel, iteratif, dan paling banyak digunakan dalam proyek data mining maupun analisis prediktif.

3.2.1 Alur Penelitian

Penelitian ini dilaksanakan dengan mengikuti alur sistematis yang berbasis pemrograman dan menggunakan pendekatan *time series forecasting* melalui model *deep learning*. Tahapan penelitian meliputi proses pengambilan data historis harga saham dari sumber terpercaya, dilanjutkan dengan tahap pengolahan dan persiapan data (*preprocessing*) untuk memastikan kualitas serta konsistensi data. Selanjutnya, dilakukan pembangunan dan pelatihan model prediktif menggunakan tiga jenis algoritma *deep learning*, yaitu LSTM, GRU, dan RNN. Proses *tuning hyperparameter* juga dilakukan untuk mengoptimalkan performa masing-masing model. Setelah itu, model dievaluasi menggunakan berbagai metrik akurasi guna menilai kemampuan prediksi. Tahap akhir melibatkan penerapan model terbaik untuk melakukan prediksi harga saham pada periode mendatang (*forecasting*), sehingga menghasilkan output yang dapat dianalisis secara komprehensif.

3.2.2 Metode *Data Mining*

Pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*) dipilih dalam penelitian ini karena kecocokannya dengan karakteristik penelitian prediksi pergerakan harga saham. CRISP-DM memberikan kerangka kerja yang terstruktur dan sistematis untuk mengatasi tahapan-tahapan penting dalam proses penambangan data, mulai dari

pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, hingga implementasi model. Dengan fokus pada prediksi harga saham, CRISP-DM mendukung analisis yang holistik terhadap data historis, memungkinkan pemilihan fitur yang relevan, dan memudahkan evaluasi kinerja model dengan metrik yang sesuai. Keandalan dan fleksibilitas CRISP-DM menjadikannya pilihan yang tepat untuk menangani tantangan kompleks yang terkait dengan prediksi pergerakan harga saham.

Dalam proses *data mining*, terdapat beberapa *framework* yang dapat dijadikan acuan. Berikut adalah Tabel 3.1 yang menyajikan perbandingan singkat antara *framework* CRISP-DM dengan *framework* lainnya yaitu SEMMA (*Sample, Explore, Modify, Model, Assess*) dan KDD (*Knowledge Discovery in Databases*) [51], [52]:

Tabel 3. 1 Perbandingan Framework CRISP-DM, SEMMA, dan KDD

Aspek	CRISP-DM	SEMMA	KDD
Tahapan	<ol style="list-style-type: none"> 1. <i>Business Understanding</i> 2. <i>Data Understanding</i> 3. <i>Data Preparation</i> 4. <i>Modeling</i> 5. <i>Evaluation</i> 6. <i>Deployment</i> 	<ol style="list-style-type: none"> 1. <i>Sample</i> 2. <i>Explore</i> 3. <i>Modify</i> 4. <i>Model</i> 5. <i>Assess</i> 	<ol style="list-style-type: none"> 1. <i>Selection</i> 2. <i>Preprocessing</i> 3. <i>Transformation</i> 4. <i>Data Mining</i> 5. <i>Interpretation / Evaluation</i>
Kelebihan	<ul style="list-style-type: none"> - Fokus pada pemahaman tujuan bisnis, tidak hanya aspek teknis. - Fleksibel, dapat 	<ul style="list-style-type: none"> - Menyediakan alur kerja yang logis untuk tugas-tugas inti <i>data mining</i>. - Terstruktur 	<ul style="list-style-type: none"> - Menjadi kerangka kerja konseptual yang mendasari metodologi lain seperti CRISP-

Aspek	CRISP-DM	SEMMA	KDD
	diadaptasi, dan independen dari teknologi atau vendor tertentu.	dengan baik untuk digunakan dalam <i>toolset</i> SAS Enterprise Miner.	DM. - Mencakup proses luas dari data hingga menjadi sebuah pengetahuan yang dapat ditindaklanjuti.
Kekurangan	- Tidak secara spesifik mengatur manajemen proyek atau praktik <i>agile</i> . - Proses iterasinya tidak didefinisikan secara jelas (kapan dan bagaimana).	- Kurang menekankan pada fase pemahaman bisnis. - Terikat pada vendor spesifik (SAS), yang membatasi penerapan umumnya.	- Tidak mendefinisikan fase Pemahaman Bisnis dan <i>Deployment</i> secara eksplisit. - Dianggap sebagai proses yang kurang detail dibandingkan evolusi praktisnya, yaitu CRISP-DM.

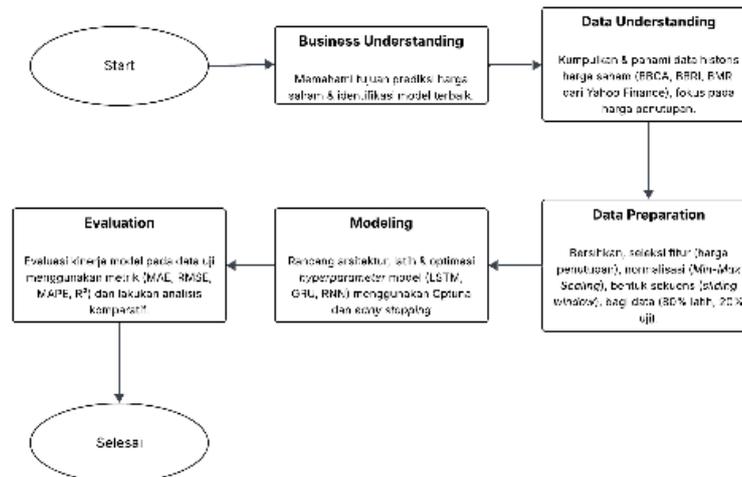
Penelitian ini menggunakan kerangka kerja CRISP-DM karena kerangka ini menyediakan struktur tahapan yang sistematis, fleksibel, dan iteratif, yang sangat sesuai untuk menangani kompleksitas prediksi harga saham di sektor perbankan Indonesia. Tahapan *Business Understanding* dalam CRISP-DM menjadi nilai lebih yang krusial karena membantu peneliti merumuskan permasalahan dari sisi praktis dunia investasi bukan hanya dari aspek teknis data. Dalam konteks ini, fluktuasi harga saham bank seperti BBKA, BBRI, dan BMRI tidak hanya merupakan fenomena matematis, tetapi berkaitan erat

dengan dinamika pasar, psikologi investor, dan sentimen ekonomi nasional. CRISP-DM juga menyediakan *business understanding* yang dimana membantu memahami permasalahan untuk menghitung estimasi kerugian per lot dari saham perbankan (BBCA, BBRI, BMRI) untuk masing-masing algoritma yaitu LSTM, GRU dan RNN. Oleh karena itu, pendekatan yang diawali dengan pemahaman mendalam terhadap kebutuhan investor dan karakteristik pasar modal menjadi sangat penting.

Berbeda dengan SEMMA dan KDD, CRISP-DM memiliki keunggulan karena dimulai dengan tahapan *Business Understanding* yang eksplisit, yang memungkinkan pemetaan masalah dilakukan sejak awal dengan mempertimbangkan konteks bisnis dan kebutuhan analisis pasar. SEMMA lebih menitikberatkan pada eksplorasi dan pemodelan tanpa penggalian kebutuhan bisnis yang mendalam, sementara KDD tidak secara tegas membedakan tahapan pemahaman bisnis dari proses eksplorasi data. Selain itu, sifat CRISP-DM yang iteratif sangat cocok dengan kebutuhan penelitian ini yang melibatkan proses penyempurnaan model secara bertahap, seperti *tuning hyperparameter* dan evaluasi performa algoritma.

Dengan kerangka ini, proses penelitian menjadi lebih terarah, mulai dari pemahaman konteks bisnis, eksplorasi data historis saham, hingga pemodelan dan evaluasi kinerja algoritma. Oleh karena itu, CRISP-DM dipilih karena mampu menjembatani kebutuhan teknis dan tujuan praktis dalam pengambilan keputusan berbasis data di sektor keuangan.

UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3. 2 Diagram Alur Penelitian

Berikut lima Langkah CRISP-DM yang digunakan dalam penelitian ini berdasarkan Gambar 3.2:

3.2.2.1 *Business Understanding*

Tahap awal ini difokuskan pada pemahaman tujuan utama penelitian dan kebutuhan dari perspektif praktis di pasar modal. Konteksnya adalah tingginya volatilitas harga saham di sektor perbankan Indonesia dan pentingnya prediksi yang akurat bagi investor dan analis pasar. Tujuan penelitian secara spesifik adalah untuk melakukan analisis perbandingan dan optimasi kinerja tiga model *deep learning* (LSTM, GRU, dan RNN) dalam memprediksi harga saham harian tiga bank *blue-chip* utama, yaitu BBCA, BBRI, dan BMRI. Permasalahan utama yang akan dijawab adalah model mana yang menunjukkan akurasi prediksi tertinggi setelah melalui proses optimasi *hyperparameter*. Keberhasilan model akan dinilai berdasarkan serangkaian metrik evaluasi standar yang relevan untuk tugas prediksi finansial.

3.2.2.2 *Data Understanding*

Setelah tujuan penelitian dirumuskan, tahap ini akan melibatkan proses pengumpulan dan eksplorasi data awal guna memahami karakteristik, kualitas, dan struktur data yang akan digunakan. Data historis harga saham harian untuk BBCA, BBRI, dan BMRI akan dikumpulkan dari platform

Yahoo Finance untuk rentang waktu 20 tahun (1 Januari 2005 – 1 Maret 2025). Aktivitas dalam tahap ini meliputi identifikasi atribut data yang tersedia seperti harga pembukaan, tertinggi, terendah, penutupan, dan volume transaksi, dengan fokus utama pada atribut harga penutupan sebagai variabel target prediksi. Tahap ini juga mencakup pelaksanaan analisis deskriptif awal dan visualisasi data untuk mengidentifikasi tren historis, volatilitas, serta potensi anomali yang dapat memengaruhi proses pemodelan selanjutnya.

3.2.2.3 *Data Preparation*

Tahap persiapan data bertujuan untuk mentransformasi data mentah menjadi format yang optimal dan sesuai untuk digunakan dalam pelatihan model *deep learning*. Proses ini sangat krusial karena kualitas data input akan sangat memengaruhi kinerja model. Langkah-langkah yang akan dilakukan meliputi:

1. Pembersihan Data: Pemeriksaan dan penanganan data yang hilang (*missing values*), data yang duplikat, dan pengecekan ketersediaan data pada hari bursa aktif untuk memastikan konsistensi *dataset*.
2. Seleksi Fitur: Pembatasan fitur input hanya pada harga penutupan (*Close Price*) harian. Harga penutupan dipilih karena merepresentasikan nilai konsensus pasar pada akhir hari perdagangan dan menjadi titik data tunggal yang paling signifikan untuk merefleksikan sentimen pasar harian. Data saham diperoleh melalui *platform* Google Colab menggunakan *library* *yfinance*, yang mengambil data langsung dari Yahoo Finance. Meskipun di situs Yahoo Finance tersedia kolom *Adjusted Close* yang telah disesuaikan terhadap aksi korporasi seperti dividen atau *stock split*, namun pada hasil pengambilan data menggunakan *yfinance*, kolom tersebut tidak tersedia secara *default*. Oleh karena itu, digunakanlah kolom *Close* sebagai representasi harga saham yang valid dan konsisten antar emiten. Penggunaan *Close Price* sebagai fitur utama

merupakan pendekatan yang telah divalidasi dan diterapkan dalam penelitian terdahulu [11], [36], [53].

3. Normalisasi Data: Penerapan teknik *Min-Max Scaling* untuk mengubah skala nilai harga penutupan ke rentang antara 0 dan 1. Normalisasi data adalah langkah pra-pemrosesan yang krusial, khususnya untuk data saham yang memiliki *noise* dan frekuensi yang tinggi. Teknik ini juga berperan penting dalam menyelaraskan skala data antar saham. Dalam penelitian ini, ketiga saham bank (BBCA, BBRI, dan BMRI) memiliki rentang harga yang berbeda, sehingga normalisasi dengan *Min-Max Scaler* dilakukan untuk memastikan bahwa seluruh data berada dalam skala yang seragam yaitu dengan rentang 0 hingga 1 dan dapat diproses secara konsisten oleh model. Praktik normalisasi data ini merupakan standar dalam pemodelan deret waktu finansial menggunakan *deep learning* dan diterapkan juga dalam studi terkait [11], [14], [36], [54].
4. Pembentukan *Sequence*: Transformasi data deret waktu menjadi format *sequence* input (fitur) dan target *output* menggunakan metode *sliding window*. Dalam penelitian ini, panjang *time_step* ditetapkan sebesar 30, yang berarti model akan menggunakan data harga historis dari 30 hari terakhir untuk memprediksi harga pada hari berikutnya. Pemilihan periode ini bertujuan untuk menangkap pola harga jangka pendek hingga menengah yang relevan, sekaligus menjaga efisiensi komputasi model. Pendekatan ini dinilai dapat memberikan konteks historis yang memadai dan didukung oleh temuan pada penelitian [8], yang menunjukkan bahwa *time_step* 30 hari menghasilkan performa prediksi yang lebih baik dibandingkan periode yang lebih pendek. Data *sequence* input yang terbentuk kemudian disesuaikan dimensinya agar sesuai dengan format yang dibutuhkan oleh model *recurrent*.
5. Pembagian Data: Pembagian *dataset* yang telah diproses menjadi data latih 80% yang akan digunakan untuk melatih dan mengoptimasi model, serta data uji 20% untuk evaluasi akhir

performa model pada data independen. Proporsi 80:20 ini dipilih karena merupakan standar umum yang menyeimbangkan antara kecukupan data untuk pelatihan model yang efektif dan volume data yang representatif untuk pengujian objektif, sebuah pendekatan yang juga diterapkan dalam penelitian-penelitian sebelumnya seperti pada [8] dan [55].

3.2.2.4 *Modeling*

Pada tahap ini, akan dilakukan perancangan, pelatihan, dan optimasi model-model prediktif. Tiga jenis algoritma *deep learning* untuk data deret waktu LSTM, GRU, dan RNN akan diimplementasikan menggunakan *library* Keras dengan *backend* TensorFlow. Proses utama dalam tahap ini adalah:

1. Perancangan Model: Untuk setiap algoritma, akan dirancang model awal yang terdiri dari lapisan *recurrent* (LSTM, GRU, atau RNN), lapisan *dropout* untuk regularisasi, dan lapisan *output Dense* untuk menghasilkan nilai prediksi.
2. Optimasi *Hyperparameter*: Akan dilakukan proses optimasi *hyperparameter* secara sistematis dan otomatis menggunakan *framework* Optuna untuk setiap model dan setiap saham. Optimasi akan difokuskan pada *hyperparameter* kunci yang secara langsung memengaruhi kapasitas model dan risiko *overfitting* seperti jumlah unit pada lapisan *recurrent*, *dropout rate*, dan *learning rate*. Jumlah unit menentukan kapasitas representasi model untuk menangkap pola temporal, *dropout rate* berfungsi sebagai teknik regularisasi penting untuk mencegah *overfitting* dengan menonaktifkan sebagian koneksi selama pelatihan, dan *learning rate* krusial dalam mengatur kecepatan dan stabilitas konvergensi pelatihan [56]. Optuna akan menjalankan sejumlah *trial* untuk menemukan kombinasi *hyperparameter* yang menghasilkan kinerja terbaik. Keunggulan Optuna dibanding metode manual atau *Grid Search* telah dibuktikan oleh penelitian [57] dalam studi LSTM untuk prediksi harga saham,

yang menunjukkan penurunan RMSE lebih baik dibanding optimasi manual. Selain itu, penelitian [58] menunjukkan *tuning* lebih cepat dengan Optuna daripada *Grid Search*. Oleh karena itu, Optuna digunakan untuk menghemat waktu dalam pemilihan *hyperparameter* secara efisien. Hal ini penting karena waktu penggunaan GPU di Google Colab terbatas, dan jika habis, proses akan dialihkan ke CPU yang jauh lebih lambat dalam melatih model *deep learning*.

3. Pelatihan Model Final: Setelah konfigurasi *hyperparameter* optimal diperoleh dari Optuna, model final untuk setiap algoritma akan dibangun kembali dengan konfigurasi tersebut dan dilatih menggunakan seluruh data latih. Mekanisme *early stopping* akan diterapkan selama pelatihan untuk mencegah *overfitting* dan memastikan model memiliki kemampuan generalisasi yang baik. Mekanisme ini bekerja dengan memantau *loss* pada data validasi dan menghentikan pelatihan jika tidak ada perbaikan dalam periode *patience*, sehingga memastikan model yang disimpan adalah versi dengan kemampuan generalisasi terbaik.
4. Verifikasi Model dan Diagnosis *Overfitting*: Setelah pelatihan model final, dilakukan verifikasi model untuk mendiagnosis *overfitting*. Langkah ini penting karena sebuah model prediktif yang andal harus memiliki kemampuan generalisasi yang baik, bukan sekadar mencapai nilai *error* (seperti MAE atau RMSE) yang rendah dengan cara menghafal data historis. Oleh karena itu, verifikasi terhadap *overfitting* adalah langkah wajib untuk memastikan bahwa model yang dihasilkan baik dan dapat dipercaya. Proses verifikasi dilakukan dengan metode analisis kualitatif terhadap kurva *loss* yang dihasilkan selama pelatihan model final. Praktik memonitor kurva *training* dan *validation loss* ini merupakan pendekatan yang divalidasi oleh beberapa penelitian seperti [29], [59], [60]. Kondisi model kemudian disimpulkan berdasarkan apakah perilaku kurva menunjukkan konvergensi, divergensi, atau stagnasi.

Untuk memberikan gambaran yang lebih jelas mengenai model, tabel 3.2 menunjukkan perbandingan yang merangkum kelebihan, kekurangan, serta karakteristik utama dari masing-masing algoritma yang digunakan dalam penelitian ini [39], [42], [43], [44], [61], [45], [46]:

Tabel 3. 2 Perbandingan Algoritma RNN, LSTM, dan GRU

Algoritma	Kelebihan	Kekurangan	Karakteristik Utama
RNN	<ul style="list-style-type: none"> - Arsitektur sederhana dan komputasi ringan. - Mampu memproses input dengan panjang yang bervariasi. 	<ul style="list-style-type: none"> - Mengalami masalah <i>vanishing gradient</i> dan <i>exploding gradient</i>. - Kesulitan mengingat informasi jangka panjang. 	<ul style="list-style-type: none"> - Memiliki <i>feedback loop</i> di mana output dari langkah sebelumnya digunakan sebagai input untuk langkah saat ini.
LSTM	<ul style="list-style-type: none"> - Mampu mengatasi masalah <i>vanishing gradient</i>. - Mampu mempelajari dan mengingat dependensi jangka panjang dalam data. 	<ul style="list-style-type: none"> - Membutuhkan waktu <i>training</i> yang lebih lama karena arsitekturnya yang kompleks. 	<ul style="list-style-type: none"> - Menggunakan tiga gerbang (<i>gates</i>): <i>forget gate</i>, <i>input gate</i>, dan <i>output gate</i> untuk mengatur informasi yang disimpan dan diteruskan.
GRU	<ul style="list-style-type: none"> - Memiliki arsitektur yang lebih sederhana 	<ul style="list-style-type: none"> - Kemampuan nya terkadang sedikit di bawah 	<ul style="list-style-type: none"> - Menggunakan dua gerbang, yaitu <i>reset gate</i> dan <i>update gate</i>, yang

Algoritma	Kelebihan	Kekurangan	Karakteristik Utama
	dibandingkan LSTM. - Komputasi lebih cepat dan efisien daripada LSTM.	LSTM pada beberapa kasus.	menyederhanakan arsitektur LSTM dengan menggabungkan fungsi dari <i>input gate</i> dan <i>forget gate</i> .

3.2.2.5 Evaluation

Setelah tahap pemodelan dan optimasi *hyperparameter* untuk model LSTM, GRU, dan RNN selesai, tahap evaluasi bertujuan untuk mengukur kinerja prediktif model LSTM, GRU, dan RNN yang telah dioptimasi, dengan menggunakan data uji yang independen guna memastikan penilaian objektif terhadap kemampuan generalisasi model. Setiap model yang optimal akan menghasilkan prediksi harga saham pada data uji, yang nilainya kemudian dikembalikan ke skala harga asli melalui proses *inverse transform* untuk perbandingan yang valid. Kinerja kuantitatif selanjutnya akan diukur menggunakan serangkaian metrik standar, yaitu MAE, RMSE, MAPE, dan R^2 . Hasil dari metrik-metrik ini akan menjadi dasar analisis komparatif untuk mengidentifikasi model dengan akurasi tertinggi untuk setiap saham (BBCA, BBRI, dan BMRI), yang juga akan dilengkapi dengan visualisasi perbandingan prediksi terhadap harga aktual untuk interpretasi kualitatif. Selain melakukan analisis komparatif berdasarkan metrik evaluasi, penelitian ini juga menginterpretasikan nilai MAPE dalam konteks finansial untuk mengestimasi potensi risiko kerugian per lot saham. Penggunaan MAPE dipilih karena metrik ini menyajikan kesalahan prediksi dalam bentuk persentase, sehingga lebih mudah dipahami secara praktis dan relevan untuk menggambarkan estimasi kerugian relatif terhadap harga saham aktual.

3.3 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan melalui metode pengumpulan data sekunder, yaitu data yang diperoleh dari sumber yang telah tersedia sebelumnya. Penelitian ini menggunakan data historis harga saham dari tiga perusahaan perbankan besar di Indonesia, yaitu PT Bank Central Asia Tbk (BBCA), PT Bank Rakyat Indonesia Tbk (BBRI), dan PT Bank Mandiri Tbk (BMRI). Data diambil dari layanan penyedia data keuangan *Yahoo Finance*, yang diakses secara otomatis menggunakan *library* Python bernama *yfinance*.

Pemilihan *Yahoo Finance* sebagai sumber data utama didasarkan pada beberapa pertimbangan kunci. *Platform* ini menyediakan akses publik gratis ke data harga saham historis yang komprehensif dengan rentang waktu yang panjang (dalam penelitian ini dari 1 Januari 2005 hingga 1 Maret 2025), yang sangat esensial untuk melatih model *deep learning*. Selain itu, kemudahan pengambilan data secara otomatis melalui *library* seperti *yfinance* serta statusnya sebagai sumber yang lazim digunakan dalam berbagai penelitian akademis turut mendukung keputusan penggunaan platform ini.

Data yang dikumpulkan mencakup harga saham harian dalam rentang waktu 1 Januari 2005 hingga 1 Maret 2025. Pengambilan data ini dilakukan dengan menggunakan skrip Python yang memanggil ticker saham masing-masing perusahaan, yaitu BBCA.JK, BBRI.JK, dan BMRI.JK. Atribut yang digunakan dalam penelitian ini hanya terbatas pada harga penutupan saham, mengingat fokus utama model prediksi adalah terhadap nilai penutupan sebagai representasi nilai akhir transaksi harian.

Rentang waktu data selama 20 tahun (1 Januari 2005 – 1 Maret 2025) dipilih untuk memastikan analisis mencakup berbagai siklus dan kondisi pasar. Periode yang panjang mencakup beragam siklus dan kondisi pasar, termasuk periode pertumbuhan stabil, krisis finansial global pada tahun 2008, serta disrupsi pasar yang signifikan akibat pandemi COVID-19 pada tahun 2020. Dengan menyertakan data dari berbagai kondisi pasar tersebut, model *deep learning* dilatih untuk mengenali tidak hanya tren jangka panjang, tetapi juga pola volatilitas ekstrem. Hal

ini penting untuk membangun model yang lebih robust dan memiliki kemampuan generalisasi yang baik.

Pendekatan ini didukung oleh temuan penelitian [62] yang menunjukkan bahwa cakupan data historis yang lebih panjang berkorelasi positif dengan akurasi prediksi. Penelitian [62] menunjukkan bahwa penggunaan data selama 7 tahun menghasilkan prediksi yang lebih akurat secara signifikan dibandingkan dengan data yang hanya mencakup 2 tahun. Oleh karena itu, rentang waktu 20 tahun dalam penelitian ini menyediakan dataset yang komprehensif untuk melatih model.

3.4 Teknik Analisis Data

Dalam Teknik analisis data dalam penelitian ini mengacu pada kerangka kerja CRISP-DM. Seluruh proses analisis dirancang untuk mengikuti tahapan-tahapan tersebut secara sistematis guna menghasilkan model prediksi harga saham yang optimal dan aplikatif. Dalam tahap *Modeling*, digunakan pendekatan *deep learning* dengan menerapkan tiga algoritma utama, yaitu LSTM, GRU, dan RNN, yang sangat sesuai untuk menangani pola data historis yang bersifat sekuensial seperti harga saham.

Ketiga model tersebut diimplementasikan menggunakan bahasa pemrograman Python, dengan bantuan pustaka seperti TensorFlow dan Keras yang secara luas digunakan dalam pengembangan model *deep learning*. Pemilihan Python sebagai alat bantu utama dalam penelitian ini didasarkan pada keunggulannya dibandingkan bahasa pemrograman lain, khususnya R, dalam hal fleksibilitas, ekosistem *library*, serta performa dalam skenario pemodelan *machine learning* berskala besar. Berikut adalah tabel perbandingan antara bahasa Python dan R yang disajikan untuk memperkuat alasan pemilihan tersebut [63]:

Tabel 3. 3 Perbandingan Bahasa Python dan R

Aspek	Python	R
Kemudahan Penggunaan	Sintaksis lebih sederhana dan mudah dipahami, cocok untuk pemula.	Sintaksis lebih kompleks dan berorientasi pada statistik.
Fleksibilitas	Bahasa pemrograman umum yang mendukung berbagai aplikasi, termasuk ML dan pengembangan web.	Fokus utama pada analisis statistik dan visualisasi data.
Komunitas dan Dukungan	Komunitas lebih besar untuk <i>Machine Learning</i> dan produksi.	Komunitas kuat di statistik dan riset akademik
<i>Library</i> untuk <i>Machine Learning</i>	Memiliki <i>library</i> lengkap seperti Scikit-learn, TensorFlow, dan PyTorch untuk pemodelan canggih.	<i>Library</i> caret dan randomForest tersedia, tetapi tidak selengkap Python.
Visualisasi Data	<i>Library</i> seperti Matplotlib dan Seaborn memungkinkan visualisasi yang baik.	ggplot2 unggul dalam visualisasi statistik, tetapi kurang fleksibel untuk kebutuhan non-statistik.
Kecepatan Eksekusi	Lebih cepat dalam menangani dataset besar dan komputasi intensif.	Lebih lambat saat menangani dataset besar.

Berdasarkan Tabel 3.3, Python dinilai lebih unggul untuk kebutuhan analisis data dan pemodelan *machine learning* dalam penelitian ini, terutama dalam konteks prediksi harga saham yang memerlukan pemrosesan data dalam jumlah besar dan pembangunan model yang kompleks. Oleh karena itu, Python dipilih sebagai bahasa pemrograman utama yang digunakan dalam keseluruhan tahapan CRISP-DM.

3.4.1 Variabel Penelitian

Dalam penelitian ini, terdapat dua jenis variabel yang digunakan, yaitu variabel independen (*input*) dan variabel dependen (*output*).

1. Variabel Independen

Variabel ini merupakan data historis harga saham yang digunakan sebagai masukan (fitur) untuk model prediksi. Fitur yang digunakan dalam penelitian ini adalah harga penutupan (*Close*) saham harian. Namun, karena model *recurrent* seperti LSTM, GRU, dan RNN memproses data secara sekuensial, variabel input ini tidak dimasukkan sebagai nilai tunggal, melainkan distrukturkan menjadi sebuah urutan dengan panjang waktu (*time_step*) 30 hari. Dengan demikian, *input shape* yang diterima oleh model adalah sebuah tensor 3D dengan format (jumlah sampel, 30, 1), di mana 30 adalah panjang *sequence* dan 1 merepresentasikan satu-satunya fitur yang digunakan.

2. Variabel Dependen

Variabel dependen merupakan nilai yang akan diprediksi oleh model. Variabel tersebut adalah harga penutupan di masa depan (*Next Close*), yaitu nilai harga saham pada hari ke-31 yang diperkirakan berdasarkan *sequence* 30 hari sebelumnya. Model akan menggunakan data input historis untuk mempelajari pola pergerakan harga, kemudian menghasilkan prediksi terhadap harga penutupan pada periode selanjutnya. Variabel ini berperan sebagai target output yang menjadi dasar dalam mengevaluasi kinerja model prediktif yang dibangun.

