

**Perbandingan Penerapan Metode *Long Short Term Memory* dan
Autoregressive Integrated Moving Average dalam Prediksi Harga
Saham**

TUGAS AKHIR

**Yoel Agustinus
1118042**



**PROGRAM STUDI INFORMATIKA
INSTITUT TEKNOLOGI HARAPAN BANGSA
BANDUNG
2021**

**Perbandingan Penerapan Metode *Long Short Term Memory* dan
Autoregressive Integrated Moving Average dalam Prediksi Harga
Saham**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh
gelar sarjana dalam bidang Informatika**

**Yoel Agustinus
1118042**



**PROGRAM STUDI INFORMATIKA
INSTITUT TEKNOLOGI HARAPAN BANGSA
BANDUNG
2021**

HALAMAN PERNYATAAN ORISINALITAS

**Saya menyatakan bahwa Tugas Akhir yang saya susun ini
adalah hasil karya saya sendiri.**

**Semua sumber yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.**

**Saya bersedia menerima sanksi pencabutan gelar akademik
apabila di kemudian hari Tugas Akhir ini terbukti plagiat.**

Bandung, Tanggal Bulan Tahun

A handwritten signature in black ink, appearing to be 'J. M. S.', written in a cursive style.

**Nama Pengarang
NIM**

HALAMAN PENGESAHAN TUGAS AKHIR

Tugas Akhir dengan judul:

JUDUL TUGAS AKHIR

yang disusun oleh:

Nama Pengarang

NIM

telah berhasil dipertahankan di hadapan Dewan Penguji Sidang Tugas Akhir yang dilaksanakan pada:

Hari / tanggal : Hari, Tanggal Bulan Tahun

Waktu : Jam (24-HOUR FORMAT, contoh 16.00 WIB) WIB

Menyetujui

Pembimbing Utama:

Pembimbing Pendamping:

Nama Dosen

NIK

Nama Dosen

NIK

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Institut Teknologi Harapan Bangsa, saya yang bertandatangan di bawah ini:

Nama : Nama Pengarang

NIM : NIM

Program Studi : Informatika

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Institut Teknologi Harapan Bangsa **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty Free Rights*) atas karya ilmiah saya yang berjudul:

JUDUL TUGAS AKHIR

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Institut Teknologi Harapan Bangsa berhak menyimpan, mengalihmediakan, mengelola dalam pangkalan data, dan memublikasikan karya ilmiah saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Bandung, Tanggal Bulan Tahun

Yang menyatakan



Nama Pengarang

ABSTRAK

Nama : Nama Pengarang
Program Studi : Informatika
Judul : Judul Tugas Akhir dalam Bahasa Indonesia

Lorem ipsum dolor sit amet, quidam dicunt blandit duo in. Cu sed dictas vidisse admodum, at qualisque scripserit est, est case salutandi ea. No quot ornatus probatus nec, movet quodsi forensibus pri ad. His esse wisi vocent et, ex est mazim libris quaeque. Habeo brute vel id, inani volumus adolescens et mei, solet mediocrem te sit. At sonet dolore atomorum sit, tistique sapientem contentiones no vix, dolore iriure ex vix. Vim commune appetere dissentiet ne, aperiri patrioque similique sed eu, nam facilisis neglegentur ex. Qui ut tistique voluptua. Ei utroque electram gubergren per. Laudem nonumes an vis, cum veniam eligendi liberavisse eu. Etiam graecis id mel. An quo rebum iracundia definitionem. At quo congrue graeco explicari. Cu eos wisi legimus patrioque. Cum iisque offendit ei. Ei eruditi lobortis pericula sea, te graeco salutatus sed, ne integre insolens mei. Mea tale aliquam minimum te. Eu mel putant virtute, essent inermis nominavi mea no. Laoreet indoctum sea te. Te scripta fabulas duo, pro doming recusabo voluptaria at. Cu sed numquam inciderint, ei minim altera disputando cum, te nec graeco maiorum convenire. Cu mel putent rationibus dissentiet. Per vidisse scaevola oportere ei, qui solet molestie eu. Hinc diceret nominati per at, nec dico denique laboramus et. Legere regione his at, aequae decore in mei. Lorem ipsum dolor sit amet, quidam dicunt blandit duo in. Cu sed dictas vidisse admodum, at qualisque scripserit est, est case salutandi ea. No quot ornatus probatus nec, movet quodsi forensibus pri ad. His esse wisi vocent et.

Kata kunci: Sonet, dolore, atomorum, tistique, sapientem.

ABSTRACT

Name : Nama Pengarang
Department : Informatics
Title : Judul Tugas Akhir dalam Bahasa Inggris

Lorem ipsum dolor sit amet, quidam dicunt blandit duo in. Cu sed dictas vidisse admodum, at qualisque scripserit est, est case salutandi ea. No quot ornatus probatus nec, movet quodsi forensibus pri ad. His esse wisi vocent et, ex est mazim libris quaeque. Habeo brute vel id, inani volumus adolescens et mei, solet mediocrem te sit. At sonet dolore atomorum sit, tistique sapientem contentiones no vix, dolore iriure ex vix. Vim commune appetere dissentiet ne, aperiri patrioque similique sed eu, nam facilisis neglegentur ex. Qui ut tistique voluptua. Ei utroque electram gubergren per. Laudem nonumes an vis, cum veniam eligendi liberavisse eu. Etiam graecis id mel. An quo rebum iracundia definitionem. At quo congue graeco explicari. Cu eos wisi legimus patrioque. Cum iisque offendit ei. Ei eruditi lobortis pericula sea, te graeco salutatus sed, ne integre insolens mei. Mea tale aliquam minimum te. Eu mel putant virtute, essent inermis nominavi mea no. Laoreet indoctum sea te. Te scripta fabulas duo, pro doming recusabo voluptaria at. Cu sed numquam inciderint, ei minim altera disputando cum, te nec graeco maiorum convenire. Cu mel putent rationibus dissentiet. Per vidisse scaevola oportere ei, qui solet molestie eu. Hinc diceret nominati per at, nec dico denique laboramus et. Legere regione his at, aequae decore in mei. Lorem ipsum dolor sit amet, quidam dicunt blandit duo in. Cu sed dictas vidisse admodum, at qualisque scripserit est, est case salutandi ea. No quot ornatus probatus nec, movet quodsi forensibus pri ad. His esse wisi vocent et.

Keywords: Sonet, dolore, atomorum, tistique, sapientem.

KATA PENGANTAR

Lorem ipsum dolor sit amet, quidam dicunt blandit duo in. Cu sed dictas vidisse admodum, at qualisque scripserit est, est case salutandi ea. No quot ornatus probatus nec, movet quodsi forensibus pri ad. His esse wisi vocent et, ex est mazim libris quaeque. Habeo brute vel id, inani volumus adolescens et mei, solet mediocrem te sit. At sonet dolore atomorum sit, tibiue sapientem contentiones no vix, dolore iriure ex vix. Vim commune appetere dissentiet ne, aperiri patrioque similique sed eu, nam facilisis neglegentur ex. Qui ut tibiue voluptua. Ei utroque electram gubergren per. Laudem nonumes an vis, cum veniam eligendi liberavisse eu. Etiam graecis id mel. An quo rebum iracundia definitionem. At quo congue graeco explicari. Cu eos wisi legimus patrioque. Cum iisque offendit ei. Ei eruditi lobortis pericula sea, te graeco salutatus sed, ne integre insolens mei. Mea tale aliquam minimum te. Eu mel putant virtute, essent inermis nominavi mea no. Laoreet indoctum sea te. Te scripta fabulas duo, pro doming recusabo voluptaria at. Cu sed numquam inciderint, ei minim altera disputando cum, te nec graeco maiorum convenire. Cu mel putent rationibus dissentiet. Per vidisse scaevola oportere ei, qui solet molestie eu. Hinc diceret nominati per at, nec dico denique laboramus et. Legere regione his at, aequae decore in mei.

Bandung, Tanggal Bulan Tahun

Hormat penulis,



Nama Pengarang

DAFTAR ISI

ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR ALGORITMA	x
DAFTAR LAMPIRAN	xi
BAB 1 PENDAHULUAN	1-1
1.1 Latar Belakang	1-1
1.2 Rumusan Masalah	1-3
1.3 Tujuan Penelitian	1-3
1.4 Batasan Masalah	1-3
1.5 Kontribusi Penelitian	1-4
1.6 Metodologi Penelitian	1-4
1.7 Sistematika Pembahasan	1-4
BAB 2 LANDASAN TEORI	2-1
2.1 Tinjauan Pustaka	2-1
2.1.1 Harga Saham	2-1
2.1.1.1 <i>Candlestick</i>	2-2
2.1.2 <i>Autoregressive Integrated Moving Average</i> (ARIMA)	2-2
2.1.3 <i>Artificial Neural Network</i> (ANN)	2-4
2.1.4 <i>Recurrent Neural Network</i> (RNN)	2-5
2.1.5 Fungsi Aktivasi	2-8
2.1.6 <i>Long Short Term Memory</i>	2-9
2.1.7 Evaluasi <i>Forecasting</i>	2-13
2.1.7.1 <i>Root Mean Square Error</i> (RMSE)	2-13
2.1.7.2 <i>Mean Absolute Error</i> (MAE)	2-14

2.1.7.3	<i>Mean Absolute Percentage Error (MAPE)</i>	2-14
2.1.8	Pustaka Python	2-15
2.1.8.1	Pandas	2-15
2.1.8.2	Matplotlib	2-16
2.2	Tinjauan Studi	2-17
2.3	Tinjauan Objek	2-21
2.3.1	<i>Time series Dataset</i>	2-21
2.3.1.1	<i>Trend</i>	2-21
2.3.1.2	<i>Seasonal</i>	2-22
2.3.1.3	<i>Cyclical</i>	2-22
2.3.1.4	<i>Irregular</i>	2-23
2.3.2	<i>Forecasting Horizon</i>	2-23
2.3.3	Dataset Saham	2-24
BAB 3	ANALISIS DAN PERANCANGAN SISTEM	3-1
3.1	Analisis Masalah	3-1
3.2	Kerangka Pemikiran	3-1
3.3	Urutan Proses Global	3-3
3.3.1	Proses Training	3-5
3.3.2	Proses Testing	3-6
3.4	Analisis Manual	3-6
3.4.1	Dataset	3-6
3.4.2	Normalisasi Data	3-7
3.4.3	<i>Splitting</i> Dataset untuk <i>Training</i> dan <i>Testing</i>	3-7
3.4.4	Perhitungan Manual	3-8
3.4.4.1	Perhitungan <i>Long Short Term Memory</i>	3-8
3.4.4.2	Perhitungan <i>Autoregressive Integrated Moving Average</i>	3-16
3.4.4.3	Perbandingan Evaluasi <i>Forecasting</i>	3-20

DAFTAR TABEL

2.1	Daftar metode <i>library</i> Pandas yang digunakan2-15
2.2	Daftar metode <i>library</i> Matplotlib yang digunakan2-17
2.3	Tinjauan Studi2-17

DAFTAR GAMBAR

2.1	Candlestick	2-2
2.2	<i>Artificial Neural Network Structure</i> [22]	2-5
2.3	<i>Recurrent Neural Network Loop</i>	2-6
2.4	<i>Recurrent Neural Network</i>	2-6
2.5	<i>Sigmoid Function</i> [23]	2-8
2.6	<i>tanh function</i>	2-9
2.7	<i>Long Short Term Memory unit</i>	2-10
2.8	Downtrend	2-22
2.9	Seasonal Time series [20]	2-22
2.10	Cyclical Time series [21]	2-23
2.11	Data Sampling [18]	2-24
3.1	Kerangka Pemikiran	3-1
3.2	Urutan Proses Global	3-4
3.3	<i>Flowchart proses training</i>	3-5
3.4	<i>Flowchart proses Testing</i>	3-6
3.5	Normalisasi Data <i>MinMaxScaler</i>	3-7
3.6	Splitting Dataset	3-8

DAFTAR ALGORITMA

DAFTAR LAMPIRAN

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Perekonomian Indonesia tidak lepas dari pasar modal. Banyak perusahaan yang menawarkan kepemilikan perusahaan dalam bentuk saham kepada publik [1]. Harga saham merupakan faktor yang sangat penting dan harus diperhatikan oleh investor dalam melakukan investasi karena harga saham menunjukkan prestasi emiten sebuah perusahaan. Terdapat empat komponen utama dalam Harga saham dipasar modal, yaitu: harga tertinggi (*high price*), harga terendah (*low price*), harga penutupan (*close price*), harga penutupan yang sudah disesuaikan (*adjusted close*). Harga tertinggi atau terendah merupakan harga yang paling tinggi atau paling rendah yang terjadi pada satu hari bursa, harga penutupan merupakan harga yang terjadi terakhir pada saat akhir jam bursa, harga penutupan yang disesuaikan merupakan harga penutupan yang telah disesuaikan dengan aksi korporasi seperti *right issue*, *stock split* atau *stock reverse*.

Ada penelitian yang menggunakan metode ARIMA untuk memprediksi harga saham Garuda Indonesia sejak 22 April 2019 sampai dengan 20 April 2020 dengan jumlah 255 data [2]. Peneliti melakukan banyak eksperimen dengan mengubah parameter p,d,q dan menemukan nilai parameter terbaik dengan orde 3,1,2 dengan mendapatkan nilai RMSE 38.03.

Pada penelitian [3], dijelaskan bahwa penulis memprediksi harga saham dengan menggunakan dataset *Amman Stock Exchange* dari Januari 2010 sampai Januari 2018 dengan menggunakan metode ARIMA. Lalu, peneliti mencari nilai parameter ARIMA(p,d,q) dengan nilai RMSE paling rendah yaitu 4 dengan menggunakan nilai parameter ARIMA(2,1,1).

Ada penelitian yang menggunakan berbagai macam metode *Machine Learning* dan membandingkan dengan metode *deep learning* untuk prediksi harga saham dengan menggunakan dataset dari *iShares MSCI United Kingdom* sejak Januari 2015 sampai Juni 2018 [4]. Metodenya antara lain: *Artificial Neural Network*, *Random Forest*, *Support Vector Regression*, dan *Long Short Term Memory* dengan mendapatkan nilai RMSE berturut-turut ialah: 0.454131, 0.389482, 0.340657, 0.306543. Namun, yang mempunyai RMSE terendah adalah model yang dibangun dengan metode *Long Short Term Memory*.

Ada penelitian lain yang membandingkan beberapa macam metode untuk

prediksi harga saham antara lain: *Decision Tree*, *Bagging*, *Random Forest*, *AdaBoost*, *Gradient Boosting*, *XGBoost*, *ANN*, *RNN*, dan *LSTM* [5]. Dataset yang digunakan merupakan dataset dari berbagai macam sektor saham yang diambil dalam rentang waktu 10 tahun. Model *AdaBoost* merupakan hasil metode perwakilan dari *Machine Learning* yang mendapatkan nilai RMSE terendah yaitu 0.0374 diantara metode *Machine Learning* lainnya. Lalu, model *LSTM* merupakan hasil metode perwakilan dari *deep learning* yang mendapatkan nilai RMSE terendah yaitu 0.0121 diantara metode *deep learning* lainnya.

Ada penelitian prediksi saham di masa pandemi COVID-19, dengan menggabungkan metode *Long Short Term Memory* dan *Gated Recurrent Unit* [6]. Penelitian ini menggunakan data saham *blue chip* yang ada di Indonesia antara lain: *BBCA*, *INDF*, *BBRI*, *ASII*, *TLKM*, jangka waktu pengambilan data saham mulai sejak 2 Februari 2015 sampai 19 Juni 2020 dengan jumlah 1357 data. Hasil RMSE terkecil dengan nilai 71.658 dan dengan *epoch* 35 dimiliki oleh emiten *TLKM* dengan hasil *Hidden Layer* *GRU-LSTM-GRU-LSTM*.

Ada penelitian lain yang menggunakan metode dari *Neural network based* seperti *LSTM*, *GRU*, *WLSTM* metode *LSTM* dengan menggunakan *Wavelet Transform* dan *WLSTM* dengan *Attention Model* [7]. Dataset yang digunakan merupakan harga data dari *S&P 500*, *DJIA*, dan *HSI* pada rentang waktu dari tahun 2000 sampai 2019. Hasil evaluasi dengan menggunakan model *LSTM* untuk masing-masing dataset berturut-turut antara lain: 0.2337, 0.1971, 0.3429.

Dalam penelitian lain yang membandingkan metode *Autoregressive Integrated Moving Average* dengan metode *Long Short Term Memory* dengan menggunakan emiten dari banyak perusahaan [8]. Setelah dilakukan penelitian disimpulkan bahwa rata-rata nilai RMSE dari model *LSTM* jauh lebih kecil dari rata-rata nilai RMSE dari model *ARIMA* dengan menunjukkan nilai RMSE dari model *ARIMA* adalah 511.381 dan nilai RMSE dari model *LSTM* adalah 64.445 yang bisa diartikan model *LSTM* lebih baik 87% dari model *ARIMA*.

Penelitian lain oleh Hyeong Kyuu Choi melakukan prediksi harga saham dengan menggunakan gabungan dari metode *ARIMA* dan *LSTM* [9]. Dataset yang digunakan merupakan dataset dari indeks *S&P 500* pada rentang waktu dari tahun 2008 sampai 2017. *Hybrid* model yang dibuat mendapatkan *MSE*, *RMSE*, *MAE* dengan hasil yang berturut-turut ialah 0.1786, 0.1889, 0.2154.

Dalam penelitian ini, akan membandingkan model *Long Short Term Memory* (*LSTM*) dengan *Autoregressive Integrated Moving Average* (*ARIMA*),

untuk menemukan metode terbaik bagi prediksi harga saham. Dataset yang digunakan adalah data *time-series* dari berbagai macam sektor saham di Indonesia yang diambil dari *Yahoo Finance* dalam rentang waktu 1 tahun dari 13 Feb 2021 sampai 13 Feb 2022. Sektor saham yang digunakan berasal dari beberapa sektor, antara lain: sektor perbankan, sektor infrastruktur, sektor tambang, sektor *Fast-moving Consumer Goods* (FMCG).

1.2 Rumusan Masalah

Dalam pembahasan ini peneliti mencoba untuk merumuskan masalah sebagai berikut:

1. Berapakah perbandingan nilai *MAPE*, *MAE*, *RMSE* dengan mengimplementasikan metode *Long Short Term Memory* dan *Autoregressive Integrated Moving Average* untuk memprediksi penutupan harga saham?
2. Berapa hasil prediksi harga saham dengan mengimplementasikan LSTM dan ARIMA untuk *short-term*, *mid-term*, *long-term forecasting*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, berikut adalah tujuan dari penelitian ini.

1. Mendapatkan model prediksi serta melakukan prediksi penutupan harga saham di Indonesia dengan menggunakan metode *Long Short Term Memory* dan *Autoregressive Integrated Moving Average*.
2. Mengevaluasi hasil kinerja model yang dibangun oleh metode *Long Short Term Memory* dan *Autoregressive Integrated Moving Average* untuk memprediksi harga saham.

1.4 Batasan Masalah

Agar penelitian ini terfokus dan terarah, peneliti memberikan beberapa batasan masalah sebagai berikut:

1. Dataset yang digunakan diambil dari *Yahoo Finance* dimana seluruh data saham diambil dari berbagai macam sektor saham, antara lain: BBKA, TLKM, ANTM, INDF.
2. Untuk melihat performa model yang dibuat akan dilihat dari RMSE, MAE, MAPE.

1.5 Kontribusi Penelitian

Kontribusi yang diberikan dari penelitian ini adalah membuat sebuah mesin untuk memprediksi harga saham dengan dari sektor saham yang telah ditentukan. Penelitian ini menggunakan metode LSTM dan ARIMA lalu menentukan nilai paramater yang tepat dari setiap metode, yang nantinya akan dilakukan mengevaluasi hasil prediksi dengan menggunakan RMSE, MAE, MAPE untuk menentukan hasil manakah yang terbaik.

1.6 Metodologi Penelitian

1. Studi Literatur

Penelitian ini diawali dengan melakukan studi kepustakaan yang bersumber dari jurnal penelitian terkait saham, *Long Short Term Memory*, *Autoregressive Integrated Moving Average*, atau metode deep learning yang lain.

2. Analisis Masalah

Tahap ini penulis melakukan analisis terhadap permasalahan yang ada, serta menentukan tujuan dan batasan yang ada pada penelitian ini.

3. Pencarian Data

Data yang digunakan untuk penelitian ini berupa data-data dari sektor saham yang bersumber dari *Yahoo Finance*, antara lain: BBKA yang merupakan perwakilan dari sektor perbankan, TLKM yang merupakan perwakilan dari sektor infrastruktur, ANTM yang merupakan perwakilan dari sektor tambang, INDF yang merupakan perwakilan dari sektor FMCG.

4. Training Data

Tahap ini merupakan proses pelatihan kedua model untuk memprediksi penutupan harga saham menggunakan data latih yang telah dimodelkan sebelumnya.

5. Testing Data

Tahap ini merupakan tahap pengujian kedua model yang telah dilatih sebelumnya menggunakan data latih untuk memprediksi penutupan harga saham.

6. Evaluasi

Tahap ini merupakan proses untuk mengevaluasi hasil dari prediksi harga penutupan saham dari kedua model dengan menggunakan RMSE, MAE, MAPE.

1.7 Sistematika Pembahasan

Pada penelitian ini, penulis menyusun berdasarkan sistematika pembahasan sebagai berikut

BAB I PENDAHULUAN

Bab ini berisi pendahuluan yang terdiri dari latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, kontribusi penelitian, dan metodologi penelitian.

BAB II LANDASAN TEORI

Bab ini berisi penjelasan dasar mengenai teori yang mendukung implementasi penelitian ini.

BAB III ANALISIS DAN PERANCANGAN

Bab ini berisi analisis algoritme yang digunakan dalam penelitian ini.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi implementasi dan pengujian yang dilakukan menggunakan metode terkait beserta data testing dan hasilnya.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian yang dilakukan dan saran untuk penelitian lebih lanjut di masa mendatang.

BAB 2 LANDASAN TEORI

2.1 Tinjauan Pustaka

Penelitian ini menggunakan beberapa teori terkait yang diperlukan dalam pengerjaan yang dilakukan. Penjelasan mengenai teori-teori tersebut akan dijelaskan sebagai berikut.

2.1.1 Harga Saham

Saham adalah surat yang menjadi bukti seseorang memiliki bagian modal suatu perusahaan. Saham (*stock*) merupakan salah satu instrumen pasar keuangan yang paling populer. Menerbitkan saham merupakan salah satu pilihan perusahaan ketika memutuskan untuk pendanaan perusahaan [10]. Pada sisi yang lain, saham merupakan instrument investasi yang banyak dipilih para investor karena saham mampu memberikan tingkat keuntungan yang menarik. Seorang investor umumnya dituntut untuk selalu mengikuti perkembangan informasi pasar dan harga pasar. Terdapat empat komponen utama dalam harga saham dipasar modal, yaitu: harga tertinggi (*high price*), harga terendah (*low price*), harga penutupan (*close price*), harga penutupan yang sudah disesuaikan (*adjusted close*). Harga tertinggi atau terendah merupakan harga yang paling tinggi atau paling rendah yang terjadi pada satu hari bursa, harga penutupan merupakan harga yang terjadi terakhir pada saat akhir jam bursa, harga penutupan yang disesuaikan merupakan harga penutupan yang telah disesuaikan

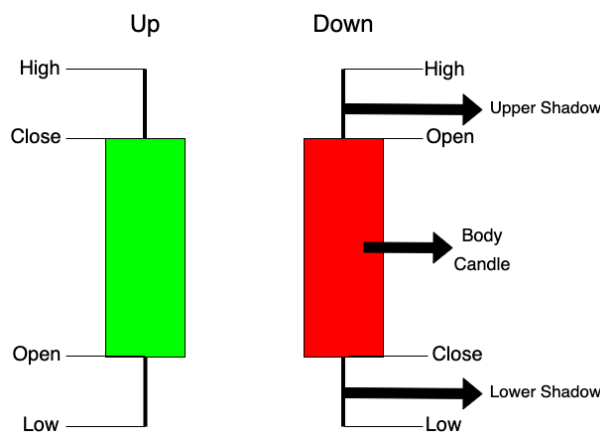
Untuk mendapatkan keuntungan tentunya diperlukan suatu analisis untuk memprediksi arah maupun harga saham. Secara garis besar analisis tersebut dapat digolongkan menjadi dua, yaitu analisis fundamental dan analisis teknikal. Analisis fundamental dalam memperkirakan harga saham di masa yang akan datang dilakukan dengan menggunakan nilai faktor-faktor fundamental, meliputi juga kinerja perusahaan, misalnya *Earning Per Share* (EPS), *Debt to Equity Ratio* (DER), *Return on Equity* (ROE), dan lainnya. Sementara, analisis teknikal adalah suatu metodologi peramalan fluktuasi harga saham yang datanya diambil dari data perdagangan saham yang terjadi di pasar saham (bursa efek) pada saat masa lampau.

Investor atau *trader* menganalisis pergerakan harga saham menggunakan perangkat statistik, seperti grafik dan rumus matematis untuk mendapatkan keuntungan baik dalam jangka pendek ataupun dalam jangka panjang. Analisis

teknikal berperang penting dalam menunjukkan *chart* harga saham, *trend* yang terjadi, *support & resistance*, serta waktu yang tepat untuk menjual maupun membeli dengan bantuan indikator.

2.1.1.1 *Candlestick*

Candlestick adalah salah satu jenis grafik harga saham yang digunakan dalam analisis teknikal yang menunjukkan harga tertinggi, terendah, pembukaan dan penutupan dari suatu saham pada periode waktu tertentu. *Candlestick* yang saat ini digunakan sebagai analisis teknikal di pasar saham memiliki dua komponen utama, yaitu: *Body Candle*, *Shadow Candle*. *Body Candle* adalah bagian dari *candlestick* yang menunjukkan harga pembukaan dan harga penutupan pada titik waktu tertentu yang ditunjukkan dari bentuk persegi empat berwarna merah atau hijau. *Shadow Candle* adalah bagian yang menunjukkan harga tertinggi dan harga terendah saham pada titik waktu tertentu yang ditunjukkan dari garis lurus yang membentang di atas dan dibawah tubuh *candlestick* dan berwarna sama seperti bentuk *candle*.



Gambar 2.1 Candlestick

2.1.2 *Autoregressive Integrated Moving Average (ARIMA)*

Autoregressive Integrated Moving Average atau yang biasa disingkat menjadi ARIMA, merupakan metode yang menghasilkan ramalan berdasarkan sintesis dari pola data secara historis [2]. Metode ARIMA, juga dikenal sebagai peramalan *Box and Jenkins*. Metode ARIMA sangat baik untuk memprediksi data *time series* karena berdasarkan asumsi bahwa data *time series* tersebut stationer yang berarti rata-rata dan varian suatu data *time series* konstan. Metode ARIMA dibagi dalam 3 unsur, yaitu: *Autoregressive*(AR), *Moving Average*(MA), dan *Integrated*(I). Ketiga unsur ini dimodifikasi secara digabungkan dan menjadi ARIMA(p,d,q), p dinyatakan sebagai ordo AR, d dinyatakan sebagai ordo

Integrated atau *difference*, dan q dinyatakan sebagai ordo MA. Apabila ada nilai parameter p, d, q berturut-turut ialah ARIMA(1,0,0) akan menjadi model yang menjalankan *Autoregressive*, dan nilai parameter p, d, q berturut-turut ialah ARIMA(0,0,1) akan menjadi model yang menjalankan *Moving Average*. Berikut merupakan persamaan model ARIMA.

$$Y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + e_t \quad (2.1)$$

Keterangan :

- Y_t : nilai yang diamati saat t
 β_p : *coefficient* untuk AR
 ϵ_t : nilai rata-rata dari periode tersebut pada t
 p : nilai *input* dari ordo p
 q : nilai *input* dari ordo q
 e_t : nilai *error white noise*

1. Autoregressive (AR)

Model *autoregressive* merupakan model dari *multiple regression* dengan nilai *lagged* dari y_t sebagai prediktor [19]. Bentuk umum persamaan dari model *Autoregressive* adalah sebagai berikut:

$$Y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t \quad (2.2)$$

Keterangan :

- Y_t : nilai yang diamati saat t
 β_p : *coefficient*
 p : nilai *input* dari ordo p
 e_t : nilai *error white noise*

2. Integrated (I)

Integrated merupakan proses pembedaan (*differencing*) agar data yang tidak stasioner menjadi stasioner [2]. Pembedaan pertama (*first difference*) dari

suatu deret waktu. Bentuk umum persamaan dari model *first difference* adalah sebagai berikut:

$$Df_t = y_t - y_{t-1} \quad (2.3)$$

Keterangan :

t : urutan nilai

y_t : nilai yang diamati saat t

Df_t : hasil nilai *difference*

3. Moving Average (MA)

Sebuah pendekatan yang memperhitungkan dengan menggabungkan ketergantungan antara nilai observasi dan nilai *residual error* dari menggunakan model rata-rata yang diterapkan pada nilai *lagged observations*. Bentuk umum persamaan dari model *Moving Average* adalah sebagai berikut:

$$Y_t = c + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \dots + \phi_q \varepsilon_{t-q} \quad (2.4)$$

Keterangan :

Y_t : Nilai yang diamati saat t

ϕ : parameter model komponen MA(q) (dimulai dengan lag terendah)

ε_t : nilai rata-rata dari periode tersebut pada t

q : Nilai input dari ordo q

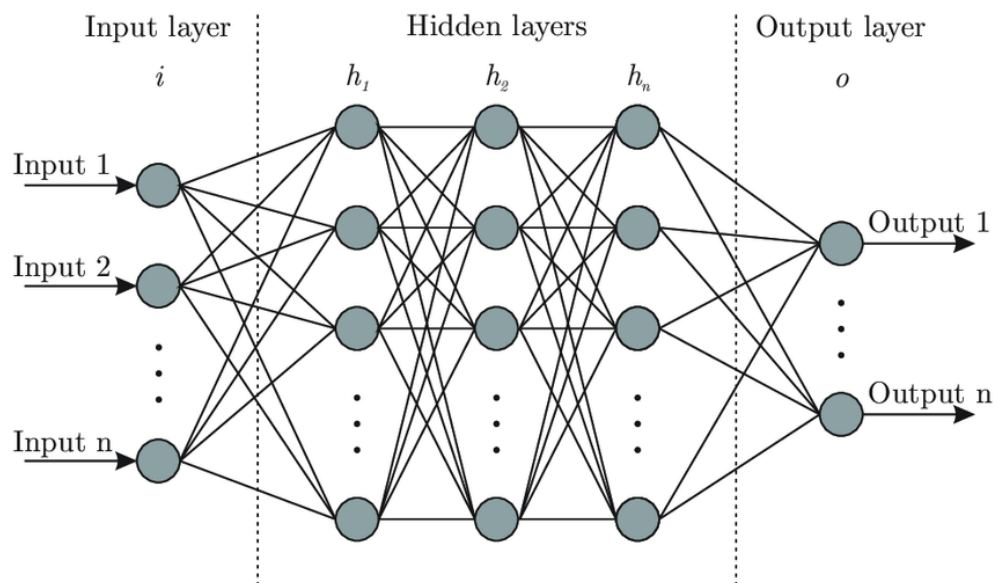
2.1.3 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) adalah kecerdasan buatan yang mengadopsi kinerja otak manusia seperti memberi stimulus atau rangsangan, melakukan proses dan memberikan *output*. *Neural Network* mempunyai tiga lapisan layer yang disebut *input layer*, *hidden layer*, dan *output layer*. Jumlah fitur dari kumpulan data menentukan dimensi atau jumlah node di *input layer* [8]. Antar *node* tersebut terhubung melalui garis yang disebut "*Synapses*" dan terhubung ke *node* yang berada di *hidden layer* yang nantinya untuk diproses.

Selanjutnya, akan diteruskan ke *output layer* atau layer paling akhir, yang berfungsi untuk mengeluarkan *output* neuron.

Neural Network terdapat 3 *layer* dan beserta elemen dasar dalam layer, yaitu [8]:

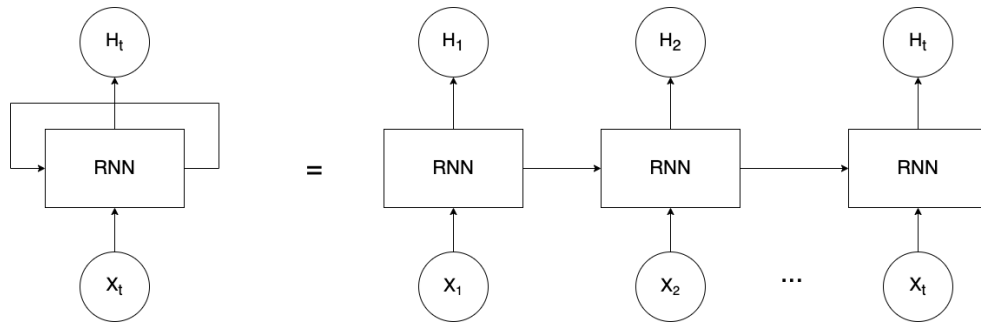
1. Kumpulan *Synapses* masing-masing mempunyai beberapa *weight*. Sementara, *weight* mempunyai peran untuk mengambil keputusan untuk memutuskan sinyal atau input mana yang dapat dilewati dan mana yang tidak
2. Di *hidden layer*, node menerapkan *activation function* (seperti, *sigmoid* dan *tanh*) pada jumlah *input* yang nanti diproses dan menjadi nilai prediksi.
3. *Output layer* menghasilkan vektor probabilitas untuk berbagai *output* dan mendapatkan nilai *error rate*



Gambar 2.2 Artificial Neural Network Structure [22]

2.1.4 Recurrent Neural Network (RNN)

RNN adalah salah satu kelas dari ANN dimana koneksi antar *node* membentuk sebuah grafik yang bersambung [14]. RNN melakukan fungsi yang sama untuk setiap *input* data. Setelah menghasilkan *output* data, data tersebut kemudian dikirim kembali pada unit berikutnya yang disebut sebagai *recurrent network*. Dalam membuat keputusan, RNN mempertimbangkan *input* titik data saat ini dan *output* dari unit sebelumnya. Hal inilah yang membedakan RNN dari ANN, *input* dan *output* pada ANN tidak bergantung sama lain sedangkan RNN memiliki sebuah memori berisikan hasil perhitungan informasi yang dihasilkan sebelumnya.



Gambar 2.3 *Recurrent Neural Network Loop*

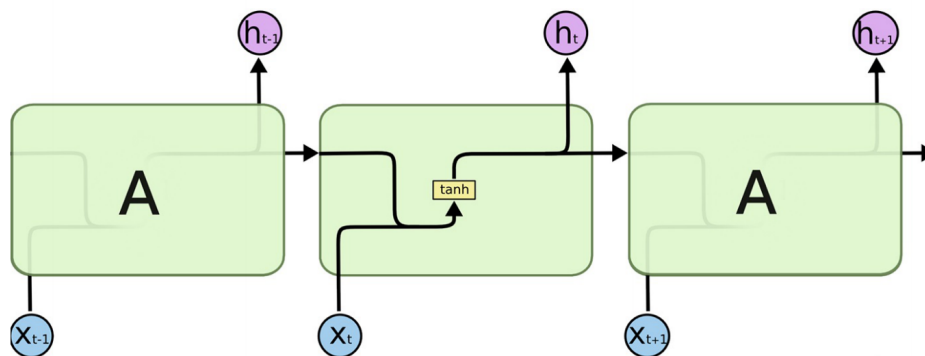
Keterangan :

X_t : nilai *input* pada *timestep* t

H_t : nilai *output* pada *timestep* t

Gambar 2.3 menunjukkan sebuah proses unit RNN menerima *input* X_t dan menghasilkan *output* H_t , dan menunjukkan RNN akan memproses *output* dari unit RNN sebelumnya sebagai *input*. Sifat dasar RNN yang sekuensial tersebut menunjukkan bahwa model RNN memiliki arsitektur yang cocok untuk data berbentuk *sequence* atau *list* [16]. Dalam gambar 2.3, dijelaskan:

1. Pertama, unit 0 akan memproses *input* X_0 dan menghasilkan *output* H_0 .
2. Setelah perhitungan unit 0 selesai, *output* 0 H_0 dan *input* unit 1 X_1 keduanya akan menjadi *input* untuk unit 1 dan demikian selanjutnya.
3. Demikian, unit RNN akan tetap mengingat konteks dari hasil pembelajaran sebelumnya.



Gambar 2.4 *Recurrent Neural Network*

Dalam gambar 2.4 ditunjukkan bahwa *loop* yang diciptakan oleh unit RNN memungkinkan informasi untuk diteruskan dari satu unit RNN ke unit RNN berikutnya.

Berikut persamaan untuk menghitung memori *layer* pada RNN.

$$H_t = f(H_{t-1}, X_t) \quad (2.5)$$

$$H_t = \tanh f(W_{hh}H_{t-1} + W_{xh}X_t) \quad (2.6)$$

Keterangan :

W : Bobot

H : Memori

W_{hh} : Bobot memori sebelumnya

W_{hx} : Bobot *input*

\tanh : Fungsi aktivasi *Hyperbolic Tangent*

Fungsi pada persamaan 2.5 dijelaskan lebih rinci pada persamaan 2.6 yang menunjukkan bahwa *input* pada titik data saat ini akan diterapkan fungsi aktivasi *Hyperbolic Tangent* bersamaan dengan menggunakan nilai memori dari unit sebelumnya.

Berikut persamaan 2.7 untuk menghitung *output* dari sistem RNN dengan mengaktifkan nilai memori terhadap bobot *output*.

$$Y_t = W_{hy}H_t \quad (2.7)$$

Keterangan :

Y_t : *output*

H : Memori

W_{hy} : Bobot *output*

2.1.5 Fungsi Aktivasi

Fungsi aktivasi dalam *neural network* digunakan untuk mengubah data *input* yang telah diberikan bobot menjadi *neuron* tersembunyi. Berikut merupakan beberapa fungsi aktivasi pada *neural network* yang akan digunakan dalam perhitungan LSTM pada penelitian ini:

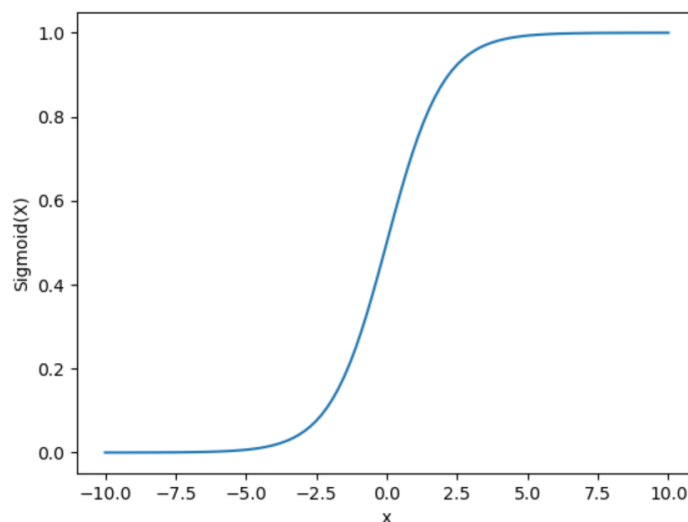
1. Sigmoid Function

Sigmoid function merupakan fungsi aktivasi yang menghasilkan nilai dengan skala 0 sampai dengan 1. Kelebihan dari fungsi aktivasi sigmoid yang menghasilkan nilai antara 0 dan 1 ini cocok digunakan untuk memrepresentasikan kemungkinan terjadinya suatu kondisi. Berikut merupakan persamaan 2.8 untuk *sigmoid function*.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.8)$$

Keterangan :

x : *input* bilangan *real*



Gambar 2.5 Sigmoid Function [23]

2. Hyperbolic Tangent Function (*tanh*)

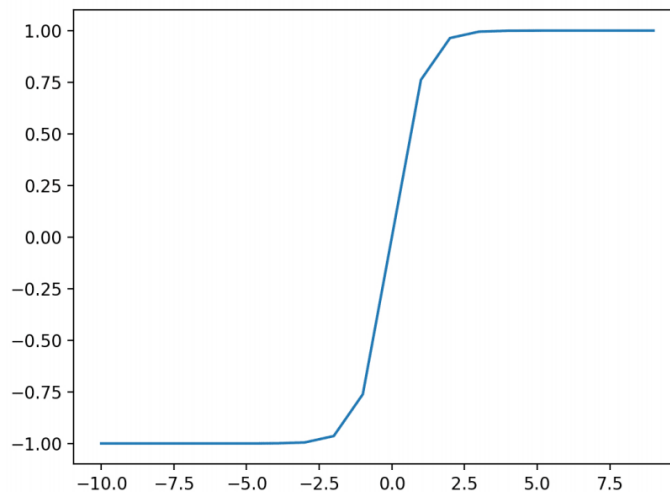
Hyperbolic Tangent Function merupakan fungsi aktivasi yang menghasilkan nilai dengan rentang -1 sampai dengan 1. Fungsi aktivasi ini bersifat non-linear dan kelebihanannya dibandingkan sigmoid function adalah kemampuannya

menghasilkan nilai negatif sampai -1. Berikut merupakan persamaan 2.9 untuk *hyperbolic tangent function*.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.9)$$

Keterangan :

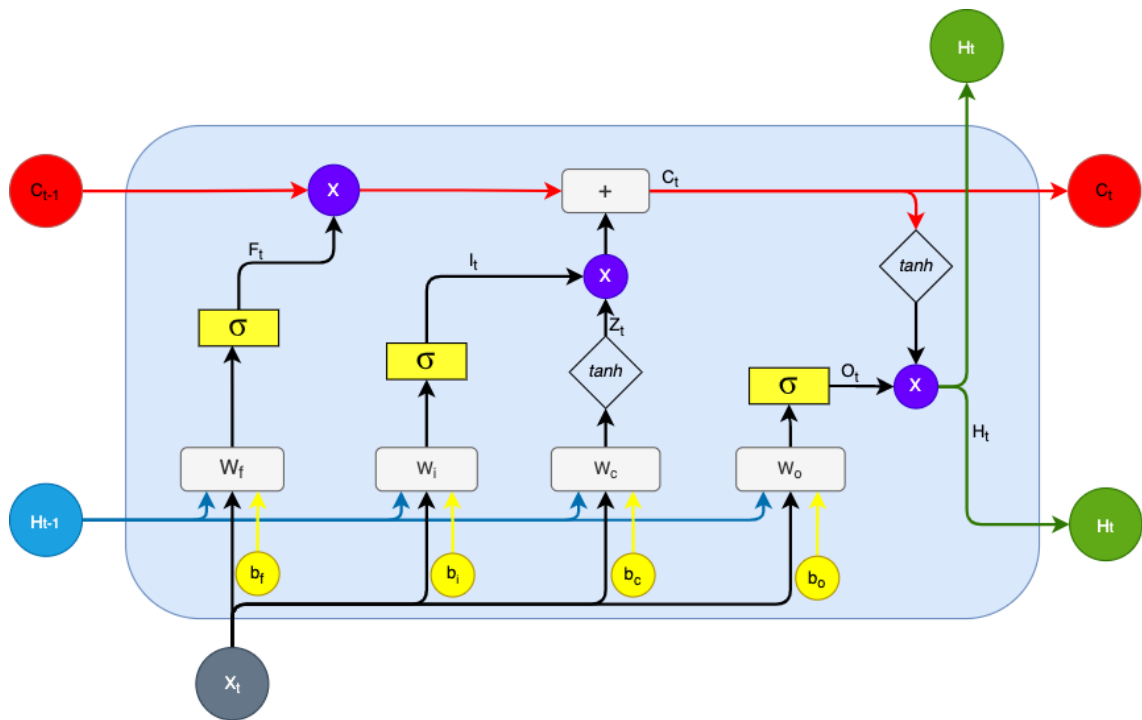
x : *input* bilangan *real*



Gambar 2.6 *tanh function*

2.1.6 Long Short Term Memory

Long Short Term Memory (LSTM) merupakan arsitektur *Recurrent Neural Network* (RNN) yang digunakan dalam bidang *deep learning* untuk melakukan pembelajaran jangka panjang dengan memanfaatkan sebuah unit memori yang disebut *cell* [6]. Arsitektur yang dimiliki oleh LSTM memungkinkan informasi dari memori sebuah unit diteruskan tidak hanya pada unit berikutnya, melainkan dapat terus memberikan kontribusi pada unit-unit selanjutnya. LSTM didesain untuk mengatasi *vanishing gradient problem* yang terdapat pada RNN dengan memiliki *Constant Error Carousel* yang memungkinkan *error* untuk melakukan *backpropagation* tanpa terjadi *vanishing gradient* atau disebut *backpropagation through time* [11]. *Constant Error Carousel* mempertahankan aktivasi internal yang disebut sebagai (*state*) dengan *recurrent connection* dengan bobot tetap yang bernilai 1.0, yang dapat diatur ulang oleh *forget gate* [24].



Gambar 2.7 Long Short Term Memory unit

Keterangan :

- X_t : nilai *input* pada *timestep* t
- C_t : memori pada *timestep* t
- C_{t-1} : memori pada *timestep* $t-1$
- H_t : *output* pada *timestep* t
- H_{t-1} : *output* pada *timestep* $t-1$
- F_t : nilai aktivasi *forget gate*
- I_t : nilai aktivasi *input gate*
- Z_t : nilai aktivasi *memory gate*
- O_t : nilai aktivasi *output gate*
- σ : fungsi aktivasi sigmoid
- \tanh : fungsi aktivasi *hyperbolic tangent*
- W_f : bobot *forget gate*
- W_i : bobot *input gate*
- W_c : bobot *memory gate*
- W_o : bobot *output gate*
- b_f : bias *forget gate*
- b_i : bias *input gate*
- b_c : bias *memory gate*
- b_o : bias *output gate*
- X : fungsi perkalian

Unit LSTM terbagi menjadi 4 *gate* yang berinteraksi untuk menambahkan atau mengeluarkan informasi ke dalam memori setiap unit. Gambar 2.7 menjelaskan setiap unit pada jaringan LSTM menerima 2 *input*, yaitu X_t unit *input* dan h_{t-1} *output* dari unit sebelumnya. Langkah pertama, LSTM akan memutuskan seberapa penting sebuah informasi untuk disimpan. Langkah ini dilakukan oleh *layer sigmoid* "forget gate" dengan persamaan 2.10 berikut:

$$F_t = \sigma(W_f \cdot [H_{t-1}, X_t] + b_f) \quad (2.10)$$

Keterangan :

F_t : nilai aktivasi *forget gate*
 H_{t-1} : *output* dari *timestep* t sebelumnya
 X_t : *input* pada *timestep* t
 σ : fungsi aktivasi *sigmoid*
 W_f : bobot *forget gate*
 b_f : bias *forget gate*

Fungsi aktivasi sigmoid akan menggunakan nilai h_{t-1} dan x_t menghasilkan nilai antara 0 dan 1, dimana hasil 1 menunjukkan bahwa nilai memori akan disimpan seluruhnya sedangkan hasil 0 dihapus selamanya [11].

$$I_t = \sigma(W_i \cdot [H_{t-1}, X_t] + b_i) \quad (2.11)$$

$$Z_t = \tanh(W_c \cdot [H_{t-1}, X_t] + b_c) \quad (2.12)$$

$$C_t = F_t \cdot C_{t-1} + I_t \cdot Z_t \quad (2.13)$$

Keterangan :

I_t	: nilai aktivasi <i>input gate</i>
W_i	: bobot <i>input gate</i>
H_{t-1}	: <i>output</i> dari <i>timestep</i> t sebelumnya
X_t	: <i>input</i> pada <i>timestep</i> t
Z_t	: nilai aktivasi <i>memory gate</i>
C_t	: <i>memory cell</i> pada <i>timestep</i> t
F_t	: nilai aktivasi <i>forget gate</i>
C_{t-1}	: <i>memory cell</i> dari <i>timestep</i> t sebelumnya
b_i	: bias <i>input gate</i>
W_c	: bobot <i>memory gate</i>
b_c	: bias <i>memory gate</i>
\tanh	: fungsi aktivasi <i>hyperbolic tangent</i>
σ	: fungsi aktivasi <i>sigmoid</i>

Langkah berikutnya adalah melakukan perubahan terhadap nilai dari memori unit yang terbagi menjadi 3 tahap. Pertama, *input gate* akan menentukan berapa besar informasi dengan melakukan fungsi aktivasi *sigmoid* terhadap *input* dari unit yang akan dijadikan memori dengan menggunakan persamaan 2.11. Tahap kedua, *memory gate* akan melakukan fungsi aktivasi *hyperbolic tangent* terhadap *input* unit untuk menghasilkan nilai memori yang akan digunakan dengan persamaan 2.12.

Tahap terakhir adalah untuk menghasilkan nilai memori baru yang diperoleh dari penjumlahan dari hasil perkalian *forget gate* (hasil dari persamaan 2.10) terhadap nilai memori unit sebelumnya (hasil dari persamaan 2.11) dengan hasil perkalian *input gate* (hasil dari persamaan 2.12) dan *memory gate* menghasilkan nilai memori unit saat ini yang akan digunakan untuk perhitungan selanjutnya pada persamaan 2.13.

$$O_t = \sigma(W_o \cdot [H_{t-1}, X_t] + b_o) \quad (2.14)$$

$$H_t = O_t \cdot \tanh(C_t) \quad (2.15)$$

Keterangan :

O_t	: nilai aktivasi <i>output gate</i>
H_{t-1}	: <i>output</i> dari <i>timestep</i> t sebelumnya
H_t	: <i>output</i> pada <i>timestep</i> t
X_t	: <i>input</i> pada <i>timestep</i> t
C_{t-1}	: <i>memory cell</i> dari <i>timestep</i> t sebelumnya
C_t	: <i>memory cell</i> pada <i>timestep</i> t
b_f	: bias <i>output gate</i>
W_f	: bobot <i>output gate</i>
\tanh	: fungsi aktivasi <i>hyperbolic tangent</i>
σ	: fungsi aktivasi <i>sigmoid</i>

Langkah terakhir adalah menghasilkan *output* H_t untuk unit saat ini (t). *Output* yang dihasilkan merupakan nilai dari *memory cell* yang telah diseleksi. Pertama, *output gate* akan menentukan jumlah informasi yang akan dijadikan *output* H_t dengan menggunakan fungsi aktivasi *sigmoid* terhadap input unit saat ini dan *output* dari unit sebelumnya (persamaan 2.14). Kemudian nilai dari *output gate* akan dikalikan dengan hasil fungsi aktivasi *hyperbolic tangent* terhadap nilai memori sesuai persamaan 2.15, menghasilkan *output* untuk unit saat ini.

Demikian diperoleh hasil *output* dan memori untuk satu unit LSTM yang akan digunakan oleh unit LSTM berikutnya. Dengan struktur arsitektur ini, maka nilai dari memori unit akan terus tersimpan dan mempengaruhi pengambilan keputusan untuk unit selanjutnya sehingga dapat mengatasi ketidakmampuan RNN dalam mengolah data dengan rentang yang besar.

2.1.7 Evaluasi *Forecasting*

Bagian ini akan menjelaskan berbagai macam pengukuran evaluasi untuk hasil prediksi harga saham.

2.1.7.1 *Root Mean Square Error (RMSE)*

RMSE adalah aturan penilaian kuadrat yang mengukur besarnya rata-rata kesalahan. Persamaan untuk RMSE diberikan di kedua referensi. Atau penjelasan lainnya yaitu, RMSE adalah perbedaan antara nilai *forecasting* dan nilai yang diamati yang dikuadratkan dan kemudian dirata-ratakan pada sampel. Lalu, akar kuadrat dari rata-rata diambil. Karena kesalahan dikuadratkan sebelum dirata-ratakan, RMSE memberikan bobot yang relatif tinggi untuk nilai kesalahan yang besar. Ini berarti RMSE paling berguna ketika nilai kesalahan yang besar sangat tidak diinginkan. Agar lebih jelas bisa dilihat pada persamaan 2.16 berikut.

$$RMSE = \sqrt{\sum_{i=1}^N (y_i - y_p)^2 / N} \quad (2.16)$$

Keterangan :

Σ : penjumlahan

y_i : nilai aktual

y_p : nilai prediksi

N : jumlah data

2.1.7.2 Mean Absolute Error (MAE)

MAE mengukur besarnya rata-rata kesalahan dalam saat melakukan *forecasting*, tanpa mempertimbangkan arahnya. MAE mengukur akurasi untuk variabel yang *continue*. Atau penjelasan lainnya yaitu, MAE adalah rata-rata atas sampel verifikasi dari nilai absolut dari perbedaan antara nilai *forecasting* dan nilai observasi. MAE adalah skor linier yang berarti bahwa semua perbedaan individu diberi bobot yang sama dalam rata-rata. Agar lebih jelas bisa dilihat pada persamaan 2.17 berikut.

$$MAE = \frac{\sum_{i=1}^N |y_i - y_p|}{N} \quad (2.17)$$

Keterangan :

Σ : penjumlahan

y_i : nilai aktual

y_p : nilai prediksi

N : jumlah data

2.1.7.3 Mean Absolute Percentage Error (MAPE)

MAPE merupakan teknik pengukuran seberapa akurat sistem *forecasting* yang sudah dibuat [15]. Pengukuran ini mengukur akurasi sebagai persentase, dan dapat dihitung sebagai *Mean Absolute Error* dalam bentuk persentase untuk setiap periode waktu dikurangi nilai aktual dibagi dengan nilai aktual. MAE adalah ukuran paling umum yang digunakan untuk memperkirakan kesalahan, dan

bekerja paling baik jika tidak ada nilai ekstrem pada data. Agar lebih jelas bisa dilihat pada persamaan 2.17 berikut.

$$MAPE = 100\% * \frac{1}{N} * \sum_{i=1}^N \frac{|y_i - y_p|}{y_i} \quad (2.18)$$

Keterangan :

Σ : penjumlahan
 y_i : nilai aktual
 y_p : nilai prediksi
 N : jumlah data

RMSE, MAE, MAPE dapat digunakan bersama-sama untuk mendiagnosis variasi kesalahan dalam serangkaian prakiraan. RMSE akan selalu lebih besar atau sama dengan MAE; semakin besar perbedaan di antara nilai RMSE dan MAE, semakin besar varians dalam kesalahan individu dalam sampel. Jika RMSE=MAE, maka semua kesalahan memiliki besar yang sama Baik MAE dan RMSE dapat berkisar dari 0 hingga tak hingga. Mereka adalah skor yang berorientasi negatif. Nilai yang lebih rendah atau yang mendekati 0 lebih baik.

2.1.8 Pustaka Python

Pada bagian ini, dijelaskan mengenai pustaka atau *library* yang digunakan dalam penelitian.

2.1.8.1 Pandas

Pandas adalah *library* python yang digunakan untuk *preprocessing* dan analisis data.

Tabel 2.1 Daftar metode *library* Pandas yang digunakan

No	Metode	Input	Output	Keterangan
----	--------	-------	--------	------------

1	<code>read_csv</code>	file path: string	DataFrame	Membaca data dalam bentuk file .csv dan mengeluarkan hasil tabel data frame.
2	<code>drop</code>	label: string, array, axis: int, inplace: boolean	DataFrame	Menghapus baris atau kolom dari sebuah DataFrame.
3	<code>sum</code>	-	Series atau DataFrame	Menjumlahkan suatu nilai dalam DataFrame.
4	<code>describe</code>	-	Series atau DataFrame	Mengeluarkan informasi statistik mengenai data.

2.1.8.2 Matplotlib

Matplotlib adalah *library* python yang digunakan untuk membuat visualisasi data.

Tabel 2.2 Daftar metode *library* Matplotlib yang digunakan

No	Metode	<i>Input</i>	<i>Output</i>	Keterangan
1	plot	X: array atau scalar, y: array atau scalar	list	Melakukan <i>plot</i> data dari sumbu x dan sumbu y
2	show	-	Figure	Menampilkan suatu visualisasi dari data yang telah didefinisikan sebelumnya.

2.2 Tinjauan Studi

Pada Tabel 2.3 diberikan penjelasan perbandingan dari berbagai penelitian terkait dengan prediksi harga saham.

Tabel 2.3 Tinjauan Studi

No	Judul	Rumusan Masalah	Metode	Hasil
----	-------	-----------------	--------	-------

1	<p>M. Nikou, G. Mansourfar, J. Bagherzadeh, "Stock Price Prediction using Deep Learning Algorithm and its Comparison with Machine Learning Algorithm," Intelligent Systems in Accounting Finance & Management, December 2019.</p> <p>[4]</p>	<p>Bagaimana hasil dari menggunakan metode LSTM, SVM, <i>Random Forest</i>, ANN?</p>	<p>Membandingkan hasil dari metode LSTM, SVM, <i>Random Forest</i>, ANN.</p>	<p>Membandingkan hasil prediksi dari model yang sudah dibangun dan hasilnya model yang dibangun dengan metode LSTM adalah yang terbaik dengan dibantu menggunakan metode <i>Dropout</i> untuk mengatasi <i>overfitting</i> dengan nilai $p = 0.8$ dan mendapatkan hasil MAE, MSE, RMSE berturut-turut ialah: 0.210350, 0.093969, 0.306543.</p>
---	--	--	--	---

2	<p>S. A. Wadi, M. Almasarweh, A. Alsaraireh, "Predicting Close Price Time Series Data Using ARIMA Model," Modern Applied Science, Vol. 12, No. 11, 2018. [3]</p>	<p>Berapakah nilai parameter ARIMA(p,d,q) untuk mencari RMSE yang terendah pada prediksi harga saham?</p>	<p><i>Autoregressive Integrated Moving Average</i></p>	<p>Mencoba nilai parameter ARIMA(p,d,q) antara 0 sampai 2 untuk mencari RMSE yang terendah. Dan mendapatkan RMSE 4.00, dengan nilai parameter ARIMA(2,1,1).</p>
---	--	---	--	---

3	<p>S. S. Namini, N. Tavakoli, A. S. Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," 2018 17th IEEE International Conference on Machine Learning and Applications. [8]</p>	<p>Apakah model yang terbaik yang dibangun dengan menggunakan metode LSTM atau ARIMA untuk prediksi harga saham?</p>	<p><i>Autoregressive Integrated Moving Average & Long Short Term Memory</i></p>	<p>Membandingkan hasil dari model yang dibangun oleh metode ARIMA atau LSTM untuk memprediksi berbagai macam harga saham. Dan hasilnya model yang dibangun menggunakan LSTM 87% lebih baik dari pada model yang dibangun dengan ARIMA, dengan nilai RMSE berturut-turut ialah: 64.213, 511.481.</p>
---	---	--	---	---

2.3 Tinjauan Objek

Bagian ini akan dipaparkan objek yang digunakan terkait dengan prediksi harga saham.

2.3.1 *Time series Dataset*

Dataset Time series adalah kumpulan pengamatan yang diperoleh melalui pengurukan berulang berdasarkan waktu. *Time series* dapat diambil pada variabel apapun yang berubah seiring waktu dengan satuan waktu yaitu jam, hari, minggu, bulan, tahun. Dalam berinvestasi, sangat umum untuk menggunakan data *time series* untuk mendapatkan harga sekuritas dari waktu ke waktu. Dapat dilacak dalam jangka pendek, seperti harga sekuritas pada jam selama hari kerja, atau jangka panjang, seperti harga sekuritas pada penutupan pada hari terakhir setiap bulan selama lima tahun. *Time series* sangat sering diplot melalui grafik *line chart*. Dengan menggunakan dataset *time series*, bisa melakukan prediksi nilai masa depan dengan memahami, menafsirkan perubahan kronologi dalam nilai suatu variabel di masa lalu. Data *time series* dapat dikategorikan menjadi data yang *stationer* atau *non stationer*. Data *time series* dikatakan stasioner jika rata-rata dan variannya konstan. Stationer bisa dikatakan tidak ada *fluktuasi* yang signifikan terhadap suatu emiten harga saham. Terdapat 4 macam komponen untuk melakukan analisis *time series*:

2.3.1.1 *Trend*

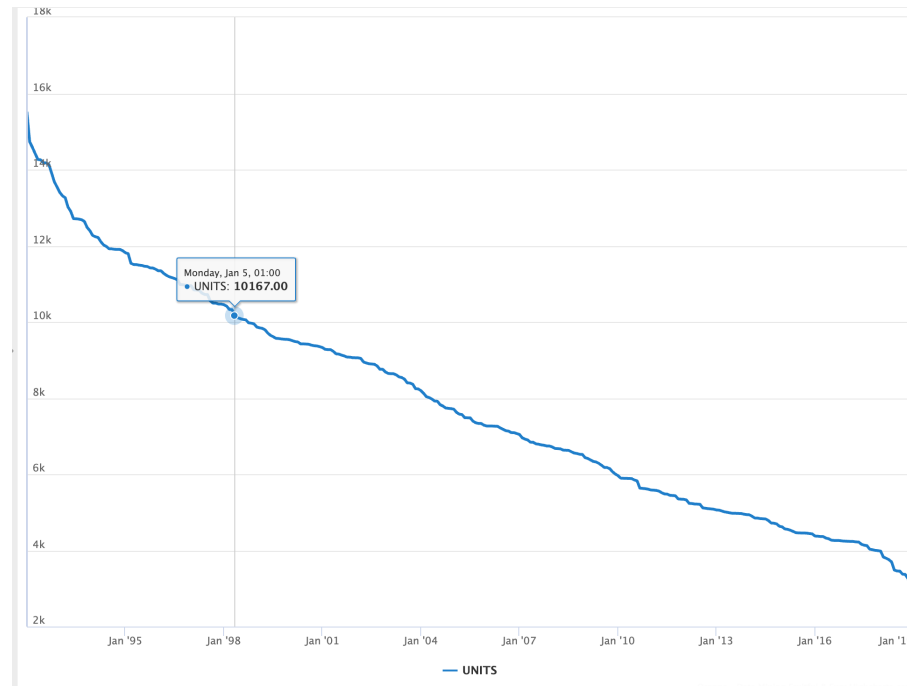
Trend adalah pola dalam data yang menunjukkan pergerakan suatu rangkaian ke nilai yang relatif lebih tinggi atau lebih rendah dalam jangka waktu yang lama. Dengan kata lain, *trend* diamati ketika ada peningkatan atau penurunan dalam *time series*. *Trend* biasanya terjadi untuk beberapa waktu dan kemudian menghilang. Misalnya, beberapa lagu baru datang, menjadi trending untuk sementara waktu, dan kemudian menghilang, dan ada kemungkinan bahwa itu menjadi *trend* lagi. Sebuah *trend* bisa berupa:

1. *Uptrend*

Analisis *time series* menunjukkan pola umum yang naik maka itu adalah *Uptrend*.

2. *Downtrend*

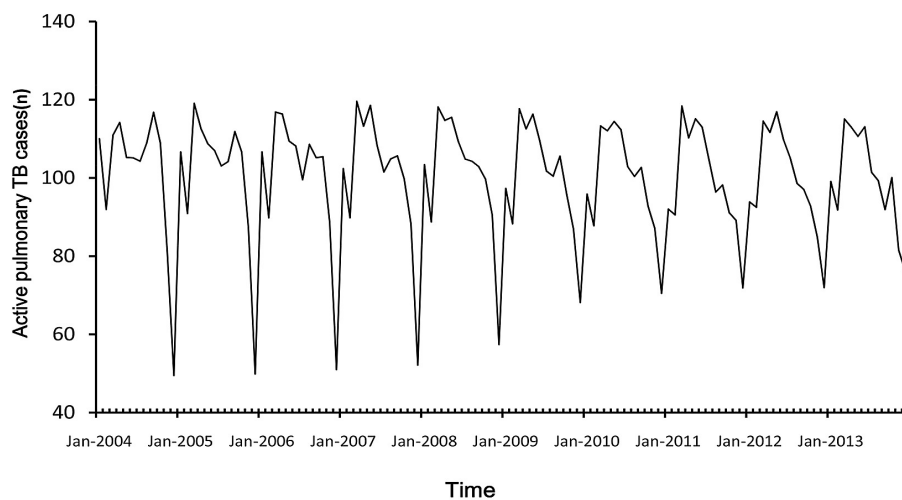
Analisis *time series* menunjukkan pola yang turun maka itu adalah *Downtrend*, seperti yang ada pada gambar 2.8.



Gambar 2.8 Downtrend

2.3.1.2 Seasonal

Seasonal adalah karakteristik pola dalam data *time series* dimana data mengalami perubahan yang teratur dan dapat diprediksi berulang setiap tahun kalender, seperti yang ada pada gambar 2.9.

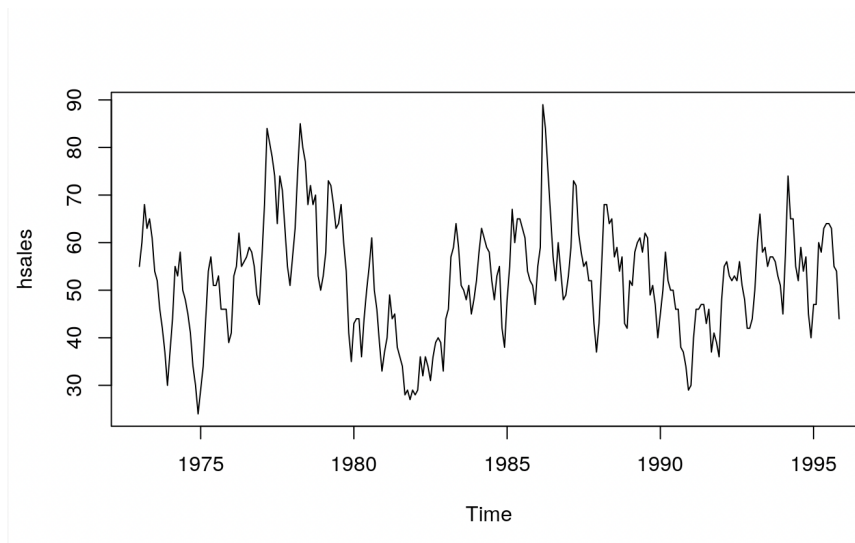


Gambar 2.9 Seasonal Time series [20]

2.3.1.3 Cyclical

Cyclical adalah karakteristik pola dalam data *time series* dimana siklus terjadi ketika data menunjukkan naik dan turun yang tidak tetap. Fluktuasi ini biasanya disebabkan oleh kondisi ekonomi. Durasi fluktuasi ini biasanya minimal 2 tahun. Seperti pada gambar 2.10 merupakan data penjualan rumah sebelum

tahun 2000. Pada tahun 1975, penjualan rumah mengalami penurunan, dan terjadi lagi saat tahun 1982, ini yang dinamakan pola data *time series cyclical*.



Gambar 2.10 Cyclical Time series [21]

2.3.1.4 Irregular

Time series biasanya diasumsikan dihasilkan pada interval waktu yang teratur dan disebut *reguler time series*. Sementara, *irregular time series* adalah kebalikan dari *regular time series* adalah urutan waktu dan nilai pengamatan yang jarak waktu pengamatannya tidak konstan.

2.3.2 Forecasting Horizon

Forecasting arti umumnya adalah untuk memprediksi sesuatu di masa yang akan datang. Banyak bidang yang memakai *forecasting* ini seperti bidang teknologi, ekonomi dan bisnis. Rencana bisnis bisa berjalan efektif di masa depan saat keadaan yang berlaku sesuai dengan perkiraan/peramalan di masa mendatang. Hal ini tidak bisa menjadi acuan pasti, tetapi *forecasting* bisa memberi gambaran agar ekonomi suatu perusahaan dapat dipersiapkan.

Forecasting Horizon secara global merupakan prediksi yang diklasifikasikan menurut periode waktu dan digolongkan menjadi *short-term*, *medium-term*, dan *long-term*. *Short-term forecasting* meramalkan kejadian hanya beberapa periode waktu ke depan (harian, mingguan). *Mid-term forecasting* meramalkan sebulan sampai 3 bulan ke depan, dan *long-term forecasting* bisa meramalkan 3 bulan sampai setahun kedepan.

2.3.3 Dataset Saham

Data yang digunakan untuk penelitian ini berupa data-data dari sektor saham yang bersumber dari *Yahoo Finance* dalam rentang waktu 1 tahun dari 13 Feb 2021 sampai 13 Feb 2022 [18], dataset harga saham dari perusahaan dan sektornya antara lain: BBKA atau Bank BCA yang merupakan perwakilan dari sektor perbankan, TLKM atau PT. Telkom Indonesia yang merupakan perwakilan dari sektor infrastruktur, ANTM atau PT. Aneka Tambang yang merupakan perwakilan dari sektor pertambangan, INDF atau PT. Indofood yang merupakan perwakilan dari sektor FMCG. Dataset harga saham memiliki 7 variabel yaitu *Date*, *Open*, *High*, *Low*, *Close*, *Adj Close*, *Volume*, contoh data dapat dilihat pada gambar 2.11.

Date	Open	High	Low	Close*	Adj Close**	Volume
Dec 03, 2021	6,588.01	6,600.17	6,536.91	6,538.51	6,538.51	-
Dec 02, 2021	6,517.34	6,586.90	6,484.58	6,583.82	6,583.82	210,837,500
Dec 01, 2021	6,544.52	6,593.07	6,494.50	6,507.68	6,507.68	195,068,500
Nov 30, 2021	6,605.80	6,647.48	6,533.93	6,533.93	6,533.93	216,798,600
Nov 29, 2021	6,552.80	6,617.29	6,487.75	6,608.29	6,608.29	-
Nov 26, 2021	6,688.57	6,691.71	6,544.90	6,561.55	6,561.55	242,523,500

Gambar 2.11 Data Sampling [18]

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini berisi analisis masalah yang diatasi beserta metode yang digunakan dan hasil yang akan ditampilkan.

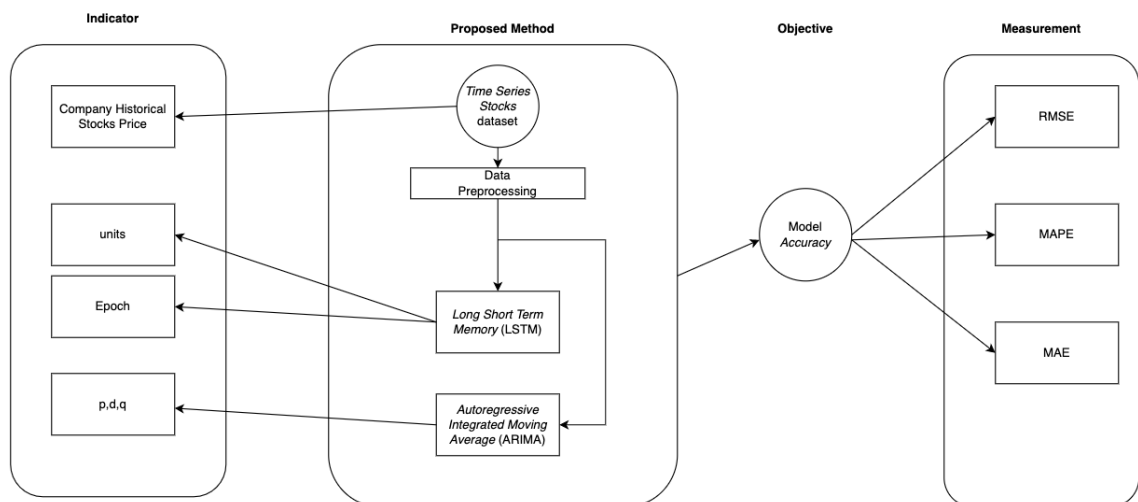
3.1 Analisis Masalah

Seperti yang dijelaskan pada bab 1, beberapa penelitian untuk memprediksi harga saham telah menggunakan beberapa metode *forecasting* oleh *machine learning*, *deep learning* ataupun *traditional forecasting*. Pada penelitian ini akan membangun sebuah sistem prediksi harga saham dengan menggunakan dataset harga saham berbagai macam sektor saham di Indonesia yang diambil dari *Yahoo Finance* dalam rentang waktu 1 tahun dari 13 Februari 2021 sampai 13 Februari 2022.

Penelitian ini akan membangun, menguji, dan membandingkan antara model *Long Short Term Memory* (LSTM) dan model *Autoregressive Integrated Moving Average* (ARIMA). *Input* pada sistem ini adalah variabel-variabel dari dataset yang sudah ditentukan, seperti *Date*, *Open*, *High*, *Low*, *Close*, *Adj Close*. *Output* pada sistem ini berupa harga dari hasil prediksi dari setiap variabel dan mendapatkan hasil pengukuran berupa RMSE, MAE, MAPE.

3.2 Kerangka Pemikiran

Berikut ini adalah kerangka pemikiran dari metode yang diusulkan untuk membangun sistem prediksi harga saham.



Gambar 3.1 Kerangka Pemikiran

Berikut akan dijelaskan setiap bagian yang ada pada gambar 3.1:

1. *Indicators* adalah variabel-variabel yang digunakan dan akan memengaruhi hasil akhir. *Indicators* yang digunakan dalam penelitian ini adalah sebagai berikut.

(a) *Company Historical Stocks Price*

Company Historical Stocks Price merupakan berbagai macam dataset historis harga saham perusahaan yang dari berbagai macam sektor, yang bersumber dari *Yahoo Finance* dalam rentang waktu 1 tahun dari 13 Feb 2021 sampai 13 Feb 2022. Dataset harga saham dari perusahaan dan sektornya antara lain: BBCA atau Bank BCA yang merupakan perwakilan dari sektor perbankan, TLKM atau PT. Telkom Indonesia yang merupakan perwakilan dari sektor infrastruktur, ANTM atau PT. Aneka Tambang yang merupakan perwakilan dari sektor pertambangan, INDF atau PT. Indofood yang merupakan perwakilan dari sektor FMCG.

(b) *Units*

Units merupakan jumlah atau ukuran dimensi dari *hidden state* atau *output*, atau yang bisa dianggap sebagai *neuron* yang berada di *hidden layer*. Pada penelitian ini akan menggunakan jumlah sebesar 1, 2, 10. Karena pada penelitian [4, 5, 6] menggunakan jumlah *unit* yang sudah ditentukan sebelumnya.

(c) *Epoch*

Epoch merupakan iterasi 1 siklus program selesai dijalankan. Semakin besar *epoch*, semakin bisa meningkatkan akurasi, namun akan semakin lama prosesnya berjalan. *Epoch* yang akan digunakan dalam penelitian ini adalah sebesar 10, 100, 1000, karena pada Penelitian [4], epoch yang digunakan sebesar 100, sedangkan pada Penelitian [6] menggunakan beragam *epoch*.

(d) p,d,q

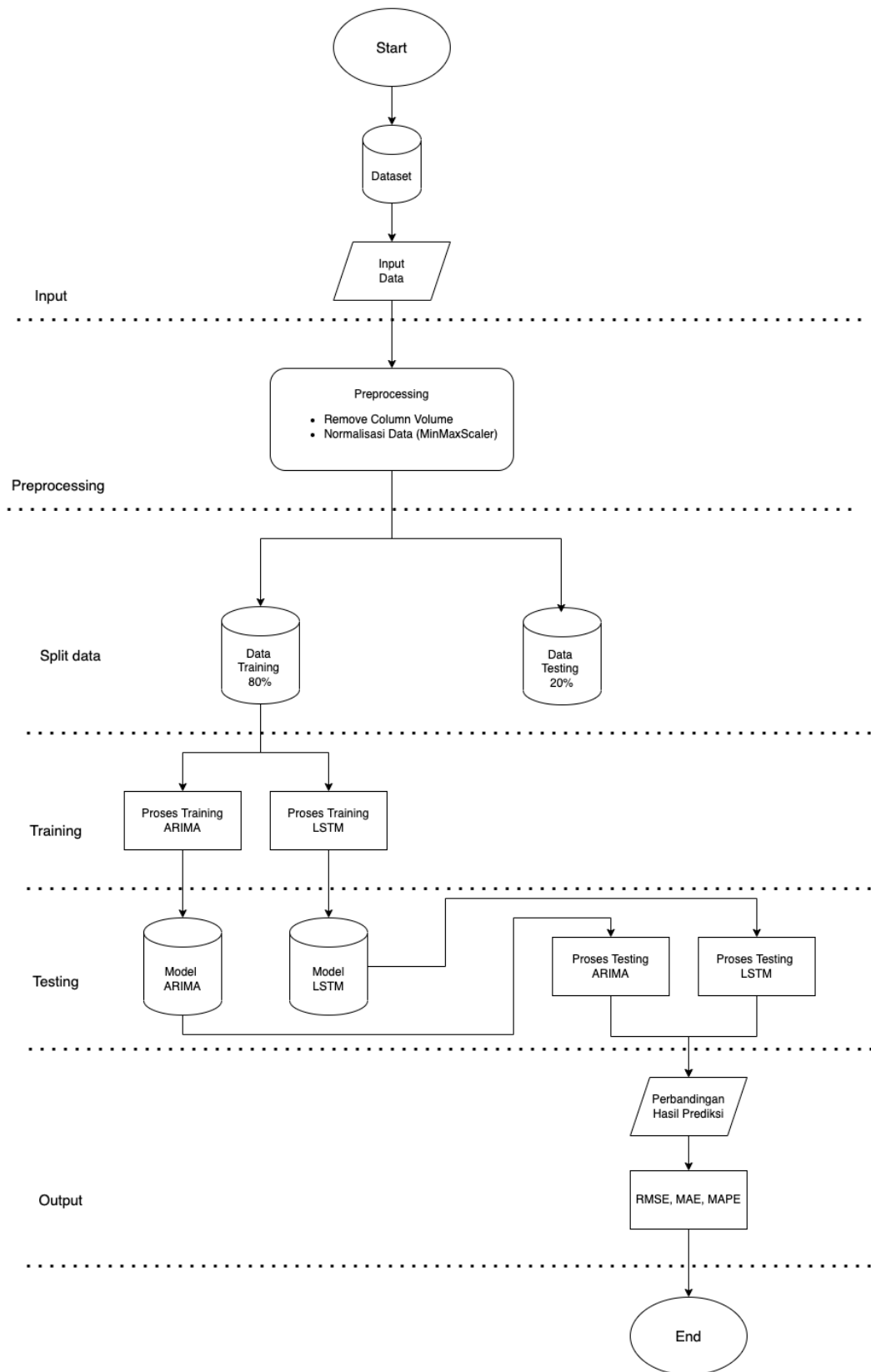
Ordo p merupakan nilai parameter untuk *Autoregressive*(AR), ordo d merupakan nilai parameter untuk *Integrated* (I), dan ordo q merupakan nilai parameter untuk *Moving Average* (MA). Pada penelitian [2], nilai ordo d yang digunakan untuk melakukan *difference* adalah 1. Ordo p

dan q dengan menggunakan nilai 1 sampai 3. Pada penelitian ini menggunakan nilai ordo d adalah 1, dan ordo p dan q menggunakan nilai dengan rentang 1 sampai 3.

2. *Proposed Method* adalah bagian yang menjelaskan proses penelitian dari awal hingga akhir. Setelah mendapatkan dataset saham maka dilakukan *preprocessing*, dengan menghapus variabel yang tidak terpakai yang terdapat dalam dataset seperti *volume*. Setelah itu akan dibuat model dengan metode *Autoregressive Integrated Moving Average* dan *Long Short Term Memory*.
3. *Objectives* adalah bagian yang menjelaskan acuan pengukuran. Penelitian ini menggunakan acuan performa dari model yang sudah dibuat.
4. *Measurement* adalah bagian yang menjelaskan ukuran yang dipakai pada bagian *objectives*, Penelitian ini menggunakan RMSE, MAPE, dan MAE.

3.3 Urutan Proses Global

Pada gambar 3.2 diberikan *flowchart* mengenai urutan proses dalam penelitian ini.



Gambar 3.2 Urutan Proses Global

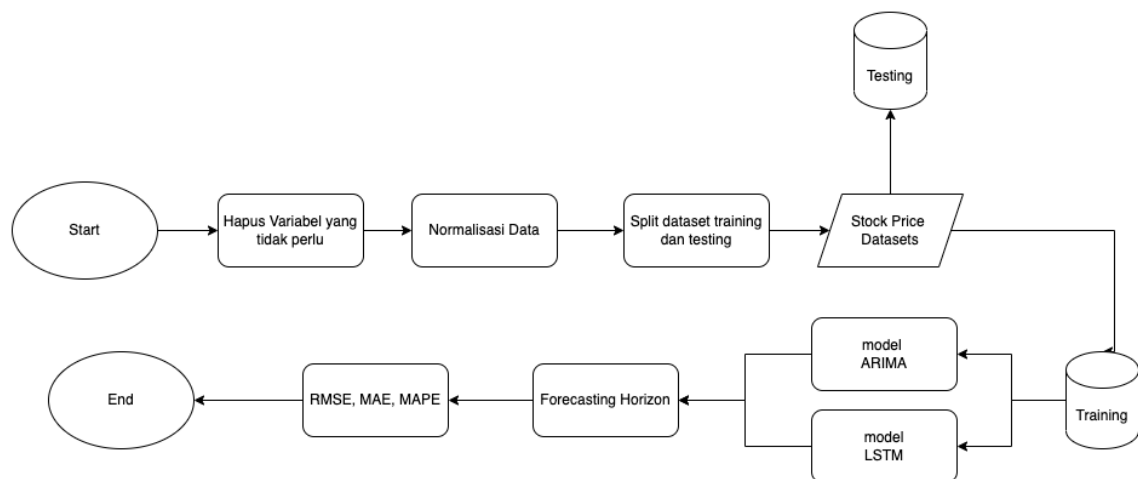
Berikut adalah uraian proses global yang dilakukan dalam penelitian ini yang

terlihat pada gambar 3.2:

Model prediksi harga saham ini dibanding menggunakan metode ARIMA dan LSTM dengan bertujuan membandingkan hasil model prediksi dari kedua metode tersebut. Setelah dilakukan *training*, model diharapkan dapat memprediksi harga saham dalam waktu jangka pendek, jangka menengah atau jangka panjang. Seperti pada gambar 3.2 dimulai dari *preprocessing* dataset, lalu melakukan *splitting* data untuk melakukan *training* dan *testing* untuk membuat model. Setelah model selesai dibuat, maka akan dilakukan proses *testing* dengan data yang bersumber dari data *testing*, maka sistem akan menghasilkan prediksi, dan performa yang akan di evaluasi dengan menggunakan RMSE, MAE, MAPE.

3.3.1 Proses Training

Pada penelitian ini, proses *training* model digambarkan pada gambar 3.3.



Gambar 3.3 Flowchart proses *training*

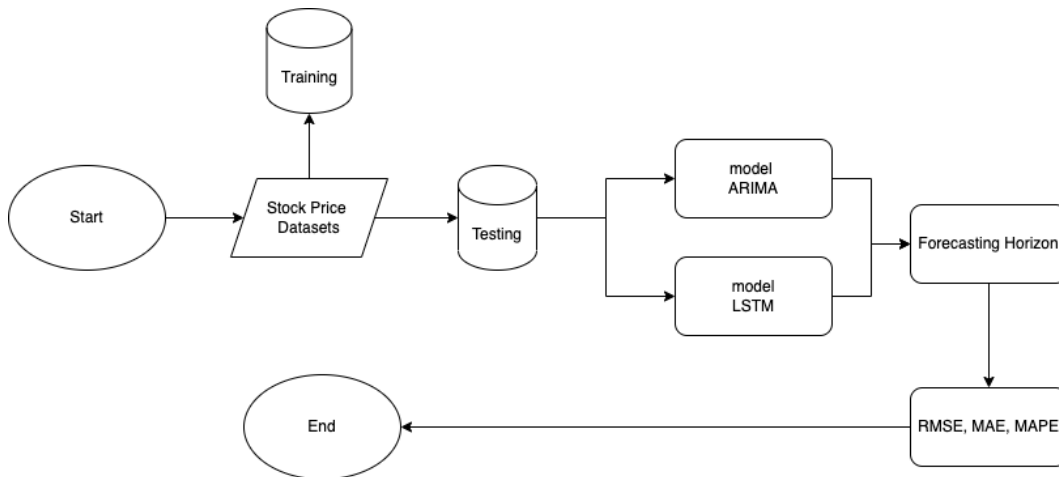
Berikut adalah uraian proses global yang dilakukan dalam penelitian ini yang terlihat pada gambar 3.3:

1. Dataset yang digunakan terdiri dari 246 data yang memiliki 7 variabel, yaitu *Date*, *Open*, *High*, *low close*, *Adj close*, *Volume*.
2. Selanjutnya, akan dilakukan penghapusan terhadap variabel *volume*.
3. Melakukan Normalisasi data dengan menggunakan *MinMaxScaler*.
4. Membagi dataset sebesar 80% untuk proses training metode ARIMA dan LSTM.
5. Dan akan menghasilkan 2 model, yaitu model ARIMA dan model LSTM.

6. Terakhir, akan diukur performa dari masing-masing model dengan menggunakan RMSE, MAE, MAPE.

3.3.2 Proses Testing

Pada penelitian ini, proses *testing* model digambarkan pada gambar 3.4.



Gambar 3.4 Flowchart proses *Testing*

Berikut adalah uraian proses global yang dilakukan dalam penelitian ini yang terlihat pada gambar 3.3:

1. Dengan menggunakan dataset yang setelah di *splitting* dengan sebesar 20% dari jumlah seluruh data.
2. Selanjutnya, akan dilakukan *testing* terhadap data yang sudah menghasilkan model ARIMA dan model LSTM.
3. Terakhir, akan melakukan prediksi *forecasting horizon* dan diukur performa dari masing-masing model dengan menggunakan RMSE, MAE, MAPE.

3.4 Analisis Manual

Pada bagian ini akan dijelaskan analisis tahapan proses yang dilakukan dalam sistem.

3.4.1 Dataset

Dalam penelitian ini, dataset yang digunakan berjumlah 4 macam dataset historis harga saham yang diambil sejak 13 Februari 2021 sampai 13 Februari 2022. Macam-macam dataset harga saham yang berasal dari: BBKA atau Bank BCA yang merupakan perwakilan dari sektor perbankan, TLKM atau PT. Telkom Indonesia yang merupakan perwakilan dari sektor infrastruktur, ANTM atau PT. Aneka Tambang yang merupakan perwakilan dari sektor pertambangan, INDF atau

PT. Indofood yang merupakan perwakilan dari sektor FMCG. Dataset yang digunakan dalam penelitian ini semua dengan format *Comma Separated Value* (CSV). Data yang akan dipakai dari dataset ini sejumlah 246 data dan memiliki 7 variabel, tetapi variabel *Volume* tidak digunakan dan akan dihapus.

3.4.2 Normalisasi Data

Pada tahap ini, data akan dilakukan proses normalisasi dengan menggunakan *MinMaxScaler*. Normalisasi data diperlukan, karena LSTM sangat sensitif terhadap data *input* yang relatif besar. *MinMaxScaler* akan mengubah nilai rentang data antara 0 sampai 1. Gambar 3.5 merupakan kode untuk melakukan normalisasi terhadap data menggunakan *MinMaxScaler* pada python.

```
from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler(feature_range=(0,1))
df1=scaler.fit_transform(np.array(df1).reshape(-1,1))
```

```
print(df1)
```

```
[[0.  ]
 [0.12]
 [0.2  ]
 [0.24]
 [0.52]
 [0.44]
 [0.6  ]
 [0.6  ]
 [0.6  ]
 [0.84]
 [0.68]
 [0.56]
 [0.56]
 [0.72]
 [1.  ]]
```

Gambar 3.5 Normalisasi Data *MinMaxScaler*

3.4.3 *Splitting* Dataset untuk *Training* dan *Testing*

Pada tahap ini dilakukan pembagian dataset untuk proses *training* dan *testing* yang akan digunakan untuk membuat model ARIMA dan LSTM. Pembagian dataset adalah 80% untuk data *training* dan 20% untuk data *testing*. Gambar 3.6 merupakan kode Python untuk melakukan proses *splitting* dataset.

```
n = int(len(df) * 0.8)
train = df.Close[:n]
test = df.Close[n:]

print(len(train))
print(len(test))

12
3

train
0      7325.0
1      7400.0
2      7362.5
3      7425.0
4      7462.5
5      7562.5
6      7625.0
7      7650.0
8      7700.0
9      7700.0
10     7775.0
11     7800.0
Name: Close, dtype: float64

test
12     7712.5
13     7675.0
14     7725.0
Name: Close, dtype: float64
```

Gambar 3.6 Splitting Dataset

3.4.4 Perhitungan Manual

Perhitungan manual akan menggunakan 12 hari data harga saham BBCA dengan memprediksi harga saham BBCA hari ke-13 sampai hari ke-15 dengan menggunakan metode LSTM dan ARIMA, yang nanti hasil prediksi hari ke-13 sampai hari ke-15 akan dilakukan evaluasi *forecasting* untuk mendapatkan nilai RMSE.

3.4.4.1 Perhitungan *Long Short Term Memory*

Pada contoh perhitungan LSTM, akan dilakukan sekali proses epoch pada 2 unit LSTM. *Input* X_t merupakan harga yang sudah dinormalisasi dan diinisialisasi pada setiap *timestep* yang dibutuhkan untuk menghasilkan *output* H_t . Nilai bobot dan bias akan diinisialisasi secara acak.

Inisialisasi dan Normalisasi Data:

Date	Close Price	Hasil Normalisasi
03/01/2022	7325	0
04/01/2022	7400	0.12
05/01/2022	7450	0.2
06/01/2022	7475	0.24
07/01/2022	7650	0.52
10/01/2022	7600	0.44
11/01/2022	7700	0.6
12/01/2022	7700	0.6
13/01/2022	7700	0.6
14/01/2022	7850	0.84
17/01/2022	7750	0.68
18/01/2022	7675	0.56
19/01/2022	P_{13}	0.56
20/01/2022	P_{14}	0.72
21/01/2022	P_{15}	1

untuk memprediksi harga P_{13} ,

maka: $X_t = [0, 0.12, 0.2, 0.24, 0.52, 0.44, 0.6, 0.6, 0.6, 0.84, 0.68, 0.56]$

untuk memprediksi harga P_{14} ,

maka: $X_t = [0.12, 0.2, 0.24, 0.52, 0.44, 0.6, 0.6, 0.6, 0.84, 0.68, 0.56, 0.56]$

untuk memprediksi harga P_{15} ,

maka: $X_t = [0.2, 0.24, 0.52, 0.44, 0.6, 0.6, 0.6, 0.84, 0.68, 0.56, 0.56, 0.72]$

Rentang *weight* = -0.4082482904638631 sampai dengan 0.4082482904638631

Maka, nilai *weight* yang didapatkan adalah:

$w_f = -0.086234$

$w_i = -0.237363$

$w_c = 0.2134627$

$w_o = 0.035626$

$b_f = 0$

$b_i = 0$

$b_c = 0$

$b_o = 0$

Berikut adalah inisialisasi data yang sudah disiapkan untuk melakukan perhitungan LSTM dan ARIMA. Terdapat 12 data penutupan harga saham yang akan di input pada perhitungan LSTM sebagai X_t yang sudah dinormalisasi dengan menggunakan *MinMaxScaler* dan ingin memprediksi 3 data harga saham dengan variabel P_{13} , P_{14} , P_{15} . Nilai bobot dan bias dengan notasi f yang merupakan nilai aktivasi yang akan digunakan oleh *forget gate*, notasi i yang merupakan nilai aktivasi dari *input gate*, notasi c yang merupakan nilai aktivasi dari *memory gate* dan notasi o yang merupakan nilai aktiasi dari *output gate*.

Perhitungan LSTM unit ke-1, prediksi harga P_1

$$F_t = \sigma(W_f * H_{t-1} + (W_f * X_t) + b_f)$$

$$F_1 = \sigma(-0.47397988232)$$

$$F_1 = 0.38367469245489$$

$$I_t = \sigma(W_i * H_{t-1} + W_i * X_t + b_i)$$

$$I_1 = \sigma(-1.2817602)$$

$$I_1 = 0.2172507475618199792226$$

$$Z_t = \tanh(W_c * H_{t-1} + W_c * X_t + b_c)$$

$$Z_1 = \tanh(1.15269858)$$

$$Z_1 = 0.8186460882475417390127$$

$$C_t = (F_t * C_{t-1}) + (I_t * Z_t)$$

$$C_1 = (0.321898269341276900773 * 0) + (0.2172507475618199792226 * 0.8186460882475417390127)$$

$$C_1 = 0.17785147466$$

$$O_t = \sigma(W_o * H_{t-1} + W_o * X_t + b_o)$$

$$O_1 = \sigma(0.1923804)$$

$$O_1 = 0.54794731276051369431$$

$$H_t = O_t * \tanh(C_t)$$

$$H_1 = O_1 * \tanh(C_1)$$

$$H_1 = 0.54794731276051369431 * \tanh(0.17785147466)$$

$$H_1 = 0.54794731276051369431 * 0.175999685440692072161$$

$$H_1 = 0.09643855468$$

Pada perhitungan di atas, merupakan perhitungan LSTM unit pertama dengan menggunakan nilai *input* X_t yang sudah ditentukan sebelumnya dengan bertujuan memprediksi harga saham P_{13} . Sesuai dengan persamaan 2.15, maka nilai H_1 yang didapatkan adalah 0.09643855468. Yang nantinya nilai H_1 akan digunakan pada perhitungan unit selanjutnya, sebagai H_{t-1} sesuai perhitungan yang berada dibawah.

Perhitungan LSTM unit ke-2, prediksi harga P_{13}

$$F_t = \sigma(W_f * H_{t-1} + (W_f * X_t) + b_f)$$

$$F_2 = \sigma((-0.086234 * 0.09643855468) + (-0.47397988232) + 0)$$

$$F_2 = \sigma(-0.48229616464)$$

$$F_2 = 0.38171006618165$$

$$I_t = \sigma(W_i * H_{t-1} + W_i * X_t + b_i)$$

$$I_2 = \sigma((-0.237363 * 0.09643855468) + (-1.2817602) + 0)$$

$$I_2 = \sigma(-1.30465114465)$$

$$I_2 = 0.213383277629580516166$$

$$Z_t = \tanh(W_c * H_{t-1} + W_c * X_t + b_c)$$

$$Z_2 = \tanh((0.2134627 * 0.09643855468) + (1.15269858) + 0)$$

$$Z_2 = \tanh(1.17328461427)$$

$$Z_2 = 0.8253222899521318863332$$

$$C_2 = (F_t * C_1) + (I_t * Z_t)$$

$$C_2 = (0.38171006618165 * 0.17785147466) + (0.213383277629580516166 * 0.8253222899521318863332)$$

$$C_2 = 0.24399767349$$

$$O_t = \sigma(W_o * H_{t-1} + W_o * X_t + b_o)$$

$$O_2 = \sigma((0.035626 * 0.09643855468) + (0.1923804) + 0)$$

$$O_2 = \sigma(0.19581611994)$$

$$O_2 = 0.5487982032078212506819$$

$$H_t = O_t * \tanh(C_t)$$

$$H_2 = O_2 * \tanh(C_2)$$

$$H_2 = 0.5525869285931001622338 * \tanh(0.24399767349)$$

$$H_2 = 0.5525869285931001622338 * 0.2392681474965215136581$$

$$H_2 = 0.13221645073$$

Pada perhitungan di atas, *output* H_2 didapatkan dengan nilai 0.13221645073. Perlu diingat, karena hasil H_2 merupakan hasil dari pemodelan LSTM dari nilai *input* yang berupa normalisasi dari *MinMaxScaler*. Maka, nilai *output* perlu di *Inverse Scaler Transform* yang berfungsi untuk mengubah nilai normalisasi *MinMaxScaler* menjadi nilai asli yang kita paham. Maka, hasil H_2 setelah dilakukan *Inverse Scaler Transform* untuk prediksi harga P_{13} adalah 7407.63.

Perhitungan LSTM unit ke-1, prediksi harga P_{14}

$$F_t = \sigma(W_f * H_{t-1} + (W_f * X_t) + b_f)$$

$$F_1 = \sigma(-0.51395464)$$

$$F_1 = 0.37426692414405$$

$$I_t = \sigma(W_i * H_{t-1} + W_i * X_t + b_i)$$

$$I_1 = \sigma(-1.41468348)$$

$$I_1 = 0.19549639942089$$

$$Z_t = \tanh(W_c * H_{t-1} + W_c * X_t + b_c)$$

$$Z_1 = \tanh(1.272237692)$$

$$Z_1 = 0.85440297634793$$

$$C_t = (F_t * C_{t-1}) + (I_t * Z_t)$$

$$C_1 = (0.37426692414405 * 0) + (0.85440297634793 * 0.19549639942089)$$

$$C_1 = 0.16703270553051$$

$$O_t = \sigma(W_o * H_{t-1} + W_o * X_t + b_o)$$

$$O_1 = \sigma((0.035626 * 0) + 0.21233096)$$

$$O_1 = 0.55288420126728$$

$$H_t = O_t * \tanh(C_t)$$

$$H_1 = O_1 * \tanh(C_1)$$

$$H_1 = 0.55288420126728 * \tanh(0.16703270553051)$$

$$H_1 = 0.55288420126728 * 0.16549644787488$$

$$H_1 = 0.091500371395874$$

Pada perhitungan di atas, merupakan perhitungan LSTM unit pertama dengan menggunakan nilai *input* X_t yang sudah ditentukan sebelumnya untuk memprediksi harga saham P_{14} . Sesuai dengan persamaan 2.15, maka nilai H_1 yang

didapatkan adalah 0.091500371395874. Yang nantinya nilai H_1 akan digunakan pada perhitungan unit selanjutnya, sebagai H_{t-1} sesuai perhitungan yang berada dibawah.

Perhitungan LSTM unit ke-2, prediksi harga P_14

$$F_t = \sigma(W_f * H_{t-1} + (W_f * X_t) + b_f)$$

$$F_2 = \sigma((-0.086234 * 0.091500371395874) + -0.51395464)$$

$$F_2 = \sigma(-0.52184508302695)$$

$$F_2 = 0.37242089287948$$

$$I_t = \sigma(W_i * H_{t-1} + W_i * X_t + b_i)$$

$$I_2 = \sigma((-0.237363 * 0.091500371395874) + -1.41468348)$$

$$I_2 = \sigma(-1.4364022826556)$$

$$I_2 = 0.19210309404676$$

$$Z_t = \tanh(W_c * H_{t-1} + W_c * X_t + b_c)$$

$$Z_2 = \tanh((0.2134627 * 0.091500371395874) + 1.272237692)$$

$$Z_2 = \tanh(1.2917696083292)$$

$$Z_2 = 0.85958929742785$$

$$C_2 = (F_t * C_1) + (I_t * Z_t)$$

$$C_2 = (0.37242089287948 * 0.16703270553051) + (0.85958929742785 * 0.19210309404676)$$

$$C_2 = 0.22733623297912$$

$$O_t = \sigma(W_o * H_{t-1} + W_o * X_t + b_o)$$

$$O_2 = \sigma((0.035626 * 0.091500371395874) + 0.21233096)$$

$$O_2 = \sigma(0.21559075223135)$$

$$O_2 = 0.55368989293003$$

$$H_t = O_t * \tanh(C_t)$$

$$H_2 = O_2 * \tanh(C_2)$$

$$H_2 = 0.55368989293003 * \tanh(0.22733623297912)$$

$$H_2 = 0.55368989293003 * 0.22349915714799$$

$$H_2 = 0.12374922439122$$

Pada perhitungan di atas, *output* H_2 didapatkan dengan nilai 0.12374922439122. Perlu diingat, karena hasil H_2 merupakan hasil dari

pemodelan LSTM dari nilai *input* yang berupa normalisasi dari *MinMaxScaler*. Maka, nilai *output* perlu di *Inverse Scaler Transform* yang berfungsi untuk mengubah nilai normalisasi *MinMaxScaler* menjadi nilai asli yang kita paham. Maka, hasil H_2 setelah dilakukan *Inverse Scaler Transform* untuk prediksi harga P_{14} adalah 7402.34.

Perhitungan LSTM unit ke-1, prediksi harga P_{15}

$$F_t = \sigma(W_f * H_{t-1} + (W_f * X_t) + b_f)$$

$$F_1 = \sigma(-0.56569504)$$

$$F_1 = 0.36223076411892$$

$$I_t = \sigma(W_i * H_{t-1} + W_i * X_t + b_i)$$

$$I_1 = \sigma(-1.55710128)$$

$$I_1 = 0.17406298803542$$

$$Z_t = \tanh(W_c * H_{t-1} + W_c * X_t + b_c)$$

$$Z_1 = \tanh(1.400315312)$$

$$Z_1 = 0.88541978464306$$

$$C_t = (F_t * C_{t-1}) + (I_t * Z_t)$$

$$C_1 = (0.36223076411892 * 0) + (0.88541978464306 * 0.17406298803542)$$

$$C_1 = 0.15411881338065$$

$$O_t = \sigma(W_o * H_{t-1} + W_o * X_t + b_o)$$

$$O_1 = \sigma(0.23370656)$$

$$O_1 = 0.55816215197308$$

$$H_t = O_t * \tanh(C_t)$$

$$H_1 = O_1 * \tanh(C_1)$$

$$H_1 = 0.55816215197308 * \tanh(0.15411881338065)$$

$$H_1 = 0.55816215197308 * 0.15291005526048$$

$$H_1 = 0.08534860550251$$

Pada perhitungan di atas, merupakan perhitungan LSTM unit pertama dengan menggunakan nilai *input* X_t yang sudah ditentukan sebelumnya untuk memprediksi harga saham P_{15} . Sesuai dengan persamaan 2.15, maka nilai H_1 yang didapatkan adalah 0.091500371395874. Yang nantinya nilai H_1 akan digunakan pada perhitungan unit selanjutnya, sebagai H_{t-1} sesuai perhitungan yang berada

dibawah.

Perhitungan LSTM unit ke-2, prediksi harga P_{15}

$$F_t = \sigma(W_f * H_{t-1} + (W_f * X_t) + b_f)$$

$$F_2 = \sigma((-0.086234 * 0.08534860550251) + -0.56569504)$$

$$F_2 = \sigma(-0.5730549916469)$$

$$F_2 = 0.36053220072448$$

$$I_t = \sigma(W_i * H_{t-1} + W_i * X_t + b_i)$$

$$I_2 = \sigma((-0.237363 * 0.08534860550251) + -1.55710128)$$

$$I_2 = \sigma(-1.5773598810479)$$

$$I_2 = 0.17116971228698$$

$$Z_t = \tanh(W_c * H_{t-1} + W_c * X_t + b_c)$$

$$Z_2 = \tanh((0.2134627 * 0.08534860550251) + 1.400315312)$$

$$Z_2 = \tanh(1.4185340557718)$$

$$Z_2 = 0.88929270918936$$

$$C_2 = (F_t * C_1) + (I_t * Z_t)$$

$$C_2 = (0.36053220072448 * 0.15411881338065) + (0.88929270918936 * 0.17116971228698)$$

$$C_2 = 0.20778477213202$$

$$O_t = \sigma(W_o * H_{t-1} + W_o * X_t + b_o)$$

$$O_2 = \sigma((0.035626 * 0.08534860550251) + 0.23370656)$$

$$O_2 = \sigma(0.023674718941963)$$

$$O_2 = 0.55891189020906$$

$$H_t = O_t * \tanh(C_t)$$

$$H_2 = O_2 * \tanh(C_2)$$

$$H_2 = 0.55891189020906 * \tanh(0.20778477213202)$$

$$H_2 = 0.55891189020906 * 0.20484519233687$$

$$H_2 = 0.11449041364923$$

Pada perhitungan di atas dengan menggunakan pemodelan dari LSTM maka mendapatkan nilai *output* H_2 0.11449041364923. Maka, hasil H_2 setelah dilakukan *Inverse Scaler Transform* untuk prediksi harga P_{15} adalah 7396.56. Maka hasil prediksi P_{13} , P_{14} , P_{15} berturut-turut adalah: 7407.63, 7402.34, 7396.56.

3.4.4.2 Perhitungan *Autoregressive Integrated Moving Average*

Pehitungan Moving Average dengan menggunakan 2 hari kebelakang:
coeffients untuk model *Autoregressive*:

<i>Coefficients</i>	
Intercept	-53.9451
ϕ_1	0.5894
ϕ_2	0.4256

Day	Close Price
<i>Day1</i>	7325
<i>Day2</i>	7400
<i>Day3</i>	7450
<i>Day4</i>	7475
<i>Day5</i>	7650
<i>Day6</i>	7600
<i>Day7</i>	7700
<i>Day8</i>	7700
<i>Day9</i>	7700
<i>Day10</i>	7850
<i>Day11</i>	7750
<i>Day12</i>	7675
<i>Day13</i>	7675
<i>Day14</i>	7775
<i>Day15</i>	7950

Date	Close Price
19/01/2022	P_{13}
20/01/2022	P_{14}
21/01/2022	P_{15}

Berikut adalah inisialisasi data untuk perhitungan ARIMA. Dengan menggunakan nilai parameter p,d,q ialah 2,1,2. Pertama-tama akan dilakukan *Moving Average* dengan bertujuan *smoothing close price*. Lalu, nilai yang setelah dilakukan smooting akan digunakan lagi untuk memprediksi harga P_{13},P_{14},P_{15} dengan menggunakan *Autoregressive* yang sudah dilakukan *difference*.

Perhitungan *Moving Average*

Ordo $q = 2$

$$Y_3 = (Day_2 + Day_1)/2$$

$$Y_3 = (7400 + 7325)/2$$

$$Y_3 = 7362.5$$

$$Y_4 = (Day_3 + Day_2)/2$$

$$Y_4 = (7450 + 7400)/2$$

$$Y_4 = 7425$$

$$Y_5 = (Day_4 + Day_3)/2$$

$$Y_5 = (7475 + 7450)/2$$

$$Y_5 = 7462.5$$

$$Y_6 = (Day_5 + Day_4)/2$$

$$Y_6 = (7650 + 7475)/2$$

$$Y_6 = 7562.5$$

$$Y_7 = (Day_6 + Day_5)/2$$

$$Y_7 = (7600 + 7650)/2$$

$$Y_7 = 7625$$

$$Y_8 = (Day_7 + Day_6)/2$$

$$Y_8 = (7700 + 7600)/2$$

$$Y_8 = 7650$$

$$Y_9 = (Day_8 + Day_7)/2$$

$$Y_9 = (7700 + 7700)/2$$

$$Y_9 = 7700$$

$$Y_{10} = (Day_9 + Day_8)/2$$

$$Y_{10} = (7700 + 7700)/2$$

$$Y_{10} = 7700$$

$$Y_{11} = (Day_{10} + Day_9)/2$$

$$Y_{11} = (7850 + 7700)/2$$

$$Y_{11} = 7775$$

$$Y_{12} = (Day_{11} + Day_{10})/2$$

$$Y_{12} = (7750 + 7850)/2$$

$$Y_{12} = 7800$$

$$Y_{13} = (Day_{12} + Day_{11})/2$$

$$Y_{13} = (7675 + 7750)/2$$

$$Y_{13} = 7712.5$$

$$Y_{14} = (Day_{13} + Day_{12})/2$$

$$Y_{14} = (7675 + 7675)/2$$

$$Y_{14} = 7675$$

$$Y_{15} = (Day_{14} + Day_{13})/2$$

$$Y_{15} = (7775 + 7675)/2$$

$$Y_{15} = 7725$$

Pada perhitungan di atas, adalah hasil dari *Moving Average* dengan nilai parameter ordo q ialah 2. Jadi, setiap perhitungan untuk mencari Y_t akan menggunakan nilai rata-rata dari ϵ_{t-q} atau $\epsilon_{t-1} + \epsilon_{t-2}$.

Perhitungan *difference*

Ordo $q = 1$

$$Df_t = y_t - y_{t-1}$$

$$Df_{13} = y_{13} - y_{12}$$

$$Df_{13} = 7712.5 - 7800$$

$$Df_{13} = -87.5$$

$$Df_{14} = y_{14} - y_{13}$$

$$Df_{14} = 7675 - 7712.5$$

$$Df_{14} = -37.5$$

$$Df_{15} = y_{15} - y_{14}$$

$$Df_{15} = 7725 - 7675$$

$$Df_{15} = 50$$

Pada perhitungan di atas, adalah hasil dari *Difference* dengan nilai ordo d ialah 1. Karena yang ingin diprediksi saat hari ke-13 sampai ke-15, maka dilakukan *difference* Df_{13} sampai Df_{15} dan mendapatkan nilai -87.5, -37.5, 50, yang nantinya nilai *difference* tersebut akan digunakan untuk memprediksi harga saham dengan menggunakan perhitungan *Autoregressive* yang berada dibawah.

Perhitungan Autoregressive

Ordo $p = 2$

Coefficients	
Intercept	-53.9451
ϕ_1	0.5894
ϕ_2	0.4256

Nilai yang sudah di *Smoothing Moving Average*

Y_t	Close
Y_{11}	7775
Y_{12}	7800
Y_{13}	7712.5
Y_{14}	7675
Y_{15}	7725

Rumus: $Y_t = c + Df_{13} + \phi_1 * y_{t-1} + \phi_2 * y_{t-2} + \dots + \phi_p y_{t-p}$

$$Y_{13} = c + Df_{13} + \phi_1 * y_{12} + \phi_2 * y_{11}$$

$$Y_{13} = -53.9451 + -87.5 + (7800 * 0.5894) + (7775 * 0.4256)$$

$$Y_{13} = 7764.92$$

$$Y_{14} = c + Df_{14} + \phi_1 * y_{13} + \phi_2 * y_{12}$$

$$Y_{14} = -53.9451 + -37.5 + (7712.5 * 0.5894) + (7800 * 0.4256)$$

$$Y_{14} = 7773.98$$

$$Y_{15} = c + Df_{15} + \phi_1 * y_{14} + \phi_2 * y_{13}$$

$$Y_{15} = -53.9451 + 50 + (7675 * 0.5894) + (7712.5 * 0.4256)$$

$$Y_{15} = 7802.14$$

Pada perhitungan di atas, adalah hasil prediksi yang menggunakan model ARIMA(2,1,2) yang didapatkan setelah melakukan perhitungan manual dengan

rumus 2.2. Maka, hasil prediksi P_{13} sampai P_{15} berturut-turut ialah 7764.92, 7773.98, 7802.14.

3.4.4.3 Perbandingan Evaluasi *Forecasting*

Pada di bagian ini akan dijelaskan perbandingan Evaluasi *Forecasting* dengan RMSE antara hasil prediksi dari LSTM dan ARIMA yang sudah hitung sebelumnya.

Evaluasi *Forecasting* dengan RMSE

<i>Close Price</i> aktual:	Day_t	Close
	Day_{13}	7675
	Day_{14}	7775
	Day_{15}	7950

<i>Close Price</i> hasil prediksi:	P_t	Close (LSTM)	Close (ARIMA)
	P_{13}	7407.63	7764.92
	P_{14}	7402.34	7773.98
	P_{15}	7396.56	7802.14

$$RMSE = \sqrt{\sum_{i=1}^N (y_i - y_p)^2 / N}$$

perhitungan RMSE untuk LSTM:

$$(y_{i13} - y_{p13})^2 = (7675 - 7407.63)^2 = 71486.7169$$

$$(y_{i14} - y_{p14})^2 = (7775 - 7402.34)^2 = 138875.4756$$

$$(y_{i15} - y_{p15})^2 = (7950 - 7396.56)^2 = 306295.8336$$

$$\sum_{i=1}^N (y_i - y_p)^2 / N$$

$$= (71486.7169 + 138875.4756 + 306295.8336) / 3$$

$$= \sqrt{172219.342033}$$

$$RMSE_{LSTM} = 414.993$$

perhitungan RMSE untuk ARIMA:

$$(y_{i13} - y_{p13})^2 = (7675 - 7764.92)^2 = 8085.6064$$

$$(y_{i14} - y_{p14})^2 = (7775 - 7773.98)^2 = 1.0404$$

$$(y_{i15} - y_{p15})^2 = (7950 - 7802.14)^2 = 21862.5796$$

$$\sum_{i=1}^N (y_i - y_p)^2 / N$$

$$= (8085.6064 + 1.0404 + 21862.5796) / 3$$

$$= \sqrt{9983.07546667}$$

$$RMSE_{ARIMA} = 99.915341498$$

Pada perhitungan Evaluasi *Forecasting* di atas dengan menggunakan rumus 2.16, maka dapat diketahui bahwa model ARIMA mempunyai hasil model prediksi lebih baik dengan mendapatkan RMSE 99.9153 dibandingkan dengan model LSTM yang mendapatkan RMSE 414.993 dalam memprediksi harga saham BBKA dengan jumlah data *training* 12 rows dan data *testing* 3 rows.

DAFTAR REFERENSI

- [1] M. Taufiq, W. R. Dharmawan, N. Jannah, "Peran dan Kontribusi Pasar Modal terhadap Pertumbuhan Ekonomi Indonesia," TRIANGLE : Journal of Management, Accounting, Economic and Business, vol 2, no. 4, 2021.
- [2] W. Y. Rusyida, V. Y. Pratama, "Prediksi Harga Saham Garuda Indonesia di Tengah Pandemi COVID-19 Menggunakan Metode ARIMA," SQUARE : Journal of Mathematics and Mathematics Education vol 2, no. 1, pp. 73-81, 2020.
- [3] S. A. Wadi, M. Almasarweh, A. A. Alsaraireh, "Predicting Close Price Time Series Data Using ARIMA Model," Modern Applied Science, Published by Canadian Center of Science and Education, vol. 12, no. 11, 2018.
- [4] M. Nikou, G. Mansourfar, J. Bagherzadeh, "Stock Price Prediction using Deep Learning Algorithm and its Comparison with Machine Learning Algorithm," Intelligent Systems in Accounting Finance & Management, December 2019.
- [5] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, Shahab S, "Deep Learning for Stock Market Prediction," Entropy, vol. 22, 2020.
- [6] W. Hastomo, A. S. B. Karno, N. Kalbuana, E. Nisfiani, E. T. P. Lussiana, "Optimasi Deep Learning untuk Prediksi Saham di Masa Pandemi COVID-19," Jurnal Edukasi dan Penelitian Informatika, vol. 7, 2021.
- [7] J. Qiu, B. Wang, C. Zhou, "Forecasting Stock Prices with Long-Short Term Memory Neural Network Based on Attention Mechanism," Journal Plos One, 3 Januari 2020.
- [8] S. S. Namini, N. Tavakoli, A. S. Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," 17th IEEE International Conference on Machine Learning and Applications, 2018.
- [9] H. K. Choi, "Stock Price Correlation Coefficient Prediction with ARIMA-LSTM Hybrid Model," 1 Oktober 2018.
- [10] A. Geron, *Hands On Machine Learning with ScikitLearn, Keras & Tensorflow*, O'Reilly Media, Inc, 2019

- [11] A. Zhang, Z. C. Lipton, M. Li, and Alexander, *Dive into Deep learning*, Release 0.17.0, 2021
- [12] Z. Hu, Y. Zhao, M. Khushi, "A Survey of Forex and Stock Price Prediction Using Deep Learning," *Applied System Innovation*, 2020
- [13] "Monthly sales of new one-family houses sold in the USA since 1973,". <https://pkg.robjhyndman.com/fma/reference/hsales.html>
- [14] Ian G., Yoshua B., Aaron C., *Deep Learning*. MIT Press, 2016
- [15] Statistics How to. "RMSE: Root Mean Square Error". [Online]. Available at: <https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/> [Accessed: Feb. 20, 2022]
- [16] A Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, Illustrated Edition. Springer, 2012
- [17] S. Maniath, A. Ashok, "Deep Learning LSTM Based Ransomware Detection", IEEE, 2017.
- [18] Yahoo Finance, 2022. "Historical Price Data Indonesia Equity". [Online]. Available: <https://finance.yahoo.com/> [Accessed: Jan. 31, 2022].
- [19] R. J. Hyndman, *Forecasting: Principles & Practice*, Illustrated Edition. Springer, 2014
- [20] X. Yang, Q. Duan, J. Wang, Z. Zhang, G. Jiang, 2014. The seasonal component of time series. [Online]. Available: https://figshare.com/articles/figure/_The_seasonal_component_of_time_series_/1200957/1 [Accessed: Feb. 18, 2022]
- [21] Robjhyndman, 2011. Cyclic and seasonal time series. [Online]. Available: <https://robjhyndman.com/hyndsight/cyclicts/> [Accessed: Feb. 18, 2022]
- [22] F Bre, J. M. Gimenez, V. D. Fachinotti. Prediction of wind pressure coefficients on building surfaces using Artificial Neural Networks. [Online]. Available: https://www.researchgate.net/figure/Artificial-neural-network-architecture-ANN-i-h-1-h-2-h-n-o_fig1_321259051 [Accessed: Feb. 22, 2022]

DAFTAR REFERENSI

- [23] Geeksforgeeks, 2019. Implement sigmoid function using Numpy. [Online]. Available: <https://www.geeksforgeeks.org/implement-sigmoid-function-using-numpy/> [Accessed: Feb. 24, 2022]
- [24] X. Wang, Y. Liu, C. Sun, B. Wang and X. Wang, "Predicting Polarities of Tweets by Composing Word Embeddings with Long Short-Term Memory," Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pages 1343–1353, Beijing, China, July 26-31, 2015.