

BAB I PENDAHULUAN

1.1 Latar Belakang

Di era modern ini, kesadaran akan pentingnya perencanaan keuangan semakin meningkat, mendorong minat masyarakat terhadap investasi, khususnya di pasar modal. Saham menjadi salah satu instrumen investasi yang populer karena potensi keuntungannya yang tinggi meskipun memiliki risiko yang besar. Berdasarkan data Kustodian Sentral Efek Indonesia (KSEI), jumlah investor di pasar modal terus meningkat, mencapai 13,98 juta per Desember 2024. Tren ini menunjukkan bahwa masyarakat Indonesia semakin tertarik berinvestasi, yang berdampak signifikan terhadap perekonomian Indonesia.

Saham adalah surat berharga yang membuktikan kepemilikan modal atau penyertaan modal dalam suatu perusahaan. Saham menjadi salah satu produk pasar keuangan dan investasi yang paling diminati oleh investor. Dengan meningkatnya partisipasi masyarakat dalam investasi saham, diperlukan strategi yang lebih baik dalam menganalisis pergerakan harga saham untuk meminimalkan risiko dan memaksimalkan keuntungan (Darmadji & Fakhruddin, 2008).

Pasar saham menawarkan peluang keuntungan yang besar, tetapi juga disertai dengan risiko tinggi. Risiko yang dihadapi investor mencakup penurunan harga saham dan kemungkinan likuidasi. Pergerakan harga saham dipengaruhi oleh berbagai faktor, seperti dinamika penawaran dan permintaan, tingkat suku bunga, inflasi, serta kinerja perusahaan. Oleh karena itu, sebelum membeli saham, investor perlu melakukan analisis menyeluruh untuk memaksimalkan potensi keuntungan. Salah satu cara untuk mendukung pengambilan keputusan investasi adalah dengan melakukan prediksi. Prediksi merupakan proses memperkirakan kejadian di masa depan berdasarkan data historis dan kondisi saat ini (Emanuella et al., 2022).

Saham sektor teknologi menjadi salah satu pilihan investasi yang menarik karena pertumbuhannya yang pesat seiring dengan perkembangan digitalisasi di Indonesia. Saham sektor teknologi masih tergolong baru dibandingkan sektor-sektor lainnya. Perusahaan teknologi memiliki peran penting dalam ekonomi modern, terutama dengan meningkatnya adopsi layanan digital, e-commerce, dan teknologi finansial. Selain itu, saham sektor teknologi cenderung memiliki volatilitas tinggi dibandingkan sektor lainnya, yang membuatnya menarik bagi investor yang mencari peluang keuntungan besar, namun juga memiliki risiko yang lebih tinggi. Tak heran jika saham sektor teknologi menjadi incaran para investor.

Beberapa saham sektor teknologi yang sering didengar seperti GOTO (GoTo) atau BUKA (Bukalapak.com Tbk) mencerminkan dinamika pasar saham teknologi di Indonesia. Selain itu, terdapat pula saham lain seperti BELI (Global Bio Farma Tbk), ELANG (Elang Mahkota Teknologi Tbk), dan WIRG (WIR Asia Tbk). Saham-saham teknologi ini menarik bagi investor karena memiliki potensi pertumbuhan yang tinggi seiring dengan perkembangan teknologi dalam kehidupan sehari-hari, baik



dalam e-commerce, layanan digital, maupun solusi berbasis teknologi lainnya. Namun, volatilitas harga saham di sektor ini juga cukup tinggi, dipengaruhi oleh faktor seperti persaingan pasar, inovasi teknologi, kebijakan pemerintah, serta kondisi ekonomi global. Oleh karena itu, diperlukan metode prediksi yang akurat untuk menganalisis pergerakan harga saham sektor teknologi guna membantu investor dalam mengelola risiko dan mengambil keputusan investasi yang lebih baik.

Metode prediksi ada berbagai macam, seperti metode statistik yang biasa dikenal dengan *Autoregressive Integrated Moving Average* (ARIMA), yang mampu menangkap pola waktu dalam data historis. ARIMA bekerja dengan mengidentifikasi tren dan pola musiman berdasarkan data masa lalu, sehingga sering digunakan dalam analisis deret waktu keuangan. Namun, model ini memiliki keterbatasan dalam menangani data dengan pola yang kompleks dan non-linear, seperti yang sering terjadi pada saham sektor teknologi yang memiliki volatilitas tinggi. Di sisi lain, terdapat pendekatan berbasis deep learning seperti Long Short-Term Memory (LSTM) dan Bidirectional Long Short-Term Memory (BLSTM) yang menawarkan keunggulan dalam menangani pola yang lebih kompleks. LSTM adalah jenis jaringan saraf berulang (RNN) yang dirancang untuk mengatasi masalah long-term dependencies dalam data deret waktu. LSTM dapat mengenali pola dalam data historis dan membuat prediksi lebih akurat dibandingkan metode statistik tradisional, terutama untuk data dengan karakteristik non-linear dan fluktuasi tinggi. Sementara itu, BLSTM merupakan pengembangan dari LSTM yang memproses data dalam dua arah, yaitu maju (forward) dan mundur (backward), sehingga mampu menangkap lebih banyak informasi kontekstual dari data historis.

State of the Art dalam penelitian ini terletak pada perbandingan antara model ARIMA, LSTM, dan BLSTM dalam memprediksi harga saham sektor teknologi (IDXTECHNO) di Indonesia. Meskipun ARIMA telah banyak digunakan dalam prediksi harga saham, penelitian ini ingin membandingkannya dengan model deep learning yang lebih fleksibel dalam menangani pola data yang lebih kompleks. Dengan melakukan perbandingan antara ketiga model tersebut, penelitian ini diharapkan dapat memberikan wawasan baru dalam pemilihan metode prediksi harga saham yang lebih akurat dan efektif. Selain itu, dalam penelitian ini diasumsikan bahwa pemilihan saham sektor teknologi (IDXTECHNO) tidak didasarkan pada harga saham atau nilai nominalnya, melainkan pada jumlah saham yang beredar di pasar. Hal ini bertujuan agar analisis lebih fokus pada dinamika pergerakan harga saham tanpa dipengaruhi bias harga saham tertentu.

Berdasarkan uraian latar belakang dan penelitian sebelumnya, peneliti tertarik untuk melakukan penelitian yang berjudul "ANALISIS PERBANDINGAN MODEL



BLSTM DALAM MEMPREDIKSI HARGA SAHAM TERENDAH GI (IDXTECHNO) DI INDONESIA". Penelitian ini diharapkan dalam memberikan rekomendasi model prediksi yang lebih lan pelaku pasar modal dalam mengambil keputusan investasi tor teknologi.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana performa model ARIMA, LSTM, dan BLSTM dalam memprediksi harga saham terendah sektor teknologi (IDXTECHNO) di Indonesia?
2. Model manakah yang memiliki akurasi terbaik dalam memprediksi harga saham terendah sektor teknologi (IDXTECHNO) di Indonesia?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menganalisis performa model ARIMA, LSTM, dan BLSTM dalam memprediksi harga saham terendah sektor teknologi (IDXTECHNO) di Indonesia.
2. Menentukan model terbaik dalam memprediksi harga saham terendah sektor teknologi (IDXTECHNO) berdasarkan akurasi prediksi yang dihasilkan.

1.3.2 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Bagi Penulis, dapat menambah wawasan dan pemahaman tentang model prediksi harga saham, khususnya ARIMA, LSTM, dan BLSTM.
2. Bagi Pembaca, dapat menjadi referensi dalam mencari metode alternatif untuk prediksi harga saham, terutama di sektor teknologi.
3. Bagi Universitas, dapat menjadi acuan dalam pengembangan penelitian di bidang kecerdasan buatan dan pemodelan prediktif.
4. Bagi Investor, dapat membantu dalam pengambilan keputusan investasi dengan memberikan informasi mengenai metode prediksi yang lebih akurat untuk saham sektor teknologi.

1.4 Landasan Teori

1.4.1 Saham

Saham merupakan instrumen keuangan yang menunjukkan kepemilikan modal seorang investor dalam perusahaan penerbit saham. Saham hanya dapat diterbitkan oleh emiten, yaitu perusahaan yang berbadan hukum Perseroan Terbatas (PT). Dengan berinvestasi dalam saham suatu perusahaan, investor secara otomatis menjadi bagian dari pemilik perusahaan dan memiliki hak untuk memberikan suara



usaha (Lestari dan Jasuni, 2023). Investasi, saham memiliki potensi keuntungan yang tinggi, tetapi risiko yang besar dibandingkan instrumen investasi lainnya. Hal ini terjadi akibat dinamika penawaran dan permintaan, yang menyebabkan fluktuasi harga secara cepat (Tenggehi et al., 2022). Pergerakan harga saham memiliki karakteristik yang kompleks, yaitu bersifat non-linear, non-

stationary, dan memiliki long memory dependency. Faktor utama yang memengaruhi pergerakan saham meliputi kondisi ekonomi, kinerja perusahaan, serta psikologi pasar. Faktor-faktor tersebut berkontribusi terhadap volatilitas saham, menjadikannya instrumen investasi dengan tingkat ketidakpastian yang tinggi (Afif et al., 2021).

1.4.2 Prediksi

Peramalan atau prediksi adalah proses sistematis dalam memperkirakan kejadian yang paling mungkin terjadi di masa depan berdasarkan data historis dan kondisi saat ini. Dalam konteks investasi saham, prediksi memainkan peran penting dalam menganalisis pergerakan harga saham untuk membantu investor dalam mengambil keputusan yang lebih akurat (Emanuella et al., 2022). Setiap investor maupun perusahaan tidak dapat mengetahui secara pasti bagaimana pergerakan harga saham di masa mendatang, sehingga diperlukan strategi yang tepat untuk mengantisipasi volatilitas pasar.

Metode prediksi sangat krusial dalam pengambilan keputusan investasi karena dapat membantu investor memahami ketidakpastian pasar saham dan mengurangi potensi risiko kerugian. Dengan menggunakan teknik prediksi yang sesuai, investor dapat memperoleh gambaran mengenai tren harga saham yang lebih akurat, sehingga dapat memaksimalkan keuntungan dan meminimalkan kerugian (Muttaqin et al., 2015). Pemilihan metode prediksi yang tepat bergantung pada karakteristik data saham yang digunakan serta tujuan investasi yang ingin dicapai.

1.4.3 Time Series

Deret waktu (time series) merupakan kumpulan data yang dicatat secara berurutan berdasarkan waktu dan mencerminkan karakteristik kuantitatif dari suatu kejadian dalam periode tertentu. Dalam konteks pasar saham, data deret waktu sering digunakan untuk menganalisis pergerakan harga saham guna mengidentifikasi tren dan pola yang dapat membantu dalam pengambilan keputusan investasi (Chandra dan Setia, 2020).

Analisis deret waktu adalah metode statistik yang digunakan untuk mengolah data historis suatu aset keuangan, seperti saham, yang memiliki pola pergerakan harga tertentu. Tujuan utama analisis ini adalah mengidentifikasi fluktuasi harga di masa lalu guna memperkirakan nilai di masa mendatang, sehingga investor dapat mengurangi risiko dan memaksimalkan keuntungan. Salah satu teknik yang sering digunakan dalam analisis ini adalah peramalan, yaitu proses memperkirakan harga

berdasarkan pola historis. Dalam proses peramalan, faktor kondisi ekonomi, kebijakan pemerintah, dan tingkat suku bunga akan karena dapat memengaruhi harga saham (Chandra dan



Peramalan berbasis deret waktu dapat dilakukan dengan dua pendekatan utama, yaitu analisis univariat dan multivariat. Pada analisis univariat, pergerakan harga saham diprediksi hanya berdasarkan data historis harga itu sendiri, sedangkan pada analisis multivariat, prediksi mempertimbangkan lebih dari satu faktor, seperti volume perdagangan, inflasi, atau sentimen pasar. Dengan memanfaatkan metode yang tepat, analisis deret waktu dapat menjadi alat yang efektif bagi investor dalam mengambil keputusan investasi yang lebih cerdas dan strategis di pasar saham (Fitrinanda et al., 2022).

1.4.4 Data Pre-processing

Persiapan data atau data pre-processing merupakan tahapan dalam mengubah data mentah menjadi format yang lebih terstruktur agar lebih mudah dipahami dan siap untuk dimodelkan. Langkah ini sangat penting karena data mentah sering kali tidak memiliki format yang baik. Proses ini melibatkan beberapa tahap yang dirancang secara sistematis untuk meningkatkan kualitas data serta memastikan bahwa data siap digunakan dalam proses pelatihan dan evaluasi model. Berikut adalah beberapa langkah dalam mempersiapkan data sebelum digunakan dalam model.

1.4.4.1 *Missing Value*

Dalam analisis harga saham, penanganan missing value dalam tahap pre-processing sangat penting untuk menjaga kualitas data dan meningkatkan akurasi model prediksi. Salah satu metode yang digunakan adalah menghapus data yang hilang jika jumlahnya relatif kecil dan tidak mempengaruhi pola keseluruhan. Data yang tidak lengkap dapat menyebabkan bias dalam prediksi pergerakan harga, terutama pada sektor teknologi (IDXTECHNO) yang memiliki volatilitas tinggi. Dengan menghapus data yang hilang, model dapat bekerja lebih optimal tanpa terganggu oleh inkonsistensi dalam dataset.

1.4.4.2 *Extract Date*

Dalam analisis harga saham, proses extract date atau ekstraksi tanggal sangat penting untuk meningkatkan keterbacaan dan pemanfaatan data waktu dalam model prediksi. Dengan mengubah variabel date menjadi indeks, data dapat diorganisir dengan lebih sistematis, sehingga mempermudah analisis pola pergerakan harga saham. Selain itu, representasi waktu pada sumbu x dalam grafik akan lebih jelas, memungkinkan investor atau analis untuk mengidentifikasi tren serta pola fluktuasi harga saham dengan lebih efektif.

1.4.4.3 *Pembagian Data*



Sebelum prediksi harga saham, pembagian data merupakan langkah penting untuk memastikan model dapat belajar dari pola historis dan diuji dengan data yang belum pernah dilihat sebelumnya. Data yang digunakan akan dibagi menjadi dua bagian utama, yaitu train data dan test data. Train data digunakan untuk melatih model agar dapat mengenali pola dalam data historis, sedangkan test data digunakan untuk mengevaluasi performa model terhadap data baru.

yang belum pernah dipelajari sebelumnya. Pembagian ini bertujuan untuk menghindari overfitting, yaitu kondisi di mana model terlalu menyesuaikan diri dengan data pelatihan sehingga kurang mampu melakukan generalisasi terhadap data baru.

Pada penelitian ini, data akan dibagi dengan rasio 80% untuk train data dan 20% untuk test data (Puteri, 2023). Dengan rasio ini, model diharapkan memiliki cukup data untuk belajar secara optimal, sekaligus tetap memiliki data yang cukup untuk menguji performanya dalam memprediksi harga saham sektor teknologi (IDXTECHNO).

1.4.4.4 Normalisasi Data

Normalisasi data bertujuan untuk memastikan keseragaman skala di seluruh fitur dalam dataset. Salah satu metode yang sering digunakan adalah Min-Max Scaling, yang mengubah nilai data harga penutupan ke dalam rentang tertentu, umumnya antara 0 dan 1, tanpa mengubah hubungan relatif antar data (Pipin, 2023).

Adapun persamaan untuk *Min-Max Scaling* adalah sebagai berikut:

$$x_{norm,i(t)} = \frac{x_{i(t)} - x_{i(min)}}{x_{i(max)} - x_{i(min)}} \quad (1)$$

Dengan:

- $x_{norm,i(t)}$ = Harga saham i periode ke- t setelah dinormalisasi
- $x_{i(t)}$ = Harga saham aktual saham i periode ke- t
- $x_{i(max)}$ = Harga tertinggi saham i
- $x_{i(min)}$ = Harga terendah saham i

1.4.4.5 Denormalisasi Data

Denormalisasi adalah langkah untuk mengembalikan hasil prediksi yang telah dinormalisasi ke skala aslinya. Dalam model yang menggunakan teknik normalisasi, hasil prediksi sering kali berada dalam rentang 0-1, sehingga perlu dikonversi kembali ke skala harga saham yang sebenarnya.

Proses denormalisasi menggunakan persamaan berikut (Budiprasetyo et al., 2022 dalam Puteri, 2023):

$$d_{i(t)} = \hat{y}_{i(t)} \times (x_{i(max)} - x_{i(min)}) + x_{i(min)} \quad (2)$$



iksi saham i periode ke- t sebelum denormalisasi
iksi saham i periode ke- t setelah denormalisasi

1.4.5 Model Time Series

Dalam analisis data keuangan, khususnya pergerakan harga saham, metode deret waktu (time series) menjadi pendekatan yang umum digunakan untuk memprediksi tren di masa depan. Model time series bekerja dengan menganalisis pola historis dari suatu data berdasarkan urutan waktu, sehingga dapat mengidentifikasi tren, musiman, dan pola fluktuatif lainnya. Salah satu metode yang banyak digunakan dalam peramalan deret waktu adalah *Autoregressive Integrated Moving Average* (ARIMA), yang merupakan kombinasi dari beberapa konsep statistik, seperti *Autoregressive* (AR), *Moving Average* (MA), dan *Autoregressive Moving Average* (ARMA).

1.4.5.1 Autoregressive (AR)

Konsep *Autoregressive* (AR) pertama kali diperkenalkan oleh Yule pada tahun 1926 dan kemudian disempurnakan lebih lanjut oleh Walker pada tahun 1931. Model ini menggambarkan bagaimana nilai saat ini dari suatu variabel dapat dipengaruhi oleh nilai sebelumnya dalam suatu pola regresi. Dengan kata lain, model AR menggunakan hubungan linier antara observasi masa lalu dan nilai saat ini untuk membuat prediksi ke depan. Model AR dengan orde p dinotasikan $AR(p)$ atau $ARIMA(p,0,0)$.

Bentuk umum dari model ini adalah sebagai berikut (Alhafiz, 2023):

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (3)$$

Atau, dapat dinyatakan dalam persamaan berikut:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) y_t = C + \varepsilon_t \quad (4)$$

$$\phi_p(B) y_t = C + \varepsilon_t \quad (5)$$

Dengan:

- p = Orde model *Autoregressive*
- C = Konstanta
- ϕ = Parameter model *Autoregressive*
- B = Operator Lag
- $\phi_p(B)$ = Polinomial lag untuk komponen AR
- ε_t = Residual pada periode ke- t



Moving Average (MA)

Metode *Moving Average* (MA) merupakan metode yang pertama kali diterapkan oleh Slutsky. Model ini berfokus pada ketergantungan nilai saat ini terhadap nilai rata-rata dari periode sebelumnya. Dalam pendekatan MA, fluktuasi historis diperhitungkan untuk menghasilkan estimasi yang lebih

akurat terhadap pola waktu tertentu (Slutzky, 1937). Model MA dengan orde q dinotasikan MA(q) atau ARIMA(0,0, q).

Bentuk umum dari model ini adalah sebagai berikut:

$$y_t = C + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (6)$$

Atau, dapat dinyatakan dalam persamaan berikut:

$$y_t = C + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (7)$$

$$y_t = C + \theta_q(B) \varepsilon_t \quad (8)$$

Dengan:

q = Orde model *moving average*

θ = Parameter model *moving average*

$\theta_q(B)$ = Polinomial lag untuk komponen MA

1.4.5.3 Autoregressive Moving Average (ARMA)

Fondasi teoritis dari kombinasi antara model AR dan MA dikembangkan oleh Wold pada tahun 1938. Wold merancang dasar-dasar dalam proses ARMA, yang menggabungkan elemen regresi dari AR dan ketergantungan residual dari MA untuk membentuk suatu model yang lebih fleksibel dalam menganalisis data deret waktu. Pengembangan lebih lanjut dari model ARMA meliputi prosedur identifikasi efisien, teknik estimasi untuk proses AR, MA, dan ARMA campuran, serta aplikasi dalam analisis deret waktu musiman (Wold, 1938). Model ARMA dengan orde p dan q dinotasikan ARMA(p,q) atau ARIMA($p,0,q$). Secara umum, model ini dapat dinyatakan sebagai berikut (Pratami et al., 2023):

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (9)$$

Atau, dapat dinyatakan dalam persamaan berikut:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) y_t = C + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (10)$$

$$\phi_p(B) y_t = C + \theta_q(B) \varepsilon_t \quad (11)$$

1.4.5.4 Autoregressive Integrated Moving Average (ARIMA)



akan pengembangan lebih lanjut dari ARMA, yang dirancang untuk data non-stasioner. Wold (1938) mengembangkan konsep dasar untuk menyusun pendekatan yang mencakup transformasi data menjadi stasioner, yang kemudian diterapkan ke dalam model ARMA. ARIMA menjadi salah satu model yang banyak digunakan dalam analisis deret waktu karena kemampuannya menangkap pola tren dan musiman dengan lebih baik.

dibandingkan model statistik lainnya. Model ARIMA adalah kombinasi dari tiga komponen utama, yaitu *Autoregressive* (p), *integrated* (d), dan *moving average* (q). Oleh karena itu, bentuk umum model ARIMA (p,d,q) dapat dijelaskan sebagai berikut (Pratami et al., 2023):

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \Delta^d y_t \quad (12)$$

$$\Delta^d y_t = (1 - B)^d y_t \quad (13)$$

Dengan:

d = Orde *differencing*

$\Delta^d y_t$ = Nilai *differencing* ke- d

Atau, dapat dinyatakan dalam persamaan berikut:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d y_t = C + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (14)$$

$$\phi_p(B)(1 - B)^d y_t = C + \theta_q(B) \varepsilon_t \quad (15)$$

1.4.6 Neural Network

Jaringan Saraf Tiruan (Neural Network/NN) adalah metode komputasi yang meniru fungsi saraf manusia, terdiri dari lapisan masukan (input layer) dan keluaran (output layer), dengan neuron yang menggunakan fungsi aktivasi untuk menentukan outputnya (Ahmad, 2017). Untuk meningkatkan kinerja NN, dapat ditambahkan lapisan tersembunyi (hidden layer), dan pelatihan menggunakan lebih banyak data dapat meningkatkan akurasi. Namun, model dasar NN hanya efektif untuk regresi dan klasifikasi serta kurang mampu menghubungkan informasi baru dengan data sebelumnya (Karno, 2020).

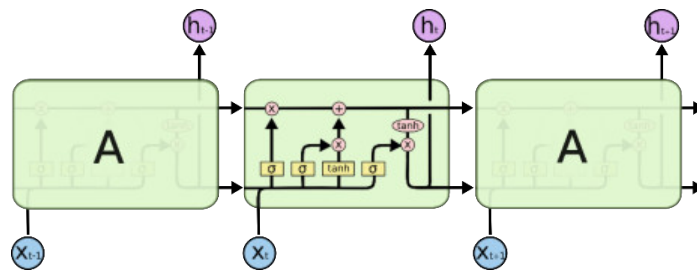
Untuk menangani data deret waktu, digunakan Recurrent Neural Network (RNN), tetapi model ini memiliki keterbatasan dalam mengingat informasi jangka panjang akibat masalah vanishing gradient. Oleh karena itu, dikembangkan Long Short-Term Memory (LSTM), yang memiliki mekanisme gerbang kompleks untuk meningkatkan kapasitas memorinya, sehingga lebih efektif dalam analisis data deret waktu (Hochreiter & Schmidhuber, 1997 dalam Karno, 2020).

1.4.7 Long-Short Term Memory

Long Short-Term Memory (LSTM) adalah varian dari Recurrent Neural Network (RNN) yang diperkenalkan oleh Sepp Hochreiter dan Jürgen Schmidhuber pada



i dirancang untuk mengatasi masalah long-term dependencies (2014 dalam Fauzi, 2019). Dalam prediksi harga saham IHSG, LSTM digunakan untuk mengidentifikasi pola data historis guna menghasilkan prediksi yang akurat. Struktur dari LSTM dapat diamati pada Gambar 1.



Gambar 1. Struktur LSTM

Sumber: Christoper, 2015. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Model LSTM dirancang untuk menyaring informasi melalui mekanisme gerbang yang berfungsi dalam mempertahankan serta memperbarui kondisi *memory cell*. Struktur gerbang dalam LSTM terdiri dari *input gate*, *forget gate*, dan *output gate*. Setiap *memory cell* memiliki tiga lapisan aktivasi sigmoid dan satu lapisan aktivasi tanh. Fungsi dari sigmoid dapat dirumuskan sebagai berikut (Aulia, 2020):

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (16)$$

Sedangkan fungsi aktivasi tanh dinyatakan dengan persamaan berikut:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (17)$$

Jika fungsi tanh pada persamaan (17) dihubungkan dengan fungsi sigmoid pada persamaan (16), maka hubungan tersebut dapat dinyatakan melalui persamaan berikut:

$$\tanh(x) = 2\sigma(2x) - 1 \quad (18)$$

Dengan:

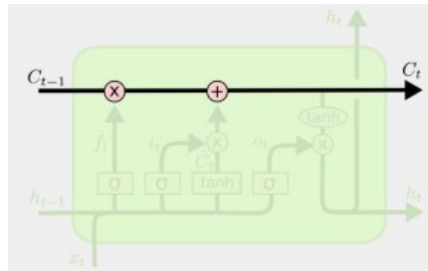
$\sigma(x)$ = Fungsi aktivasi sigmoid

x = Data *input*



Konsep utama dalam LSTM adalah adanya jalur koneksi antara *memory cell* dengan *memory cell* yang baru (C_t). *Memory cell* berperan yang menghubungkan seluruh *output layer* pada bagian atas ini memungkinkan nilai dari *memory cell* lama diteruskan ke bagian sedikit perubahan sesuai dengan kebutuhan model (Aulia,

2020). Representasi *memory cell* berdasarkan Christopher (2015) dapat dilihat pada Gambar 2.

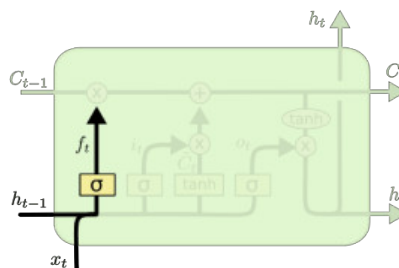


Gambar 2. Cell state

Sumber: Christopher, 2015. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

LSTM mampu menambah atau menghapus informasi dalam *cell state* melalui mekanisme yang dikendalikan oleh struktur yang disebut *gates*. *Gates* ini berfungsi untuk menentukan informasi mana yang akan diteruskan. Komponen tersebut terdiri dari lapisan jaringan saraf sigmoid serta operasi perkalian cepat. Output dari lapisan sigmoid berupa nilai antara 0 dan 1, yang menunjukkan seberapa banyak informasi yang perlu dipertahankan. Jika nilainya 0, maka informasi tidak akan diteruskan, sedangkan nilai 1 berarti informasi sepenuhnya diteruskan (Aulia, 2020).

Menurut Christopher (2015), proses dalam LSTM terdiri dari beberapa tahapan. Tahap awal dalam LSTM adalah menentukan informasi yang harus dihapus dari *cell state*. Keputusan ini dibuat oleh lapisan sigmoid yang dikenal sebagai *forget gate layer*. Lapisan ini menerima masukan dari h_{t-1} dan x_t , lalu menghasilkan output berupa angka antara 0 hingga 1 untuk setiap elemen dalam *cell state* (C_{t-1}). Nilai 0 menunjukkan bahwa informasi tersebut harus sepenuhnya dihapus, sedangkan nilai 1 menunjukkan bahwa informasi tersebut harus tetap dipertahankan. Representasi dari *forget gate* dapat diamati pada Gambar 3.



Gambar 3. Struktur LSTM forget gate

2015. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>



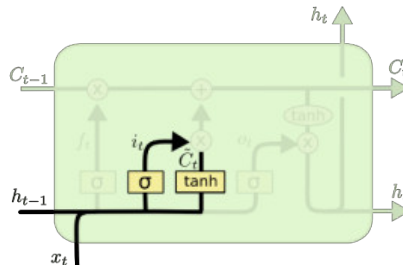
Dimana, f_t dapat dirumuskan sebagai berikut:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (19)$$

Dengan:

- f_t = Forget gate ke- t
- W_f = Nilai *weight* untuk *forget gate*
- h_{t-1} = Nilai hidden state orde ke- $(t-1)$
- x_t = *Input* pada orde ke- t
- b_f = Nilai bias pada *forget gate*

Tahap berikutnya dalam mekanisme kerja LSTM adalah menentukan informasi yang akan disimpan dalam *cell state*. Proses ini terdiri dari dua langkah utama. Pertama, lapisan sigmoid yang disebut sebagai *input gate* (i_t) berfungsi untuk mengidentifikasi nilai-nilai yang perlu diperbarui dalam *cell state*. Kedua, lapisan tanh membentuk vektor kandidat baru ($\tilde{C}_t$) yang nantinya akan ditambahkan ke *cell state*. Gabungan dari kedua langkah ini menghasilkan pembaruan pada *cell state*. *Input gate* memiliki peran penting dalam mengatur sejauh mana informasi dari input pada waktu tertentu (x_t) akan disimpan dalam *cell state* (C_t). Mekanisme ini berfungsi untuk menyaring informasi yang masuk, sehingga hanya data yang relevan yang dipertahankan dalam memori, sementara informasi yang tidak signifikan dapat diabaikan. Representasi dari *input gate* dapat diamati pada Gambar 4.



Gambar 4. Struktur LSTM *input gate*

Sumber: Christoper, 2015. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Dalam proses ini, *input gate* bekerja melalui persamaan berikut:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (20)$$



ke- t
 t untuk *input gate*
pada *input gate*

Setelah nilai *input gate* dihitung, lapisan tanh membuat vektor kandidat baru untuk *cell state* ($\tilde{C}_t$) melalui persamaan berikut:

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (21)$$

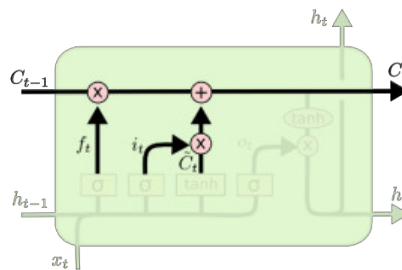
Dengan:

$\tilde{C}_t$ = Nilai baru yang dapat ditambahkan ke *cell state*

W_c = Nilai *weight* untuk *cell state*

b_c = Nilai bias pada *cell state*

Langkah berikutnya adalah memperbarui *cell state* dari keadaan sebelumnya (C_{t-1}) menjadi keadaan baru (C_t). Proses ini dilakukan dengan mengalikan *cell state* lama dengan f_t , sehingga informasi yang tidak lagi diperlukan dapat dihilangkan. Setelah itu, nilai baru diperoleh dengan menambahkan hasil perkalian antara i_t dan $\tilde{C}_t$. Nilai ini kemudian disesuaikan sesuai dengan keputusan pembaruan pada setiap elemen *state*. Representasi dari *update cell state* dapat diamati pada Gambar 5.



Gambar 5. Struktur LSTM pembaruan *cell state*

Sumber: Christoper, 2015. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Dapat dirumuskan secara matematis sebagai berikut:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (22)$$

Dengan:

C_t = *Cell state* orde ke- t

C_{t-1} = *Cell state* sebelum orde ke- t



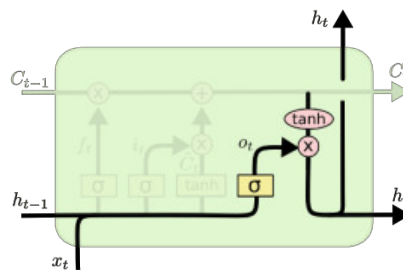
Optimized using
trial version
www.balesio.com

di ($\odot$) adalah operasi biner antara dua vektor (atau matriks) yang identik. Operasi ini menghasilkan sebuah vektor baru dengan di mana setiap elemennya merupakan hasil perkalian dari masing-masing elemen pada kedua vektor awal.

Salah satu LSTM terletak pada kemampuannya menyimpan informasi yang diterima melalui unit memori yang diperluas. Hal ini memungkinkan

model untuk mengatasi masalah vanishing gradient yang sering terjadi pada RNN saat memproses data sekuensial panjang (Pontoh et al., 2022 dalam Puteri, 2023). Cell state dalam LSTM berfungsi sebagai jalur utama yang menghubungkan informasi dari lapisan sebelumnya (C_{t-1}) ke lapisan berikutnya (C_t), yang memastikan aliran informasi yang lebih stabil.

Tahap terakhir dalam proses ini adalah menentukan output yang dihasilkan. Output tersebut bergantung pada *cell state* yang telah melalui proses penyaringan. Pertama, lapisan aktivasi sigmoid akan menentukan bagian dari *cell state* yang akan digunakan sebagai output. Setelah itu, *cell state* akan diproses melalui fungsi aktivasi *tanh*, yang akan menyesuaikan nilainya dalam rentang -1 hingga 1. Hasil ini kemudian dikalikan dengan output dari gerbang sigmoid untuk menghasilkan keluaran akhir. Representasi dari *output gate* dapat diamati pada Gambar 6.



Gambar 6. Struktur LSTM output gate

Sumber: Christopher, 2015. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Output gate pertama-tama menghitung nilai sigmoid menggunakan persamaan berikut:

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (23)$$

Dengan:

- O_t = Nilai *output gate*
- W_o = Nilai *weight* untuk *output gate*
- b_o = Nilai bias pada *output gate*

Kemudian, hasil akhir dari *cell state* (C_t) yang telah diproses menggunakan fungsi aktivasi *tanh* dikalikan dengan O_t untuk memperoleh nilai output akhir dari sel.

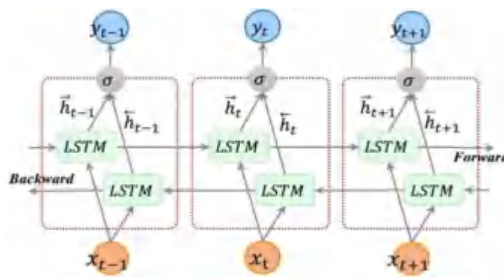


$$h_t = O_t \odot \tanh(C_t) \quad (24)$$

Output gate mengombinasikan informasi dari *cell state* dan *output gate* untuk menentukan informasi yang paling relevan pada setiap langkah (Pontoh et al., 2024).

1.4.8 Bidirectional Long-Short Term Memory

Bidirectional Long-Short Term Memory (BiLSTM) merupakan pengembangan dari model LSTM yang dirancang untuk mengatasi keterbatasannya dengan memungkinkan pemrosesan data dalam dua arah (Alghifari et al., 2022). BiLSTM terdiri dari dua jaringan LSTM, di mana jaringan pertama memproses data dari awal ke akhir, sedangkan jaringan kedua memproses data dari akhir ke awal. Hasil keluaran dari kedua jaringan tersebut kemudian digabungkan pada setiap langkah waktu, sehingga model dapat menangkap informasi dari masa lalu maupun masa depan untuk setiap sequence input (Puteri, 2023). Dengan adanya dua lapisan yang bekerja secara berlawanan arah, BiLSTM mampu memberikan pemahaman kontekstual yang lebih baik dibandingkan LSTM standar.



Gambar 7. Struktur BiLSTM

Sumber: Puteri, 2023.

Proses pada *forward* LSTM dapat dituliskan sebagai berikut:

$$\vec{h}_t = LSTM[x_t, h_{t-1}] \quad (25)$$

Selanjutnya untuk proses *backward* LSTM dapat dituliskan sebagai berikut:

$$\overleftarrow{h}_t = LSTM[x_t, h_{t+1}] \quad (26)$$

Kemudian, hasil akhir dari BiLSTM merupakan kombinasi dari keluaran *forward* LSTM dan *backward* LSTM yang dituliskan dalam bentuk:

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (27)$$



del

apkan untuk memprediksi data, tahap selanjutnya adalah ukur tingkat akurasi hasil prediksi yang diperoleh. Proses ini

bertujuan untuk menguji kinerja model BiLSTM yang telah dikembangkan. Dalam penelitian ini, pengujian dilakukan menggunakan dua metrik evaluasi, yaitu Root Mean Square Error (RMSE) dan Mean Absolute Percentage Error (MAPE).

RMSE dihitung sebagai akar dari rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi. Semakin kecil nilai RMSE, semakin rendah tingkat kesalahan prediksi (Fitrinanda et al., 2022). Rumus perhitungan RMSE adalah sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (28)$$

Dengan:

n = Jumlah data

y_t = Nilai aktual pada waktu ke- t

$\hat{y}_t$ = Nilai prediksi pada waktu ke- t

Di sisi lain, MAPE dihitung sebagai rata-rata persentase kesalahan antara nilai aktual dan nilai prediksi. Semakin kecil nilai MAPE, semakin akurat model dalam melakukan prediksi (Puteri, 2023). Rumus MAPE dapat dinyatakan sebagai berikut:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{(y_t - \hat{y}_t)}{y_t} \times 100 \right| \quad (29)$$

Nilai MAPE juga dapat dikategorikan berdasarkan tingkat akurasi prediksi sebagai berikut:

Tabel 1. Kriteria nilai MAPE

Skala MAPE	Interpretasi
$x < 10\%$	Sangat Akurat
$10\% \leq x < 20\%$	Akurat
$20\% \leq x < 50\%$	Cukup
$x \geq 50\%$	Tidak Akurat



BAB II METODOLOGI PENELITIAN

2.1 Pendekatan dan Jenis Penelitian

Pendekatan yang digunakan dalam penelitian ini adalah kuantitatif deskriptif dengan metode penelitian eksperimental bersifat prediktif. Pendekatan kuantitatif deskriptif dipilih karena penelitian ini berfokus pada pengolahan data numerik berupa harga saham harian, yang dianalisis menggunakan teknik statistik dan algoritma machine learning untuk menggambarkan pola serta karakteristik data secara matematis. Metode penelitian eksperimental prediktif digunakan untuk menguji model prediksi guna memperkirakan harga saham masa depan berdasarkan data historis.

Data yang digunakan dalam penelitian ini adalah data saham sektor teknologi yang terdaftar di Bursa Efek Indonesia (BEI) dengan kode IDXTECHNO. Data diperoleh dari Investing.com, yang merupakan salah satu platform penyedia data keuangan dan pasar saham. Pemilihan saham sektor teknologi dalam penelitian ini didasarkan pada beberapa pertimbangan yaitu saham sektor teknologi tergolong baru di pasar saham, sehingga masih terdapat sedikit penelitian yang membahasnya, sektor ini menunjukkan pertumbuhan pesat serta memiliki volatilitas tinggi dibandingkan dengan sektor lainnya, dan saham sektor teknologi memiliki relevansi tinggi dengan tren global karena pergerakannya sangat dipengaruhi oleh inovasi dan regulasi industri global.

Sampel yang digunakan dalam penelitian ini terdiri dari lima saham sektor teknologi dengan jumlah peredaran terbanyak pada saat penelitian dilakukan. Kelima saham tersebut adalah GOTO (GoTo Gojek Tokopedia Tbk.), BELI (Global Digital Niaga Tbk.), BUKA (Bukalapak.com Tbk.), EMTK (Elang Mahkota Teknologi Tbk.), dan WIRG (WIR Asia Tbk.). Pemilihan lima saham ini dilakukan dengan harapan dapat mewakili kondisi saham sektor teknologi secara umum.

Periode data yang digunakan dalam penelitian ini menyesuaikan dengan ketersediaan data masing-masing saham, mengingat mayoritas saham sektor teknologi yang dianalisis baru melakukan Initial Public Offering (IPO) pada tahun 2022. Oleh karena itu, rentang waktu yang digunakan dalam penelitian ini dimulai dari tanggal IPO masing-masing saham hingga Desember 2024. Namun, khusus untuk saham EMTK, data digunakan dari Januari 2021 hingga Desember 2024 agar perbandingan antar model dan antar saham menjadi lebih adil.

Jenis data yang digunakan dalam penelitian ini adalah data kuantitatif yaitu harga terendah harian (Low) dari masing-masing saham. Pemilihan variabel ini didasarkan pada pertimbangan bahwa harga terendah dapat memberikan insight bagi investor strategi investasi dengan risiko minimal. Dengan menggunakan data ini, kerugian dapat diminimalkan karena pada titik terendah suatu saham dalam satu hari perdagangan.



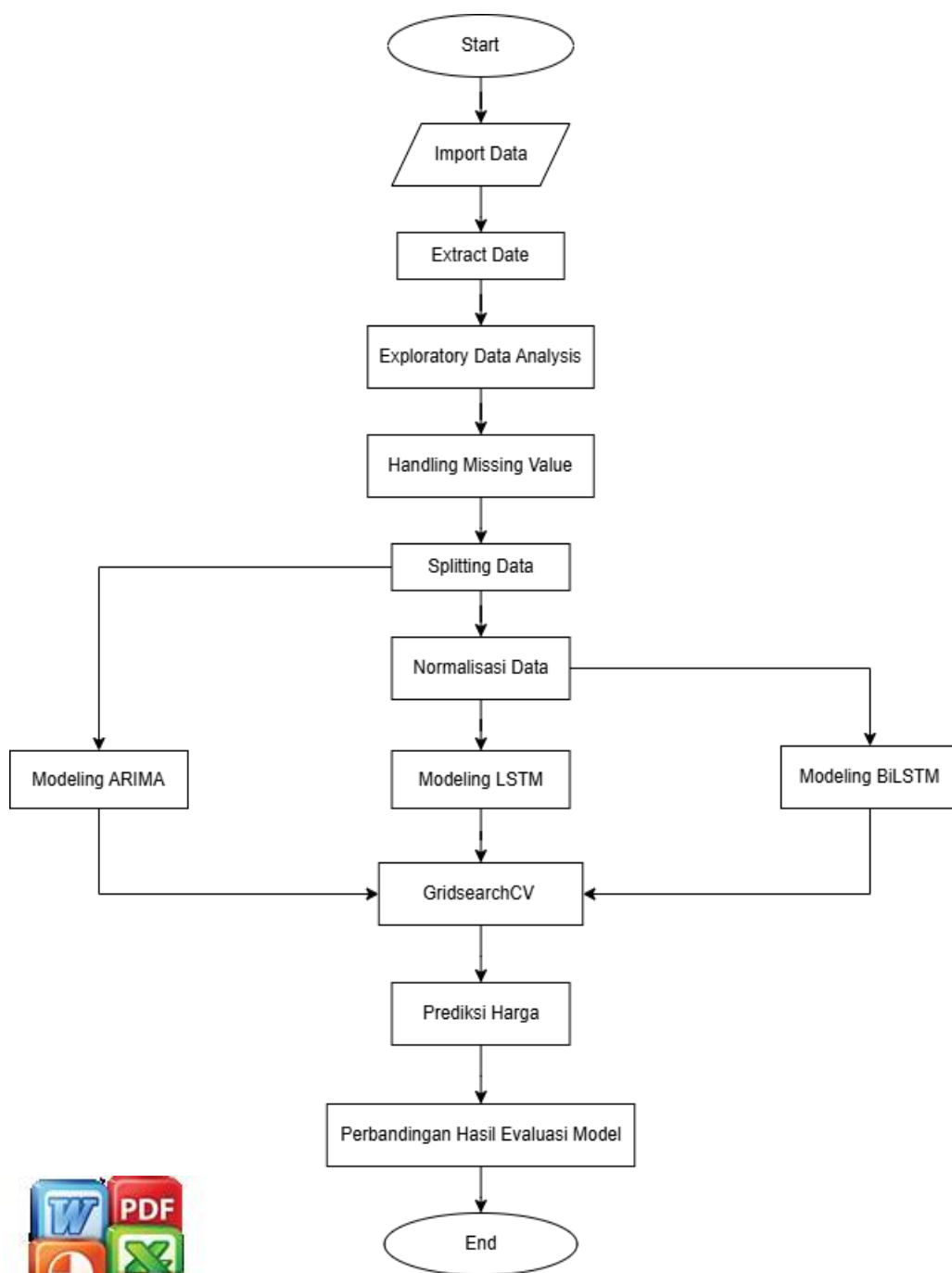
Data yang telah dikumpulkan akan dianalisis menggunakan teknik statistik dan algoritma machine learning untuk menggambarkan pola serta karakteristik data secara matematis. Model yang digunakan dalam penelitian ini meliputi:

1. Autoregressive Integrated Moving Average (ARIMA) yaitu model statistik yang digunakan untuk analisis deret waktu dalam memprediksi harga saham berdasarkan pola historis.
2. Long Short-Term Memory (LSTM) yaitu model jaringan saraf tiruan yang dirancang untuk menangani data sekuensial dan memiliki kemampuan mengingat informasi dalam jangka panjang, sehingga cocok untuk analisis data saham.
3. Bidirectional Long Short-Term Memory (BiLSTM) yaitu pengembangan dari model LSTM yang mempertimbangkan informasi dari kedua arah dalam data sekuensial, yang diharapkan dapat meningkatkan akurasi prediksi.

Model-model tersebut akan diuji dan dibandingkan untuk menentukan model terbaik dalam memprediksi harga saham sektor teknologi. Evaluasi model akan dilakukan dengan menggunakan metrik kesalahan prediksi seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE). Penelitian ini dilakukan di Laboratorium Big Data, Program Studi Ilmu Aktuaria, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin.



2.2 Alur kerja



Gambar 8. Alur Kerja

