

# ANALISIS DAN HASIL PREDIKSI EMAS 1

Hendrico Kristiawan - 1906350912  
Jonathan Amadeus Hartman - 1906400261



SCRAPING

# Scraping

- Terdapat 2 data yang dapat di scraping, yaitu tabel dan grafik.
- Data pada tabel akan dinamakan data\_table yang diambil dari website static sehingga cukup dengan menggunakan **Beautiful Soup**.
- Data pada grafik akan dinamakan data\_plot yang diambil dari website dinamis sehingga perlu menggunakan **selenium** dan **pyautogui**.
- Dilakukan scraping pada notebook berbeda yang diupload ke dalam Github beserta data hasil scraping.

([https://raw.githubusercontent.com/loriU/macrotrends-gold-scrap/main/emas\\_data.csv](https://raw.githubusercontent.com/loriU/macrotrends-gold-scrap/main/emas_data.csv))

# data\_table

Year	Average Closing Price	Year Open	Year High	Year Low	Year Close	Annual % Change
1969	\$41.10	\$41.80	\$43.75	\$35.00	\$35.21	-16.07%
1970	\$35.96	\$35.13	\$39.19	\$34.78	\$37.38	6.16%
1971	\$40.80	\$37.33	\$43.90	\$37.33	\$43.50	16.37%
1972	\$58.17	\$43.73	\$70.00	\$43.73	\$64.70	48.74%
1973	\$97.12	\$64.99	\$127.00	\$64.10	\$112.25	73.49%

## data\_plot

	Time	Value
0	Jun, 1965	\$323.05
1	May, 1965	\$325.10
2	Apr, 1965	\$325.10
3	Mar, 1965	\$326.16
4	Feb, 1965	\$327.18

Pada data\_plot hanya terdapat harga emas setiap bulannya.



# PREPROCESSING

# Tahap Preprocessing

- Menghilangkan tanda ‘,’ ‘%’, dan ‘\$’
- Menghilangkan data duplikat pada data\_plot
- Membuat dataframe baru yang berisi annual % change berdasarkan data\_plot yang dinamakan change\_plot

## change\_plot

Time	Annual_Change
1969-12-31	0.000000
1970-12-31	-17.414263
1971-12-31	6.688825
1972-12-31	37.157344
1973-12-31	73.151381

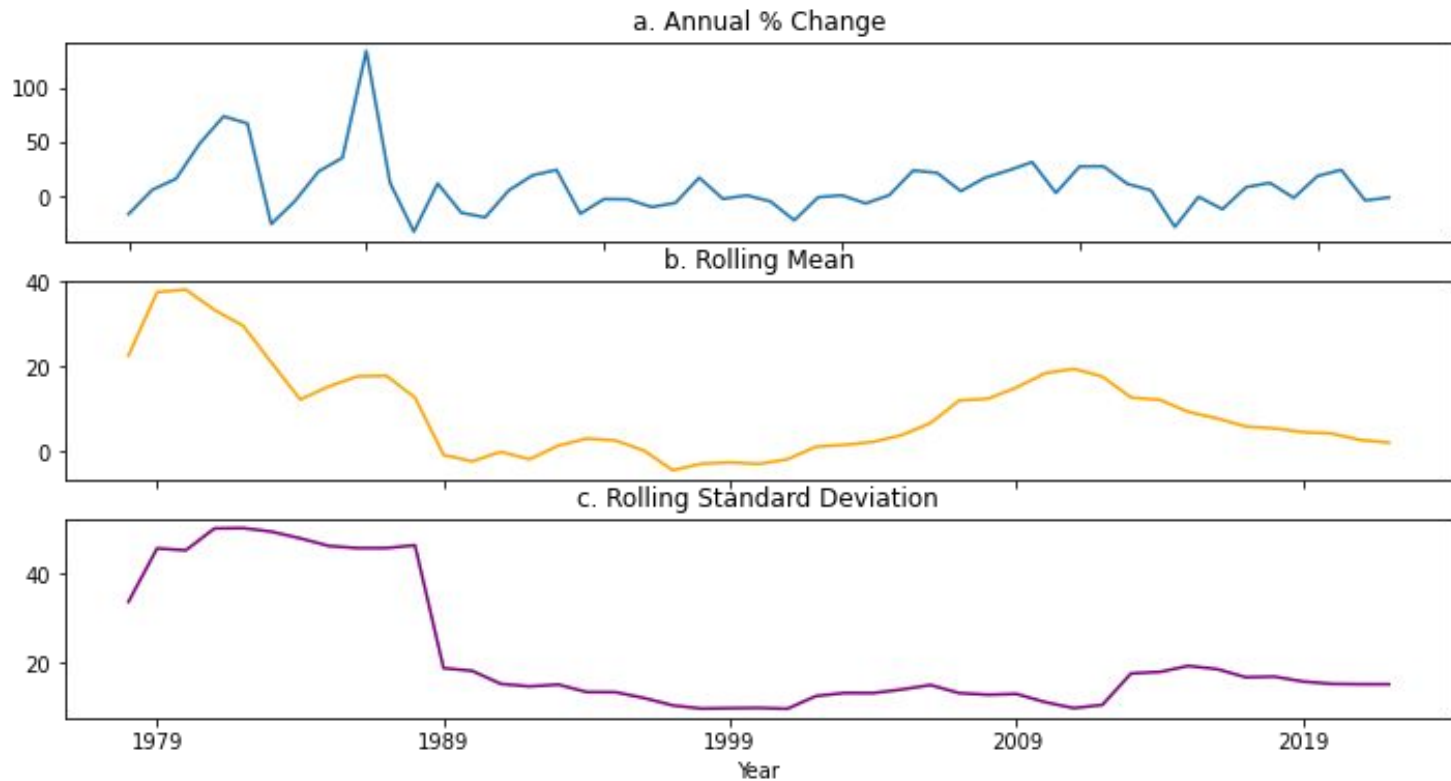
Terdapat adanya perbedaan annual % change yang didapatkan dari harga emas per bulan dengan hasil scraping dari tabel.





EDA

# Visualisasi



# Penjelasan

- Berdasarkan grafik tersebut, dapat dilihat bahwa tidak ada trends dan seasonal yang dapat ditemukan.
- Berdasarkan rata-rata dan varian pada data, terlihat bahwa data tersebut merupakan stasioner. Untuk lebih pasti, akan menggunakan **Augmented Dickey Fuller Test**.
- Berdasarkan Augmented Dickey Fuller test, didapatkan nilai  **$p < 0.05$** .
- Sehingga, data annual % change merupakan data yang **stasioner**.

# Stationarity

```
[27] from statsmodels.tsa.stattools import adfuller
      from numpy import log
      result = adfuller(df['Annual % Change'].values.flatten())
      print('ADF Statistic: %f' % result[0])
      print('p-value: %f' % result[1])

ADF Statistic: -4.103697
p-value: 0.000955
```

Nilai  $p < 0.05$  sehingga data annual % change merupakan stasioner



MODEL LSTM

# Arsitektur LSTM

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(1, 10, 100)	42800
dropout (Dropout)	(1, 10, 100)	0
lstm_1 (LSTM)	(1, 10, 100)	80400
dropout_1 (Dropout)	(1, 10, 100)	0
lstm_2 (LSTM)	(1, 100)	80400
dense (Dense)	(1, 5)	505
=====		
Total params: 204,105		
Trainable params: 204,105		
Non-trainable params: 0		

Kami membuat arsitektur LSTM dengan menggunakan library **Keras**. Arsitektur yang dibuat dengan menghubungkan 3 layer LSTM yang diakhiri dengan dense yang akan mengeluarkan nilai Annual % Change untuk 5 tahun kedepan.

# Dataset yang digunakan

Dengan model arsitektur yang sama, kami melatih dan memprediksi model tersebut dengan 3 data yang berbeda, yaitu:

- Multivariate data\_table: Memprediksi annual % change untuk 5 tahun kedepan berdasarkan Average Closing Price, Year Open, Year High, Year Low, Year Close, dan Annual % Change 10 tahun terakhir
- Univariate data\_table: Memprediksi annual % change untuk 5 tahun kedepan berdasarkan Annual % Change 10 tahun terakhir
- Univariate data\_plot: Memprediksi annual % change untuk 5 tahun kedepan berdasarkan Annual % Change 10 tahun terakhir

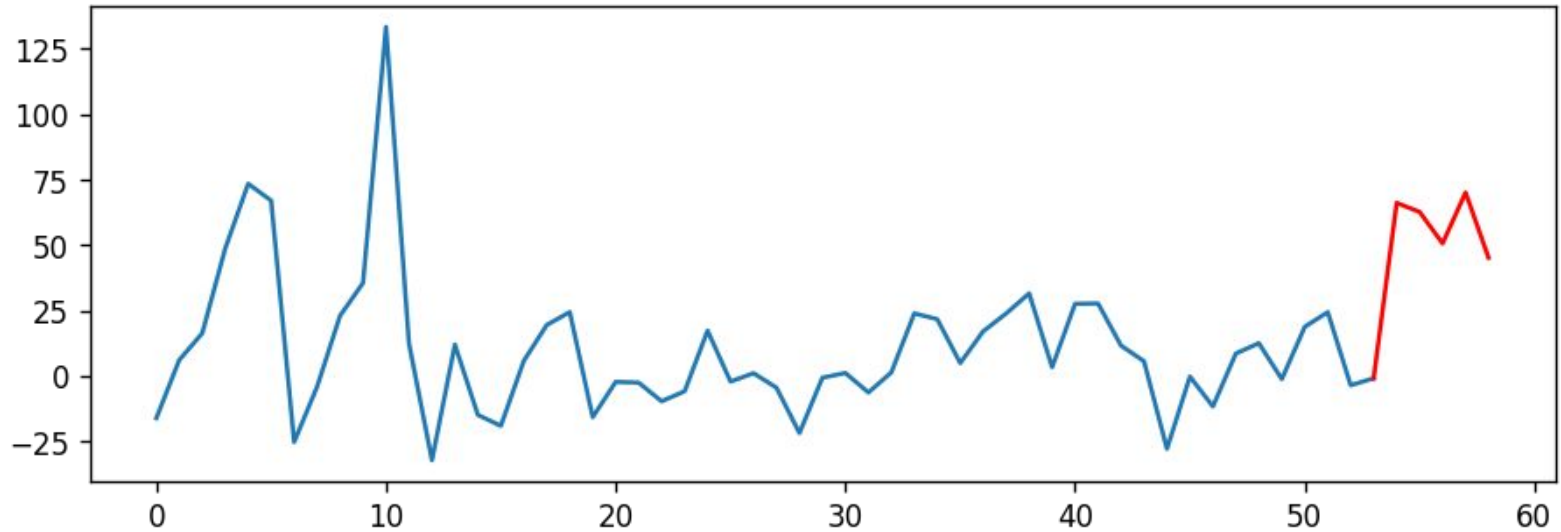
**Setiap data yang akan digunakan akan diubah menjadi lag feature.**

# Contoh Lag Feature

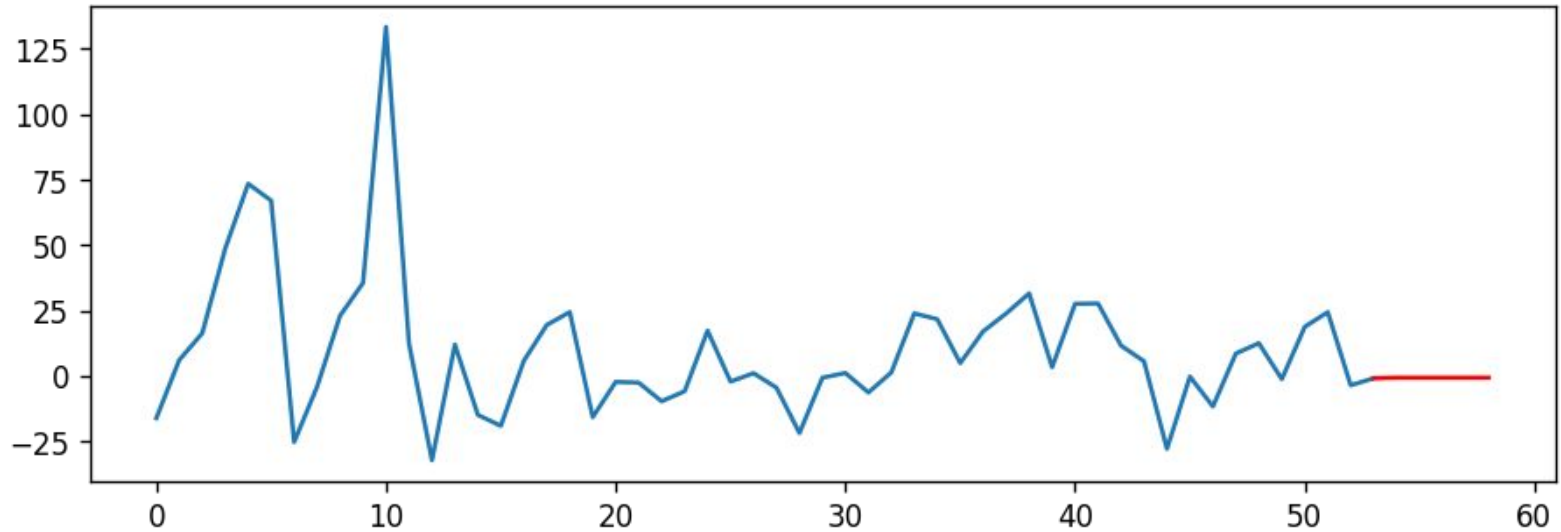
	Annual_Change(t-3)	Annual_Change(t-2)	Annual_Change(t-1)	Annual_Change(t)	Annual_Change(t+1)
0	0.000000	-17.414263	6.688825	37.157344	73.151381
1	-17.414263	6.688825	37.157344	73.151381	35.460321
2	6.688825	37.157344	73.151381	35.460321	-7.910337
3	37.157344	73.151381	35.460321	-7.910337	-25.895206



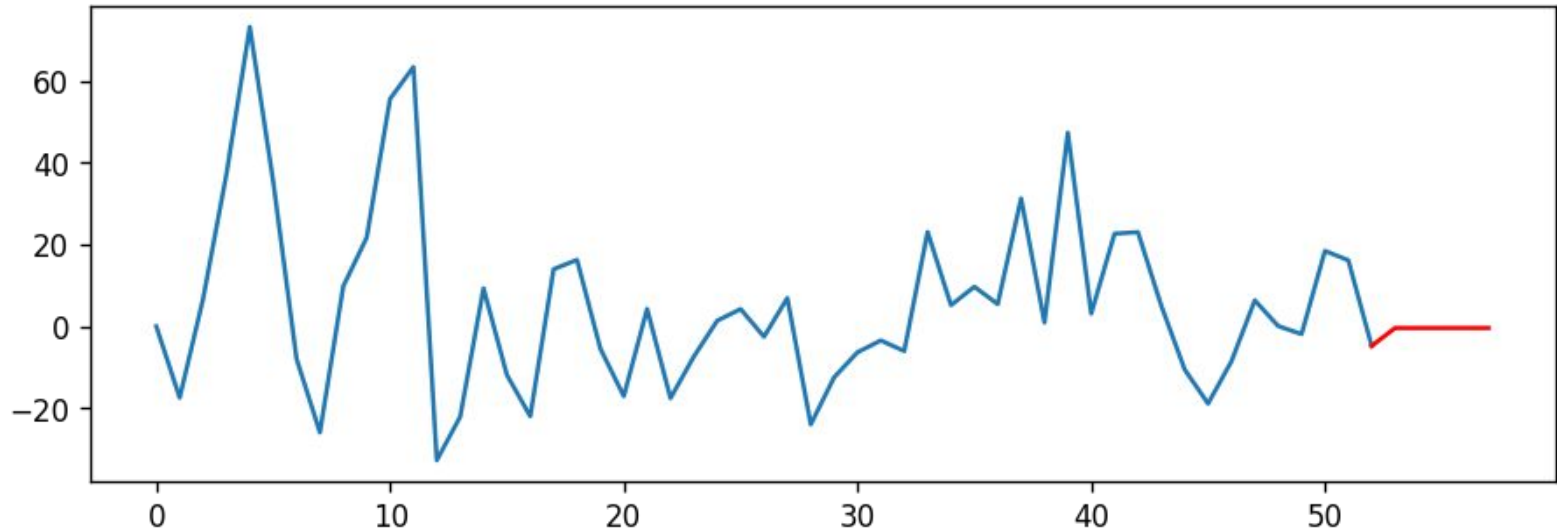
## Hasil Prediksi Multivariate data\_table



## Hasil Prediksi Univariate data\_table



## Hasil Prediksi Univariate data\_plot





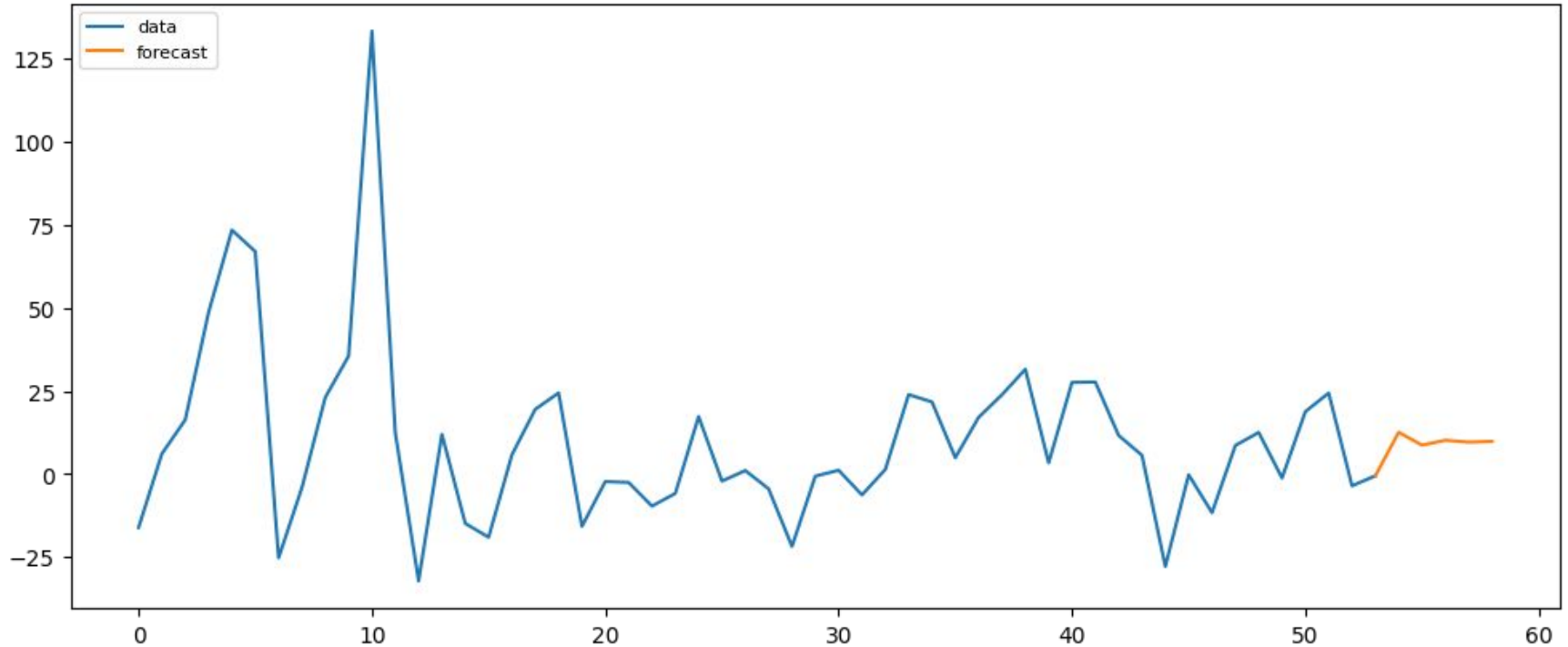
# MODEL ARIMA

# Parameter ARIMA (p, d, q) data\_table

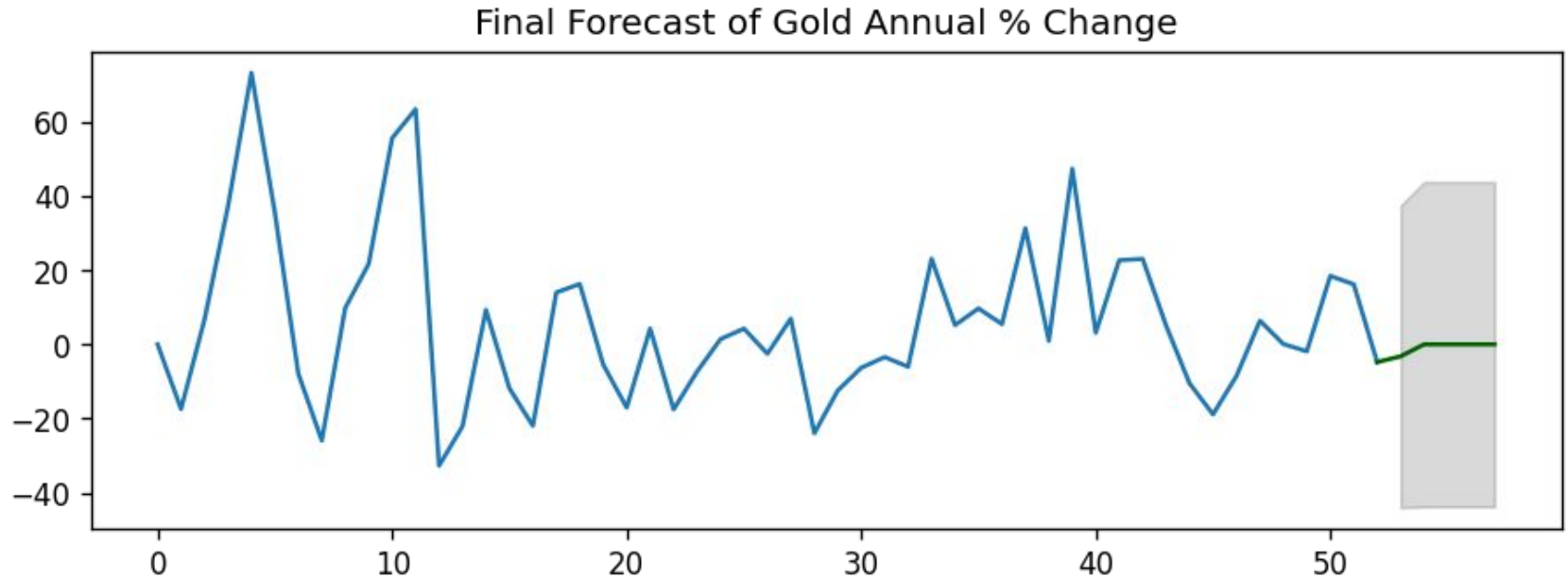
- Tidak ada nonseasonal differences yang dibutuhkan untuk mencapai stationarity menurut Augmented Dickey–Fuller test. Maka, **d bernilai 0**.
- Menurut grafik auto-correlation function (ACF) dan partial auto-correlation function (PACF), autoregressive term berjumlah 1 dan jumlah lagged forecast error pada prediksi berjumlah 1. Sehingga, didapatkan variabel **p bernilai 1** dan variabel **q bernilai 1**.
- Parameter yang digunakan = **(1, 0, 1)**

# Hasil Prediksi ARIMA (1, 0, 1) data\_table

Final Forecast



# Hasil Prediksi ARIMA (0, 0, 1) data\_plot





# EVALUASI



# Pemilihan Metrik

- Akan dibandingkan error dari masing-masing model dengan data test yang sudah dibagi.
- Metrik: **mean absolute error (MAE)** dan **root mean square error (RMSE)**.
- Kedua error tersebut dipilih karena sensitif terhadap outlier yang dihasilkan.

# Evaluasi data\_plot

- Model LSTM univariate data\_plot memiliki error yang lebih kecil dibandingkan model-model lainnya.
- Namun dikarenakan data\_plot menggunakan dataset yang berbeda, kemungkinan errornya memang lebih kecil.

# Evaluasi data\_table

- Model LSTM univariate data\_table memiliki error MAE dan RMSE yang lebih kecil dibandingkan model lainnya yang menggunakan data\_table juga.
- Hasil dari model LSTM multivariate data\_table kurang stabil dibanding model lainnya karena mempertimbangkan variable yang lebih banyak dibanding univariate, sehingga lebih mudah terpengaruh.



# KESIMPULAN

# Kesimpulan

- Model LSTM univariate data\_plot memiliki hasil yang lebih baik dibandingkan dengan model lainnya dalam hal error MAE dan RMSE.
- Untuk data\_table, model LSTM univariate data\_table lebih baik dalam hal error MAE dan RMSE dibandingkan model multivariate ataupun ARIMA.
- **Model LSTM dengan epoch dan arsitektur yang baik dapat menghasilkan prediksi yang lebih bagus dibandingkan model ARIMA.**



TERIMA KASIH