

**ANALISIS DATA DERET WAKTU MENGGUNAKAN METODE *HYBRID*
AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) DAN
LONG SHORT TERM MEMORY (LSTM) DALAM MEMPREDIKSI
HARGA SAHAM**

Skripsi

Oleh

ZIDNY ILMA ZAIN



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

**ANALISIS DATA DERET WAKTU MENGGUNAKAN METODE *HYBRID*
AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) DAN
LONG SHORT TERM MEMORY (LSTM) DALAM MEMPREDIKSI
HARGA SAHAM**

Oleh

Zidny Ilma Zain

Skripsi

**Sebagai salah satu syarat untuk Mencapai Gelar
SARJANA MATEMATIKA**

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRACT

TIME SERIES ANALYSIS USING HYBRID AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) AND LONG SHORT TERM MEMORY (LSTM) METHOD TO PREDICTION MARKET PRICE

By

ZIDNY ILMA ZAIN

Time series analysis examines how one time series relates to another in order to determine the best model to explain the structured time series relationship between various data, which was followed by time series predictions of a single time series or many time series data. The Box-Jenkins ARIMA approach, one of the conventional techniques, was useful for modeling linear interactions but insufficient for modeling non-linear ones. The LSTM technique, another approach based on ANNs or neural networks, could simulate both linear and non-linear interactions but could not guarantee the same outcomes for every data set. The hybrid model, which combines ARIMA with LSTM, could help to identify patterns that one model alone cannot provide if the other model receives support from it. In this research, the Hybrid ARIMA-LSTM approach was used to forecast conventional media stock prices more accurately using test and training data that were respectively 80% and 20% and 70% and 30%. Based on MAPE accuracy, the ARIMA-LSTM hybrid model's best prediction for both training and test data was 99.99%.

Key words: Time Series Analysis, *Hybrid* ARIMA-LSTM

ABSTRAK

ANALISIS DATA DERET WAKTU MENGGUNAKAN METODE *HYBRID AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)* DAN *LONG SHORT TERM MEMORY (LSTM)* DALAM MEMPREDIKSI HARGA SAHAM

By

ZIDNY ILMA ZAIN

Analisis deret waktu mempelajari hubungan deret waktu yang satu dengan waktu yang lain untuk menemukan cara dan model yang tepat dalam mengekspresikan hubungan periode waktu yang terstruktur antara beberapa data yang kemudian dilakukan prediksi deret waktu dari satu atau lebih data deret waktu. Salah satu metode tradisional adalah metode ARIMA atau dikenal sebagai *BoxJenkins* baik digunakan dalam memodelkan hubungan linear, tetapi tidak cukup untuk memodelkan hubungan yang tidak linear. Metode lain berdasar pada ANN atau jaringan syaraf tiruan adalah metode LSTM, dapat memodelkan hubungan yang linear dan tidak linear tetapi tidak dapat memberikan hasil yang sama untuk setiap kumpulan data yang digunakan. Model *Hybrid* merupakan kombinasi dari ARIMA dan LSTM yang dapat membantu untuk mendapatkan pola tertentu yang tidak bisa diberikan oleh salah satu model jika tidak mendapatkan dukungan dari yang lain. Penelitian ini untuk penerapan metode *Hybrid ARIMA – LSTM* dalam meningkatkan akurasi dan memprediksi harga saham media konvensional dengan data uji dan data latih 80% dan 20% serta 70% dan 30%. Diperoleh hasil prediksi terbaik model *hybrid ARIMA-LSTM* sebesar 99,99% untuk kedua data latih dan data uji berdasarkan akurasi MAPE.

Kata Kunci: Analisis Deret Waktu, *Hybrid ARIMA-LSTM*

Judul Skripsi : **ANALISIS DATA DERET WAKTU
MENGUNAKAN METODE *HYBRID*
AUTOREGRESSIVE INTEGRATED
MOVING AVERAGE (ARIMA) DAN LONG
SHORT TERM MEMORY (LSTM) DALAM
MEMPREDIKSI HARGA SAHAM**

Nama Mahasiswa : **Zidny Irma Zain**

Nomor Pokok Mahasiswa : **1917031093**

Program Studi : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**



1. Komisi Pembimbing

Drs. Nusyirwan, M.Si.
NIP. 196610101992031028

Subian Saidi, S.Si., M.Si.
NIP. 198008212008121001

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman S.Si. M.Si.
NIP. 197403162005011001

MENGESAHKAN

1. Tim Penguji

Ketua

: Drs. Nusyirwan, M.Si.

Sekretaris

: Sublian Saidi, S.Si., M.Si.

Penguji

Bukan Pembimbing

: Dr. Khoirin Nisa, S.Si., M.Si.

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.

NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi : 17 April 2023

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama : **Zidny Ilma Zain**
Nomor Pokok Mahasiswa : **1917031093**
Jurusan : **Matematika**
Judul Skripsi : **Analisis Data Deret Waktu Menggunakan Metode *Hybrid Autoregressive Integrated Moving Average (ARIMA) Dan Long Short Term Memory (LSTM)* Dalam Memprediksi Harga Saham**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 17 April 2023

Yang menyatakan



Zidny Ilma Zain
NPM. 1917031093

RIWAYAT HIDUP

Penulis bernama lengkap Zidny Ilma Zain dilahirkan di Tangerang pada tanggal 24 Oktober 2001. Penulis terlahir dari pasangan Imam Khanafi dan Rohmawati dan terlahir sebagai anak kedua dari dua bersaudara, merupakan adik dari Almaula Faruqi Zain.

Penulis memulai pendidikan pertama di Taman Kanak-Kanak TK Putra pada tahun 2006. Kemudian, setelah menyelesaikan pendidikan TK dilanjutkan pendidikan Sekolah Dasar di SD Islamic Village pada tahun 2007. Kemudian dilanjutkan, pendidikan Sekolah Menengah Pertama di SMP Islamic Center pada tahun 2013. Lalu, melanjutkan pendidikan Sekolah Menengah Atas di SMA Negeri 5 Tangerang pada tahun 2016. Pendidikan sekolah selesai di tahun 2019 dan penulis memutuskan untuk melanjutkan Pendidikan S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung melalui jalur SBMPTN.

Selama menjadi mahasiswa penulis aktif mengikuti kegiatan kemahasiswaan dan organisasi diantaranya adalah menjadi anggota bidang Eksternal Himpunan Mahasiswa Matematika (HIMATIKA), penulis juga pernah aktif menjadi anggota divisi acara DINAMIKA XXI HIMATIKA.

Sebagai bentuk aplikasi bidang ilmu di dunia kerja, penulis telah melaksanakan Kerja Praktik (KP) di PT Global Informasi Bermutu (GTV) di divisi riset dan developmen pada bulan Januari sampai Maret 2022. Dan juga, sebagai bentuk aplikasi bidang ilmu kepada masyarakat, penulis telah melakukan Kuliah Kerja Nyata (KKN) selama 40 hari di Desa Pempen, Kecamatan Gunung Pelindung, Kabupaten Lampung Timur.

PERSEMBAHAN

Puji dan syukur kepada Tuhanku yaitu ALLAH SWT., atas rahmat, berkat dan karunianya sehingga mampu menyelesaikan skripsi ini dengan baik.

Kuepersembahkan karya tulis ini kepada mamah dan papah yang telah berkorban banyak dan mendoakan perjalanan panjang dalam penyusunan skripsi ini, juga sudah ikut dalam suka dan duka prosesnya. Serta kakaku yang telah membantu banyak baik dengan semangat dan tenaga.

Dan untuk almamaterku Universitas Lampung yang telah memberikan banyak ilmu, pengalaman, teman dan kenangan yang berharga

KATA INSPIRASI

“Allah tidak membebani seseorang melainkan sesuai kesanggupannya” (QS. Al-Baqarah: 286)

“Janganlah kamu khawatir, sesungguhnya aku bersama kamu, aku mendengar dan melihat” (QS. At-Taha: 46)

“Maka ingatlah kepadaku, aku pun akan ingat kepadamu” (QS. Al-Baqarah: 152)

“Masa depan bergantung pada apa yang kamu lakukan hari ini” (Mahatma Gandhi)

“Dunia ini gurun pasir jika setiap hari cerah. Harus turun hujan dan salju agar rumput bisa tumbuh” (Seo Dal Mi, Start Up)

“Tidak apa untuk istirahat, tidak apa jika lelah, tetapi jangan terlalu lama. Karena dunia ini terus berputar dan tidak menunggumu” (Zet)

SANWACANA

Puji syukur penulis panjatkan kepada Allah SWT, atas segala berkat, rahmat dan karunia-Nya yang melimpah sehingga penulis mampu menyelesaikan skripsi yang berjudul “**Analisis Data Deret Waktu Menggunakan Metode *Hybrid Autoregressive Integrated Moving Average (ARIMA) Dan Long Short Term Memory (LSTM) Dalam Memprediksi Harga Saham***”. Skripsi ini sebagai salah satu syarat untuk mendapatkan gelar Sarjana Matematika pada jurusan Matematika FMIPA Unila.

Masih banyak kekurangan yang dihadapi penulis dalam proses penulisan skripsi ini tetapi dengan bantuan dan semangat yang diberikan oleh orang-orang yang sangat berjasa dalam penyusunan skripsi ini penulis mampu menyelesaikan skripsi dengan baik. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada orang-orang yang sudah mau membantu, membimbing dan bekerjasama dalam penulisan skripsi ini, yaitu kepada:

1. Bapak Drs. Nusyirwan, M.Si., selalu dosen pembimbing utama yang selalu bersedia memberikan waktu untuk membimbing, memberikan saran dan arahan kepada penulis dalam menyelesaikan skripsi ini.
2. Bapak Subian Saidi, S.Si., M.Si., selaku dosen pembimbing kedua yang juga selalu bersedia memberikan waktu untuk membimbing, memberikan saran dan arahan kepada penulis dalam menyelesaikan skripsi ini.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si., selaku dosen pembahas pada ujian skripsi dan dosen pembimbing akademik. Terima kasih telah memberikan pembelajaran, ilmu yang sangat bermanfaat, saran dan evaluasi yang sangat baik kepada penulis sehingga dapat mengerjakan dan menyelesaikan dengan

baik. Terimakasih untuk semua waktu yang telah diberikan dan juga pemahaman baru yang penulis dapatkan.

4. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Bapak. Dr. Eng. Heri Satria, S.Si., M.Si., selaku dekan FMIPA Universitas Lampung.
6. Seluruh dosen, staff dan karyawan Jurusan Matematika FMIPA Universitas Lampung.
7. Kepada mamah dan papah yang sangat saya cintai atas segala dukungan, tenaga, waktu, pemikiran, dan juga doa yang tidak pernah berhenti dipanjatkan. Semoga penulis bisa mebalas jasa kalian kelak.
8. Kakakku yang paling aku sayang Almaula Faruqi Zain, atas tenaga dan motivasi yang telah diberikan dari awal perkuliahan sampai saat ini
9. Fathurrahman Saleh, yang selalu ada di masa senang dan sedih penulis dalam menulis skripsi ini. Terima kasih untuk tetap tinggal dan tidak pergi, semoga bisa bersama-sama terus sampai akhir.
10. Sahabat-sahabat ku tercinta, Bintang Gitano, Deva Randri, Sandra Puspa, Dinda Ayu, Nabilla Yolanda, Fadhilah Gustriandini, Citra Puspa, Putri Aisyah, Sund Grace, dan Adzkia A yang selalu memberikan semangat, membantu serta meyakinkan penulis untuk bisa menyelesaikan skripsi ini.
11. Teman-teman Matematika 2019 atas bantuan, arahan dan kebersamaan yang diberikan kepada penulis selama menempuh pendidikan di Universitas Lampung.

Penulis berharap dalam penyusunan skripsi yang masih belum sempurna ini bisa bermanfaat dan juga bisa menjadi pembelajaran kedepannya bagi penulis.

Bandar Lampung, 17 April 2023
Penulis,

Zidny Ilma Zain

DAFTAR ISI

DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian.....	4
1.3 Manfaat Penelitian.....	4
II. TINJAUAN PUSTAKA	6
2.1 Prediksi	6
2.2 Analisis Deret Waktu.....	6
2.3 <i>Data Mining</i>	7
2.4 Stationeritas.....	9
2.5 <i>Autocorrelation Function</i>	10
2.6 <i>Partial Autocorrelation Function</i>	10
2.7 <i>Autoregressive Integrated Moving Average (ARIMA)</i>	11
2.8 Uji <i>White Noise</i>	14
2.9 <i>Machine Learning</i>	15
2.10 <i>Deep Learning</i>	16
2.11 Artificial Neural Network	16
2.12 <i>Recurrent Neural Network</i>	18
2.13 <i>Long Short Term Memory (LSTM)</i>	19
2.14 <i>Hybrid ARIMA – LSTM</i>	25
2.15 Akurasi Model	26
III. METODOLOGI PENELITIAN	29
3.1 Waktu dan Tempat Penelitian	29

3.2	Data Penelitian.....	29
3.3	Metode Penelitian.....	29
IV. HASIL DAN PEMBAHASAN.....		32
4.1	Karakteristik Data	32
4.2	<i>Preprocessing</i> Data	33
4.3	Prediksi Data ARIMA	34
4.3.1	Uji Stationer	34
4.3.2	Pemilihan Model ARIMA.....	37
4.3.3	Estimasi Model ARIMA.....	38
4.3.4	Estimasi Parameter ARIMA	40
4.4.	Prediksi Data LSTM.....	42
4.4.1	Normalisasi Data	42
4.4.2	<i>Training</i> dan <i>Testing</i> Data.....	43
4.4.3	Segmentasi Data.....	44
4.4.4	Konstruksi Model LSTM	44
4.4.5	Pengujian Model LSTM	46
4.5	<i>Hybrid</i> ARIMA-LSTM.....	49
4.5 1.	Prediksi Residual LSTM.....	50
4.5 2.	Kombinasi Prediksi dan Residual.....	51
4.6	Perbandingan Akurasi.....	53
V. KESIMPULAN		57
DAFTAR PUSTAKA		58
LAMPIRAN		62

DAFTAR TABEL

Tabel	Halaman
1. Data Harga Saham Media Konvensional.....	32
2. Uji <i>Augmented Dickey-Fuller</i>	35
3. Uji <i>Augmented Dickey-Fuller</i> setelah <i>differencing</i>	35
4. Perbandingan model ARIMA.....	39
5. Perhitungan MAPE ARIMA(0,1,1)	39
6. Uji <i>Ljung-Box</i>	40
7. Uji <i>Kolmogorov-Sminorv</i>	41
8. Data Rata-rata dan Standar Deviasi	42
9. <i>Standar scaling</i> harga saham.....	43
10. Data <i>training</i> dan data <i>testing</i> harga saham	44
11. Prediksi harga saham LSTM.....	47
12. Dornormalisasi prediksi harga saham LSTM	48
13. Akurasi prediksi harga saham model LSTM.....	48
14. Akurasi prediksi berdasarkan pembagian training dan testing.....	49
15. Data Prediksi Residual.....	50
16. Akurasi predikis residual dengan LSTM.....	50
17. Prediksi <i>Hybrid</i> ARIMA-LSTM.....	52
18. Akurasi <i>Hybrid</i> ARIMA-LSTM	52
19. Perbandingan akurasi RMSE dan MAPE.....	54
20. Perbandingan akurasi RMSE dan MAPE.....	55

DAFTAR GAMBAR

Gambar	Halaman
1. Arsitekur dari jaringan syaraf.....	17
2. Struktur sederhana <i>Recurrent Neural Network</i>	18
3. Struktur <i>time step</i> dari LSTM.....	19
4. <i>Forget Gate</i> dari struktur LSTM.....	21
5. <i>Input gate</i> untuk struktur LSTM	22
6. <i>Cell state</i> untuk struktur LSTM	23
7. <i>Output gate</i> untuk struktur LSTM.....	24
8. Diagram alir proses <i>Hybrid</i> ARIMA-LSTM.....	31
9. Plot Kolom harga pada data harga saham media konvensional.....	33
10. Plot data harga saham yang sudah diisi dengan nilai <i>median</i>	34
11. Plot ACF dan PACF.....	36
12. Plot <i>Autocorrelation Function</i> (ACF).....	37
13. Plot <i>Partial Autocorrelation Function</i> (PACF).	38
14. Proses segmentasi data harga saham.....	44
15. Model LSTM harga saham yang terbentuk	45
16. Plot prediksi LSTM data harga saham.....	46
17. Plot ARIMA	54
18. Plot LSTM 70% data training dan 30% data testing.....	54
19. Plot <i>Hybrid</i> ARIMA-LSTM 70% data training dan 30% data testing	55

20. Plot ARIMA	55
21. Plot LSTM 80% data training dan 20% data testing.....	56
22. Plot <i>Hybrid</i> ARIMA-LSTM 80% data training dan 20% data testing	56

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Menurut Pankartz (1991), analisis deret waktu adalah sebuah analisis yang mempelajari hubungan deret waktu yaitu hubungan waktu yang satu dengan waktu yang lain. Salah satu tujuan dari analisis deret waktu adalah untuk menemukan cara dan model yang tepat untuk mengekspresikan hubungan periode waktu yang terstruktur antara beberapa data yang kemudian dilakukan evaluasi dari hubungan waktu tersebut dan dilakukan prediksi deret waktu dari satu atau lebih data deret waktu.

Salah satu metode tradisional dalam prediksi data deret waktu adalah metode ARIMA atau dikenal sebagai *BoxJenkins* pada umumnya merupakan model statistik peramalan yang termasuk *Autoregressive (AR)*, *Moving Average (MA)*, model *Autoregressive Integrated Moving Average (ARIMA)* dan angka dari variansnya. ARIMA sendiri merupakan salah satu metode paling populer dan sering digunakan untuk prediksi deret waktu yang dikembangkan oleh George Box dan Gwilym Jenkins (Zheng, dkk., 2017). Metode ini bekerja dengan sebuah asumsi bahwa deret waktu di masa sebelumnya dan deret waktu di masa depan adalah saling berhubungan satu sama lain, dimana membuat metode ini menjadi kurang efektif untuk deret waktu dengan karakteristik yang sangat tidak linear. Walaupun begitu model statistik ini hanya bisa memprediksi dengan baik apabila data yang digunakan merupakan data stasioner.

Metode lain dari metode prediksi adalah berdasar pada *Artificial Neural Networks*

(ANN) atau jaringan syaraf tiruan. Salah satunya adalah metode *Long Short Term Memory* (LSTM), yang merupakan arsitektur khusus dari *Recurrent Neural Network* (RNN) yang dapat dimanfaatkan untuk mempelajari urutan temporal dan dependensi *long term* lebih akurat dari *deep neural network* dan RNN konvensional. Menurut Qiu, dkk. (2020), *Long Short Term Memory* menggunakan bentuk yang paling umum dari RNN. Bentuk yang dimaksudkan merupakan bentuk yang digunakan untuk menghindari masalah dalam memproses dan memprediksi data deret waktu yaitu berupa masalah ketergantungan jangka panjang. Pada tahun 1997 Sepp Hochreiter dan Jurgen Schmidhuber pernah mengusulkan model LSTM yaitu sebuah model jaringan syaraf tiruan yang terdiri dari serangkaian sel memori yang unik dimana sel memori tersebut menggantikan *neuron* atau lapisan tersembunyi dari beberapa kunci sel memori dan RNN. Model LSTM sendiri dapat menyaring informasi melalui sebuah struktur *gate* dimana berfungsi untuk memperbarui dan mempertahankan keadaan sel memori. Struktur dari *gate* ini terdiri dari input, forgotten dan output gates. Dan terdapat tiga lapisan sigmoid dan satu lapisan tanh pada setiap sel memori.

Banyak terdapat model deret waktu yang menyertakan hubungan linear dan hubungan tidak linear. Model ARIMA baik digunakan dalam memodelkan hubungan linear dalam rangkaian deret waktu, tetapi tidak cukup untuk memodelkan hubungan yang tidak linear. Sedangkan untuk jaringan syaraf tiruan LSTM dapat memodelkan hubungan yang linear dan tidak linear tetapi tidak dapat memberikan hasil yang sama untuk setiap kumpulan data (Opitz dan Maclin, 1999). Maka dari itu, untuk mendapatkan hasil prediksi terbaik, digunakan model gabungan berdasarkan prinsip pemodelan terpisah dari komponen deret waktu linear dan tidak linear.

Dalam penelitian ini akan diuji coba efek dari kedua bentuk model untuk menghasilkan sebuah hasil yang dapat melebihi kinerja individu dari masing-masing model. Dibentuk model *Hybrid* untuk analisis deret waktu (Fathi, 2019). Yang merupakan kombinasi dari ARIMA dan LSTM yang dapat membantu untuk mendapatkan pola tertentu yang tidak bisa diberikan oleh salah satu model jika

tidak mendapatkan dukungan dari yang lain dan memberikan hasil yang memuaskan dalam prediksi data deret waktu.

Penelitian mengenai *Hybrid* ARIMA – LSTM pernah dilakukan untuk meramalkan kasus Covid-19 di Jawa Timur yang dilakukan oleh Rowan tahun 2022. Menunjukkan hasil bahwa *Hybrid* ARIMA-LSTM menghasilkan akurasi yang tinggi pada peramalan jangka pendek.

Selanjutnya penelitian mengenai Hybrid ARIMA – LSTM pernah dilakukan untuk meramalkan penjualan rumah di Turki yang dilakukan oleh Temur pada tahun 2019 dan hasil yang diperoleh model hybrid mencapai prediksi 34% lebih baik daripada model ARIMA dan 40% lebih baik daripada model LSTM dalam hal kriteria keberhasilan MAPE.

Televisi merupakan media konvensional yang masih diminati sampai saat ini mendapat gempuran dari media baru. Dari hasil survey litbang Kompas didapatkan informasi bahwa masih terdapat separuh responden dari 1.436 warga Jakarta yang berusia 13 tahun keatas masih menonton televisi paling lama dua jam dalam sehari (Haqqu dan Ersyad, 2020). Akan tetapi media baru semakin berkembang tiap tahunnya. Salah satunya keberadaan internet menunjukkan adanya pola perkembangan dalam media konvensional. Akses informasi yang cepat, tepat dan akurat serta murah sangat memenuhi kebutuhan masyarakat dan membuat industri pertelevisian Indonesia harus mulai berbenah dan mengikuti perkembangan zaman saat ini.

Tentu saja hal ini sangat berpengaruh kepada fluktuasi saham media konvensional. Dimana saham sendiri merupakan sebuah tanda penyertaan modal badan usaha atau dari seseorang dalam suatu perusahaan atau perseorangan terbatas. Sebuah investasi saham memungkinkan untuk memperoleh keuntungan besar bagi investor saham, tetapi juga dapat memberikan kerugian besar dalam waktu yang singkat. Sehingga prediksi pada fluktuasi harga saham sangat penting dan

dibutuhkan oleh para investor untuk melakukan investasi sebelum memutuskan (Herawati, 2013).

Berdasarkan hasil pemaparan di atas penulis tertarik untuk melakukan penelitian tentang penerapan metode *Hybrid* ARIMA – LSTM untuk meningkatkan akurasi dan meramalkan harga saham media konvensional agar dapat mengetahui kapan terjadi kenaikan dan juga penurunan yang sangat signifikan terhadap perkembangan teknologi dan bagaimana mempengaruhi pergerakan nilai saham media konvensional di era saat ini.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Untuk mengetahui prediksi dari model *Hybrid* ARIMA-LSTM terhadap harga saham media konvensional.
2. Mengetahui performa metode *Hybrid* ARIMA-LSTM dalam memprediksi harga saham media konvensional.
3. Untuk mendapatkan jaringan arsitektur terbaik yang diperoleh dari metode *Hybrid* ARIMA-LSTM.

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Dapat memberikan pengetahuan terkait metode *Hybrid* ARIMA – LSTM.
2. Dapat menjadi sarana pengembangan minat kilmuan, terkhusus di bidang prediksi nilai saham dengan menggunakan metode *Hybrid* ARIMA – LSTM.

3. Sebagai bahan referensi bagi peneliti selanjutnya tentang metode *Hybrid* ARIMA-LSTM dan juga dalam memprediksi harga saham media konvensional.

II. TINJAUAN PUSTAKA

2.1 Prediksi

Menurut Brilliant, dkk (2021), terdapat dua dalam peramalan yaitu prediksi dan prakiraan. Prakiraan didefinisikan sebagai proses memprediksi peristiwa masa depan berdasarkan masa lalu. Data masa sebelumnya tersebut secara sistematis diproses bersama dengan metode tertentu untuk mendapatkan prediksi peristiwa masa depan. Sementara prediksi adalah proses memprediksi variabel masa depan berdasarkan pertimbangan intuitif dari data sebelumnya, meskipun lebih didasarkan pada perkiraan intuitif, prediksi juga sering digunakan sebagai tambahan informasi dan digunakan untuk membuat prediksi data kuantitatif.

2.2 Analisis Deret Waktu

Menurut Pankartz (1991), analisis deret waktu adalah sebuah analisis yang mempelajari hubungan deret waktu yaitu waktu yang satu dengan waktu yang lain. Salah satu tujuan dari analisis seri waktu adalah untuk menemukan cara dan model yang tepat untuk mengekspresikan hubungan periode terstruktur antara beberapa data, yang kemudian dinilai dari hubungan deret waktu dan prediksi deret waktu yang dilakukan dari data satu atau lebih deret waktu.

Menurut Durbin dan Koopman (2012), dari suatu deret waktu dapat diketahui pola perkembangan dari suatu peristiwa, variabel atau kejadian. Jika perkembangan suatu peristiwa dapat mengikuti suatu pola yang teratur, maka dari pola tersebut

akan didapat peramalan suatu peristiwa yang akan terjadi di masa depan. Jika suatu observasi dilambangkan dengan y_1, y_2, \dots, y_n , maka deret waktu memiliki model dasar sebagai berikut:

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad t = 1, \dots, n \quad (2.1)$$

Dimana:

μ_t = komponen yang bervariasi secara perlahan (tren)

γ_t = periodik komponen periode tetap (musiman)

ε_t = komponen tidak beraturan yang disebut kesalahan (error)

Terdapat 4 pola data yang sering dijumpai dalam analisis deret waktu (Makridakis, dkk., 1999):

1. Pola Horizontal

Pola data horizontal dapat terjadi jika fluktuasi data deret waktu berada di sekitar nilai konstan atau nilai rata-ratanya.

2. Pola Musiman

Pola data musiman dapat terjadi jika nilai dari data deret waktu dipengaruhi oleh suatu faktor musiman

3. Pola Siklis

Pola data siklis dapat terjadi jika data deret waktu dipengaruhi oleh fluktuasi ekonomi jangka waktu yang panjang.

4. Pola *Trend*

Pola data *trend* dapat terjadi jika terdapat kenaikan atau penurunan sekuler jangka panjang pada data deret waktu.

2.3 *Data Mining*

Data Mining disebut juga dengan *Knowledge Discovery in Database (KDD)* merupakan sebuah aktivitas untuk mengumpulkan data, pemakaian data historis untuk dapat menemukan informasi, pengetahuan, keteraturan, hubungan atau pola

dalam sebuah data yang berukuran besar. *Output* dalam *data mining* sendiri dapat digunakan untuk dijadikan alternatif dalam mengambil keputusan atau untuk memperbaiki pengambilan keputusan di masa depan (Bulolo, 2020).

Terdapat tahapan KDD dalam data mining sebagai berikut (Fayyad, 1996):

1. Seleksi Data

Seleksi data adalah proses pengambilan data yang berhubungan dengan analisis dari data dan dilakukan teknik perolehan sebuah pengurangan representasi dari data yang bertujuan untuk meminimalkan hilangnya informasi data. Pemilihan data atau seleksi data dari sekumpulan data operasional dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data dari hasil seleksi tersebut akan digunakan untuk proses *data mining*.

2. *Pre-processing Data*

Sebelum dilakukan proses *data mining* perlu terlebih dahulu dilakukan *pre-processing data* yang mencakup membuang duplikasi data, memeriksa data yang hilang, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data.

3. Transformasi Data

Transformasi dilakukan setelah data telah dipilih, proses transformasi dilakukan untuk menyesuaikannya dengan proses *data mining*. Salah satu teknik pemrosesan data yang disebut normalisasi adalah untuk meminimalkan rentang nilai *output* atau fungsi untuk mencegah anomali dan ketidaksesuaian data. Ada beberapa teknik dalam normalisasi data, salah satunya adalah (Lewis, 2017):

- a. *Normalization*

Teknik *Standar Scaling* merupakan teknik digunakan untuk mengatasi perbedaan nilai yang cukup besar antara dataset. Nilai data aktual diubah menjadi nilai dengan skala (0,1) tanpa mengubah informasi sebelumnya.

4. *Data Mining*

Data mining adalah cara atau proses untuk menemukan pola data yang cocok dengan data yang ditampilkan menggunakan metode atau teknik tertentu.

5. Interpretasi atau Evaluasi

Pola data atau informasi yang dihasilkan oleh proses ekstraksi data harus ditampilkan dengan cara yang mudah dimengerti oleh para yang berkepentingan.

2.4 Stationeritas

Menurut makridaris (1999), dalam analisis deret waktu, kestasioneran adalah asumsi penting bahwa data tidak berubah secara fluktuatif. Data deret waktu dianggap stationer jika nilai rata-rata dan variansnya tetap konstan selama periode waktu tertentu.

Uji *Augmented Dickey Fuller* (ADF) adalah salah satu uji yang digunakan untuk menentukan apakah data stasioner atau tidak. Uji ini menguji apakah *unit root* ada dalam model autoregresif dengan menguji hipotesis nol (H_0), dan persamaan ADF adalah sebagai berikut:

$$\Delta Y_t = \phi Y_{t-1} + \sum_{j=1}^{p-1} \alpha_j^* \Delta Y_{t-j} + \varepsilon_t \quad (2.2)$$

dengan:

ΔY_t = bentuk dari pembeda pertama

Y_{t-1} = Nilai variabel pada waktu ke t-1

ϕ = parameter

p = panjang lag

α_j^* = suatu konstanta

ε_t = residual pada waktu ke-t

Hipotesis yang diuji

$H_0: \phi \geq 1$ (data tidak stationer)

$H_1: \phi < 1$ (data stationer)

Statistik uji:

$$t_{hitung} = \frac{\hat{\phi} - 1}{sd(\hat{\phi})} \quad (2.3)$$

dengan:

$\hat{\phi}$ = Koefisien *autoregressive* (AR)

$sd(\hat{\phi})$ = varians dari $\hat{\phi}$

Taraf signifikansi: $\alpha = 5\%$

Dengan kriteria uji:

Hipotesis H_0 ditolak apabila nilai p-value $< \alpha$ yang dapat disimpulkan bahwa data deret waktu sudah stationer.

2.5 *Autocorrelation Function*

Fungsi autokorelasi, juga dikenal sebagai ACF, adalah hubungan antara satu atau lebih variabel yang menunjukkan korelasi antara data deret waktu terhadap lag atau selang waktu yang berbeda. Menurut Analita dan Irhamah (2015), lag pada ACF yang menghasilkan model *moving average* yang akurat hanya 10%, yang mencakup lag awal. Dan dapat digunakan untuk mengetahui apakah data stationer.

2.6 *Partial Autocorrelation Function*

Menurut Analita dan Irhamah (2015), Autokorelasi parsial, juga dikenal sebagai PACF, adalah ukuran seberapa dekat suatu variabel dengan variabel bebas lainnya. lag pada PACF, yang menghasilkan model rata-rata pergerakan yang akurat hanya 10% yang mencakup lag awal. Dan dapat digunakan untuk mengetahui apakah data stationer.

2.7 *Autoregressive Integrated Moving Average (ARIMA)*

Menurut Rezaldi dan Sugiman (2021), Metode ARIMA adalah metode yang dapat menghasilkan prediksi berdasarkan pola data historis. Metode ARIMA adalah kombinasi dari model AR atau model autoregresif yang merupakan model untuk menjelaskan pergerakan variabel melalui variabel itu sendiri di masa lalu dan model MA atau model rata-rata bergerak yang merupakan model untuk melihat pergerakan residual dari masa lalu.

Meunurut Purwanto, dkk (2020), terdapat klasifikasi dari model ARIMA:

1. Model *Autoregressive* (AR)

Model *Autoregressive* diperkenalkan oleh Yule pada tahun 1926 dan kemudian dikembangkan oleh Walker pada tahun 1931 Model regresi adalah bentuk regresi yang menghubungkan nilai saat ini dari pengamatan dengan nilai pengamatan sebelumnya pada interval waktu tertentu. Berikut merupakan bentuk persamaan umum dari model AR dengan ordo p (AR(p)) atau model ARIMA ($p,0,0$):

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \quad (2.3)$$

Keterangan:

Y_t = data pada periode ke- t , dimana $t = 1, 2, \dots, n$

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ = nilai lampau data yang bersangkutan

ϕ_0 = konstanta

$\phi_1, \phi_2, \dots, \phi_p$ = parameter koefisien *Autoregressive*

e_t = galat

2. Model *Moving Average* (MA)

Moving Average atau MA digunakan untuk menjelaskan fenomena di mana pengamatan dalam t waktu dinyatakan sebagai kombinasi linear dari sejumlah

kesalahan acak. *Moving Average* bergerak diperkenalkan oleh Slutsky pada tahun 1973 dan dikembangkan oleh Wadsworth pada tahun 1989. Berikut adalah bentuk umum dari persamaan *Moving Average* bergerak dengan urutan q (MA(q)) atau model ARIMA (0,0, q):

$$Y_t = \theta_0 + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (2.4)$$

Keterangan:

Y_t	= data pada periode ke- t
e_t	= galat pada periode ke- t
θ_0	= konstanta
$\theta_1, \theta_2, \dots, \theta_p$	= parameter koefisien <i>Moving Average</i>

3. Model *Autoregressive Moving Average* (ARMA)

Metode ARMA sering disebut sebagai metode Box-Jenkins karena dikembangkan oleh George Box dan Gwilym Jenkins pada tahun 1976. Model *Autoregressive Integrated Moving Average* (ARMA) mengasumsikan bahwa data deret waktu periode berjalan dipengaruhi oleh data periode sebelumnya dan nilai residual periode sebelumnya. Berikut merupakan persamaan bentuk umum dari model rata-rata bergerak autoregresif dengan ordo p (AR(p)) dan ordo q (MA(q)) atau model ARMA:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} \dots - \theta_q e_{t-q} \quad (2.4)$$

Keterangan:

Y_t	= data pada periode ke- t , dimana $t = 1, 2, \dots, n$
Y_{t-1}	= nilai lampau series yang bersangkutan
e_{t-1}	= kesalahan (error)
e_t	= kesalahan peramalan
ϕ_0	= konstanta
ϕ_1, \dots, ϕ_p	= parameter koefisien <i>Autoregressive</i>
$\theta_1, \dots, \theta_p$	= parameter koefisien <i>Moving Average</i>

4. ARIMA dikenal sebagai metode deret waktu *Box-Jenkins*. ARIMA memiliki akurasi yang sangat baik untuk peramalan jangka pendek, dan kurang baik untuk akurasi peramalan jangka panjang. Biasanya, akan datar atau konstan untuk jangka waktu yang lama.

Model ARIMA adalah model yang sama sekali mengabaikan variabel independen dalam membuat peramalan. ARIMA menggunakan nilai masa lalu dan masa sekarang dari variabel dependen untuk menghasilkan perkiraan jangka pendek yang akurat. Metode ARIMA akan tepat jika pengamatan deret waktu secara statistik berhubungan satu dengan yang lain atau saling tergantung.

Sebelum digunakan untuk peramalan metode ARIMA memiliki tiga langkah dasar yaitu berupa tahap identifikasi, tahap penilaian, dan tahap pengujian, serta terdapat juga tahap dianostik. Peramalan moel ARIMA dapat digunakan jika model yang diperoleh sudah memnuhi.

Terdapat sebuah asumsi yang digunakan Model AR, MA dan ARMA dimana asumsi yang digunakan adalah data deret waktu stationer. Faktanya data deret waktu sebagian besar tidak stasioner. Jika data deret waktu tidak stasioner, cara yang dapat digunakan untuk membuat data tersebut stasioner yaitu melakukan pembedaan untuk data yang rata-rata tidak stasioner dan dilakukan proses transformasi data untuk data yang tidak stasioner dalam varians. Bentuk persamaan umum model ARIMA (p,d,q) sebagai berikut:

$$\phi_p(B)(1-B)^d Y_t = \phi_0 + \theta_q(B)e_t$$

Dimana $p = 1, 2, \dots, p$ dan $q = 1, 2, \dots, q$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d Y_t = \phi_0 + e_t - \theta_1 e_{t-1} \dots - \theta_q e_{t-q}$$

$$1 - B^d + \phi_1 B^d + \phi_1 B^{1+d} + \dots + \phi_p B^{p+d}) Y_t = \phi_0 + e_t - \theta_1 e_{t-1} \dots - \theta_q e_{t-q}$$

Menggunakan persamaan:

$$B^i Y_t = Y_{t-i}$$

Maka diperoleh:

$$\begin{aligned} Y_t + Y_{t-i} - \phi_1 Y_{t-1-i} - \dots - \phi_p Y_{t-p-i} &= \phi_0 + e_t - \theta_1 e_{t-1} \dots - \theta_q e_{t-q} \\ Y_t &= \phi_0 + Y_{t-i} - \phi_1 Y_{t-1-i} - \dots - \phi_p Y_{t-p-i} + e_t - \theta_1 e_{t-1} \dots - \theta_q e_{t-q} \end{aligned}$$

Sehingga diperoleh persamaan model ARIMA:

$$Y_t = \phi_0 + Y_{t-i} - \phi_1 Y_{t-1-i} - \dots - \phi_p Y_{t-p-i} + e_t - \theta_1 e_{t-1} \dots - \theta_q e_{t-q} \quad (2.5)$$

Keterangan:

Y_t = nilai variabel pada waktu ke-t

ϕ_0 = konstanta

e_t = galat pada periode ke-t

$(1 - B)^d$ = proses pembedaan

ϕ_p = operator langkah mundur untuk AR

θ_q = operator langkah mundur untuk MA

2.8 Uji *White Noise*

Menurut Ekananda (2014), Uji *white noise* adalah uji yang paling umum digunakan untuk memeriksa kecocokan model dengan melihat apakah hasil residual dari pemodelan data deret waktu memenuhi uji *white noise*. Uji L-Jung-Box menunjukkan bahwa jika p-value lebih besar dari taraf signifikansi atau alpha memenuhi asumsi uji *white noise*. Coba uji suara putih ini: Uji *white noise* sebagai berikut:

$$Q = n(n+2) \sum_{k=1}^k \frac{\hat{\rho}_k^2}{n-k} \quad (2.6)$$

Keterangan:

- Q = Uji *White Noise*
 n = banyak data
 k = nilai pada lag
 K = nilai maksimum lag
 ρ_k = nilai perkiraan autokorelasi residual lag-k

2.9 *Machine Learning*

Pembelajaran mesin adalah bagian dari kecerdasan buatan yang memiliki fungsi untuk membuat komputer belajar tentang data baru tanpa harus secara eksplisit diprogram. Fokus utama dari pembelajaran ini adalah membuat komputer membangun aplikasi yang dapat mempelajari data, kemudian membuat model pembelajaran yang siap dan dapat digunakan untuk memecahkan kasus tertentu. Terdapat proses yang digunakan dalam pembelajaran mesin yang biasa disebut proses klasifikasi. Di mana data yang digunakan adalah data pelatihan, maka Anda harus memiliki informasi yang dilengkapi dengan data kategori atau label. Pelatihan data adalah proses untuk menemukan fungsi pembatasan. (Daqiqil, 2021).

Machine learning dapat dikelompokkan dengan berdasarkan bagaimana cara belajar sehingga mampu dalam melakukan tugasnya, berdasarkan cara belajarnya yaitu:

1. *Supervised Learning*

Supervised Learning adalah pembelajaran yang diarahkan atau diawasi.

Dalam inti proses pembelajaran, sistem menerima data yang berisi solusi atau hasil yang diinginkan atau contoh kasus yang akan diselesaikan oleh sistem.

2. *Unsupervised Learning*

Unsupervised learning merupakan pembelajaran tanpa petunjuk dalam prosesnya. Dan berupa algoritma dalam komputer yang dapat bekerja untuk menemukan beberapa pola dalam data. Secara sistematis, pembelajaran yang

tidak diawasi dapat terjadi jika ada sejumlah data input (X) dan tidak ada variabel output yang terkait. Pembelajaran yang tidak diawasi dapat dibagi menjadi dua jenis: asosiasi dan kelompok.

2.10 *Deep Learning*

Deep learning atau pembelajaran mendalam merupakan suatu cabang dari *Machine Learning* yang berupa algoritma pemodelan menggunakan fungsi transformasi tidak linear dengan abstraksi tingkat tinggi yang ditata berlapis-lapis dan mendalam (Cholissodin, dkk., 2019). *Deep learning* adalah pembelajaran yang berbasis pada fitur yang berbentuk hirarki.

Deep learning memiliki beberapa algoritma diantaranya *Convolution Neural Network* (CNN) dan *Recurrent Neural Network* (RNN) (Rowan, dkk., 2022).

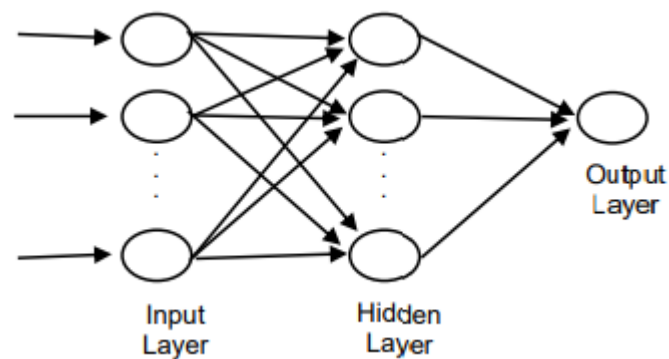
2.11 **Artificial Neural Network**

Jaringan saraf buatan, juga dikenal sebagai *Artificial Neural Network* (ANN), adalah sistem pemrosesan informasi yang menyerupai jaringan saraf biologis. ANN adalah representasi dari jaringan otak manusia, yang terdiri dari jutaan neuron atau sel saraf yang berfungsi untuk memproses informasi, dan ANN terdiri dari kelompok yang disebut lapisan. Neuron di satu lapisan terhubung ke lapisan lain. Berat atau kekuatan hubungan menunjukkan kekuatan hubungan antara neuron terdekat.

Sebuah ANN biasanya terdiri dari tiga layer yaitu terdiri dari *input layer*, *hidden layer*, dan *output layer*. *Input layer* terdiri dari neuron yang menerima input dari lingkungan eksternal. Input adalah deskripsi masalah. *Hidden layer* terdiri dari neuron yang menerima input dari lapisan input, dan kemudian membawa output

ke lapisan berikutnya. *Output layer* juga disebut unit output, yang terdiri dari lapisan tersembunyi dan mengirimkannya ke pengguna.

Langkah penting dalam pengembangan model ANN adalah menentukan berat matriks melalui pelatihan atau pelatihan. Ada dua jenis pelatihan: pelatihan yang diawasi dan pelatihan tanpa pengawasan. Pelatihan yang diawasi membutuhkan pengawasan eksternal untuk membimbing proses pelatihan. Algoritma ini menggunakan sejumlah pasang data input dan output yang digunakan sebagai contoh, di mana data yang digunakan adalah yang terbaik menggunakan data yang sudah diketahui benar. Hasil jaringan dibandingkan dengan data output yang diharapkan atau contoh output untuk mendapatkan perbedaan antara output jaringan dan output aktual. Perbedaan ini digunakan untuk mengubah berat jaringan sehingga *output* sama atau dekat dengan target. Mekanisme pelatihan yang diawasi yang disebut pelatihan umpan balik umumnya digunakan dalam aplikasi teknik. Karena ANN tidak mempertimbangkan fisik masalah mereka, ANN adalah model *black box*, tetapi dapat mendeteksi proses fisik pada model ANN yang telah dilatih. (Dharma, dkk., 2011).



Gambar 1. Arsitekur dari jaringan syaraf

Sumber: (Ali, dkk.; 2014)

Namun pada ANN tidak dapat mengingat informasi yang diberikan sebelumnya. Hal ini diatasi oleh *Recurrent Neural Network* (RNN). RNN merupakan jaringan dengan *loop* di dalamnya, memungkinkan untuk informasi bertahan.

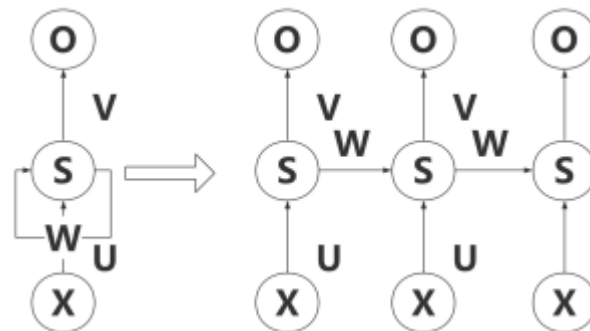
2.12 Recurrent Neural Network

Recurrent Neural Network merupakan sejenis dari *Artificial Neural Network* yang memiliki kemampuan untuk melihat korelasi tersembunyi pada data dalam aplikasi untuk pemrosesan bahasa alami, pengenalan suara, dan prediksi data deret waktu. RNN ini sangat cocok digunakan untuk masalah *sequence* atau data beruntun dengan beroperasi pada informasi input serta jejak informasi yang diperoleh sebelumnya karena berulang. Model RNN dimodelkan sebagai berikut:

$$S_t = f((U * X_t) + (W * S_{t-1})) \quad (2.7)$$

$$O_t = g(V * S_t) \quad (2.8)$$

Dimana S_t merupakan memori dari jaringan pada waktu t ; U , W , dan V merupakan matriks bobot yang terbagi di setiap lapisan; X_t dan O_t representasi dari *input* dan *output* pada waktu t ; dan $f(\cdot)$ dan $g(\cdot)$ representasi dari fungsi tidak linear.



Gambar 2. Struktur sederhana *Recurrent Neural Network*

Sumber : (Tian, dkk., 2018)

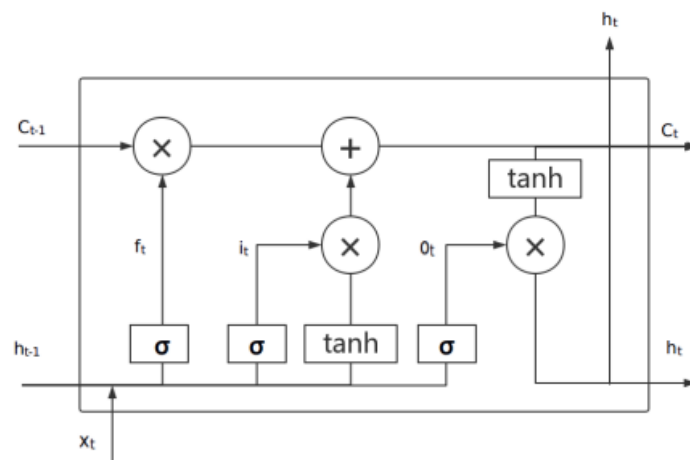
Dalam memproses urutan input, RNN dapat menggunakan keadaan internal. Ini berbeda dengan koneksi bobot yang dibangun antara lapisan jaringan saraf dasar. Status tersembunyi menyimpan memori sebelumnya. Ketika *output* dekat dengan input yang terkait dengan informasi dari node sebelumnya dikirim kembali ke

node berikutnya, RNN bekerja dengan baik. Namun, dalam aplikasi praktis, karena masalah gradient menghilang, RNN tidak dapat menyimpan informasi sebelumnya dengan akurat dalam interval waktu yang lama. Konstruksi RNN khusus yang disebut LSTM diusulkan untuk mengatasi kelemahan ini dan meningkatkan kinerja.. (Tian, dkk., 2018).

2.13 Long Short Term Memory (LSTM)

Long Short Term Memory memakai bentuk yang paling umum dari RNN. Yang dimaksudkan adalah untuk menghindari masalah ketergantungan jangka panjang dan sesuai untuk memproses serta memprediksi deret waktu. Seperti telah pada tahun 1997 oleh Sepp Hochreiter dan Jurgen Schmidhuber mengusulkan model serangkaian sel memori unik dari LSTM untuk menggantikan neuron lapisan tersembunyi dari RNN, dan kunci dari sel memori.

Model LSTM menggunakan struktur *gate* dalam menyaring informasi untuk memperbarui dan mempertahankan keadaan sel memori. Struktur dari pintu ini terdiri dari *input*, *forget* dan *output gates*. Setiap memori sel terdapat sebanyak tiga lapisan sigmoid dan satu lapisan tanh (Qiu, dkk., 2020).



Gambar 3. Struktur *time step* dari LSTM
Sumber : (Tian, dkk., 2018)

Persamaan fungsi sigmoid dan tanh sebagai berikut:

1. Fungsi Sigmoid

Lapisan sigmoid menunjukkan antara angka nol dan satu, angka ini menggambarkan seberapa banyak masing-masing komponen yang harus dilewati. Nilai nol sendiri memiliki arti “jangan biarkan apa pun lewat” dan angka satu memiliki arti yaitu “biarkan semuanya lewat”. Sigmoid memiliki rentang dimana jika terletak di rentang $0 < \sigma \leq 0,5$ artinya bernilai nol dan jika terletak pada rentang $0,5 < \sigma \leq 1$ artinya bernilai satu (R, 2017).

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.9)$$

Keterangan:

x = data input

e = konstanta matematika

2. Fungsi tanh

Menurut Lewis (2017), fungsi tanh atau disebut dengan tangen hiperbolik memiliki rentang nilai -1 sampai dengan 1. Fungsi tanh memiliki rentang nilai yang lebih luas dari fungsi sigmoid, sehingga lebih efektif digunakan untuk pemodelan tidak linier yang kompleks. Persamaan fungsi tanh sebagai berikut:

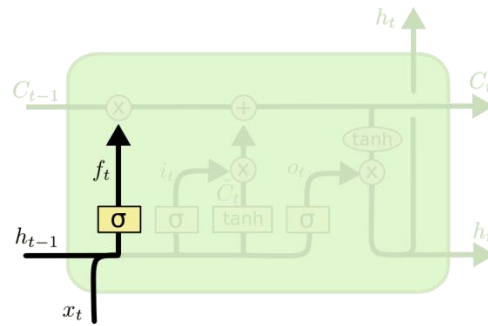
$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.10)$$

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} \quad (2.11)$$

Keterangan:

x = data input

e = konstanta matematika



Gambar 4. *Forget Gate* dari struktur LSTM

Sumber : (Syahram, dkk., 2021)

Langkah awal, menentukan informasi yang akan dibuang atau yang akan disimpan. Keputusan ini ditentukan dengan menggunakan lapisan sigmoid yang disebut *forget gate* (Syahram, dkk., 2021). *Forget gate* pada LSTM unit digunakan dalam menentukan informasi mana pada status sel yang akan dibuang dari model. Seperti yang dapat dilihat pada gambar, *output* diterima oleh sel memori h_{t-1} dari informasi eksteral dan momen sebelumnya atau dari luar x_t dari momen yang biasanya sebagai *input* dan mengkombinasikannya di dalam vektor $[h_{t-1}, x_t]$ melalui transformasi σ dan menjadi bentuk persamaan: (Qiu, dkk., 2020).

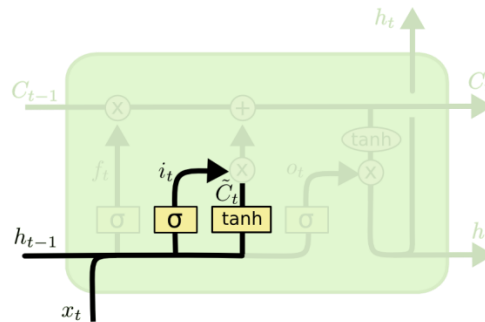
$$f_t = \sigma(W_t \cdot [h_{t-1}, x_t] + b_f) \quad (2.12)$$

Keterangan:

- σ = fungsi sigmoid
- W_t = nilai *weight* untuk *forget gate*
- b_f = nilai *bias* untuk *forget gate*
- x_t = nilai *input* sebelum orde ke-t
- h_{t-1} = nilai *output* sebelum orde ke-t
- f_t = *forget gate*

W_t dan b_f diperjelas sebagai matriks *weight* dan *bias* dari *forgotten gate*, dan σ adalah fungsi sigmoid. *Forgotten gate* memiliki fungsi utama untuk merekam seberapa banyak status sel C_{t-1} dari waktu dicadangkan sebelumnya ke status sel

C_t ke waktu sekarang. Lalu *gates* akan memberikan nilai *output* antara nilai 0 dan 1 berdasar pada h_{t-1} dan x_t , dimana 1 berupa reservasi lengkap dan 0 berupa pembangunan lengkap (Qiu, dkk., 2020).



Gambar 5. *Input gate* untuk struktur LSTM

Sumber : (Syahram, dkk., 2021)

Selanjutnya, *input gate* merupakan lapisan neuron dengan fungsi aktivasi berupa fungsi sigmoid dan memiliki fungsi aktivasi berupa fungsi tanh (Syahram, dkk., 2021). *Input gate* menentukan seberapa banyak jaringan *input* pada waktu x_t atau waktu sekarang dicadangkan ke status sel C_t , dimana untuk mencegah konten yang memasuki sel memori tidak signifikan. Gerbang *input* ini mempunyai dua fungsi. Satu untuk menemukan status dari sel yang harus selalu diperbarui; nilai yang diperbarui ini dipilih oleh lapisan sigmoid, dalam persamaan sebagai berikut: (Qiu, dkk., 2020).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.13)$$

Keterangan:

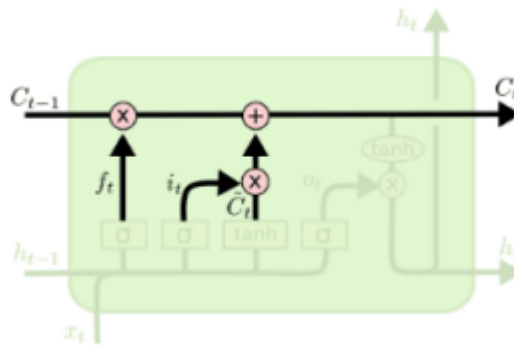
σ = fungsi sigmoid

W_i = nilai *weight* untuk *input gate*

b_i = nilai *bias* untuk *input gate*

x_t = nilai *input* pada orde ke-t

h_{t-1} = nilai *output* pada orde ke-t



Gambar 6. *Cell state* untuk struktur LSTM

Sumber : (Syahram, dkk., 2021)

Selanjutnya, status sel yang lama (C_{t-1}) untuk kemudian dimasukkan ke dalam status sel yang baru (C_t). Dengan cara dikalikan dengan f_t atau kondisi lama untuk dilupakan (Syahram, dkk., 2021). Fungsi lainnya memperbarui informasi ke b diperbarui kembali ke keadaan sel. Terdapat vektor kandidat baru \hat{C} yang dibuat melalui lapisan tanh untuk mengontrol bertambahnya informasi. Persamaan untuk memperbarui keadaan sel (Qiu, dkk., 2020).

$$\hat{C} = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2.14)$$

Keterangan:

\hat{C} = nilai baru ditambahkan ke *cell state*

\tanh = fungsi tanh

W_c = nilai *weight cell state*

h_{t-1} = nilai output sebelum ordo ke-t

x_t = nilai input pada ordo ke-t

b_c = nilai *bias cell state*

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \quad (2.15)$$

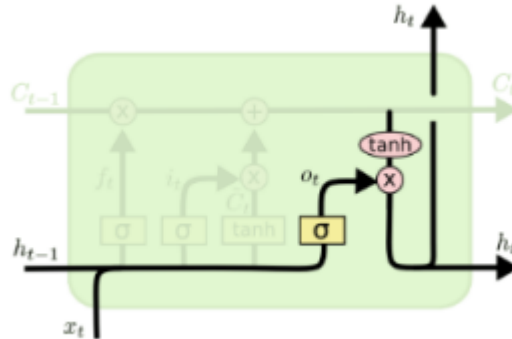
C_t = nilai baru ditambahkan ke *cell state*

f_t = *forget gate*

C_{t-1} = *cell state* sebelum ordo ke-t

i_t = *input gate*

\hat{C}_t = nilai baru yang ditambahkan ke *cell state*



Gambar 7. *Output gate* untuk struktur LSTM
Sumber : (Syahram, dkk., 2021)

Banyak keadaan sel saat ini dibuang akan dikontrol oleh *Output gate*. *Output gate*, menjalankan lapisan sigmoid dimana untuk menentukan bagian sel yang menjadi *output*. Berikutnya, pada tanh memproses *cell state* untuk mengubah nilai menjadi interval 1 dan -1 kemudian akan dikalikan dengan *output* dari lapisan sigmoid (Syahram, dkk., 2021).

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2.16)$$

Keterangan:

O_t = *output gate*

σ = fungsi sigmoid

W_o = nilai *weight* untuk *output gate*

h_{t-1} = nilai *output* sebelum ordo ke-t

x_t = nilai *input* pada ordo ke-t

b_o = nilai *bias* pada *output gate*

Nilai output akhir sel didefinisikan sebagai berikut:

$$h_t = O_t * \tanh(C_t) \quad (2.17)$$

Keterangan:

h_t = nilai output orde ke-t

O_t = output gate

$Tanh$ = fungsi tanh

C_t = *cell state*

2.14 Hybrid ARIMA – LSTM

Menurut Zhang (2001), model *Hybrid* adalah suatu metode penggabungan dua atau lebih model dalam satu sistem. Alasan metode *hybrid* dikembangkan adalah adanya permasalahan deret waktu yang murni linier maupun tidak linier yang jarang terjadi. Pola data deret waktu sering mengandung keduanya. Jika terjadi pola deret waktu seperti itu sering terjadi maka menggunakan metode tunggal bisa dinilai kurang efektif untuk meramalkan deret waktu karena tidak dapat menangkap pola campuran.

Menurut Opitz dan Maclin (1999), banyak terdapat model deret waktu yang menyertakan hubungan linear dan hubungan tidak linear. Model ARIMA baik digunakan dalam memodelkan hubungan linear dalam rangkaian deret waktu, tetapi tidak cukup untuk memodelkan hubungan yang tidak linear. Sedangkan untuk jaringan syaraf tiruan LSTM dapat memodelkan hubungan yang linier dan tidak linier tetapi tidak dapat memberikan hasil yang sama untuk setiap kumpulan data. Maka dari itu, untuk mendapatkan hasil prediksi terbaik, digunakan model gabungan berdasarkan prinsip pemodelan terpisah dari komponen deret waktu linier dan tidak linier.

Rumus prediksi deret waktu dari model yang dihasilkan secara umum dapat dinyatakan sebagai penjumlahan komponen linier dan tidak linier, dan memiliki persamaan sebagai berikut (Zhang, 2001):

$$y_t = L_t + N_t \quad (2.18)$$

Keterangan:

L_t = komponen linear dari deret waktu

N_t = komponen tidak linear dari deret waktu

Langkah pertama adalah menggunakan ARIMA untuk memodelkan komponen linear, kemudian residual yang didapatkan dari model linear akan berisi hubungan yang tidak linear dari data deret waktu.

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (2.19)$$

Metodologi yang diusulkan oleh *hybrid* terdiri dari dua langkah. Yaitu pada langkah awal, bagian linear dari masalah akan dianalisis oleh model ARIMA. Langkah kedua, dikembangkan, model jaringan syaraf tiruan yang berguna dalam memodelkan residual dari hasil model ARIMA. Dikarenakan model ARIMA tidak mampu menangkap struktur komponen tidak linear, residual dari model linear memberikan informasi yang tidak linear tersebut. Hasil dari jaringan syaraf tiruan berupa prediksi akan digunakan dari *error term* untuk model ARIMA.

2.15 Akurasi Model

Ukuran akurasi diperlukan untuk membandingkan satu mode-model alternatif dengan yang lain dan untuk menentukan nilai parameter yang muncul dari ekspresi untuk fungsi prediksi F. Setiap model pada data yang lebih dulu digunakan untuk mengidentifikasi yang paling akurat, dan model dengan kesalahan terkecil dipilih. Setelah model prediktif telah dikembangkan dan digunakan untuk memberikan prediksi untuk waktu yang akan datang, diperlukan menilai keakuratan agar dapat mendeteksi kekurangan atau kelainan pada data, dan mengembangkan model prediksi yang mungkin muncul dalam periode waktu

berikutnya (Budiman, 2016):. Untuk menilai keakuratan dan konsistensi dari asumsi model kinerja yang berbeda, digunakan evaluasi berikut.

1. *Akaike Information Criterion (AIC)*

Merupakan ukuran perbandingan yang digunakan untuk memilih model sebagai penentuan model terbaik dari data deret waktu dan juga untuk mendeskripsikan data yang memiliki sejumlah parameter terbaik untuk digunakan dalam meramalkan data. Jika nilai AIC yang diperoleh semakin kecil maka model tersebut semakin baik digunakan untuk peramalan.

2. *Mean Square Error (MSE)*

merupakan penjumlahan kuadrat error atau selisih antara nilai sebenarnya dengan nilai prediksi, kemudian membagi hasil penjumlahan tersebut dengan banyaknya waktu peramalan. Dengan persamaan sebagai berikut:

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} \quad (2.20)$$

Keterangan:

Y_i = nilai data asli

\hat{Y}_i = nilai data peramalan

n = jumlah periode prediksi

3. *Root Mean Square Error (RMSE)*

Merupakan penjumlahan kuadrat error atau selisih antara nilai sebenarnya dan nilai prediksi, dan membagi hasil penjumlahan tersebut dengan banyaknya waktu peramalan dan menarik akarnya. Dengan persamaan sebagai berikut:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (2.21)$$

Keterangan:

n : jumlah periode prediksi

Y_i : nilai data asli

\hat{Y}_i : nilai data peramalan

4. *Mean Absolute Percentage Error* (MAPE)

Merupakan nilai yang *absolute* dari presentase error data terhadap rata-rata.

Dengan persamaan sebagai berikut:

$$MAPE = \frac{\sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i}}{n} \times 100\% \quad (2.22)$$

Keterangan:

n = banyaknya periode prediksi

Y_i = nilai data actual

\hat{Y}_i = nilai data peramalan

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung, pada semester ganjil Tahun Ajaran 2022/2023.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari situs IDN *Financial* dengan link <https://www.idnfinancials.com/id/mncn/pt-media-nusantara-citra-tbk> yang merupakan data time series mengenai nilai saham PT. Media Nusantara Tbk terhitung berjangka selama 8 tahun terhitung sejak November 2015 sampai dengan Februari 2023 dalam skala mingguan. Dengan jumlah data sebanyak 380.

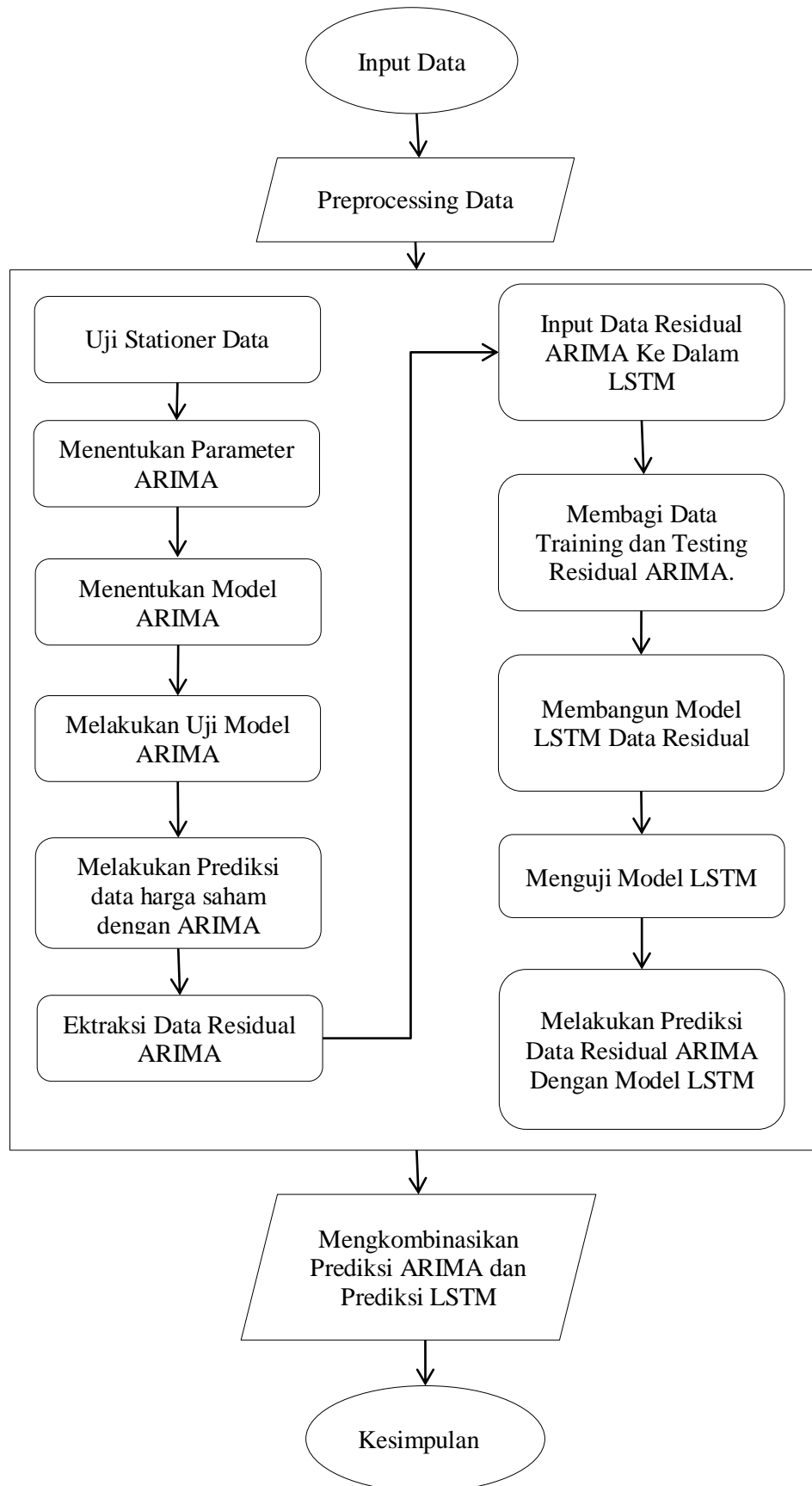
3.3 Metode Penelitian

Proses analisis data dilakukan dengan menggunakan bantuan *Google Colab* guna memudahkan peneliti dalam menganalisa.

Langkah-langkah yang dilakukan dalam penelitian ini adalah:

1. Mengumpulkan data yang akan digunakan untuk proses peramalan

2. Melakukan *pre-processing* data agar dapat diproses dengan menggunakan ARIMA dan LSTM
3. Melakukan pembentukan model ARIMA
4. Berdasarkan hasil dari pemodelan ARIMA, kemudian dilakukan prediksi nilai harga saham
5. Melakukan ekstraksi data residual yang merupakan selisih antar data prediksi ARIMA dengan data aktual
6. Data residual yang didapat dari model ARIMA digunakan sebagai input dengan menggunakan LSTM.
7. Melakukan *splitting* untuk data residual.
8. Membuat model utama data residual dengan menggunakan LSTM
9. Penentuan parameter model LSTM untuk prediksi dari data residual ARIMA
10. Melakukan prediksi dengan model LSTM menggunakan data residual ARIMA.
11. Menggabungkan hasil prediksi dari kedua model utama ARIMA dan model LSTM dengan proses penjumlahan
12. Hasil dari penggabungan data prediksi model ARIMA dan model LSTM merupakan model *hybrid* ARIMA-LSTM.



Gambar 8. Diagram alir proses *Hybrid* ARIMA-LSTM

V. KESIMPULAN

Berdasarkan hasil pembahasan yang sudah diperoleh didapatkan kesimpulan sebagai berikut:

1. Metode *Hybrid* ARIMA-LSTM dapat digunakan untuk memprediksi harga saham media konvensional pada indeks “Harga” dengan trend turun.
2. Parameter terbaik yang digunakan untuk memprediksi data dengan ARIMA adalah model ARIMA(0,1,1) dengan parameter untuk residual menggunakan LSTM adalah 32 untuk LSTM units, 64 batch size, 0,1 untuk dropout dan 100 untuk epoch.
3. Dari perbandingan akurasi menggunakan RMSE dan MAPE dapat diketahui bahwa model *hybrid* ARIMA-LSTM dapat memberikan akurasi terbaik dengan nilai error atau kesalahan paling kecil yaitu RMSE sebesar 43,15281 dan nilai MAPE sebesar 0,0074 atau 0,0074%. Hal ini dapat diartikan bahwa tingkat akurasi dengan model *hybrid* ARIMA-LSTM sebesar 99,99%. Oleh karena itu, model *hybrid* ARIMA-LSTM dapat digunakan sebaik mungkin untuk meramalan dan memprediksi harga saham yang lain.
4. Didapatkan model *Hybrid* ARIMA-LSTM dapat memprediksi lebih baik dari model LSTM tetapi kurang dari model ARIMA diukur dengan keberhasilan MAPE.

DAFTAR PUSTAKA

- Andalita, I. dan Irhamah. 2015. *Peramalan jumlah penumpang kereta api kelas ekonomi kertajaya menggunakan ARIMA dan ANFIS*. Jurnal Sains dan Seni ITS. **2**(4): 311-316.
- Ali, Y.H., Rahman, R.A., dan Hamzah, R.I.R. 2014. *Acoustic Emission Signal Analysis and Artificial Intelligence Techniques In Machine Condition Monitoring and Fault Diagnosis: A Review*. Jurnal Teknologi, Sains dan Teknik Universitas Tenknik Malaysia. **69**(2): 121-126.
- Brilliant, M., Nurhasanah, I.A., dan Oktaria, H. 2021. *Perbandingan Metode Deep Learning Recurrent Neural Networks, ARIMA dan Holt-Winters Exponential Smoothing Dalam Prediksi Pola Transaksi Pelanggan*. Journal of Software Engineering And Technology. **1**(1): 29-36.
- Budiman, H. 2016. *Analisis dan Perbandingan Akurasi Model Prediksi Rentet Waktu Support Vector Machines Dengan Support Vector Machines Particle Swarm Optimization Untuk Arus Lalu Lintas Jangka Pendek*. Systemic: Information System and Informatics Journal. **2**(1): 19–24.
- Buulolo, E. 2020. *Data Mining Untuk Perguruan Tinggi*. Deepublish, Yogyakarta.

- Cholissodin, I., Sutrisno, Soebroto, A.A., Hasanah, U., dan Febiola, Y.I. 2019. *AI, Machine Learning & Deep Learning (Teori & Implementasi)*. Fakultas Ilmu Komputer Universitas Brawijaya, Malang.
- Daqiqil, I. 2021. *Machine Learning: Teori, Studi, Kasus dan Implementasi Menggunakan Python*. Pekanbaru, Riau.
- Dharma, I. S., Putera, I. A., & Ardana, P. D. 2011. *Artificial Neural Networks untuk Pemodelan Curah Hujan Limpasan pada Daerah Aliran Sungai (DAS) di Pulau Bali*. Journal of Environment Universitas Udayana. (11)1: 9-22.
- Durbin, J., dan Koopman, S.J. 2012. *Time Series Analysis by State Space Methods 2nd edition*. Oxford University Press, United Kingdom.
- Ekananda. 2014. *Analisis Data Time Series*. Mitra Wacana Media, Jakarta.
- Fathi, O. 2019. *Time series forecasting using a ahybrid ARIMA and LSTM model*. Velvet Consulting.
- Fayyad, U. 1996. *Advances in Knowledge Discovery and Data Mining*. MIT Press, USA.
- Haquq, R., dan Ersyad, F.A. 2020. *Eksistensi Media Televisi Era Digital Dikalangan Remaja*. Jurnal Dinamika Sosial Budaya USM. (22)1: 38-44.
- Herawati, S. 2013. *Peramalan Harga Saham Menggunakan Integrasi Empirical Mode Decomposition dan Jaringan Syaraf Tiruan*. Jurnal Mikrotek Universitas Trunojoyo. (1)1: 23-28

- Lewis, N.D. 2017. *Neural Network For Time Series Forecasting with R*. Pearson Education, New York.
- Makridakis, S., Wheelwright, S.C., and McGee, V.E. 1999. *Metode dan Aplikasi Peramalan Edisi Kedua*. Erlangga, Jakarta.
- Opitz, D., dan Maclin, R. 1999. *Metode ansambel populer: studi empiris*. Jurnal Buatan industri Universitas Sakarya. 169-198.
- Pankratz, A. 1991. *Forecasting with Dynamic Regression Models*. Willey Intersciences Publication. Canada.
- Porwanto, A., Pramono, R., Asbari, M., Hyun, C., Wijayanti, L., Putri, R., dan Santoso, P. 2020. *Studi Eksploratif Dampak Pandemi COVID-19 Terhadap Proses Pembelajaran Online di Sekolah Dasar*. Journal of Education, Psychology and Counseling UPH. (2)1: 2716-4446.
- Qiu, J., Wang, B., & Zhou, C. 2020. *Forecasting Stock Prices with Long-Short Term Memory Neural Network Based on Attention Mechanism*. PLOS ONE. (15)1: 2-3.
- Rezaldi, D.A., Sugiman. 2021. *Peramalan Metode ARIMA Data Saham PT. Telekomunikasi Indonesia*. PRISMA, Prosiding Seminar Nasional Matematika. (4): 611-620.
- Rowan, Muflikhah, L., dan Cholissodin, I. 2022. *Peramalan Kasus Positif COVID19 di Jawa Timur menggunakan Metode Hybrid ARIMA-LSTM*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. 6(9):4146-4153.

- Syahram, E.F., Effendy, M.M., dan Setyawan, N. 2021. *Sun Position Forecasting Menggunakan Metode RNN – LSTM Sebagai Referensi Pengendalian Daya Solar Cell*. JEETech. 65-72.
- Temur, A.S., Akgun, M., dan Temur, G. 2019. *Predicting Housing In Turkey Using ARIMA, LSTM and Hybrid Models*. Journal of Business Economics and Management. (5)20: 920-938.
- Tian, C., Ma, J., Zhang, C., dan Zhan, P. 2018. *A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network*. Journal Energies Beijing University. 1-13.
- Zheng, J., Xu, C., Zhang, Z., dan Li, X. 2017. *Electric Load Forecast in Smart Grid Using Long-Short-Term-Memory based Recurrent Neural Network*. Journal of Engineering CISS.
- Zhang, G.P. 2001. *Time Series Forecasting Using A Hybrid ARIMA and Neural Network Model*. Journal Elsevier Neuro Computing. 50: 159-175.