

ANALISIS PERAMALAN JUMLAH KUNJUNGAN WISATAWAN DOMESTIK MENGGUNAKAN MODEL LONG SHORT-TERM MEMORY (LSTM) DI KOTA PANGKAL PINANG

Muhammad Irfan Hidayat¹, Mei Dita Kumala²

Program Studi Matematika, Universitas Bangka Belitung, Bangka, Indonesia ¹

Bangka, Indonesia²

Email: irfan-hidayat@ubb.ac.id¹, meiditakumala@gmail.com²

Corresponding author: Muhammad Irfan Hidayat, Email. irfan-hidayat@ubb.ac.id

Abstrak. Sektor pariwisata berperan penting sebagai pendorong utama pertumbuhan ekonomi suatu daerah, dengan jumlah wisatawan menjadi indikator yang patut diperhatikan. Peningkatan kunjungan wisatawan berimplikasi pada meningkatnya potensi risiko, sehingga diperlukan upaya peramalan yang tepat. Dalam penelitian ini, metode *Long Short-Term Memory (LSTM)* diterapkan untuk memprediksi jumlah perjalanan wisatawan domestik di Kota Pangkal Pinang dan dibandingkan dengan metode ARMA serta SARIMA. Hasil perbandingan menunjukkan bahwa metode LSTM menghasilkan nilai RMSE sebesar 32.431,219, yang lebih rendah dibandingkan ARMA (41.273,347) dan SARIMA (101.884,554). Hasil studi menunjukkan bahwa LSTM memiliki performa prediksi yang lebih efektif, sehingga metode ini lebih direkomendasikan.

Kata Kunci: Peramalan, Long Short-Term Memory (LSTM), Jaringan Saraf Tiruan, Root Mean Square Error (RMSE).

Abstract. Tourism is one of the main pillars of economic growth in a region, with the number of tourists being a critical factor to consider. As the number of visitors increases, so do the potential risks, making forecasting essential. This study applied the Long Short-Term Memory (LSTM) method to forecast the number of domestic tourist trips in Pangkal Pinang City. The results were then compared with two alternative methods, namely ARMA and SARIMA. The prediction using the LSTM method yielded an RMSE value of 32431.219245444987. For comparison, the ARMA and SARIMA methods were also applied, resulting in RMSE values of 41273.34779355672 and 101884.5544766975. The findings show that LSTM delivers superior predictive performance, thereby making it the more recommended method.

Keywords: Forecasting, Long Short-Term Memory (LSTM), Neural Networks, RMSE.

A. Pendahuluan

Pendahuluan Sektor pariwisata telah menjadi salah satu pilar utama dalam pembangunan ekonomi (Indrizal et al., 2024). Hal ini dapat dipengaruhi oleh jumlah perjalanan wisatawan domestik. Perjalanan wisatawan domestik didefinisikan sebagai pergerakan warga negara Indonesia untuk tujuan wisata di dalam wilayah Indonesia. Peningkatan frekuensi perjalanan wisatawan domestik memiliki korelasi langsung dengan penguatan ekonomi melalui sektor pariwisata. Jumlah kunjungan wisatawan domestik berfluktuasi setiap tahunnya di berbagai provinsi di Indonesia, mengalami peningkatan maupun penurunan (Riski et al., 2024). Di antara provinsi-provinsi tersebut, Provinsi Kepulauan Bangka Belitung menunjukkan pertumbuhan yang berkelanjutan dalam jumlah kunjungan wisatawan domestik.

Badan Pusat Statistik (BPS) Provinsi Kepulauan Bangka Belitung melaporkan bahwa jumlah wisatawan domestik yang berkunjung ke provinsi tersebut pada periode Januari hingga Juni 2025 mencapai 2,26 juta orang, meningkat sebesar 52,70 persen dibandingkan dengan 1,48



juta orang pada periode yang sama tahun 2024. Kepala BPS Provinsi Kepulauan Bangka Belitung, Toto Haryanto Silitonga, menyatakan bahwa jumlah perjalanan wisatawan domestik berdasarkan tujuan pada Juni 2025 tercatat sebanyak 431,01 ribu perjalanan, atau meningkat 12,33 persen dibandingkan Mei 2025 dan naik 82,96 persen dibandingkan Juni 2024. Ia juga menambahkan bahwa pada Juni 2025, wisatawan domestik sebagian besar berkunjung ke Kota Pangkal Pinang, Kabupaten Bangka, dan Kabupaten Bangka Tengah, yang secara keseluruhan menyumbang sekitar 66,63 persen dari total perjalanan wisatawan domestik ke Provinsi Kepulauan Bangka Belitung (BPS, 2025).

Merujuk pada publikasi yang disampaikan oleh Badan Pusat Statistik (BPS) Provinsi Kepulauan Bangka Belitung, pemerintah provinsi tersebut telah berhasil mengembangkan dan mengelola sektor pariwisata secara efektif sehingga mampu menarik minat wisatawan domestik. Namun demikian, peningkatan jumlah pengunjung menjadi tantangan tersendiri bagi otoritas setempat. Apabila arus wisatawan tidak dikelola dengan baik, hal ini dapat menimbulkan berbagai permasalahan, seperti meningkatnya risiko kecelakaan serta jumlah wisatawan yang melebihi kapasitas wilayah (Manullang et al., 2020). Pemahaman terhadap pola kunjungan wisatawan menjadi hal yang penting, karena memungkinkan identifikasi tren kedatangan wisatawan pada destinasi tertentu (Sugiartawan et al., 2018). Permasalahan ini menjadi dasar penelitian yang bertujuan untuk melakukan peramalan terhadap jumlah perjalanan wisatawan domestik, dengan fokus khusus pada Kota Pangkal Pinang.

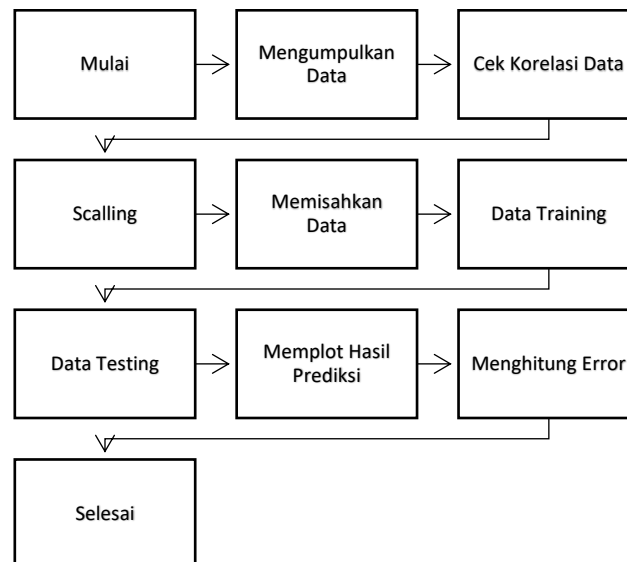
Sebelumnya, E. Marcellina dkk. (2024) telah melakukan penelitian mengenai peramalan jumlah perjalanan di Kabupaten Belitung dengan menggunakan Metode Monte Carlo (Marcelina et al., 2024). Penelitian tersebut berfokus pada salah satu wilayah di Provinsi Kepulauan Bangka Belitung, yaitu Kabupaten Belitung. Berbeda dengan penelitian tersebut, studi ini berfokus pada kota utama Provinsi Kepulauan Bangka Belitung, yakni Kota Pangkal Pinang. Dengan menggunakan data yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Kepulauan Bangka Belitung, peneliti mengumpulkan informasi mengenai perjalanan wisatawan domestik di wilayah tersebut dari Januari 2019 hingga Juli 2025. Data yang dihimpun menunjukkan adanya korelasi antara suatu kejadian dengan kejadian sebelumnya.

Peramalan bertujuan untuk mempelajari tren data berdasarkan pola yang diamati dari data yang dikumpulkan secara berkala dari waktu ke waktu. Metode yang digunakan dalam penelitian ini merupakan metode peramalan lanjutan, yaitu model *Long Short-Term Memory (LSTM)*, yang merupakan varian yang lebih kuat dari model *Recurrent Neural Network (RNN)* (Sugiartawan et al., 2018). Penggunaan metode LSTM dalam penelitian ini dianggap sangat relevan karena metode tersebut dapat diterapkan secara efektif untuk memprediksi jumlah kunjungan wisatawan domestik di Kota Pangkal Pinang, mengingat data yang digunakan memiliki keterkaitan antarwaktu. Selain itu, penelitian ini juga akan membandingkan metode LSTM dengan metode lainnya, seperti ARMA dan SARIMA.

B. Metodologi Penelitian

Proses peramalan dalam penelitian ini terdiri atas beberapa tahapan yang digambarkan dalam bagan alur penelitian yang disajikan pada **Gambar 1**.





Gambar 1. Bagan Alur Penelitian

Alur proses dalam penelitian ini terdiri atas beberapa tahapan, antara lain:

- Pengumpulan data dari situs resmi Badan Pusat Statistik (BPS) Provinsi Kepulauan Bangka Belitung.
- Seleksi data relevan menggunakan peta korelasi (correlation heatmap) untuk mengidentifikasi parameter optimal dalam pengembangan model.
- Penskalaan data (data scaling) menggunakan metode Min-Max Scaling, yang juga dikenal sebagai normalisasi data.
- Pembagian data menjadi dua bagian dengan rasio 80% data latih dan 20% data uji.
- Pelatihan data untuk memperoleh model peramalan.
- Perbandingan antara hasil prediksi yang diperoleh dari model dengan data aktual (data uji).
- Pembuatan plot yang menampilkan perbandingan antara data aktual dan data hasil prediksi.

1. Heatmap Korelasi

Korelasi antarfitur dalam himpunan data dianalisis menggunakan diagram heatmap. Heatmap merupakan alat visualisasi yang digunakan untuk menampilkan matriks korelasi, yang menggambarkan kekuatan serta arah hubungan linear antarvariabel numerik dalam himpunan data (Syarifudin & Hapsoro, 2025).

2. Min-Max Scaling

Salah satu teknik penskalaan data yang mentransformasikan data secara linear adalah *Min-Max Scaling*. Hasil dari proses penskalaan ini berupa data yang diubah ke dalam rentang skala antara 0 dan 1. Rumus untuk melakukan penskalaan *Min-Max* ditunjukkan pada Persamaan (1) di bawah ini. Dalam persamaan tersebut, x_i merepresentasikan data yang akan diskalakan, x' menunjukkan hasil penskalaan, $\min(x)$ merupakan nilai minimum suatu fitur, dan $\max(x)$ adalah nilai maksimumnya (Allorerung et al., 2024). Penskalaan *Min-Max* dapat dinyatakan sebagai berikut:

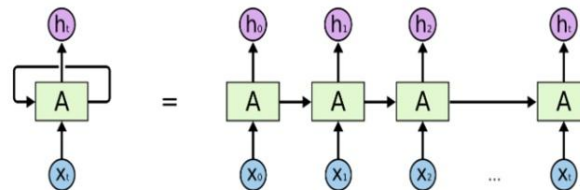
$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$



3. Recurrent Neural Network (RNN)

Salah satu jenis algoritma machine learning yang berbasis pada pemrosesan informasi dari data dan termasuk dalam kategori *deep learning* adalah *Recurrent Neural Network (RNN)*. Algoritma RNN mampu mempertahankan pola informasi masa lalu melalui proses iterasi dalam arsitekturnya, sehingga informasi tersebut dapat tersimpan. RNN merupakan pengembangan dari *Artificial Neural Networks (ANN)* dan memiliki arsitektur yang sangat mirip dengan Multilayer Perceptron (MLP) (Nurashila et al., 2023).

Secara mendasar, *Recurrent Neural Network (RNN)* memiliki parameter yang serupa dengan jaringan saraf standar, dengan perbedaan utama terletak pada konsep dasarnya. Lapisan tersembunyi (hidden layer) pada RNN dari langkah waktu (timestep) sebelumnya digunakan sebagai informasi atau masukan pada proses berikutnya. Lapisan tersembunyi ini menyimpan informasi dari tahap awal ekstraksi fitur, sehingga ketika digunakan sebagai input pada tahap selanjutnya, lapisan tersebut membawa memori dari input sebelumnya [9].

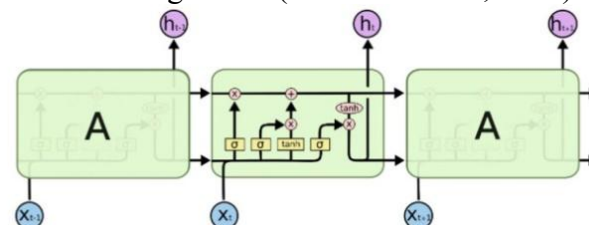


Gambar 2. Recurrent Neural Network

Gambar 2 di atas menggambarkan proses kerja *Recurrent Neural Network (RNN)*. Notasi X_t merepresentasikan input pada waktu ke- t , sedangkan keluaran dari proses RNN dinyatakan dengan notasi h_t . Setelah proses RNN diuraikan (unrolled), keseluruhan jaringan dapat terlihat secara menyeluruh. Proses RNN dimulai dengan mengambil X_0 sebagai input awal untuk menghasilkan h_0 , yang kemudian secara bersamaan menerima X_1 sebagai input untuk langkah berikutnya. Dengan demikian, h_0 dan X_1 digunakan sebagai input untuk tahap selanjutnya, dan pola ini terus berlanjut. Mekanisme tersebut memungkinkan RNN untuk mengingat konteks selama proses pelatihan. RNN menjalankan tugas yang sama secara iteratif, dengan hasil keluarannya bergantung pada perhitungan sebelumnya (Nurashila et al., 2023).

4. Long Short-Term Memory (LSTM)

Salah satu bentuk lanjutan dari arsitektur *Recurrent Neural Network (RNN)* adalah *Long Short-Term Memory (LSTM)*, yang diperkenalkan oleh Hochreiter dan Schmidhuber pada tahun 1997. Dibandingkan dengan RNN konvensional, LSTM memiliki keunggulan signifikan melalui mekanisme *memory* dan *forget*, yang memungkinkan keluaran digunakan kembali sebagai masukan serta mendukung pemrosesan informasi yang lebih akurat. Selain itu, LSTM mampu mempertahankan sinyal kesalahan selama proses *backpropagation*, sehingga mencegah terjadinya akumulasi kesalahan seiring waktu (Nurashila et al., 2023).



Gambar 3. Long Short-Term Memory (LSTM)

Sebagaimana terlihat pada gambar sebelumnya, struktur dasar *Long Short-Term Memory (LSTM)* mencakup tiga gerbang utama: *forget gate*, *input gate*, dan *output gate*. Peran *forget gate* adalah menyaring informasi yang tidak relevan menggunakan fungsi aktivasi sigmoid. Jika

hasil keluarannya bernilai 1, semua data akan dipertahankan; sebaliknya, nilai 0 menunjukkan bahwa data tersebut diabaikan (Nurashila et al., 2023).

5. *Root Mean Squared Error (RMSE)*

Penelitian ini menerapkan *Root Mean Squared Error (RMSE)* sebagai ukuran untuk menilai seberapa besar perbedaan antara nilai hasil prediksi dan nilai aktual (data uji). Persamaan yang digunakan untuk menghitung RMSE disajikan sebagai berikut:

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (X_t - Y_t)^2}{N}} \quad (2)$$

Dalam persamaan tersebut, X_t merepresentasikan nilai data uji pada waktu ke- t , yaitu data aktual yang digunakan sebagai acuan untuk mengevaluasi kinerja model prediksi. Sementara itu, Y_t merupakan notasi yang menggambarkan nilai hasil prediksi pada waktu ke- t , yaitu output yang dihasilkan oleh model berdasarkan proses pembelajaran yang telah dilakukan. Adapun menunjukkan jumlah keseluruhan data hasil prediksi yang digunakan dalam perhitungan, yang berfungsi sebagai dasar untuk menentukan nilai rata-rata penyimpangan antara data prediksi dan data aktual (Khaira et al., 2020).

6. *Autoregressive Moving Average (ARMA)*

Autoregressive Moving Average (ARMA) merupakan salah satu metode yang paling banyak digunakan dalam analisis deret waktu. Model ini terdiri atas dua komponen, yaitu *Autoregressive (AR)* yang mengasumsikan adanya pengaruh nilai masa lalu terhadap nilai saat ini, serta *Moving Average (MA)* yang mengasumsikan bahwa nilai saat ini ditentukan oleh residual dari periode sebelumnya. Deret waktu sendiri merupakan kumpulan observasi yang dicatat secara kronologis. Model ARMA dikembangkan oleh George Edward Pelham Box dan Gwilym Meirion Jenkins pada tahun 1976, sehingga dikenal pula sebagai metode *Box-Jenkins*. Bentuk umum dari model Autoregressive (AR(p)) disajikan sebagai berikut:

$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \dots + b_n y_{t-n} + \epsilon_t \quad (3)$$

Di sisi lain, model *Moving Average (MA)*, yang dinotasikan sebagai MA(q), menggambarkan hubungan antara nilai data saat ini dengan kesalahan prediksi (error term) dari beberapa periode sebelumnya. Model ini berperan penting dalam memperhalus fluktuasi data serta menangkap pola acak yang muncul akibat ketidakpastian atau gangguan eksternal dalam proses pengamatan. Dengan demikian, nilai pada periode sekarang merupakan hasil dari kombinasi linier residual masa lalu dengan koefisien tertentu. Bentuk matematis umum dari model MA dengan orde q dapat dinyatakan sebagai berikut:

$$y_t = a_0 - a_1 \epsilon_{t-1} + a_2 \epsilon_{t-2} + \dots + a_n \epsilon_{t-n} + \epsilon_t \quad (4)$$

Keterangan:

y_t : nilai data pada periode waktu ke- t

$y_{(t-1)}, y_{(t-2)}, \dots, y_{(t-n)}$: nilai data pada periode waktu sebelumnya

ϵ_t : nilai kesalahan (*error*) pada periode waktu ke- t

$\epsilon_{(t-1)}, \epsilon_{(t-2)}, \dots, \epsilon_{(t-n)}$: nilai kesalahan pada periode waktu sebelumnya

$b_0, b_1, \dots, b_n$: koefisien konstan pada komponen *Autoregressive (AR)*

$a_0, a_1, \dots, a_n$: koefisien konstan pada komponen *Moving Average (MA)* (Melyani et al., 2021).

7. *Seasonal Autoregressive Integrated Moving Average (SARIMA)*

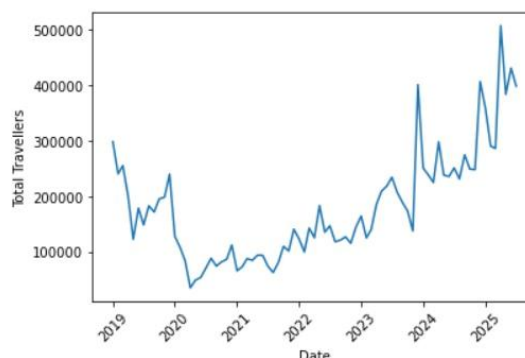


Model *Seasonal Autoregressive Integrated Moving Average (SARIMA)* merupakan bentuk pengembangan dari model ARIMA yang dirancang untuk menangani data deret waktu yang menunjukkan pola musiman atau berulang secara periodik. Model ini efektif digunakan ketika data menunjukkan fluktuasi yang mengikuti siklus waktu tertentu, misalnya perubahan bulanan atau tahunan. Secara konseptual, SARIMA tersusun atas dua bagian utama, yaitu komponen non-musiman, yang mencerminkan pola umum data tanpa mempertimbangkan efek musiman, serta komponen musiman, yang menggambarkan pola berulang yang muncul pada interval waktu tertentu. Umumnya, model SARIMA ditulis dalam bentuk notasi sebagai berikut:

$$\phi_p(B)\Phi(B^s)(1-B)^d(1-B^s)^D\dot{Z}_t = \theta_q(B)\Theta_Q(B^s)\alpha_t \quad (\text{Marisa et al., 2025}). \quad (5)$$

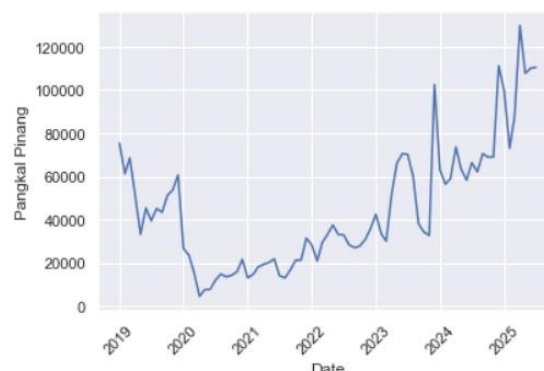
C. Hasil Penelitian dan Pembahasan

Pada bagian ketiga dari makalah ini disajikan hasil prediksi jumlah perjalanan wisatawan domestik di Kota Pangkal Pinang, Provinsi Kepulauan Bangka Belitung, dengan menggunakan metode *Long Short-Term Memory (LSTM)*. Data yang digunakan dalam penelitian ini diperoleh dari Badan Pusat Statistik (BPS), khususnya dataset mengenai jumlah perjalanan wisatawan domestik berdasarkan kota tujuan. Dataset tersebut mencakup data bulanan mengenai perjalanan wisatawan domestik di Provinsi Kepulauan Bangka Belitung, yang dikumpulkan sejak bulan pertama tahun 2019 hingga Juli 2025. Jumlah total wisatawan domestik di provinsi tersebut divisualisasikan dalam Gambar 4.



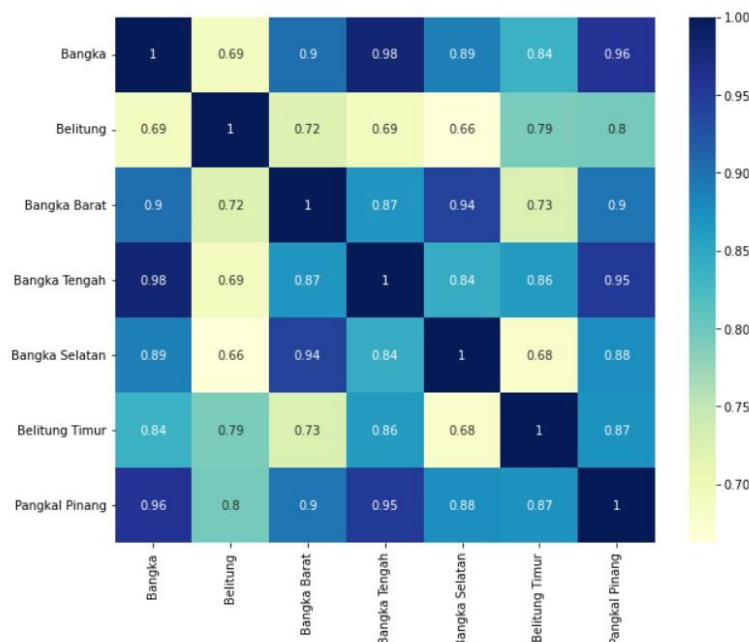
Gambar 4. Jumlah Total Wisatawan di Provinsi Kepulauan Bangka Belitung

Mengingat bahwa penelitian ini berfokus pada Kota Pangkal Pinang, maka data terkait jumlah perjalanan wisatawan domestik di wilayah tersebut ditampilkan pada gambar di bawah ini.



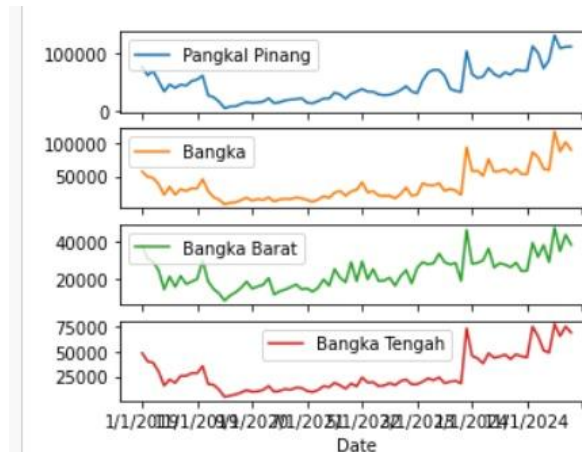
Gambar 5. Jumlah Wisatawan di Kota Pangkal Pinang

Setelah seluruh data berhasil dikumpulkan, penelitian ini hanya memanfaatkan data dari beberapa kabupaten/kota di Provinsi Kepulauan Bangka Belitung yang memiliki korelasi tinggi dengan Kota Pangkal Pinang, dan tidak mencakup seluruh wilayah provinsi. Analisis hubungan antarwilayah dilakukan dengan menggunakan metode heatmap correlation, yang hasil visualisasinya disajikan pada Gambar 6.



Gambar 6. Heatmap Korelasi Jumlah Kunjungan Wisatawan pada Kota-Kota di Provinsi Kepulauan Bangka Belitung

Dalam visualisasi heatmap, gradasi warna menggambarkan besarnya (magnitude) dan arah (direction) hubungan korelasi antarvariabel, sedangkan nilai numerik yang tercantum pada setiap sel menunjukkan koefisien korelasi, dengan rentang nilai antara -1 hingga $+1$. Berdasarkan Gambar 6, dapat diamati bahwa semakin besar nilai koefisien korelasi suatu kabupaten, maka semakin kuat pula tingkat keterkaitannya dengan Kota Pangkal Pinang. Oleh karena itu, dalam penelitian ini hanya kabupaten yang memiliki nilai koefisien korelasi sebesar 0,9 atau lebih tinggi yang dipilih sebagai bagian dari analisis, yaitu Kabupaten Bangka, Bangka Barat, dan Bangka Tengah. Data yang berasal dari kabupaten/kota tersebut kemudian divisualisasikan dalam Gambar 7.



Gambar 7. Plot dari Kota/Kabupaten yang dipilih

Setelah data yang relevan ditentukan untuk digunakan sebagai input dalam proses pemodelan, langkah selanjutnya dalam penelitian ini adalah menerapkan proses normalisasi data, atau data scaling. Tahapan ini memiliki peranan penting dalam mempercepat proses konvergensi algoritmik selama pelatihan model serta meningkatkan kinerja dan ketepatan hasil prediksi. Contoh dari data yang telah dinormalisasi disajikan pada Gambar 8 sebagai ilustrasi dari hasil transformasi tersebut.

```
In [20]: data_scaled
Out[20]: array([[0.44929359, 0.76356032, 0.6015157 ],
 [0.38182775, 0.57289243, 0.48676182],
 [0.36347176, 0.53530257, 0.4706938 ],
 [0.27873809, 0.41633773, 0.34673256],
 [0.13084662, 0.15592733, 0.1554485 ],
 [0.24231153, 0.33057117, 0.24023357],
 [0.13486924, 0.19129526, 0.1934237 ],
 [0.21008598, 0.33870082, 0.28916927],
 [0.18438046, 0.22925108, 0.28943155],
 [0.22149381, 0.26239707, 0.32801414],
 [0.21956723, 0.29619658, 0.32533613],
 [0.34741696, 0.55164031, 0.42358024],
 [0.18350637, 0.26294602, 0.17724524],
 [0.09938813, 0.16084172, 0.16632616],
 [0.06719826, 0.09792184, 0.09857541],
 [0. , 0. , 0. ],
 [0.02381462, 0.07175533, 0.01775213],
 [0.0305398 , 0.11504379, 0.0346208 ],
 [0.06151664, 0.17804209, 0.06013086],
 [0.09303757, 0.26226637, 0.09450319],
 [0.05293624, 0.16570383, 0.07111896],
 [0.07448535, 0.19471964, 0.07680627],
 [0.06206964, 0.21853353, 0.09563513],
 [0.09762211, 0.3121422 , 0.15132106],
 [0.04238467, 0.08547902, 0.07063582],
 [0.06875022, 0.12874134, 0.08304575],
 [0.07776767, 0.15336557, 0.10830734],
 [0.07491348, 0.18679911, 0.09751249],
 [0.09476792, 0.222533 , 0.12757792],
 [0.08462664, 0.15775716, 0.12310538],
 [0.0614988 , 0.16857927, 0.08140305],
 [0.0374166 , 0.12429748, 0.06809586],
```

Gambar 8. Data yang Telah Ditransformasi

Selanjutnya, dataset yang telah melalui tahap normalisasi dibagi menjadi dua bagian, yaitu 80% data untuk pelatihan (*training data*) dan 20% data untuk pengujian (*testing data*). Pembagian ini bertujuan untuk memungkinkan proses pelatihan model dilakukan secara optimal sekaligus menyediakan data independen untuk mengukur kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya. Adapun data uji (*testing data*) yang digunakan dalam penelitian ini disajikan pada bagian berikut.

```
xtest
array([[0.62161511, 0.71397203, 0.59904475],
 [0.44910628, 0.45588812, 0.53698131],
 [0.45818616, 0.51368449, 0.55368433],
 [0.4837133 , 0.49488956, 0.58109936],
 [0.42712904, 0.45868514, 0.51920157],
 [0.48902922, 0.52157888, 0.58613788],
 [0.41673802, 0.40708404, 0.55645896],
 [0.41656855, 0.41194615, 0.5436211 ],
 [0.71955831, 0.79377859, 0.9660556 ],
 [0.64717792, 0.59950333, 0.83168604],
 [0.4903136 , 0.7637433 , 0.63826233],
 [0.4723679 , 0.53096327, 0.6065266 ],
 [1. , 1. , 1. ],
 [0.72640836, 0.67967586, 0.83401894],
 [0.85460594, 0.90560711, 0.9677121 ],
 [0.75021406, 0.7678212 , 0.88205737]])
```

Gambar 9. Data Tes

Data yang telah dibagi sebelumnya kemudian diproses pada tahap pelatihan (*training stage*). Proses pelatihan model dilakukan secara berulang sesuai dengan jumlah *epoch* yang telah ditentukan sebelumnya. Berdasarkan serangkaian percobaan dan pengujian parameter, diperoleh konfigurasi pelatihan yang paling optimal pada 500 *epoch*. Hasil pelatihan menunjukkan bahwa metode *Long Short-Term Memory (LSTM)* memiliki kemampuan yang baik dalam menangani jumlah epoch yang besar serta mampu memproses sejumlah besar sampel tanpa kehilangan akurasi secara signifikan. Setelah proses pelatihan selesai, hasil prediksi diperoleh secara otomatis melalui perintah yang dijalankan menggunakan bahasa pemrograman Python sebagaimana ditunjukkan pada bagian berikut

predictions = model.predict_generator(test_generator)

Perintah yang dijalankan tersebut secara otomatis menghasilkan nilai prediksi, tetapi hasil yang diperoleh masih berupa data yang telah melalui proses skala (normalisasi). Bentuk keluaran ini belum dikembalikan ke skala aslinya, sebagaimana dapat dilihat pada ilustrasi di bawah ini.

```
In [45]: predictions
Out[45]: array([[0.46891683],
                [0.44243327],
                [0.44941568],
                [0.45887542],
                [0.42958835],
                [0.46019587],
                [0.44401142],
                [0.4390166 ],
                [0.47106436],
                [0.47025284],
                [0.4666549 ],
                [0.4618922 ],
                [0.4725565 ],
                [0.47095457],
                [0.47191072]], dtype=float32)
```

Gambar 10. Hasil Prediksi (dalam skala)

Langkah selanjutnya dalam penelitian ini adalah melakukan transformasi balik (*inverse transformation*) terhadap data prediksi yang masih terskala agar dapat dikembalikan ke bentuk aslinya. Proses ini dijalankan dengan menggunakan skrip perintah dalam bahasa pemrograman Python, yang berfungsi untuk menormalisasi ulang hasil prediksi menjadi nilai yang setara dengan data sebenarnya. Adapun perintahnya adalah sebagai berikut:

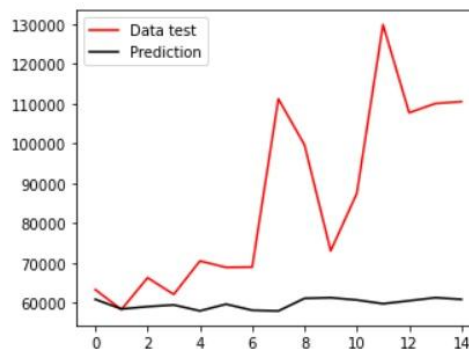
pangkalpinang_prediction = scaler_output.inverse_transform(predictions)

Sebagai hasil dari proses tersebut, data yang telah dikonversi kembali ke bentuk aslinya disajikan pada Gambar 11.

	Bangka	Bangka Barat	Bangka Tengah	Pangkal Pinang_pred	Pangkal Pinang
0	56903.0	26133.0	43590.0	60927.421875	63355.0
1	57921.0	28344.0	44800.0	58514.933594	58380.0
2	60783.0	27625.0	46786.0	59100.882812	66339.0
3	54439.0	26240.0	42302.0	59527.019531	62159.0
4	61379.0	28646.0	47151.0	58034.085938	70547.0
5	53274.0	24266.0	45001.0	59716.996094	68942.0
6	53255.0	24452.0	44071.0	58186.507812	69021.0
7	87225.0	39059.0	74673.0	58016.710938	111175.0
8	79110.0	31627.0	64939.0	61170.460938	99639.0
9	61523.0	37910.0	50927.0	61339.007812	73069.0
10	59511.0	29005.0	48628.0	60761.335938	87512.0
11	118667.0	46948.0	77132.0	59827.792969	129795.0
12	87993.0	34694.0	65108.0	60570.351562	107712.0
13	102366.0	43337.0	74793.0	61347.273438	110021.0
14	90662.0	38066.0	68588.0	60927.035156	110486.0

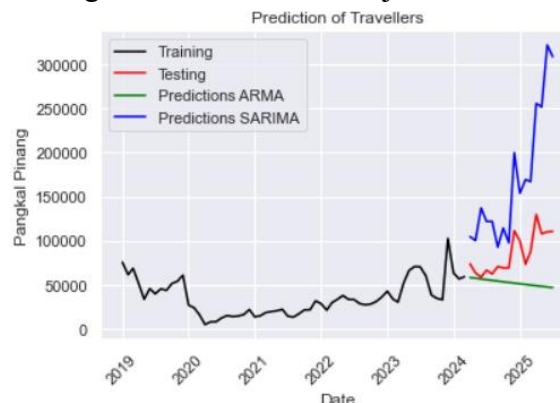
Gambar 11. Hasil Prediksi

Apabila data disajikan dalam bentuk grafik, maka perbandingan antara nilai prediksi dan nilai aktual terlihat secara lebih jelas, sebagaimana diperlihatkan pada Gambar 12.



Gambar 12. Grafik Perbandingan Hasil Prediksi dengan Data Asli

Sebagaimana ditampilkan pada Gambar 12, garis merah menunjukkan data aktual, sedangkan garis hitam menunjukkan hasil prediksi yang diperoleh melalui model LSTM. Setelah dilakukan pengukuran tingkat kesalahan menggunakan *Root Mean Squared Error (RMSE)*, diperoleh nilai sebesar 32.431,2192, yang menunjukkan tingkat akurasi prediksi yang tinggi. Ketika dibandingkan dengan dua pendekatan lainnya, yaitu ARMA dengan nilai RMSE sebesar 41.273,3478 dan SARIMA dengan nilai 101.884,5545, metode LSTM terbukti memberikan hasil prediksi yang lebih akurat. Untuk memberikan gambaran yang lebih jelas, perbandingan grafis antara ketiga metode tersebut disajikan dalam Gambar 13.



Gambar 13. Perbandingan Antara Data Aktual, ARMA, dan SARIMA

Pada Gambar 13, garis hitam merepresentasikan data pelatihan (*training data*) yang digunakan untuk membangun model prediksi, sedangkan garis merah menggambarkan data aktual yang berfungsi sebagai acuan atau pembandingan terhadap hasil prediksi. Sementara itu, garis hijau menunjukkan hasil prediksi yang dihasilkan melalui metode *Autoregressive Moving Average (ARMA)*, dan garis biru merepresentasikan hasil prediksi yang diperoleh dari penerapan metode *Seasonal Autoregressive Integrated Moving Average (SARIMA)*. Melalui visualisasi ini, dapat diamati dengan jelas perbedaan pola prediksi antar metode serta tingkat kedekatannya terhadap data aktual yang menjadi dasar penilaian performa model.

Temuan penelitian ini, yang menunjukkan bahwa model LSTM memberikan akurasi prediksi jumlah perjalanan wisatawan domestik lebih tinggi dibandingkan ARMA dan SARIMA, memiliki implikasi signifikan bagi pengelolaan dan pengembangan sektor pariwisata. Prediksi yang lebih memiliki nilai *error* lebih kecil memungkinkan pemerintah daerah dan pemangku kepentingan untuk melakukan perencanaan strategis yang lebih tepat, khususnya dalam mengantisipasi fluktuasi jumlah kunjungan wisatawan. Informasi mengenai potensi lonjakan kunjungan dapat dimanfaatkan untuk menyiapkan kapasitas layanan yang memadai, mengelola risiko kelebihan kapasitas, serta memastikan ketersediaan infrastruktur

pendukung seperti transportasi, fasilitas publik, dan akomodasi. Kemampuan LSTM dalam mempelajari pola temporal jangka panjang juga memberikan nilai tambah dalam perencanaan program dan kebijakan pariwisata, termasuk perencanaan event berbasis data, penentuan periode kampanye promosi, dan evaluasi efektivitas strategi pemasaran. Di sisi lain, prediksi yang lebih presisi dapat membantu mencegah tekanan berlebih pada lingkungan dan objek wisata, sehingga mendukung upaya mewujudkan konsep pariwisata berkelanjutan.

Dalam konteks Kota Pangkal Pinang, hasil peramalan ini memberikan sejumlah rekomendasi strategis yang dapat diimplementasikan untuk meningkatkan kualitas dan daya saing sektor pariwisata daerah. Pemerintah kota dapat memanfaatkan hasil prediksi untuk memperkuat infrastruktur wisata dan layanan pendukung, terutama pada destinasi yang berpotensi mengalami peningkatan intensitas kunjungan. Selain itu, pemahaman terhadap pola kunjungan bulanan membuka peluang bagi pemerintah dan pelaku industri pariwisata untuk mengelola arus wisatawan secara lebih efektif melalui pengaturan kapasitas, penyediaan alternatif rute wisata, dan pengembangan aktivitas wisata yang tersebar secara temporal. Informasi prediktif ini juga berguna sebagai dasar perencanaan event dan kegiatan wisata, sehingga promosi dapat dilakukan pada periode yang paling tepat untuk mengoptimalkan jumlah kunjungan. Kota Pangkal Pinang, dengan potensi wisata bahari dan budaya yang kuat, dapat mengarahkan strategi promosi yang lebih spesifik dan terukur berdasarkan periode prediksi peningkatan wisatawan. Lebih jauh, hasil penelitian ini memberikan landasan kuat bagi penyusunan kebijakan mitigasi risiko serta pengendalian dampak lingkungan yang mungkin timbul akibat pertumbuhan jumlah wisatawan. Dengan demikian, model LSTM tidak hanya berfungsi sebagai alat analisis deret waktu, tetapi juga sebagai instrumen pendukung pengambilan keputusan yang relevan bagi pembangunan pariwisata yang terencana, berkelanjutan, dan berbasis data di Kota Pangkal Pinang.

D. Kesimpulan

Dalam penelitian ini, dilakukan prediksi jumlah perjalanan wisatawan domestik di Kota Pangkal Pinang menggunakan metode *Long Short-Term Memory (LSTM)*. Metode ini dipilih karena memiliki kemampuan untuk mempelajari pola temporal pada data deret waktu, di mana setiap data saling berkaitan dengan data sebelumnya. Data yang digunakan berasal dari Kabupaten Bangka, Bangka Barat, dan Bangka Tengah, yang dipilih berdasarkan hasil analisis heatmap korelasi dengan Kota Pangkal Pinang. Setelah melalui serangkaian percobaan, konfigurasi terbaik diperoleh pada 500 *epochs*, menunjukkan performa pelatihan model yang optimal.

Hasil pengujian menunjukkan bahwa metode LSTM menghasilkan nilai RMSE sebesar 32.431,2192, lebih rendah dibandingkan dengan metode ARMA (41.273,3478) dan SARIMA (101.884,5545). Nilai kesalahan yang lebih kecil ini menegaskan bahwa LSTM lebih unggul dan akurat dalam memprediksi jumlah perjalanan wisatawan domestik di Kota Pangkal Pinang. Dengan demikian, LSTM dapat dianggap sebagai pendekatan yang lebih efektif dan andal dibandingkan metode konvensional dalam analisis deret waktu pariwisata.

Selain itu, hasil penelitian ini juga memberikan implikasi strategis bagi sektor pariwisata Kota Pangkal Pinang. Prediksi jumlah kunjungan wisatawan yang lebih akurat memungkinkan pemerintah daerah dan pemangku kepentingan untuk melakukan perencanaan yang lebih tepat dalam pengelolaan destinasi, penyediaan infrastruktur, serta penyusunan strategi promosi berbasis data. Dengan memahami pola kunjungan melalui model LSTM, Kota Pangkal Pinang dapat menerapkan kebijakan pariwisata yang lebih adaptif dan berkelanjutan, sehingga mampu meningkatkan kualitas layanan pariwisata sekaligus meminimalkan risiko *overcapacity* pada periode kunjungan puncak.



DAFTAR PUSTAKA

- Allorerung, P. P., Erna, A., & Bagussahrir, M. (2024). *Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit*. 9(3), 178–191.
- Badan Pusat Statistik Provinsi Kepulauan Bangka Belitung (2025). *Jumlah perjalanan Wisatawan Nusantara (Wisnus) tujuan Provinsi Kepulauan Bangka Belitung pada Juni 2025 tercatat sebanyak 431,01 ribu perjalanan*. <https://babel.bps.go.id/id/pressrelease/2025/08/01/1210/jumlah-perjalanan-wisatawan-nusantara--wisnus--tujuan-provinsi-kepulauan-bangka-belitung-pada-juni-2025-tercatat-sebanyak-431-01-ribu-perjalanan.html>
- Indrizal, E., Nurti, Y., & Irwandi, A. (2024). *Museum Hidup: Perkampungan Adat Nagari Sijunjung Dalam Kancah Industri Pariwisata*. 22(3), 1950–1957. <https://doi.org/10.33087/jiubj.v22i3.2834>
- Khaira, U., Alfalah, M., Claudia, P., Gulo, S., Purnomo, R., Studi, P., Informasi, S., Jambi, U., Darat, M., & Jambi, K. (2020). *Prediksi Kemunculan Titik Panas Di Lahan Gambut Provinsi Riau Menggunakan Long Short Term Memory*. 5(3), 77–82.
- Marcelina, E., Agustin, T., Luthfiyaturrohman, K., Octaviani, J., Pramita, A., Monika, I., Dalimunthe, D. Y., & Nasrun, A. (2024). *Peramalan Jumlah Wisatawan Kabupaten Belitung Menggunakan Simulasi Monte Carlo. Euler : Jurnal Ilmiah Matematika, Sains Dan Teknologi*, 12(1), 57–62. <https://doi.org/10.37905/euler.v12i1.25153>
- Marisa, La Pimpi, & Alfian. (2025). *Analisis Metode Seasonal Autoregressive Integrated Moving Average (Sarima) Dan Penerapannya Untuk Meramalkan Penjualan Motor Yamaha Di Indonesia. Jurnal Matematika Komputasi Dan Statistika*, 5(1), 848–856. <https://doi.org/10.33772/jmks.v5i1.117>
- Melyani, C. A., Nurtsabita, A., Shafa, G. Z., & Widodo, E. (2021). *Peramalan Inflasi Di Indonesia Menggunakan Metode Autoregressive Moving Average (ARMA)*. 4(2), 67–74.
- Nurashila, S. S., Hamami, F., & Kusumasari, T. F. (2023). *Perbandingan kinerja algoritma recurrent neural network (rnn) dan long short-term memory (lstm): studi kasus prediksi kemacetan lalu lintas jaringan pt xyz*. 8(3), 864–877.
- Riski, G., Asido, V., & Panggabean, E. (2024). *Penerapan Jaringan Syaraf Tiruan Dalam Memprediksi Tingkat Jumlah Perjalanan Wisatawan Nusantara Menurut Provinsi Tujuan Menggunakan Algoritma Backpropagation Conjugate Gradient. Jurnal JISIILKOM (Jurnal Inovasi Sistem Informasi & Ilmu Komputer)*, 2(1), 3025–4868.
- Sugiartawan, P., Aan, A., Permana, J., & Iman, P. (2018). *Forecasting Kunjungan Wisatawan Dengan Long Short Term Memory (LSTM)*. 1(1).
- Syaifudin, A., & Hapsoro, H. W. (2025). *Implementasi Exploratory Data Analysis Untuk Analisis*. XX(1).