Amazon Stock Price

Time-series forecasting

Presented by:

kelompok 9 DBA-A9

Ruangguru CAMP 2022

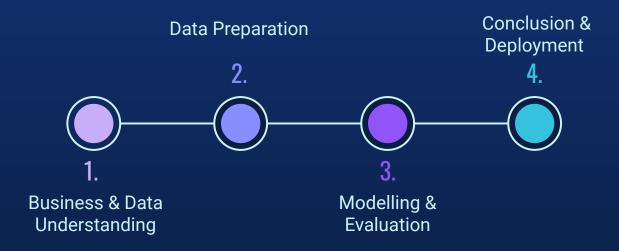


Presented by:

- Richardo Zefanya Damarjanan (DBA2215795)
- Silvia Nurlaili Agustina (DBA2151534)
- Muhammad Cakrawala Persada (DBA2262545)
- Fanny Maulidina Rizki (DBA2073703)
- Muhammad Raushan Fikri (DBA2145318)



Outline:





01
Business &
Data Understanding



Business Understanding



1. What's Amazon?

Amazon.com, Inc. adalah multinasional teknologi yang berfokus pada e-niaga, komputasi, streaming digital dan kecerdasan buatan. Amazon terlibat dalam penjualan eceran produk konsumen dan langganan di Amerika Utara serta internasional.



2. Amazon Stock Record

Di tahun 1997 Amazon telah menjadi Go Public dimana sahamnya bisa diperdagangkan oleh investor publik.

IPO Amazon diterbitkan 15 Mei 1997, dengan harga \$18 per lembar saham.

Yang unik dari Amazon di mata investor adalah sahamnya sudah naik lebih dari **1000 kali** hanya dalam waktu 20 tahun.



3.Objective

Saham merupakan surat berharga yang dikeluarkan oleh perusahaan dan sebagai tanda kepemilikan suatu perusahaan.

Semakin tinggi permintaan maka harga saham akan meningkat dan semakin tinggi penawaran maka harga saham akan menurun. Pergerakan harga saham Amazon dapat dikatakan cukup fluktuatif.

Dengan Dataset Amazon, kami akan **meprediksi Harga Penutupan Saham Amazon kurang lebih dalam 1 bulan.** Hal ini dilakukan untuk mengetahui kondisi dan mempersiapkan strategi untuk menghadapi penurunan atau pelonjakan harga saham pada saat close price.

Data Understanding

Source

https://finance.yahoo.com/quote/AMZN/history?p=AMZN

- Dataset ini terdiri dari (kolom) Date, rata-rata harga High, Low, Open, Close dan Volume serta Adj Close dalam kurun waktu selama 5 tahun lebih mulai dari tahun 2016 hingga 2021.
- Deskripsi dari masing-masing coloumns dapat dilihat pada tabel disamping.

Data explanation

Variable(Key Data Points)	Description
Data Start Date	2016-04-26
Data End Date	2021-04-26
Time Period	6 Years
Key Fields	Open,High,Low,Close,Volume,Adj Close

Variable(Key Data Points)	Description	
Date	as day-month-date	
Open	as-traded opening price for the day	
High	as-traded high price for the day	
Low	as-traded low price for the day	
Close	as-traded close price for the day	
Volume	trading volume for the day	
Adj Close	split & dividend adjusted closing price of the day	

Data type dan Null values

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1258 entries, 0 to 1257
Data columns (total 7 columns):
              Non-Null Count Dtype
    Column
    Date
           1258 non-null
                             object
           1258 non-null float64
    Open
              1258 non-null
    High
                             float64
    Low
              1258 non-null float64
4 Close
              1258 non-null float64
   Adi Close 1258 non-null float64
    Volume
              1258 non-null
                             int64
dtypes: float64(5), int64(1), object(1)
memory usage: 68.9+ KB
```

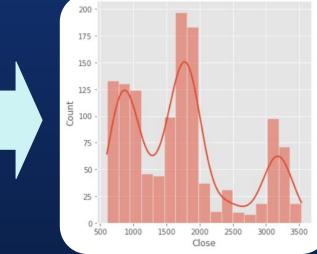
- Data ini memiliki jumlah baris
 sebanyak 1258 dengan data kolom
 sebanyak 7
- Dapat dilihat bahwa dataset ini tidak
 memiliki Missing Value

Statistic Summary:

	0pen	High	Low	Close	Adj Close	Volume
count	1258.000000	1258.000000	1258.000000	1258.000000	1258.000000	1.258000e+03
mean	1719.277043	1737.104458	1698.973275	1718.622010	1718.622010	4.316484e+06
std	799.483202	810.242682	787.075320	798.374373	798.374373	2.091110e+06
min	611.799988	615.950012	599.200012	602.000000	602.000000	8.813000e+05
25%	978.662491	984.500000	969.157517	978.197495	978.197495	2.913225e+06
50%	1698.900024	1714.195007	1678.570007	1695.960022	1695.960022	3.708350e+06
75%	1950.875000	1977.477509	1933.629975	1957.605042	1957.605042	5.123225e+06
max	3547.000000	3552.250000	3486.689941	3531.449951	3531.449951	1.656500e+07

Rata-rata harga penutupan saham (closing price) setiap harinya yaitu sebesar US\$ 1.718

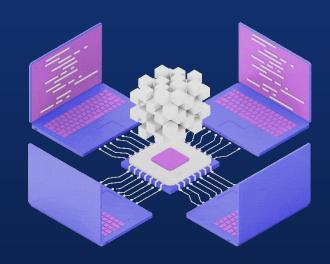
Distribusi harga penutupan saham selama periode 2016 - 2021





02

Data Preparation



```
df_close = df[["Date", "Close"]].copy()
df_close["Date"] = pd.to_datetime(df_close["Date"])
df_close.set_index("Date", inplace = True)
df_close = df_close.asfreq("b")
df_close = df_close.fillna(method = "bfill")
df_close.head(5)
```

f.	head(5) #ta	mpilkan 5 d	ata pertama				
	Date	Open	High	Low	Close	Adj Close	Volume
0	2016-04-26	626.169983	626.750000	614.880005	616.880005	616.880005	2521400
1	2016-04-27	611.799988	615.950012	601.280029	606.570007	606.570007	4068800
2	2016-04-28	615.539978	626.799988	599.200012	602.000000	602.000000	7872600
3	2016-04-29	666.000000	669.979980	654.000000	659.590027	659.590027	10310700
4	2016-05-02	663.919983	685.500000	662.030029	683.849976	683.849976	6578500

	Close
Date	
2016-04-26	616.880005
2016-04-27	606.570007
2016-04-28	602.000000
2016-04-29	659.590027
2016-05-02	683.849976

- Menampilkan 5 Data pertama kemudian
- Mensetting Kolom 'Date' sebagai index dan membuat variabel baru yang menyimpan kolom 'Date' dan 'Close' yang akan kita olah nanti

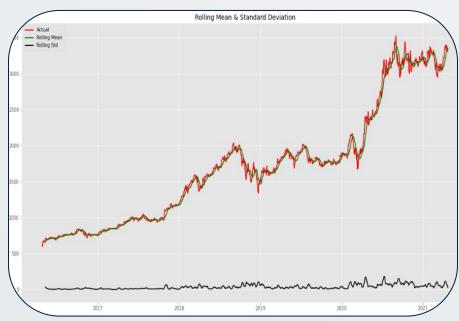
Stationarity Check



Dari grafik disamping, kita melihat bahwa pola penutupan saham (closing price) memiliki tren



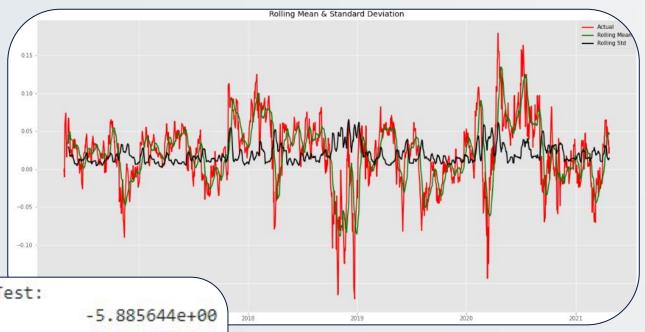
Pada grafik di samping, kita melihat bahwa data actual dan rolling (mean dan std) cukup berbeda, menandakan bahwa series ini tidak stasioner.



Dickey-Fuller Test:	
Test Statistic	0.202532
p-value	0.972403
Lags Used	19.000000
No. of Obs	1284.000000
Critical Value (1%)	-3.435453
Critical Value (5%)	-2.863794
Critical Value (10%)	-2.567970
dtype: float64	

ADF test:

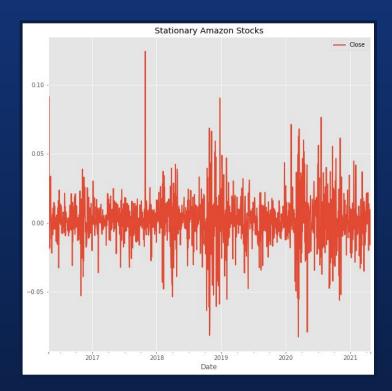
Nilai ADF statistic yang jauh lebih besar dari critical values pada semua level, dan juga p-value yang lebih besar dari 0.05 turut menandakan, kita gagal menolak null hypothesis pada 90%, 95%, and 99% confidence, yang berarti bahwa data time series ini strongly non-stationary.



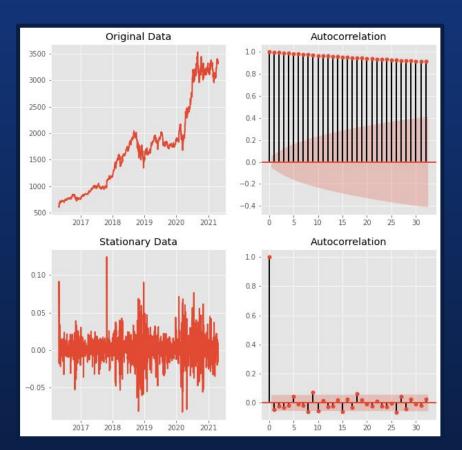
Dickey-Fuller Test:
Test Statistic -5.885644e+00
p-value 3.006160e-07
Lags Used 1.000000e+01
No. of Obs 1.293000e+03
Critical Value (1%) -3.435418e+00
Critical Value (5%) -2.863778e+00
Critical Value (10%) -2.567961e+00
dtype: float64

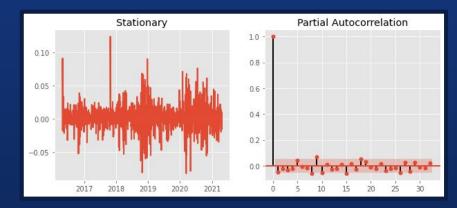
Diambil kesimpulan bahwa nilai t-value (3.006160e-07) < DF critical value 1% (-3.435418) maka kita menolak null hypothesis, dengan 99% tingkat keyakinan menyatakan bahwa **series ini sudah stasioner.**

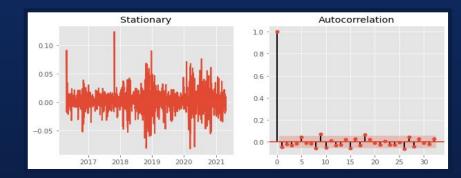
Grafik Stasioner Harga Penutupan Saham Amazon











03
Modeling & Evaluation



MODELING

1. ARMA MODEL

from statsmodels.tsa.arima_model import ARMA
model = ARMA(df['Close'], order=(0,1)).fit()
df['ARMA'] = model.predict()







```
modelValid1 = ARMA(train,order=(0,1))
fitModelValid1 = modelValid1.fit(disp= -1)

fc1,se1,conf1 = fitModelValid1.forecast(len(df_close) - length)
forecast1 = pd.Series(fc1, index = test.index)
```



- Dapat dilihat pada gambar Forecasting menggunakan data testing pada model ini dapat dikatakan berbeda jauh dengan Data Actual
- 90% dari data digunakan sebagai data train yaitu 1173 rows dan sisanya yaitu sebagai data test yaitu 131 rows



```
#Predicting
modelPred1 = ARMA(df_close,order=(0,1))
fitModelPred1 = modelPred1.fit(disp= -1)

fitModelPred1.plot_predict(1,len(df_close) + 1000)
x = fitModelPred1.forecast(1000)
x = plt.title("Amazon Stock Forecast")
x = plt.xlabel("Year")
x = plt.ylabel("Close Price")
plt.grid(False)
plt.show()
```

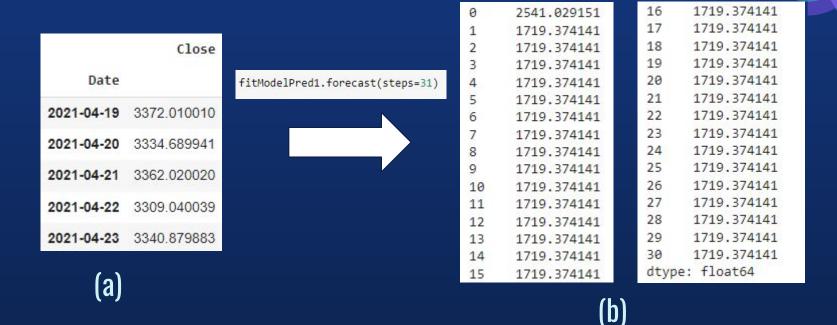




Model ARMA - MAE: 323.26904088310675 | RMSE: 413.5233160445343

- Pada gambar diatas kami melakukan Forecast untuk 1000 hari kedepan dan juga menampilkan MAE dan RMSE dan model tersebut.





 Gambar (a) merupakan 5 data terakhir pada tabel 'CLOSE' dan gambar (b) merupakan hasil forecasting untuk 1 bulan kedepan dimulai dari tanggal '24-04-2021' sampai '24-05-2021'.

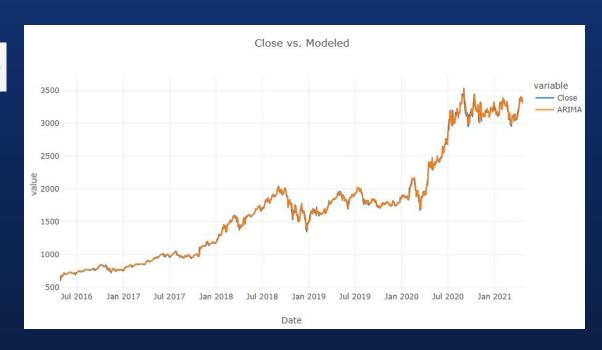
MODELING

2. ARIMA MODEL

from statsmodels.tsa.arima_model import ARIMA
model = ARIMA(df['Close'], order=(5,1,5)).fit()
df['ARIMA'] = model.predict()



Visualisasi





```
modelValid2 = ARIMA(train,order=(5,1,5))
fitModelValid2 = modelValid2.fit(disp= -1)

fc2,se2,conf1 = fitModelValid2.forecast(len(df_close) - length)
forecast2 = pd.Series(fc2, index = test.index)
```



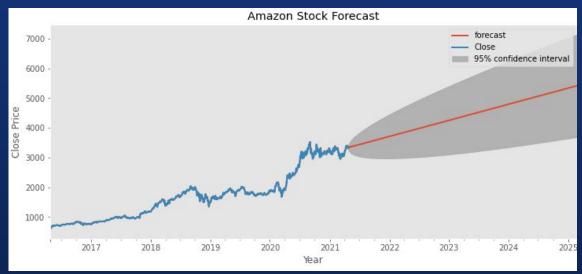
Dapat dilihat pada gambar Forecasting menggunakan data testing pada model ini dapat dikatakan cukup baik karena cukup sesuai dengan Data Actual



```
#Predicting
modelPred2 = ARIMA(df_close,order=(5,1,5))
fitModelPred2 = modelPred2.fit(disp= -1)

fitModelPred2.plot_predict(1,len(df_close) + 1000)
x = fitModelPred2.forecast(1000)
x = plt.title("Amazon Stock Forecast")
x = plt.xlabel("Year")
x = plt.ylabel("Close Price")
plt.grid(False)
plt.show()
```





Model ARIMA - MAE: 23.14604697077924 | RMSE: 36.52419851517407

- Pada gambar diatas kami melakukan Forecast untuk 1000 hari kedepan dan juga menampilkan MAE dan RMSE dan model tersebut.

	Close
Date	
2021-04-19	3372.010010
2021-04-20	3334.689941
2021-04-21	3362.020020
2021-04-22	3309.040039
2021-04-23	3340.879883



(a)

0	3346.707189	16 3373.264841
1	3350.011843	17 3373.228030
2	3344.448759	18 3378.688597
3	3355.827473	19 3376.090835
4	3348.955986	20 3381.354556
5	3355.224253	21 3381.973707
6	3356.941000	22 3383.234068
7	3353.452695	23 3388.058938
8	3361.968530	24 3386.893405
9	3356.460283	25 3391.834742
10	3362.530988	26 3392.365243
11	3363.791502	27 3393.964060
12	3362.828101	28 3397.594475
13	3370.209116	29 3396.952591
14	3366.725259	30 3401.242595
15	3372.606731	dtype: float64

(b)

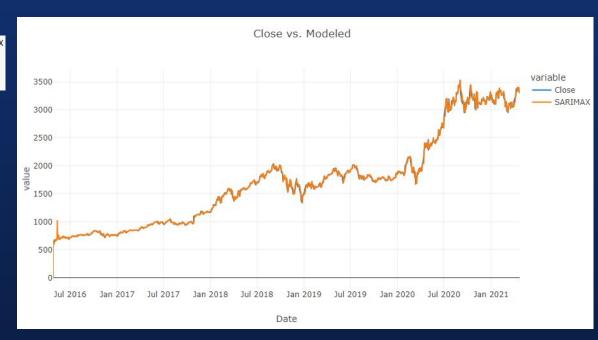
 Gambar (a) merupakan 5 data terakhir pada tabel 'CLOSE' dan gambar (b) merupakan hasil forecasting untuk 30 hari kedepan dimulai dari tanggal '24-04-2021' sampai '24-05-2021'.

MODELING

3. SARIMAX MODEL



Visualisasi









Model SARIMAX - MAE: 24.5874892494192 | RMSE: 42.32467354560117

 Pada gambar diatas kami melakukan Forecast untuk 150 hari kedepan dan juga menampilkan MAE dan RMSE dan model tersebut.



	Close
Date	
2021-04-19	3372.010010
2021-04-20	3334.689941
2021-04-21	3362.020020
2021-04-22	3309.040039
2021-04-23	3340.879883

results.get_forecast(steps=31)



(a)

Gambar (a) merupakan 5 data terakhir pada tabel 'CLOSE' dan gambar (b) merupakan hasil forecasting untuk 31 hari kedepan dimulai dari tanggal '24-04-2021' sampai '24-05-2021'.

2021-04-26	3464.217872
2021-04-27	3485.252809
2021-04-28	3510.628150
2021-04-29	3520.409070
2021-04-30	3575.621627
2021-05-03	3581.207919
2021-05-04	3566.883129
2021-05-05	3589.486430
2021-05-06	3626.209083
2021-05-07	3605.582151
2021-05-10	3622.221064
2021-05-11	3641.933439
2021-05-12	3681.422515
2021-05-13	3693.361523
2021-05-14	3704.598125

2021-05-17	3708.121959
2021-05-18	3736.302293
2021-05-19	3743.469718
2021-05-20	3730.545367
2021-05-21	3740.848410
2021-05-24	3761.698392
2021-05-25	3746.206705
2021-05-26	3748.441034
2021-05-27	3767.835619
2021-05-28	3807.316071
2021-05-31	3816.360792
2021-06-01	3826.888464
2021-06-02	3829.019475
2021-06-03	3859.724186
2021-06-04	3864.870569
2021-06-07	3850.177648

(b)

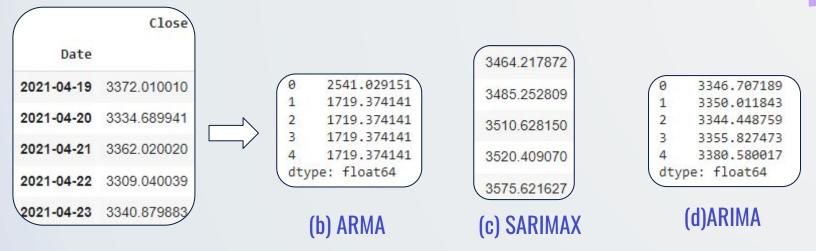
EVALUATION

```
models = ['ARMA', 'ARIMA', 'SARIMAX']

for model in models:
    diff = df['Close'] - df[model]
    mae = diff.abs().mean()
    rmse = np.sqrt(np.mean(diff**2))
    print(f'Model {model} - MAE: {mae} | RMSE: {rmse}')
```

```
Model ARMA - MAE: 323.26904088310675 | RMSE: 413.5233160445343
Model ARIMA - MAE: 23.14604697077924 | RMSE: 36.52419851517407
Model SARIMAX - MAE: 24.5874892494192 | RMSE: 42.32467354560117
```

Dari Hasil diatas dapat dilihat bahwa Nilai Mean Absolute Error (MAE) dan Root Mean Square Error (RMSE) Dari Model ARIMA **memiliki nilai error yang lebih kecil**. Oleh karena itu model ini yang akan dipakai karena jarak antara nilai aktual dengan nilai prediksi dapat dikatakan cukup dekat dibandingkan Model yang lain



(a) 5 Data Terakhir

- Jika diamati baik-baik terlihat perbedaan hasil forecasting yang sangat jauh antara ARMA dan kedua Model lainnya.
- Hasil Forecasting pada ARMA dihari kedua dan setelahnya nilainya telihat sama seperti pada grafik sebelumnya membentuk garis yang sejajar dengan sumbu X
- Hasil pada Model SARIMAX dan ARIMA terbilang lebih baik karena angkanya tidak terlalu jauh dan masih sesuai dengan pola trend



Berdasarkan diagram diatas dapat disimpulkan bahwa:

- Prediksi menggunakan modell ARIMA dan SARIMAX sangat baik karena mendekati nilai aktual (CLOSE)
- Sedangkan pada ARMA kurang baik karena cenderung menjauhi nilai aktual

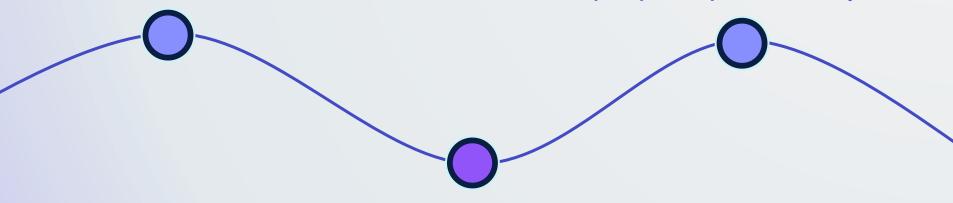
O4
Conclusion & Deployment



Deployment

Berdasarkan hasil modeling, jenis model yang digunakan dalam analisa time series-forecasting kali ini yaitu model ARIMA.

Melalui prediksi harga penutupan harga (closing price) saham amazon ini dapat membantu para pelaku saham untuk mempresiksi harga saham pada periode-periode berikutnya.



Dari model tersebut kami memprediksi bahwa rata-rata harga closing price saham amazon dalam 1 bulan kedepan yaitu sebesar **US\$ 3.371**

Average History in 5 years

US\$ 1718



Average Forecasting 31 Days

US\$ 3371

*Harga closing price (rata-rata) saham amazon selama periode 2016 - 2021

*Hasil forecasting closing price saham amazon selama 1 bulan kedepan dari tanggal '24-04-2021' sampai '24-05-2021'. (ARIMA model)



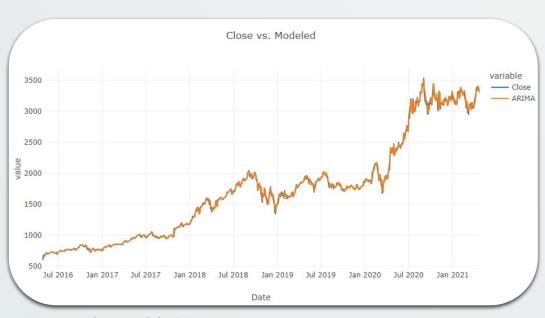
Forecast Overview

0	3346.707189	16	373.264841
1	3350.011843	17	373.228030
2	3344.448759	18	378.688597
3	3355.827473	19 3	376.090835
4	3348.955986	20 3	381.354556
5	3355.224253	21 3	3381.973707
6	3356.941000	22	383.234068
7	3353.452695	23 3	388.058938
8	3361.968530	24 3	386.893405
9	3356.460283	25 3	3391.834742
10	3362.530988	26	392.365243
11	3363.791502	27	393.964060
12	3362.828101	28	397.594475
13	3370.209116	29 3	396.952591
14	3366.725259	30 3	401.242595
15	3372.606731	dtype:	float64

^{*}Forecast 31 hari

US\$ 3371

Avg Forecasting 31 Days



*Menggunakan model ARIMA

46

The goal of forecasting is not to predict the future but to tell you what you need to know to take meaningful action in the present

7

-Paul Saffo



