



**MODEL *HYBRID* SARIMAX-BILSTM DENGAN DATA HISTORIS  
DAN *NON-LINEAR* UNTUK PREDIKSI KUNJUNGAN WISATAWAN**

**SKRIPSI**

**Oleh:**

**ARYADEWA SATYAGRAHA**

**20211000022**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS BUDDHI DHARMA**

**TANGERANG**

**2025**



**MODEL *HYBRID* SARIMAX-BILSTM DENGAN DATA HISTORIS  
DAN *NON-LINEAR* UNTUK PREDIKSI KUNJUNGAN WISATAWAN**

**SKRIPSI**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program  
Studi Teknik Informatika Fakultas Sains dan Teknologi**

**Universitas Buddhi Dharma  
Jenjang Pendidikan Strata 1**

**Oleh:**

**ARYADEWA SATYAGRAHA**

**20211000022**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS BUDDHI DHARMA**

**TANGERANG**

**2025**

## LEMBAR PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini,

NIM : 20211000022  
Nama : Aryadewa Satyagraha  
Jenjang Studi : Strata 1  
Program Studi : Teknik Informatika  
Peminatan : Data Science

Dengan ini saya menyatakan bahwa:

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapat gelar akademik Sarjana atau kelengkapan studi, baik di Universitas Buddhi Dharma maupun di Perguruan Tinggi lainnya.
2. Skripsi ini saya buat sendiri tanpa bantuan dari pihak lain, kecuali arahan dosen pembimbing.
3. Dalam Skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dan dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan daftar pustaka.
4. Dalam Skripsi ini tidak terdapat pemalsuan (kebohongan), seperti buku, artikel, jurnal, data sekunder, pengolahan data, dan pemalsuan tanda tangan dosen atau Ketua Program Studi Universitas Buddhi Dharma yang dibuktikan dengan keasliannya.
5. Lembar pernyataan ini saya buat dengan sesungguhnya, tanpa paksaan dan apabila dikemudian hari atau pada waktu lainnya terdapat penyimpangan dan ketidak benaran dalam pernyataan ini, saya bersedia menerima sanksi akademik berupa pencabutan gelar akademik yang telah saya peroleh karena Skripsi ini serta sanksi lainnya sesuai dengan peraturan dan norma yang berlaku.

Tangerang, 06-08-2025

Yang membuat pernyataan,



Aryadewa Satyagraha  
20211000022

## LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini,

|               |                       |
|---------------|-----------------------|
| NIM           | : 20211000022         |
| Nama          | : Aryadewa Satyagraha |
| Jenjang Studi | : Strata 1            |
| Program Studi | : Teknik Informatika  |
| Peminatan     | : Data Science        |

Dengan ini menyetujui untuk memberikan ijin kepada pihak Universitas Buddhi Dharma, Hak Bebas Royalti Non – Eksklusif (Non-exclusive Royalty-Free Right) atas karya ilmiah kami yang berjudul: “MODEL *HYBRID* SARIMAX-BILSTM DENGAN DATA HISTORIS DAN *NON-LINEAR* UNTUK PREDIKSI KUNJUNGAN WISATAWAN”.

Dengan Hak Bebas Royalti Non – Eksklusif ini pihak Universitas Buddhi Dharma berhak menyimpan, mengalih-media atau format-kan, mengelolanya dalam pangkalan data (database), mendistribusikannya, dan menampilkan atau mempublikasikannya di internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis pertama atau pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Universitas Buddhi Dharma, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 06-08-2025

Yang membuat pernyataan,



Aryadewa Satyagraha  
20211000022

## LEMBAR PENGESAHAN PEMBIMBING

### **MODEL *HYBRID* SARIMAX-BILSTM DENGAN DATA HISTORIS DAN *NON-LINEAR* UNTUK PREDIKSI KUNJUNGAN WISATAWAN**

Dibuat Oleh:

NIM : 20211000022

Nama : Aryadewa Satyagraha

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian

Komprehensif

Program Studi Teknik Informatika

Data Science

Tahun Akademik 2024/2025

Disahkan oleh,

Tangerang, 22-07-2025

Pembimbing

**Yusuf Kurnia, M.Kom. M.M.**

NUPTK. 4551765666130223

**LEMBAR PENGESAHAN SKRIPSI**  
**MODEL *HYBRID* SARIMAX-BILSTM DENGAN DATA HISTORIS**  
**DAN *NON-LINEAR* UNTUK PREDIKSI KUNJUNGAN WISATAWAN**

Dibuat Oleh:

NIM : 20211000022

Nama : Aryadewa Satyagraha

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian

Komprehensif

Program Studi Teknik Informatika

Data Science

Tahun Akademik 2024/2025

Disahkan oleh,

Tangerang, 06-08-2025

**Dekan,**

**Ketua Program Studi**

  
Dr. Yakub, M.Kom., M.M.

NUPTK. 1836747648130172

  
Hartana Wijaya, M.Kom.

NUPTK. 3844759660131192

## LEMBAR PENGESAHAN TIM PENGUJI

Nama : Aryadewa Satyagraha  
NIM : 2021100022  
Fakultas : Sains dan Teknologi  
Judul Skripsi : MODEL *HYBRID* SARIMAX-BILSTM DENGAN DATA HISTORIS DAN *NON-LINEAR* UNTUK PREDIKSI KUNJUNGAN WISATAWAN

Dinyatakan LULUS setelah mempertahankan di depan Tim Penguji pada hari Rabu, 06 Agustus 2025

Nama penguji:  
Ketua Sidang : **Dr. Yakub, M.Kom., M.M.**  
NUPTK. 1836747648130172  
Penguji I : **Muhammad Subhana, M.Kom.**  
NUPTK. 8133772673130323  
Penguji II : **Yusuf Kurnia, M.Kom., M.M.**  
NUPTK. 4551765666130223

Tanda Tangan:



Mengetahui,

Dekan Fakultas Sains dan Teknologi



**Dr. Yakub, M.Kom., M.M.**  
NUPTK. 1836747648130172



## KATA PENGANTAR

Dengan mengucapkan Puji Syukur kepada Tuhan Yang Maha Esa, yang telah memberikan Rahmat dan karunia-Nya kepada penulis sehingga dapat menyusun dan menyelesaikan Skripsi ini dengan judul *MODEL HYBRID SARIMAX-BILSTM DENGAN DATA HISTORIS DAN NON-LINEAR UNTUK PREDIKSI KUNJUNGAN WISATAWAN* . Tujuan utama dari pembuatan Skripsi ini adalah sebagai salah satu syarat kelengkapan dalam menyelesaikan program pendidikan Strata 1 Program Studi Teknik Informatika di Universitas Buddhi Dharma. Dalam penyusunan Skripsi ini penulis banyak menerima bantuan dan dorongan baik moril maupun materiil dari berbagai pihak, maka pada kesempatan ini penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Limajatini, S.E., M.M., B.K.P, C.T.C., sebagai Rektor Universitas Buddhi Dharma
2. Bapak Dr. Yakub, S.Kom., M.Kom., M.M., Dekan Fakultas Sains dan Teknologi
3. Bapak Hartana Wijaya, S.Kom., M.Kom., sebagai Ketua Program Studi Teknik Informatika
4. Bapak Yusuf Kurnia, S.Kom., M.Kom., M.M., sebagai pembimbing yang telah membantu dan memberikan dukungan serta harapan untuk menyelesaikan penulisan Skripsi ini.
5. Bapak Aditya Hermawan, S.Kom., M.Kom., selaku dosen yang telah memberikan dukungan dan masukan selama proses penulisan Skripsi ini. .
6. Orang tua dan keluarga yang selalu memberikan dukungan baik moril dan materiil.
7. Meliyana Cannilia, yang senantiasa memberikan dukungan, semangat, serta pengertian selama proses penyelesaian Skripsi ini.

Serta semua pihak yang terlalu banyak untuk disebutkan satu-persatu sehingga terwujudnya penulisan ini. Penulis menyadari bahwa penulisan Skripsi ini masih belum sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan di masa yang akan datang.

Akhir kata semoga Skripsi ini dapat berguna bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Tangerang, 06-08-2025

Penulis



## ABSTRAK

Industri pariwisata, khususnya di Indonesia, menjadi salah satu sektor penting yang memberikan kontribusi signifikan terhadap perekonomian negara. Namun, prediksi kedatangan wisatawan seringkali mengalami keterlambatan akibat data yang tidak tersedia secara real-time. Penelitian ini bertujuan untuk mengembangkan model prediksi yang lebih akurat dengan mengintegrasikan data historis kedatangan wisatawan dari Badan Pusat Statistik (BPS) dan data pencarian daring dari *Google Trends*. Penelitian ini mengusulkan penggunaan model *hybrid* SARIMAX-BiLSTM yang menggabungkan kekuatan SARIMAX dalam menangkap pola musiman dan hubungan linear, serta kemampuan BiLSTM dalam memodelkan pola *non-linear* dan hubungan temporal yang kompleks. Metode yang digunakan melibatkan tahap *pre-processing data* untuk memastikan kualitas dataset, diikuti dengan pemodelan menggunakan SARIMAX dan BiLSTM. Hasil evaluasi model menunjukkan bahwa pendekatan *hybrid* ini mengungguli model SARIMAX dan BiLSTM tunggal dalam hal akurasi prediksi, yang dibuktikan dengan nilai *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), dan *Root Mean Squared Error* (RMSE) yang lebih rendah. Dengan demikian, model *hybrid* SARIMAX-BiLSTM dapat memberikan prediksi yang lebih akurat dan adaptif terhadap perubahan tren kunjungan wisatawan, serta mendukung pengambilan keputusan yang lebih tepat dalam perencanaan dan pengelolaan pariwisata di Indonesia.

**Kata kunci:** SARIMAX, BiLSTM, Prediksi Kunjungan Wisatawan, *Google Trends*, *Hybrid Model*

## **ABSTRACT**

The tourism industry, particularly in Indonesia, constitutes a vital economic sector with significant contributions to national development. However, forecasting tourist arrivals remains a persistent challenge due to the unavailability of real-time data. This study aims to develop a more robust and accurate predictive model by integrating historical tourist arrival data from Statistics Indonesia (Badan Pusat Statistik, BPS) with online search data obtained from Google Trends. To this end, the research proposes a hybrid SARIMAX-BiLSTM model that leverages the seasonal and linear pattern-capturing capabilities of the Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX) model and the nonlinear, temporally-aware Learning potential of the Bidirectional Long Short-Term Memory (BiLSTM) network. The methodology encompasses comprehensive data preprocessing to ensure dataset quality, followed by independent and hybrid modeling. Model performance was evaluated using standard forecasting accuracy metrics, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). Empirical results demonstrate that the proposed hybrid approach significantly outperforms standalone SARIMAX and BiLSTM models in predictive accuracy. The SARIMAX-BiLSTM hybrid model thus offers a more adaptive and reliable forecasting tool for capturing evolving trends in tourist arrivals, with direct implications for strategic tourism planning and policy formulation in Indonesia.

**Keywords:** SARIMAX, BiLSTM, Tourist Arrival Forecasting, Google Trends, Hybrid Model

## DAFTAR ISI

|                                                 |      |
|-------------------------------------------------|------|
| LEMBAR PERNYATAAN KEASLIAN SKRIPSI.....         | i    |
| LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH ..... | ii   |
| LEMBAR PENGESAHAN PEMBIMBING .....              | iii  |
| LEMBAR PENGESAHAN SKRIPSI .....                 | iv   |
| LEMBAR PENGESAHAN TIM PENGUJI .....             | v    |
| KATA PENGANTAR.....                             | vi   |
| ABSTRAK.....                                    | vii  |
| ABSTRACT .....                                  | viii |
| DAFTAR ISI .....                                | ix   |
| DAFTAR GAMBAR.....                              | xiii |
| DAFTAR TABEL .....                              | xiv  |
| DAFTAR LAMPIRAN .....                           | xv   |
| BAB I PENDAHULUAN .....                         | 1    |
| 1.1 Latar Belakang.....                         | 1    |
| 1.2 Identifikasi Masalah.....                   | 5    |
| 1.3 Ruang Lingkup.....                          | 5    |
| 1.4 Tujuan Penelitian .....                     | 6    |
| 1.4.1 Tujuan.....                               | 6    |
| 1.4.2 Manfaat.....                              | 6    |
| 1.5 Sistematika Penulisan .....                 | 7    |
| BAB II LANDASAN TEORI.....                      | 8    |
| 2.1 Data .....                                  | 8    |
| 2.2 Website .....                               | 8    |

|                                 |                                                                                                   |    |
|---------------------------------|---------------------------------------------------------------------------------------------------|----|
| 2.3                             | Prediksi .....                                                                                    | 9  |
| 2.4                             | <i>Machine Learning</i> .....                                                                     | 10 |
| 2.5                             | <i>Deep Learning</i> .....                                                                        | 11 |
| 2.6                             | <i>Time Series</i> .....                                                                          | 12 |
| 2.7                             | <i>Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX)</i> ..... | 12 |
| 2.8                             | <i>Bidirectional Long Short-Term Memory (BiLSTM)</i> .....                                        | 13 |
| 2.9                             | <i>Google Trends</i> .....                                                                        | 14 |
| 2.10                            | Python .....                                                                                      | 14 |
| 2.11                            | <i>Flowchart</i> .....                                                                            | 15 |
| 2.12                            | Tinjauan Studi .....                                                                              | 16 |
| 2.13                            | Kesimpulan Tinjauan Studi .....                                                                   | 18 |
| BAB III METODE PENELITIAN ..... |                                                                                                   | 20 |
| 3.1                             | Kerangka Penelitian .....                                                                         | 20 |
| 3.2                             | Pengumpulan Data .....                                                                            | 23 |
| 3.2.1                           | Data Kedatangan Wisatawan Mancanegara .....                                                       | 23 |
| 3.2.2                           | <i>Data Google Trends</i> .....                                                                   | 24 |
| 3.3                             | <i>Data Understanding</i> .....                                                                   | 28 |
| 3.3.1                           | Data Kedatangan Wisatawan Mancanegara .....                                                       | 28 |
| 3.3.2                           | <i>Data Google Trends</i> .....                                                                   | 28 |
| 3.4                             | <i>Data Preprocessing</i> .....                                                                   | 30 |
| 3.4.1                           | Pembersihan Data .....                                                                            | 30 |

|                    |                                                                                                   |    |
|--------------------|---------------------------------------------------------------------------------------------------|----|
| 3.4.2              | Transformasi Data .....                                                                           | 33 |
| 3.4.3              | Uji Stasioner .....                                                                               | 35 |
| 3.4.4              | Uji ACF dan PACF .....                                                                            | 36 |
| 3.4.5              | Analisis Lag.....                                                                                 | 38 |
| 3.4.6              | Normalisasi Data .....                                                                            | 40 |
| 3.5                | <i>Data Spliting</i> .....                                                                        | 41 |
| 3.6                | <i>Model Training</i> .....                                                                       | 43 |
| 3.6.1              | <i>Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX)</i> ..... | 43 |
| 3.6.2              | <i>Bidirectional Long Short-Term Memory (BiLSTM)</i> .....                                        | 45 |
| 3.6.3              | <i>SARIMAX-BILSTM</i> .....                                                                       | 47 |
| 3.7                | <i>Metode Evaluasi</i> .....                                                                      | 49 |
| 3.7.1              | <i>Mean Absolute Error (MAE)</i> .....                                                            | 50 |
| 3.7.2              | <i>Mean Absolute Percentage Error (MAPE)</i> .....                                                | 50 |
| 3.7.3              | <i>Root Mean Squared Error (RMSE)</i> .....                                                       | 50 |
| BAB IV HASIL ..... |                                                                                                   | 51 |
| 4.1                | <i>Evaluasi Model</i> .....                                                                       | 51 |
| 4.1.1              | <i>Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX)</i> ..... | 51 |
| 4.1.2              | <i>Bidirectional Long Short-Term Memory (BiLSTM)</i> .....                                        | 52 |
| 4.1.3              | <i>SARIMAX-BILSTM</i> .....                                                                       | 54 |
| 4.2                | <i>Analisis Hasil Evaluasi</i> .....                                                              | 56 |

|                                |                                                |    |
|--------------------------------|------------------------------------------------|----|
| 4.2.1                          | Analisis Evaluasi SARIMAX.....                 | 56 |
| 4.2.2                          | Analisis Evaluasi BiLSTM.....                  | 58 |
| 4.2.3                          | Analisis Evaluasi SARIMAX-BiLSTM.....          | 59 |
| 4.3                            | Visualisasi Hasil Prediksi .....               | 61 |
| 4.3.1                          | Visualisasi Hasil Prediksi SARIMAX .....       | 62 |
| 4.3.2                          | Visualisasi Hasil Prediksi BiLSTM .....        | 63 |
| 4.3.3                          | Visualisasi Hasil Prediksi SARIMAX-BiLSTM..... | 64 |
| 4.4                            | Implementasi Web .....                         | 65 |
| 4.4.1                          | Halaman Utama.....                             | 70 |
| 4.4.2                          | Bagian Hasil Prediksi di Halaman Utama.....    | 71 |
| 4.4.3                          | Halaman <i>About Predict</i> .....             | 75 |
| 4.5                            | Pembahasan.....                                | 79 |
| BAB V SIMPULAN DAN SARAN ..... |                                                | 83 |
| 5.1                            | Simpulan .....                                 | 83 |
| 5.2                            | Saran .....                                    | 84 |
| DAFTAR PUSTAKA.....            |                                                | 86 |
| DAFTAR RIWAYAT HIDUP .....     |                                                | 92 |

## DAFTAR GAMBAR

|                                                                   |    |
|-------------------------------------------------------------------|----|
| Gambar 2.1 Struktur BiLSTM.....                                   | 13 |
| Gambar 3.1 Kerangka Penelitian.....                               | 20 |
| Gambar 3.2 ACF Kedatangan Wisatawan Mancanegara .....             | 37 |
| Gambar 3.3 PACF Kedatangan Wisatawan Mancanegara .....            | 38 |
| Gambar 3.4 <i>Flowchart</i> SARIMAX .....                         | 44 |
| Gambar 3.5 <i>Flowchart</i> BiLSTM .....                          | 46 |
| Gambar 3.6 <i>Flowchart</i> SARIMX-BiLSTM .....                   | 48 |
| Gambar 4.1 Hasil Prediksi SARIMAX .....                           | 62 |
| Gambar 4.2 Hasil Prediksi BiLSTM .....                            | 63 |
| Gambar 4.3 Hasil Prediksi SARIMAX-BiLSTM.....                     | 64 |
| Gambar 4.4 Halaman Utama .....                                    | 70 |
| Gambar 4.5 Bagian Halaman Hasil Prediksi ( <i>default</i> ) ..... | 71 |
| Gambar 4.6 Bagian Halaman Hasil Prediksi (prediksi 1 bulan).....  | 72 |
| Gambar 4.7 Bagian Halaman Hasil Prediksi (prediksi 3 bulan).....  | 72 |
| Gambar 4.8 Bagian Halaman Hasil Prediksi (prediksi 6 bulan).....  | 72 |
| Gambar 4.9 Bagian Halaman Hasil Prediksi (prediksi 12 bulan)..... | 73 |
| Gambar 4.10 Halaman <i>About Predict</i> (bagian 1) .....         | 75 |
| Gambar 4.11 Halaman <i>About Predict</i> (bagian 2) .....         | 75 |
| Gambar 4.12 Halaman <i>About Predict</i> (bagian 3) .....         | 76 |
| Gambar 4.13 Halaman <i>About Predict</i> (bagian 4) .....         | 76 |
| Gambar 4.14 Halaman <i>About Predict</i> (bagian 5) .....         | 77 |
| Gambar 4.15 Halaman <i>About Predict</i> (bagian 6) .....         | 77 |
| Gambar 4.16 Halaman <i>About Predict</i> (bagian 7) .....         | 78 |



## DAFTAR TABEL

|                                                                                                        |    |
|--------------------------------------------------------------------------------------------------------|----|
| Tabel 2.1 Simbol <i>Flowchart</i> .....                                                                | 15 |
| Tabel 3.1 Rata-rata Kedatangan Wisatawan Mancanegara ke Bali 2014-2022 .....                           | 25 |
| Tabel 3.2 Jumlah <i>Score Keyword Google Trends</i> Destinasi Wisata Bali .....                        | 25 |
| Tabel 3.3 Data Wisatawan Mancanegara Bulanan ke Bali Menurut Pintu Masuk 2009 ....                     | 27 |
| Tabel 3.4 Data Wisatawan Mancanegara Bulanan ke Bali Menurut Pintu Masuk 2009<br><i>Cleaning</i> ..... | 27 |
| Tabel 3.5 Data Kedatangan Wisatawan Mancanegara ke Bali 2009-2024 .....                                | 27 |
| Tabel 3.6 Data <i>Google Trends</i> Kuta di Australia .....                                            | 30 |
| Tabel 3.7 Data <i>Google Trends</i> Kuta di Australia <i>Cleaning</i> .....                            | 32 |
| Tabel 3.8 Data <i>Google Trends</i> 2009-2024 .....                                                    | 35 |
| Tabel 3.9 Uji Stasioner .....                                                                          | 36 |
| Tabel 3.10 Uji Stasioner <i>Differencing</i> .....                                                     | 36 |
| Tabel 3.11 Analisis Lag.....                                                                           | 39 |
| Tabel 3.12 Perbandingan Sebelum dan Sesudah <i>Shifting</i> .....                                      | 39 |
| Tabel 3.13 Data Setelah Normalisasi .....                                                              | 40 |
| Tabel 3.14 <i>Data Training</i> .....                                                                  | 42 |
| Tabel 3.15 <i>Data Testing</i> .....                                                                   | 42 |
| Tabel 4.1 Hasil Evaluasi Model .....                                                                   | 56 |
| Tabel 4.2 Hasil Evaluasi SARIMAX .....                                                                 | 56 |
| Tabel 4.3 Hasil Evaluasi BiLSTM .....                                                                  | 58 |
| Tabel 4.4 Hasil Evaluasi SARIMAX-BiLSTM.....                                                           | 59 |

## DAFTAR LAMPIRAN

|                       |    |
|-----------------------|----|
| Kartu Bimbingan ..... | 96 |
|-----------------------|----|



# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Industri Pariwisata telah menjadi sektor penting yang berkontribusi besar terhadap ekonomi banyak negara, baik negara maju maupun berkembang, termasuk negara Indonesia. Menurut data Kementerian Pariwisata dan Ekonomi Kreatif (Kemenparekraf), kontribusi sektor pariwisata terhadap Produk Domestik Bruto (PDB) nasional pada tahun 2023 mencapai 3,9%, dengan nilai devisa pariwisata sebesar USD 14 miliar atau sekitar Rp 218,4 triliun (Kemenparekraf/Baparekraf RI, 2024). Capaian ini menempatkan pariwisata sebagai salah satu sektor utama penyumbang devisa negara, menunjukkan peran signifikan dalam perekonomian nasional. Salah satu provinsi yang menjadi penyumbang devisa terbesar dalam sektor ini adalah Bali, yang dikenal secara global sebagai destinasi wisata unggulan Indonesia.

Bali merupakan salah satu daerah wisata di Indonesia yang sering dikunjungi oleh wisatawan mancanegara (Salimu & Yunus, 2020). Dengan berbagai daya tarik yang dimiliki, Bali mampu menarik perhatian wisatawan lokal maupun mancanegara. Bali tidak hanya menyumbang devisa dalam jumlah besar, tetapi juga menjadi pusat pengembangan pariwisata nasional yang memikat jutaan pengunjung setiap tahunnya. Menurut data dari Badan Pusat Statistik, jumlah wisatawan mancanegara yang berkunjung ke Bali pada tahun 2023 mencapai 5.273.258 kunjungan, ini meningkat signifikan sebesar 144,61% dari tahun 2022 (Badan Pusat Statistik (BPS) Provinsi Bali, 2024). Meski menunjukkan tren positif yang mengindikasikan tingginya minat wisatawan terhadap Bali, data ini masih memiliki keterbatasan dalam hal ketepatan waktu penyajiannya.

Data kedatangan wisatawan mancanegara yang disediakan oleh Badan Pusat Statistik (BPS) memiliki keterbatasan dalam hal ketepatan waktu, terutama terkait jeda penerbitan

laporan. Laporan resmi sering kali dirilis dengan jeda waktu yang signifikan, sehingga kurang mampu merespons dinamika pariwisata yang cepat berubah. Sebagai contoh, laporan BPS mengenai perkembangan pariwisata Provinsi Bali untuk bulan Maret 2024 baru dirilis pada Mei 2024 (Badan Pusat Statistik Provinsi Bali, 2024), menunjukkan adanya keterlambatan dalam penyajian data. Keterbatasan ini dapat menghambat pengambilan keputusan yang tepat waktu oleh pemangku kepentingan di sektor pariwisata.

Untuk mengatasi keterbatasan tersebut, solusi dapat atasi dengan memanfaatkan data pencarian daring seperti *Google Trends* (GT). GT menawarkan data *real-time* tentang kebutuhan, keinginan, permintaan, dan minat pengguna terhadap berbagai topik, termasuk destinasi wisata seperti Bali (Costa & Leite, 2022). GT dapat memberikan wawasan mengenai minat pariwisata global terhadap Bali, yang dapat digunakan untuk memprediksi tren kunjungan wisatawan (Shafa Anasya et al., 2024). Dengan integrasi data kedatangan langsung dari BPS dan data pencarian daring dari *Google Trends* memungkinkan identifikasi pola kedatangan wisatawan yang lebih komprehensif, sehingga dapat meningkatkan akurasi prediksi dan mendukung pengambilan keputusan strategis oleh pemangku kepentingan di sektor pariwisata.

Integrasi data kedatangan wisatawan dengan data pencarian daring dari *Google Trends* adalah salah satu pendekatan yang semakin banyak digunakan dalam prediksi pariwisata. Data statistik kedatangan wisatawan yang disediakan oleh Badan Pusat Statistik memberikan gambaran historis mengenai pola musiman dan tren jangka panjang, sementara data *Google Trends* mampu menangkap pola perilaku wisatawan potensial secara real-time. Kombinasi data kedatangan wisatawan dan data *Google Trends* dapat memberikan hasil prediksi yang lebih akurat (Mukhtar et al., 2023). Dengan integrasi ini, penggunaan model prediksi yang tepat, seperti SARIMAX (*Seasonal Autoregressive Integrated Moving Average with Exogenous Variables*) untuk menangkap pola *linear* dan musiman

serta BiLSTM (*Bidirectional Long Short-Term Memory*) untuk mendeteksi pola *non-linear* dan dinamis, menjadi sangat penting dalam mengolah data yang kompleks.

SARIMAX merupakan model statistik yang dirancang untuk menganalisis data *Time Series* dengan mempertimbangkan pola musiman, tren jangka panjang, dan variabel eksogen. Model ini sangat efektif dalam menangkap hubungan antara data historis dan faktor eksternal seperti data *Google Trends*, yang sering digunakan untuk memproyeksikan perubahan tren pariwisata. Seperti pada penelitian sebelumnya menunjukkan bahwa SARIMAX mampu meningkatkan akurasi prediksi kunjungan wisatawan dengan memasukkan variabel eksogen seperti tren pencarian daring dan dampak pandemi COVID-19, menghasilkan nilai MAPE yang lebih rendah dibandingkan model statistik tradisional lainnya (Prilistya et al., 2021). Meskipun SARIMAX unggul dalam menangkap pola *linear* dan musiman, model ini memiliki keterbatasan dalam mengolah pola *non-linear* yang sering muncul pada data pencarian daring, sehingga pendekatan berbasis pembelajaran mendalam seperti BiLSTM menjadi solusi yang relevan untuk melengkapinya.

Penggunaan BiLSTM dalam penelitian ini bertujuan untuk menangkap pola *non-linear* dan hubungan temporal yang kompleks, yang sering kali tidak dapat dimodelkan secara optimal oleh SARIMAX. Penggunaan BiLSTM dalam penelitian ini bertujuan untuk menangkap pola *non-linear* dan hubungan temporal yang kompleks, yang sering kali tidak dapat dimodelkan secara optimal oleh SARIMAX. BiLSTM merupakan pengembangan lanjutan dari LSTM (*Long Short-Term Memory*) yang dirancang untuk menangkap dependensi temporal jangka panjang baik dari arah masa lalu maupun masa depan dan memiliki arsitektur jaringan yang mampu memproses data secara *bidirectional*, sehingga memberikan kemampuan yang lebih baik dalam memahami pola *non-linear* yang rumit. BiLSTM mengungguli LSTM dan model pembelajaran mesin tradisional dalam memperkirakan kedatangan wisatawan Singapura (Kulshrestha et al., 2020). Selain itu,

BiLSTM juga terbukti sangat efektif dalam memodelkan data *non-linear*, terutama ketika digunakan pada data pencarian daring yang menunjukkan fluktuasi dan pola yang sulit diprediksi dengan pendekatan *linear* (Laaroussi et al., 2023). Dengan demikian, penggabungan model SARIMAX-BiLSTM memungkinkan pemodelan yang lebih akurat dan komprehensif terhadap data kedatangan wisatawan dan pencarian daring, memanfaatkan kekuatan SARIMAX dalam menangkap pola musiman dan tren *linear* serta kemampuan BiLSTM dalam mengidentifikasi pola *non-linear* dan dinamis.

Dengan menggabungkan SARIMAX dan BiLSTM, pendekatan *hybrid* ini menawarkan solusi yang lebih efektif dalam menangani kompleksitas data pariwisata yang dipengaruhi oleh faktor eksternal dan perubahan tren perilaku wisatawan. Model *hibrida* SARIMAX-BiLSTM ini menggabungkan kekuatan SARIMAX dalam menangkap pola musiman dan hubungan jangka panjang antara data kedatangan wisatawan dan variabel *eksogen*, sementara BiLSTM unggul dalam mengidentifikasi pola *non-linear* dan dinamika yang lebih kompleks, terutama pada data yang fluktuatif dan dinamis. Model *hybrid* menunjukkan kinerja yang lebih baik dalam meramalkan permintaan pariwisata, terutama dalam menangani pola musiman dan ketergantungan *non-linear* yang kompleks, sehingga mengungguli model tradisional (He et al., 2021a). Dengan demikian, kombinasi kedua model ini memungkinkan prediksi yang lebih akurat dan komprehensif.

Berdasarkan latar belakang yang telah diuraikan, penelitian ini berfokus pada pengoptimalan prediksi kedatangan wisatawan mancanegara melalui pendekatan model *hybrid* SARIMAX-BiLSTM. Model ini menggabungkan kekuatan SARIMAX dalam menangkap pola musiman dan hubungan jangka panjang antara data kedatangan wisatawan dan variabel eksternal, dengan BiLSTM yang mampu memodelkan pola *non-linear* dan dinamika kompleks. Optimasi ini dilakukan dengan menambahkan BiLSTM untuk mengatasi keterbatasan model tradisional seperti SARIMAX dalam menangani pola non-

linear, khususnya pada prediksi kedatangan wisatawan di Indonesia. Pendekatan *hybrid* ini diharapkan mampu memberikan solusi yang lebih akurat dan adaptif dibandingkan model tradisional, sehingga mendukung pengambilan keputusan strategis yang lebih tepat berbasis data. Oleh karena itu, penelitian ini diberi judul "**MODEL *HYBRID* SARIMAX-BILSTM DENGAN DATA HISTORIS DAN *NON-LINEAR* UNTUK PREDIKSI KUNJUNGAN WISATAWAN**".

## **1.2 Identifikasi Masalah**

Berdasarkan latar belakang masalah di atas, dapat diidentifikasi masalah sebagai berikut:

- a. Keterlambatan penyajian data statistik yang menghambat pengambilan keputusan yang tepat waktu.
- b. Kesulitan mengolah pola *non-linear* dalam data pariwisata dan tantangan mengintegrasikan data kedatangan wisatawan dengan data pencarian daring secara efektif.
- c. Kurangnya penggunaan model *hybrid* dalam prediksi pariwisata di Indonesia.

## **1.3 Ruang Lingkup**

Ruang lingkup penelitian ini adalah sebagai berikut:

- a. *Dataset* yang digunakan adalah data kedatangan wisatawan mancanegara langsung ke Bali dari tahun 2008 hingga 2023 yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Bali, serta data *Google Trends* dari 10 negara dengan jumlah rata-rata kedatangan tertinggi ke Bali, menggunakan *Keyword* destinasi wisata di Bali.



- b. Fokus penelitian adalah menganalisis dan memprediksi tren kedatangan wisatawan mancanegara ke Bali berdasarkan data historis kedatangan dan data pencarian daring (*Google Trends*).
- c. Pengembangan aplikasi berbasis web yang akan menampilkan prediksi kedatangan wisatawan mancanegara ke Bali

## 1.4 Tujuan Penelitian

### 1.4.1 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

- a. Mengembangkan model prediksi yang lebih akurat dengan mengintegrasikan data *real-time* dari *Google Trends* dan data historis kedatangan wisatawan.
- b. Menerapkan model *hybrid* SARIMAX-BiLSTM untuk menggabungkan data historis dan *real-time*, menangani pola *non-linear*, serta meningkatkan akurasi prediksi kedatangan wisatawan.
- c. Mengeksplorasi dan menguji efektivitas model *hybrid* SARIMAX-BiLSTM dalam memprediksi kedatangan wisatawan mancanegara di Indonesia.

### 1.4.2 Manfaat

Manfaat yang dapat diperoleh dalam penelitian ini adalah sebagai berikut:

- a. Mempercepat pengambilan keputusan di sektor pariwisata dengan menyediakan data prediksi kedatangan wisatawan yang lebih akurat dan tepat waktu.
- b. Menyediakan kerangka kerja prediksi yang lebih efektif melalui penerapan model *hybrid* SARIMAX-BiLSTM, yang mampu menangkap pola musiman dan hubungan *non-linear* dalam data wisatawan.

- c. Memberikan pendekatan baru dalam prediksi kedatangan wisatawan yang lebih tepat guna untuk perencanaan dan pengelolaan pariwisata di Indonesia, serta menjadi referensi untuk penelitian lebih lanjut di sektor pariwisata.

## **1.5 Sistematika Penulisan**

Penulisan sistematis dimaksudkan untuk mempermudah pemahaman dan penelaahan penelitian. Dalam laporan penelitian ini, bahan yang ditulis dikelompokkan menjadi beberapa sub bab, yang secara garis besar dapat dijelaskan sebagai berikut:

### **BAB I PENDAHULUAN**

Bab ini memuat penjelasan mengenai latar belakang masalah, identifikasi masalah, ruang lingkup penelitian, tujuan dan manfaat penelitian, serta sistematika penulisan.

### **BAB II LANDASAN TEORI**

Bab ini memuat paparan mengenai teori-teori yang berhubungan dengan penelitian ini, yang terbagi menjadi 4 bagian, di antaranya adalah teori umum, teori khusus, teori rancangan, dan teori pengujian, serta mencakup tinjauan studi.

### **BAB III METODE PENELITIAN**

Bab ini memuat paparan terkait metode yang digunakan dalam penelitian.

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini memuat hasil evaluasi terhadap model yang diangkat, tampilan program, dan pengujian aplikasi.

### **BAB V SIMPULAN DAN SARAN**

Bab ini memuat penjelasan mengenai kesimpulan penelitian berdasarkan hasil evaluasi dan penerapan yang telah dilakukan, serta rekomendasi yang dapat dikembangkan untuk penelitian di masa mendatang.

## BAB II

### LANDASAN TEORI

#### 2.1 Data

Data dapat didefinisikan sebagai sekumpulan fakta, informasi, atau pengamatan yang disajikan dalam bentuk *interval*, *nominal*, *ordinal*, maupun persentase, yang dapat dimanfaatkan untuk keperluan penelitian, pengambilan keputusan, dan berbagai tujuan lainnya (Umar Hussain, 2023). Data dapat di kategorikan menjadi tiga jenis, antara lain data tidak terstruktur, data terstruktur, dan data semi-terstruktur. Data tidak terstruktur adalah data yang tidak mengikuti aturan tertentu dan tidak dapat diakses menggunakan *query*, sementara data terstruktur dapat diakses baik oleh manusia maupun mesin melalui *query*, dan di antara keduanya terdapat data semi-terstruktur, seperti yang ditemukan pada format JSON dan XML. (Breitinger & Jotterand, 2023).

#### 2.2 Website

*Website* merupakan sekumpulan halaman web yang berisi teks dalam format *Hyper Text Markup Language* (HTML), yang disimpan di *server hosting* dan dapat diakses melalui *browser* dengan koneksi internet menggunakan *Uniform Resource Locator* (URL) (Widia & Asriningtias, 2021). *Website* dapat dibagi menjadi dua jenis, yaitu *Website* statis dan *Website* dinamis. *Website* statis memerlukan perubahan manual pada *source code* untuk mengubah konten, sementara *Website* dinamis memungkinkan pembaruan konten secara otomatis tanpa perlu mengubah *source code*, karena kontennya disimpan dalam database dan biasanya dikembangkan menggunakan bahasa pemrograman *server-side* seperti PHP, ASP, JSP, dan lainnya (Widia & Asriningtias, 2021).

### 2.3 Prediksi

Peramalan atau prediksi merupakan suatu teknik analisa perhitungan dengan melakukan pendekatan kualitatif maupun kuantitatif untuk memperkirakan kejadian di masa mendatang dengan menggunakan referensi data-data yang terdahulu (Ariyanto et al., 2020). Prediksi berfokus pada meramalkan kejadian-kejadian yang mungkin terjadi di masa depan, dengan mengandalkan informasi yang diperoleh dari data masa lalu dan saat ini. Proses prediksi muncul sebagai respons terhadap ketidaksesuaian antara pemahaman terhadap suatu peristiwa dengan kebutuhan akan informasi relevan yang dapat membantu pengambilan keputusan di masa depan. Dengan demikian, melalui prediksi, pengambil keputusan dapat merencanakan tindakan yang lebih efektif dan efisien untuk menghadapi kejadian-kejadian di masa depan (Julianto et al., 2021).

Namun, prediksi tidak selalu memberikan jawaban pasti dan benar mengenai apa yang akan terjadi. Sebaliknya, prediksi berusaha memberikan gambaran yang paling mendekati kemungkinan peristiwa yang akan datang. Prediksi berfungsi untuk memberikan informasi yang dapat digunakan untuk merencanakan dan mengambil keputusan strategis, meskipun tetap ada elemen ketidakpastian. Proses prediksi berfokus pada proyeksi keadaan masa depan dengan menggunakan data dan pola yang ada dari masa lalu dan saat ini. Peramalan (*forecasting*) merupakan bagian dari prediksi yang digunakan untuk memproyeksikan keadaan di masa depan berdasarkan pengujian keadaan yang pernah terjadi di masa lalu. Dalam konteks ini, peramalan menciptakan informasi berdasarkan fakta dan data historis untuk memahami situasi sosial masa depan, yang dapat memengaruhi kebijakan yang akan diterapkan (Kafil, 2019).

Prediksi juga dapat dilihat sebagai metode yang digunakan untuk meramalkan kejadian-kejadian yang belum terjadi. Metode ini bersifat terstruktur, dengan tujuan untuk meminimalkan kesalahan dalam memproyeksikan peristiwa masa depan. Berdasarkan

analisis data masa lalu dan informasi yang ada saat ini, prediksi berusaha untuk memperkecil ketidakpastian dan kesalahan yang mungkin timbul dalam peramalan keadaan di masa mendatang (Eka Pratiwi et al., 2022).

Secara keseluruhan, prediksi berperan penting dalam membantu pengambilan keputusan dengan memberikan gambaran mengenai potensi perubahan atau peristiwa yang mungkin terjadi. Dalam aplikasi sistem prediksi (*forecasting system*), gambaran masa lalu diperoleh melalui analisis data yang dikumpulkan dari penelitian yang telah dilakukan, yang kemudian digunakan untuk memproyeksikan peristiwa yang serupa di masa depan (Rohman Hariri & Mashuri, 2022).

#### **2.4 Machine Learning**

*Machine Learning* merupakan bagian dari kecerdasan buatan yang menggunakan berbagai pendekatan probabilistik dan statistik untuk memungkinkan komputer memperoleh pengetahuan dari pengalaman. Pendekatan ini juga memungkinkan komputer untuk mengidentifikasi hubungan dan pola antara *input* dan *output* dalam suatu *dataset* (Chowdhury et al., 2021). *Machine Learning* dapat diklasifikasikan ke dalam empat kategori utama, yaitu *supervised learning*, *unsupervised learning*, *semi-supervised learning*, dan *reinforcement Learning* (Pugliese et al., 2021).

*Supervised Learning* melibatkan pelatihan model dengan data yang diberi label untuk memetakan *input* ke *output* yang diketahui, seperti pada klasifikasi dan regresi. *Unsupervised Learning* menggunakan data tanpa label, memungkinkan model untuk menemukan pola atau struktur tersembunyi, seperti pada *clustering* dan reduksi dimensi. *Semi-Supervised Learning* menggabungkan keduanya, menggunakan sebagian kecil data berlabel dan sebagian besar data tak berlabel untuk meningkatkan akurasi model dengan data terbatas. Sementara itu, *Reinforcement Learning* melibatkan agen yang belajar melalui

interaksi dengan lingkungan, mendapatkan umpan balik berupa *reward* atau *punishment* untuk memaksimalkan keuntungan jangka panjang (Pugliese et al., 2021).

## 2.5 Deep Learning

*Deep Learning*, yang mulai banyak digunakan sejak tahun 2006, menerapkan pendekatan arsitektur pembelajaran mendalam atau pembelajaran hierarkis. Proses pembelajaran ini melibatkan estimasi parameter model agar algoritme yang dirancang mampu menyelesaikan tugas atau masalah tertentu secara efektif (Alom et al., 2019). Metode ini terdiri dari beberapa lapisan yang terletak di antara lapisan *input* dan *output*. Struktur ini memungkinkan pemrosesan data secara *non-linear* melalui beberapa tahapan, yang hasilnya dapat dimanfaatkan untuk pembelajaran fitur (*feature learning*) dan klasifikasi pola (*pattern classification*) (Haris et al., 2021).

*Deep Learning* juga dapat diartikan sebagai salah satu kelas algoritme dalam *Machine Learning* yang memanfaatkan sejumlah lapisan pemrosesan *non-linear* yang tersusun secara berurutan (*cascade*). Proses ini bertujuan untuk melakukan ekstraksi fitur (*feature extraction*) dan transformasi, di mana setiap lapisan mengambil keluaran dari lapisan sebelumnya sebagai inputnya.. Algoritme yang digunakan dapat bertipe *supervised* dan *unsupervised* serta implementasinya dapat digunakan sebagai *pattern analysis* (*unsupervised*) dan klasifikasi (*supervised*) (Haris et al., 2021).

Beberapa teknik dalam i dapat dikategorikan menjadi *supervised*, *semi-supervised*, dan *unsupervised* (Alom et al., 2019).

1. *Deep Supervised Learning*: Teknik *Learning* yang digunakan dalam kategori ini menggunakan data yang telah diberi label sebelumnya (*labeled data*). Contoh yang populer dalam kategori ini adalah *Deep Neural Networks* (DNN), CNN, (RNN), termasuk juga LSTM, dan *Gated Recurrent Unit* (GRU).



2. *Deep Semi-Supervised Learning: Semi-supervised Learning* menggunakan teknik *Learning* yang menggunakan sebagian data yang telah diberi label sebelumnya (partially labeled data). Pada beberapa kasus, DRL, *Generative Adversarial Networks* (GAN), serta RNN, termasuk LSTM, dan GRU juga menggunakan teknik *Learning* ini.
3. *Deep Unsupervised Learning*: Teknik *Learning* ini menggunakan data yang tidak diberi label sebelumnya (*unlabeled data*). *Auto Encoders* (AE), *Restricted Boltzmann Machines* (RBM), dan generasi terbaru dari GAN menggunakan teknik *Learning* ini pada implementasinya.

## 2.6 Time Series

Deret waktu (*Time Series*) adalah sekumpulan data berupa pengamatan yang diukur dalam jangka waktu tertentu berdasarkan selang waktu yang tetap. Model ini berupaya memprediksi kondisi masa depan berdasarkan data masa lalu dan merupakan contoh metode prediksi kuantitatif. (Julianto et al., 2021).

## 2.7 Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX)

Model SARIMA merupakan perluasan dari model SARIMA dengan penambahan peubah lain sebagai peubah *eksogen* atau kovariat (Rochayati et al., 2019). Dalam model ini factor-faktor yang mempengaruhi peubah dependent Y pada waktu ke-t tidak hanya oleh fungsi peubah Y dalam waktu, tetapi juga peubah-peubah independent lainnya pada waktu ke-t. Bentuk umum dari model SARIMAX terdapat pada persamaan (1).

$$\phi_p(B)\phi_p(B^S)(1-B)^d(1-B^S)^DY_t = \mu' + \theta_q(B)\theta_q(B)^S a_t + \alpha_1 X_{1,t} + \dots + \alpha_k X_{k,t} \quad (1)$$

Dengan

$Y_t$  : Nilai pengamatan pada waktu ke-t

$\phi_p, \phi_p$  : Parameter *autoregresif* ke-p



$B$  : Operator backshift

$\theta_q, \theta_Q$  : Parameter moving average

$a_t$  : Nilai kesalahan pada saat  $t$

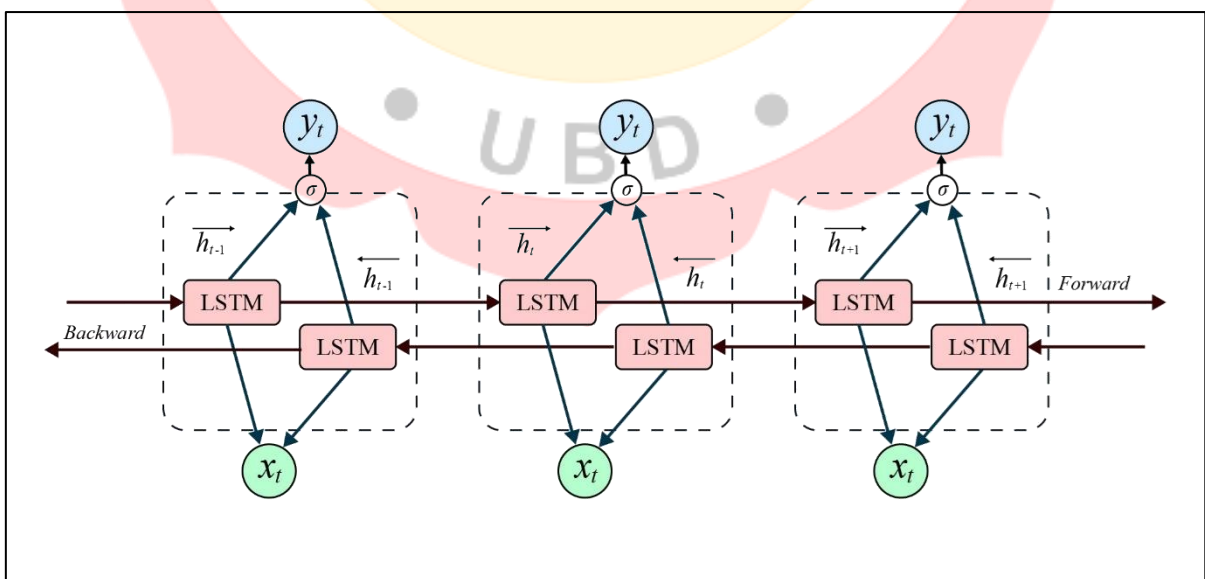
$\mu'$  : Konstanta

## 2.8 Bidirectional Long Short-Term Memory (BiLSTM)

*Bidirectional Long Short-Term Memory* (BiLSTM) merupakan perpanjangan dari jaringan LSTM di mana dua LSTM diterapkan. Satu diterapkan sebelum data masukan, dan yang lainnya diterapkan setelah data masukan (Méndez et al., 2023). BiLSTM memiliki kemampuan untuk secara selektif memberikan bobot pada kata-kata dengan memperhitungkan variasi konteks (Aziz et al., 2023). Model ini terdiri dari tiga komponen utama: Lapisan *Embedding*, Lapisan BiLSTM, dan Lapisan *Attention* (Jbene et al., 2022). Prinsip kerja tiap lapisan BiLSTM serupa dengan LSTM, namun *output* akhirnya merupakan kombinasi dari LSTM yang bekerja dalam dua arah, sebagaimana dijelaskan dalam Persamaan (2):

$$h_t = \xrightarrow{LSTM} (h_{t-1}, x_t, c_{t-1}) + \xleftarrow{L} STM(h_{t+1}, x_t, c_{t+1}) \quad (2)$$

di mana  $h_t$  adalah *output* (Wang et al., 2023).



Gambar 2.1 Struktur BiLSTM

Pada gambar 2.1 yang dirancang untuk mengatasi keterbatasan dengan memproses urutan *input* baik secara maju maupun mundur, lalu menggabungkan informasi dari kedua arah tersebut menjadi satu representasi tunggal. Berbeda dengan LSTM yang hanya mempertimbangkan konteks dari masa lalu, BiLSTM mampu menangkap pola berurutan yang lebih kompleks dalam kedua arah, sehingga menghasilkan representasi fitur yang lebih optimal untuk query input. Matriks bobot pada BiLSTM diinisialisasi secara acak, dan *dropout* diterapkan untuk mencegah overfitting pada model.

## 2.9 Google Trends

*Google Trends* adalah layanan publik di Internet milik Google LLC yang melacak tren permintaan pencarian selama periode waktu tertentu atau ditentukan oleh seberapa sering istilah dicari dan dibandingkan di berbagai wilayah dan bahasa melalui layanan pencarian Google. Google LLC membuat data query penelusuran tersedia untuk umum mulai tahun 2006 hingga saat ini. *Google Trends* memberikan informasi tentang popularitas permintaan pencarian, hanya kumpulan hasil pencarian dari pengguna layanan pencarian Google, tanpa disaring sepenuhnya. Selain sebagai sumber informasi, Google memberikan akses dan mengharapakan layanan *Google Trends* miliknya dapat digunakan untuk berbagai keperluan penelitian. (Julianto et al., 2021; Purnaningrum & Ariqoh, 2019).

## 2.10 Python

*Python* adalah bahasa pemrograman yang bersifat terinterpretasi, interaktif, dan berorientasi objek, dengan kemampuan yang kuat berkat penggabungan modul, pengecualian, tipe data tingkat tinggi, dan kelas. *Python* memiliki sintaks yang sangat jelas, sehingga mudah dipahami (Python Software Foundation, 2024). Keunggulannya mencakup dukungan lintas *platform*, termasuk Linux yang umumnya menyertakan *Python* sebagai bagian standar. *Python* juga menyediakan beragam modul siap pakai, seperti Matplotlib,

Numpy, Pandas, Scikit-learn, dan banyak pustaka lainnya yang mendukung berbagai kebutuhan pengembangan (Kurniawan, 2022)

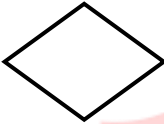
*Python* memiliki beragam kegunaan, termasuk pengembangan aplikasi web, pengelolaan alur kerja, integrasi dengan basis data, pembacaan dan modifikasi file, pemrosesan data dalam skala besar, pelaksanaan operasi matematika, serta mempercepat proses pembuatan prototipe dan pengembangan lebih lanjut perangkat lunak yang telah ada (Ma'arif, 2020).

### 2.11 Flowchart

*Tools* berupa *Flowchart* digunakan agar algoritma lebih mudah dipahami oleh orang lain, terutama programmer yang bertanggung jawab mengimplementasikan program. *Flowchart* adalah representasi grafis yang menggunakan simbol-simbol standar untuk menunjukkan serangkaian langkah atau proses dalam suatu sistem atau program (Putri & Jarti, 2022). Setiap simbol dalam diagram alur mewakili proses tertentu, dan hubungan antar proses ditunjukkan dengan garis penghubung. Diagram alur membantu memperjelas dan memahami alur proses kegiatan. Tabel 2.1 mencantumkan simbol-simbol standar yang umum digunakan dalam diagram alur.

**Tabel 2.1 Simbol Flowchart**

| Simbol                                                                              | Name                | Deskripsi                                              |
|-------------------------------------------------------------------------------------|---------------------|--------------------------------------------------------|
|  | <i>Start/End</i>    | Menunjukkan Dimana sebuah proses dimulai atau berakhir |
|  | <i>Input/Output</i> | Menandakan proses <i>input</i> dan <i>output</i> data  |
|  | <i>Process</i>      | Menunjukkan proses yang dilakukan oleh komputer        |

|                                                                                     |                           |                                                                           |
|-------------------------------------------------------------------------------------|---------------------------|---------------------------------------------------------------------------|
|                                                                                     |                           |                                                                           |
|    | <i>Predefined Process</i> | Menunjukkan pelaksanaan suatu bagian (sub-program)                        |
|    | <i>Flow Line</i>          | Digunakan untuk menghubungkan simbol                                      |
|    | <i>Decision</i>           | Digunakan untuk memberi nilai awal pada suatu variabel.                   |
|    | <i>Preparation</i>        | Digunakan untuk memberi nilai awal pada suatu variabel.                   |
|  | <i>Database</i>           | Digunakan untuk menyatakan database atau basis data.                      |
|  | <i>On-page Reference</i>  | Digunakan untuk menghubungkan <i>Flowchart</i> pada halaman yang sama.    |
|  | <i>On-page Reference</i>  | Digunakan untuk menghubungkan <i>Flowchart</i> pada halaman yang berbeda. |

## 2.12 Tinjauan Studi

Penelitian mengenai prediksi permintaan wisatawan telah berkembang pesat dalam beberapa tahun terakhir, terutama dengan kemajuan dalam metode statistik dan pembelajaran mesin (*machine learning*). Banyak studi yang telah mengeksplorasi berbagai teknik dalam prediksi wisata, termasuk penggunaan model statistik tradisional seperti

ARIMA dan SARIMA, serta metode berbasis kecerdasan buatan seperti LSTM dan BiLSTM.

Beberapa penelitian terdahulu telah mengeksplorasi penggunaan model statistik seperti ARIMA dan SARIMA dalam prediksi wisatawan. Misalnya, penelitian yang dilakukan oleh (Devi et al., 2024) dan (Makoni et al., 2023) menggunakan pendekatan SARIMA dalam memprediksi kedatangan wisatawan, dengan mempertimbangkan pola musiman dan tren jangka panjang. Hasil dari studi tersebut menunjukkan bahwa model ini cukup akurat dalam menangkap pola musiman, tetapi memiliki keterbatasan dalam menangani pola *non-linear* yang sering muncul dalam data pencarian daring dan perilaku wisatawan yang dinamis.

Untuk meningkatkan akurasi prediksi, beberapa penelitian mengusulkan model *hybrid* yang mengombinasikan metode statistik dengan pembelajaran mesin. Misalnya, penelitian oleh (He et al., 2021a) mengembangkan model SARIMA-CNN-LSTM yang berhasil menangkap pola *linear* dan *non-linear* dalam data wisatawan harian. Studi lain oleh (Laaroussi et al., 2023) menggunakan kombinasi Autoencoder dan LSTM untuk meningkatkan akurasi prediksi permintaan wisata berbasis data pencarian *Google Trends*. Hasil penelitian ini menunjukkan bahwa kombinasi model statistik dengan *Deep Learning* dapat meningkatkan ketepatan prediksi dibandingkan model tunggal.

Seiring meningkatnya penggunaan data pencarian daring dalam prediksi wisata, beberapa penelitian telah mengintegrasikan *Google Trends* dan Baidu Index ke dalam model prediksi. Misalnya, penelitian oleh (Costa & Leite, 2022) serta (Wickramasinghe & Ratnasiri, 2021) menunjukkan bahwa data pencarian memiliki korelasi yang kuat dengan kedatangan wisatawan di beberapa negara. Mereka menggunakan model SARIMAX untuk menggabungkan variabel *eksogen* dari data pencarian daring dengan data historis wisatawan.

Hasilnya menunjukkan peningkatan akurasi prediksi dibandingkan dengan model yang hanya menggunakan data historis.

Selain itu, penelitian lain oleh (Hu et al., 2025) memperkenalkan model BiLSTM-MIDAS yang mengombinasikan data berfrekuensi campuran, seperti volume pencabera studi seperti (Chen et al., 2024) menggunakan Boruta untuk seleksi fitur, tetapi belum banyak penelitian yang mengeksplorasi bagaimana interaksi antar fitur dapat mempengaruhi prediksi wisatawan.

Keterbatasan dalam menangani volatilitas data pascapandemi – Studi seperti (Prilistya et al., 2021) telah mengeksplorasi dampak pandemi terhadap prediksi wisatawan, tetapi belum banyak penelitian yang mengembangkan model adaptif yang mampu menyesuaikan diri dengan perubahan pola perjalanan wisatawan setelah pandemi.

Minimnya penelitian tentang interpretabilitas model *hybrid* – Meskipun model berbasis *Deep Learning* telah menunjukkan hasil yang baik dalam menangkap pola non-linear, pemahaman mengenai bagaimana model ini mengambil keputusan masih terbatas, yang dapat mengurangi kepercayaan pengguna dalam menerapkannya di dunia nyata.

Berdasarkan kesenjangan penelitian ini, penelitian ini mengusulkan pendekatan *hybrid* SARIMAX-BiLSTM, yang bertujuan untuk menggabungkan kekuatan model SARIMAX dalam menangkap pola musiman dan hubungan jangka panjang dengan kemampuan BiLSTM dalam mempelajari pola *non-linear* dan dinamis dalam data wisatawan. Dengan menggunakan data historis kedatangan wisatawan serta data pencarian daring dari *Google Trends*, penelitian ini diharapkan dapat memberikan model prediksi yang lebih akurat dan adaptif terhadap perubahan tren pariwisata.

### **2.13 Kesimpulan Tinjauan Studi**

Berdasarkan penelitian sebelumnya, berbagai pendekatan telah digunakan untuk memprediksi permintaan wisatawan. Model klasik seperti SARIMA sering diaplikasikan

untuk menangkap pola musiman, sementara model berbasis *Deep Learning* seperti LSTM unggul dalam memodelkan hubungan temporal yang kompleks. Beberapa penelitian mengembangkan model *hybrid* seperti SARIMAX-CNN, yang memanfaatkan kekuatan SARIMAX untuk pola musiman dan CNN untuk analisis fitur spasial. Selain itu, integrasi data eksternal seperti *Google Trends* terbukti meningkatkan akurasi prediksi dengan menangkap perubahan perilaku pasar secara real-time.

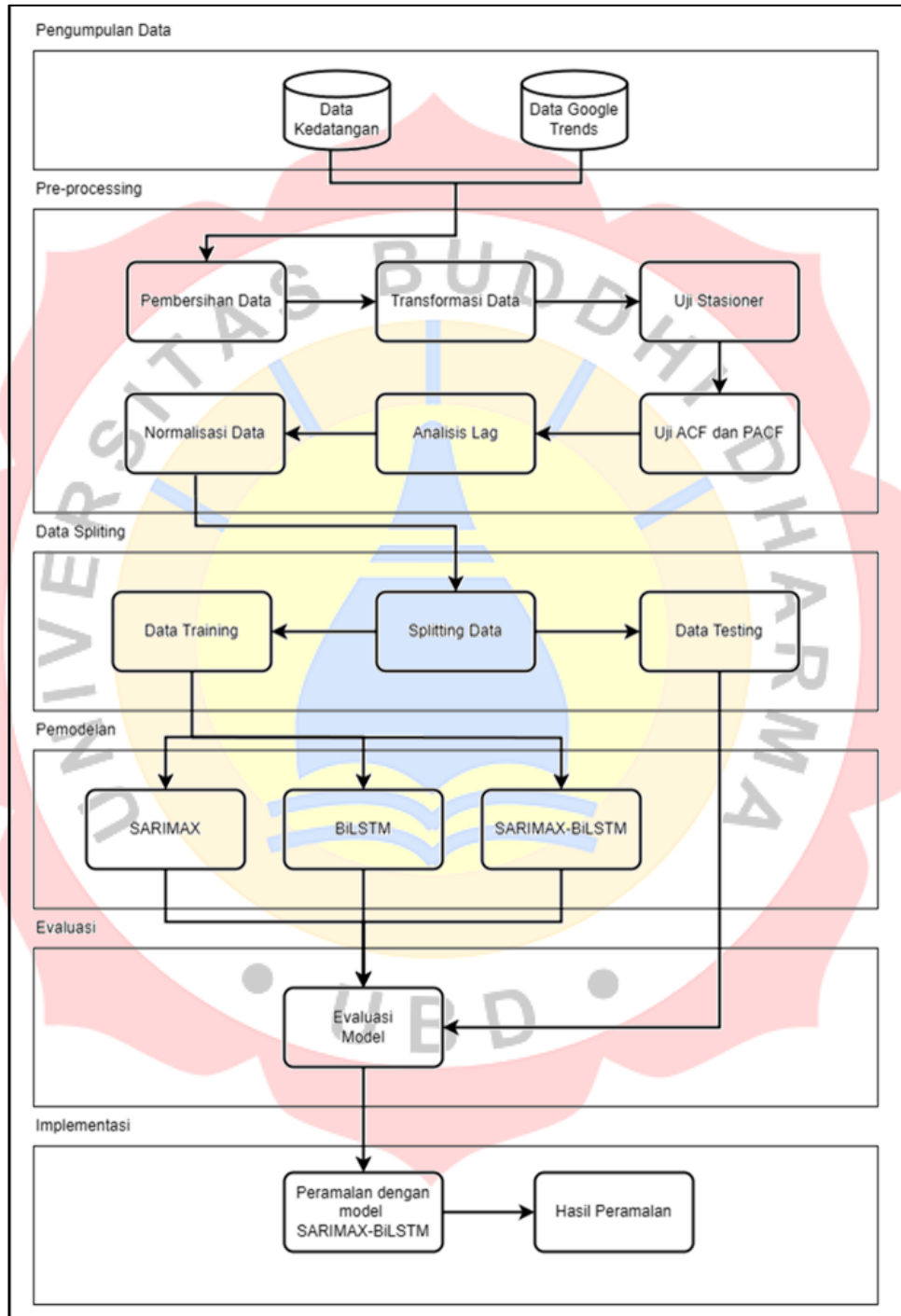
Penelitian ini bertujuan mengeksplorasi pendekatan *hybrid* SARIMAX-BiLSTM, di mana SARIMAX digunakan untuk pola musiman dan pengaruh variabel eksternal, sementara BiLSTM ditambahkan untuk menangkap pola *non-linear* yang tidak dapat ditangani model tradisional seperti SARIMAX, khususnya dalam prediksi kedatangan wisatawan mancanegara di Indonesia. Dengan mengombinasikan data kedatangan wisatawan dari BPS dan *Google Trends*, model ini diharapkan menghasilkan prediksi yang lebih akurat dan adaptif terhadap dinamika pasar pariwisata, memberikan kontribusi baru dalam metode prediksi yang lebih presisi dan relevan.



# BAB III

## METODE PENELITIAN

### 3.1 Kerangka Penelitian



Gambar 3.1 Kerangka Penelitian

Gambar 3.1 menunjukkan alur sistematis pengolahan data hingga implementasi model dalam penelitian ini, yang bertujuan untuk memprediksi kunjungan wisatawan mancanegara menggunakan model *hybrid* SARIMAX-BiLSTM. Proses penelitian dimulai dengan tahap pengumpulan data, yang terdiri dari data kedatangan wisatawan mancanegara dari Badan Pusat Statistik (BPS) dan data minat pencarian daring dari *Google Trends*. Kombinasi kedua jenis data ini bertujuan untuk memberikan gambaran yang lebih komprehensif mengenai pola kunjungan wisatawan, baik berdasarkan data historis maupun perilaku pencarian daring.

Tahap berikutnya adalah *pre-processing*, yang merupakan serangkaian prosedur krusial yang dilakukan untuk mempersiapkan data agar memenuhi syarat penggunaan dalam pemodelan. Proses ini diawali dengan pembersihan data yang bertujuan untuk mengeliminasi anomali serta data yang tidak valid, sehingga kualitas *dataset* dapat terjaga. Selanjutnya, dilakukan transformasi data guna menyesuaikan format dan struktur data sesuai dengan kebutuhan analisis. Tahap selanjutnya adalah pengujian stasioneritas data untuk memastikan bahwa data deret waktu memiliki sifat kestabilan yang diperlukan agar model dapat menghasilkan prediksi yang akurat dan dapat diandalkan. Setelah itu, pengujian *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) diterapkan untuk menentukan parameter optimal yang akan digunakan dalam pemodelan deret waktu. Analisis lag juga dilaksanakan untuk memahami hubungan temporal antar variabel dan menentukan nilai lag terbaik yang akan dimasukkan ke dalam model. Sebagai tahap akhir, normalisasi data dilakukan untuk menyamakan skala variabel sehingga dapat meningkatkan efisiensi dan efektivitas proses pelatihan model.

Setelah proses *pre-processing* selesai, langkah selanjutnya adalah melakukan pembagian data (*data splitting*) dengan membagi *dataset* menjadi dua bagian utama, yaitu *Data Training* sebanyak 70% dan *Data Testing* sebanyak 30%. Pembagian ini dilakukan

agar model dapat dilatih secara optimal menggunakan data historis yang telah diproses, sehingga dapat mempelajari pola dan karakteristik dari data tersebut. Selanjutnya, *Data Testing* yang merupakan data baru dan belum pernah digunakan dalam proses pelatihan dimanfaatkan untuk menguji dan mengevaluasi kinerja model secara objektif, sehingga dapat diketahui seberapa baik model tersebut mampu melakukan prediksi pada data yang tidak dikenal sebelumnya.

Tahap pemodelan merupakan inti dari penelitian ini. Pada tahap ini, penelitian menggunakan tiga pendekatan utama: SARIMAX, BiLSTM, dan model *hybrid* SARIMAX-BiLSTM. Model SARIMAX digunakan untuk menangkap pola musiman dan pengaruh variabel *eksogen* seperti data *Google Trends*. Di sisi lain, model BiLSTM dirancang untuk mempelajari hubungan temporal yang kompleks dan *non-linear* dalam data, yang tidak dapat ditangkap oleh SARIMAX. Kedua model ini digunakan sebagai pembanding untuk mengevaluasi efektivitas model baru, yaitu SARIMAX-BiLSTM. Model *hybrid* ini memadukan kekuatan SARIMAX dalam menangkap pola musiman dan pengaruh variabel eksternal dengan kemampuan BiLSTM dalam memodelkan hubungan temporal yang *non-linear*. SARIMAX-BiLSTM dirancang untuk mengintegrasikan residual SARIMAX dan fitur *eksogen* (seperti data *Google Trends* dengan lag optimal) sebagai *input* ke BiLSTM, menciptakan pendekatan prediktif yang lebih akurat dan adaptif.

Tahap selanjutnya adalah evaluasi model, di mana performa SARIMAX, BiLSTM, dan SARIMAX-BiLSTM dibandingkan untuk menentukan model dengan kinerja terbaik. Evaluasi ini mencakup pengukuran keakuratan prediksi untuk memastikan bahwa model *hybrid* dapat memberikan hasil yang lebih unggul dibandingkan pendekatan individual. Akhirnya, pada tahap implementasi, model SARIMAX-BiLSTM yang telah dioptimalkan digunakan untuk memprediksi kunjungan wisatawan mancanegara secara lebih presisi.

### 3.2 Pengumpulan Data

Pada penelitian ini, data yang digunakan untuk membangun model prediksi kedatangan wisatawan mancanegara ke Bali diperoleh dari dua sumber utama, yaitu data historis kedatangan wisatawan dari Badan Pusat Statistik (BPS) Provinsi Bali dan data minat pencarian daring dari *Google Trends*. Berikut adalah penjelasan detail tentang pengumpulan data:

#### 3.2.1 Data Kedatangan Wisatawan Mancanegara

Data kedatangan wisatawan mancanegara ke Bali diperoleh dari *Website* resmi Badan Pusat Statistik (BPS) Provinsi Bali, yang merupakan sumber data resmi dan terpercaya dalam bidang statistik pariwisata di Indonesia. Data yang digunakan mencakup jumlah kedatangan wisatawan mancanegara setiap bulan, dimulai dari Januari 2009 hingga Desember 2024, sehingga mencakup periode waktu selama 16 tahun secara berturut-turut. Dengan rentang waktu yang cukup panjang ini, data diharapkan mampu memberikan gambaran yang komprehensif dan representatif mengenai tren dan pola kedatangan wisatawan ke Bali selama periode tersebut.

Total data yang terkumpul mencapai 192 titik data bulanan, yang masing-masing merepresentasikan jumlah kedatangan wisatawan mancanegara di Bali setiap bulan dalam kurun waktu tersebut. Jumlah data yang cukup besar ini memungkinkan analisis deret waktu yang lebih mendalam dan pemodelan statistik maupun *Machine Learning* yang lebih akurat, karena model dapat mempelajari pola musiman, tren jangka panjang, serta fluktuasi yang terjadi sepanjang waktu. Data ini menjadi fondasi utama dalam penelitian karena langsung mencerminkan kondisi nyata dan aktual terkait kunjungan wisatawan internasional ke Bali.

Selain itu, data dari BPS ini dianggap sangat valid dan mencakup kedatangan wisatawan melalui berbagai pintu masuk utama, baik melalui jalur udara maupun jalur

laut. Hal ini menjadikan data tersebut sebagai sumber yang komprehensif untuk mengukur aktivitas pariwisata secara keseluruhan di Bali. Dengan demikian, penggunaan data ini dalam penelitian memberikan landasan yang kuat untuk melakukan analisis dan prediksi kedatangan wisatawan yang dapat diandalkan serta relevan dengan kondisi riil di lapangan.

### **3.2.2 Data *Google Trends***

Selain menggunakan data historis dari Badan Pusat Statistik (BPS), penelitian ini juga mengintegrasikan data dari *Google Trends* untuk menangkap pola minat pencarian daring yang berkaitan dengan pariwisata di Bali. Penggunaan *Google Trends* memungkinkan analisis yang lebih dinamis karena data ini dapat mencerminkan perubahan minat dan perilaku calon wisatawan secara real-time, yang sulit ditangkap oleh data statistik konvensional. Dengan demikian, data *Google Trends* menjadi pelengkap penting dalam memperkaya pemahaman terhadap tren wisatawan internasional yang berkunjung ke Bali.

Fokus pengumpulan data *Google Trends* diarahkan pada sepuluh negara utama yang memiliki rata-rata kunjungan wisatawan tertinggi ke Bali selama periode 2019 hingga 2022, sesuai laporan resmi BPS dan ditampilkan pada Tabel 3.1. Negara-negara tersebut meliputi Australia, Cina, India, Jepang, United Kingdom, Amerika Serikat, Malaysia, Perancis, Korea Selatan, dan Jerman. Pemilihan negara-negara ini bertujuan agar analisis dapat difokuskan pada pasar-pasar utama yang memberikan kontribusi signifikan terhadap jumlah wisatawan yang datang ke Bali.

**Tabel 3.1 Rata-rata Kedatangan Wisatawan Mancanegara ke Bali 2014-2022**

| No | Negara                      | Rata-rata |
|----|-----------------------------|-----------|
| 1  | Australia / Australia       | 922359    |
| 2  | RRC / China                 | 788153    |
| 3  | India / India               | 203858    |
| 4  | Jepang / Japan              | 189336    |
| 5  | Inggris / United Kingdom    | 185846    |
| 6  | Amerika Serikat / USA       | 159255    |
| 7  | Malaysia / Malaysia         | 157503    |
| 8  | Perancis / France           | 141143    |
| 9  | Korea Selatan / South Korea | 133295    |
| 10 | Jerman / Germany            | 132802    |
| 11 | Singapura / Singapore       | 130743    |
| 12 | Taiwan / Taiwan             | 88968     |
| 13 | Belanda / Holland           | 82208     |
| 14 | Rusia / Russia              | 81947     |
| 15 | Selandia Baru / New Zealand | 53737     |

Untuk setiap negara, pengumpulan data pencarian dilakukan dengan menggunakan kata kunci utama “Bali” serta sembilan destinasi wisata populer di Bali, antara lain “Kuta”, “Ubud”, “Seminyak”, “Denpasar”, “Sanur”, “Nusa Dua”, “Jimbaran”, “Amed”, dan “Uluwatu”, sebagaimana tercantum pada Tabel 3.2. Data ini dikumpulkan dalam skala waktu bulanan, mencakup periode Januari 2004 hingga Desember 2024, sehingga cakupan waktunya konsisten dengan data kedatangan wisatawan dari BPS. Dengan menggabungkan data *Google Trends* dan data BPS, penelitian ini mampu menangkap perubahan minat pasar secara lebih akurat dan real-time, sekaligus memberikan fondasi yang kuat untuk analisis dan prediksi kedatangan wisatawan ke Bali.

**Tabel 3.2 Jumlah Score Keyword *Google Trends* Destinasi Wisata Bali**

| No | Keyword   | Jumlah Score Google Trends |
|----|-----------|----------------------------|
| 1  | Bali      | 106872                     |
| 2  | Kuta      | 73232                      |
| 3  | Ubud      | 69450                      |
| 4  | Nusa Dua  | 62697                      |
| 5  | Denpasar  | 61098                      |
| 6  | Seminyak  | 60354                      |
| 7  | Sanur     | 57258                      |
| 8  | Jimbaran  | 56571                      |
| 9  | Amed      | 48569                      |
| 10 | Uluwatu   | 41776                      |
| 11 | Lovina    | 40000                      |
| 12 | Canggu    | 38616                      |
| 13 | Kintamani | 25999                      |
| 14 | Sidemen   | 17825                      |
| 15 | Munduk    | 17245                      |
| 16 | Bedugul   | 8808                       |
| 17 | Medewi    | 4225                       |



**Tabel 3.3 Data Wisatawan Mancanegara Bulanan ke Bali Menurut Pintu Masuk 2009**

| Pintu Masuk | Januari | Februari | Maret  | April  | Mei    | Juni   | Juli   | Agustus | September | Oktober | November | Desember |
|-------------|---------|----------|--------|--------|--------|--------|--------|---------|-----------|---------|----------|----------|
| Bandara     | 173867  | 146115   | 167954 | 188189 | 190638 | 200503 | 234984 | 232090  | 218199    | 221059  | 184622   | 221604   |
| Pelabuhan   | 674     | 1589     | 251    | 587    | 165    | 63     | 214    | 165     | 244       | 223     | 181      | 942      |
| Jumlah      | 174541  | 147704   | 168205 | 188776 | 190803 | 200566 | 235198 | 232255  | 218443    | 221282  | 184803   | 222546   |

**Tabel 3.4 Data Wisatawan Mancanegara Bulanan ke Bali Menurut Pintu Masuk 2009 *Cleaning***

| Januari | Februari | Maret  | April  | Mei    | Juni   | Juli   | Agustus | September | Oktober | November | Desember |
|---------|----------|--------|--------|--------|--------|--------|---------|-----------|---------|----------|----------|
| 174541  | 147704   | 168205 | 188776 | 190803 | 200566 | 235198 | 232255  | 218443    | 221282  | 184803   | 222546   |

**Tabel 3.5 Data Kedatangan Wisatawan Mancanegara ke Bali 2009-2024**

| Data       | Arrival |
|------------|---------|
| 2009-01-01 | 174541  |
| 2009-02-01 | 147704  |
| 2009-03-01 | 168205  |
| 2009-04-01 | 188776  |
| 2009-05-01 | 190803  |
| 2009-06-01 | 200566  |
| ...        | ...     |
| 2024-12-01 | 551100  |

### 3.3 Data Understanding

#### 3.3.1 Data Kedatangan Wisatawan Mancanegara

Tabel 3.3 menyajikan gambaran rinci mengenai struktur *dataset* yang digunakan dalam penelitian ini. *Dataset* tersebut diperoleh dari Badan Pusat Statistik (BPS), yang merupakan lembaga resmi penyedia data statistik di Indonesia. Oleh karena itu, data yang digunakan dalam penelitian ini dapat dianggap valid dan dapat dipercaya sebagai representasi akurat dari kedatangan wisatawan mancanegara ke Bali. Data ini mencakup jumlah wisatawan yang datang ke Bali setiap bulan, dengan rincian berdasarkan pintu masuk utama yaitu bandara dan pelabuhan, serta total kedatangan setiap bulannya selama periode 2009 hingga 2024.

Secara keseluruhan, *dataset* ini terdiri dari 180 poin data yang mewakili jumlah kedatangan wisatawan per bulan selama 16 tahun tersebut. Keberadaan data yang cukup lengkap dan terperinci ini memungkinkan analisis tren musiman, pola fluktuasi, dan perubahan jangka panjang yang dapat dianalisis secara menyeluruh untuk kepentingan prediksi. Selain itu, dengan adanya data terpisah berdasarkan pintu masuk melalui udara dan laut, penelitian dapat lebih mendalam dalam mengevaluasi kontribusi masing-masing jalur terhadap total kedatangan wisatawan.

*Dataset* ini memiliki atribut berikut:

- a. Januari – Desember : Mewakili jumlah wisatawan yang datang setiap bulan.
- b. Bandara : Jumlah wisatawan yang masuk melalui jalur udara.
- c. Pelabuhan : Jumlah wisatawan yang masuk melalui jalur laut.
- d. Jumlah : Total jumlah kedatangan wisatawan (bandara + pelabuhan).

#### 3.3.2 Data Google Trends

*Dataset Google Trends* digunakan untuk mencerminkan minat pencarian daring terkait pariwisata Bali dari sepuluh negara utama yang memiliki rata-rata kedatangan

wisatawan tertinggi. Negara-negara tersebut meliputi Australia, China, India, Jepang, Amerika Serikat, Inggris, Malaysia, Prancis, Korea Selatan, dan Jerman. Data ini disusun dalam format bulanan dan mencakup rentang waktu yang panjang, yaitu dari Januari 2004 hingga Desember 2024, sebagaimana dijelaskan lebih lanjut dalam Tabel 3.6. Pendekatan ini memungkinkan analisis tren minat pencarian wisatawan secara rinci dan konsisten selama dua dekade terakhir, sekaligus menyediakan data *real-time* yang relevan untuk prediksi pariwisata.

Jumlah data untuk setiap kata kunci terdiri dari 180 baris, yang merepresentasikan jumlah bulan selama periode pengamatan, dikalikan dengan sepuluh negara, sehingga menghasilkan total 1.800 baris data untuk satu kata kunci. Data tersebut kemudian dikalikan dengan sepuluh kata kunci pencarian yang digunakan dalam penelitian, yang mencakup kata kunci utama “Bali” dan sembilan destinasi wisata populer di Bali. Dengan demikian, secara keseluruhan *dataset* ini berjumlah sekitar 18.000 baris data, yang mencerminkan intensitas pencarian wisatawan dari berbagai negara terhadap Bali dan destinasi terkait secara detail.

Ketersediaan *dataset Google Trends* dalam jumlah besar ini memungkinkan pemodelan yang lebih komprehensif dan dinamis, karena data ini tidak hanya memberikan gambaran mengenai perilaku wisatawan yang sudah terjadi, tetapi juga dapat merefleksikan minat dan kecenderungan yang sedang berkembang. Hal ini sangat penting untuk mendukung prediksi kedatangan wisatawan secara lebih akurat, terutama dalam menghadapi perubahan perilaku wisatawan yang cepat dan dinamis di era digital saat ini.

*Dataset* ini memiliki dua atribut utama sebagai berikut:

- a. Bulan : Menunjukkan waktu pencarian dalam format bulanan.

- b. *Keyword* : Destinasi wisata di Bali yang menjadi kata kunci pencarian, disesuaikan dengan bahasa masing-masing negara.

**Tabel 3.6 Data Google Trends Kuta di Australia**

| Month   | Kuta: (Australia) |
|---------|-------------------|
| 2004-01 | 64                |
| 2004-02 | 44                |
| 2004-03 | 38                |
| 2004-04 | 47                |
| 2004-05 | 40                |
| 2004-06 | 53                |
| ...     | ...               |
| 2024-12 | 34                |

### 3.4 Data Preprocessing

#### 3.4.1 Pembersihan Data

Dalam penelitian ini, pembersihan dilakukan pada dua jenis data utama, yaitu data kedatangan wisatawan mancanegara yang diperoleh dari Badan Pusat Statistik (BPS) dan data *Google Trends* yang menangkap minat pencarian daring terkait pariwisata di Bali.

##### a. Data Kedatangan Wisatawan Mancanegara

Berdasarkan hasil pemeriksaan menyeluruh terhadap *dataset* yang diperoleh dari Badan Pusat Statistik (BPS), tidak ditemukan adanya data yang hilang (*missing values*) maupun inkonsistensi yang signifikan. Kondisi ini menunjukkan bahwa kualitas data yang digunakan dalam penelitian sudah cukup baik dan tidak memerlukan penanganan khusus terkait aspek kebersihan dan kelengkapan data. Keberadaan data yang lengkap dan konsisten ini sangat penting untuk memastikan bahwa analisis yang dilakukan dapat berlangsung tanpa kendala dan memberikan hasil yang valid serta dapat dipercaya.

Namun demikian, untuk menyederhanakan proses analisis dan memfokuskan penelitian pada aspek yang paling relevan, dilakukan penghapusan beberapa fitur atau kolom yang berisi rincian kedatangan wisatawan berdasarkan moda transportasi, yaitu kedatangan melalui bandara dan pelabuhan laut. Langkah ini diambil dengan pertimbangan bahwa fokus utama penelitian adalah pada total jumlah kedatangan wisatawan mancanegara ke Bali, tanpa memerlukan detail yang memisahkan moda kedatangan. Penghapusan fitur ini juga bertujuan agar analisis menjadi lebih efisien dan mudah dipahami, tanpa mengurangi esensi dari data yang dianalisis.

Proses pembersihan dan penyederhanaan *dataset* ini memastikan bahwa data yang digunakan dalam penelitian sudah sesuai dengan tujuan dan ruang lingkup analisis. Dengan demikian, hasil yang diperoleh diharapkan lebih akurat dan relevan dengan konteks penelitian. Sebagai gambaran, Tabel 3.3 menunjukkan struktur *dataset* sebelum proses pembersihan, di mana data masih mencakup rincian kedatangan melalui bandara dan pelabuhan laut. Sedangkan Tabel 3.4 memperlihatkan struktur *dataset* setelah proses pembersihan, yang kini hanya memuat total kedatangan wisatawan dari seluruh pintu masuk. Kedua tabel tersebut memberikan ilustrasi visual yang jelas mengenai perubahan struktur data yang dilakukan demi memenuhi kebutuhan analisis penelitian ini.

#### **b. Data Google Trends**

Pemeriksaan menyeluruh terhadap *dataset Google Trends* dilakukan untuk memastikan kualitas dan kelengkapan data yang akan digunakan dalam analisis. Dari hasil pemeriksaan tersebut, tidak ditemukan adanya data yang hilang (*missing values*) maupun inkonsistensi yang signifikan dalam *dataset*. Kondisi ini menunjukkan bahwa *dataset Google Trends* yang digunakan memiliki integritas

data yang baik dan siap untuk dianalisis lebih lanjut tanpa perlu dilakukan penanganan khusus terkait masalah kualitas data.

Untuk menjaga konsistensi analisis dan kesesuaian dengan data kedatangan wisatawan yang diperoleh dari Badan Pusat Statistik (BPS), dilakukan penyaringan data *Google Trends* dengan menghapus data yang berada di luar rentang waktu penelitian, yaitu sebelum Januari 2009 dan setelah Desember 2023. Langkah ini diambil agar fokus penelitian tetap terjaga pada periode yang sama dengan data utama, sehingga hasil analisis dapat dipertanggungjawabkan dan relevan dengan konteks waktu yang dianalisis. Awalnya, *dataset Google Trends* mencakup data mulai dari Januari 2004 hingga Desember 2024, sehingga proses ini merupakan upaya untuk menyaring data agar selaras dengan cakupan waktu penelitian.

Perubahan dan penyesuaian *dataset Google Trends* ini dijelaskan secara visual melalui dua tabel utama, yaitu Tabel 3.6 dan Tabel 3.7. Tabel 3.6 memperlihatkan struktur *dataset* asli yang mencakup periode Januari 2004 hingga Desember 2024, sedangkan Tabel 3.7 menunjukkan *dataset* akhir setelah dilakukan penyaringan sehingga hanya mencakup data dari Januari 2009 hingga Desember 2023. Kedua tabel ini memberikan gambaran yang jelas mengenai proses transformasi dan penyesuaian *dataset Google Trends*, sehingga memastikan bahwa data yang dianalisis benar-benar relevan dan konsisten dengan tujuan penelitian.

**Tabel 3.7 Data *Google Trends* Kuta di Australia *Cleaning***

| Month | Kuta: (Australia) |
|-------|-------------------|
|-------|-------------------|

|         |     |
|---------|-----|
| 2009-01 | 63  |
| 2009-02 | 61  |
| 2009-03 | 64  |
| 2009-04 | 59  |
| 2009-05 | 67  |
| 2009-06 | 69  |
| ...     | ... |
| 2024-12 | 31  |

### 3.4.2 Transformasi Data

#### a. Data Kedatangan Wisatawan Mancanegara

Proses transformasi *dataset* kedatangan wisatawan mancanegara dilakukan dengan tujuan utama untuk mempermudah analisis deret waktu (*Time Series*) yang merupakan metode utama dalam penelitian ini. Pada langkah awal, data yang sebelumnya disusun dalam format kolom berdasarkan bulan (dari Januari hingga Desember) diubah menjadi format baris. Transformasi ini penting agar data menjadi lebih terstruktur dan kompatibel dengan alat analisis deret waktu yang umumnya memproses data dalam format baris sesuai urutan waktu. Dengan demikian, pengolahan dan pemodelan data dapat dilakukan dengan lebih efisien dan tepat.

Selain itu, untuk menyelaraskan data dengan standar analisis waktu yang umum digunakan, nama bulan pada *dataset* diubah menjadi format tanggal lengkap. Contohnya, data untuk bulan Januari diubah menjadi “2009-01-01”. Pengubahan ini memudahkan dalam pengindeksan waktu, pemrosesan, serta visualisasi data dalam bentuk grafik atau model prediksi yang memerlukan informasi tanggal yang spesifik dan terstandarisasi. Format tanggal ini juga memperlancar proses integrasi data dengan variabel lain yang memiliki skala waktu serupa.



Selanjutnya, seluruh data yang sudah tertransformasi dari periode Januari 2009 hingga Desember 2024 digabungkan menjadi satu file tunggal, yang memungkinkan pengelolaan *dataset* secara menyeluruh dan konsisten. Proses penggabungan ini ditampilkan secara jelas pada Tabel 3.5, yang memperlihatkan struktur data akhir yang seragam dan siap untuk dianalisis lebih lanjut. Dengan format data yang konsisten ini, analisis deret waktu dapat dilakukan dengan lebih efektif dan hasil prediksi yang dihasilkan menjadi lebih akurat.

#### **b. Data Google Trends**

Transformasi pada data *Google Trends* dilakukan untuk menyelaraskan struktur data dengan kebutuhan analisis dalam penelitian ini. Proses ini melibatkan penggabungan seluruh file data yang berasal dari berbagai kata kunci pencarian dan negara asal wisatawan, sehingga membentuk satu *dataset* yang terintegrasi dan komprehensif. Dengan demikian, data menjadi lebih terorganisir dan siap untuk dianalisis secara menyeluruh, memungkinkan pemahaman yang lebih baik terhadap pola minat pencarian yang terjadi.

Selain penggabungan data, dilakukan agregasi skor *Google Trends* pada tingkat bulanan dengan menjumlahkan nilai pencarian untuk setiap kata kunci dan negara dalam satu bulan tertentu. Langkah ini menghasilkan representasi minat pencarian agregat yang lebih sederhana namun tetap informatif. Hasil agregasi ini memudahkan dalam mengidentifikasi tren dan pola minat pencarian secara keseluruhan maupun secara spesifik per negara asal wisatawan, seperti yang dapat dilihat pada Tabel 3.8.

Dengan melakukan transformasi dan agregasi data ini, penelitian dapat menggambarkan gambaran yang lebih komprehensif dan akurat tentang minat pencarian daring yang berkaitan dengan pariwisata Bali. Periode analisis yang

mencakup Januari 2009 hingga Desember 2024 memungkinkan evaluasi yang konsisten dengan data kedatangan wisatawan, sehingga integrasi kedua sumber data tersebut dapat meningkatkan kualitas dan relevansi hasil analisis.

**Tabel 3.8 Data Google Trends 2009-2024**

| Month      | Total |
|------------|-------|
| 2009-01-01 | 2008  |
| 2009-02-01 | 2185  |
| 2009-03-01 | 2310  |
| 2009-04-01 | 2667  |
| 2009-05-01 | 2718  |
| ...        | ...   |
| 2024-12-01 | 3449  |

### 3.4.3 Uji Stasioner

Uji stasioner dilakukan untuk memverifikasi apakah data yang digunakan dalam penelitian ini memenuhi syarat stasioneritas, yang merupakan prasyarat penting dalam analisis deret waktu. Metode *Augmented Dickey-Fuller* (ADF), salah satu pendekatan uji akar unit, digunakan untuk memeriksa stasioneritas data deret waktu (Devi et al., 2024). Uji ADF diterapkan pada data kedatangan wisatawan mancanegara dari BPS dan data *Google Trends* untuk memastikan bahwa data memenuhi kriteria stasioneritas.

Dalam pemodelan deret waktu, tahap identifikasi memainkan peran penting dalam menentukan apakah data deret waktu bersifat stasioner atau non-stasioner. Jika data tersebut non-stasioner, maka diperlukan evaluasi untuk menentukan berapa banyak perbedaan (*differencing*) yang diperlukan agar data menjadi stasioner (Devi et al., 2024). Proses ini memastikan bahwa data yang digunakan

dalam pemodelan telah memenuhi persyaratan analisis deret waktu sehingga hasil prediksi lebih akurat dan valid.

Berdasarkan hasil uji awal, data kedatangan wisatawan menunjukkan nilai ADF sebesar **-2.6221** dengan p-value **0.0885**, yang lebih besar dari tingkat signifikansi 5%. Dengan demikian, data ini tidak stasioner pada tingkat awal, sebagaimana ditunjukkan pada Tabel 3.9. Oleh karena itu, dilakukan proses *differencing* untuk mengatasi masalah ini. Setelah *differencing*, data menunjukkan nilai ADF sebesar **-2.9479** dengan p-value **0.0400**, yang berada di bawah tingkat signifikansi 5%, sebagaimana ditampilkan pada Tabel 3.10. Hasil ini menunjukkan bahwa data kedatangan wisatawan mancanegara telah memenuhi syarat stasioneritas setelah proses *differencing*.

**Tabel 3.9 Uji Stasioner**

| ADF Statistic      | lag order | p-value             |
|--------------------|-----------|---------------------|
| -2.622113637018492 | 12        | 0.08851241290237721 |

**Tabel 3.10 Uji Stasioner *Differencing***

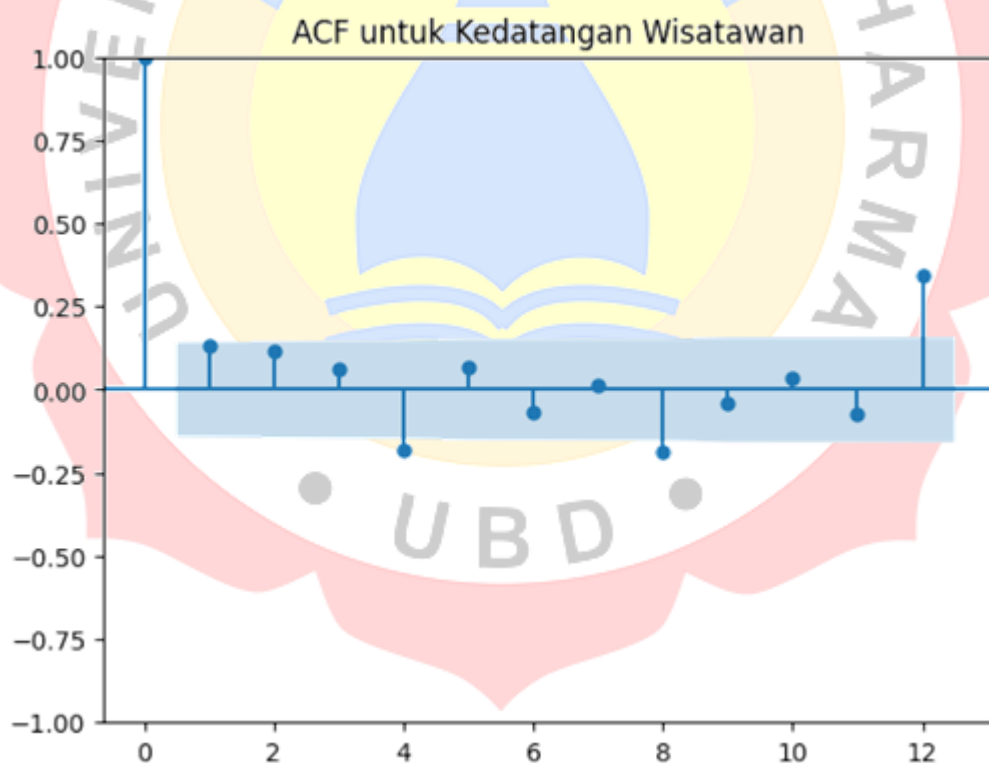
| ADF Statistic      | lag order | p-value             |
|--------------------|-----------|---------------------|
| -2.947921194930253 | 11        | 0.04005249190620248 |

#### 3.4.4 Uji ACF dan PACF

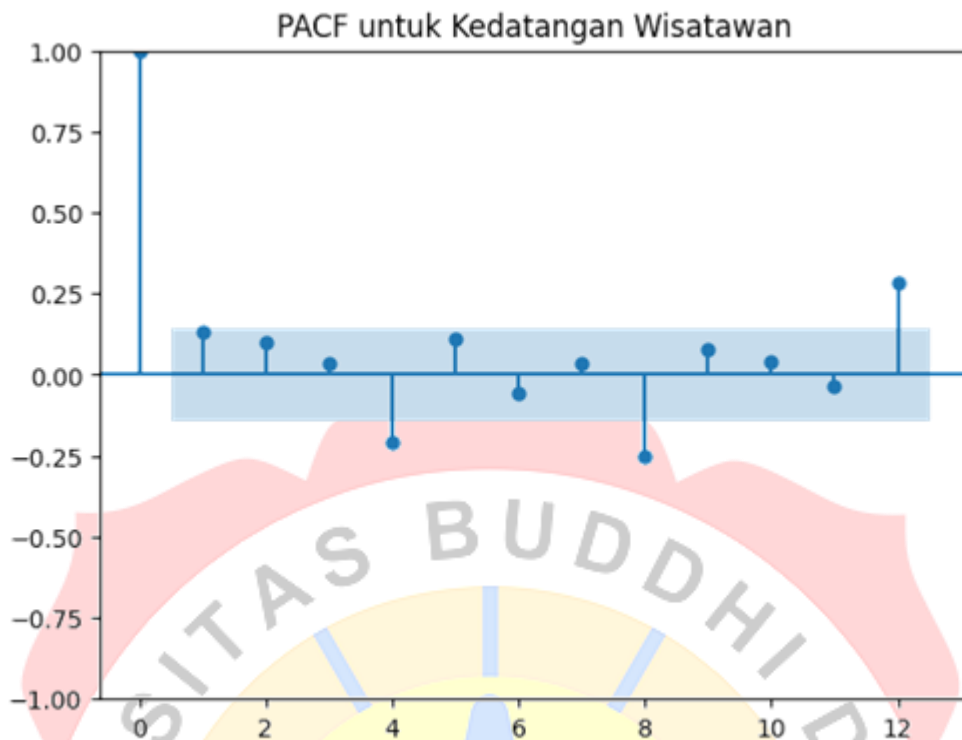
Setelah data dinormalisasi, langkah selanjutnya adalah melakukan uji *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) untuk menentukan parameter yang sesuai dalam pemodelan deret waktu, khususnya untuk model SARIMAX-BiLSTM. Hasil analisis ACF tergambar pada Gambar 3.2 dan

Gambar 3.2 menunjukkan bahwa terdapat korelasi yang sangat tinggi pada lag 1, yang mengindikasikan bahwa nilai saat ini memiliki hubungan yang kuat dengan nilai pada lag 1. Sebagian besar lag lainnya berada dalam interval kepercayaan, sehingga dianggap tidak signifikan. Berdasarkan hasil ini, parameter  $q$  (*Moving Average*) adalah 1.

Pada uji PACF hasil menunjukkan korelasi parsial yang sangat tinggi pada lag 1 tergambar pada Gambar 3.3, yang mengindikasikan pengaruh langsung lag 1 terhadap nilai saat ini. Sebagian besar lag lainnya juga berada dalam interval kepercayaan dan tidak signifikan. Berdasarkan hasil ini, parameter  $p$  (*Autoregressive*) adalah 1. Dengan demikian, hasil uji ACF dan PACF memberikan panduan penting dalam menentukan struktur parameter model SARIMAX untuk memastikan akurasi dalam pemodelan prediksi.



**Gambar 3.2 ACF Kedatangan Wisatawan Mancanegara**



**Gambar 3.3 PACF Kedatangan Wisatawan Mancanegara**

### 3.4.5 Analisis Lag

Uji lag menggunakan metode cross-correlation dilakukan untuk memahami hubungan temporal antar data yang relevan dalam pemodelan deret waktu dengan pendekatan SARIMAX-BiLSTM. Hasil uji, sebagaimana ditunjukkan pada Tabel 3.11, menunjukkan bahwa korelasi positif tertinggi ditemukan pada lag 1 dengan nilai 0.250843. Ini mengindikasikan bahwa nilai data pada satu periode sebelumnya memiliki pengaruh yang cukup kuat terhadap nilai saat ini. Sementara itu, lag 0 menunjukkan korelasi positif sebesar 0.075175, yang turut mendukung adanya hubungan langsung antara nilai sekarang dan periode sebelumnya. Selain itu, lag 2 dan lag 3 masing-masing mencatat nilai korelasi positif sebesar 0.188481 dan 0.140522, meskipun kontribusinya lebih kecil dibandingkan lag 1. Sebaliknya, pada lag negatif seperti lag -1, -2, dan -3, nilai korelasi ditemukan kecil atau bahkan negatif, masing-masing sebesar -0.029965, -0.120452, dan -0.140734. Ini menunjukkan bahwa nilai pada masa lalu sebelum periode

saat ini memiliki hubungan yang lemah atau bahkan berlawanan arah terhadap nilai saat ini.

Untuk memperdalam pemahaman hubungan ini, dilakukan pula pergeseran (shifting) data *Google Trends*, dengan tujuan mengkaji apakah pergeseran waktu dapat meningkatkan korelasi terhadap data kunjungan wisatawan. Hasil perbandingan sebelum dan sesudah shifting ditampilkan pada Tabel 3.12. Berdasarkan hasil ini, terlihat bahwa setelah dilakukan shifting, korelasi antara data *Google Trends* dan data kunjungan wisatawan meningkat, yang mendukung penggunaan variabel exogenous *Google Trends* dalam model SARIMAX-BiLSTM. Dengan demikian, lag 1 diidentifikasi sebagai lag paling signifikan yang memengaruhi pola data, dan shifting data *Google Trends* memperkuat hubungan tersebut. Penyesuaian ini diharapkan dapat meningkatkan akurasi prediksi kedatangan wisatawan serta mengoptimalkan integrasi data minat pencarian daring ke dalam model prediktif.

**Tabel 3.11 Analisis Lag**

| Lag | Cross-Correlation |
|-----|-------------------|
| -3  | -0.140734         |
| -2  | -0.120452         |
| -1  | -0.029965         |
| 0   | 0.075175          |
| 1   | 0.250843          |
| 2   | 0.188481          |
| 3   | 0.140522          |

**Tabel 3.12 Perbandingan Sebelum dan Sesudah *Shifting***

| date       | google_trends | google_trends_lag1 |
|------------|---------------|--------------------|
| 2009-01-01 | 2016          | 0                  |
| 2009-02-01 | 2313          | 2016               |
| 2009-03-01 | 2499          | 2313               |

|            |      |      |
|------------|------|------|
| 2009-04-01 | 2867 | 2499 |
| 2009-05-01 | 2809 | 2867 |
| ...        | ...  |      |
| 2023-12-01 | 3024 | 2839 |

### 3.4.6 Normalisasi Data

Setelah memastikan bahwa data bersifat stasioner melalui uji stasioner dan proses *differencing*, langkah selanjutnya adalah melakukan normalisasi data untuk memastikan bahwa semua variabel memiliki skala yang seragam, sehingga mempermudah proses analisis dan meningkatkan stabilitas model yang digunakan. Normalisasi dilakukan dengan metode Min-Max Scaling, yang mengubah nilai-nilai dalam *dataset* ke dalam rentang [0,1] untuk mengurangi variabilitas skala antara variabel yang berbeda. Hal ini penting karena *dataset* mencakup data kedatangan wisatawan mancanegara dan data *Google Trends*, yang memiliki rentang nilai yang sangat berbeda.

Proses normalisasi ini tidak hanya bertujuan untuk menyederhanakan kompleksitas pemodelan, tetapi juga untuk memetakan semua fitur ke dalam interval [0,1], sehingga membantu melatih model dengan lebih baik dan mempercepat konvergensi selama proses pelatihan (Chen et al., 2024; Laaroussi et al., 2023). Data yang telah dinormalisasi menjadi seperti yang tergambar pada Tabel 3.13. Proses ini juga bertujuan untuk mencegah bias akibat perbedaan skala besar pada data, serta meningkatkan kinerja model analisis, khususnya dalam penerapan model SARIMAX-BiLSTM. Dengan langkah ini, data menjadi siap untuk digunakan dalam pemodelan prediksi menggunakan pendekatan SARIMAX-BiLSTM.

**Tabel 3.13 Data Setelah Normalisasi**

| date       | arrival_diff_scaled | google_trends_lag1_scaled |
|------------|---------------------|---------------------------|
| 2009-01-01 | 0.6385167357        | 0                         |



|            |              |              |
|------------|--------------|--------------|
| 2009-02-01 | 0.5362265087 | 0.4137646816 |
| 2009-03-01 | 0.7166570622 | 0.4502369668 |
| 2009-04-01 | 0.7169238698 | 0.4759942304 |
| 2009-05-01 | 0.6462427218 | 0.5495569751 |
| ...        | ...          |              |
| 2024-12-01 | 0.9365789539 | 0.7582938389 |

### 3.5 Data Splitting

*Dataset* yang mencakup data kedatangan wisatawan mancanegara dan data *Google Trends* terlebih dahulu digabungkan untuk membentuk satu *dataset* komprehensif yang mencakup berbagai aspek baik dari data historis maupun data real-time. Setelah proses penggabungan selesai, *dataset* tersebut kemudian dibagi menjadi dua bagian utama, yaitu *Data Training* dan *Data Testing*, yang masing-masing memiliki peran penting dalam proses pelatihan dan evaluasi model. Pembagian data dilakukan berdasarkan periode waktu, di mana *Data Training* meliputi rentang waktu dari Januari 2009 hingga Desember 2019, sedangkan *Data Testing* mencakup periode Januari 2020 hingga Desember 2024. Dengan pembagian ini, dihasilkan sebanyak 132 *Data Training* dan 60 *Data Testing* yang secara proporsional mendekati rasio 70:30 yang umum digunakan dalam praktik data science, sehingga diharapkan dapat memberikan keseimbangan yang optimal antara proses pelatihan dan pengujian model (Joseph, 2022). Rincian lengkap mengenai distribusi data pada masing-masing subset ini dapat dilihat secara terperinci pada Tabel 3.14 untuk *Data Training* dan Tabel 3.15 untuk *Data Testing*, yang memperlihatkan susunan serta cakupan waktu dari masing-masing data tersebut.

Pemilihan periode sebelum pandemi sebagai *Data Training* dipertimbangkan secara khusus dengan tujuan agar model dapat membangun pemahaman yang kuat dan stabil dari

pola historis tanpa gangguan yang disebabkan oleh perubahan drastis selama pandemi. Sebaliknya, penggunaan data setelah tahun 2020 sebagai *Data Testing* dimaksudkan untuk menguji kemampuan generalisasi model dalam menghadapi situasi baru dan pola-pola yang berubah akibat dampak pandemi COVID-19 serta tren dinamis dalam minat pencarian daring. Dengan metode splitting ini, model SARIMAX-BiLSTM dapat dilatih menggunakan data yang representatif dan konsisten, sekaligus divalidasi pada data yang mencerminkan kondisi nyata terkini yang penuh tantangan. Pendekatan ini diharapkan mampu meningkatkan keandalan serta ketepatan prediksi kedatangan wisatawan dan tren minat pencarian daring yang menjadi fokus utama penelitian ini.

**Tabel 3.14 Data Training**

| date       | arrival_diff_scaled | google_trends_lag1_scaled |
|------------|---------------------|---------------------------|
| 2009-01-01 | 0.6385167357        | 0                         |
| 2009-02-01 | 0.5362265087        | 0.4137646816              |
| 2009-03-01 | 0.7166570622        | 0.4502369668              |
| 2009-04-01 | 0.7169238698        | 0.4759942304              |
| 2009-05-01 | 0.6462427218        | 0.5495569751              |
| ...        | ...                 |                           |
| 2019-12-01 | 0.8461616657        | 0.7267669483              |

**Tabel 3.15 Data Testing**

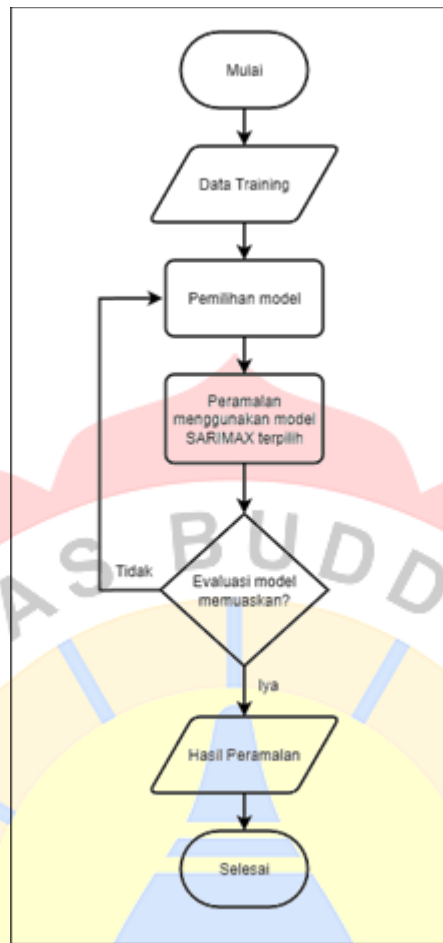
| date       | arrival_diff_scaled | google_trends_lag1_scaled |
|------------|---------------------|---------------------------|
| 2009-01-01 | 0.6385167357        | 0                         |
| 2009-02-01 | 0.5362265087        | 0.4137646816              |
| 2009-03-01 | 0.7166570622        | 0.4502369668              |
| 2009-04-01 | 0.7169238698        | 0.4759942304              |
| 2009-05-01 | 0.6462427218        | 0.5495569751              |
| ...        | ...                 |                           |

|            |              |              |
|------------|--------------|--------------|
| 2019-12-01 | 0.8461616657 | 0.7267669483 |
|------------|--------------|--------------|

### 3.6 *Model Training*

#### 3.6.1 *Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX)*

Gambar 3.4 menggambarkan alur penelitian secara menyeluruh yang dimulai dari tahap Data Preprocessing hingga proses pemilihan model yang akan digunakan untuk prediksi kedatangan wisatawan. Tahapan Data Preprocessing merupakan bagian krusial dalam penelitian ini karena mencakup berbagai langkah penting yang bertujuan untuk memastikan bahwa data yang digunakan telah memenuhi standar kualitas dan siap untuk dimasukkan ke dalam proses pemodelan. Proses ini meliputi pembersihan data (*Cleaning*) yang bertujuan untuk mengatasi nilai kosong (*missing values*) serta mendeteksi dan menangani outlier yang dapat mengganggu keakuratan model. Selain itu, transformasi data dilakukan untuk menyesuaikan format dan struktur data sesuai kebutuhan analisis.



**Gambar 3.4 Flowchart SARIMAX**

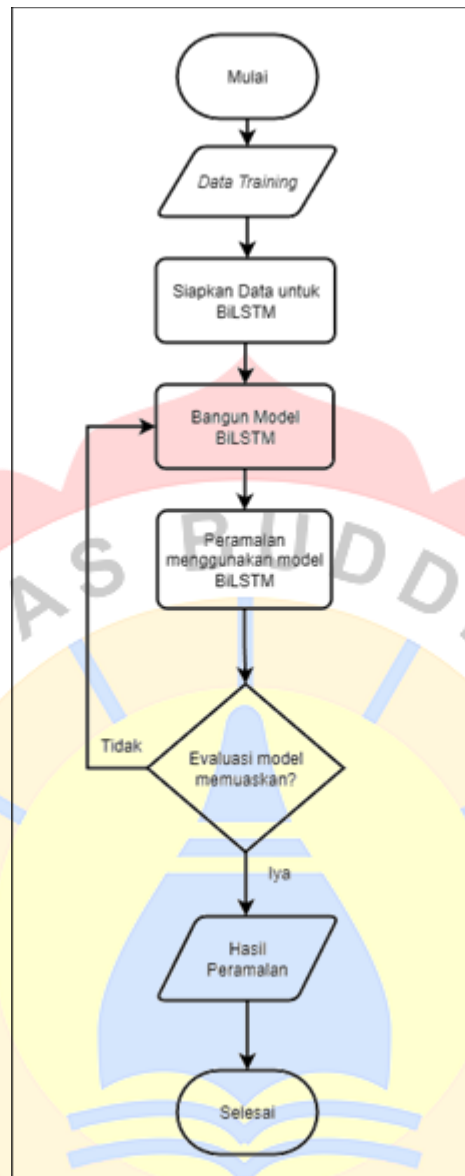
Selanjutnya, dilakukan uji stasioneritas guna memastikan bahwa data memiliki sifat yang stabil dari waktu ke waktu, yang merupakan syarat penting dalam pemodelan deret waktu. Setelah itu, normalisasi diterapkan untuk menyelaraskan skala variabel sehingga model dapat berfungsi lebih optimal tanpa bias yang disebabkan oleh perbedaan skala data. Tahap berikutnya adalah analisis *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF), yang berfungsi untuk membantu dalam menentukan parameter awal model yang paling sesuai, sehingga dapat menangkap pola data dengan tepat.

Hasil dari seluruh proses pre-processing ini menunjukkan bahwa data kedatangan wisatawan memerlukan *differencing* sebanyak satu kali agar menjadi stasioner. Analisis ACF dan PACF menghasilkan parameter model dengan nilai  $p=1$ ,  $d=1$ , dan  $q=1$ , yang

menjadi dasar dalam pembentukan model SARIMAX. Dengan parameter tersebut, model SARIMAX dapat secara efektif memodelkan pola linear, pola musiman, serta pengaruh variabel *eksogen* yang telah dimasukkan ke dalam model. Setelah parameter ditentukan, proses pelatihan model SARIMAX dilakukan menggunakan data yang telah melalui tahap preprocessing, menghasilkan model yang mampu merepresentasikan tren *linear* dan pola musiman dengan baik. Prediksi yang dihasilkan dari model ini akurat dan andal, mencerminkan pola historis serta pengaruh variabel *eksogen* secara optimal, sehingga hasil prediksi dapat digunakan sebagai dasar pengambilan keputusan dan analisis lebih lanjut dalam konteks pengelolaan pariwisata Bali.

### **3.6.2 Bidirectional Long Short-Term Memory (BiLSTM)**

Alur proses pembangunan dan prediksi menggunakan model BiLSTM (*Bidirectional Long Short-Term Memory*) dirancang khusus untuk menangkap pola temporal yang kompleks pada data deret waktu (*Time Series*), seperti yang tergambar pada Gambar 3.5. Proses ini diawali dengan data yang telah melewati tahapan preprocessing yang meliputi pembersihan data dari nilai kosong dan anomali, transformasi untuk menyesuaikan format data, normalisasi untuk menyamakan skala variabel, serta analisis lag dan uji stasioneritas guna memastikan kestabilan data. Setelah data siap, langkah berikutnya adalah memformat data menjadi bentuk tiga dimensi [*samples, timesteps, features*], sehingga model BiLSTM dapat memahami hubungan temporal dan ketergantungan antar titik data secara optimal.



**Gambar 3.5 Flowchart BiLSTM**

Model BiLSTM kemudian dibangun dengan menggabungkan beberapa *layer* BiLSTM yang berfungsi untuk menangkap pola temporal secara dua arah—baik dari masa lalu ke masa depan maupun sebaliknya. Penambahan *layer* Dense di bagian akhir model bertujuan untuk mengolah informasi hasil ekstraksi pola temporal menjadi prediksi akhir yang dapat digunakan untuk analisis. Struktur ini memungkinkan model menangani ketergantungan jangka panjang dalam data sekaligus mengenali pola-pola kompleks yang mungkin tersembunyi dalam hubungan antar variabel dan waktu.

Setelah proses pembangunan model selesai, dilakukan tahap prediksi dengan menggunakan model BiLSTM yang telah dilatih menggunakan data historis. Model ini memanfaatkan pola temporal dan informasi kompleks yang telah dipelajari selama pelatihan untuk menghasilkan prediksi yang akurat dan dapat diandalkan. Hasil prediksi tersebut kemudian digunakan dalam proses pengambilan keputusan atau analisis lanjutan, memberikan dasar yang kuat untuk memahami tren kedatangan wisatawan berdasarkan data historis dan pola temporal yang telah dimodelkan dengan baik.

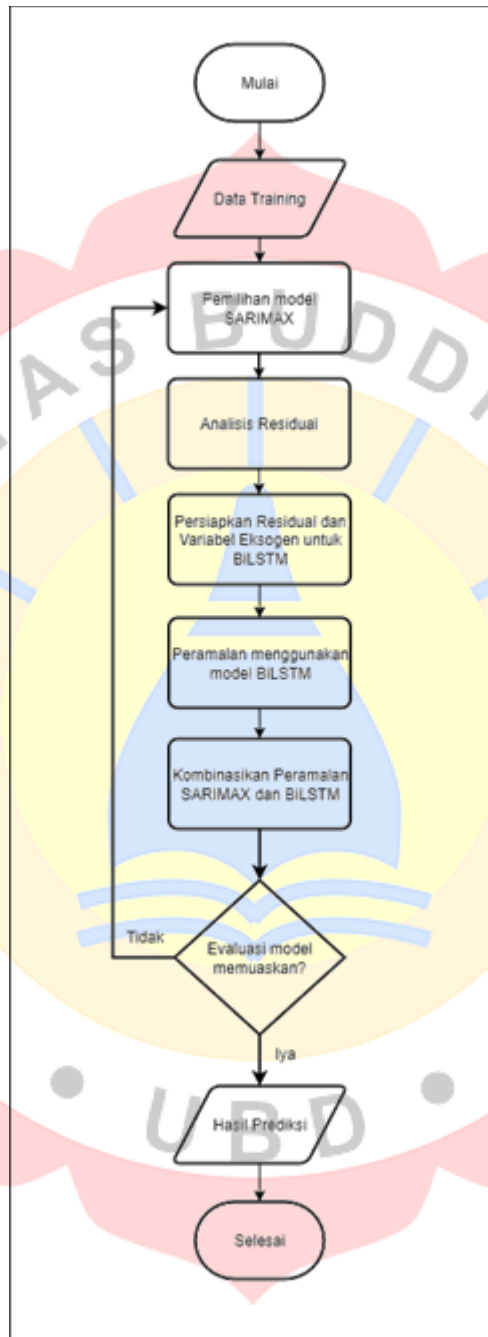
### 3.6.3 SARIMAX-BILSTM

Gambar 3.6 menjelaskan alur proses pemodelan menggunakan pendekatan *hybrid* SARIMAX-BiLSTM yang mengintegrasikan keunggulan kedua model untuk memproses data deret waktu (*Time Series*). Model SARIMAX bertugas menangkap pola *linear* dan musiman yang ada dalam data, sementara model BiLSTM berfungsi untuk mengenali dan memodelkan pola *non-linear* yang mungkin tidak dapat dijelaskan oleh SARIMAX. Proses dimulai dari tahap Data Preprocessing, yang mencakup beberapa langkah penting seperti pembersihan data dari nilai hilang dan outlier, transformasi data agar sesuai dengan kebutuhan analisis, normalisasi untuk menyamakan skala variabel, serta analisis *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) guna menentukan parameter awal model yang optimal.

Parameter model SARIMAX, yaitu  $p$ ,  $d$ , dan  $q$ , ditentukan berdasarkan hasil uji *differencing* dan analisis ACF/PACF pada data, sementara parameter musiman  $P$ ,  $D$ ,  $Q$ , dan  $s$  disesuaikan berdasarkan pola musiman yang teridentifikasi dari data. Setelah parameter ini ditetapkan, model SARIMAX dilatih menggunakan data target untuk menghasilkan prediksi awal. Selain prediksi SARIMAX, model ini juga menghasilkan residual, yaitu selisih antara nilai aktual dengan nilai prediksi SARIMAX. Residual ini kemudian dianalisis untuk memastikan bahwa pola yang tersisa adalah white noise, atau



tanpa pola signifikan. Namun, jika residual masih menunjukkan adanya pola *non-linear* yang tersisa, maka tahap selanjutnya adalah menggunakan model BiLSTM untuk menangkap pola *non-linear* tersebut.



**Gambar 3.6 Flowchart SARIMX-BiLSTM**

Pada tahap penggunaan BiLSTM, residual hasil prediksi SARIMAX digunakan sebagai variabel target yang akan dipelajari, sedangkan variabel *eksogen* dengan lag optimal dipakai sebagai *input* model. Data disusun dalam format tiga dimensi [samples,

timesteps, features] agar model BiLSTM dapat memahami dengan baik hubungan temporal dan kompleksitas pola dalam data residual. Setelah model BiLSTM dilatih, diperoleh prediksi hasil pembelajaran pola *non-linear* tersebut. Prediksi akhir dari pendekatan *hybrid* ini diperoleh dengan menggabungkan prediksi SARIMAX dan BiLSTM sesuai dengan metode yang dijelaskan dalam persamaan 3 pada studi (He et al., 2021)Click or tap here to enter text., sehingga menghasilkan model prediksi yang lebih akurat dengan mengintegrasikan kelebihan masing-masing model.

$$\text{Hasil Prediksi} = \text{Prediksi SARIMAX} + \text{Prediksi BiLSTM} \quad (3)$$

### 3.7 Metode Evaluasi

Dalam evaluasi model, tiga pendekatan—SARIMAX, BiLSTM, dan SARIMAX-BiLSTM—diukur kinerjanya menggunakan metrik MAE (*Mean Absolute Error*), MAPE (*Mean Absolute Percentage Error*), dan RMSE (*Root Mean Squared Error*). Penggunaan ketiga metrik ini memberikan pendekatan yang holistik untuk mengukur akurasi prediksi. MAPE dipilih karena sifatnya yang sederhana dan kemampuannya me

mberikan interpretasi yang jelas jika semua data bernilai positif dan lebih besar dari nol. MAE digunakan karena memberikan rata-rata kesalahan absolut yang mudah dipahami, terutama ketika data memiliki skala yang sama. Sebagai pembanding, RMSE sering digunakan dalam pemodelan statistik karena skalanya yang konsisten dengan data asli dan memberikan bobot lebih besar pada kesalahan besar (Prilistya et al., 2021). Evaluasi ini bertujuan untuk mengidentifikasi model dengan performa terbaik, di mana SARIMAX-BiLSTM diharapkan mampu memberikan prediksi yang paling akurat dengan mengintegrasikan kekuatan SARIMAX dalam menangkap pola linier dan musiman serta BiLSTM dalam menangkap pola non-linear.

Nilai yang lebih kecil dari ketiga metrik evaluasi ini mencerminkan kinerja model yang lebih baik. Ketiga metrik dihitung menggunakan rumus berikut:

### 3.7.1 Mean Absolute Error (MAE)

MAE mengukur rata-rata kesalahan absolut antara nilai aktual ( $y_i$ ) dan nilai prediksi ( $\hat{y}_i$ ) tanpa mempertimbangkan arah kesalahan, seperti yang ditunjukkan pada Persamaan (4):

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (4)$$

Di mana  $n$  adalah jumlah data.

### 3.7.2 Mean Absolute Percentage Error (MAPE)

MAPE memberikan kesalahan rata-rata dalam bentuk persentase relatif terhadap nilai aktual, seperti yang ditunjukkan pada Persamaan (5):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (5)$$

MAPE berguna untuk interpretasi yang jelas jika semua nilai aktual ( $y_i$ ) bernilai positif dan tidak nol.

### 3.7.3 Root Mean Squared Error (RMSE)

RMSE mengukur akar rata-rata kesalahan kuadrat, memberikan bobot lebih besar pada kesalahan besar, seperti yang ditunjukkan pada Persamaan (6):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

RMSE sangat berguna untuk mengukur kesalahan dalam satuan yang sama dengan data asli.

Ketiga metrik ini memberikan pandangan yang komprehensif terhadap kinerja model, dengan MAE dan MAPE memberikan gambaran rata-rata kesalahan, sedangkan RMSE lebih sensitif terhadap kesalahan besar.