

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

Cornelius Justin Satryo Hadi, Lutfia Maulidina, Muhammad Habil Amardias, Muhammad Rafly Januar, Rachel Octaviani Putri, Safira Raissa Rahmi

Dibimbing oleh: Dr. Dipo Aldila, S.Si. M.Si.

Departemen Matematika, Universitas Indonesia, Depok 16424, Indonesia

Dikumpulkan sebagai laporan akhir kuliah Pemodelan Matematis, Semester Ganjil 2022

Abstrak

Pesawat terbang merupakan mode transportasi yang digunakan untuk bepergian dari suatu tempat ke tempat lain yang memiliki jarak yang jauh. Banyaknya permintaan penggunaan pesawat terbang, maka menjadi banyaknya maskapai penerbangan dan membuat adanya variasi harga tiket pesawat. Penelitian ini akan memprediksi harga tiket pesawat dengan menggunakan algoritma LSTM dan XGBoost sampai h-1 keberangkatan untuk mendapatkan waktu terbaik membeli tiket pesawat. Dalam penelitian ini menggunakan penerbangan dengan 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. Data diambil dari aplikasi web Google Flights pada tanggal 15 November 2022.

Kata kunci: tiket pesawat, prediksi, *forecasting*, LSTM, XGBoost

1. Pendahuluan

1.1. Latar Belakang

Penjualan harga tiket pesawat berubah secara dinamis setiap harinya. Hal ini disebut dengan *Dynamic Pricing* dimana harga tiket pesawat ditentukan oleh pihak maskapai untuk mencapai keuntungan yang maksimal. Harga tiket pesawat bisa berubah secara drastis di hari setelahnya atau bahkan di hari yang sama. Hal ini bisa dipengaruhi oleh beberapa hal seperti harga kompetitor, pola pemesanan konsumen, kegiatan ternama, harga avtur, dan lain sebagainya. Fluktuasi harga ini menimbulkan kesulitan bagi konsumen untuk menentukan kapan waktu pembelian harga tiket pesawat yang tepat.

Dengan adanya prediksi harga tiket pesawat termurah akan membantu para pengguna mode transportasi pesawat terbang dalam melakukan pertimbangan untuk melakukan pembelian terhadap tiket pesawat sehingga mendapatkan harga penjualan tiket pesawat yang tidak pada tarif tinggi.

1.2. Rumusan Masalah

Pemodelan yang kami lakukan harapannya dapat menjawab premis masalah yang dirumuskan sebagai berikut:

1. Manakah algoritma terbaik antara XGBoost dan LSTM untuk memprediksi harga pembelian tiket pesawat liburan tahun baru pulang pergi CGK - DPS dan CGK - KUL?
2. Kapan waktu terbaik untuk membeli tiket pesawat liburan tahun baru pulang pergi CGK - DPS dan CGK - KUL?

1.3. 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. Memprediksi kapan waktu terbaik untuk membeli tiket pesawat untuk berlibur tahun baru pulang pergi CGK - DPS dan CGK - KUL

1.4. Batasan Masalah

Pemodelan yang kami lakukan dibatasi dengan mengambil rute pulang pergi Jakarta - CGK ke penerbangan domestik (Bali - DPS) dan internasional (Kuala Lumpur - KUL). Fluktuasi yang diamati adalah harga tiket pada liburan akhir tahun berupa rute tersebut. Dibatasi juga bahwa yang berlibur berangkat dari bandara Soekarno-Hatta yaitu warga Jakarta dan sekitarnya. Tanggal yang diamati adalah tanggal 30 Desember 2022 untuk keberangkatan dan 2 Januari 2023 untuk kepulangan.

Model yang dibandingkan akan dihitung nilai performanya menggunakan RMSE (*Root mean squared error*). Akan dicari mana model terbaik yang dapat memprediksi fluktuasi harga tiket pesawat sampai tanggal keberangkatan

2. Tinjauan Pustaka

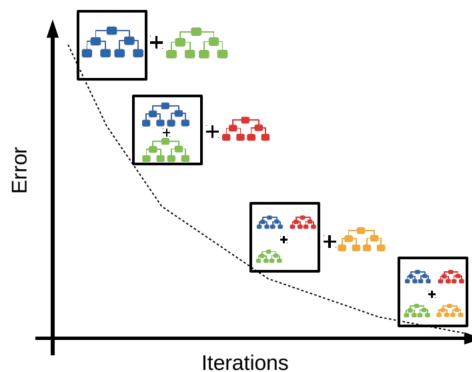
2.1. Autokorelasi

Autokorelasi (atau *autocorrelation*) adalah tingkat korelasi variabel yang sama diantara dua interval waktu yang berurutan. Autokorelasi menghitung korelasi antara *lagged version* dari nilai suatu variabel dengan versi aslinya dalam sebuah *time series*. Autokorelasi hampir sama dengan menghitung korelasi antara 2 variabel, hanya saja autokorelasi menghitung korelasi antara 2 interval waktu yang berbeda, X_t dan X_{t-n} dari *time series* yang sama, n disebut dengan *lag*.

Sama seperti korelasi pada umumnya, autokorelasi mungkin saja positif atau negatif atau 0, dengan *rangennya* dari -1 (*perfectly negative autocorrelation*) sampai 1 (*perfectly positive autocorrelation*). Autokorelasi positif menunjukkan bahwa apabila nilai pada interval waktu yang diamati meningkat maka nilai pada versi *lagged* nya juga cenderung meningkat, begitu juga sebaliknya, jika terjadi penurunan nilai pada interval waktu yang diamati maka nilai pada versi *lagged* nya juga cenderung menurun. Sedangkan autokorelasi negatif menunjukkan apabila nilai pada interval waktu yang diamati meningkat maka nilai pada versi *lagged* nya cenderung menurun, begitu juga sebaliknya. Jika autokorelasinya sama dengan 0 maka data cenderung bersifat acak atau random.

2.2. XGBoost

Gradient boosting merupakan algoritma machine learning yang menggunakan *ensemble* dari *decision tree* untuk memprediksi nilai. *Gradient boosting* dimulai dengan menghasilkan pohon klasifikasi awal dan terus menyesuaikan pohon baru melalui minimisasi fungsi kerugian (Natekin dan Knoll, 2013).



Gambar 1: Proses *training* XGBoost

1. Inisialisasi prediksi.

Hitung rata-rata nilai target:

$$mean = \frac{\sum_{i=0}^n a_i}{n}, a_i = \text{nilai target ke-}i$$

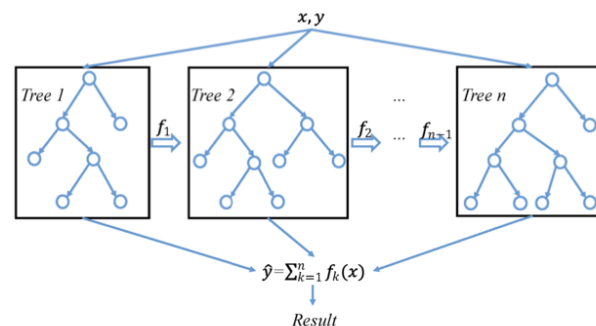
2. Hitung nilai residual.

$$residual = actual\ value_i - mean,$$

$actual\ value_i$ = nilai sebenarnya dari target ke- i

3. Bentuk pohon 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.

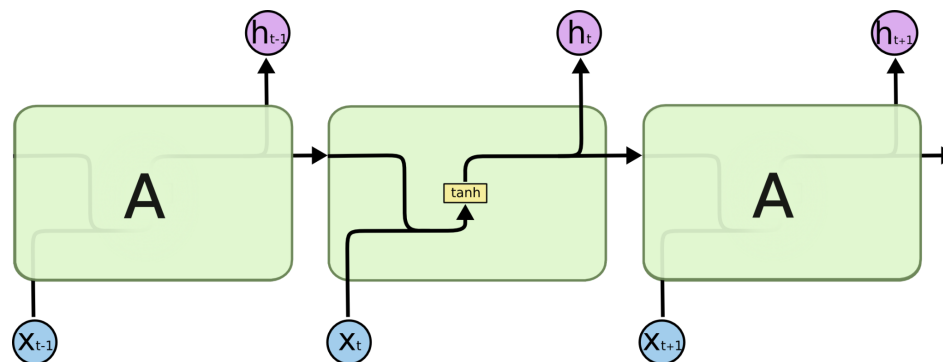
Extreme Gradient Boosting (XGBoost) merupakan pengembangan dari algoritma *Gradient Boosting Machine* klasik dan hanya digunakan untuk data yang memiliki label dalam proses latihnya (Syahrani, 2019). *XGBoost* merupakan pohon ensembles algoritma yang terdiri atas kumpulan beberapa *classification and 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*.



Gambar 2: Ilustrasi arsitektur *XGBoost*

2.3. LSTM

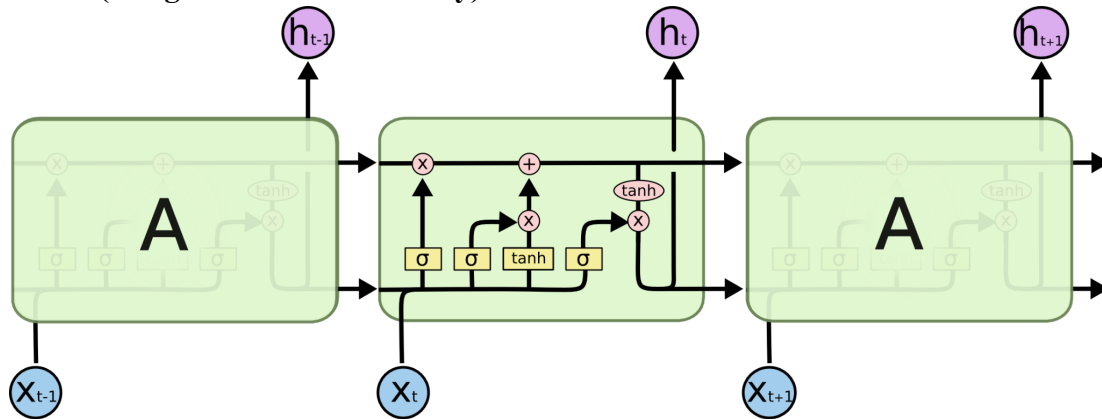
2.4.1. RNN



Gambar 3: Arsitektur RNN

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

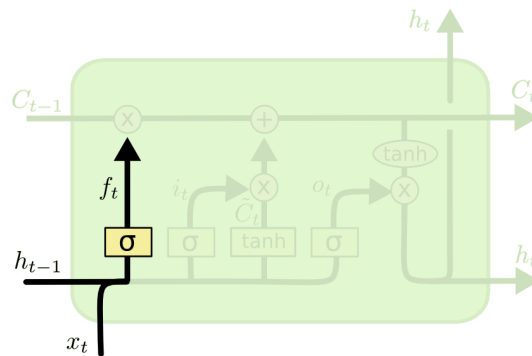
2.4.2. LSTM (Long Short Term Memory)



Gambar 4: Arsitektur LSTM

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.

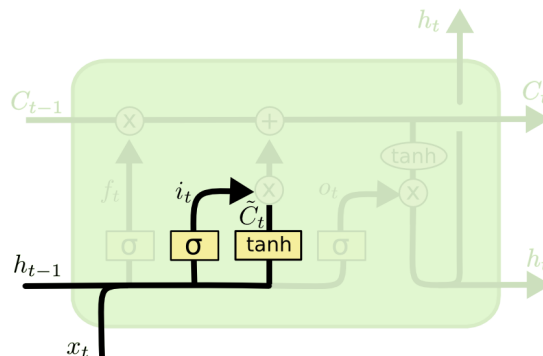
a. Forget Gate



Gambar 5: Forget gate LSTM

Gerbang yang berfungsi untuk melakukan filter informasi mana yang perlu "diingat" dan yang perlu "dilupakan" dengan memanfaatkan fungsi sigmoid.

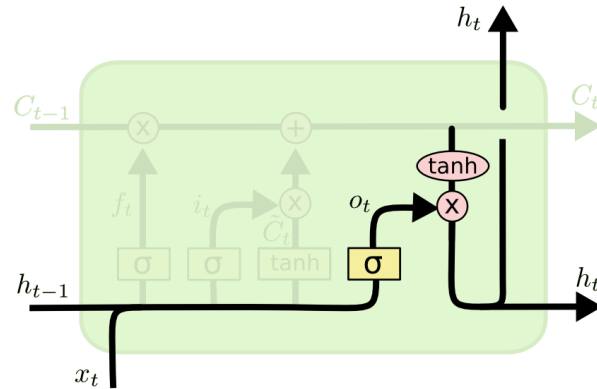
b. Input Gate



Gambar 6: *Input gate* LSTM

Gerbang yang akan menentukan informasi yang perlu untuk diperbaharui.

c. *Output Gate*



Gambar 7: *Output gate* LSTM

Gerbang yang akan menerima bagian – bagian dari cell states yang akan menjadi output.

3. Data dan Rancangan Penelitian

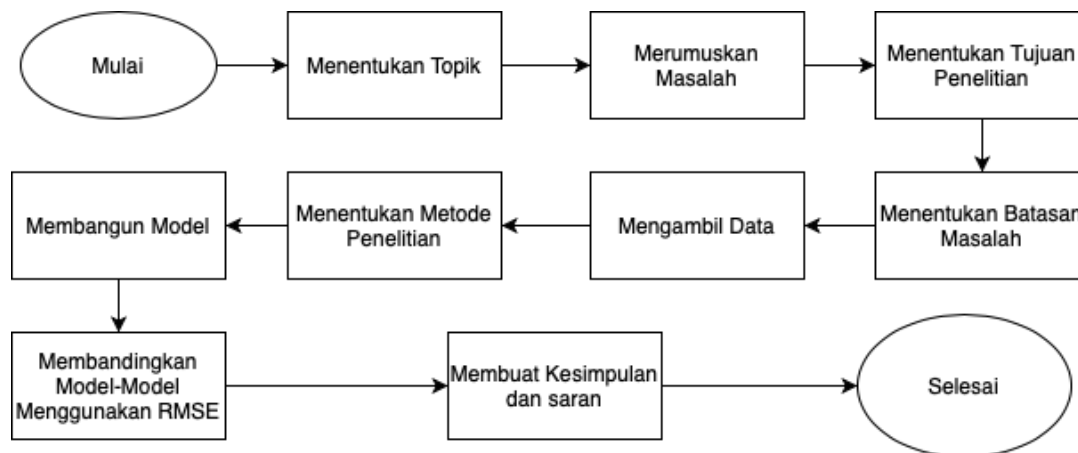
3.1. Pengambilan Data

Data diambil dari situs <https://www.google.com/travel/flights> pada tanggal 15 November 2022. Harga tiket yang diambil adalah harga termurah dari semua maskapai dan tiket *single flight* (tidak bolak-balik). Data yang dikumpulkan berupa fluktuasi harga tiket pesawat untuk penerbangan:

1. Rute Jakarta (CGK) - Denpasar (DPS): Jumat, 30 Desember 2022
Dari rute CGK-DPS didapat data histori harga tiket pesawat sejumlah 285 hari dari tanggal 3 Februari 2022 -15 November 2022.
2. Rute Denpasar (DPS) - Jakarta (CGK): Senin, 2 Januari 2023
Dari rute DPS-CGK didapat data histori harga tiket pesawat sejumlah 118 hari dari tanggal 20 Juli 2022 -15 November 2022.
3. Rute Jakarta (CGK) - Kuala Lumpur (KUL): Jumat, 30 Desember 2022
Dari rute CGK-KUL didapat data histori harga tiket pesawat sejumlah 231 hari dari tanggal 3 Maret 2022 -15 November 2022.
4. Rute Kuala Lumpur (KUL) - Jakarta (CGK): Senin, 2 Januari 2023
Dari rute CGK-KUL didapat data histori harga tiket pesawat sejumlah 69 hari dari tanggal 7 September 2022 - 15 November 2022.

3.2. Tahapan Penelitian

Langkah atau tahapan yang dilakukan pada penelitian ini diberikan pada gambar berikut.



Gambar 8. Tahapan Penelitian

4. Pembahasan

4.1. Evaluasi Kinerja Model

RMSE (*Root Mean Squared Error*) adalah sebuah ukuran (metrik) yang menunjukkan rata rata jarak atau selisih antara nilai yang diprediksi oleh model dengan nilai asli dari dataset. Semakin rendah RMSE yang dihasilkan dari suatu model berarti semakin baik model tersebut dalam memprediksi suatu nilai.

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

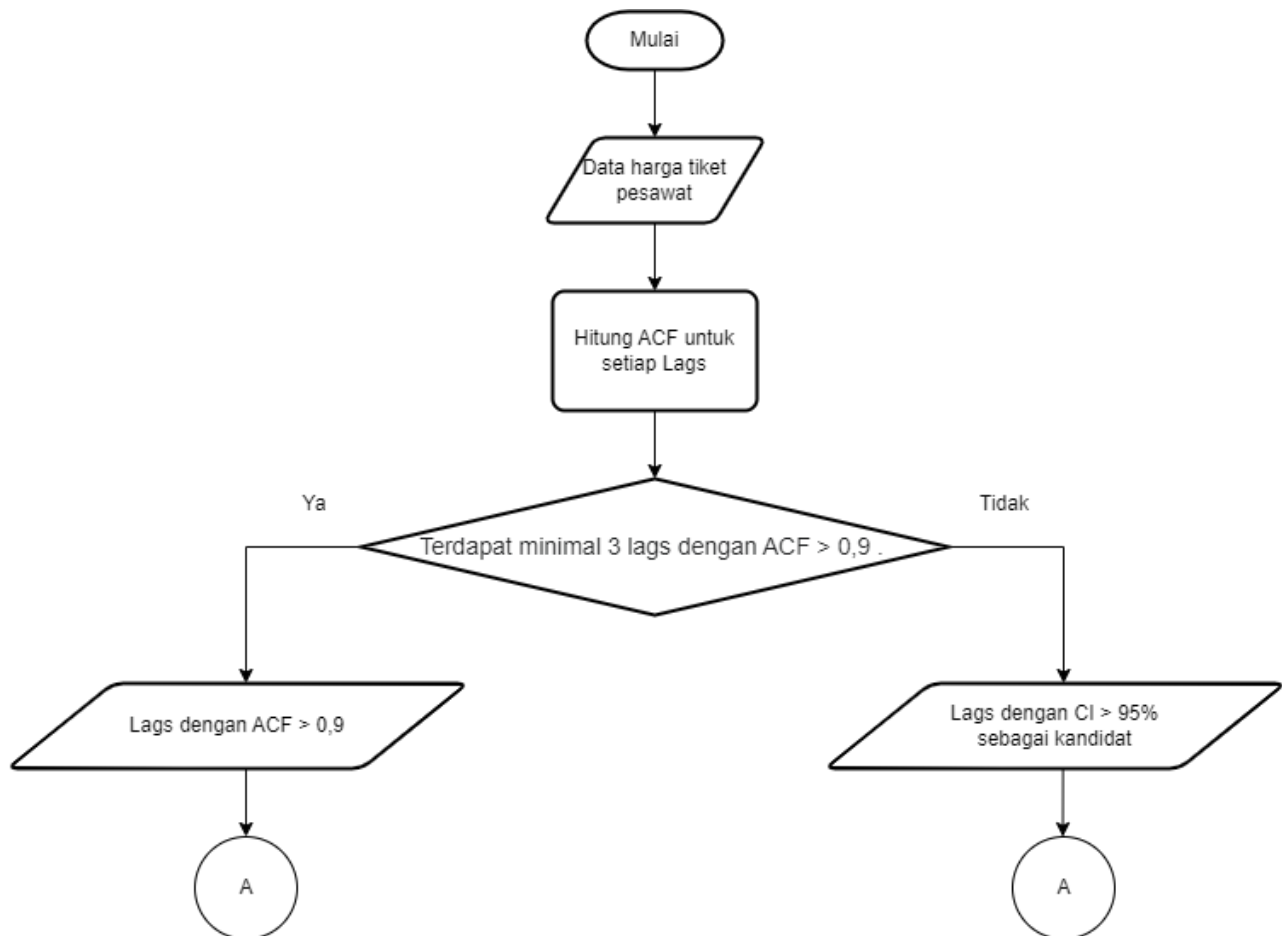
P_i : nilai yang diprediksi oleh model pada observasi ke i pada dataset

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

n : ukuran sampel

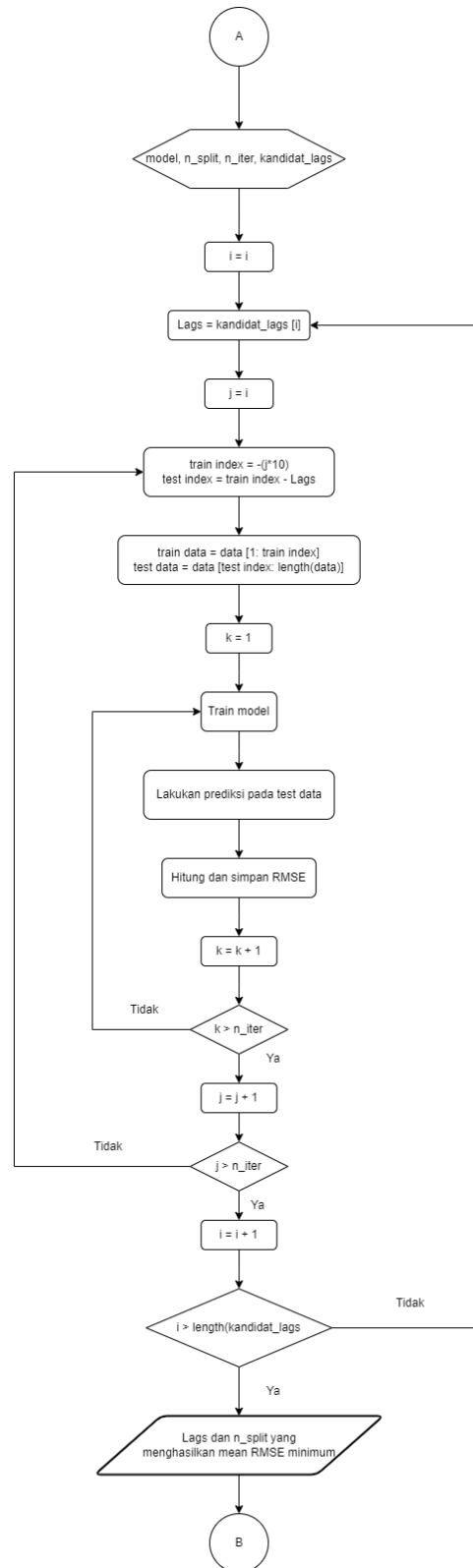
4.2. Skenario Pengujian

4.2.1. Proses Pencarian Kandidat *Lags*



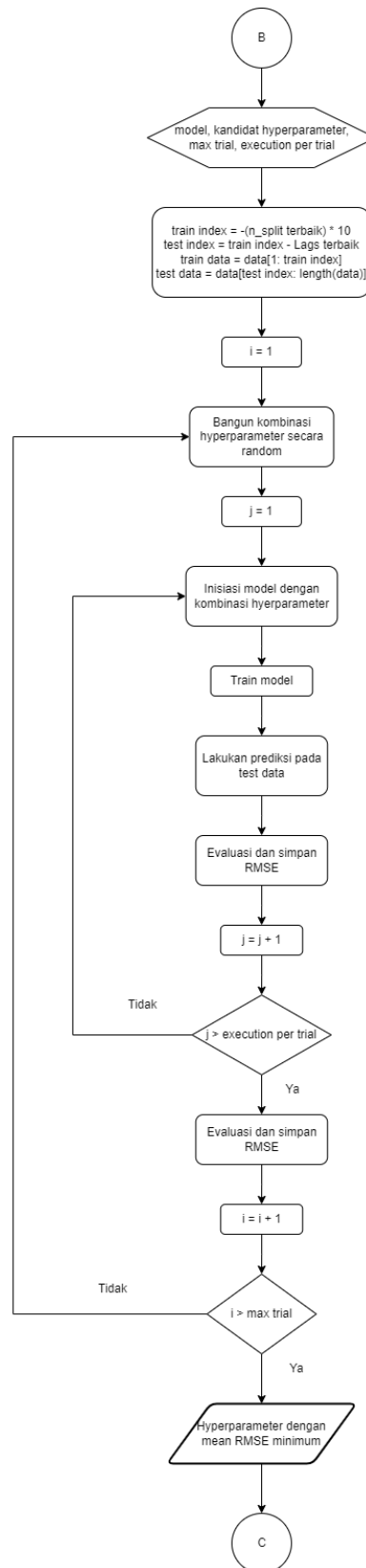
Gambar 9: *Flowchart* pencarian kandidat *lags*

4.2.2. Proses Pencarian *Lags* dan Komposisi *Split Data Train* dan *Test* Teroptimal



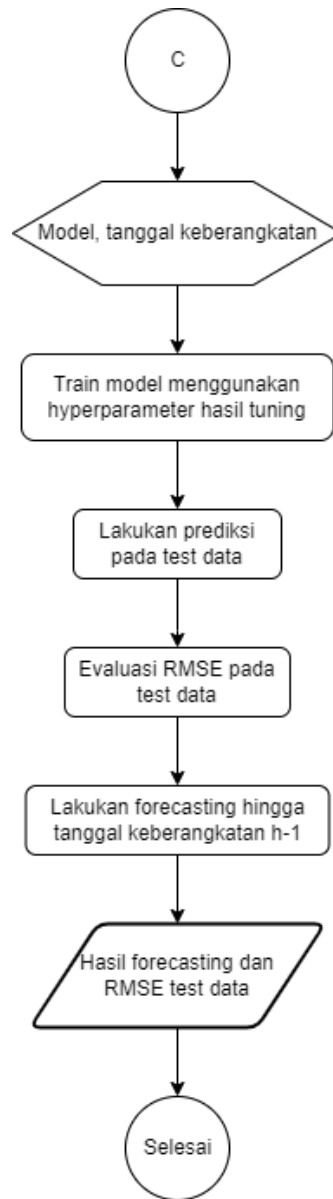
Gambar 10: Flowchart pencarian *lags* dan komposisi *split data train* dan *test* teroptimal

4.2.3. Proses *Hyperparameter Tuning (Randomized Search)*



Gambar 11: Flowchart hyperparameter tuning

4.2.4. Proses *Training*, *Testing*, dan *Forecasting*



Gambar 12: *Flowchart training, testing, dan forecasting*

4.3. Hyperparameter

Kandidat *Hyperparameter XGBoost*

Hyperparameter	Rentang Nilai	<i>Step Size</i>
Maksimum kedalaman <i>tree</i>	[3, 10]	1
Minimum bobot <i>tree</i>	[2, 6]	1
<i>Learning rate</i>	[0.1, 0.3]	0.1

<i>Gamma</i>	[0.01, 0.2]	0.01
Nilai regularisasi L1 (<i>Alpha</i>)	[0,1]	-
Nilai regularisasi L2 (<i>Lambda</i>)	[0,1]	-
<i>Subsample</i>	(0,1]	0.25
<i>Cosample by Tree</i>	(0,1]	0.25
<i>Cosample by Level</i>	(0,1]	0.25

Tabel 1: Kandidat *hyperparameter* XGBoost berdasarkan [9]

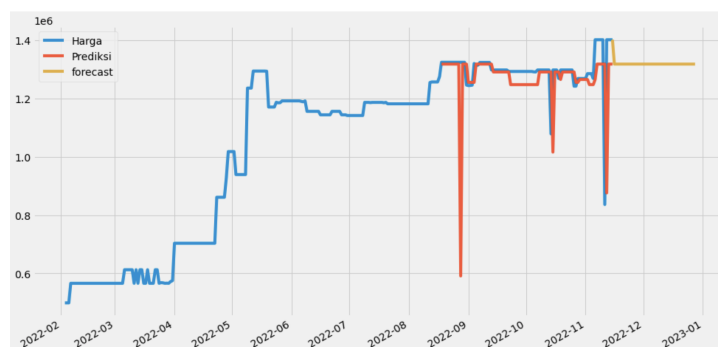
Kandidat *Hyperparameter* LSTM

Hyperparameter	Rentang Nilai	<i>Step Size</i>
Hidden layer LSTM	[1, 3]	1
Neuron pada Layer LSTM	[8, 32]	8
Hidden layer dense	[0, 2]	1
Neuron pada layer dense	[8, 32]	8
Learning Rate	[0.0001, 0.1]	-
Activation Function	['relu', 'tanh', 'sigmoid']	-
Batch Size	32	-
Max Epoch	1000	-

Tabel 2: Kandidat *hyperparameter* LSTM berdasarkan [8]

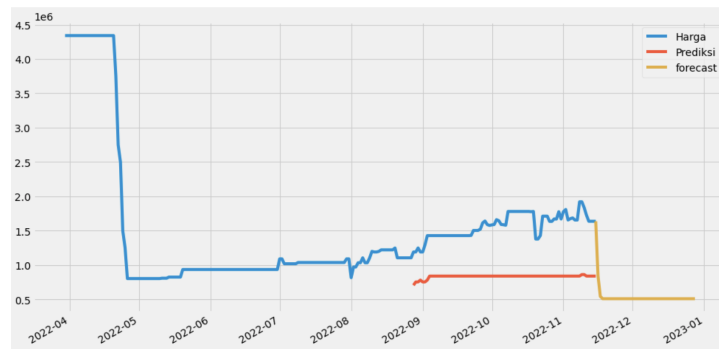
4.4. Hasil Pengujian

Hasil *forecasting* XGBoost CGK-DPS



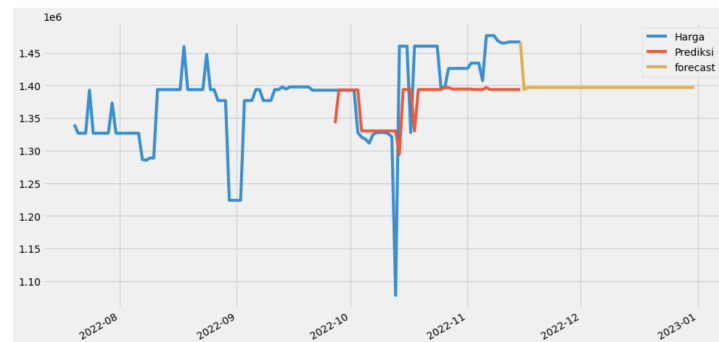
Gambar 13: Plot forecasting XGBoost CGK-DPS

Hasil forecasting *XGBoost* CGK-KUL



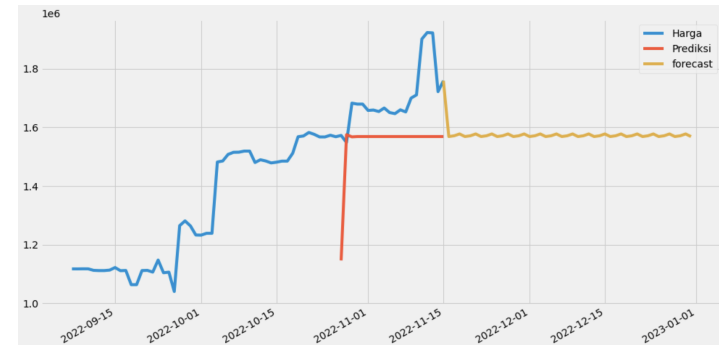
Gambar 14: Plot forecasting *XGBoost* CGK-KUL

Hasil forecasting *XGBoost* DPS-CGK



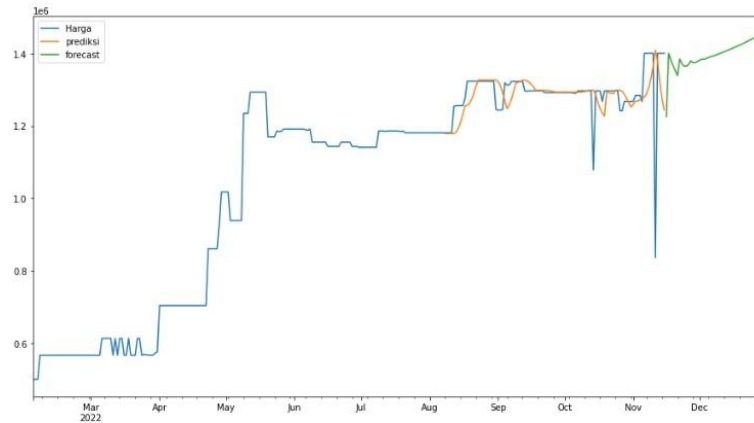
Gambar 15: Plot forecasting *XGBoost* DPS-CGK

Hasil forecasting *XGBoost* KUL-CGK



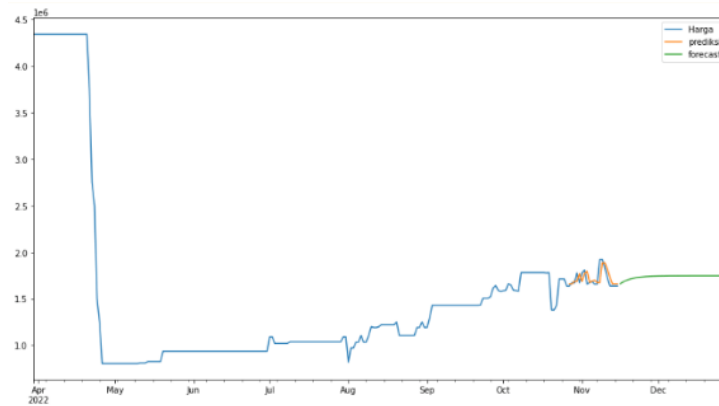
Gambar 16: Plot forecasting *XGBoost* KUL-CGK

Hasil forecasting *LSTM* CGK-DPS



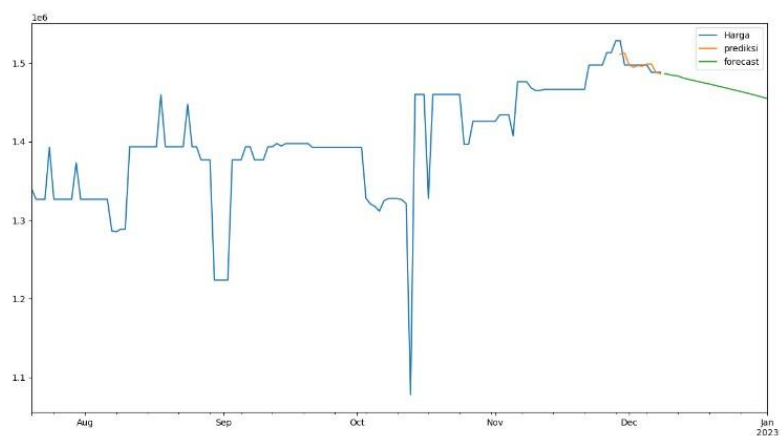
Gambar 17: Plot forecasting LSTM CGK-DPS

Hasil forecasting LSTM CGK-KUL



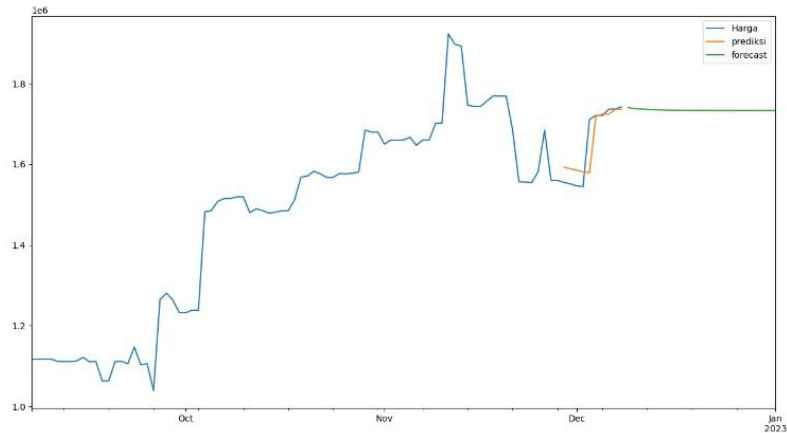
Gambar 18: Plot forecasting LSTM CGK-KUL

Hasil forecasting LSTM DPS-CGK



Gambar 19: Plot forecasting LSTM DPS-CGK

Hasil forecasting LSTM KUL-CGK



Gambar 20: Plot forecasting LSTM KUL-CGK

Tabel perbandingan RMSE

Rute	<i>XGBoost</i>	<i>LSTM</i>
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

Tabel 3: Perbandingan RMSE model *LSTM* dan *XGBoost*

5. Penutup

5.1. Kesimpulan

Dari hasil pengujian bab sebelumnya, tim peneliti menyimpulkan:

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

5.2. Saran

Pada penelitian selanjutnya disarankan untuk mencari dataset yang lebih banyak dan pada tujuan model disarankan waktu yang di-*forecast* menjadi lebih dekat. selain itu disarankan mencari data dengan fitur lain yang dapat dikorelasikan dengan fluktuasi harga.

Daftar Pustaka

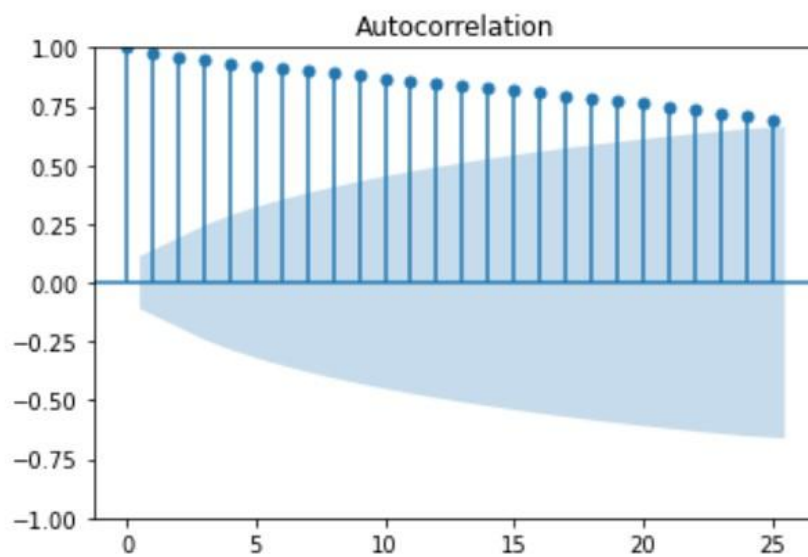
- [1] Ayu W. 2022. Analisis Kinerja Algoritma CNN dan LSTM untuk Memprediksi Tinggi Muka Air di DKI Jakarta, Indonesia. Fakultas Ilmu Komputer. Universitas Mercu Buana: Jakarta
- [2] Chen & Guestrin, 2016. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining .
- [3] lah, Christopher. 2015. Understanding LSTM Networks.<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [4] Google.com/flights, 15 November 2022,
<https://www.google.com/flights?tfs=CBoQAhojEgoyMDIyLTEyLTMwagwIAxIIL20vMDQ0cnZyBwgBEgNEUFNAAUgBUgNJRFJ6ZENqUklSR3RYYnpRNfduQTRiRzlCUWxnNE1FRkNSeTB0TFMwdExTMHRMWFJzWW1OaU9VRkRVUZCUjA0d1ZFbDNTMWwwYzBGQkVnRXhHZ3NJOEw1VkVBQWFbMGxFVWpnWGNLUkc&hl=id&gl=ID&curr=IDR&gsas=1&authuser=1&sa=X&ved=0CAUQ2Y8EahcKEwioouyn1bH7AhUAAAAAHQAAAAAQHA>
- [5] Google.com/flights, 15 November 2022,
<https://www.google.com/flights?tfs=CBoQAhoeEgoyMDIzLTAxLTAyagcIARIDS1VMcgcIARIDQ0dLQAFIAVIDSURSemRDalJJZEMxTFRWRk9hWFpYVDFGQIEwTXpRbWRDUnkwdExTMHRMUzB0TFMxMGFHSnROMEZCUVVGQIIwNHdWR05KVGTVMVkwRkJFZ0V5R2dzSWkvWnJFQUFhQTBSRVVqZ1hjTnhZ&hl=id&gl=ID&curr=IDR&gsas=1&authuser=1&sa=X&ved=0CBQQ2Y8EahcKEwjA79671rH7AhUAAAAAHQAAAAAQHA>
- [6] Google.com/flights, 15 November 2022,
<https://www.google.com/travel/flights/search?tfs=CBwQAhojagwIAxIIL20vMDQ0cnYS8BQAFIAZgBAG&tfu=CmRDalJJTjJJMVp6bFFUbHBIZWtGQIFtTnJZMUZDUnkwdExTMHRMUzB0TFMxMGJIQnNOVUZCUVVGQIIwNHdWRkpqUlhoWk9FRkJFZ0V6R2dzSTI2bGpFQUFhQTBSRVVqZ1hjSIZT>
- [7] Google.com/flights, 15 November 2022,
<https://www.google.com/travel/flights/search?tfs=CBwQAhojagcIARIDRFBTEgoyMDIzLTAxLTAycgwIAxIIL20vMDQ0cnZwAYIBCwj8BQAFIAZgBAG&tfu=CmRDalJJU25OV2VHbDJNbU40T1UxQIEweEpOa0ZDUnkwdExTMHRMUzB0ZEdoaVxzUXhPRUZCUVVGQIIwNHdWR3N3VFMxR1lrRkJFZ0V3R2dzSXlZQlpFOUFhQTBSRVVqZ1hjTXBK>
- [8] Chowdhury, Kuldeep. 2021. 10 Hyperparameters to keep an eye on for your LSTM model — and other tips
<https://medium.com/geekculture/10-hyperparameters-to-keep-an-eye-on-for-your-lstm-m>

[odel-and-other-tips-f0ff5b63fcd4](#)

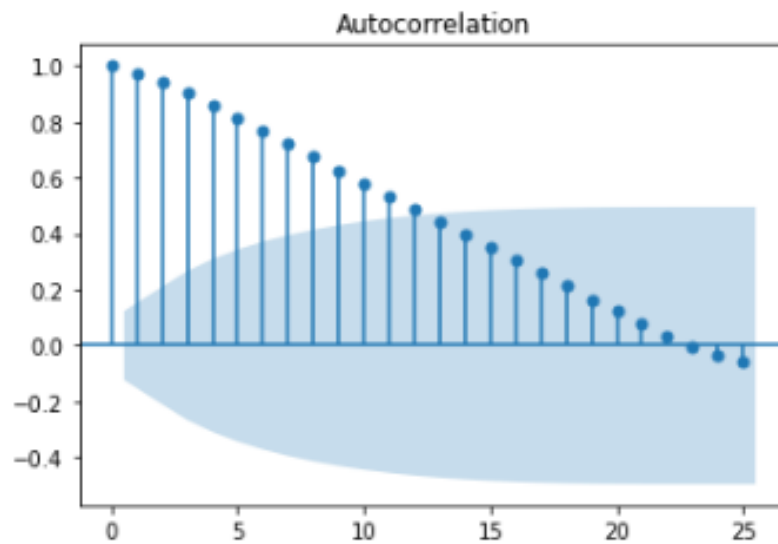
- [9] Kiran, Mustafa S. 2021. Crude Oil Price Forecasting Using XGBoost. Department of Computer Engineering. Selcuk University.

LAMPIRAN

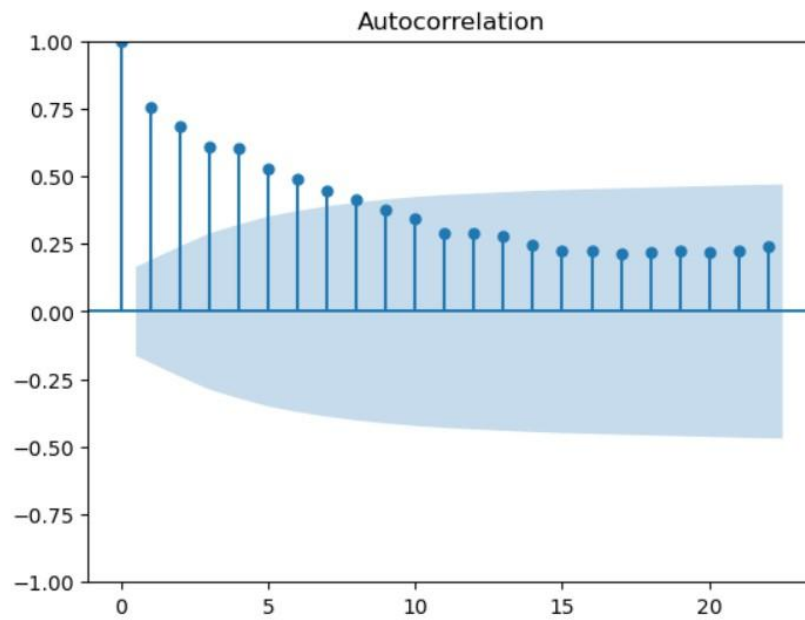
Plot Autokorelasi harga tiket pesawat rute CGK-DPS



Plot Autokorelasi harga tiket pesawat rute CGK-KUL



Plot Autokorelasi harga tiket pesawat rute DPS-CGK



Plot Autokorelasi harga tiket pesawat rute KUL-CGK

