

Responsi Analitika Data

Editha Anggrieniputri Kinbenu

19/445626/PA/19450

Link Kaggle : <https://www.kaggle.com/edithakinbenu/445626-editha-anggrieniputri-kinbenu-reponsi-andat/edit>

1. Permodelan yang kelompok saya lakukan adalah forecasting Penambahan Kasus Covid-19 Harian di Indonesia dengan Metode LSTM. Data yang digunakan di penelitian kami adalah data penambahan kasus harian Covid-19 di Indonesia yang diambil secara online melalui laman [worldometer](https://www.worldometer.com/) dan dimulai dari 2 Maret 2020 hingga 11 November 2021. Pada pelatihan dataset COVID19 dengan model LSTM kami digunakan nilai epoch sebanyak 25, ukuran batch sebesar 64, dan jumlah neuron sebesar 200. Dimana hasil forecasting peramalan yang dihasilkan oleh model LSTM pada data uji menghasilkan pola data yang cukup mirip dengan data uji asli. Sehingga, secara eksploratif dapat dikatakan bahwa model LSTM ini menghasilkan performa yang baik pada data uji

2. Penilaian teman sejawat

Kelompok : 3

No urut	Nama Lengkap	Nilai	Jobdesk dan Komentar
8	BENEDICTA YESIKA M	92	Aktif, Membantu mensortir data yang digunakan, dan mendapat landasan teori pada laprak.
17	HUSNA RAHMA YUNITA	95	Aktif, Bersama saya menemukan program dan syntax untuk Laprak UAS, dan mendapat bagian membuat pemilihan epoch pada Laprak
19	JAMES JULIUS	90	Aktif, Membantu membuat abstrak dan kesimpulan
34	RAFIDA ALAIYYA HAYYIN	93	Aktif, Membantu menyusun semua hasil laprak, mendapat bagian evaluasi model
36	SHOFIAYUMNA F P	94	Aktif, Inisiator, dan membantu menyusun laprak, membuat introduction

3. Permodelan *deep learning* mata uang kripto

Dataset

	Date	Open	High	Low	Close	Adj Close	Volume
0	21/11/2020	509.744598	550.227417	504.237762	549.486633	549.486633	20757099184
1	22/11/2020	549.486633	579.372498	514.517212	558.068115	558.068115	21967049601
2	23/11/2020	558.059509	609.987610	551.265259	608.454041	608.454041	27272302872
3	24/11/2020	608.522766	621.173401	593.835144	603.897766	603.897766	23281758100
4	25/11/2020	603.902039	605.094177	559.671387	570.686646	570.686646	20088492893

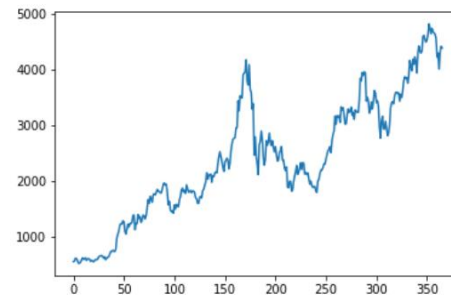
	Date	Open	High	Low	Close	Adj Close	Volume
361	17/11/2021	4213.910645	4300.306641	4107.125977	4287.593750	4287.593750	22183461850
362	18/11/2021	4287.802246	4343.557129	3959.227783	4000.650879	4000.650879	21383250893
363	19/11/2021	3995.729492	4311.713867	3982.917480	4298.306641	4298.306641	20626269711
364	20/11/2021	4298.350586	4434.388184	4209.242676	4409.931152	4409.931152	15448261277
365	21/11/2021	4411.071289	4414.897461	4308.393066	4380.979980	4380.979980	14316200960

Interpretasi:

Data yang digunakan di penelitian ini data nilai mata uang kripto Ethereum dalam US Dollar dari tanggal 21 November 2020 sampai dengan 21 November 2021. Bentuk data yang digunakan adalah data *time series*. Data yang terkumpul adalah sebanyak 365. Setiap data mengandung 7 atribut, yaitu *Date*, *Open*, *High*, *Low*, *Close*, *Adj Close* dan *Volume*. Kita akan memodelkan harga penutupan maka kita harus menyisahkan variabel *Date* dengan *Close*

```
0    549.486633
1    558.068115
2    608.454041
3    603.897766
4    570.686646
Name: Close, dtype: float64
```

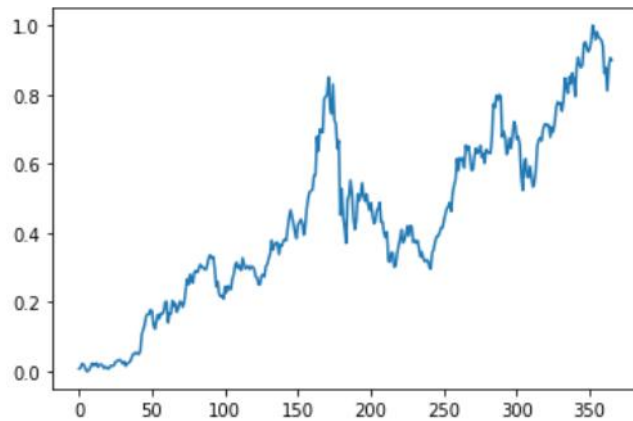
```
361    4287.593750
362    4000.650879
363    4298.306641
364    4409.931152
365    4380.979980
Name: Close, dtype: float64
```



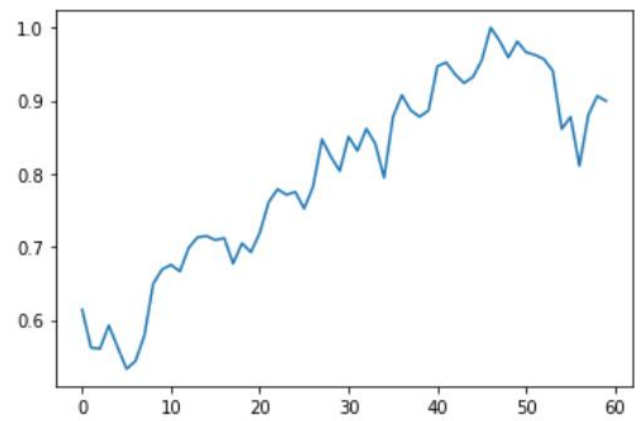
GAMBAR Plot dari *Date* dan *Close*

Data yang sudah dikumpulkan dipilah menjadi dua bagian, yaitu data training dan data testing. Pada soal sudah disebutkan bahwa Data training terdiri dari observasi dari tanggal 21 November 2020 sampai dengan 22 September 2021, sedangkan data testing terdiri dari observasi dari tanggal 23 September 2021 sampai dengan 21 November 2021. Sedangkan untuk grafiknya dapat dilihat seperti berikut.

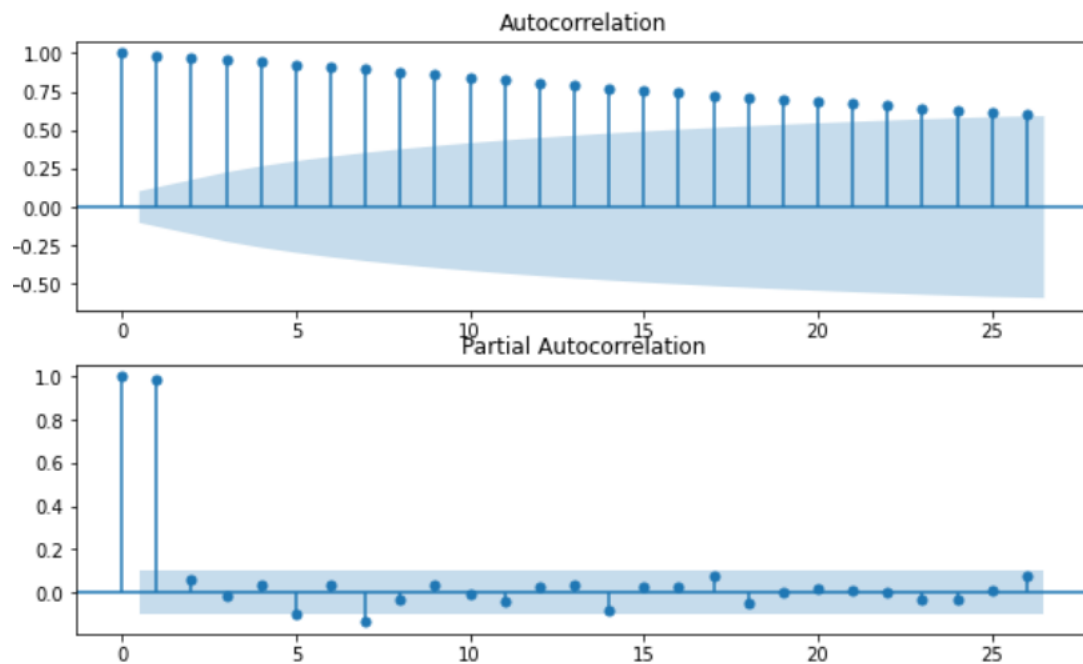
Data Training	Data Testing
306	60



Plot data *training*



Plot data *testing*



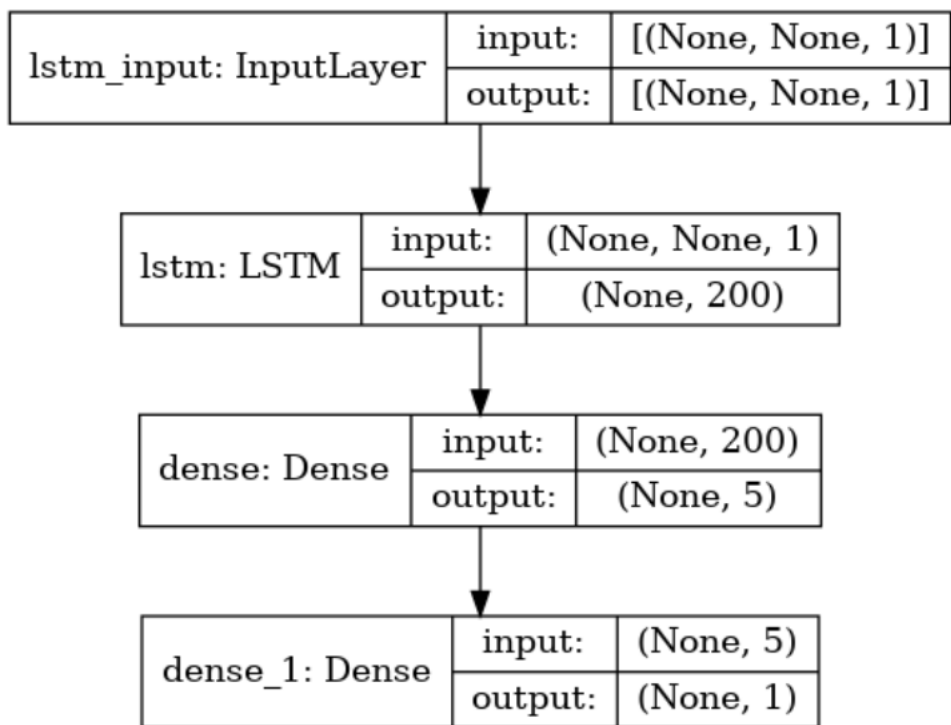
Correlogram data penutupan nilai mata uang kripto Ethereum dalam US

Pada kasus ini akan diperhatikan 5 lag pertama pada PACF. Berdasarkan correlogram di atas, diperoleh nilai lag ke-1 merupakan lag yang terakhir keluar sehingga input adalah data dari period ke $t-5$ sampai $t-1$, dan output period ke- t .

Berdasarkan correlogram, kemudian dibentuk data dengan dimensi yang sesuai dengan input dari LSTM, dengan jumlah data yang digunakan sebanyak 1, yaitu data Close, jumlah neuron 200, dan dense lag ke 5 dengan fungsi aktivasi ReLU. Diperoleh model LSTM dibentuk dengan arsitektur sebagai berikut :

Model: "sequential"

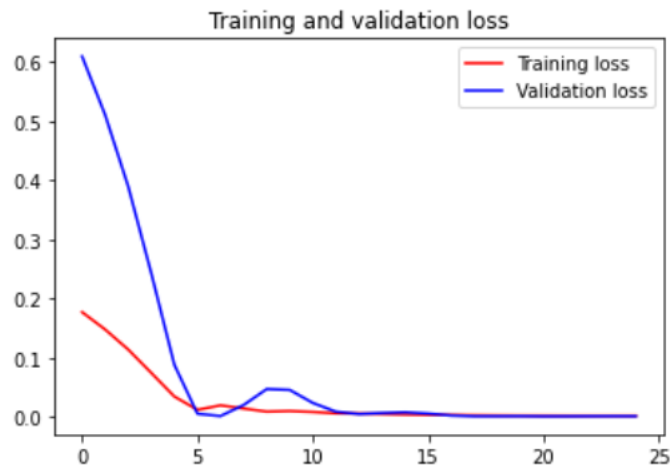
Layer (type)	Output Shape	Param #
=====		
lstm (LSTM)	(None, 200)	161600
dense (Dense)	(None, 5)	1005
dense_1 (Dense)	(None, 1)	6
=====		
Total params: 162,611		
Trainable params: 162,611		
Non-trainable params: 0		



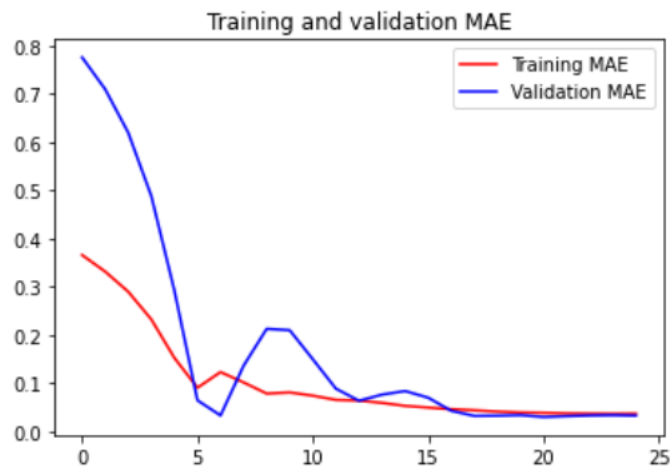
Arsitektur Model LSTM

Evaluasi Model

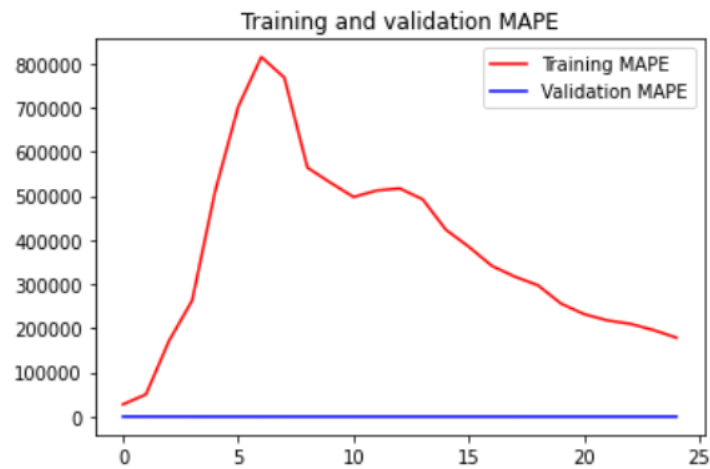
Pada pelatihan dataset COVID19 dengan model LSTM ini akan digunakan nilai epoch sebanyak 25, ukuran batch sebesar 64, dan jumlah neuron sebesar 200. Untuk plot MAPE, MAE, dan loss pada data training dan data validation, dapat dilihat pada grafik berikut.



Berdasarkan plot diatas, dapat dilihat bahwa nilai dari *training* dan *validation* loss menurun, dan menghasilkan nilai loss yang cukup kecil serta mendekati 0, artinya model yang dihasilkan oleh model LSTM pada data uji menghasilkan model yang baik.



Berdasarkan plot diatas, dapat dilihat bahwa nilai dari *training* dan *validation* MAE menurun, dan menghasilkan nilai MAE yang cukup kecil serta mendekati 0, artinya model yang dihasilkan oleh model LSTM pada data uji menghasilkan model yang baik.



Berdasarkan plot diatas, dapat dilihat bahwa nilai dari *training* dan *validation* MAPE statis dan turun naik, dan menghasilkan nilai MAPE untuk validasi statis di 0 dan untuk training sempat naik kemudian turun, artinya model yang dihasilkan oleh model LSTM pada data uji menghasilkan model yang kurang baik. Kemudian didapatkan nilai

Loss	0.001878758193925023
MAE	0.0344504788517952
MAPE	4.46992826461792
RMSE	0.043344646457500226

Dapat dilihat nilai loss dan RMSE sudah cukup kecil, sehingga dapat simpulkan bahwa model ini sudah cukup baik.

Prediksi



Hari	Data Uji	Data Prediksi
306	3155	3042
307	2932	2923
308	2926	2950
....
364	4298	4731
365	4410	4821
366	4381	4955

Interpretasi:

Berdasarkan pengujian data didapatkan hasil terbaik dengan model LSTM menggunakan jumlah epoch sebanyak 25, batch size sebanyak 64, dan jumlah neuron sebanyak 200. Plot diatas menunjukan bahwa hasil peramalan yang dihasilkan oleh model LSTM pada data uji menghasilkan pola data yang cukup mirip dengan data uji asli. Sehingga, secara eksploratif dapat dikatakan bahwa model LSTM ini menghasilkan performa yang baik pada data uji.