

BAB II TINJAUAN PUSTAKA

2.1 Kajian Relevan

Penelitian terkait bertujuan untuk memberikan gambaran mengenai perkembangan metode dan pendekatan yang telah digunakan dalam penelitian sebelumnya, khususnya yang berkaitan dengan data deret waktu (time series) dan penerapan metode deep learning. Kajian ini juga berfungsi untuk menunjukkan posisi penelitian yang dilakukan dibandingkan dengan penelitian terdahulu serta mengidentifikasi celah penelitian (research gap).

Berbagai penelitian menunjukkan bahwa metode deep learning, terutama Long Short-Term Memory (LSTM), memiliki kemampuan yang baik dalam menangani data sekuensial dan deret waktu yang bersifat non-linear. Selain LSTM, perkembangan terbaru juga menunjukkan penggunaan arsitektur deep learning lain seperti Transformer dan Autoencoder dalam memodelkan data time series.

Untuk mempermudah pemahaman dan perbandingan, kajian penelitian relevan dalam penelitian ini disajikan dalam bentuk tabel sebagai berikut.

Tabel 2. 1 Kajian Relevan 1

No	Penulis (Tahun)	Judul Penelitian	Objek & Metode	Relevansi
1	Fathurrahman & Saepudin (2021)	Prediksi Harga Saham Perbankan di Indonesia Menggunakan LSTM	BMRI, BBNI, BBRI (LSTM)	Membuktikan bahwa LSTM efektif memprediksi saham bank BUMN (Himbara) dengan akurasi tinggi.
2	Wibowo & Riana (2023)	Analisis Perbandingan Model RNN dan LSTM dalam Prediksi Harga Saham	BBNI (LSTM)	Menunjukkan bahwa LSTM lebih unggul dibandingkan RNN standar dalam menangkap tren saham BBNI.

Tabel 2. 2Kajian Relevan 2

3	Pratama & Hartawan (2022)	Implementasi Algoritma LSTM untuk Prediksi Harga Saham Bank Mandiri	BMRI (LSTM)	Fokus spesifik pada saham BMRI, merekomendasikan penggunaan <i>Epoch</i> yang cukup untuk mencapai konvergensi.
4	Moghar & Hamiche (2020)	Stock Market Prediction Using LSTM	General (LSTM)	Landasan teori mengenai pentingnya pengaturan <i>Epoch</i> untuk menghindari <i>overfitting</i> .
i5	Nugraha (2023)	Analisis Dampak Sliding Window pada Kinerja Algoritma LSTM	Time Series	Landasan teori eksperimen variabel <i>Window Size</i> .
6	Sitio et al. (2024)	Analisis Sentimen Ulasan Aplikasi Gojek Menggunakan LSTM	Ulasan aplikasi Gojek, klasifikasi sentimen, metode LSTM + Word2Vec	Menunjukkan efektivitas LSTM dalam pemodelan data teks berurutan (time series tekstual) dan klasifikasi berbasis memori historis
7	Anwar et al. (2024)	Analisis Sentimen Aplikasi PLN Mobile Review Menggunakan LSTM	Review aplikasi PLN Mobile, text mining, LSTM	Membuktikan LSTM mampu meningkatkan akurasi pada data sekuens
8	Nurrahmah et al. (2025)	Prediksi Kualitas Udara di Jakarta Menggunakan LSTM	Data ISPU (PM2.5, CO, dll), prediksi time series, LSTM	Menguatkan penggunaan LSTM untuk prediksi numerik deret waktu non-linear mirip data saham
9	Wulandari et al. (2024)	Analisis Sentimen Opini Publik Twitter Menggunakan LSTM	Tweet Twitter, klasifikasi sentimen, LSTM + Word2Vec	Menunjukkan kemampuan LSTM memproses data sekuens panjang dan kompleks

Tabel 2. 3Kajian Relevan 3

10	Fathahillah et al. (2025)	Optimasi Perlakuan Tanaman Padi Menggunakan CNN-LSTM	Data citra + lingkungan, CNN-LSTM	Menunjukkan kombinasi deep learning untuk pola temporal dan spasial, memperkuat fleksibilitas arsitektur LSTM
11	Khairunnisa et al. (2023)	Prediksi Curah Hujan Menggunakan LSTM	Data cuaca time series, LSTM, evaluasi RMSE	Serupa dengan penelitian ini karena memprediksi data numerik deret waktu menggunakan evaluasi kesalahan
12	Ridha et al. (2025)	Prediksi Konsumsi Energi Listrik Menggunakan LSTM	Konsumsi listrik bulanan, LSTM, MAPE & RMSE	Relevan dalam konteks prediksi tren historis jangka panjang seperti pergerakan saham
13	Hayani et al. (2024)	Analisis Sentimen Komentar Youtube Menggunakan LSTM	Komentar Youtube, LSTM + Word2Vec	Menguatkan efektivitas LSTM dalam pembelajaran sekuens besar
14	Fadhilah et al. (2025)	Prediksi Ketinggian Air Bendungan Menggunakan LSTM	Data ketinggian air & curah hujan, LSTM	Mirip dengan penelitian ini karena memodelkan time series numerik dengan arsitektur LSTM berlapis

Penelitian yang dilakukan oleh Sitio et al. (2024) membahas penerapan algoritma Long Short-Term Memory (LSTM) dalam melakukan analisis sentimen terhadap ulasan pengguna aplikasi Gojek. Data yang digunakan berupa teks ulasan pengguna yang dikumpulkan dari platform digital, kemudian diproses menggunakan teknik *text preprocessing* dan representasi kata Word2Vec sebelum dimasukkan ke dalam model LSTM. Pendekatan ini bertujuan untuk mengklasifikasikan sentimen pengguna ke dalam kategori positif, negatif, atau netral.

Hasil penelitian menunjukkan bahwa LSTM mampu mempelajari pola urutan kata dalam kalimat secara efektif karena memiliki memori jangka panjang yang dapat menangkap konteks linguistik. Dibandingkan metode konvensional seperti Naïve Bayes atau SVM, LSTM memberikan akurasi yang lebih tinggi karena mampu memahami dependensi antar kata dalam urutan teks. Hal ini membuktikan bahwa LSTM unggul dalam memodelkan data sekuensial.

Keterkaitan penelitian tersebut dengan penelitian ini terletak pada kesamaan penggunaan LSTM untuk memproses data berurutan. Meskipun objek penelitian berbeda, prinsip pembelajaran pola temporal pada data sekuens tetap relevan. Oleh

karena itu, penelitian Sitio et al. memperkuat dasar teoritis bahwa LSTM efektif dalam menangani data runtun waktu seperti harga saham.

Anwar et al. (2024) melakukan penelitian mengenai analisis sentimen terhadap ulasan aplikasi PLN Mobile menggunakan algoritma LSTM. Data yang dianalisis berupa teks komentar pengguna yang mencerminkan pengalaman terhadap layanan aplikasi. Tahapan penelitian meliputi pembersihan data, tokenisasi, pembobotan kata, serta pelatihan model LSTM untuk klasifikasi sentimen.

Penelitian ini menunjukkan bahwa LSTM memiliki kemampuan dalam memahami hubungan kontekstual antar kata dalam kalimat, sehingga dapat menghasilkan klasifikasi yang lebih akurat dibandingkan metode *machine learning* tradisional. Model mampu mengenali pola kalimat yang kompleks dan tidak linier, sehingga cocok untuk data dengan struktur urutan.

Relevansi penelitian ini dengan penelitian yang dilakukan terletak pada kesamaan karakteristik data sekuens. Walaupun digunakan pada teks, LSTM tetap memanfaatkan memori historis untuk menghasilkan prediksi. Konsep ini sejalan dengan prediksi harga saham, di mana informasi masa lalu digunakan untuk memprediksi nilai masa depan.

Nurrahmah et al. (2025) mengembangkan model prediksi kualitas udara di Jakarta menggunakan algoritma LSTM. Data yang digunakan berupa parameter polusi udara seperti PM2.5, CO, dan NO₂ yang direkam secara berkala. Penelitian ini bertujuan memprediksi tingkat kualitas udara pada periode selanjutnya berdasarkan data historis.

Hasil penelitian menunjukkan bahwa LSTM mampu memodelkan pola fluktuasi kualitas udara yang dipengaruhi faktor musiman dan tren jangka panjang. Evaluasi menggunakan RMSE dan MAE menunjukkan performa prediksi yang baik. Model LSTM mampu menangkap dependensi jangka panjang yang sulit dipelajari oleh metode linier.

Penelitian ini sangat relevan dengan penelitian prediksi saham karena sama-sama menggunakan data numerik time series. Pendekatan, arsitektur, serta metode

evaluasi yang digunakan hampir serupa, sehingga dapat dijadikan referensi langsung dalam perancangan model prediksi harga saham.

Wulandari et al. (2024) meneliti analisis sentimen opini publik pada media sosial Twitter menggunakan LSTM dan Word2Vec. Data berupa ribuan tweet diproses untuk mengetahui persepsi masyarakat terhadap suatu topik tertentu. LSTM digunakan untuk memahami urutan kata dan makna kontekstual dalam kalimat.

Penelitian ini menunjukkan bahwa model LSTM mampu mengelola data berukuran besar dan memiliki pola sekuens panjang dengan baik. Hasil klasifikasi yang diperoleh lebih stabil dibandingkan model konvensional. Hal ini menegaskan keunggulan LSTM dalam pembelajaran pola berurutan.

Keterkaitannya dengan penelitian ini terletak pada konsep pembelajaran memori historis dan pemrosesan data sekuens. Prinsip kerja LSTM dalam memahami urutan informasi sangat relevan untuk prediksi harga saham yang juga memiliki pola temporal.

Penelitian Fathahillah et al. (2025) menggabungkan CNN dan LSTM untuk optimasi perlakuan tanaman padi. CNN digunakan untuk ekstraksi fitur spasial dari citra, sedangkan LSTM digunakan untuk mempelajari pola temporal pertumbuhan tanaman.

Pendekatan hibrida ini menunjukkan bahwa LSTM efektif dalam memodelkan perubahan data dari waktu ke waktu. Kombinasi tersebut menghasilkan akurasi prediksi yang tinggi terhadap kondisi tanaman di masa mendatang.

Relevansi penelitian ini terletak pada fleksibilitas arsitektur LSTM dalam berbagai domain. Hal ini membuktikan bahwa LSTM dapat diterapkan tidak hanya pada teks, tetapi juga pada data numerik dan prediksi berkelanjutan seperti harga saham.

Khairunnisa et al. (2023) memprediksi curah hujan menggunakan model LSTM berbasis data meteorologi historis. Variabel yang digunakan meliputi suhu,

kelembaban, dan tekanan udara. Penelitian ini fokus pada prediksi nilai numerik masa depan berdasarkan pola historis.

Hasil eksperimen menunjukkan bahwa LSTM mampu menghasilkan prediksi dengan nilai RMSE rendah. Model berhasil menangkap pola musiman serta tren jangka panjang yang terdapat pada data cuaca.

Penelitian ini memiliki kesamaan langsung dengan penelitian prediksi saham karena sama-sama memodelkan data numerik time series menggunakan metrik evaluasi yang sama. Oleh karena itu, metode yang digunakan dapat dijadikan acuan teknis.

Ridha et al. (2025) meneliti prediksi konsumsi energi listrik menggunakan LSTM. Data konsumsi energi dianalisis untuk mengetahui tren kebutuhan listrik masa depan. Model dilatih menggunakan beberapa konfigurasi parameter untuk mendapatkan hasil optimal.

Penelitian ini menunjukkan bahwa LSTM mampu mempelajari pola konsumsi jangka panjang serta fluktuasi musiman. Evaluasi menggunakan MAPE dan RMSE menunjukkan peningkatan akurasi dibanding metode statistik.

Relevansinya terletak pada kesamaan pendekatan eksperimental dalam mengoptimasi parameter model, yang juga dilakukan dalam penelitian prediksi harga saham ini.

Hayani et al. (2024) menerapkan LSTM untuk analisis sentimen komentar Youtube. Penelitian ini bertujuan memahami opini pengguna terhadap suatu konten. Model LSTM digunakan untuk klasifikasi berbasis urutan teks.

Hasil penelitian menunjukkan LSTM mampu mengolah data besar dengan performa stabil. Model efektif dalam memahami dependensi jangka panjang antar kata.

Penelitian ini mendukung penggunaan LSTM sebagai metode pembelajaran sekuens yang andal, yang prinsipnya sama dengan prediksi harga saham berbasis histori.

Fadhilah et al. (2025) memprediksi ketinggian air bendungan menggunakan LSTM dengan data curah hujan dan tinggi air sebelumnya. Penelitian bertujuan memberikan peringatan dini terhadap potensi banjir.

Hasil menunjukkan bahwa LSTM mampu memodelkan hubungan temporal yang kompleks dan menghasilkan prediksi akurat. Evaluasi menggunakan RMSE menunjukkan performa baik.

Penelitian ini sangat relevan karena sama-sama memprediksi data numerik berbasis waktu. Pendekatan LSTM yang digunakan menjadi referensi langsung dalam penelitian prediksi harga saham.

2.2 Landasan Teori

Landasan teori membahas berbagai teori yang digunakan dalam penelitian berikut ini, teori-teori dihimpun dan didapatkan dari berbagai buku, jurnal dan artikel yang terpercaya yang membahas, mendukung dan menjadi referensi dari penelitian kali ini.

2.2.1 Profil Singkat BBNI dan BMRI

PT Bank Negara Indonesia (Persero) Tbk atau BBNI merupakan salah satu bank milik negara tertua di Indonesia yang didirikan pada tahun 1946. Sebagai bank komersial nasional, BBNI memiliki peran strategis dalam mendukung pembiayaan sektor riil, layanan perbankan ritel, serta pembiayaan korporasi dan usaha kecil menengah. Dengan pengalaman operasional yang panjang, BBNI telah berkembang menjadi salah satu bank dengan aset besar dan jaringan layanan yang luas di seluruh Indonesia.

Dalam aktivitas perdagangan di pasar modal, saham BBNI termasuk kategori *blue chip stock* yang memiliki kapitalisasi pasar tinggi serta likuiditas transaksi yang aktif. Pergerakan harga saham BBNI cukup sensitif terhadap kondisi makroekonomi nasional, kebijakan suku bunga Bank Indonesia, serta sentimen investor domestik dan asing. Karakteristik ini menyebabkan volatilitas harga saham BBNI relatif dinamis dibandingkan beberapa saham perbankan lainnya.

Dari sudut pandang analisis data, saham BBNI menunjukkan fluktuasi harga harian yang cukup tajam dan responsif terhadap informasi pasar. Pola pergerakan

tersebut menghasilkan data deret waktu yang kompleks dan non-linear, sehingga menantang untuk dimodelkan menggunakan metode statistik konvensional. Oleh karena itu, pendekatan berbasis *Deep Learning* seperti LSTM menjadi relevan untuk menangkap pola temporal jangka pendek maupun jangka panjang.

Dalam penelitian ini, saham BBNI dipilih sebagai salah satu objek studi karena karakteristik volatilitasnya yang tinggi sehingga dapat menguji kemampuan model LSTM dalam menangani data dengan perubahan cepat. Selain itu, sensitivitas BBNI terhadap variasi *window size* dan *epoch* memberikan peluang untuk menganalisis pengaruh parameter model terhadap akurasi prediksi secara lebih komprehensif.

PT Bank Mandiri (Persero) Tbk atau BMRI merupakan bank milik negara yang berdiri pada tahun 1998 melalui penggabungan beberapa bank pemerintah. Bank Mandiri saat ini dikenal sebagai salah satu bank terbesar di Indonesia berdasarkan total aset, jumlah nasabah, dan jaringan operasional. Layanan yang diberikan mencakup perbankan ritel, korporasi, investasi, serta layanan digital yang terus berkembang.

Di pasar modal Indonesia, saham BMRI termasuk salah satu saham dengan kapitalisasi terbesar dan volume perdagangan harian yang sangat tinggi. Tingkat likuiditas yang besar mencerminkan partisipasi pelaku pasar yang luas, baik investor institusi maupun individu. Kondisi ini membuat harga saham BMRI cenderung lebih stabil dan mencerminkan mekanisme *supply and demand* pasar secara lebih efisien.

Secara karakteristik data, pergerakan harga saham BMRI menunjukkan tren yang relatif lebih halus dan konsisten dibandingkan BBNI. Fluktuasi yang lebih terkendali menyebabkan pola temporal lebih mudah dipelajari oleh model prediksi. Hal ini menjadikan BMRI sebagai objek yang baik untuk menguji performa model LSTM pada saham dengan volatilitas moderat dan likuiditas tinggi.

Dalam penelitian ini, BMRI dipilih sebagai pembanding terhadap BBNI untuk menganalisis bagaimana perbedaan karakteristik pasar memengaruhi akurasi model prediksi. Dengan membandingkan dua saham perbankan BUMN yang

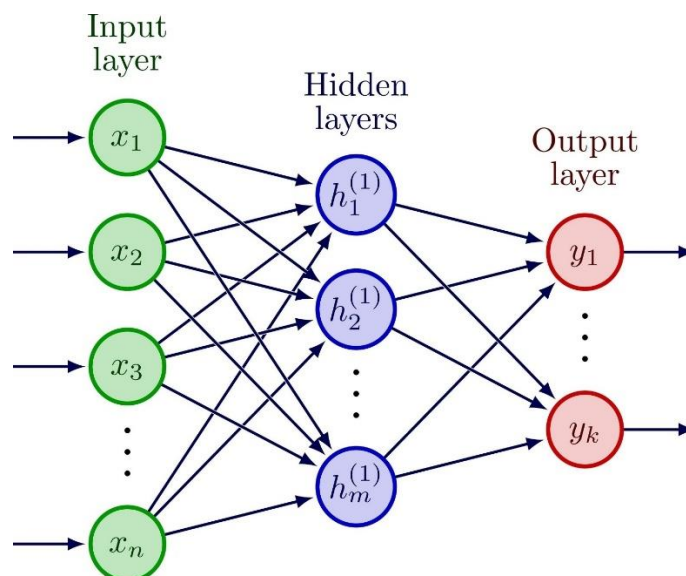
memiliki sifat volatilitas berbeda, penelitian ini dapat mengevaluasi efektivitas parameter LSTM secara lebih objektif dan komprehensif.

2.2.2 Deep Learning

Deep Learning adalah sub-bidang dari Machine Learning yang algoritmanya terinspirasi oleh struktur dan fungsi otak manusia yang disebut jaringan saraf tiruan (Artificial Neural Networks). Berbeda dengan Machine Learning tradisional, Deep Learning mampu memproses data mentah dalam jumlah besar dan mengekstraksi fitur secara otomatis melalui banyak lapisan (layers). Dalam konteks data saham yang bersifat runtun waktu (time series), di mana data saat ini sangat dipengaruhi oleh data masa lalu, metode Deep Learning seperti RNN dan LSTM menjadi pilihan utama karena kemampuannya menangkap pola sekuensial dan ketergantungan jangka panjang (Moghar & Hamiche, 2020).

2.2.3 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) adalah jenis arsitektur *Recurrent Neural Network* (RNN) yang dirancang khusus untuk mengatasi masalah *vanishing gradient* yang sering terjadi pada RNN standar ketika memproses data dengan rentang waktu yang panjang..



Gambar 2. 1Arsitektur LSTM

Arsitektur LSTM memiliki komponen inti yang disebut *Cell State* (memori sel) yang berfungsi membawa informasi di sepanjang proses sekuensial dengan sedikit modifikasi linier. Kemampuan LSTM untuk mengingat atau melupakan informasi

diatur oleh struktur yang disebut *Gates* (Gerbang):

1. Forget Gate (f_t): Menentukan informasi mana dari *cell state* sebelumnya yang harus dibuang atau dilupakan. Keputusan ini dibuat berdasarkan fungsi aktivasi *sigmoid*.
2. Input Gate (i_t): Menentukan informasi baru apa yang akan disimpan ke dalam *cell state*. Terdiri dari fungsi *sigmoid* yang menentukan nilai mana yang akan diperbarui, dan fungsi *tanh* yang membuat vektor nilai kandidat baru.
3. Output Gate (o_t): Menentukan apa yang akan menjadi *output* dari sel saat ini berdasarkan *cell state* yang telah diperbarui.

Secara matematis, proses dalam sel LSTM dapat dituliskan sebagai berikut:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

Rumus 2. 1LSTM secara Matematis

Dimana x_t adalah input pada waktu t , h_{t-1} adalah output dari langkah sebelumnya, dan C_t adalah status memori sel.

2.2.4 Hyperparameter LSTM

Dalam membangun model LSTM, terdapat parameter eksternal yang tidak dipelajari oleh model, melainkan harus diatur oleh peneliti sebelum proses pelatihan dimulai. Parameter ini disebut *Hyperparameter*. Penelitian ini memfokuskan pada dua *hyperparameter* utama:

1. Window Size (Time Steps):

Window Size atau sering disebut Look-back Period adalah jumlah langkah waktu (time steps) masa lalu yang digunakan model sebagai input untuk memprediksi satu langkah ke depan.

Contoh: Jika *Window Size* = 30, maka model akan melihat harga saham dari hari $t - 30$ sampai $t - 1$ untuk memprediksi harga pada hari t .

Pemilihan ukuran *window* sangat krusial; *window* yang terlalu kecil mungkin tidak menangkap tren musiman, sedangkan *window* yang terlalu besar dapat meningkatkan kompleksitas komputasi dan memasukkan *noise*.

2. Epoch:

Epoch adalah satu putaran penuh di mana algoritma pembelajaran telah melihat seluruh dataset pelatihan satu kali.

Underfitting: Jika jumlah *epoch* terlalu sedikit, model belum sempat mempelajari pola data dengan baik.

Overfitting: Jika jumlah *epoch* terlalu banyak, model akan mulai "menghafal" data latih, sehingga akurasi tinggi pada data latih namun buruk saat diuji dengan data baru.

2.2.5 Metrik Evaluasi Kinerja

Untuk mengukur seberapa akurat model LSTM dalam memprediksi harga saham, penelitian ini menggunakan dua metrik evaluasi standar:

1. Root Mean Squared Error (RMSE):

RMSE mengukur rata-rata besarnya kesalahan (error) model dengan memberikan bobot lebih pada kesalahan yang besar. Semakin kecil nilai RMSE, semakin akurat model tersebut.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad \text{Rumus 2. 2RMSE}$$

Dimana:

- a) n = jumlah data pengujian
- b) y_i = nilai aktual harga saham
- c) $\hat{y}_i$ = nilai prediksi harga saham

2. Mean Absolute Percentage Error (MAPE):

MAPE mengukur rata-rata persentase kesalahan absolut terhadap nilai aktualnya. Metrik ini memudahkan interpretasi karena hasilnya dalam bentuk persen.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad \text{Rumus 2. 3MAPE}$$

Nilai MAPE di bawah 10% dikategorikan sebagai kemampuan prediksi yang sangat baik.

2.2.6 Python

Python merupakan bahasa pemrograman tingkat tinggi yang sangat populer dalam pengembangan perangkat lunak, otomasi, dan analisis data ilmiah. Bahasa ini termasuk dalam kategori *general-purpose language*, artinya dapat digunakan untuk membangun berbagai jenis program dan tidak terbatas pada masalah tertentu saja. Menurut Peta (2022), popularitas Python yang tinggi didorong oleh sintaksisnya yang sederhana, serbaguna, dan mudah dipahami, menjadikannya pilihan utama bagi akademisi maupun profesional di bidang teknologi informasi.

Dalam penelitian ini, Python digunakan sebagai alat utama untuk membangun model prediksi harga saham BBNI dan BMRI. Python dipilih karena ekosistemnya yang kuat dalam mendukung implementasi *Deep Learning*, khususnya algoritma *Long Short-Term Memory* (LSTM). Berbeda dengan penelitian statistik konvensional, implementasi LSTM memerlukan pustaka khusus yang mampu menangani komputasi matriks dan jaringan saraf tiruan yang kompleks.

Kekuatan utama Python dalam penelitian ini terletak pada ketersediaan *library* (pustaka) yang komprehensif, antara lain:

1. Pandas dan NumPy: Digunakan untuk *preprocessing* data saham historis, seperti penanganan *missing values*, manipulasi *dataframe*, dan perhitungan matriks.
2. Scikit-Learn: Digunakan untuk normalisasi data (*MinMax Scaler*) dan evaluasi metrik kinerja model (RMSE dan MAPE).
3. Keras dan TensorFlow: Pustaka inti untuk membangun arsitektur jaringan saraf LSTM, mengatur lapisan (*layers*), serta menjalankan proses pelatihan (*training*) dengan variasi *epoch*.
4. Matplotlib dan Seaborn: Digunakan untuk visualisasi data, seperti grafik perbandingan harga aktual dengan harga prediksi, serta grafik *loss* selama proses pelatihan.

Dengan dukungan pustaka-pustaka tersebut, Python memungkinkan proses eksperimen variasi *Window Size* dan *Epoch* dilakukan secara efisien, terstruktur, dan akurat (Peta, 2022).

2.2.7 Karakteristik Data Time Series Finansial

Data harga saham termasuk dalam kategori data deret waktu (time series) yang memiliki ketergantungan temporal antar periode. Artinya, nilai harga pada suatu waktu dipengaruhi oleh nilai-nilai sebelumnya, baik dalam jangka pendek maupun jangka panjang. Tidak seperti data statis, data time series bersifat dinamis, berurutan, dan sangat dipengaruhi oleh faktor eksternal seperti kondisi ekonomi, kebijakan moneter, serta sentimen pasar.

Menurut Zeng et al. (2023), data time series finansial memiliki karakteristik kompleks berupa tren, musiman, siklus, dan noise yang saling tumpang tindih. Kondisi ini menyebabkan hubungan antar variabel tidak selalu linier dan sulit dimodelkan menggunakan pendekatan statistik klasik. Oleh karena itu, diperlukan metode yang mampu mempelajari pola non-linear dan dependensi jangka panjang secara adaptif.

Selain itu, data saham juga memiliki fenomena concept drift, yaitu perubahan pola distribusi data seiring waktu. Pola harga saham pada periode tertentu belum tentu relevan untuk periode berikutnya. Hal ini membuat model prediksi harus memiliki kemampuan generalisasi yang baik agar tetap akurat pada kondisi pasar yang berubah.

2.2.8 Volatilitas Saham dan Tantangan Prediksi

Volatilitas merupakan ukuran fluktuasi harga saham dalam periode tertentu dan mencerminkan tingkat ketidakpastian pasar. Saham dengan volatilitas tinggi cenderung mengalami perubahan harga yang tajam dalam waktu singkat, sehingga meningkatkan risiko kesalahan prediksi.

Chen et al. (2021) menyatakan bahwa volatilitas tinggi memperbesar noise dalam data time series, yang dapat menurunkan stabilitas model prediksi. Model yang terlalu sensitif terhadap perubahan kecil akan menghasilkan prediksi yang fluktuatif, sedangkan model yang terlalu kaku akan mengalami keterlambatan dalam merespons perubahan tren.

Dalam konteks saham perbankan BUMN seperti BBNI dan BMRI, volatilitas dipengaruhi oleh faktor makroekonomi, kebijakan suku bunga Bank

Indonesia, serta aliran dana asing. Perbedaan tingkat volatilitas antara BBNI dan BMRI menjadi alasan penting dilakukannya analisis perbandingan performa model LSTM pada kedua saham tersebut.

2.2.9 Non-Linearitas dan Noise pada Data Saham

Hubungan antara harga saham dan waktu tidak bersifat linier. Perubahan kecil pada suatu periode dapat menghasilkan dampak besar pada periode berikutnya, terutama saat terjadi peristiwa ekonomi tertentu. Selain itu, data saham mengandung noise yang berasal dari aktivitas spekulatif, rumor pasar, dan transaksi jangka pendek.

El Ouardi et al. (2024) menjelaskan bahwa algoritma deep learning seperti LSTM memiliki keunggulan dalam menangani non-linearitas karena mampu mempelajari representasi data secara hierarkis. Namun, keberadaan noise yang berlebihan dapat menyebabkan model belajar pola yang tidak relevan.

Oleh karena itu, pemilihan hyperparameter seperti window size menjadi krusial. Window yang terlalu panjang berpotensi memasukkan noise historis yang tidak lagi relevan, sedangkan window yang terlalu pendek tidak cukup merepresentasikan pola tren.

2.2.10 Overfitting dan Underfitting pada Model LSTM

Overfitting terjadi ketika model terlalu menyesuaikan diri dengan data latih sehingga kehilangan kemampuan generalisasi. Sebaliknya, underfitting terjadi ketika model belum cukup mempelajari pola data sehingga performanya rendah pada data latih maupun data uji.

Menurut Dong et al. (2025), jumlah epoch merupakan salah satu faktor utama yang memengaruhi risiko overfitting pada model LSTM. Penambahan epoch memang dapat meningkatkan akurasi pada tahap awal, namun setelah titik tertentu justru menyebabkan penurunan performa pada data uji.

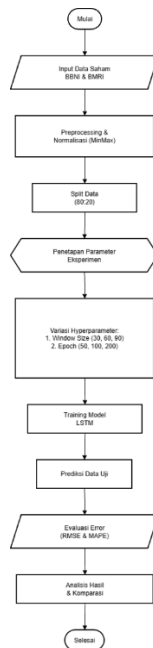
Untuk mengatasi permasalahan ini, teknik regularisasi seperti dropout serta pemantauan loss validasi sering digunakan. Dalam penelitian ini, analisis pengaruh epoch dilakukan untuk menemukan keseimbangan optimal antara proses pembelajaran dan kemampuan generalisasi model.

2.2.11 CRISP-DM (*Cross-Industry Standard Process for Data Mining*)

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) merupakan metodologi standar yang menyediakan alur kerja sistematis yang terdiri dari enam tahapan utama, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* (Chapman et al., 2000). Kerangka ini banyak digunakan dalam proyek machine learning baik pada ranah industri maupun penelitian akademik karena menekankan proses yang terstruktur, iteratif, serta terdokumentasi dengan baik sehingga meningkatkan transparansi dan replikasi penelitian (Wirth & Hipp, 2000). Selain itu, berbagai studi menunjukkan bahwa CRISP-DM tetap relevan dalam implementasi analitik modern, termasuk pada sektor finansial dan sistem prediktif berbasis data mining (Mariscal et al., 2010). Oleh karena itu, penelitian ini menggunakan CRISP-DM sebagai kerangka operasional dalam menyusun tahapan penelitian yang meliputi pengumpulan data, pembersihan data, pembentukan windowing, pemodelan LSTM, hingga evaluasi eksperimen.

2.3 Kerangka Pemikiran

Kerangka pemikiran dalam penelitian ini terdiri dari beberapa tahap yang dimulai dari pemahaman masalah hingga analisis hasil. Tahapan ini dijelaskan sebagai berikut:



Gambar 2. 2Kerangka Berfikir

Gambar di atas menunjukkan tahapan sistematis yang dilakukan dalam penelitian ini, yang terdiri dari beberapa fase utama:

1. Pengumpulan dan Persiapan Data

Mulai: Proses penelitian dimulai.

Input Data: Tahap awal adalah pengumpulan dataset historis harga saham BBNI dan BMRI yang diperoleh dari sumber publik Yahoo Finance.

Preprocessing & Normalisasi: Data mentah yang diperoleh kemudian melalui tahap *preprocessing* untuk membersihkan *noise*. Setelah itu, dilakukan Normalisasi menggunakan metode *MinMax Scaler* untuk mengubah skala data menjadi rentang 0 hingga 1, agar optimal saat diproses oleh algoritma LSTM.

Split Data: Dataset yang sudah bersih dibagi menjadi dua bagian dengan rasio 80:20, yaitu 80% sebagai data latih (*training data*) untuk melatih model, dan 20% sebagai data uji (*testing data*) untuk validasi.

2. Fase Eksperimen Parameter (Inti Penelitian)

Eksperimen Parameter: Pada tahap ini, penelitian masuk ke skenario pengujian untuk mencari konfigurasi terbaik. Terdapat dua cabang eksperimen utama:

Uji Window Size: Model dilatih dengan memvariasikan panjang jendela input (*window size*) sebesar 30, 60, dan 90 hari. Tujuannya untuk melihat seberapa jauh data masa lalu yang efektif untuk memprediksi harga besok.

Uji Epoch: Model dilatih dengan memvariasikan jumlah iterasi pelatihan (*epoch*) sebanyak 50, 100, dan 200 iterasi. Tujuannya untuk melihat pengaruh durasi pelatihan terhadap konvergensi model.

3. Pemodelan dan Prediksi

Training Model LSTM: Berdasarkan skenario parameter di atas, algoritma Long Short-Term Memory (LSTM) kemudian dilatih untuk mempelajari pola tren harga saham dari data latih.

Prediksi: Model yang telah terlatih digunakan untuk memprediksi harga saham pada data uji.

4. Evaluasi dan Analisis

Evaluasi Error: Hasil prediksi dibandingkan dengan harga aktual. Kinerja model diukur menggunakan dua metrik evaluasi, yaitu RMSE (*Root Mean Squared Error*) dan MAPE (*Mean Absolute Percentage Error*) untuk mengetahui tingkat kesalahan prediksi.

Analisis Hasil: Hasil evaluasi dari setiap skenario (variasi Window Size dan Epoch) dibandingkan dan dianalisis untuk menentukan kombinasi parameter mana yang menghasilkan akurasi tertinggi (error terendah).