

Passenger Density Prediction at the Airport Using LSTM and SARIMA: A Case Study at Radin Inten Airport, Lampung

Diaji Yugo Prasajo ^{1*}, Kurnia Muludi ^{2*}

Institut Informatika dan Bisnis Darmajaya

diaji.2321211016@mail.darmajaya.ac.id ¹, kurnia@darmajaya.ac.id ²

Article Info

Article history:

Received 2025-06-24

Revised 2025-07-28

Accepted 2025-08-09

Keyword:

*Passenger Density Prediction,
SARIMA,
LSTM,
Random Forest,
XGBoost,
Time Series,
Airport.*

ABSTRACT

Prediksi kepadatan penumpang di bandara merupakan aspek penting dalam perencanaan operasional dan pengambilan keputusan strategis. Penelitian ini bertujuan untuk mengembangkan model prediksi jumlah penumpang di Bandara Radin Inten Lampung menggunakan kombinasi model Seasonal AutoRegressive Integrated Moving Average (SARIMA) dan Long Short-Term Memory (LSTM), serta membandingkannya dengan model Random Forest dan XGBoost. Dataset yang digunakan mencakup jumlah penumpang harian dari Januari 2023 hingga Desember 2024. Penelitian ini meliputi tahapan eksplorasi data, preprocessing, pemodelan SARIMA dan LSTM secara terpisah, serta penggabungan keduanya melalui pendekatan residual learning. Hasil evaluasi menunjukkan bahwa model SARIMA memiliki performa terbaik dalam menangkap pola musiman dengan nilai MAPE sebesar 3.81%, diikuti oleh Random Forest dengan MAPE sebesar 5.81% dan XGBoost dengan MAPE sebesar 5.84%. Model LSTM menunjukkan hasil yang lebih rendah dengan MAPE sebesar 6.81%. Meskipun kombinasi SARIMA dan LSTM menjanjikan secara teoritis, hasilnya lebih buruk dengan MAPE sebesar 14.27% akibat akumulasi error pada tahap residual learning. Penelitian ini menunjukkan bahwa pemilihan model yang tepat sangat bergantung pada karakteristik data dan tujuan prediksi, serta pentingnya penggabungan multi-model untuk meningkatkan akurasi prediksi dalam aplikasi peramalan kepadatan penumpang di bandara.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Kepadatan penumpang di bandara merupakan faktor penting yang mempengaruhi operasional dan pelayanan [1]. Dengan meningkatnya jumlah penumpang, pengelola bandara perlu memiliki sistem yang efektif untuk memprediksi kepadatan penumpang agar dapat merencanakan sumber daya dan meningkatkan pengalaman penumpang [2]. Penelitian ini bertujuan untuk mengembangkan model prediksi kepadatan penumpang di Bandara Radin Inten Lampung menggunakan kombinasi model SARIMA dan LSTM.

Prediksi jumlah penumpang di bandara merupakan komponen krusial dalam perencanaan operasional dan pengambilan keputusan strategis [3]. Informasi akurat mengenai jumlah penumpang yang akan datang dapat membantu pengelola bandara dalam mengatur sumber daya,

meningkatkan layanan, dan merancang strategi bisnis yang efektif [4]. Dalam konteks ini, pemodelan deret waktu menjadi alat yang esensial untuk memproyeksikan tren dan pola musiman dalam data historis jumlah penumpang.

Seasonal AutoRegressive Integrated Moving Average (SARIMA) adalah metode statistik yang umum digunakan untuk menganalisis dan memprediksi data deret waktu yang memiliki pola musiman [5]. Model ini memperhitungkan komponen autoregresif (AR), perbedaan (I), dan moving average (MA) serta faktor musiman yang berulang secara periodik [6]. Penelitian sebelumnya telah menunjukkan bahwa SARIMA efektif dalam meramalkan jumlah penumpang pesawat di berbagai bandara. Misalnya, studi di Bandara Sultan Iskandar Muda menggunakan SARIMA untuk memprediksi jumlah penumpang dan menemukan

bahwa model ini mampu menangkap pola musiman dengan baik[7]

Di sisi lain, Long Short-Term Memory (LSTM) adalah jenis jaringan saraf tiruan yang dirancang untuk mengatasi masalah long-term dependencies dalam data deret waktu [8]. LSTM memiliki kemampuan untuk belajar dan mengingat informasi jangka panjang, sehingga cocok untuk memodelkan data dengan pola yang kompleks dan non-linier [9]. Penelitian telah menunjukkan bahwa LSTM dapat memberikan akurasi prediksi yang tinggi dalam berbagai aplikasi deret waktu, termasuk peramalan jumlah penumpang[10]. Sebagai contoh, sebuah studi menggunakan LSTM untuk memprediksi jumlah penumpang pesawat dan menemukan bahwa model ini mampu menangkap pola data yang kompleks [11].

Pendekatan hibrida yang menggabungkan model statistik tradisional seperti SARIMA dengan model pembelajaran mesin seperti LSTM telah mendapatkan perhatian dalam literatur terbaru [12]. Tujuan dari kombinasi ini adalah untuk memanfaatkan keunggulan masing-masing model dalam menangkap pola linier dan non-linier dalam data deret waktu. [13]. Misalnya, sebuah studi menunjukkan bahwa kombinasi ARIMA dan LSTM dapat memberikan hasil prediksi yang lebih akurat dibandingkan dengan masing-masing model secara terpisah [14].

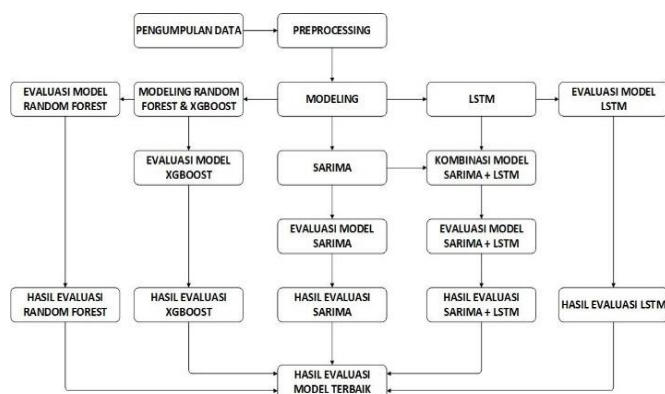
Selain pendekatan deret waktu, penelitian ini juga menguji performa model Random Forest dan XGBoost. Kedua model ini dikenal kuat dalam menangani data non-linier dengan banyak fitur, dan banyak digunakan untuk prediksi regresi [15]. Namun, berbeda dengan SARIMA dan LSTM, kedua model ini tidak secara eksplisit mempertimbangkan dimensi temporal, sehingga analisis komparatif diperlukan untuk menilai kelayakan dan efektivitasnya dalam konteks data penumpang bandara [16]. Dengan mengombinasikan pendekatan statistik, deep learning, dan machine learning klasik, penelitian ini diharapkan dapat memberikan kontribusi yang lebih menyeluruh dalam peramalan jumlah penumpang, serta memberikan insight terhadap kekuatan dan kelemahan masing-masing metode dalam konteks sistem transportasi udara di Indonesia.

Penelitian ini memiliki keunikan tersendiri karena mengangkat konteks lokal, yaitu Bandara Radin Inten Lampung, yang hingga saat ini belum banyak dikaji secara akademik terutama dalam konteks pemodelan prediksi jumlah penumpang. Sebagian besar studi terdahulu berfokus pada bandara besar nasional atau internasional, sehingga pendekatan berbasis data lokal memberikan kontribusi baru dalam literatur transportasi udara Indonesia. Selain itu, penelitian ini tidak hanya membandingkan metode deret waktu statistik seperti SARIMA dan pendekatan deep learning seperti LSTM secara individual, tetapi juga mengeksplorasi penggabungan keduanya melalui pendekatan residual learning. Penambahan model Random Forest dan XGBoost sebagai pembanding memperkuat kontribusi komparatif penelitian ini karena mempertimbangkan aspek linier, non-linier, dan kekuatan model berbasis pohon keputusan dalam menangani data time series eksogen. Kombinasi dan perbandingan lintas pendekatan ini

menjadikan penelitian ini komprehensif serta relevan untuk penerapan prediktif di sektor transportasi udara lokal.

II. METODE

Penelitian ini bertujuan untuk memprediksi kepadatan penumpang di Bandara Radin Inten Lampung menggunakan kombinasi model Seasonal AutoRegressive Integrated Moving Average (SARIMA) dan Long Short-Term Memory (LSTM). Tahapan penelitian ini terdiri dari pengumpulan data, preprocessing, pemodelan SARIMA, pemodelan LSTM, kombinasi model, serta evaluasi performa model. Berikut adalah alur proses penelitian pada gambar 1.



Gambar 1 Alur Penelitian

A. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini terdiri dari jumlah penumpang harian di Bandara Radin Inten Lampung dari Januari 2023 hingga Desember 2024. Data ini diperoleh dari Bandara Radin Inten secara langsung untuk jumlah penumpang dan juga Jumlah Penerbangan pada setiap harinya. Selain itu terdapat 3 atribut lain yang di kumpulkan yaitu suhu minimal dan suhu maksimal pada setiap harinya dan juga curah hujan di daerah Bandar Lampung yang di dapatkan langsung dari data public di open meteo. Berikut adalah atribut yang terdapat dalam dataset pada gambar 2.

1	Tanggal	Jumlah_Penumpang	Suhu_Minimal	Suhu_Maksimal	Curah_Hujan	Hari_Libur	Jumlah_Penerbangan
2	01/01/2023	1713	29.5	22.8	2.0	1	12
3	02/01/2023	2094	29.5	22.7	5.9	0	16
4	03/01/2023	1926	28.3	22.5	3.6	0	16
5	04/01/2023	1862	30.1	22.7	3.7	0	16
6	05/01/2023	1813	30.5	22.8	3.4	0	14
7	06/01/2023	1646	30.0	22.9	3.8	0	20
8	07/01/2023	1815	29.9	24.1	12.6	0	16
9	08/01/2023	2056	29.8	23.7	13.1	1	16
10	09/01/2023	1844	31.2	22.8	3.2	0	14
11	10/01/2023	1666	29.8	23.6	4.8	0	12
12	11/01/2023	1889	29.5	24.7	2.7	0	16
13	12/01/2023	1821	29.5	24.8	2.8	0	16
14	13/01/2023	1966	30.0	25.7	3.3	0	16
15	14/01/2023	1769	29.6	24.9	0.2	0	16
16	15/01/2023	1985	30.0	24.7	0.4	1	16
17	16/01/2023	1948	29.4	23.1	3.2	0	14
18	17/01/2023	1742	27.8	23.2	10.9	0	14
19	18/01/2023	2084	28.0	22.8	18.5	0	16

Gambar 2 Dataset Penelitian

Data menunjukkan adanya fluktuasi jumlah penumpang yang cenderung meningkat pada periode liburan dan musim puncak (seperti liburan Natal dan Tahun Baru, serta libur panjang lainnya). Sebagai contoh, pada periode akhir tahun,

jumlah penumpang biasanya lebih tinggi dibandingkan dengan bulan-bulan lainnya, yang mengindikasikan pola musiman yang jelas. Kemudian dalam dataset, terdapat outlier yang dapat terjadi pada waktu tertentu, seperti peristiwa besar (misalnya pandemi atau liburan nasional) yang dapat menyebabkan lonjakan atau penurunan jumlah penumpang yang sangat besar. Untuk memastikan model dapat memprediksi dengan baik, outlier ini telah ditangani menggunakan metode Interquartile Range (IQR) pada tahap preprocessing.

B. Preprocessing

Tahapan preprocessing dilakukan untuk memastikan data dalam kondisi optimal sebelum digunakan untuk pemodelan. Beberapa langkah yang dilakukan meliputi:

- Penanganan Missing Values: Nilai yang hilang diisi dengan metode interpolasi atau nilai median.
- Deteksi dan Penanganan Outlier: Menggunakan metode Interquartile Range (IQR) untuk menghilangkan data yang ekstrem.
- Normalisasi Data: Menggunakan Min-Max Scaling untuk mengonversi nilai fitur ke rentang [0,1] agar lebih optimal dalam pemrosesan oleh LSTM.
- Feature Engineering: Menambahkan fitur seperti hari dalam seminggu, bulan, serta moving average untuk menangkap pola musiman.

C. Pemodelan Sarima

Model SARIMA (Seasonal Autoregressive Integrated Moving Average) digunakan untuk memodelkan data deret waktu yang memiliki pola musiman. Model ini memperhitungkan komponen autoregresif (AR), diferensiasi (I), dan moving average (MA), serta faktor musiman yang berulang secara periodik. Model SARIMA terdiri dari dua bagian utama:

1. Model SARIMA terdiri dari dua bagian utama:
 - p (Order of Autoregression): Menyatakan jumlah lag (penundaan) yang digunakan untuk memodelkan hubungan antara variabel dengan nilai-nilai masa lalu (autocorrelation).
 - d (Degree of Differencing): Menentukan jumlah diferensiasi yang diperlukan untuk menjadikan data stasioner (menghilangkan tren).
 - q (Order of Moving Average): Menyatakan jumlah lag dari residual model sebelumnya yang digunakan untuk memprediksi nilai saat ini.
2. (P, D, Q, m) - untuk komponen musiman:
 - P (Order of Seasonal Autoregression): Menyatakan jumlah lag musiman yang digunakan untuk memodelkan hubungan musiman dalam data.
 - D (Degree of Seasonal Differencing): Menunjukkan jumlah diferensiasi musiman yang diperlukan untuk menjadikan data musiman stasioner.

- Q (Order of Seasonal Moving Average): Menyatakan jumlah lag musiman dari residual model musiman sebelumnya.
- m (Seasonal Periodicity): Menunjukkan panjang periode musiman, misalnya 12 untuk data bulanan dengan musiman tahunan.

Parameter-parameter SARIMA dioptimalkan dengan menggunakan Akaike Information Criterion (AIC), yang digunakan untuk menilai keseimbangan antara kompleksitas model dan akurasi prediksi. AIC dihitung dengan rumus:

$$AIC = 2k - 2\ln(L)$$

di mana :

k : adalah jumlah parameter dalam model,

L : adalah nilai likelihood (kemungkinan) model.

AIC yang lebih rendah menunjukkan model yang lebih baik dalam hal keseimbangan antara kecocokan model dan kompleksitasnya. Model dengan AIC yang lebih tinggi cenderung overfitting (terlalu kompleks) atau underfitting (terlalu sederhana).

D. Pemodelan LSTM

LSTM (Long Short-Term Memory) adalah jenis recurrent neural network (RNN) yang digunakan untuk memproses dan memprediksi data deret waktu (time series). LSTM memiliki kemampuan untuk menangkap dependensi jangka panjang dalam data yang memiliki urutan waktu, yang sangat berguna untuk model prediksi jumlah penumpang dalam kasus ini. Berikut adalah struktur LSTM yang digunakan dalam model ini:

1. Layer LSTM pertama:
 - Jumlah unit: 50
 - Return_sequences: True. Artinya, lapisan LSTM pertama akan mengembalikan seluruh urutan output, yang akan diteruskan ke lapisan LSTM kedua.
 - Input shape: (n_steps, 1), di mana n_steps adalah panjang input sequence (dalam kasus ini, 30 langkah waktu) dan 1 adalah jumlah fitur (di sini hanya ada satu fitur, yaitu jumlah penumpang).
2. Dropout layer pertama:

Rate: 0.2. Dropout digunakan untuk mengurangi overfitting dengan menghapus 20% dari neuron pada lapisan ini selama pelatihan.
3. Layer LSTM kedua:
 - Jumlah unit: 50
 - Return_sequences: False. Lapisan LSTM kedua hanya mengembalikan output terakhir dari urutan, yang akan diteruskan ke lapisan berikutnya, yaitu layer Dense.
4. Dropout layer kedua:

Rate: 0.2. Fungsi dropout ini bertujuan untuk mencegah overfitting.
5. Layer Dense pertama:

Jumlah unit: 25. Dense layer ini bertujuan untuk memproses output dari lapisan LSTM untuk menghasilkan output yang lebih terstruktur dan memberikan prediksi jumlah penumpang.

6. Layer Dense kedua:

Jumlah unit: 1. Lapisan ini menghasilkan output akhir (yaitu prediksi jumlah penumpang).

Pada Pengaturan Model LSTM yang digunakan adalah sebagai berikut:

1. **Optimizer:** Adam. Adam adalah salah satu optimizer yang sangat populer dalam jaringan saraf karena efisiensinya dalam menangani data besar dan kompleks. Pengaturan learning rate adalah 0.001, yang merupakan nilai umum untuk Adam dalam banyak kasus.
2. **Loss Function:** Mean Squared Error (MSE). Fungsi loss ini digunakan untuk menghitung seberapa besar perbedaan antara prediksi dan nilai aktual. MSE cocok untuk masalah regresi seperti ini.
3. **Epochs:** 50. Jumlah epoch menentukan berapa kali model akan melihat seluruh data pelatihan. Dalam hal ini, model dilatih selama 50 iterasi.
4. **Batch Size:** 16. Ini adalah jumlah sampel data yang diproses dalam satu iterasi sebelum bobot diperbarui.

E. Cross-Validation dan Pembagian Data

Untuk memastikan model dapat memprediksi data yang tidak terlihat sebelumnya, data dibagi menjadi dua bagian: data latih (train) dan data uji (test). Pembagian ini dilakukan tanpa pengacakan (shuffle=False) untuk mempertahankan urutan waktu yang penting dalam model prediksi deret waktu.

1. **Data Latih (Train):** Menggunakan 80% dari data (data hingga Desember 2023).
2. **Data Uji (Test):** Menggunakan 20% dari data (data tahun 2024).

Dengan pembagian data ini, model dapat dilatih pada data sebelumnya (tahun 2023) dan diuji pada data yang lebih baru (tahun 2024), yang memastikan bahwa prediksi dilakukan pada data yang belum pernah dilihat sebelumnya.

F. Kombinasi Model

Pendekatan kombinasi dilakukan untuk menggabungkan keunggulan SARIMA dalam menangkap pola linier dan LSTM dalam menangkap pola non-linier. Langkah-langkah kombinasi model meliputi:

- **Prediksi Awal Menggunakan SARIMA**, menghasilkan nilai dasar prediksi.
- **Ekstraksi Residual dari SARIMA**, yaitu selisih antara nilai aktual dan nilai prediksi SARIMA.
- **Pelatihan LSTM Menggunakan Residual SARIMA**, untuk memodelkan pola non-linier yang tidak dapat ditangkap oleh SARIMA.

- **Kombinasi Output**, yaitu hasil prediksi SARIMA ditambah dengan hasil prediksi residual dari LSTM untuk mendapatkan prediksi akhir.

G. Evaluasi Performa Model

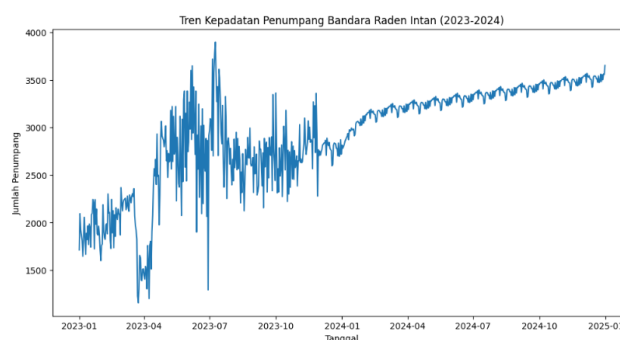
Evaluasi dilakukan untuk membandingkan performa model SARIMA, LSTM, dan kombinasi keduanya menggunakan metrik berikut:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)

III. HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini mencakup informasi jumlah penumpang harian di Bandara Radin Inten Lampung selama periode Januari 2023 hingga Desember 2024. Data ini terdiri dari beberapa atribut utama, yaitu jumlah penumpang, tanggal, jumlah penerbangan, suhu minimum, suhu maksimum, dan curah hujan. Sumber data yang digunakan berasal dari dua kategori utama: data operasional bandara yang mencakup jumlah penumpang dan jumlah penerbangan, serta data meteorologi seperti suhu dan curah hujan yang diperoleh melalui layanan Open-Meteo.

Sebagai langkah awal dalam penelitian ini, dilakukan eksplorasi dan analisis tren menggunakan visualisasi deret waktu (time series visualization) untuk mengidentifikasi pola dan karakteristik data. Hasil analisis menunjukkan adanya pola musiman yang cukup jelas serta fluktuasi harian yang dipengaruhi oleh berbagai faktor eksternal. Guna memberikan gambaran lebih jelas, berikut ditampilkan visualisasi tren jumlah penumpang pada tahun 2023 dan prediksi untuk tahun 2024 yang ditunjukkan pada Gambar 3.



Gambar 3. Visualisasi Jumlah Penumpang Pesawat

Setelah itu, dilakukan preprocessing data, yang mencakup:

- **Penanganan missing values** menggunakan metode interpolasi dan pengisian nilai median.
- **Deteksi dan penghapusan outliers** menggunakan metode Interquartile Range (IQR) untuk memastikan bahwa data tidak mengandung nilai ekstrem yang dapat mengganggu pemodelan.

- Normalisasi data menggunakan Min-Max Scaling untuk memastikan bahwa data dapat diolah dengan baik oleh model LSTM.
- Feature Engineering, yaitu menambahkan variabel tambahan seperti hari dalam seminggu, bulan, dan moving average 7 hari untuk menangkap pola musiman yang lebih kompleks.

A. Pemodelan Sarima

Model SARIMA digunakan untuk menangkap pola musiman dalam data. Setelah dilakukan eksplorasi awal menggunakan uji Augmented Dickey-Fuller (ADF), diketahui bahwa data mengalami tren dan musiman, sehingga perlu dilakukan differencing untuk menjadikannya stasioner. Melalui metode Grid Search, diperoleh parameter SARIMA terbaik, seperti pada gambar 2 berikut.

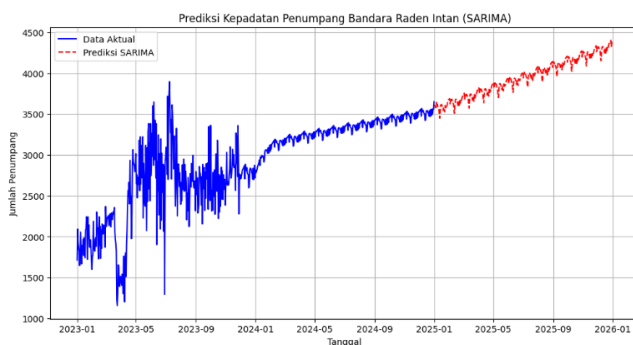
```
# Parameter terbaik yang ditemukan
print(f"Best SARIMA parameters: {best_param} with AIC: {best_aic}")

Best SARIMA parameters: (2, 1, 2, 1, 1, 30) with AIC: 9033.04637798484
```

Gambar 2 Parameter SARIMA Terbaik yang di peroleh

dengan Akaike Information Criterion (AIC) = 9033.046, yang menunjukkan bahwa model ini memiliki keseimbangan yang cukup baik antara kompleksitas dan akurasi prediksi.

Model SARIMA kemudian digunakan untuk memprediksi jumlah penumpang harian pada Januari 2025 hingga Desember 2025. Hasil prediksi menunjukkan bahwa model mampu menangkap pola musiman yang terdapat dalam dataset historis, dengan fluktuasi jumlah penumpang yang mengikuti pola asli. Berikut adalah hasil visualisasi penangkapan pola asli pada gambar 4.



Gambar 4. Visualisasi Hasil Prediksi menggunakan SARIMA

Kemudian Berikut adalah hasil dari prediksi pada jumlah penumpang harian di bandara raden inten Lampung pada gambar 5.

Tanggal	Jumlah_Penumpang
2025-01-01	3610
2025-01-02	3607
2025-01-03	3587
2025-01-04	3577
2025-01-05	3644
...	...
2025-12-27	4383
2025-12-28	4401
2025-12-29	4329
2025-12-30	4379
2025-12-31	4398

Gambar 5. Hasil Prediksi dengan SARIMA

B. Pemodelan LSTM

Model Long Short-Term Memory (LSTM) digunakan untuk menangkap pola non-linier dalam data. Dataset dibagi menjadi training dan testing, kemudian dikonversi menjadi sequences dengan window size = 10 hari. Arsitektur LSTM yang digunakan adalah sebagai berikut:

- 2 LSTM Layers, masing-masing dengan 50 unit neuron.
- Dropout layer (0.2) untuk mencegah overfitting.
- Dense layer sebagai output untuk memprediksi jumlah penumpang.
- Optimizer Adam dengan learning rate 0.001.

Model dilatih menggunakan Mean Squared Error (MSE) sebagai fungsi loss, dengan hasil konvergensi setelah 50 epoch. Berikut adalah hasil dari proses pelatihan dengan Mean Squared Error pada gambar 6.

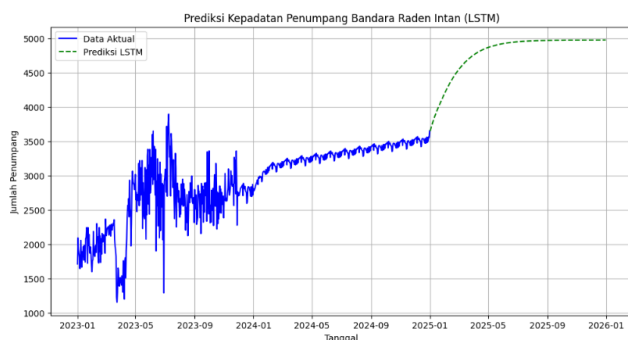
```
# Train model
history = model.fit(X_train, y_train, validation_data=(X_test, y_test), epochs=50)

Epoch 22/50
35/35 ————— 1s 25ms/step - loss: 0.0082 - val_loss: 5.0053e-04
Epoch 23/50
35/35 ————— 1s 40ms/step - loss: 0.0111 - val_loss: 2.9796e-04
Epoch 24/50
35/35 ————— 3s 50ms/step - loss: 0.0084 - val_loss: 6.2586e-04
Epoch 25/50
35/35 ————— 2s 27ms/step - loss: 0.0107 - val_loss: 6.1450e-04
Epoch 26/50
35/35 ————— 1s 27ms/step - loss: 0.0077 - val_loss: 8.6279e-04
Epoch 27/50
35/35 ————— 1s 26ms/step - loss: 0.0088 - val_loss: 0.0016
Epoch 28/50
35/35 ————— 1s 25ms/step - loss: 0.0096 - val_loss: 2.7448e-04
Epoch 29/50
35/35 ————— 1s 25ms/step - loss: 0.0092 - val_loss: 7.2450e-04
Epoch 30/50
35/35 ————— 1s 27ms/step - loss: 0.0093 - val_loss: 0.0025
Epoch 31/50
35/35 ————— 1s 33ms/step - loss: 0.0079 - val_loss: 2.5045e-04
Epoch 32/50
35/35 ————— 1s 40ms/step - loss: 0.0081 - val_loss: 2.9374e-04
Epoch 33/50
35/35 ————— 2s 28ms/step - loss: 0.0097 - val_loss: 2.4951e-04
Epoch 34/50
35/35 ————— 1s 32ms/step - loss: 0.0073 - val_loss: 2.4324e-04
Epoch 35/50
35/35 ————— 1s 25ms/step - loss: 0.0080 - val_loss: 0.0012
Epoch 36/50
---
```

Gambar 6 Hasil pelatihan dengan Mean Squared Error

Hasil prediksi menggunakan model LSTM menunjukkan bahwa model ini mampu menangkap pola fluktuatif dalam jumlah penumpang harian di Bandara Radin Inten Lampung. Namun, dibandingkan dengan model SARIMA, prediksi yang dihasilkan oleh LSTM cenderung lebih halus dan stabil,

terutama dalam menghadapi lonjakan atau penurunan jumlah penumpang yang drastis. Hal ini dikarenakan LSTM lebih unggul dalam menangani pola non-linier dengan mempertimbangkan hubungan jangka panjang dalam data deret waktu. Berikut merupakan hasil prediksi jumlah penumpang menggunakan model LSTM, yang divisualisasikan pada Gambar 7.



Gambar 7. Hasil Prediksi Menggunakan Model LSTM

Kemudian Berikut adalah hasil dari prediksi pada jumlah penumpang harian di bandara raden inten Lampung menggunakan Model LSTM pada gambar 8

```
# Tampilkan hasil prediksi dalam bentuk tabel
print(pred_df)
```

	Tanggal	Jumlah_Penumpang
0	2025-01-01	3649
1	2025-01-02	3673
2	2025-01-03	3698
3	2025-01-04	3723
4	2025-01-05	3747
..	...	...
360	2025-12-27	4977
361	2025-12-28	4977
362	2025-12-29	4977
363	2025-12-30	4977
364	2025-12-31	4977

[365 rows x 2 columns]

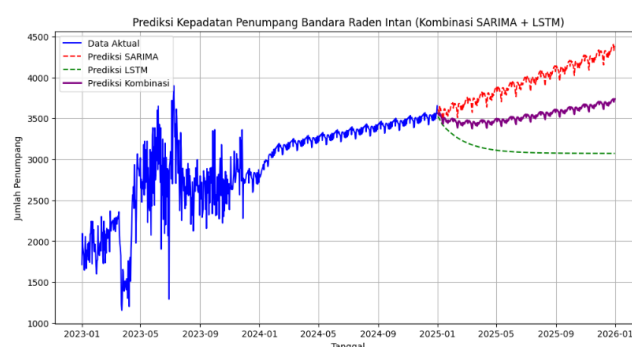
Gambar 8 Hasil Prediksi Jumlah Penumpang menggunakan LSTM

C. Kombinasi Model SARIMA dan LSTM

Kombinasi model dilakukan untuk meningkatkan akurasi prediksi dengan memanfaatkan kelebihan masing-masing model. Model SARIMA unggul dalam menangkap pola linier musiman, sedangkan LSTM mampu mempelajari hubungan non-linier dalam data deret waktu. Dengan menggabungkan keduanya, kita dapat memperoleh model prediksi yang lebih akurat dan adaptif terhadap pola data yang kompleks. Untuk meningkatkan akurasi, dilakukan kombinasi model dengan pendekatan residual learning, di mana prediksi SARIMA digunakan sebagai dasar, lalu residual-nya dimodelkan menggunakan LSTM. Langkah-langkah kombinasi:

- Prediksi awal menggunakan SARIMA untuk mendapatkan pola linier.
- Hitung residual, yaitu selisih antara data aktual dengan hasil prediksi SARIMA.
- Latih model LSTM untuk memprediksi residual, sehingga model dapat menangkap pola non-linier dalam error SARIMA.
- Gabungkan hasil prediksi SARIMA dan LSTM untuk memperoleh hasil akhir yang lebih akurat.

Pada hasil Penggabungan antara SARIMA dan LSTM mendapat hasil yang lebih konsisten dan mampu menangkap dan mengikuti pola data pada jumlah penumpang pada dataset. Berikut adalah gambar hasil prediksi SARIMA dan LSTM pada gambar 9.



Gambar 9. Hasil Prediksi SARIMA dan LSTM

Model SARIMA dilatih menggunakan dataset historis, kemudian digunakan untuk memprediksi jumlah penumpang untuk periode mendatang (Januari - Desember 2025). Hasil prediksi SARIMA menangkap pola musiman dengan baik, namun masih terdapat selisih atau error dibandingkan dengan data aktual.

Model LSTM dilatih menggunakan residual sebagai target output, dengan input berupa data deret waktu sebelumnya. Dengan demikian, LSTM akan belajar untuk mengenali pola dalam residual yang tidak dapat ditangkap oleh SARIMA. Kemudian Model dilatih menggunakan Mean Squared Error (MSE) sebagai fungsi loss hingga konvergen.

Hasil prediksi model kombinasi berhasil menangkap tren musiman yang terjadi setiap bulannya, seperti peningkatan jumlah penumpang pada musim liburan dan akhir pekan. Model kombinasi mampu mengatasi error yang sebelumnya terjadi pada prediksi SARIMA, terutama dalam periode di mana jumlah penumpang meningkat atau menurun secara tiba-tiba. Dibandingkan dengan model LSTM murni, hasil prediksi model kombinasi lebih stabil, karena SARIMA memberikan dasar prediksi yang kuat, sedangkan LSTM hanya memperbaiki bagian yang tidak bisa ditangkap SARIMA. Berikut adalah hasil Evaluasi Model Kombinasi SARIMA dan LSTM pada gambar 10.

```
# Cetak hasil evaluasi
print("Evaluasi Model Kombinasi SARIMA + LSTM:")
print(f"Mean Absolute Error (MAE): {mae:.2f}")
print(f"Mean Squared Error (MSE): {mse:.2f}")
print(f"Root Mean Squared Error (RMSE): {rmse:.2f}")
print(f"R-squared (R2): {r2:.2f}")
print(f"Mean Absolute Percentage Error (MAPE): {mape:.2f}%")
```

```
Evaluasi Model Kombinasi SARIMA + LSTM:
Mean Absolute Error (MAE): 571.90
Mean Squared Error (MSE): 361815.31
Root Mean Squared Error (RMSE): 601.51
R-squared (R2): -5.75
Mean Absolute Percentage Error (MAPE): 14.27%
```

Gambar 10. Hasil Evaluasi Model Kombinasi SARIMA dan LSTM

Pada hasil evaluasi Model Kombinasi SARIMA dan LSTM menunjukkan Mean absolute Percentage Error (MAPE) sebesar 14.27 %. Hal ini terjadi karena Error Accumulation yaitu ketika residual error dari SARIMA yang tidak bersifat acak (white noise) digunakan untuk melatih LSTM. Residual dari SARIMA ini mengandung noise atau kesalahan yang tidak dapat diprediksi atau dijelaskan oleh model SARIMA. Jika LSTM dilatih untuk memprediksi residual yang mengandung noise, LSTM dapat overfit pada noise tersebut, mempelajari pola acak yang tidak ada hubungannya dengan pola sebenarnya dalam data. Akibatnya, LSTM tidak efektif dalam menangani residual tersebut, dan model kombinasi yang dihasilkan tidak lebih baik dari model individu (SARIMA atau LSTM). Sebagai contoh, pada running program, MAPE model kombinasi SARIMA-LSTM mencapai 14.27%

D. Pemodelan *RandomForest* dan *XGBoost*

Untuk memperluas cakupan evaluasi dan membandingkan performa pendekatan time series dengan metode machine learning konvensional, penelitian ini juga menerapkan model *Random Forest* dan *XGBoost*. Keduanya merupakan algoritma berbasis decision tree yang banyak digunakan dalam berbagai tugas regresi karena kemampuannya menangani non-linearitas dan interaksi antar fitur. Dataset dibagi menjadi data training dan testing tanpa dilakukan pengacakan (shuffle=False) agar urutan waktu tetap terjaga. Fitur-fitur yang digunakan dalam model meliputi suhu minimum, suhu maksimum, curah hujan, jumlah penerbangan, hari libur, hari dalam minggu (day_of_week), bulan (month), dan moving average 7 hari terakhir (ma_7). Semua fitur dinormalisasi menggunakan Min-Max Scaling untuk meningkatkan efisiensi training.

Model *Random Forest* dibangun dengan 100 pohon keputusan (estimators), dan model *XGBoost* menggunakan parameter `n_estimators=100`, `learning_rate=0.1`, dan `max_depth=5`. Model dilatih pada data tahun 2023–2024 dan dievaluasi untuk memprediksi jumlah penumpang selama periode 1 Januari 2025 hingga 31 Desember 2025. Hasil evaluasi menunjukkan bahwa *XGBoost* memiliki performa lebih baik dibanding *Random Forest*, dengan error yang lebih rendah pada semua metrik yang dapat dilihat pada table 1.

Tabel 1
Hasil Evaluasi

No	Nama Model	MAE	RMSE	MAPE
1	Random Forest	302.36	302.53	8.51%
2	XGBoost	207.41	207.41	5.84%

Pada table 1 terlihat *XGBoost* memiliki performa yang lebih baik dari *Random Forest*. Namun demikian, kedua model cenderung mengulang pola historis secara mingguan tanpa menangkap tren jangka panjang atau pola musiman secara eksplisit. Hal ini disebabkan oleh karakteristik model yang tidak mempertimbangkan urutan waktu secara langsung seperti SARIMA atau LSTM. Berikut adalah visualisasi hasil *Random Forest* dan *XGBoost* yang dapat dilihat pada gambar 11.



Gambar 11. Visualisasi Hasil *Random Forest* dan *XGBoost*

Pada visualisasi prediksi dalam gambar 10 menunjukkan kurva hasil prediksi yang relatif konstan dan kurang adaptif terhadap fluktuasi besar. Meskipun begitu, kedua model dapat dijadikan baseline atau komponen dalam pendekatan ensemble yang lebih kompleks, khususnya untuk mengolah fitur eksogen atau memperkuat prediksi pada periode dengan anomali.

E. Pembahasan Hasil

Sebagai upaya untuk memperkuat kontribusi pada penelitian ini, dilakukan perbandingan tambahan dengan dua model machine learning populer, yaitu *Random Forest* dan *XGBoost*. Kedua model ini dilatih menggunakan fitur-fitur eksternal seperti suhu, curah hujan, jumlah penerbangan, hari libur, serta fitur waktu. Data prediksi ditujukan untuk periode Januari 2025 hingga Desember 2025. Hasil evaluasi untuk masing-masing model ditampilkan pada table 2 sebagai berikut:

No	Nama Model	MAE	RMSE	MAPE
1	Sarima	124.37	152.35	3.81%
2	LSTM	228.70	234.47	6.81%
3	Random Forest	302.36	302.53	8.51%
4	XGBoost	207.41	207.41	5.84%
5	Sarima + LSTM	571.90	601.51	14.27%

Dari hasil evaluasi, terlihat bahwa model SARIMA memberikan performa terbaik dalam memprediksi tren jangka panjang dan pola musiman. LSTM unggul dalam menangani

fluktuasi non-linier, tetapi lebih rentan terhadap overfitting. Sementara itu, model Random Forest dan XGBoost menghasilkan prediksi yang cukup baik, tetapi cenderung mengulang pola mingguan akibat tidak memperhitungkan urutan waktu secara eksplisit.

Model kombinasi SARIMA-LSTM menunjukkan performa yang kurang baik berdasarkan metrik evaluasi karena akumulasi error dari dua tahap prediksi (SARIMA dan LSTM pada residual). Hal ini terjadi karena adanya akumulasi error residual dari model SARIMA yang tidak dapat ditangkap oleh LSTM. LSTM yang berlebihan dalam mempelajari noise dalam residual dapat menyebabkan overfitting, yang pada akhirnya menghasilkan prediksi yang kurang baik. Namun, dari sisi kemampuan menangkap pola, model ini menunjukkan potensi yang besar untuk dikembangkan lebih lanjut dengan tuning parameter dan strategi ensemble yang lebih kompleks.

Kemudian pada temuan penelitian ini mendukung studi sebelumnya oleh [3] dan [4] yang menunjukkan bahwa model SARIMA mampu menangkap pola musiman secara efektif dan menghasilkan performa yang baik dalam data time series dengan tren yang stabil. Model LSTM yang digunakan juga selaras dengan hasil penelitian [8], di mana LSTM unggul dalam mendeteksi pola non-linier serta fluktuasi tajam. Namun, berbeda dengan temuan [14] yang melaporkan bahwa pendekatan hybrid SARIMA-LSTM memberikan peningkatan akurasi, pada penelitian ini justru ditemukan bahwa kombinasi tersebut menghasilkan akurasi lebih rendah akibat akumulasi error residual, menunjukkan pentingnya penyesuaian struktur model dan parameter terhadap karakteristik data lokal. Temuan ini memperluas diskursus bahwa keberhasilan model hybrid sangat kontekstual, dan tidak selalu memberikan hasil lebih baik tanpa strategi penggabungan yang optimal.

F. Visualisasi Hasil Prediksi

Berikut ini adalah perbandingan antara data aktual dan hasil prediksi untuk model-model yang diuji dalam penelitian ini:

1. Perbandingan SARIMA dengan data aktual menunjukkan hasil yang cukup baik dalam menangkap pola musiman, meskipun tidak bisa menangani lonjakan tiba-tiba.
2. Perbandingan LSTM dengan data aktual menunjukkan fluktuasi yang lebih halus, meskipun model ini gagal dalam menangkap lonjakan yang tajam.
3. Perbandingan antara SARIMA dan LSTM dalam kombinasi model menunjukkan hasil yang lebih stabil dengan beberapa kesalahan residual yang berhasil dikoreksi.
4. Perbandingan dengan model Random Forest dan XGBoost menunjukkan bahwa meskipun keduanya dapat memberikan prediksi yang stabil, XGBoost menunjukkan kinerja yang lebih baik dibandingkan Random Forest.

IV. KESIMPULAN

Penelitian ini bertujuan untuk memprediksi kepadatan penumpang harian di Bandara Radin Inten Lampung dengan memanfaatkan berbagai pendekatan pemodelan deret waktu, baik berbasis statistik maupun machine learning. Model utama yang dikembangkan adalah kombinasi Seasonal AutoRegressive Integrated Moving Average (SARIMA) dan Long Short-Term Memory (LSTM), dengan penambahan eksperimen terhadap model Random Forest dan XGBoost untuk memperkuat validitas hasil dan membandingkan performa lintas pendekatan. Dari hasil evaluasi terhadap masing-masing model, diperoleh beberapa kesimpulan sebagai berikut:

1. Model SARIMA menunjukkan performa yang baik dalam menangkap pola musiman dan tren linier yang terjadi dalam data jumlah penumpang harian. Model ini menghasilkan evaluasi yang cukup akurat dengan RMSE sebesar 152.35, MAE sebesar 124.37, dan MAPE sebesar 3.81%. Namun demikian, SARIMA memiliki keterbatasan dalam merespons perubahan mendadak atau variasi non-linier yang tidak berulang secara musiman.
2. Model LSTM, sebagai representasi metode deep learning, berhasil mempelajari relasi jangka panjang dan pola non-linier yang kompleks. Model ini memberikan hasil yang kompeten dengan RMSE sebesar 234.47, MAE sebesar 228.70, dan MAPE sebesar 6.81%. Kelebihan LSTM terletak pada fleksibilitasnya dalam mengenali fluktuasi ekstrem, tetapi kelemahannya adalah sensitivitas terhadap overfitting, terutama ketika panjang data pelatihan tidak memadai atau pemilihan parameter tidak optimal.
3. Kombinasi SARIMA dan LSTM dirancang untuk menggabungkan kelebihan masing-masing model, dengan SARIMA menangkap pola linier-musiman sebagai baseline, dan LSTM digunakan untuk mempelajari sisa kesalahan (residual) yang tidak dapat dijelaskan oleh SARIMA. Secara teoritis, pendekatan ini seharusnya memberikan akurasi terbaik. Namun, hasil evaluasi menunjukkan bahwa performa model kombinasi justru menurun dibanding model individual, dengan RMSE sebesar 601.51, MAE sebesar 571.90, dan MAPE sebesar 14.27%. Hal ini disebabkan kemungkinan adanya ketidaksesuaian antara distribusi residual dan arsitektur LSTM yang digunakan, serta potensi overfitting pada tahap koreksi residual.
4. Model Random Forest dan XGBoost ditambahkan sebagai pembanding dari ranah machine learning klasik. Random Forest memberikan hasil prediksi dengan RMSE sebesar 302.53, MAE sebesar 302.36, dan MAPE sebesar 8.51%, sementara XGBoost menunjukkan performa lebih baik dengan RMSE sebesar 216.12, MAE sebesar 207.41, dan

MAPE sebesar 5.84%. Meskipun keduanya tidak mampu menangkap pola musiman secara eksplisit, model XGBoost khususnya dapat memberikan estimasi yang cukup stabil dan efisien ketika digunakan dengan fitur-fitur eksternal yang representative.

5. Secara keseluruhan, model SARIMA tetap menjadi pilihan terbaik untuk prediksi jangka panjang dengan pola musiman yang kuat. LSTM efektif untuk menangani perubahan mendadak, tetapi membutuhkan konfigurasi yang cermat. Kombinasi SARIMA dan LSTM, meskipun secara teori menjanjikan, masih perlu perbaikan dalam teknik penggabungan dan penanganan residual. Sementara itu, model berbasis pohon keputusan seperti XGBoost dapat berfungsi sebagai alternatif cepat dan ringan, terutama ketika ketersediaan data temporal tidak lengkap.

DAFTAR PUSTAKA

- [1] F. Febriana, Y. Salim, and H. Darwis, "Implementasi Analisis Volume Capacity Ratio Untuk Memprediksi Kepadatan Lalu Lintas Di Kota Makassar," *Bul. Sist. Inf. dan Teknol. Islam*, vol. 3, no. 3, pp. 219–224, 2022, doi: 10.33096/busiti.v3i3.864.
- [2] A. Mahfud Al, D. Kurniasari, Mustofa Usman, and P. Pesawat, "Peramalan Data Time Series Seasonal Menggunakan Metode Analisis Spektral," *J. Siger Mat.*, vol. 1, no. 1, pp. 10–15, 2020.
- [3] D. C. W. Wu, L. Ji, K. He, and K. F. G. T'so, "Forecasting Tourist Daily Arrivals With A Hybrid Sarima-Lstm Approach," *J. Hosp. Tour. Res.*, vol. 45, no. 1, pp. 52–67, 2021, doi: 10.1177/1096348020934046.
- [4] W. Gu, B. Guo, Z. Zhang, and H. Lu, "Civil Aviation Passenger Traffic Forecasting: Application and Comparative Study of the Seasonal Autoregressive Integrated Moving Average Model and Backpropagation Neural Network," *Sustain.*, vol. 16, no. 10, 2024, doi: 10.3390/su16104110.
- [5] D. Kurniasari, A. Dafa Salsabila, and M. Usman, "Enhancing Weather Forecasting in Bandar Lampung: A Hybrid SARIMA-LSTM Approach," *J. Teor. dan Apl. Mat.*, vol. 9, no. 1, p. 206, 2025, [Online]. Available: <http://journal.ummat.ac.id/index.php/jtam>
- [6] Q. H. Do, S. K. Lo, J. F. Chen, C. L. Le, and L. H. Anh, "Forecasting air passenger demand: A comparison of LSTM and SARIMA," *J. Comput. Sci.*, vol. 16, no. 7, pp. 1063–1084, 2020, doi: 10.3844/JCSP.2020.1063.1084.
- [7] F. I. Durrah, Y. Yulia, T. P. Parhusip, and A. Rusyana, "Peramalan Jumlah Penumpang Pesawat Di Bandara Sultan Iskandar Muda Dengan Metode SARIMA (Seasonal Autoregressive Integrated Moving Average)," *J. Data Anal.*, vol. 1, no. 1, pp. 1–11, 2018, doi: 10.24815/jda.v1i1.11847.
- [8] M. Davi and E. Winarko, "Rancang Bangun Aplikasi Peramalan Jumlah Penumpang Menggunakan Long Short-Term Memory (LSTM)," *Infotekmesin*, vol. 14, no. 2, pp. 303–310, 2023, doi: 10.35970/infotekmesin.v14i2.1911.
- [9] A. Boateng, C. Anum Adams, and E. Kofi Akowuah, "Estimating Passenger Demand Using Machine Learning Models: A Systematic Review," *E3S Web Conf.*, vol. 418, 2023, doi: 10.1051/e3sconf/202341803002.
- [10] B. Jiang, G. Ding, J. Fu, J. Zhang, and Y. Zhang, "An Overview of Demand Analysis and Forecasting Algorithms for the Flow of Checked Baggage among Departing Passengers," *Algorithms*, vol. 17, no. 5, pp. 1–16, 2024, doi: 10.3390/a17050173.
- [11] V. Singh, S. K. Sahana, and V. Bhattacharjee, "A novel CNN-GRU-LSTM based deep learning model for accurate traffic prediction," *Discov. Comput.*, vol. 28, no. 1, 2025, doi: 10.1007/s10791-025-09526-0.
- [12] A. K. Aggarwal, "Learning Texture Features from GLCM for Classification of Brain Tumor MRI Images using Random Forest Classifier," *Wseas Trans. Signal Process.*, vol. 18, no. May, pp. 60–63, 2022, doi: 10.37394/232014.2022.18.8.
- [13] W. H. Bangyal et al., "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches," *Comput. Math. Methods Med.*, vol. 2021, 2021, doi: 10.1155/2021/5514220.
- [14] A. S. B. Karno, "Analisis Data Time Series Menggunakan LSTM (Long Short Term Memory) Dan ARIMA (Autoregressive Integrated Moving Average) Dalam Bahasa Python," *Ultim. InfoSys J. Ilmu Sist. Inf.*, vol. XI, no. 1, pp. 1–7, 2020.
- [15] N. N. Pandika Pinata, I. M. Sukarsa, and N. K. Dwi Rusjayanthi, "Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 8, no. 3, p. 188, 2020, doi: 10.24843/jim.2020.v08.i03.p04.
- [16] A. Lisanthoni, F. I. Sari, E. L. Gunawan, and C. A. Adhigadany, "Model Prediksi Kepadatan Lalu Lintas: Perbandingan Algoritma Random Forest dan XGBoost," *Pros. Semin. Nas. Sains Data*, vol. 3, no. 1, pp. 296–303, 2023, doi: 10.33005/senada.v3i1.126.