

Analisis Prediksi Harga Pembelian Tiket Pesawat Pulang Pergi CGK – DPS dan CGK – KUL pada Liburan Akhir Tahun dengan Algoritma XGBoost dan LSTM

Dosen Fasilitator:
Pak Dr. Dipo Aldila, S.Si., M.Si.

Anggota Kelompok

Rachel Octaviani Putri	2006463585
Cornelius Justin Satryo Hadi	2006529796
Lutfia Maulidina	2006483151
Safira Raissa Rahmi	2006568891
Muhammad Rafly Januar	2006568645
Muhammad Habil Amardias	2006572251

Contents

1

Latar Belakang

2

Pengambilan Data

3

Rumusan Masalah

4

Tujuan Penelitian

5

Batasan Masalah

6

Metode

7

Implementasi

8

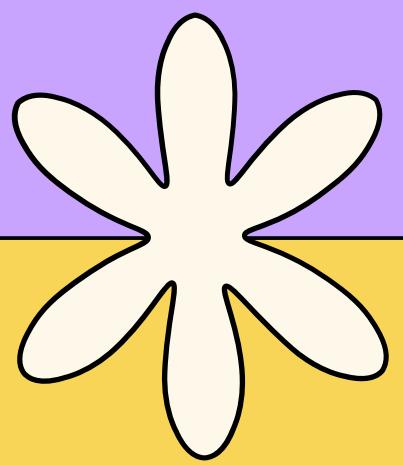
Analisis dan Hasil

9

Kesimpulan



Latar Belakang



Perubahan Harga Tiket Pesawat

Diambil tanggal 30 September (Pagi)

Tanpa henti • 1hr 50mins

CGK 07:55

Jakarta
16 Okt (Min)



SIN 10:45

Singapura
16 Okt (Min)

Singapore Airlines • SQ 953

[> Detail lebih lanjut](#)

ECONOMY

DARI IDR
9,586,400
PER DEWASA



Diambil tanggal 30 September (Malam)

Tanpa henti • 1hr 50mins

CGK 07:55

Jakarta
16 Okt (Min)



SIN 10:45

Singapura
16 Okt (Min)

Singapore Airlines • SQ 953

[> Detail lebih lanjut](#)

ECONOMY

DARI IDR
7,396,400
PER DEWASA



Diambil tanggal 1 Oktober (Pagi)

Tanpa henti • 1hr 50mins

CGK 07:55

Jakarta
16 Okt (Min)



SIN 10:45

Singapura
16 Okt (Min)

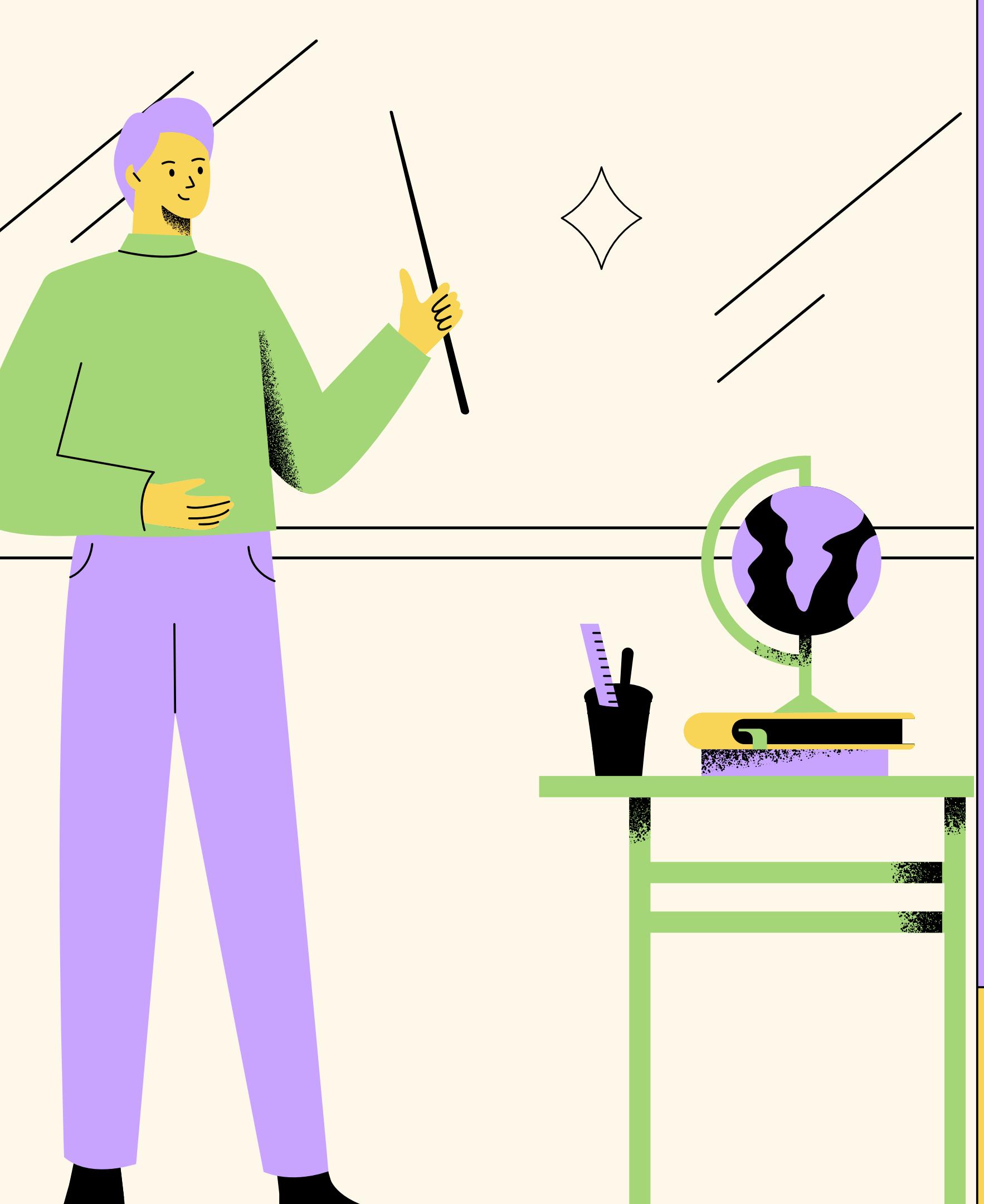
Singapore Airlines • SQ 953

[> Detail lebih lanjut](#)

ECONOMY

DARI IDR
2,786,400
PER DEWASA



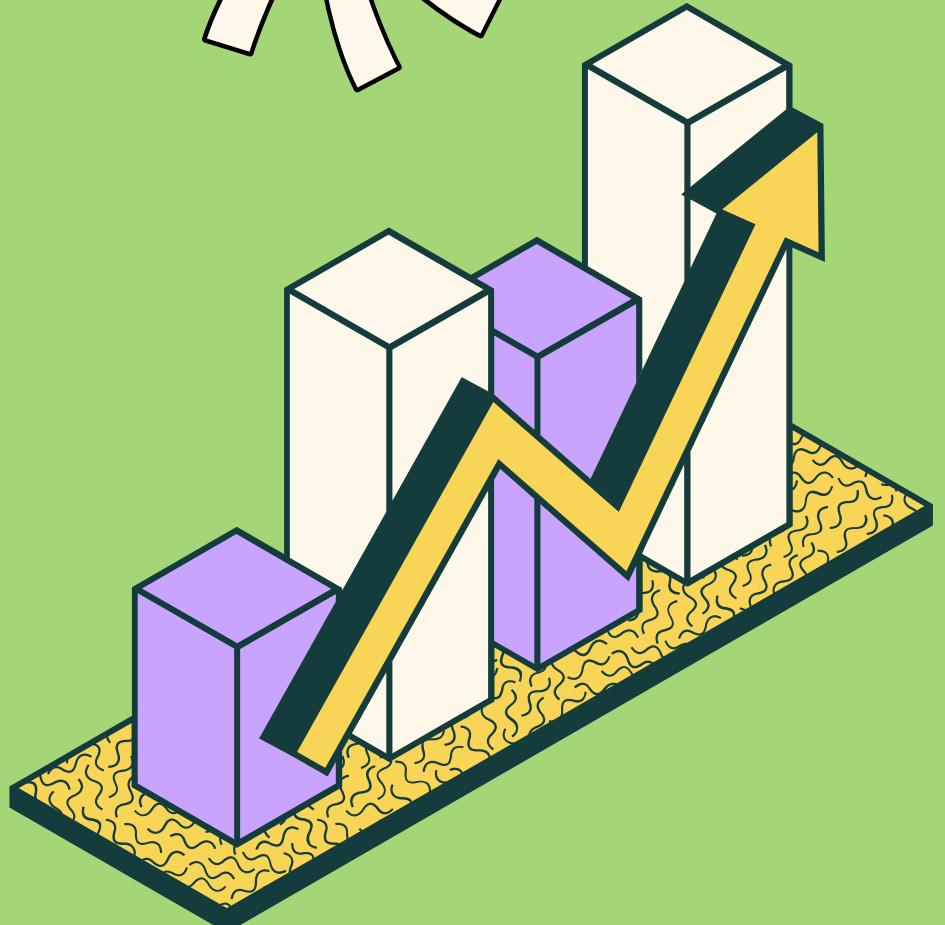


Pengambilan Data

Data yang Diambil

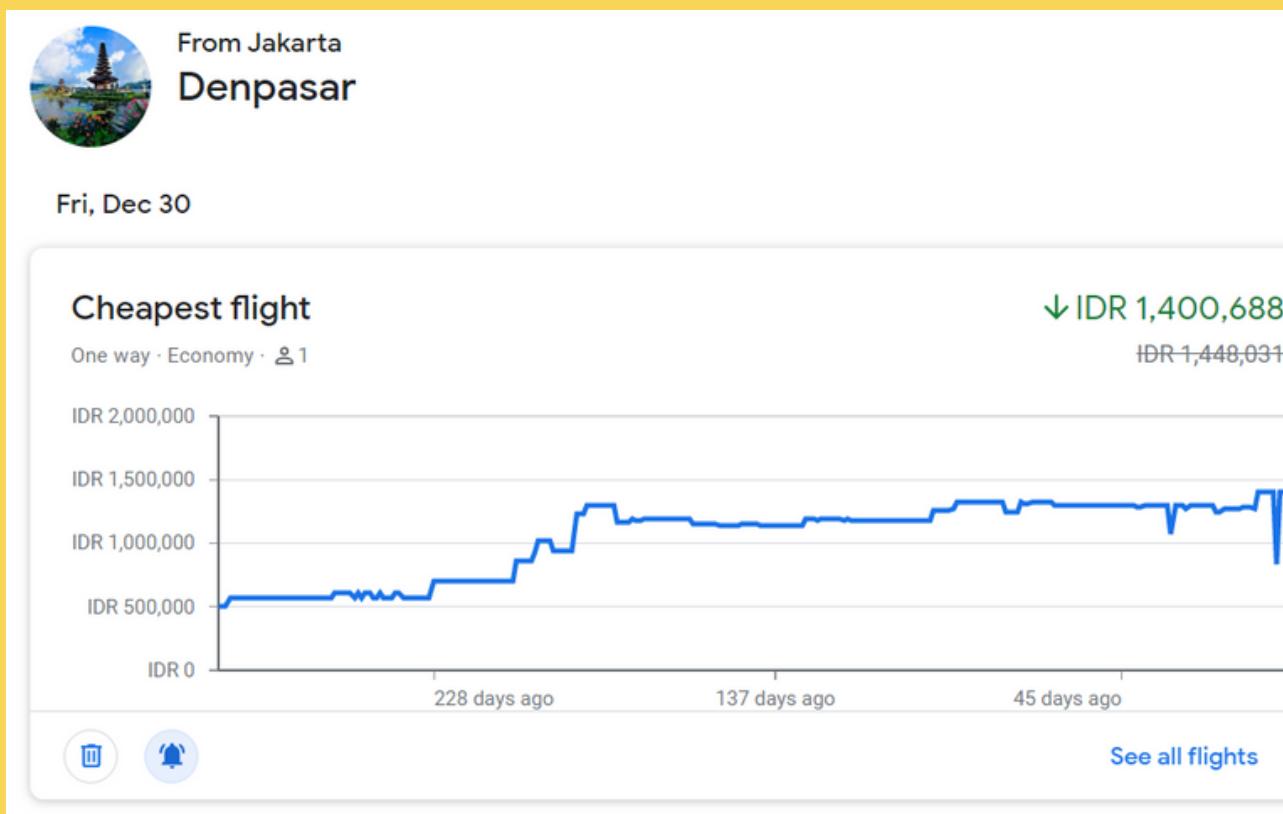
Google Flight

- Rute Jakarta (CGK) - Denpasar (DPS): Jumat, 30 Desember 2022
- Rute Denpasar (DPS) - Jakarta (CGK): Senin, 2 Januari 2023
- Rute Jakarta (CGK) - Kuala Lumpur (KUL): Jumat, 30 Desember 2022
- Rute Kuala Lumpur (KUL) - Jakarta (CGK): Senin, 2 Januari 2023



*Harga tiket yang diambil adalah harga termurah dari semua maskapai dan tiket single flight dari <https://www.google.com/travel/flights>

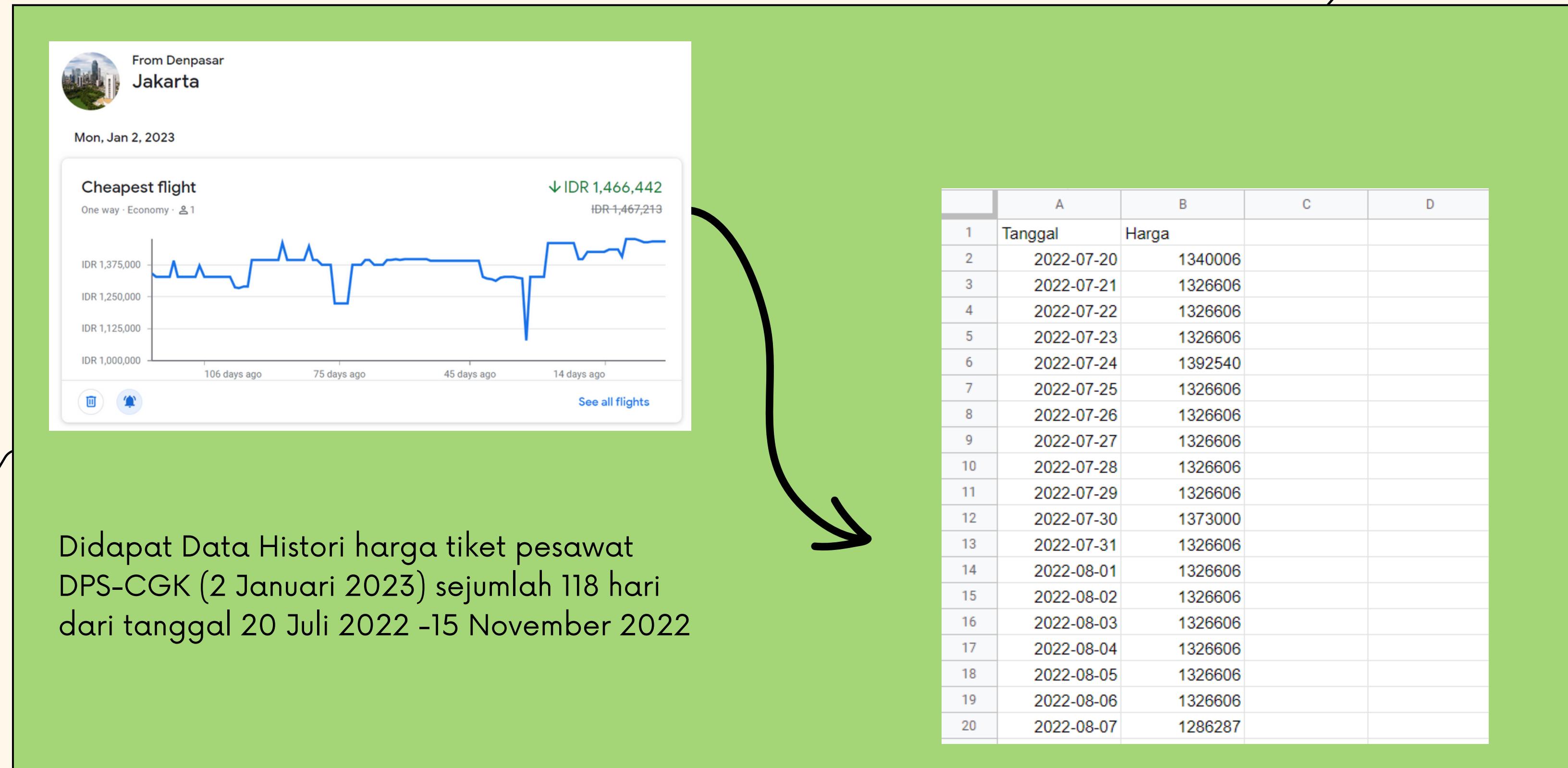
CGK - DPS : 30 Desember 2022



Didapat Data Histori harga tiket pesawat CGK - DPS (30 Desember 2022) sejumlah 285 hari dari tanggal 3 Februari 2022 - 15 November 2022

	A	B	C	D
1	Tanggal	Harga		
2	2022-02-03	499000		
3	2022-02-04	499000		
4	2022-02-05	499000		
5	2022-02-06	566100		
6	2022-02-07	566100		
7	2022-02-08	566100		
8	2022-02-09	566100		
9	2022-02-10	566100		
10	2022-02-11	566100		
11	2022-02-12	566100		
12	2022-02-13	566100		
13	2022-02-14	566100		
14	2022-02-15	566100		
15	2022-02-16	566100		
16	2022-02-17	566100		
17	2022-02-18	566100		
18	2022-02-19	566100		
19	2022-02-20	566100		
20	2022-02-21	566100		

DPS - CGK : 2 Januari 2023



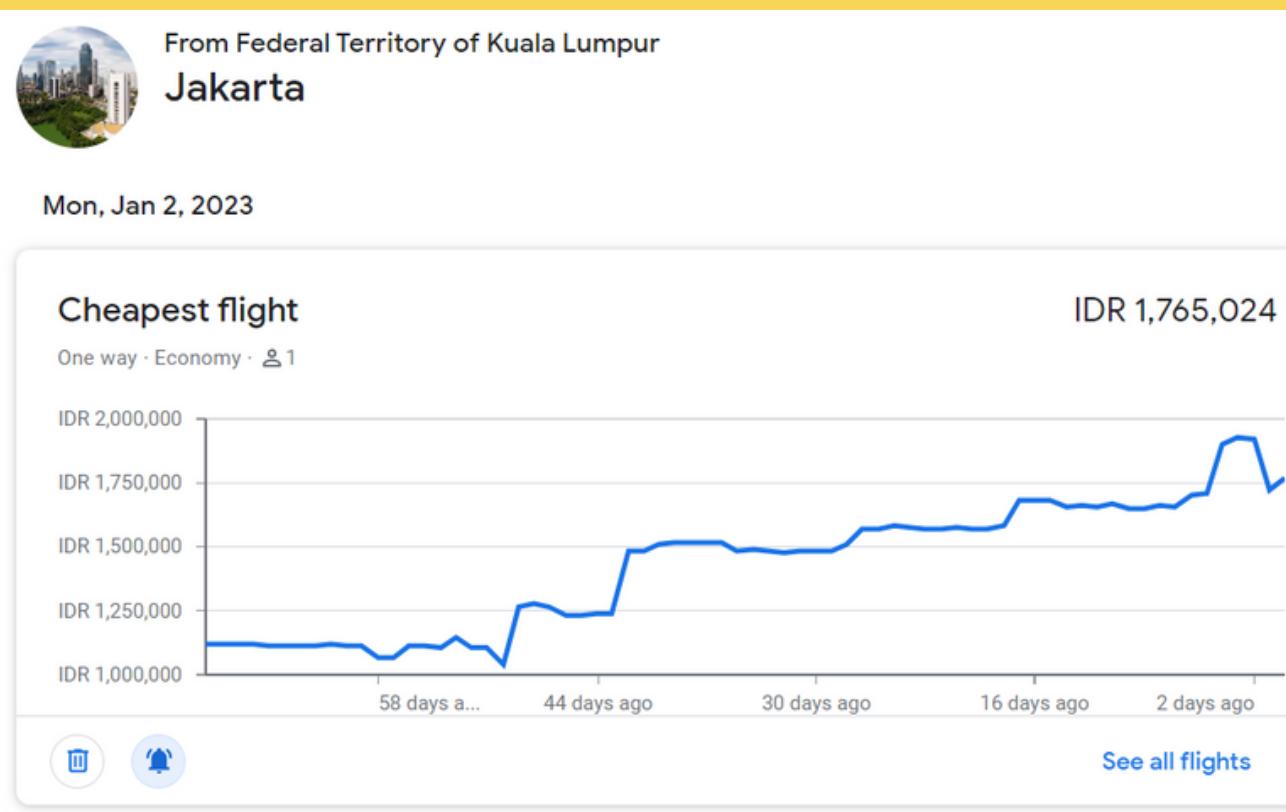
CGK - KUL : 30 Desember 2022



Didapat Data Histori harga tiket pesawat
CGK - KUL (30 Desember 2022) sejumlah 231
hari dari tanggal 3 Maret 2022 - 15
November 2022

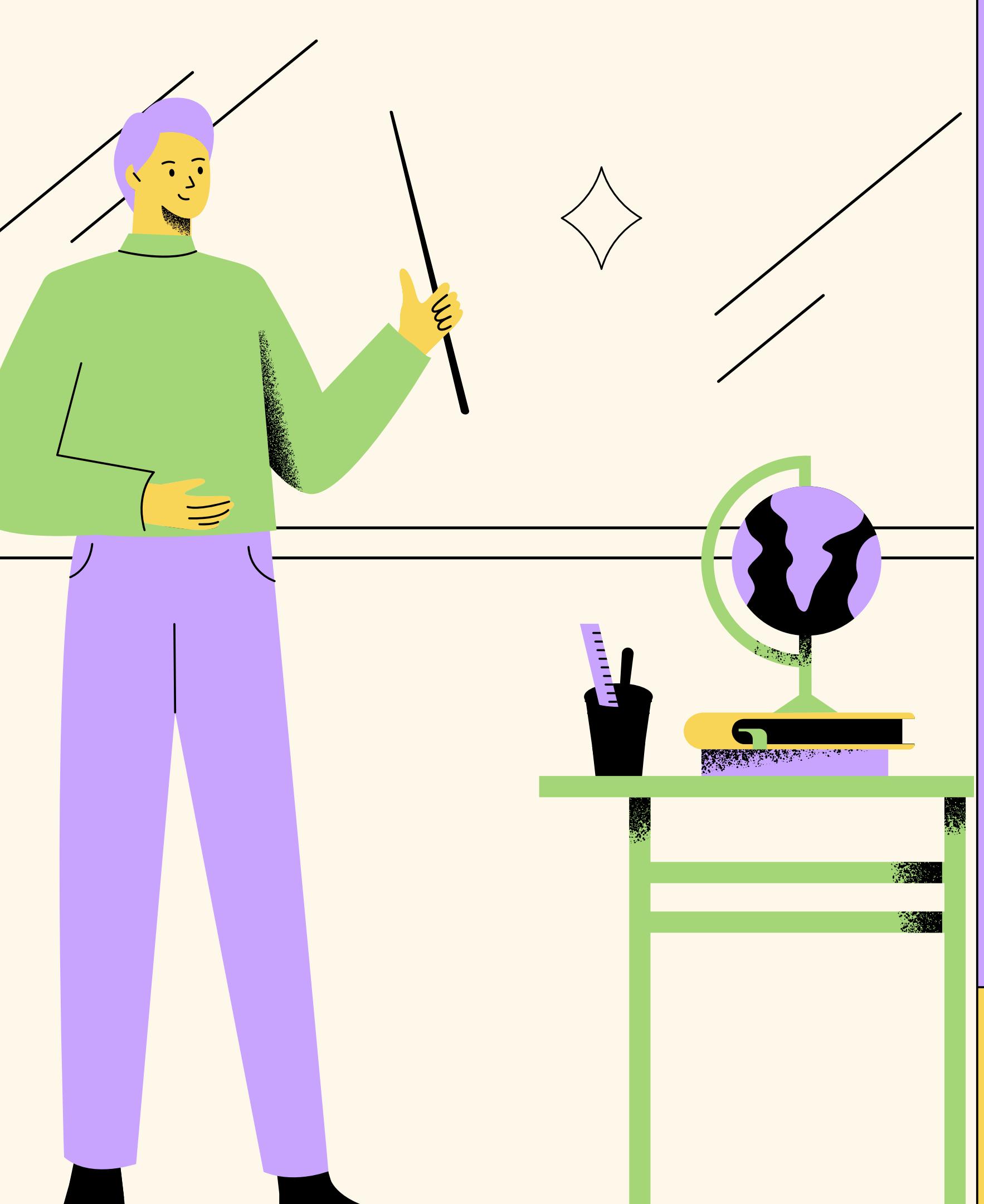
A	B	C	D
1	Tanggal	Harga	
2	2022-03-30	4341000	
3	2022-03-31	4341000	
4	2022-04-01	4341000	
5	2022-04-02	4341000	
6	2022-04-03	4341000	
7	2022-04-04	4341000	
8	2022-04-05	4341000	
9	2022-04-06	4341000	
10	2022-04-07	4341000	
11	2022-04-08	4341000	
12	2022-04-09	4341000	
13	2022-04-10	4341000	
14	2022-04-11	4341000	
15	2022-04-12	4341000	
16	2022-04-13	4341000	
17	2022-04-14	4341000	
18	2022-04-15	4341000	
19	2022-04-16	4341000	
20	2022-04-17	4341000	
21	2022-04-18	4341000	

KUL - CGK : 2 Januari 2023

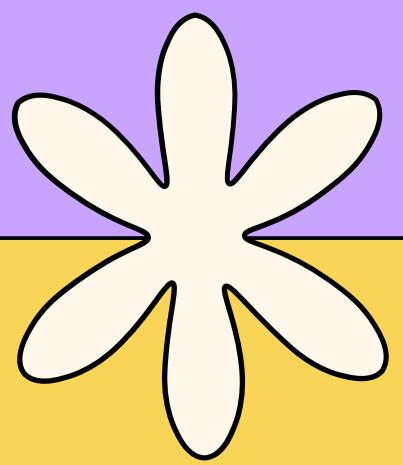


Didapat Data Histori harga tiket pesawat KUL - CGK (2 Januari 2023) sejumlah 69 hari dari tanggal 7 September 2022 - 15 November 2022

	A	B	C	D
1	Tanggal	Harga		
2	2022-09-07	1116970		
3	2022-09-08	1116855		
4	2022-09-09	1117184		
5	2022-09-10	1117088		
6	2022-09-11	1111940		
7	2022-09-12	1111096		
8	2022-09-13	1111096		
9	2022-09-14	1112733		
10	2022-09-15	1121524		
11	2022-09-16	1110685		
12	2022-09-17	1111469		
13	2022-09-18	1063021		
14	2022-09-19	1063021		
15	2022-09-20	1111171		
16	2022-09-21	1111803		
17	2022-09-22	1105947		
18	2022-09-23	1147052		
19	2022-09-24	1103713		
20	2022-09-25	1105721		

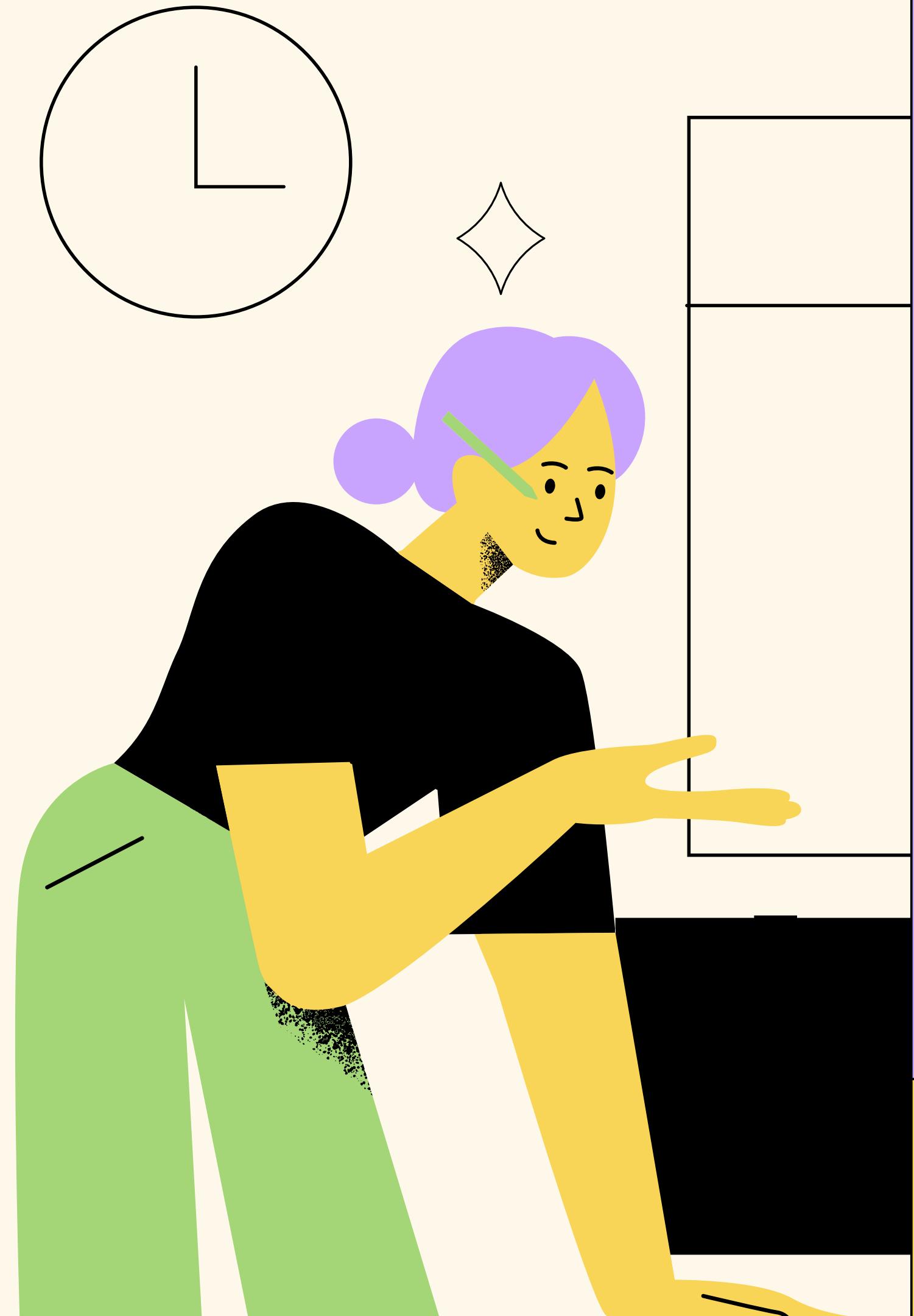


Rumusan Masalah

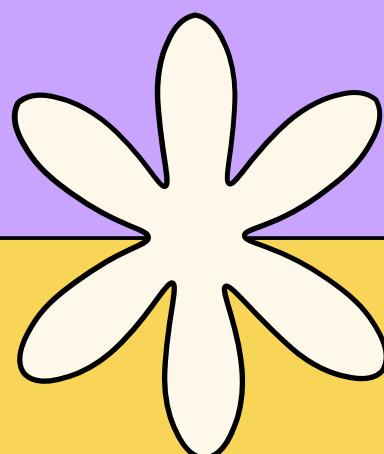


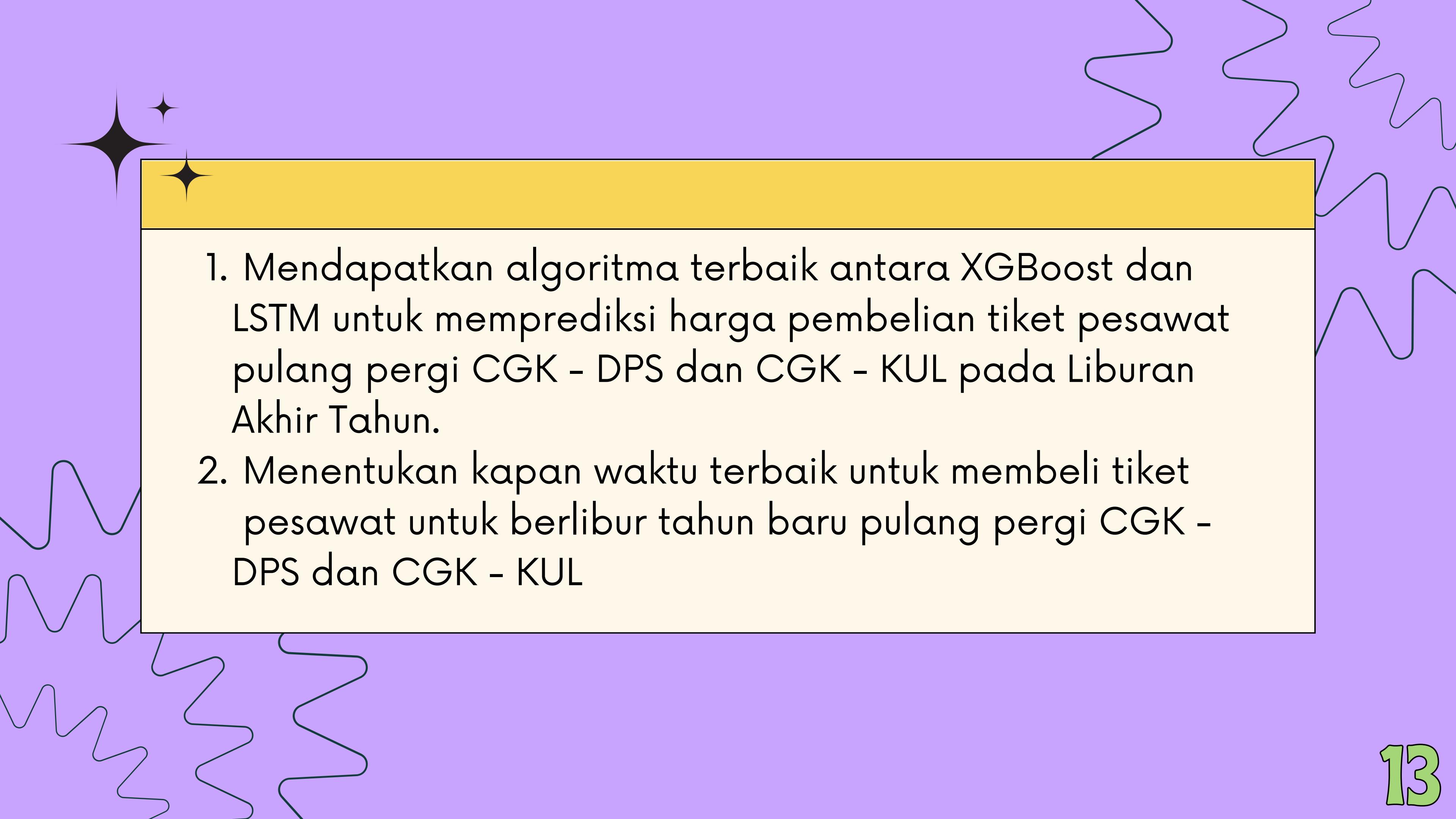
Manakah algoritma terbaik antara XGBoost dan LSTM untuk memprediksi harga pembelian tiket pesawat liburan tahun baru pulang pergi CGK - DPS dan CGK - KUL.

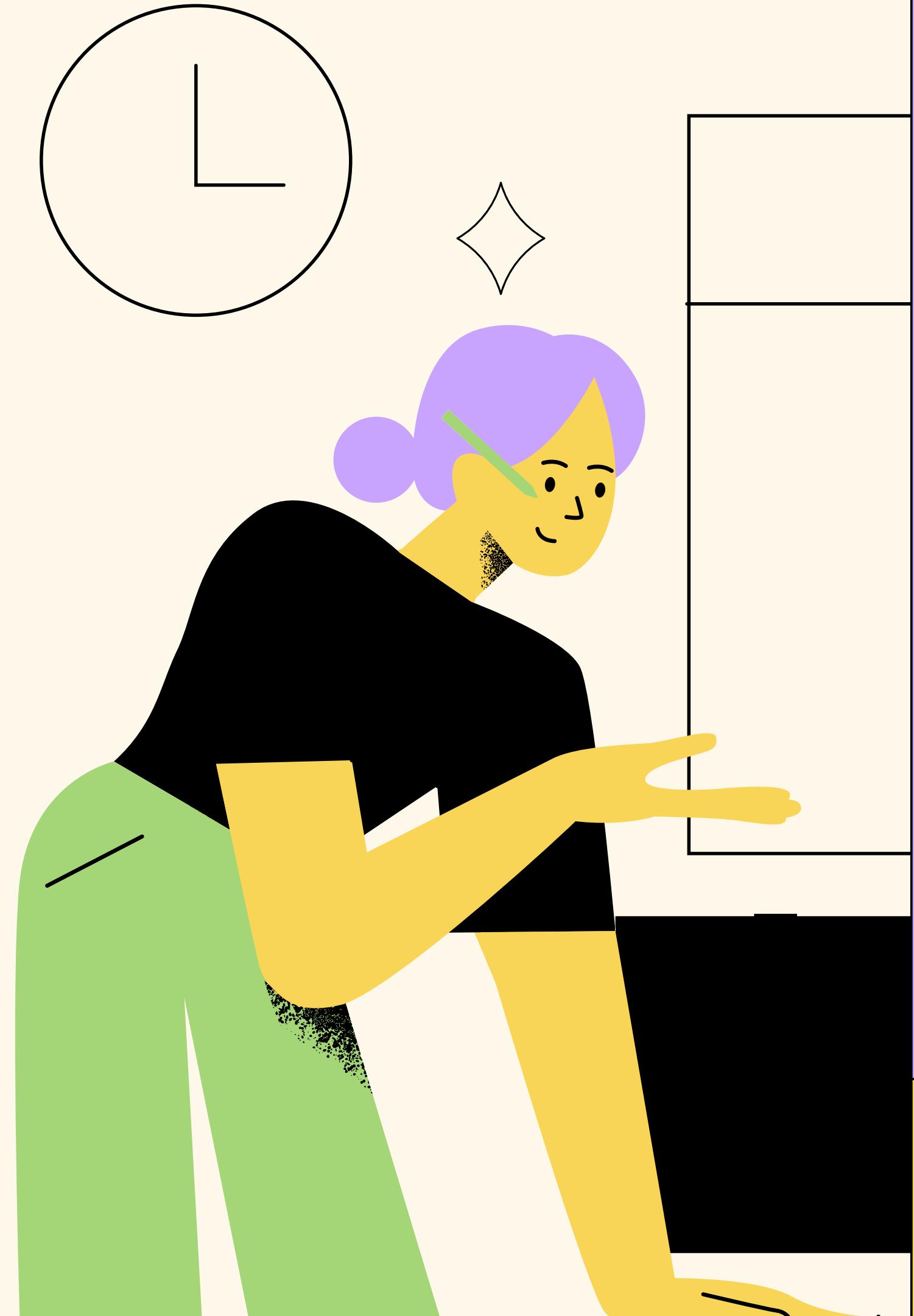
Kapan waktu terbaik untuk membeli tiket pesawat liburan tahun baru pulang pergi CGK - DPS dan CGK - KUL.



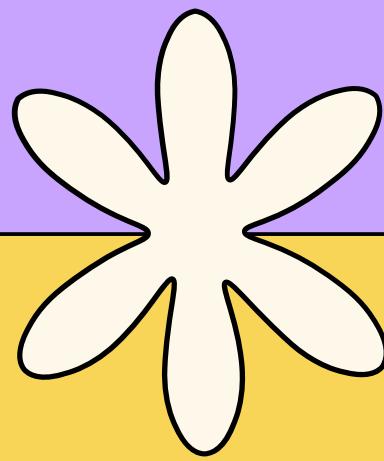
Tujuan Penelitian



- 
1. Mendapatkan algoritma terbaik antara XGBoost dan LSTM untuk memprediksi harga pembelian tiket pesawat pulang pergi CGK - DPS dan CGK - KUL pada Liburan Akhir Tahun.
 2. Menentukan kapan waktu terbaik untuk membeli tiket pesawat untuk berlibur tahun baru pulang pergi CGK - DPS dan CGK - KUL



Batasan Masalah

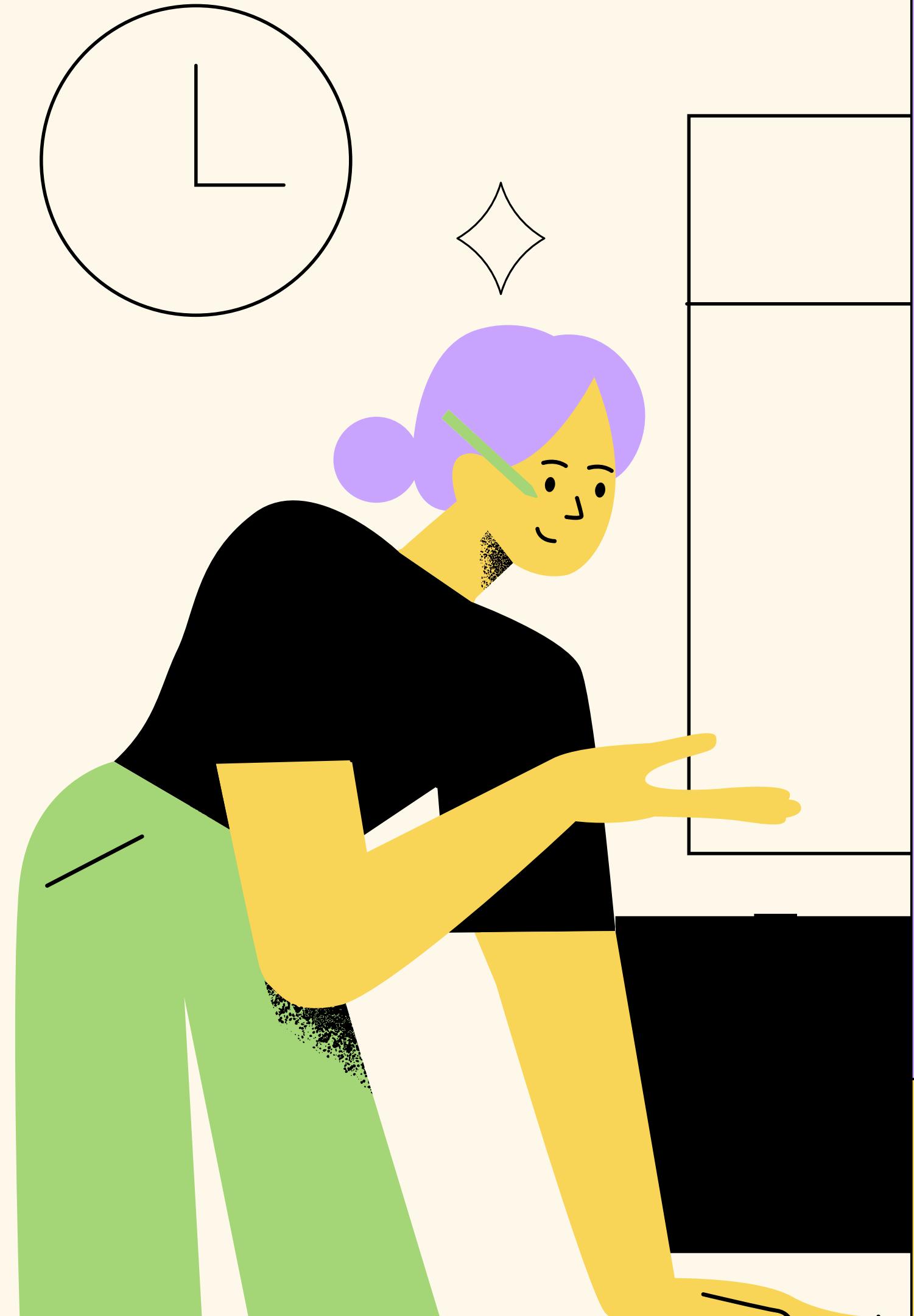


Rute pulang pergi Jakarta - CGK ke penerbangan domestik (Bali - DPS) dan internasional (Kuala Lumpur - KUL)

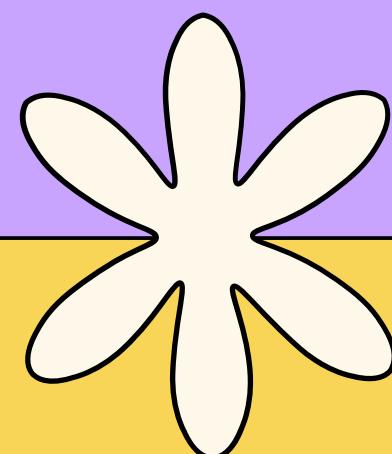
Pengamatan fluktuasi harga tiket pada liburan akhir tahun

Tanggal keberangkatan: 30 Desember 2022
Tanggal kepulangan: 2 Januari 2023

Kemampuan model diukur berdasarkan nilai RMSE



Metode (Time Series)



Autokorelasi

Autokorelasi menghitung korelasi antara lagged version dari nilai suatu variabel dengan versi aslinya dalam sebuah time series.

Autokorelasi positif menunjukkan bahwa apabila nilai pada interval waktu yang diamati meningkat maka nilai pada versi lagged nya juga cenderung meningkat.

Autokorelasi menghitung korelasi antara 2 interval waktu yang berbeda, X_t dan X_{t-n} dari time series yang sama, n disebut dengan lag.

Autokorelasi negatif menunjukkan apabila nilai pada interval waktu yang diamati meningkat maka nilai pada versi lagged nya cenderung menurun.

RMSE

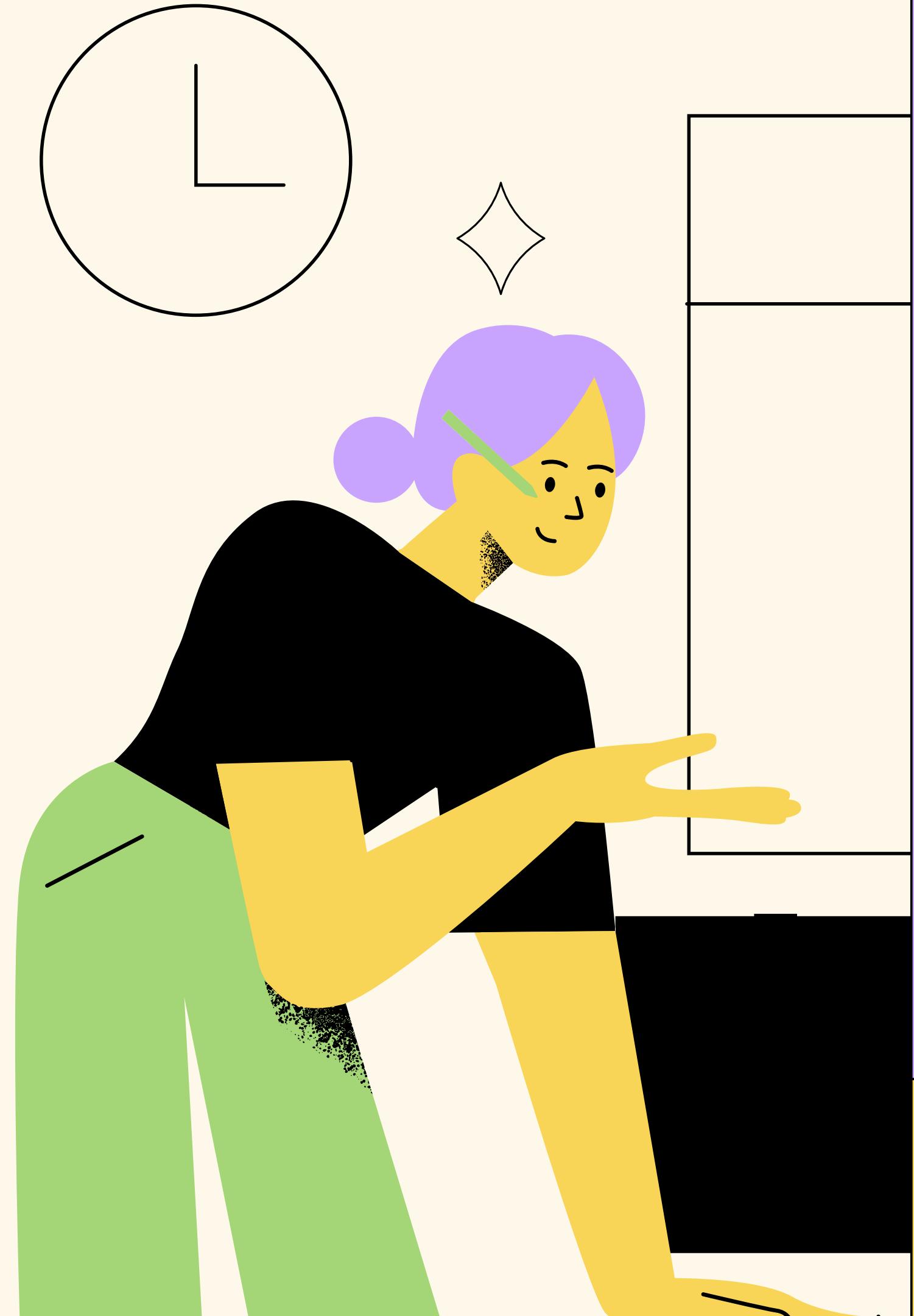
RMSE adalah sebuah ukuran (metrik) yang menunjukkan rata rata jarak atau selisih antara nilai yang diprediksi oleh model dengan nilai asli dari dataset

$$\text{RMSE} = \sqrt{\sum(P_i - O_i)^2 / n}$$

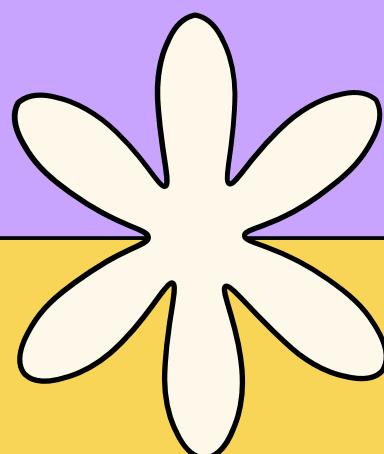
P_i: adalah nilai yang diprediksi oleh model pada observasi ke i pada dataset

O_i: nilai asli atau nilai yang diobservasi pada observasi ke i pada dataset

n: Ukuran sampel

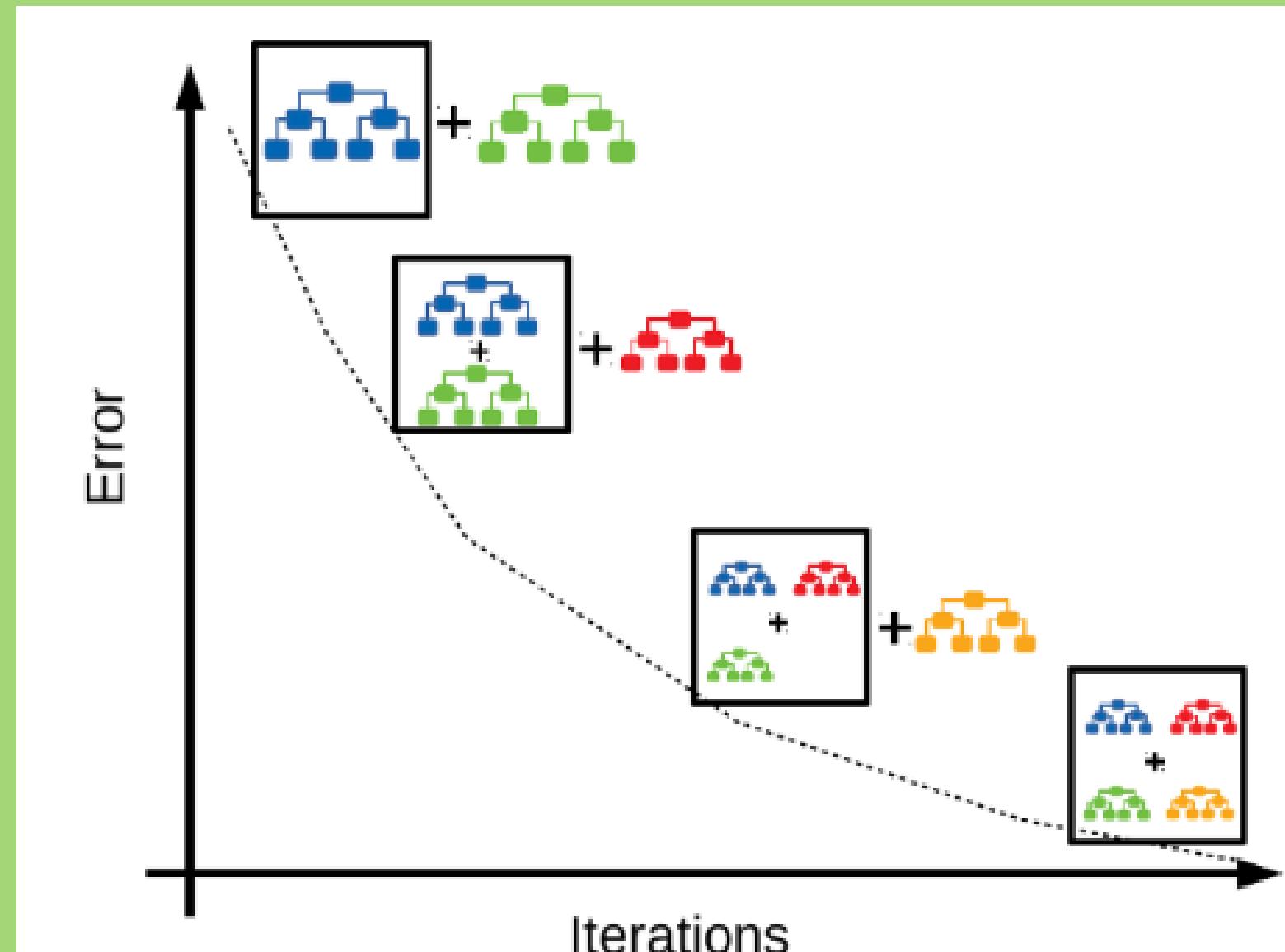


Metode (Model)



XGBoost

Extreme Gradient Boosting

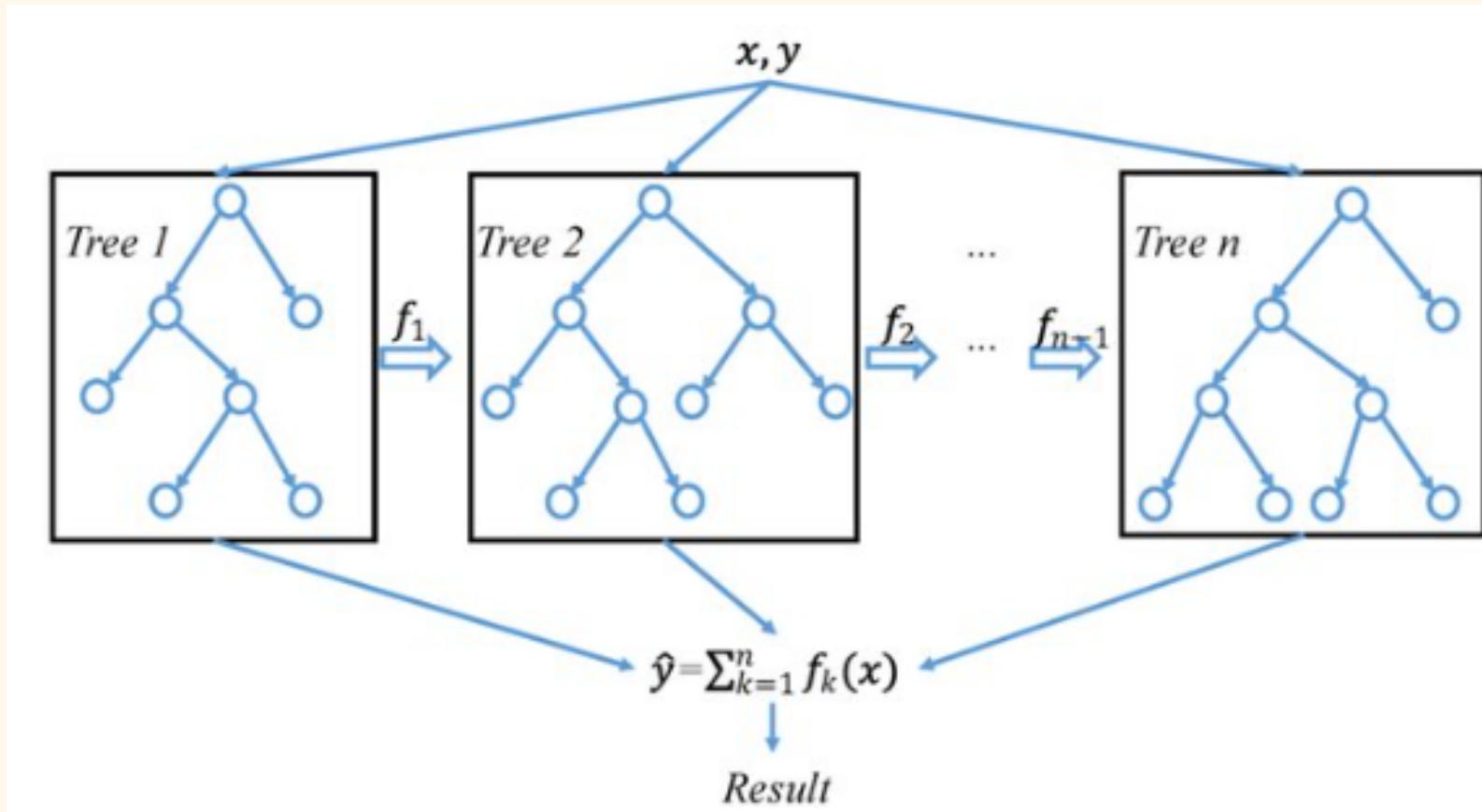


Sumber Gambar: Pal, Aratrika. 2020. Gradient Boosting Trees for Classification: A Beginner's Guide . Diakses melalui: <https://medium.com/swlh/gradient-boosting-trees-for-classification-a-beginners-guide-596b594a14ea>

- Gradient boosting merupakan algoritma machine learning yang menggunakan ensamble dari decision tree untuk memprediksi nilai.
- Gradient boosting dimulai dengan menghasilkan pohon klasifikasi awal dan terus menyesuaikan pohon baru melalui minimalisasi fungsi kerugian (Natekin dan Knoll, 2013).
- Tahapan dari algoritma gradient boosting:
 1. Inisialisasi prediksi.
 2. Hitung nilai residual.
 3. Bentuk tree baru untuk memprediksi nilai residual.
 4. Jumlahkan prediksi inisial dengan prediksi residual.
 5. Cari nilai residual baru.
 6. Ulangi langkah 3 hingga 5 sampai nilai residual mendekati nol atau jumlah iterasi sesuai dengan hyperparameter ketika menjalankan algoritmanya.

XGBoost

Extreme Gradient Boosting



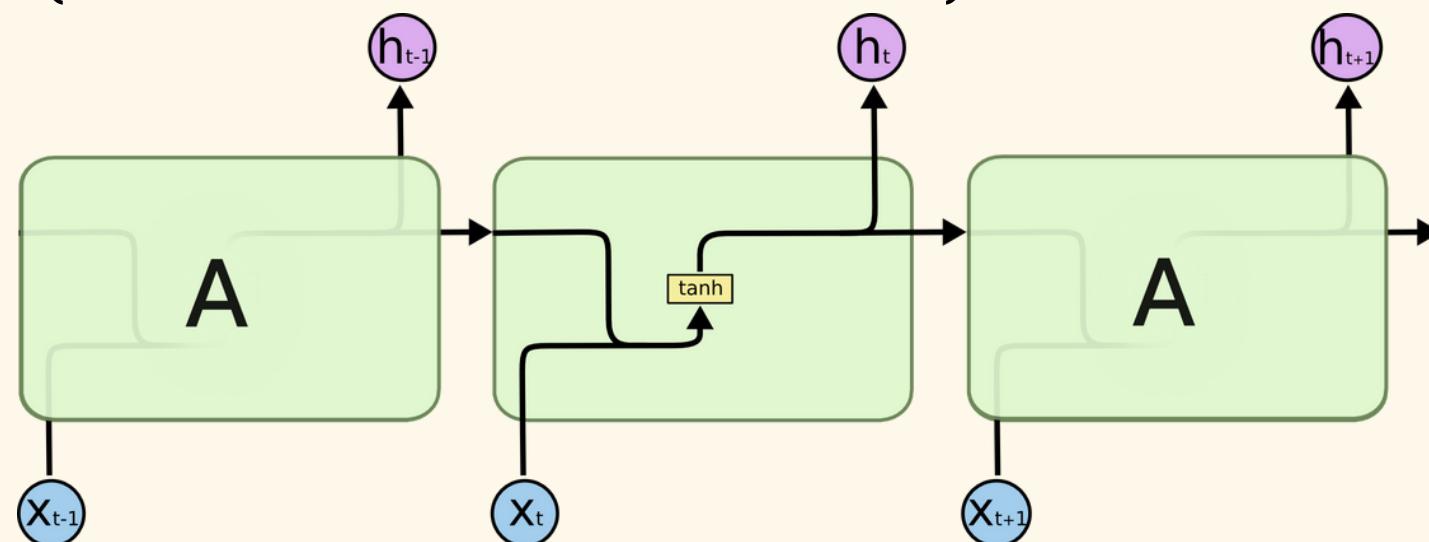
Sumber Gambar: Y, Wang et al. 2019. A hybrid ensemble method for pulsar candidate classification. *Astrophys Space Sci*.

- Extreme Gradient Boosting (XGBoost) merupakan pengembangan dari algoritma Gradient Boosting Machine klasik dan hanya digunakan untuk data yang memiliki label dalam proses latihnya (Syahrani, 2019).
- XGBoost merupakan tree ensembles algoritma yang terdiri atas kumpulan beberapa classification and regression trees (CART).
- XGBoost mampu mengerjakan berbagai fungsi seperti regresi, klasifikasi, dan ranking.
- XGBoost menggunakan algoritma Gradient Boosting yang sudah dioptimalkan, seperti memiliki hyperparameter regularisasi untuk mencegah overfitting .

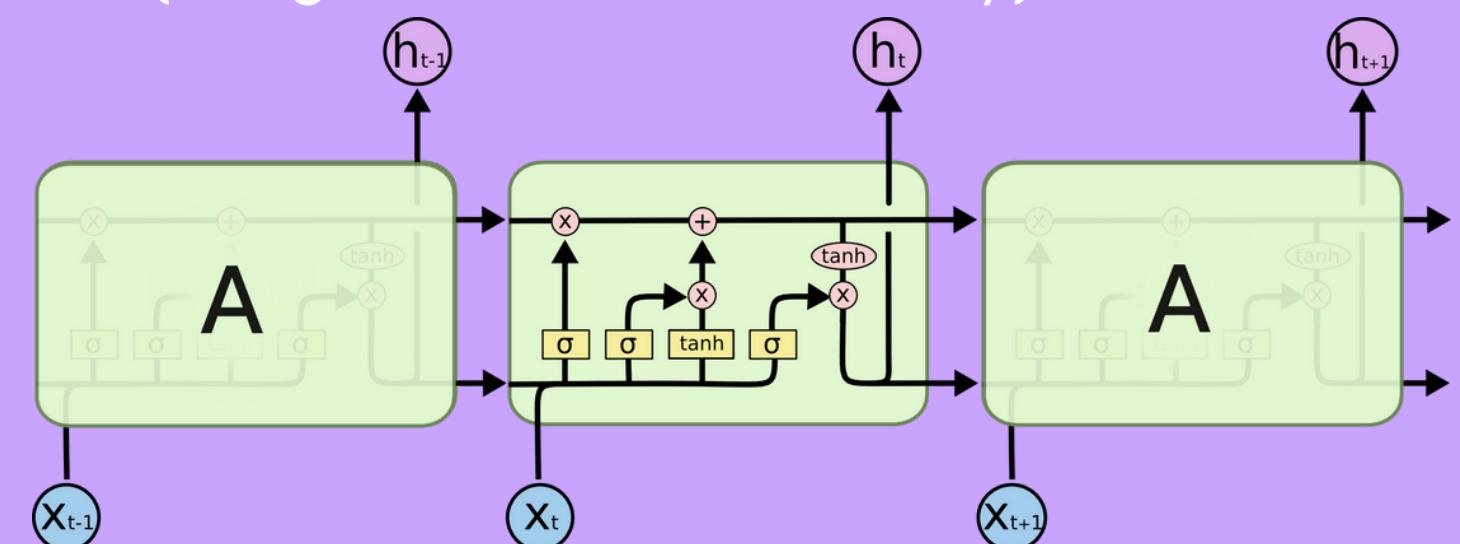
LSTM

Long Short Term Memory

RNN (Recurrent Neural Network)



LSTM (Long-Short Term Memory)



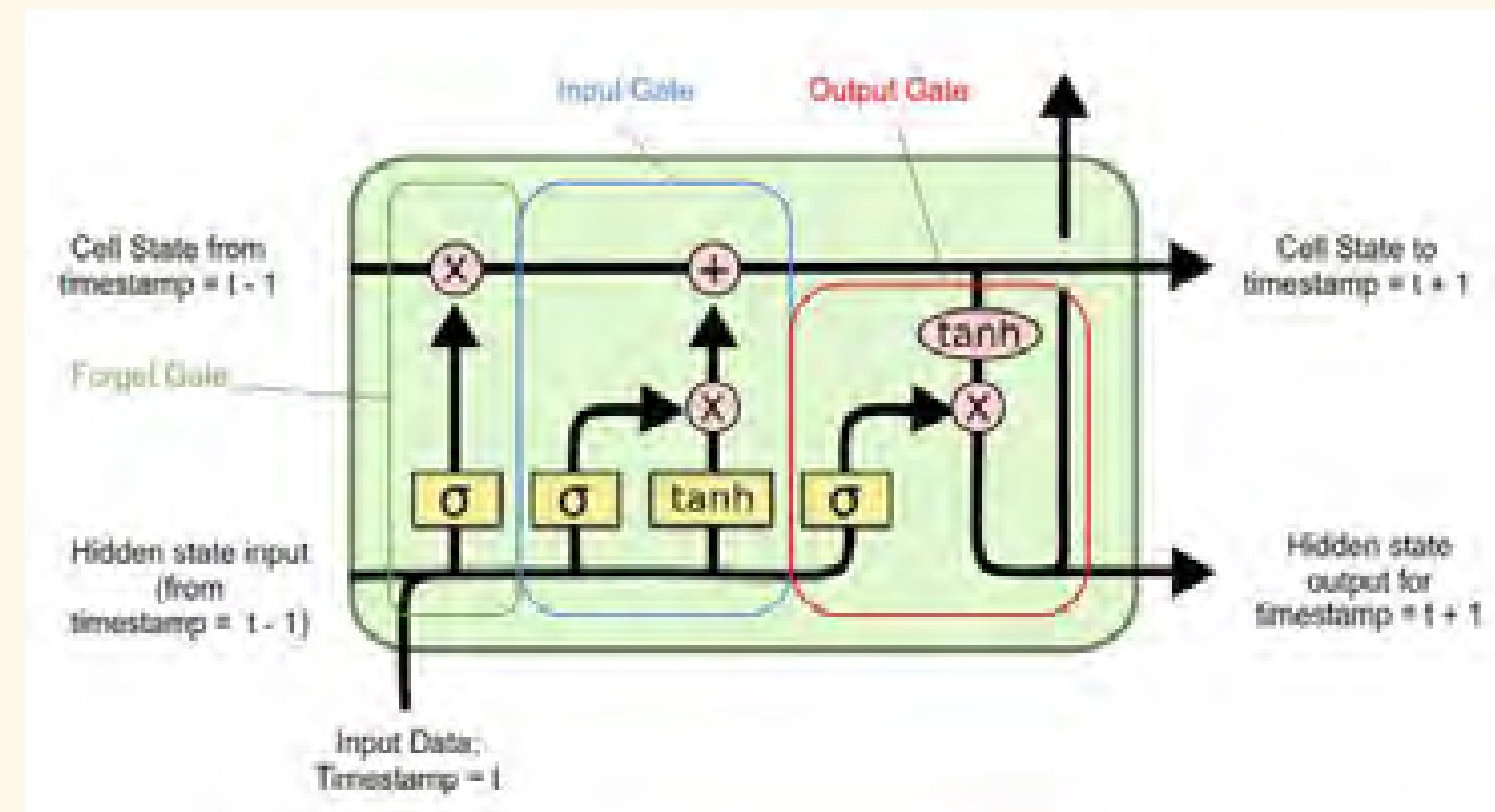
Sumber Gambar: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

- RNN merupakan salah satu jenis Neural Network dengan loop di dalamnya, memungkinkan input di waktu ke- t mempengaruhi input di $t+1$.
- RNN bekerja baik pada input berjenis urutan (sekuens).
- **Permasalahan:** Jika sekuens yang diberikan panjang, RNN cenderung "melupakan" informasi (input) yang sudah lama diberikan. (Vanishing Gradient Problem)

- LSTM merupakan modifikasi dari arsitektur RNN yang mampu "mengingat" informasi jangka panjang.
- LSTM dapat memilih informasi yang harus dibuang dan informasi yang akan tetap disimpan dalam "cell states".
- Terdapat tiga struktur pada LSTM yang mempengaruhi cell states, yaitu forget gate, input gate, dan output gate.

LSTM

Long Short Term Memory

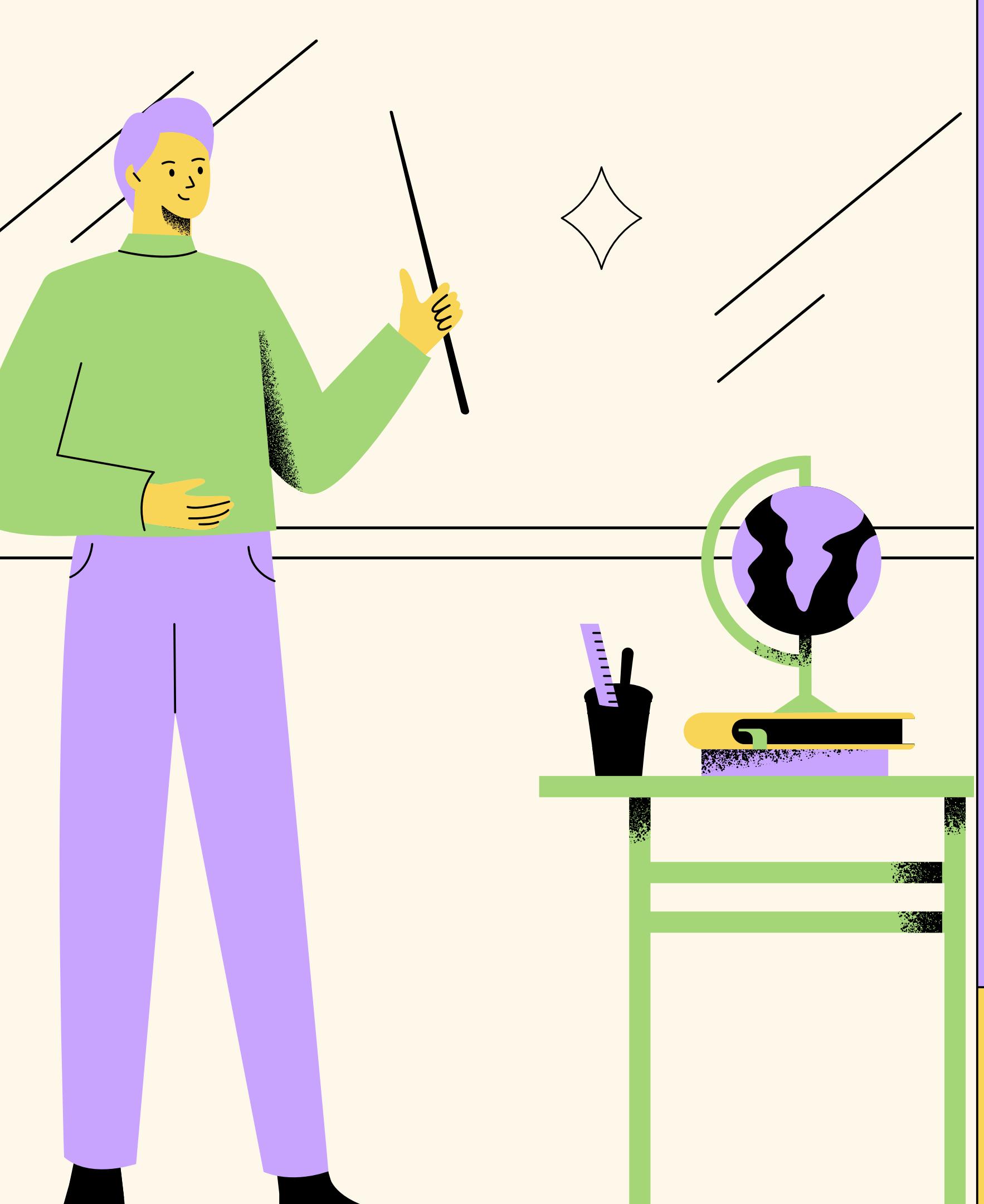


Sumber Gambar: Ayu W. 2022. Analisis Kinerja Algoritma CNN dan LSTM untuk Memprediksi Tinggi Muka Air di DKI Jakarta, Indonesia. Fakultas Ilmu Komputer. Universitas Mercu Buana: Jakarta

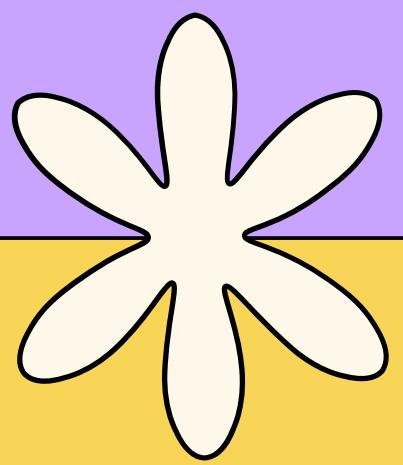
LSTM

Long Short Term Memory

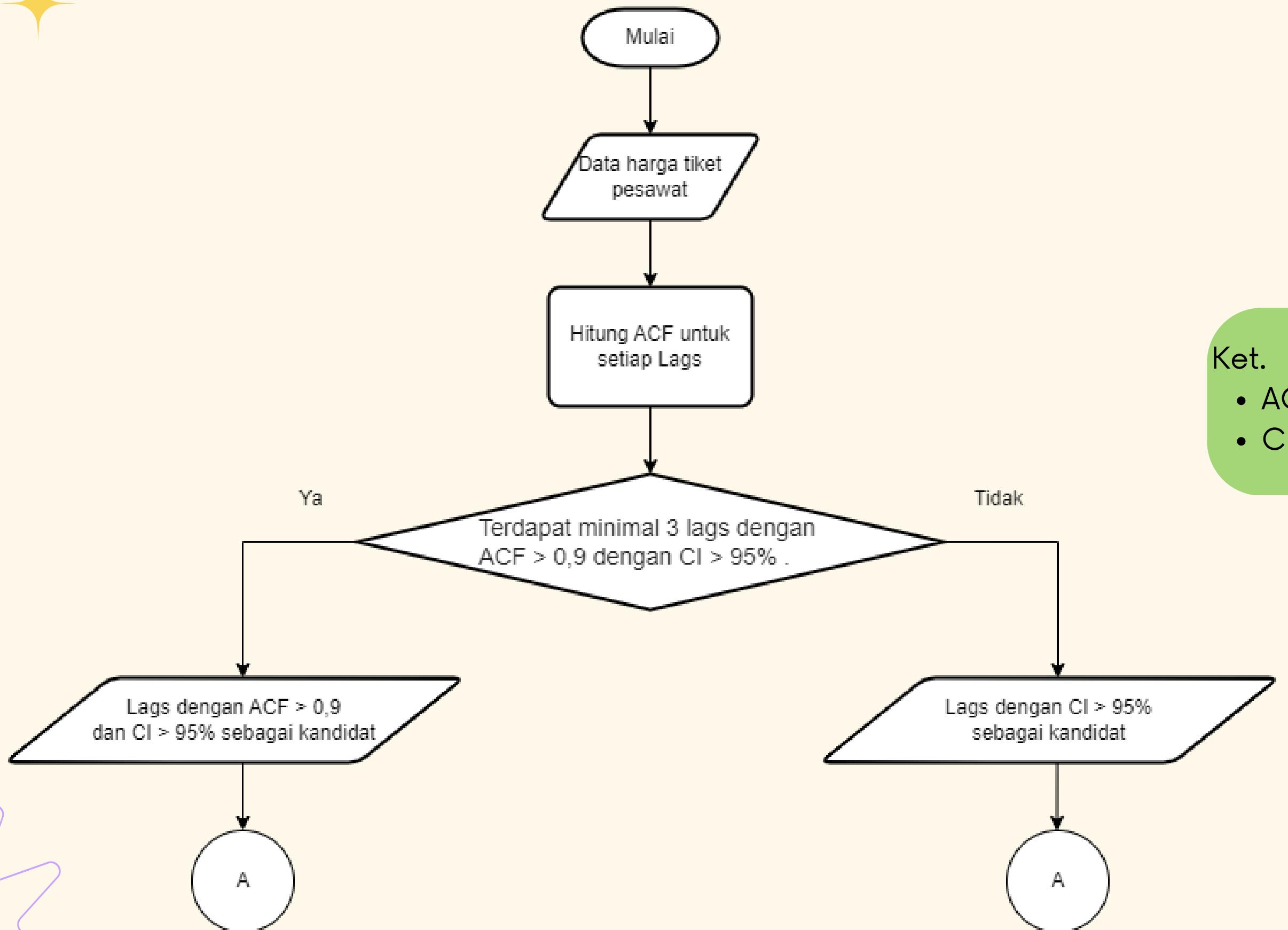




Implementasi



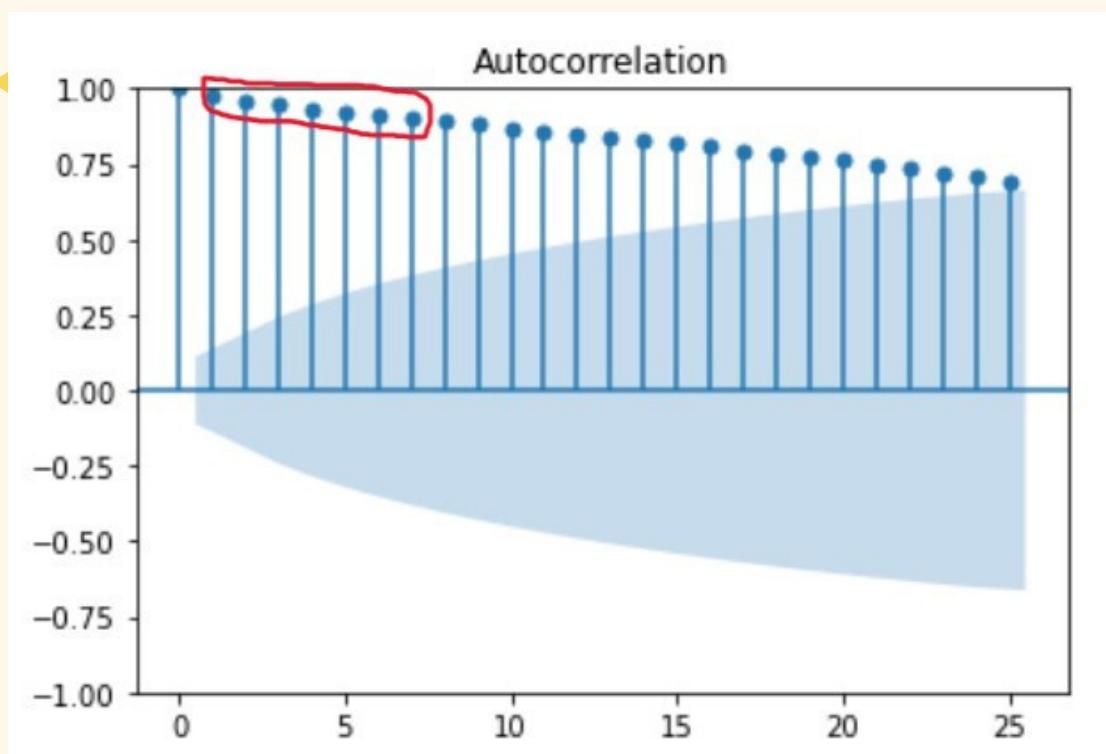
Proses Pencarian Kandidat Lags



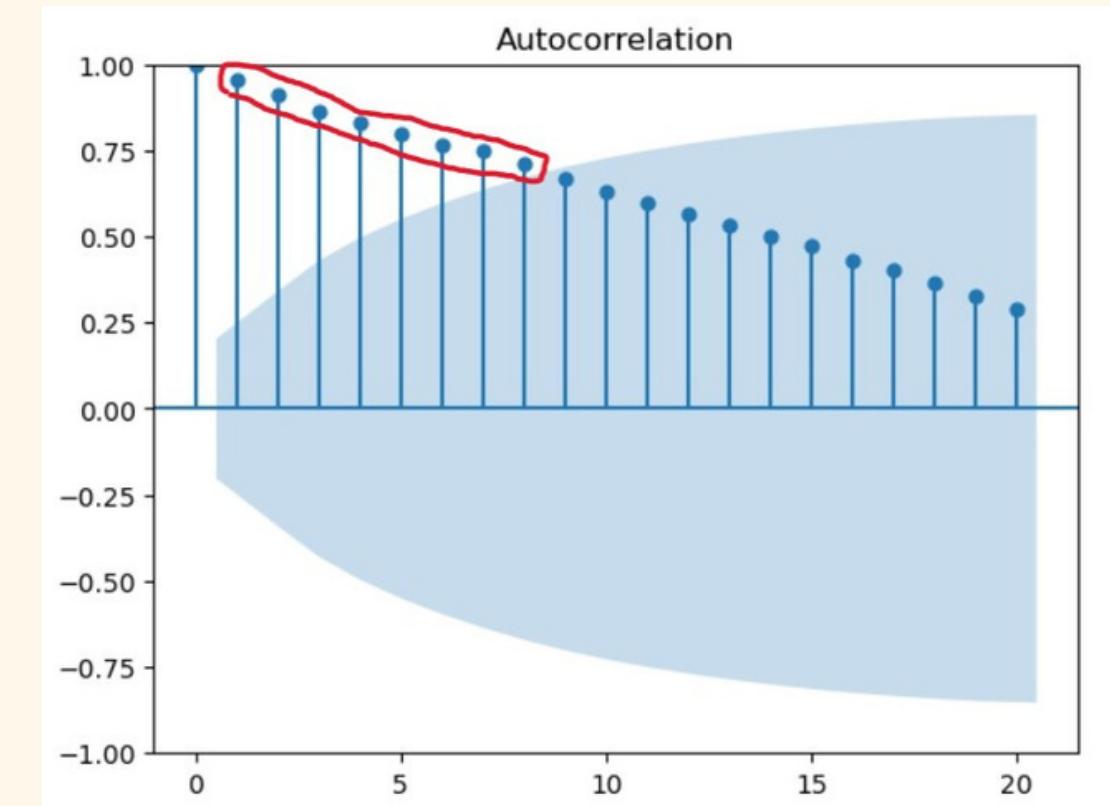
Ket.

- ACF: Auto Correlation Function
- CI: Confidence Interval

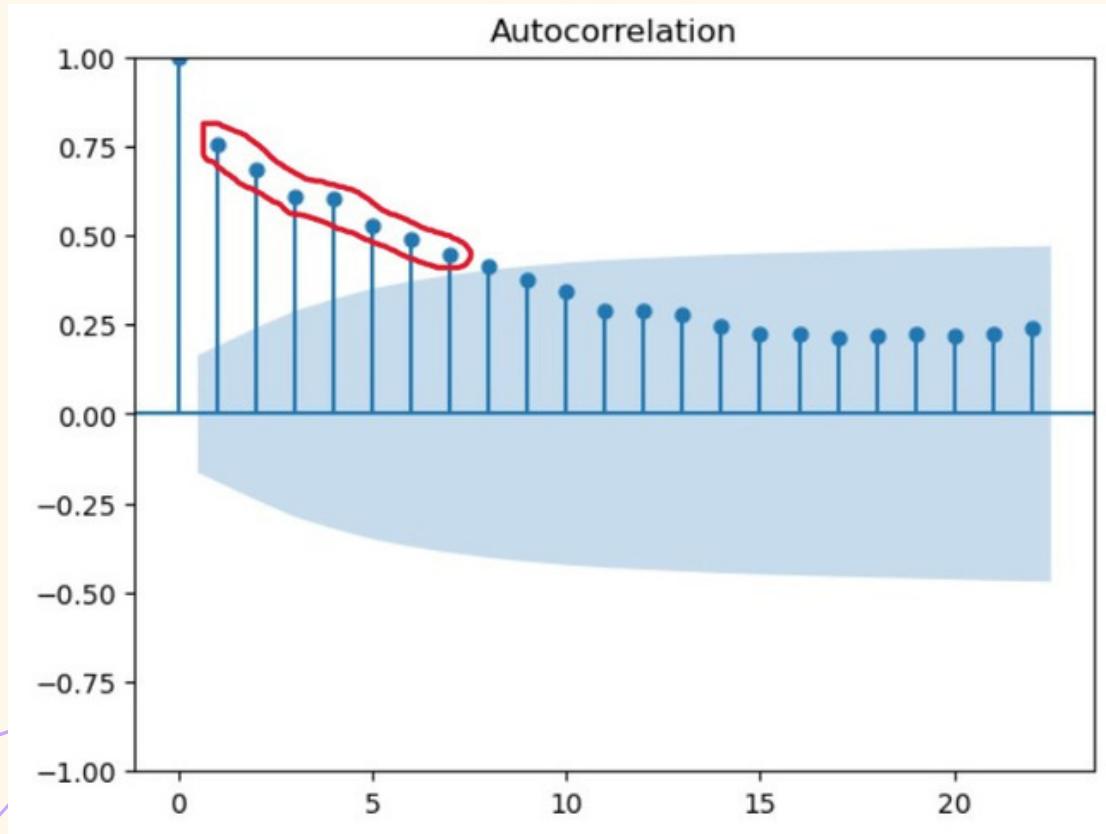
Plot Autokorelasi rute CGK-DPS



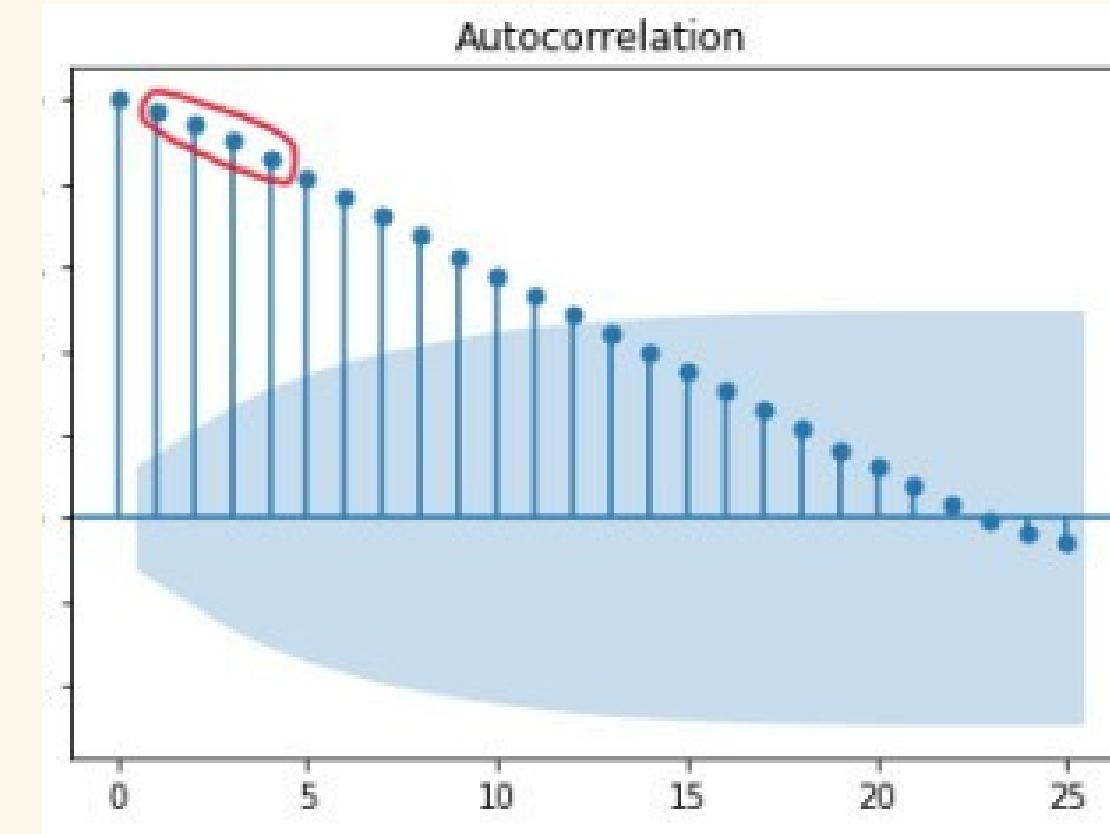
Plot Autokorelasi rute KUL-CGK



Plot Autokorelasi rute DPS-CGK



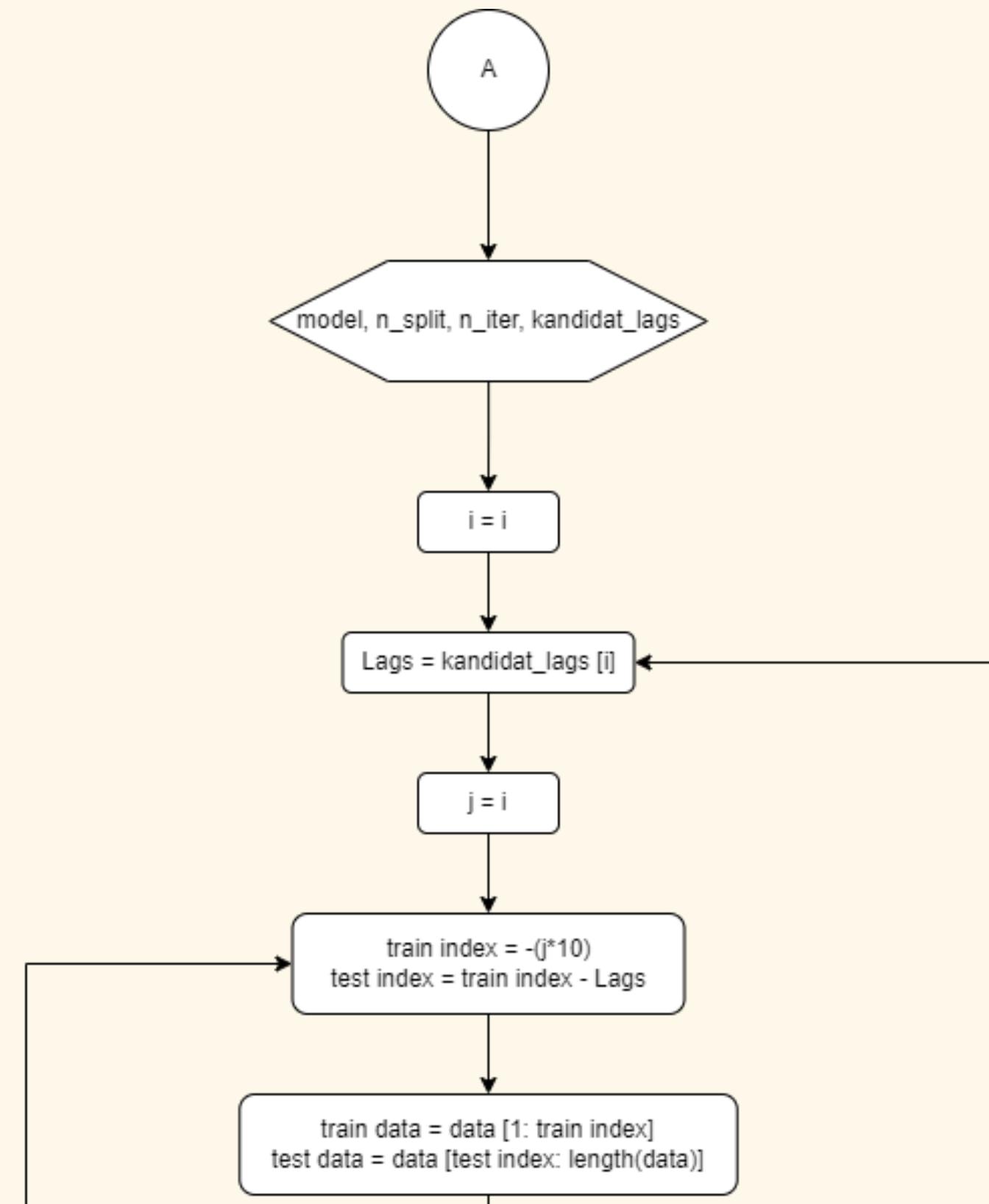
Plot Autokorelasi rute CGK-KUL



Proses Pencarian Lags dan Komposisi Split Data Train dan Test Teroptimasi

Ket.

- **n_split**: banyaknya kandidat komposisi split data
- **kandidat_lags**: list yang berisi kandidat lags
- **n_iter**: banyaknya model yang dilakukan training
- **i**: index loop untuk kandidat_lags
- **j**: index loop untuk n_split

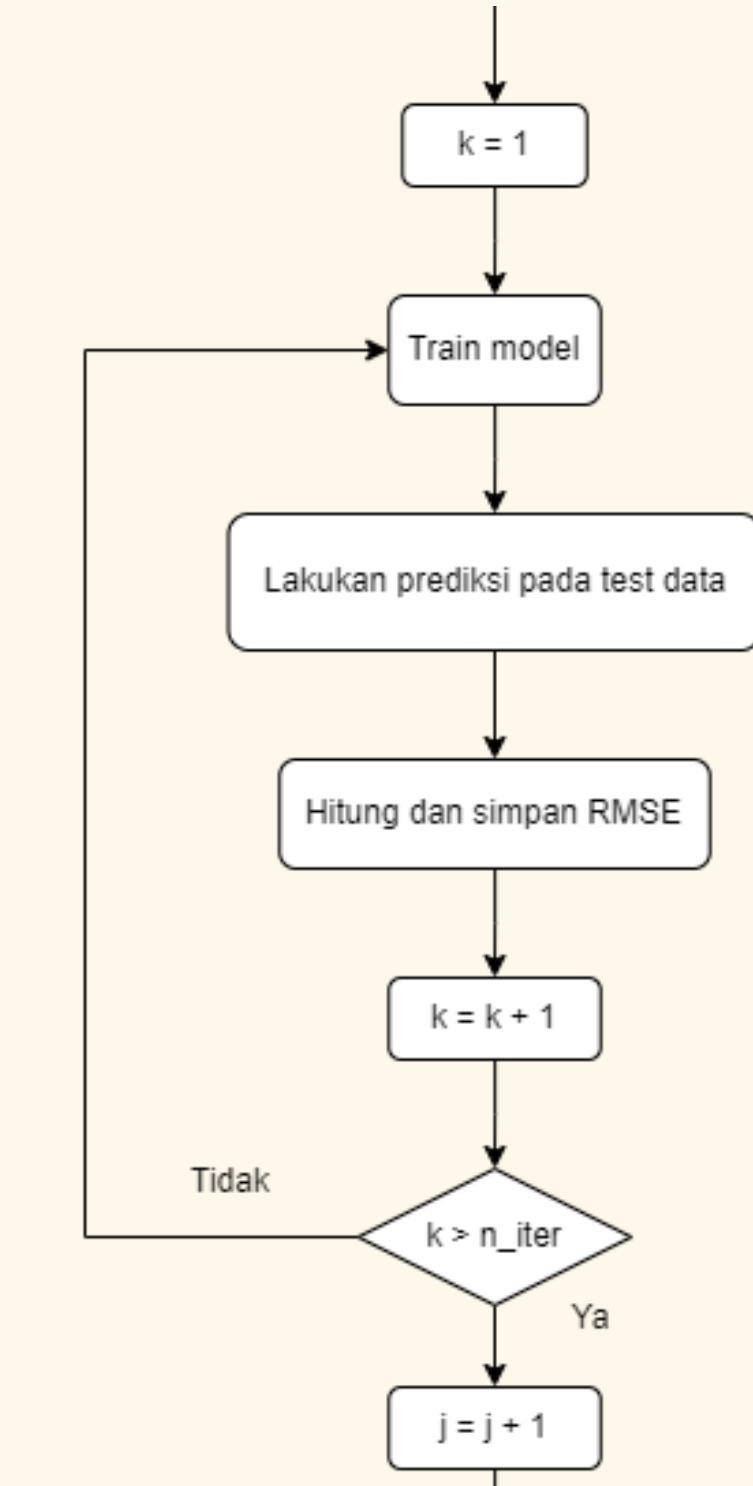


Contoh **n_split** = 3, **lags** = 5:

- jumlah observasi: 130
- Maka:
 - data train = `data[1:-30]`
 - data test = `data[-35:130]`

Proses Pencarian Lags dan Komposisi

Split Data Train dan Test Teroptimasi

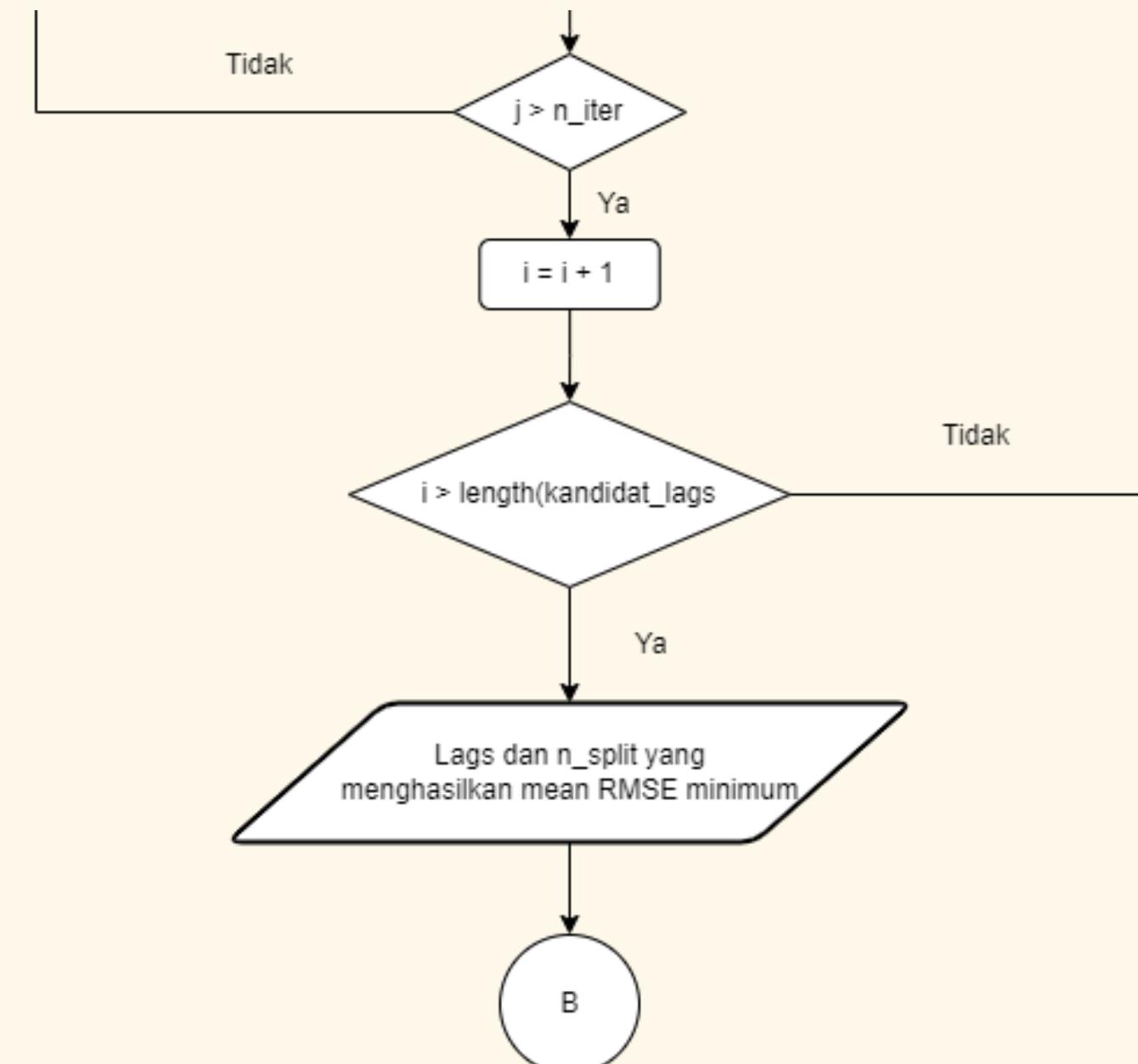


Ket.

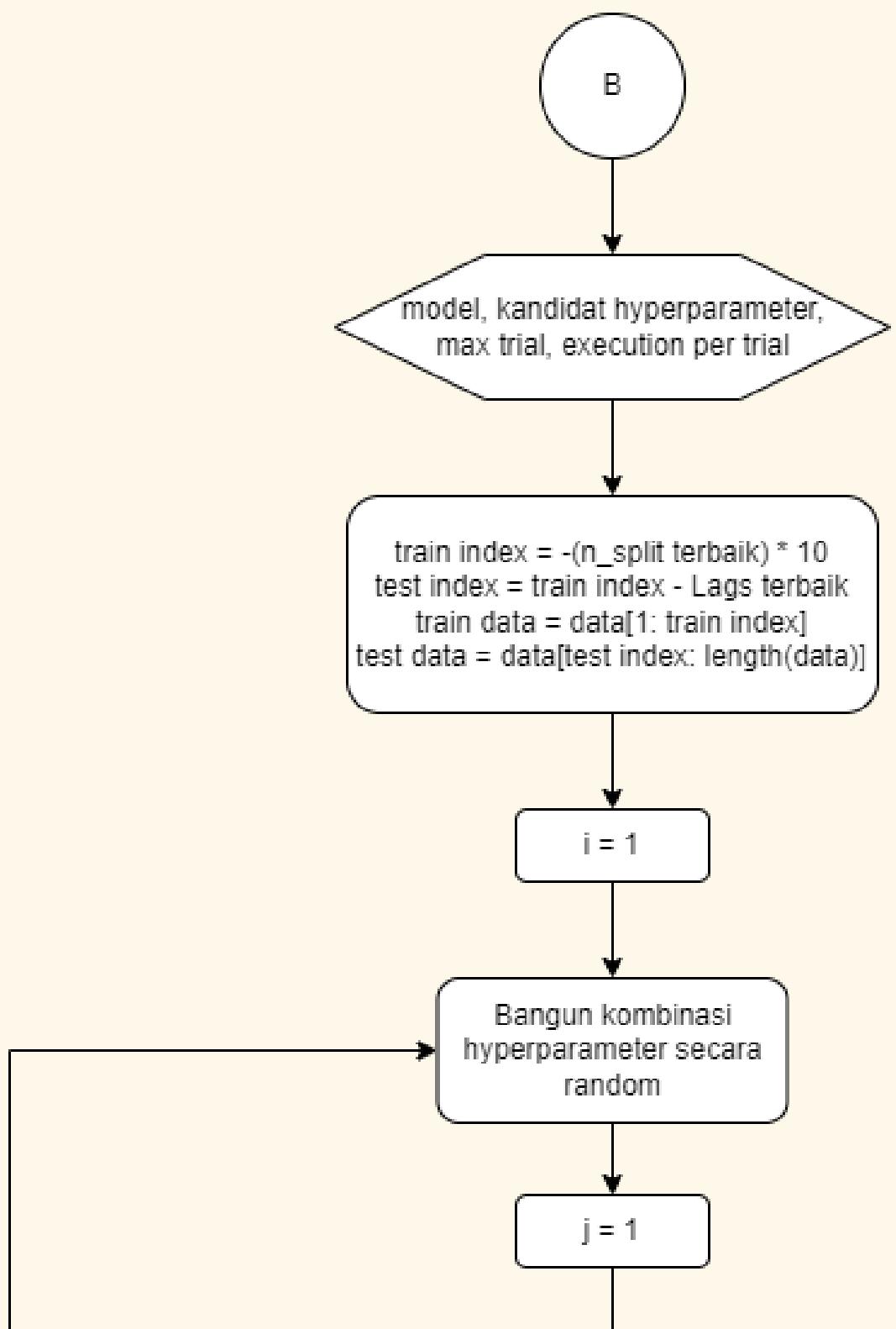
- k : index loop untuk n_{iter}

Proses Pencarian Lags dan Komposisi

Split Data Train dan Test Teroptimasi



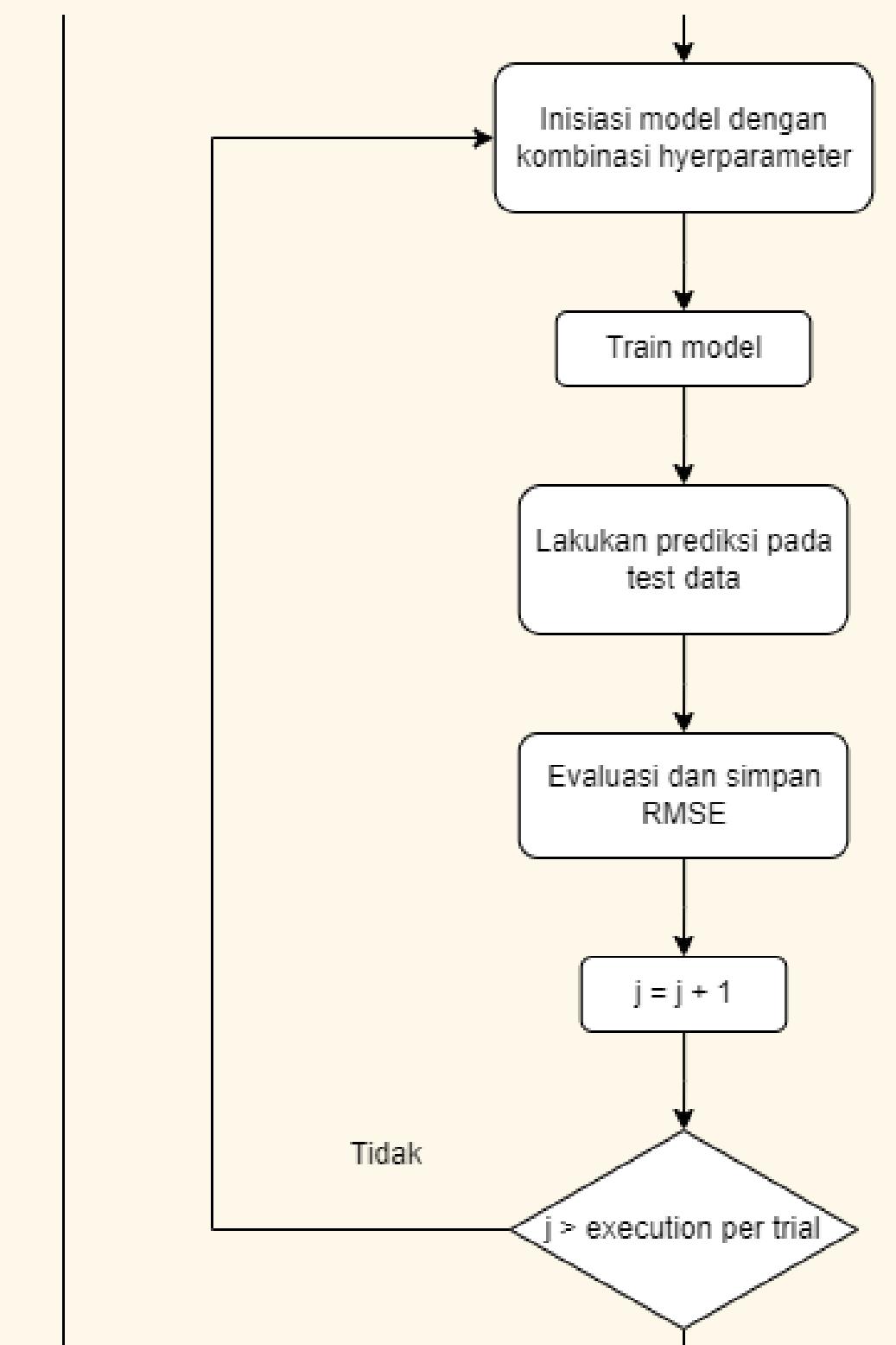
Proses Hyperparameter Tuning (Random Search)



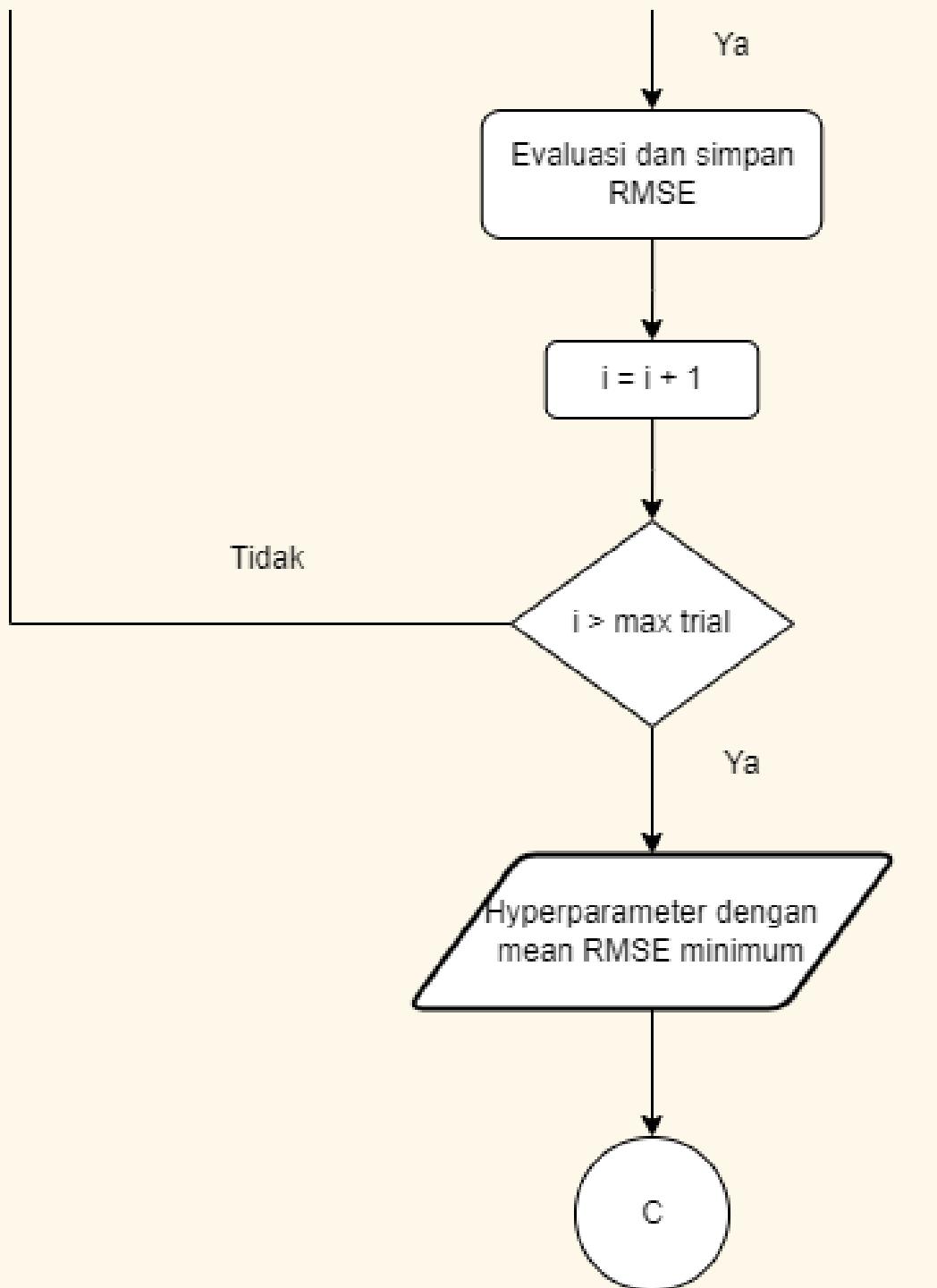
Ket.

- **max trial:** jumlah iterasi pada random search
- **execution per trial:** banyaknya model yang dilakukan training untuk setiap kombinasi hyperparameter

Proses Hyperparameter Tuning (Random Search)



Proses Hyperparameter Tuning (Random Search)



Kandidat Hyperparameter XGBoost

• Maksimum kedalaman tree	Rentang nilai: [3, 10]	step size =1
• Minimum bobot tree	Rentang nilai: [2,6]	step size =1
• Learning rate	Rentang nilai: [0.1, 0.3]	step size = 0.1
• Gamma	Rentang nilai: [0.01, 0.2]	step size = 0.01
• Nilai Regularisasi L1 (Alpha)	Rentang nilai: [0,1]	
• Nilai Regularisasi L2 (Lambda)	Rentang nilai: [0,1]	
• Subsample	Rentang nilai: (0,1]	step size = 0.25
• Cosample by Tree	Rentang nilai: (0,1]	step size = 0.25
• Cosample by Level	Rentang nilai: (0,1]	step size = 0.25

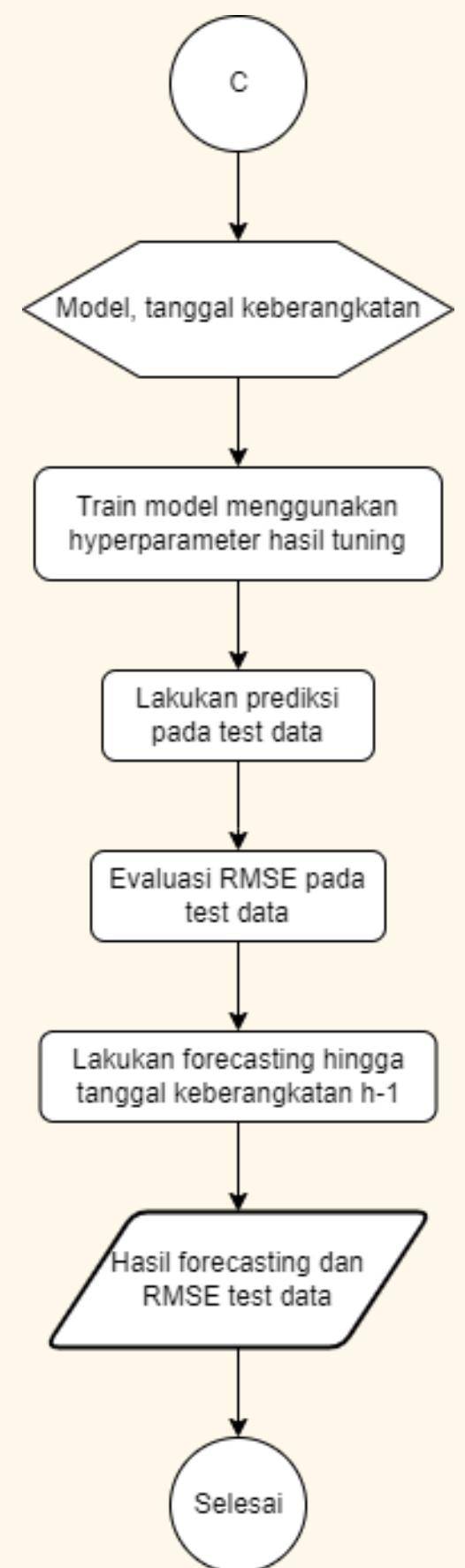
Kandidat Hyperparameter LSTM

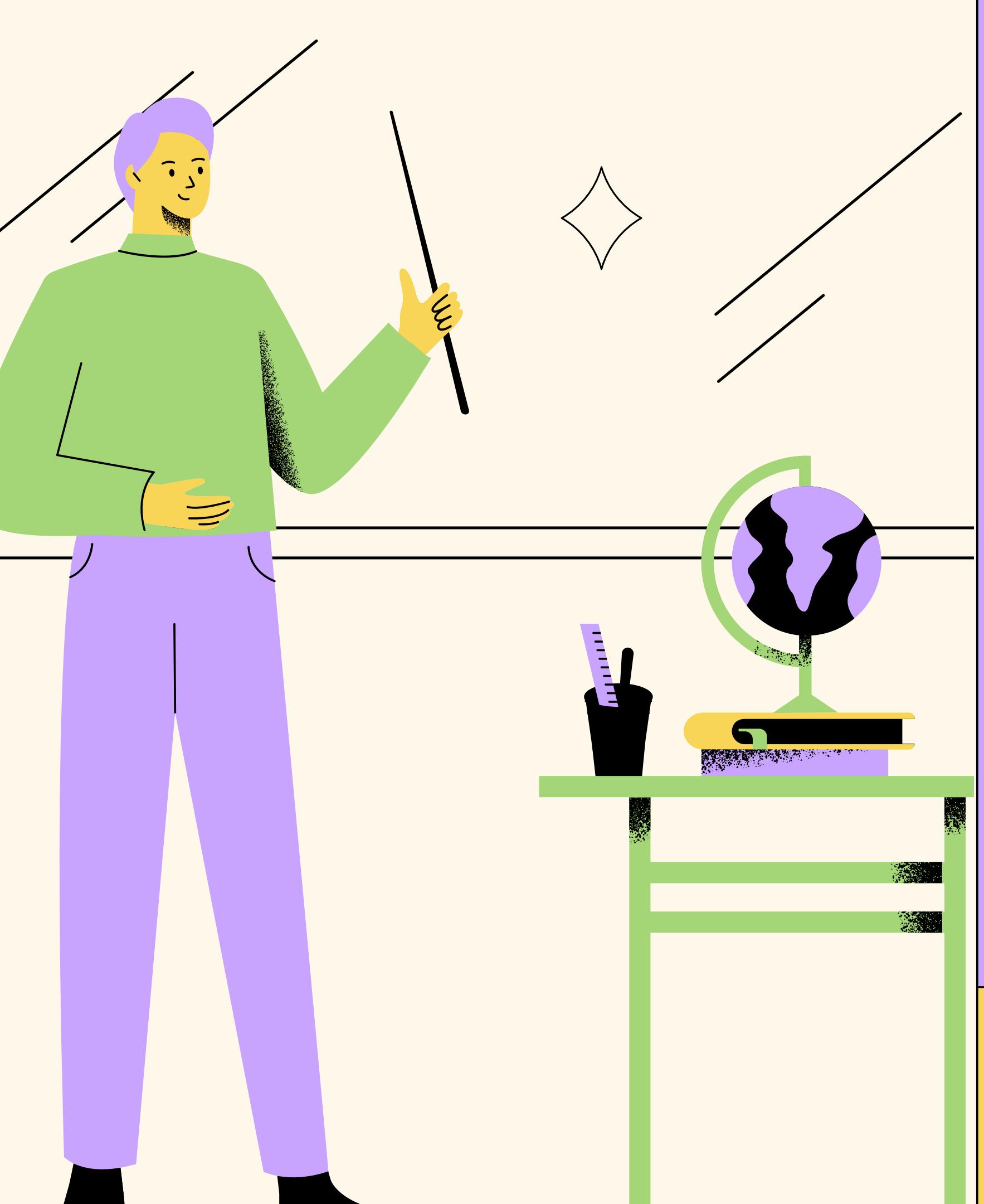
- Banyaknya hidden layer LSTM
- Banyaknya neuron pada layer LSTM
- Banyaknya hidden layer Dense
- Banyaknya neuron pada layer Dense
- Learning rate
- Activation function
- Max epoch
- Batch size

rentang nilai [1, 3]
rentang nilai [8, 32]
rentang nilai [0, 2]
rentang nilai [8, 32]
rentang nilai [0.0001, 0.1]
['relu', 'tanh', 'sigmoid']
1000
32

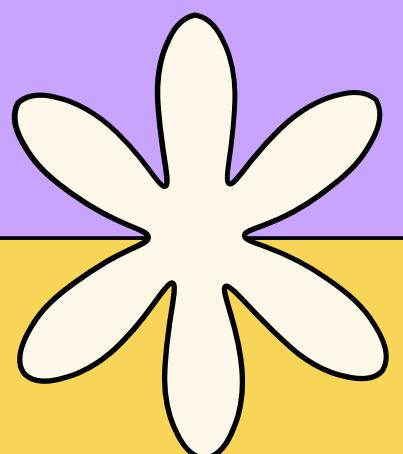
step size = 1
step size = 8
step size = 1
step size = 8

Proses Training, Testing, dan Forecasting



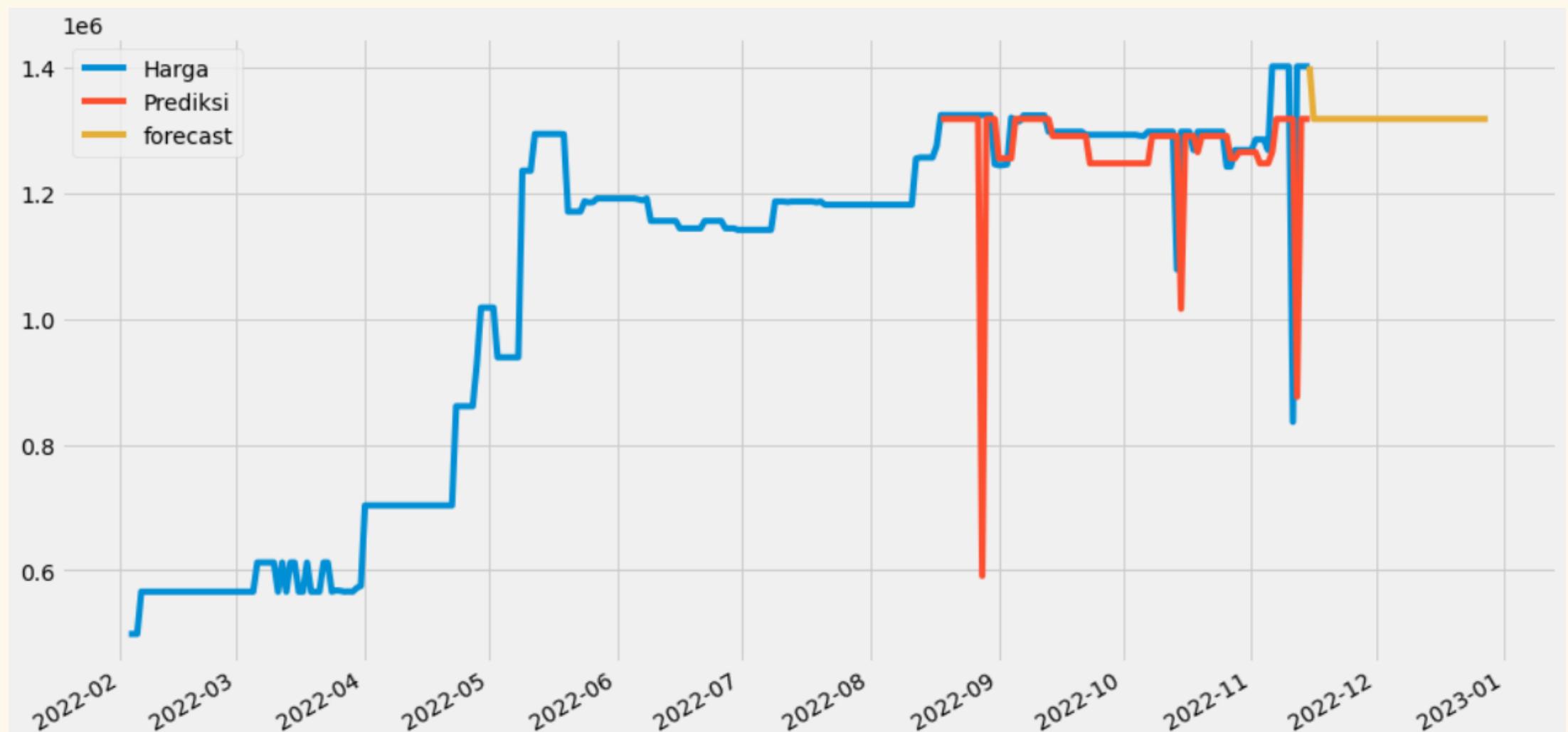


Analisis dan Hasil



Hasil XGBoost

CGK-DPS

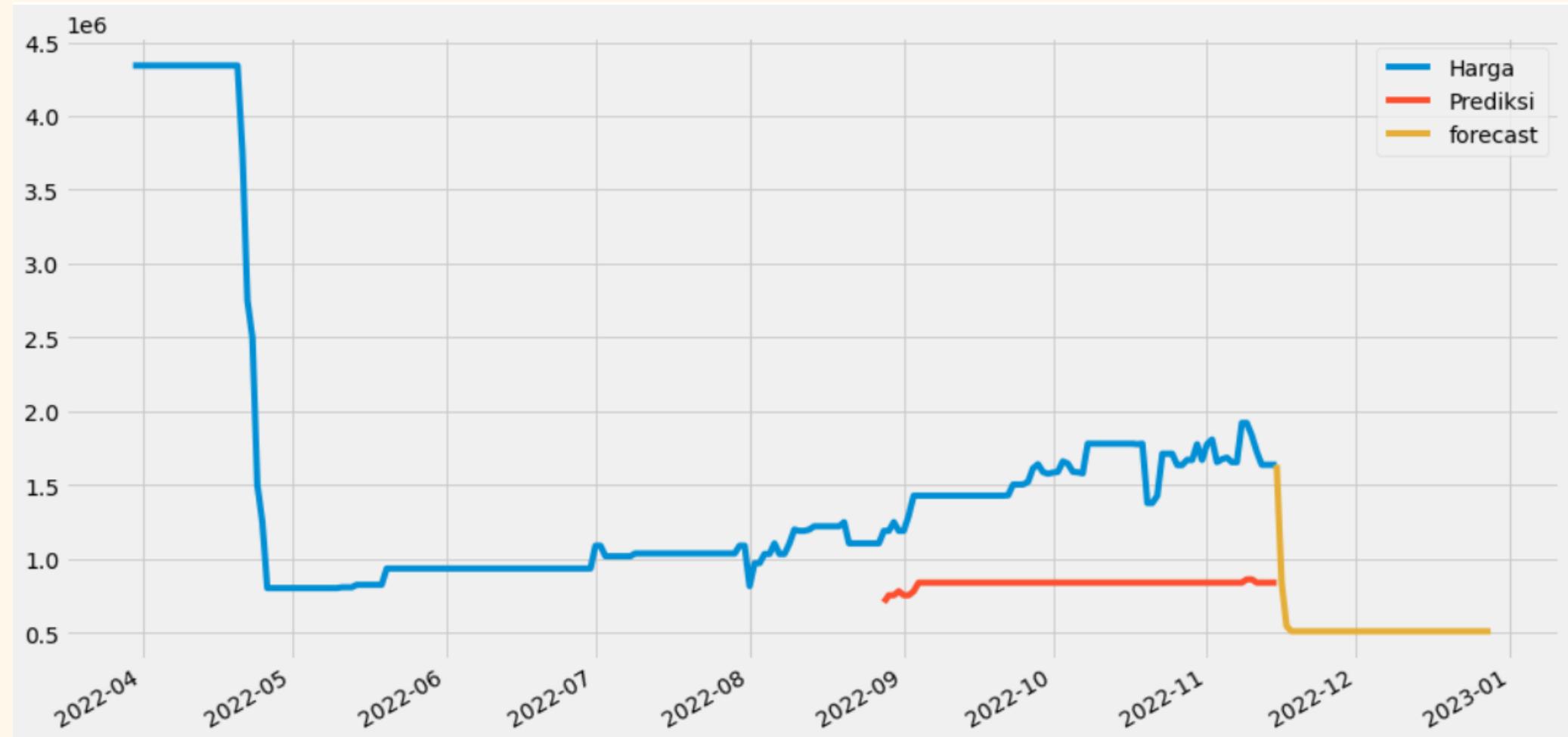


Jumlah data test: 80
Lags terpilih: 1

- Parameter terpilih:
 - Maksimum kedalaman tree: 3
 - Minimum bobot tree: 2
 - Learning rate: 0.3
 - Gamma: 0.1
 - Nilai Regularisasi L1 (Alpha): 0.0001
 - Nilai Regularisasi L2 (Lambda): 0.01
 - Subsample: 0.5
 - Cosample by Tree: 0.5
 - Cosample by Level: 0.5

Hasil XGBoost

CGK-KUL

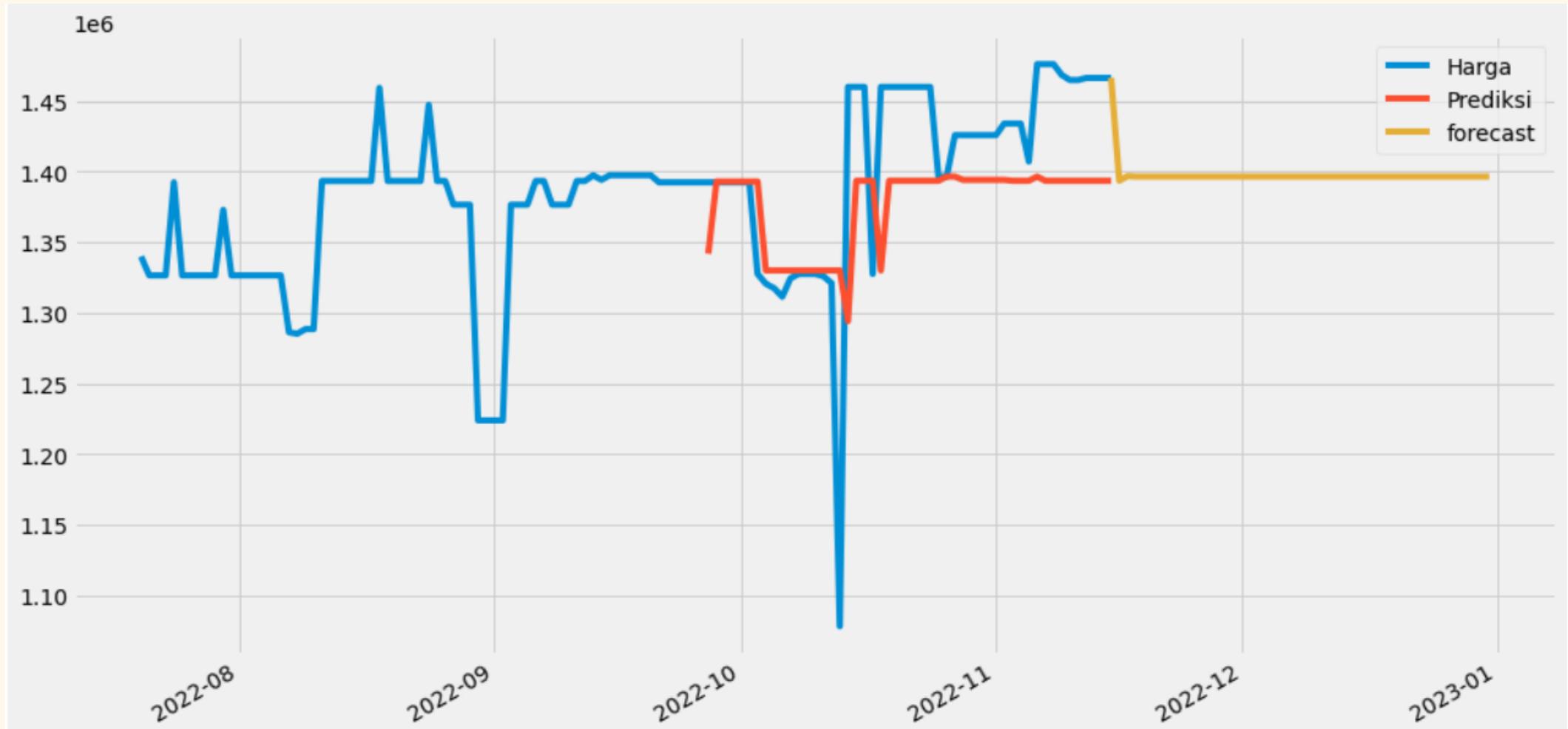


Jumlah data test: 90
Lags terpilih: 1

- Parameter terpilih:
 - Maksimum kedalaman tree: 6
 - Minimum bobot tree: 3
 - Learning rate: 0.1
 - Gamma: 0.2
 - Nilai Regularisasi L1 (Alpha): 0.0001
 - Nilai Regularisasi L2 (Lambda): 0.001
 - Subsample: 0.5
 - Cosample by Tree: 0.5
 - Cosample by Level: 0.5

Hasil XGBoost

DPS-CGK

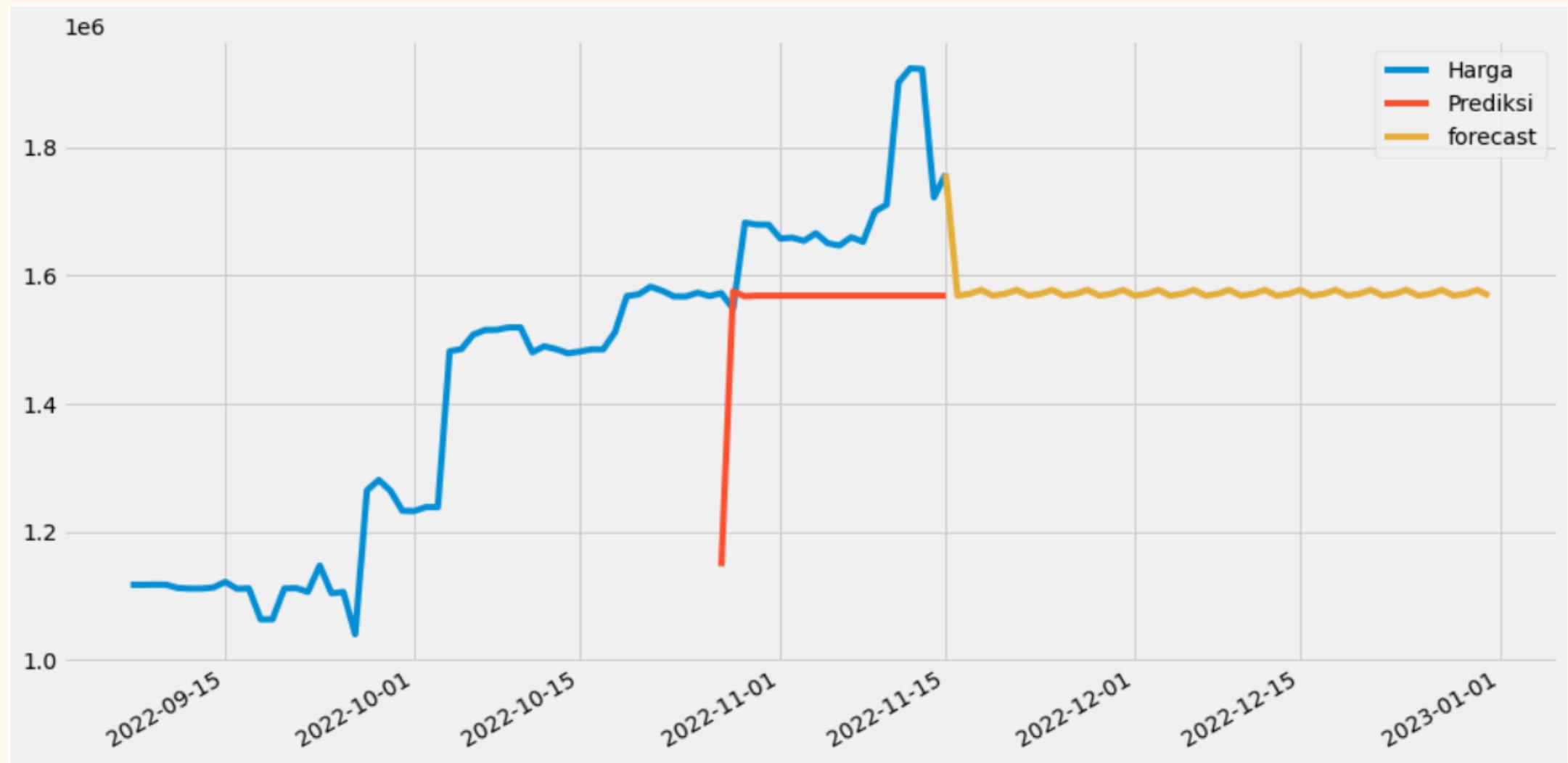


Jumlah data test: 50
Lags terpilih: 1

- Parameter terpilih:
 - Maksimum kedalaman tree: 9
 - Minimum bobot tree: 2
 - Learning rate: 0.3
 - Gamma: 0.2
 - Nilai Regularisasi L1 (Alpha): 0.01
 - Nilai Regularisasi L2 (Lambda): 0.0001
 - Subsample: 0.5
 - Cosample by Tree: 0.5
 - Cosample by Level: 0.5

Hasil XGBoost

KUL-CGK

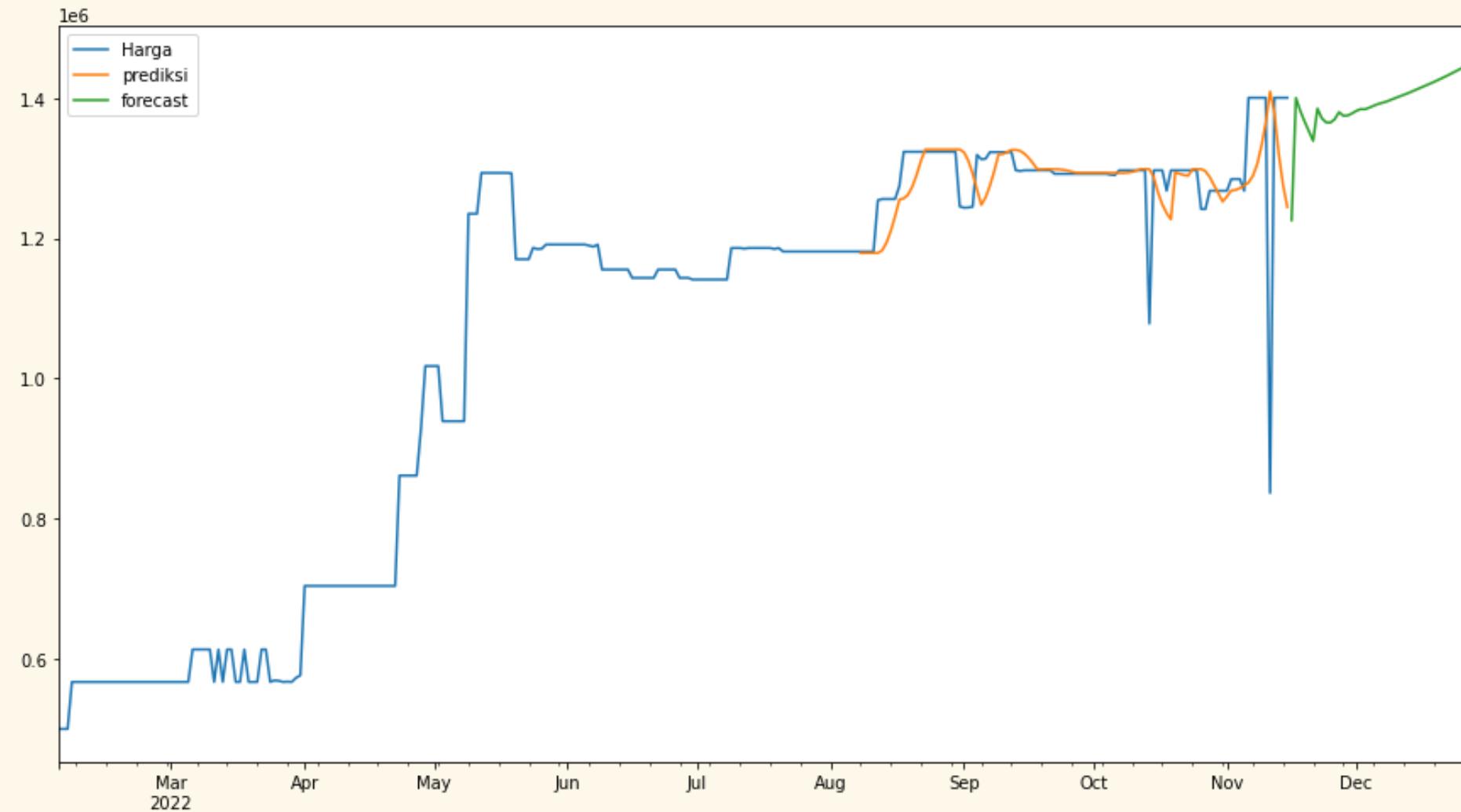


Jumlah data test: 20
Lags terpilih: 1

- Parameter terpilih:
 - Maksimum kedalaman tree: 7
 - Minimum bobot tree: 2
 - Learning rate: 0.1
 - Gamma: 0.015
 - Nilai Regularisasi L1 (Alpha): 1e-05
 - Nilai Regularisasi L2 (Lambda): 0.01
 - Subsample: 0.5
 - Cosample by Tree: 0.5
 - Cosample by Level: 0.5

Hasil LSTM

CGK-DPS

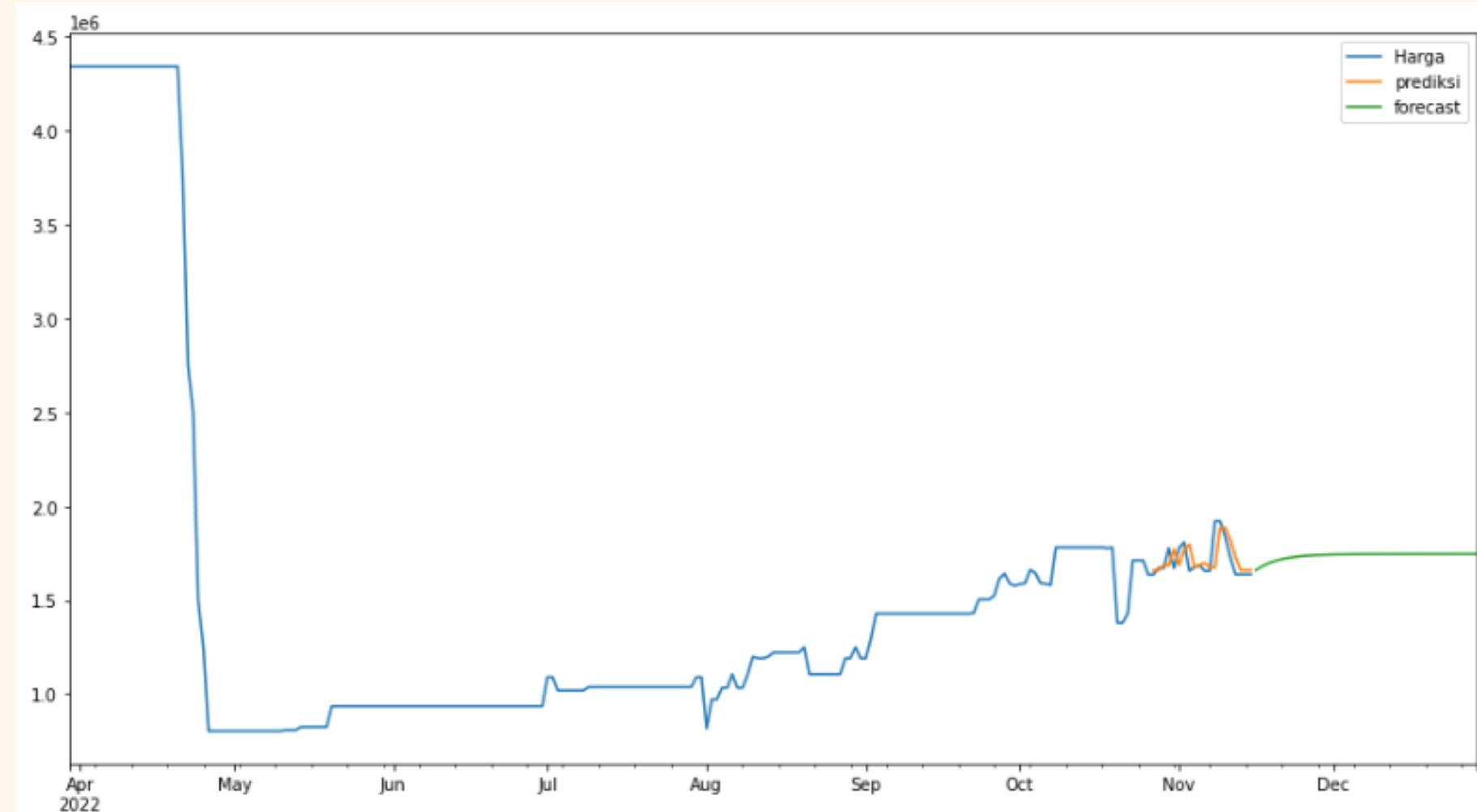


Jumlah data test: 100
Lags terpilih: 5

- Banyaknya hidden layer LSTM: 1
- Banyaknya neuron pada layer LSTM: 128
- Banyaknya hidden layer Dense: 0
- Banyaknya neuron pada layer Dense: 0
- Learning rate: 0.01
- Activation function: relu
- Max epoch: 126 (early stopping)
- Batch size: 32

Hasil LSTM

CGK-KUL

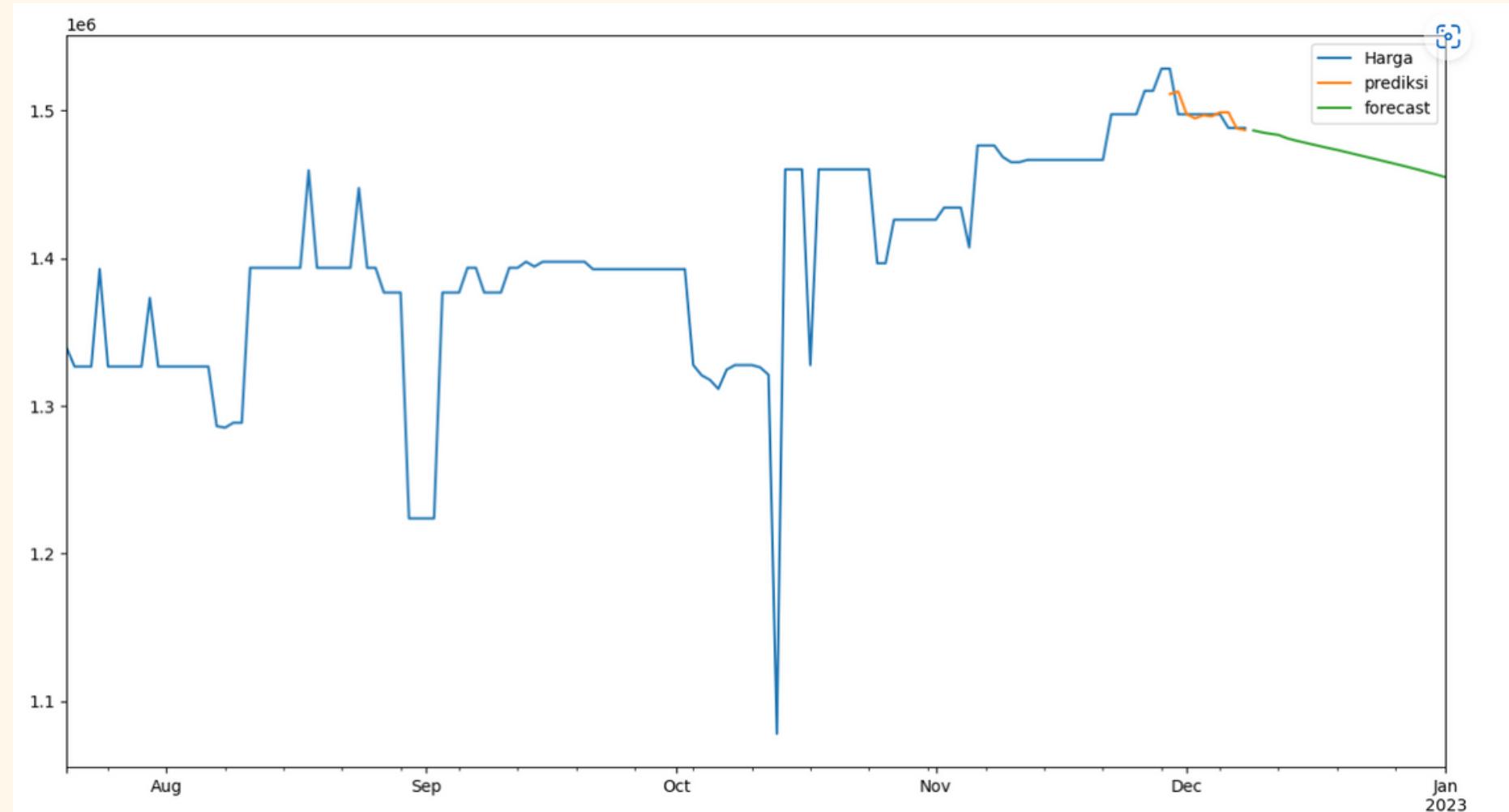


- Banyaknya hidden layer LSTM: 1
- Banyaknya neuron pada layer LSTM: 72
- Banyaknya hidden layer Dense: 0
- Banyaknya neuron pada layer Dense: 0
- Learning rate: 0.1
- Activation function: relu
- Max epoch: 161 (early stopping)
- Batch size: 32

Jumlah data test: 20
Lags terpilih: 1

Hasil LSTM

DPS-CGK

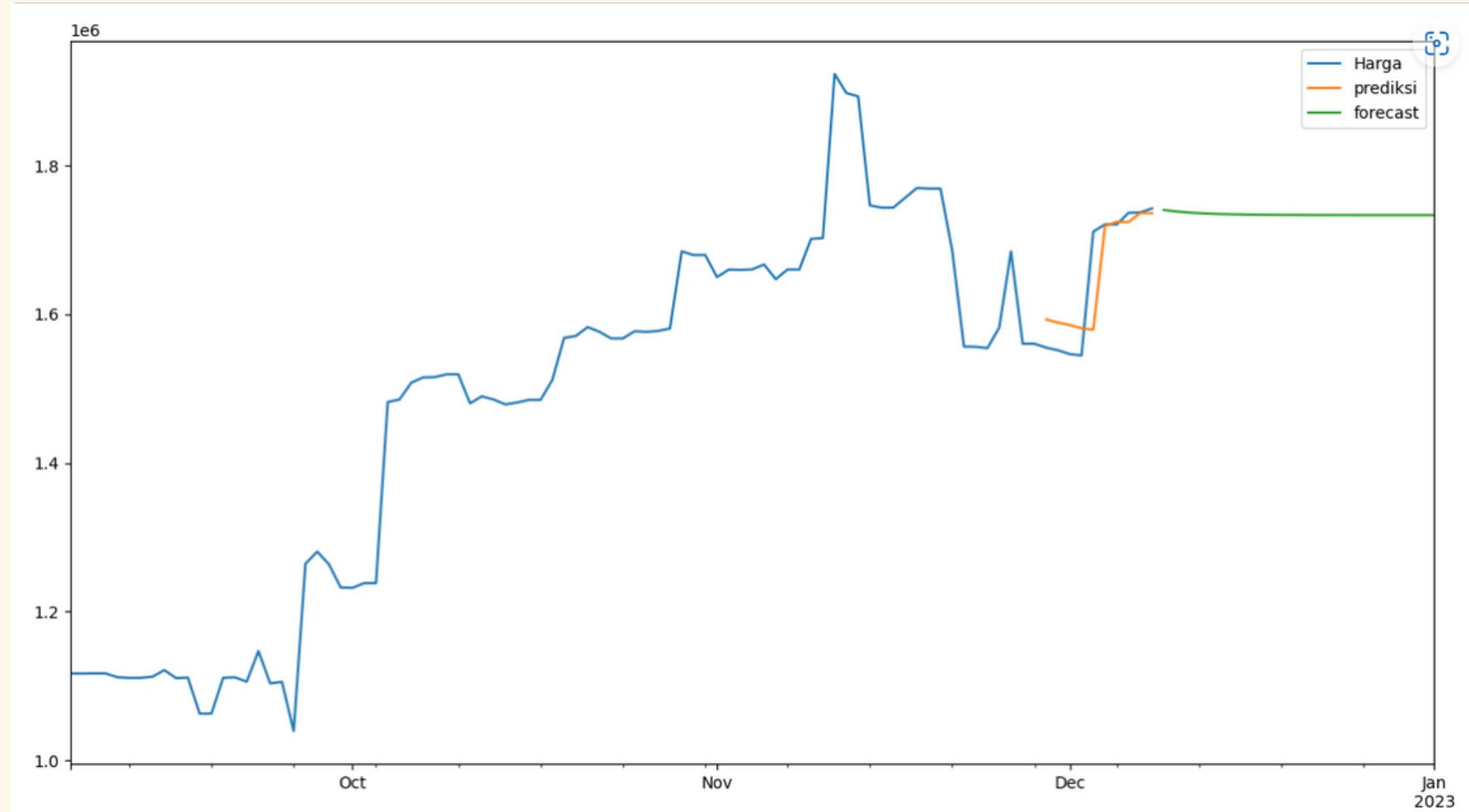


- Banyaknya hidden layer LSTM: 1
- Banyaknya neuron pada layer LSTM: 128
- Banyaknya hidden layer Dense: 0
- Banyaknya neuron pada layer Dense: 0
- Learning rate: 0.1
- Activation function: relu
- Max epoch: 433 (early stopping)
- Batch size: 32

Jumlah data test: 10
Lags terpilih: 7

Hasil LSTM

KUL-CGK

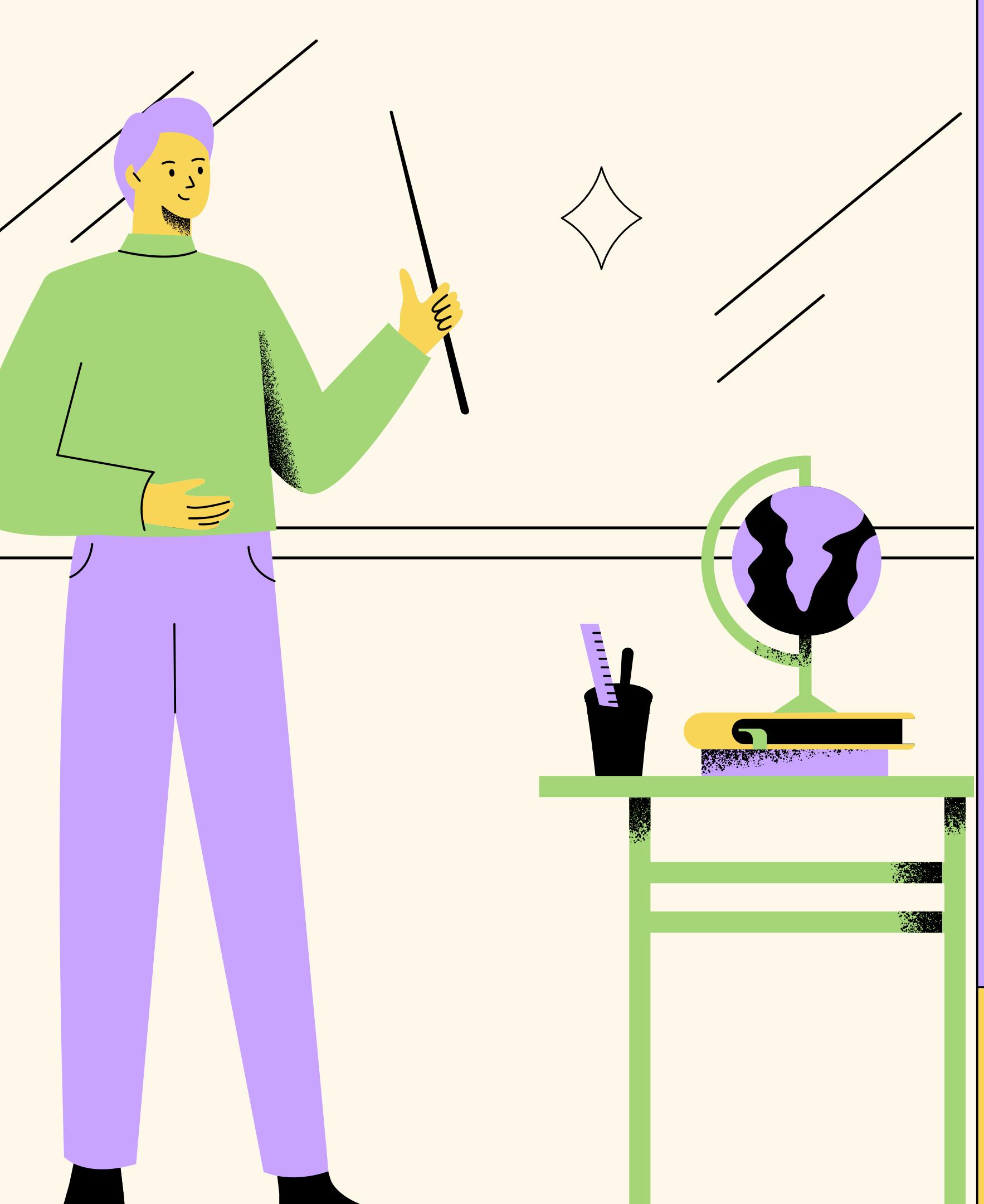


- Banyaknya hidden layer LSTM: 1
- Banyaknya neuron pada layer LSTM: 64
- Banyaknya hidden layer Dense: 0
- Banyaknya neuron pada layer Dense: 0
- Learning rate: 0.1
- Activation function: relu
- Max epoch: 307 (early stopping)
- Batch size: 32

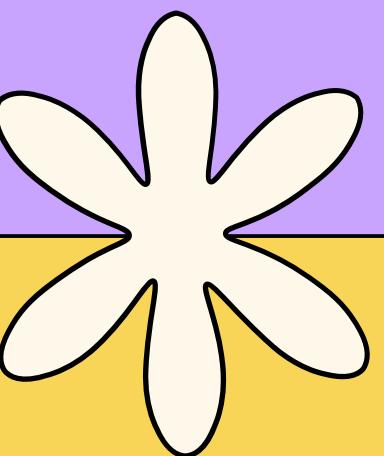
Jumlah data test: 10
Lags terpilih: 2

Tabel Perbandingan RMSE

Rute	XGBoost	LSTM
CGK - DPS	126,742.21724	74,109.12737
DPS - CGK	66,786.66356	8,124.69946
CGK - KUL	746,201.37710	81,960.325403
KUL - CGK	192,572.62623	48,235.89157



Kesimpulan



1. Model LSTM lebih baik dibandingkan model XGBoost dalam memprediksi harga tiket pesawat berdasarkan nilai RMSE nya.
2. Berdasarkan hasil forecasting, kedua model menunjukkan bahwa semakin mendekati hari keberangkatan, harga tiket pesawat cenderung tidak menurun. Oleh karena itu, harga tiket pesawat termurah tidak dapat ditentukan dari hasil forecasting untuk keempat data yang diambil
3. Berdasarkan data histori yang diambil, harga tiket pesawat termurah tidak dapat ditentukan karena masing-masing rute memiliki rentang waktu yang berbeda dalam mencapai harga termurahnya

Daftar Pustaka

Ayu W. 2022. Analisis Kinerja Algoritma CNN dan LSTM untuk Memprediksi Tinggi Muka Air di DKI Jakarta, Indonesia. Fakultas Ilmu Komputer. Universitas Mercu Buana: Jakarta

Chen & Guestrin, 2016. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

lah, Christopher.2015. Understanding LSTM Networks.<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Google.com/flights, 15 November 2022, <https://www.google.com/flights?tfs=CBoQAh0jEgoyMDIyLTEyLTMwagwIAxIIL20vMDQ0cnZyBwgBEgNEUFNAAUgBUgNJRFJ6ZENqUklSR3RYYnpRNFdUQTRiRzlCUWxnNE1FRkNSETBOTFMwdExTMHRMWFJzWW1OaU9VRkJRVUZCUjA0d1ZFbDNTMWwwYzBGQkVnRXhHZ3NJOEw1VkBQWFBMGxFVWpnWGNLUkc&hl=id&gl=ID&curr=IDR&gsas=1&authuser=1&sa=X&ved=0CAUQ2Y8EahcKEwioouyn1bH7AhUAAAAAHQAAAAAQHA>

Daftar Pustaka

Google.com/flights, 15 November 2022, <https://www.google.com/flights?tfs=CB0QAhoeEgoyMDIzLTAxLTAyagcIARIDS1VMcgcIARIDQ0dLQAFIAVIDSURSemRDalJJZEMxTFRWRk9hWFpYVDFGQlEwTXpRbWRDUnkwdExTMHRMUzBOTFMxMGFHSnROMEZCUVVGQlIwNHdWR05KVGtWMVkwRkJFZ0V5R2dzSWkvWnJFQUFhQTBsRVVqZ1hjTnhZ&hl=id&gl=ID&curr=IDR&gsas=1&authuser=1&sa=X&ved=0CBQQ2Y8EahcKEwjA79671rH7AhUAAAAAHQAAAAAQHA>

Google.com/flights, 15 November 2022, https://www.google.com/travel/flights/search?tfs=CBwQAhojagwIAxIIL20vMDQ0cnYSCjIwMjItMTItMzByBwgBEgNLVUxwAYIBCwj_____8BQAfIAZgBAg&tfu=CmRDalJJTjjJMVP6bFFUbHBIZWtGQlFtTnJZMUZDUnkwdExTMHRMUzBOTFMxMGJIQnNOVUZCUVVGQlIwNHdWRkpqUlhoWk9FRkJFZ0V6R2dzSTl2bGpFQUFhQTBsRVVqZ1hjSlZT

Google.com/flight, 15 November 2022, https://www.google.com/travel/flights/search?tfs=CBwQAhojagcIARIDFBTEgoyMDIzLTAxLTAycgwIAxIIL20vMDQ0cnZwAYIBCwj_____8BQAfIAZgBAg&tfu=CmRDalJJU25OV2VHbDJNbU40T1UxQlEweEpOa0ZDUnkwdExTMHRMUzBOZEdeoVkzUXhPRUZCUVVGQlIwNHdWR3N3VFMxR1lrRkJFZ0V3R2dzSXlzQlpFQUFhQTBsRVVqZ1hjTXBK

Terima Kasih