



Pemodelan *Time Series* Data Saham LQ45 dengan Algoritma LSTM, RNN, dan Arima

Muhammad Nur Alim^{a,*}

^aJurusan Matematika, FMIPA, UNNES, Jl. Raya Sekaran Gunungpati, Semarang 50229, Indonesia

* Alamat Surel: muhammadnuralim@students.unnes.ac.id

Abstrak

Perkembangan teknologi merubah cara manusia untuk melakukan kegiatan. Termasuk, dalam hal berinvestasi. Untuk memilih saham dalam berinvestasi dibutuhkan strategi dan analisis yang baik, analisis yang sering dipakai adalah peramalan deret waktu. Peramalan deret waktu disini dilakukan dengan memilih saham yang terdaftar pada *Liquid 45* (LQ45) yang dikeluarkan oleh Bursa Efek Indonesia (BEI) dan mengambilnya menjadi dataset melalui *finance.yahoo.com*, kemudian memodelkannya menggunakan metode tradisional *Autoregressive Integrated Moving Average* (ARIMA) serta metode modern yang populer seperti *Long-Short Term Memory* (LSTM), *Convolutional Neural Network-Long-Short Term Memory* (CNN-LSTM), dan *Convolutional Long-Short Term Memory* (ConvLSTM). Tujuan penelitian ini adalah untuk membandingkan metode Arima dengan metode LSTM, CNN-LSTM, dan ConvLSTM dalam meramalkan data saham.

Kata kunci:

Saham, *Time Series*, Arima, LSTM, ConvLSTM, CNN-LSTM.

© 2023 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

1. Pendahuluan

Pasar modal adalah salah satu ciptaan manusia yang paling berguna dalam dunia perekonomian. Tanpa adanya pasar modal banyak perusahaan tidak dapat mengembangkan usaha yang dilakukan dan akan sangat sulit melakukan ekspansi ke berbagai wilayah bahkan negara. Dari pasar modal ini didapatkan berbagai data saham, banyak investor dan peneliti yang berusaha memahami data saham ini untuk menentukan arah dan pergerakan dari data saham ini.

Fama (1995), menjelaskan bahwa pada pasar modal dengan pertukaran informasi yang sangat efektif terjadi saham akan berperilaku seperti *random walk*. Menurutnya ada tiga kategori efisiensi yaitu efisiensi lemah, efisiensi semi-kuat, dan efisiensi kuat. Ada berbagai faktor yang mempengaruhi efisiensi tersebut seperti faktor makro dan mikro ekonomi, faktor politik, permintaan dan penawaran, dan faktor *Gross Domestic Product* (GDP) dan *Consumer Price Index* (CPI).

Dalam menganalisis saham berbagai peneliti telah mengusulkan berbagai analisis antara lain analisis teknikal dan analisis dasar. Analisis dasar adalah analisis dengan melihat faktor kesehatan dari perusahaan baik dari segi keuangan dan segi struktural organisasinya. Sedangkan analisis teknikal menurut Bustos & Pomares-Quimbaya (2020), adalah dengan menggunakan input berupa data saham dan turunannya. Untuk memprediksi harga saham cukup melihat dari pola dan trend yang ditunjukkan oleh data saham.

Pada penelitian kali ini akan digunakan berbagai metode yang dapat digunakan untuk melakukan prediksi terhadap data saham, metode-metode yang dimaksud adalah sebagai berikut. Pendekatan *Long-short term memory* oleh Obthong et al (2020), pendekatan *neural network* oleh Hu et al., (2021) dengan recurrent neural network, dan pendekatan tradisional dengan menggunakan ARIMA.

1.1. Rumusah Masalah

Berdasarkan latar belakang penelitian pemodelan data saham yang telah dijabarkan sebelumnya, maka rumusan masalah yang akan diteliti adalah bagaimana pemodelan data saham LQ45 dengan menggunakan

To cite this article:

Alim, M. A. (2023). Pemodelan *Time Series* Data Saham LQ45 dengan Algoritma LSTM, CNN-LSTM, ConvLSTM, dan Arima. *PRISMA, Prosiding Seminar Nasional Matematika* 6, 694-710

algoritma Arima, LSTM, dan RNN serta perbandingan dari performa model yang dihasilkan dari algoritma Arima, LSTM, dan RNN?

1.2. Tujuan Penelitian

Adapun tujuan penelitian ini adalah untuk membangun model time series untuk pemodelan data saham dengan menggunakan algoritma Arima, LSTM, dan RNN serta mengetahui bagaimana performa model yang dihasilkan oleh Arima, LSTM, dan RNN.

2. Metode

Metode pada penelitian kali ini adalah metode *neural network* seperti RNN dan LSTM, dan metode tradisional Arima. Kedua metode ini akan dibandingkan hasilnya untuk melihat performa dari kedua metode ini. Untuk memulai proses penelitian dibutuhkan beberapa peralatan seperti laptop, python. Serta *package-package* yang diperlukan dalam analisis seperti numpy, pandas, matplotlib, statsmodels arima, sklearn.keras LSTM, sklearn.keras RNN.

2.1. Dataset data saham LQ45

Dataset yang digunakan kali ini adalah dataset yang diambil dari *yahoo finance* dimana data saham yang dipilih adalah data saham yang terdaftar pada data saham LQ45. Saham yang terdaftar didaftar LQ45 adalah saham dengan kinerja yang baik. Akan diambil dua saham dari sektor perbankan yaitu saham Aneka Tambang Tbk (ANTM) dan saham dari sektor kebutuhan primer yaitu saham Indofood Sukses Makmur Tbk (ICBP). Untuk data saham yang diambil adalah data saham sepuluh tahun terakhir dari 1 Januari 2010 hingga 1 Januari 2020.

2.2. Autoregressive Integrated Moving Average (ARIMA)

Model ARIMA adalah model yang menggabungkan antara model *autoregressive* dan model *moving average*. Dimana I pada ARIMA bermakna *intergrated* yang bertujuan untuk membedakan *time series* yang tidak stasioner. Oleh karenanya ARIMA hanya dapat menerima data yang stasioner. Bentuk matematis dari model ARIMA adalah sebagai berikut.

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \delta t$$

2.3. Recurrent Neural Network (RNN)

Model ini juga dikenal sebagai model *sequential*, karena model ini memanfaatkan informasi *sequential* yang terdapat pada data. Menurut Moolayil (2019), model ini berasumsi bahwa semua data input memiliki hubungan satu sama lain seperti pada kata prediksi yang ada pada *smartphone* yang muncul berdasarkan kata sebelumnya. RNN memiliki perulangan pada arsitekturnya, pada waktu t . RNN akan menyimpan input x_t , kemudian akan memperbarui keadaan vektor h_t dari keadaan sebelumnya h_{t-1} . Jika terdapat input baru akan dihitung berdasarkan vektor h_t dan proses ini akan berulang yang akan disebut sebagai jejak urutan (*history of sequence*)

2.4. Long-Short Term memory (LSTM)

Model ini merupakan penembangan lebih lanjut dari model RNN. Model ini terdiri dari beberapa sel yaitu input sel, output sel, dan forget sel. Sel-sel ini bertugas untuk mengingat informasi sesuai interval waktu yang berubah-ubah. LSTM terdiri dari berbagai kumpulan sub-jaringan *recurrent* yang saling terhubung, yang selanjutnya disebut sebagai blok memori. Keunggulan dari LSTM dibandingkan dengan RNN adalah kemampuan LSTM untuk mengatasi permasalahan *vanishin gradient* dan *exploding gradient* yaitu permasalahan perubahan bobot (*weight*) yang terlalu kecil atau terlalu besar. Terdapat berbagai pengembangan lebih lanjut dari metode LSTM ini berikut adalah beberapa metode LSTM baru yang sering digunakan.

2.4.1. CNN-LSTM

Metode ini adalah gabungan dari metode Convolutional Neural Network (CNN) dan LSTM. pada metode ini terdapat bagian CNN yang berfungsi mengekstrak informasi yang terdapat pada data kemudian menyederhakan data input untuk kemudian diberikan kepada bagian LSTM yang bertujuan untuk memodelkan data dan memprediksi berdasarkan model yang dibangun.

2.4.2. ConvLSTM

Berbeda dengan CNN-LSTM model ini membangun sel convolutionalnya langsung kedalam unit LSTM itu sendiri, yang bertujuan untuk membaca *two-dimensional spatial-temporal* data.

3. Hasil dan Pembahasan

Sebelum memulai pemodelan dengan *neural network* data akan diubah dengan mentransformasikan data dengan fungsi *sliding windows*. Data *time series* akan direkonstruksi ulang oleh fungsi *sliding windows* dimana y_t akan diprediksi oleh y_{t-1} dan y_{t+1} akan diprediksi oleh y_t dan seterusnya.

Tabel 1. Contoh data *time serie*.

Time	Measure
1	100
2	200
3	150
4	300
5	400
...	...

Tabel 2. Contoh data *time series* hasil rekonstruksi *sliding windows*.

X	y
...	100
100	200
200	150
150	300
300	400
400	...

Dari tabel 2 terlihat bahwa y_t menjadi x_{t+1} dan y_{t+1} menjadi x_{t+2} dan seterusnya.

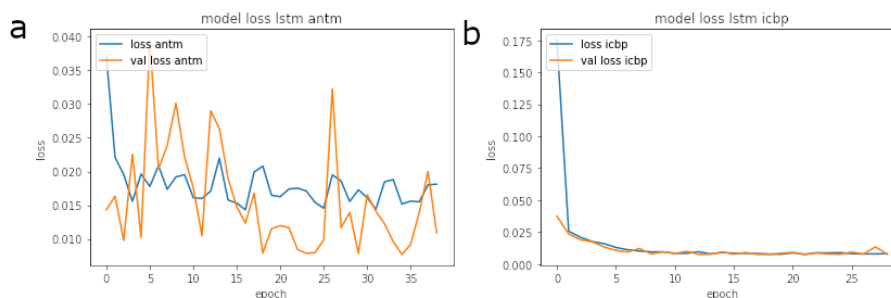
3.1. Long-shor term memory (LSTM)

Untuk proses pemodelan data saham ANTM dan ICBP dengan metode LSTM data harus direkonstruksi dengan *sliding windows function* dengan tujuan untuk mempermudah dalam proses pemodelan nantinya. Setelah proses tersebut data akan dinormalisasikan dan akan dipisahkan menjadi data latih dan data uji. Kemudian data akan direshape sesuai *input shape* yang dibutuhkan oleh metode LSTM. berikut adalah hasil pemodelan metode LSTM untuk data saham ANTM dan data saham ICBP.

Tabel 3. Hasil pemodelan data saham ANTM dan ICBP.

Nama saham/	ANTM	ICBP
Matriks evaluasi		
MAE	0.0079	0.0075
MAPE	0.0408	0.033
RMSE	0.013	0.0122
R2	0.9957	0.9981
Time	23.8	10.6

Selanjutnya untuk plotting model loss berdasarkan fungsi loss MAE dari hasil pemodelan metode LSTM diatas diperoleh plot sebagai berikut.



Gambar 1. (a) Plot model loss ANTMM (b) Plot model loss ICBP

Dari gambar 1 di atas terlihat bahwa model loss ANTMM sangat fluktuatif sedangkan untuk model loss ICBP terlihat terjadi penurunan loss semakin besar nilai epochnya. Selanjutnya hasil diatas akan disebut sebagai model *vanilla LSTM*, akan dilakukan pemodelan untuk metode CNN-LSTM dan Conv LSTM

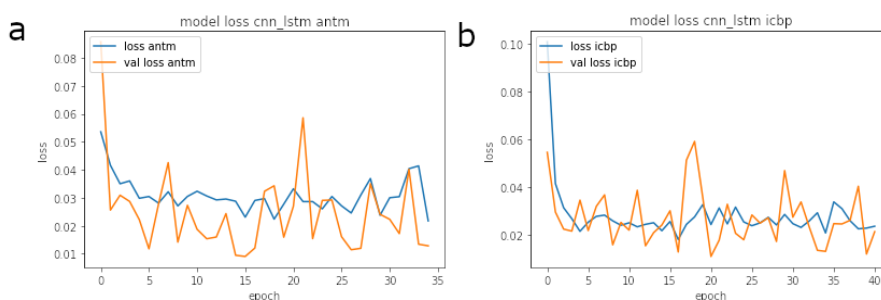
3.1.1. *CNN-LSTM*

Untuk melihat hasil pemodelan data saham ANTMM dan ICBP dengan metode CNN-LSTM akan disajikan dalam tabel berikut.

Tabel 4. Hasil pemodelan data saham ANTMM dan ICBP dengan CNN-LSTM

Nama saham/	ANTMM	ICBP
Matriks evaluasi		
MAE	0.0154	0.009
MAPE	0.0828	0.0574
RMSE	0.0216	0.0139
R2	0.989	0.9976
Time	24.8	23.4

Berikut adalah plot model loss metode CNN-LSTM.



Gambar 2. (a) Plot model loss ANTMM metode CNN-LSTM (b) Plot model loss ICBP metode CNN-LSTM

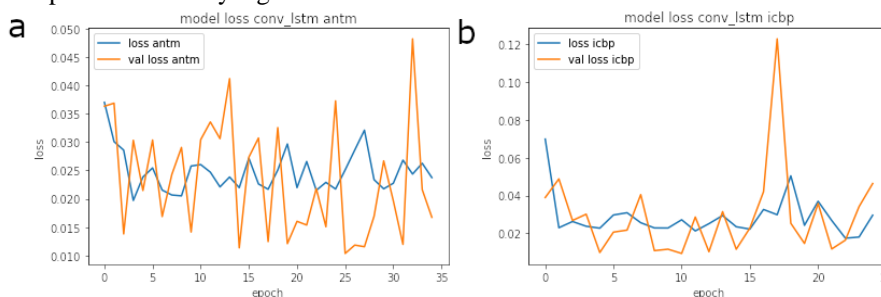
3.1.2. *ConvLSTM*

Selanjutnya adalah hasil pemodelan data saham ANTMM dan ICBP dengan metode ConvLSTM.

Tabel 5. Hasil pemodelan data saham ANTMM dan ICBP dengan metode ConvLSTM

Nama saham/ Matriks evaluasi	ANTM	ICBP
MAE	0.009	0.0093
MAPE	0.0442	0.0479
RMSE	0.0148	0.0134
R2	0.9946	0.9977
Time	24.8	22.1

Berikut adalah plot model loss yang dihasilkan oleh metode ConvLSTM.



Gambar 3. (a) Plot model loss ANTM metode ConvLSTM (b) Plot model loss ICBP metode ConvLSTM.

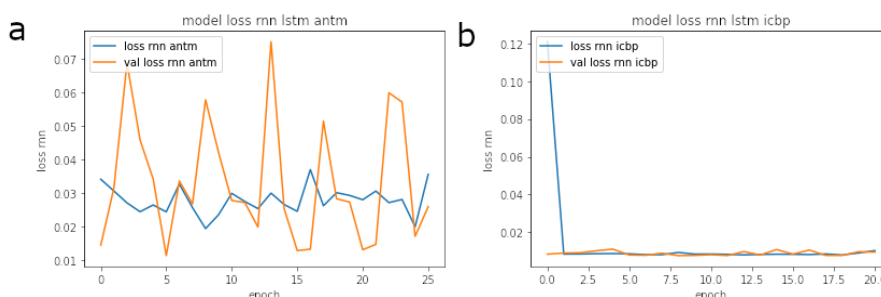
3.2. Recurrent Neural Network (RNN)

Proses pemodelan data saham ANTM dan data saham ICBP dengan metode RNN serupa dengan metode LSTM. rekonstruksi data dengan sliding windows, normalisasi data, split data, dan terakhir reshape data. Berikut adalah hasil pemodelan data saham ANTM dan data saham ICBP dengan metode RNN.

Tabel 6. Hasil pemodelan data saham ANTM dan ICBP dengan metode RNN

Nama saham/ Matriks evaluasi	ANTM	ICBP
MAE	0.0107	0.0071
MAPE	0.0601	0.0295
RMSE	0.0181	0.0112
R2	0.9919	0.9984
Time	19.8	10.6

Selanjutnya akan diplot model loss yang dihasilkan oleh metode RNN.



Gambar 4. (a) Plot model loss ANTM metode RNN (b) Plot model loss ICBP metode RNN

3.3. Perbandingan Metode Neural Network

Untuk melihat metode *neural network* terbaik antara LSTM dan RNN hasil dari kedua metode tersebut akan dibandingkan. Berikut adalah hasil perbandingan antara metode LSTM dan RNN.

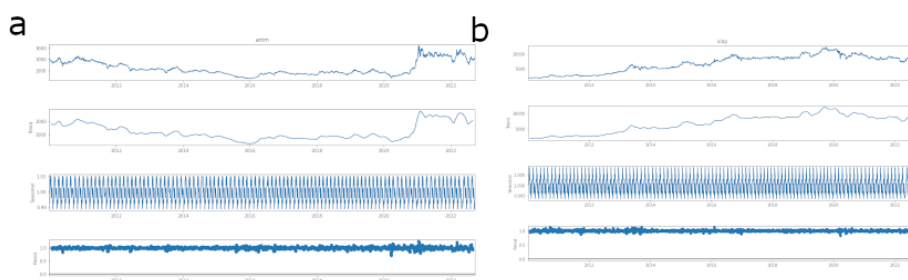
Tabel 7. Perbandingan Hasil Metode LSTM dan RNN.

Ket	ANTM				ICBP			
	LSTM	CNN-LSTM	ConvLSTM	RNN	LSTM	CNN-LSTM	ConvLSTM	RNN
MAE	0.0079	0.0154	0.009	0.0107	0.0075	0.009	0.0093	0.0071
MAPE	0.0408	0.0828	0.0442	0.0601	0.033	0.0574	0.0479	0.0295
RMSE	0.013	0.0216	0.0148	0.0181	0.0122	0.0139	0.0134	0.0112
R2	0.9957	0.989	0.9946	0.9919	0.9981	0.9976	0.9977	0.9984
Time	23.8	24.8	24.8	19.8	10.6	23.4	22.1	10.6

Dari tabel 7 diatas terlihat bahwa performa model LSTM lebih unggul dari metode CNN-LSTM, ConvLSTM dan RNN pada data ANTM. Sedangkan metode RNN unggul dari metode LSTM, CNN-LSTM, dan ConvLSTM pada data ICBP. Walaupun pada data ICBP perbedaan akurasi antara LSTM dan RNN sangat kecil, sedangkan pada data ANTM terjadi perbedaan yang cukup signifikan antara metode LSTM dan RNN untuk semua metrik evaluasi yang digunakan. Keunggulan RNN disini adalah waktu latih tercepat dari semua metode yang digunakan. Dari sini didapatkan dua metode terbaik yaitu LSTM dan RNN, yang selanjutnya akan dibandingkan dengan metode tradisional ARIMA.

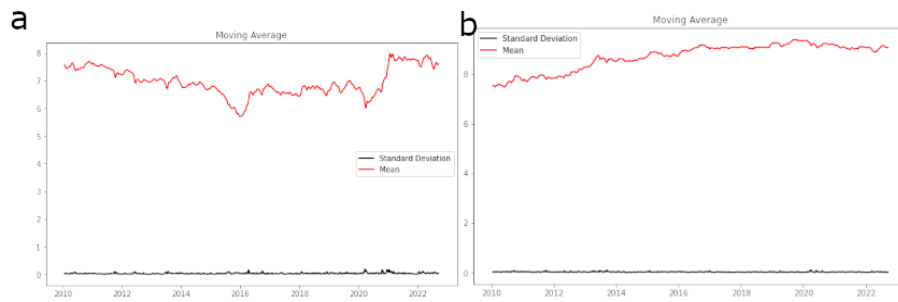
3.4. Autoregressive Integrated Moving Average (ARIMA)

Pada metode ARIMA ini terdapat perbedaan dalam memodelkan data, dimana data terlebih dahulu diuji stasionaritasnya karena ARIMA tidak dapat memodelkan data yang non-stasioner. Jika uji stationeritas menunjukan data tidak stationer maka langkah selanjutnya adalah memisahkan trend dan seasonalitas dari data. Kemudian memisahkan data menjadi data latih dan data uji, kemudian masuk dalam tahapan tuning parameter dengan *auto arima* yang akan menghasilkan arima dengan order terbaik untuk data yang digunakan.



Gambar 5. (a) Seasonal decompose untuk data saham ANTM (b) seasonal decompose untuk data saham ICBP.

Tahap berikutnya adalah pemisahan trend yang dimiliki oleh data saham ANTM dan ICBP.



Gambar 6. (a) Menghilangkan trend data ANTAM (b) menghilangkan trend pada data ICBP.

Setelah data berhasil bersih dari seasonalitas dan trend data akan dipisah menjadi data latih dan data uji. Kemudian data tersebut akan dituning menggunakan *auto arima* yang akan memberikan model arima terbaik untuk data ANTAM dan data ICBP, hasil *auto arima* yang didapatkan untuk data ANTAM adalah ARIMA(0,1,0) dan untuk data ICBP adalah ARIMA(0,1,0). Langkah selanjutnya adalah pemodelan dengan metode ARIMA, berikut adalah hasil metrik evaluasi yang dihasilkan dengan metode ARIMA.

Tabel 8. Hasil pemodelan data saham ANTAM dan ICBP dengan metode RNN

Nama saham/	ANTM	ICBP
Matriks evaluasi		
MAE	669.01	4488.77
MAPE	0.7	0.9
RMSE	906.99	4625.71

3.5. Perbandingan Hasil Metode LSTM, RNN, dan ARIMA.

Pada proses ini akan diberikan tabel yang akan membandingkan hasil pemodelan LSTM, RNN, dan ARIMA. Dari tabel tersebut dapat dibandingkan hasil performa model *neural network* dan model tradisional ARIMA.

Tabel 9. Perbandingan hasil pemodelan data saham ANTAM dan ICBP untuk metode LSTM, RNN, dan ARIMA.

Ket	ANTM			ICBP		
	LSTM	RNN	ARIMA	LSTM	RNN	ARIMA
MAE	0.0079	0.0107	669.01	0.0075	0.0071	4488.77
MAPE	0.0408	0.0601	0.7	0.033	0.0295	0.9
RMSE	0.013	0.0181	906.99	0.0122	0.0112	4625.71

4. Simpulan

Berdasarkan hasil dari tabel 9 diatas terlihat bahwa metode *neural network* unggul dari metode tradisional ARIMA. Terlihat bahwa tingkat akurasi dari model LSTM dan RNN sangat tinggi dibandingkan dengan metode ARIMA. Metode *neural network* juga lebih cepat dalam proses latih dibandingkan dengan metode ARIMA yang memerlukan tuning untuk menemukan order model terbaik untuk data yang digunakan. Untuk metode *neural network* sendiri dapat dilihat bahwa model LSTM unggul dalam tingkat akurasi. Sedangkan RNN unggul dalam proses latih yang memakan waktu paling cepat dibandingkan dengan LSTM. Untuk penelitian selanjutnya dapat menggunakan data saham yang lebih kompleks lagi seperti memodelkan data multivariate dengan metode LSTM dan RNN.

Daftar Pustaka

- Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A Systematic review. *Expert Systems with Applications*, 156. <https://doi.org/10.1016/j.eswa.2020.113464>
- Fama, E. F. (1995). Random Walks in Stock Market Prices. *Financial Analysts Journal*, 51(1), 75–80. <https://doi.org/10.2469/faj.v51.n1.1861>
- Hu, Z., Zhao, Y., & Khushi, M. (2021). A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 4(1), 1–30. <https://doi.org/10.3390/ASI4010009>
- Moolayil, J. (2019). Learn Keras for Deep Neural Networks. In *Learn Keras for Deep Neural Networks*. <https://doi.org/10.1007/978-1-4842-4240-7>
- Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W., & Wills, G. (2020). A survey on machine learning for stock price prediction: Algorithms and techniques. *FEMIB 2020 - Proceedings of the 2nd International Conference on Finance, Economics, Management and IT Business, Femib*, 63–71. <https://doi.org/10.5220/0009340700630071>
-